Particle Clustering and Sub-clustering as a Proxy for Mixing in Geophysical Flows

by

Rishiraj Chakraborty

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics

Waterloo, Ontario, Canada, 2019

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The Eulerian point of view is the traditional theoretical and numerical tool to describe fluid mechanics. Some modern computational fluid dynamics codes allow for the efficient simulation of particles, in turn facilitating a Lagrangian description of the flow. The existence and persistence of Lagrangian coherent structures in fluid flow has been a topic of considerable study. Here we focus on the ability of Lagrangian methods to characterize mixing in geophysical flows. We study the instability of a strongly non-linear double jet flow, initially in geostrophic balance, which forms quasi-coherent vortices when subjected to ageostrophic perturbations. Particle clustering techniques are applied to study the behaviour of the particles in the vicinity of coherent vortices. Changes in inter–particle distance play a key role in establishing the patterns in particle trajectories. This thesis exploits graph theory in finding particle clusters and regions of dense interactions (also known as sub-clusters). The methods discussed and results presented in this thesis can be used to identify mixing in a flow and extract information about particle behaviour in coherent structures from a Lagrangian point of view.

Acknowledgements

Dedication

This work is dedicated to nature.

Table of Contents

List of Figures

Chapter 1

Introduction

Geophysical fluid dynamics is the branch of fluid dynamics that deals with naturally occurring fluids (the Earth's interior, the oceans and the atmosphere) and their physical processes (e.g. lava flows, hurricanes, ocean eddies etc). The two main ingredients that make geophysical flows different from traditional fluid dynamics are rotation and stratification. In general fluid motions are described by certain field variables like velocity, pressure, density, temperature, salinity, magnetic field etc. In order to describe these fluid motions, we need to resort to the classical laws of mechanics viz. conservation of mass, conservation of momentum and conservation of energy. Conservation of mass ensures that the amount of fluid entering a fixed three dimensional frame in space is equal to the sum of the amount of fluid that leaves the frame and the amount that gets stored in it per unit time. For fluid dynamics, we are interested in conservation of linear momentum, which is just the balance between acceleration of fluid parcels related and the external forces acting on them. This is known as the famous Navier-Stokes equations which is Newton's second law of motion for fluids. We also have the law of conservation of energy stating that for a given fluid the sum of its kinetic, potential and internal energy is constant. In geophysical flows, due to the rotation of the Earth about its axis, there are two rotational acceleration components present in the Navier-Stokes equations which are Coriolis acceleration and the centrifugal acceleration. The centrifugal acceleration doesn't play a crucial role in geophysical flows however surprising it may be, unlike the Coriolis acceleration which delivers a certain vertical rigidity such that the large scale flows maintain columnar motion. Stratification is another frequent component present in geophysical flows because naturally occurring flows consist of fluids having different densities (e.g. warm water overlying cold water in a lake in early summer). The gravitational force has a role to play here, colloquially pushing the lighter fluid over the denser fluid, leading to the typical situation of a stratification in
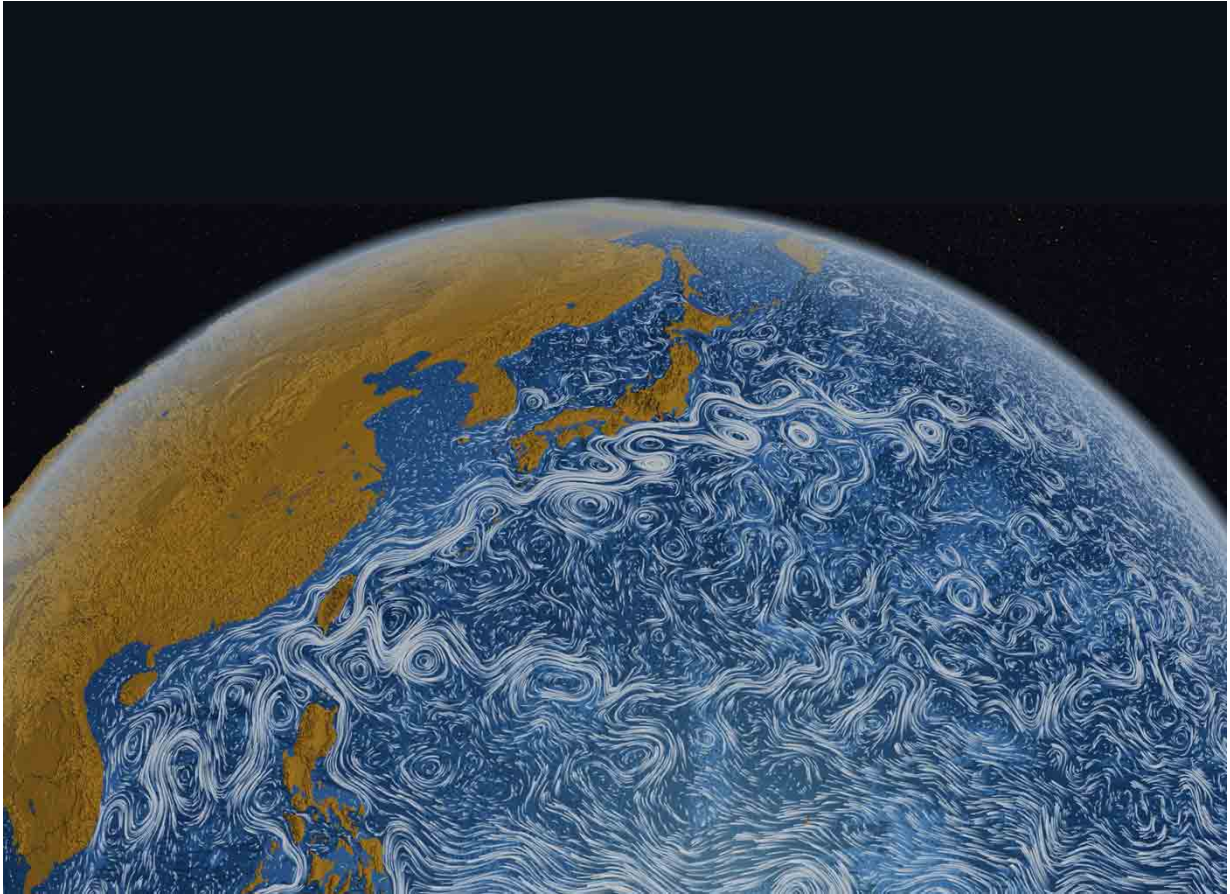
Figure 1.1: Eddies in Kuroshio Current resolved in ocean simulations (https://researchfeatures.com/2017/10/20/mesoscale-ocean-currents-improved-climate-modelling/)
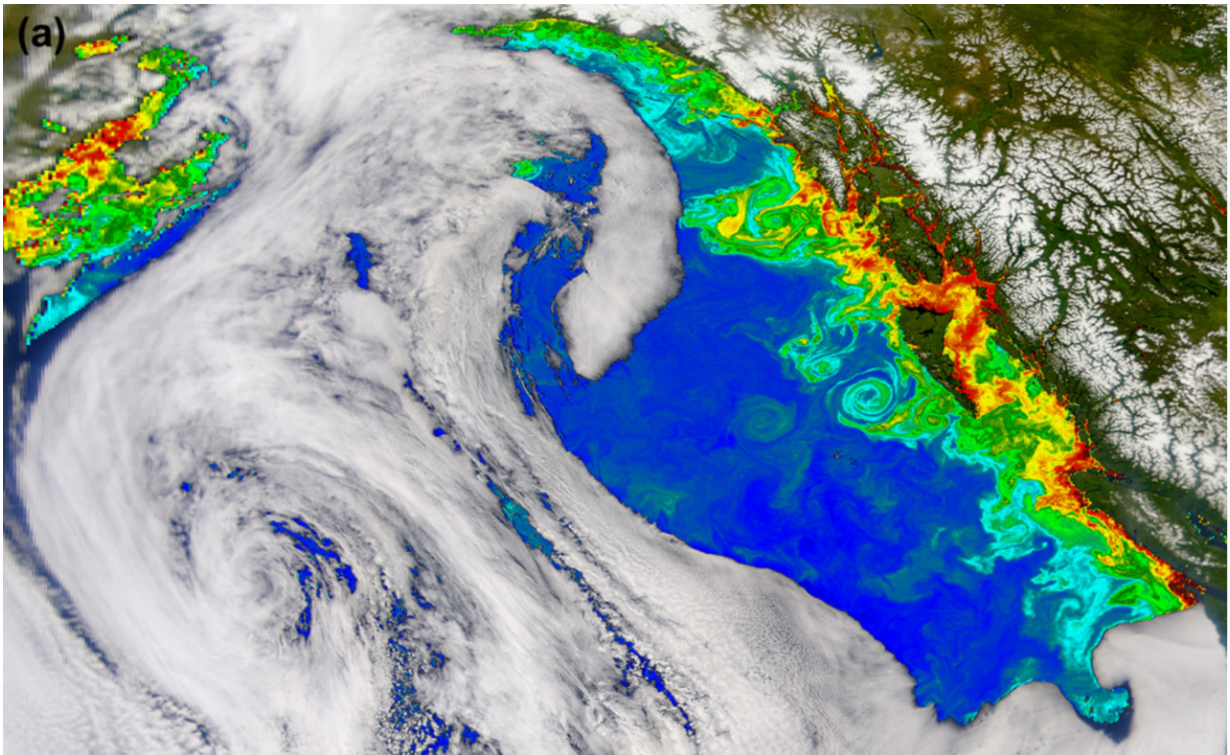
Figure 1.2: Haida eddies from NASA Satellite image. (https://en.wikipedia.org/wiki/Haida_Eddies)

stable equilibrium. However fluid motions or external forces like winds tend to perturb this equilibrium to a varying extent, giving rise to interesting phenomenon like instabilities, mixing, and waves of various type. The relative effects of various terms, e.g.rotation, in the fluid dynamics is dependant on the different scalings or dimensionless numbers, involved. For example in the ocean the horizontal length scales can stretch up to hundreds or thousands of kilometres, while the vertical length scales are only limited to less than about ten kilometres. This allows us to assume that, as a first approximation, the fluid moves in columns or put simply the horizontal velocities are independent of the vertical position or depth. The horizontal motions dominate as the horizontal velocities are much larger than the vertical velocities.The effects of rotation are important, when the local horizontal velocity scale is smaller as compared to the velocity scale due to the Coriolis acceleration. A "traditional approximation" in geophysical fluid dynamics is to neglect the local horizontal component of the Coriolis acceleration. This "traditional approximation" is reliable when the vertical length scales are very small compared to the horizontal length scales and the equations of motion can be derived using a shallow layer approximation also known as the shallow water equations. This approximation becomes questionable for motions with smaller horizontal length scales and for waves with frequencies comparable to the Coriolis frequency. This approximation also breaks down near the equator.

We look at a few examples of geophysical flows in this section. **Fig.(1.1)** and **Fig.(1.2)** show the eddies formed in oceans due to the rotation of the earth.

Fig.(1.3) shows the Labrador coast where sea ice mixes with warmer water. We can spot the formation of eddies due to the rotation of the earth as land fast ice breaks up and mixes with warm water.

Fig.(1.4) shows the problem of water contamination from agricultural sources. Here as well we see the formation of eddies due to the turbulent nature of the water sources. We see that formation of eddies is ubiquitous in large scale flows. In this thesis we introduce a Lagrangian technique of find regions of strong mixing (in the sense of diffusion) so that we can be precisely aware of the situation of large-scale water bodies in scenarios like mentioned above.

1.1 Governing Equations

Conservation of mass per unit volume for compressible flows leads to the following equation [16],

Figure 1.3: Sea Ice off the Labrador coast (https://www.canada.ca/en/environment-climate-change/services/ice-forecasts-observations)

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0 \tag{1.1}$$

For incompressible flows, which we are interested in in this work, 1.1 becomes,

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0 \tag{1.2}$$

and we note that flows in which density ρ is constant incompressibility immediately follows.

A Newtonian fluid is one, whose stress arising from viscous forces is linearly proportional to the local strain rate. For a Newtonian fluid exhibiting incompressible flows, the Navier-Stokes equation of conservation of linear momentum reads,

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\boldsymbol{\nabla} p + \rho \boldsymbol{g} + \mu \nabla^2 \boldsymbol{u} \tag{1.3}$$

5

Figure 1.4: Problem of water contamination from agricultural sources (https://www.cleantechloops.com/pollution-oceans/)

where μ is the coefficient of viscosity (i.e. the shear viscosity).

If we consider the fluid to be inviscid, equation 1.3 becomes

$$\rho\frac{D\boldsymbol{u}}{Dt} = -\boldsymbol{\nabla}p + \rho\boldsymbol{g} \tag{1.4}$$

which, along with the conservation of mass is known as the set of Euler equations.

1.2   Shallow Water Equations

The first assumption in the derivation of the shallow water equations is that fluid moves in columns. This means the velocity can be taken to be constant along the vertical coordinate. This allows us to integrate the conservation of mass 1.2 from the bottom (which for simplicity we take to be flat and at $x = -H$) to the surface $z = \eta(x,t)$,

$$\int_{-H}^{\eta}(u_x + v_y)dz + w(\eta) - w(-H) = 0.$$

Since there is no flow through the bottom $w(-H) = 0$ and from the kinematic boundary condition, which says that particles on the surface stay on the surface, we find

$$\int_{-H}^{\eta} (u_x + v_y)dz + \frac{D\eta}{Dt} = 0.$$

This can be further simplified by assuming that the horizontal velocity is independent of depth (z). This allows us to take the u_x out of the integral and find

$$\frac{D\eta}{Dt} + (H + \eta)(u_x + v_y) = 0.$$

Finally this can be rewritten to give the conservation of mass

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x}[(H + \eta)u] + \frac{\partial}{\partial y}[(H + \eta)v] = 0. \tag{1.5}$$

Since the motion is in the form of vertical columns, we drop all terms except the pressure term in the vertical momentum equation, which then becomes,

$$\frac{\partial p}{\partial z} = -\rho_0 g. \tag{1.6}$$

which yields pressure at any arbitrary height z^* from the bottom as,

$$p(z^*) = \rho_0 g \eta - \rho_0 g z^* - p_{atm}. \tag{1.7}$$

Letting

$$\nabla_H = (\partial_x, \partial_y)$$

we get

$$-\nabla_H p = -\rho_0 g \nabla_H \eta.$$

Thus for the horizontal momentum equations we have,

$$\rho_0 \frac{D(u, v)}{Dt} = -\nabla_H p \tag{1.8}$$

Using 1.7, 1.8 becomes,

$$\frac{D(u, v)}{Dt} = -g \nabla_H \eta. \tag{1.9}$$

7

1.3 Rotating Shallow Water Equations

In ocean and atmospheric flows, the frame of reference from which we observe the flows is a rotating reference frame because of the fact that the Earth rotates about its axis. Hence we need to incorporate the components of the Coriolis acceleration in the shallow water momentum equations. As specified earlier, the centrifugal components of acceleration are not important. We use local Cartesian coordinates such that the vertical coordinate is aligned with the local normal. If we represent Earth's angular velocity as Ω, we are only interested in the local vertical component of the ω, i.e. $\Omega \sin(\theta)$, where θ is the latitude, also known as the f-plane approximation as shown in Fig.(1.5). This let's us take the Coriolis parameter as $f = 2\Omega \sin(\theta)$ and hence to write the rotating shallow water equations as follows,

$$\frac{D\vec{u}}{Dt} + (-fv, fu) = -g\nabla_H \eta$$
$$\frac{\partial \eta}{\partial t} + \nabla \cdot [(H + \eta)\vec{u}] = 0. \tag{1.10}$$

1.4 Non-dimensionalising the Equations

Considering the momentum equations in 1.10, let $\vec{u} = U\tilde{u}$, $\eta = He$, $(x, y) = L(\tilde{x}, \tilde{y})$ and choose the advective time scale $T = L/U$, where L the horizontal length scale is much larger than the vertical length scale H.

The momentum equations thus read,

$$\frac{U^2}{L}\frac{D\tilde{u}}{Dt} + fU(-\tilde{v}, \tilde{u}) = -g\frac{H}{L}\tilde{\nabla}e$$

rearranged to read

$$\frac{U}{fL}\frac{D\tilde{u}}{Dt} + (-\tilde{v}, \tilde{u}) = -g\frac{H}{LfU}\tilde{\nabla}e. \tag{1.11}$$

The dimensionless number $\frac{U}{fL}$ is called the Rossby number Ro, responsible for measuring the effects of rotation. With U and L fixed, the higher the value of f, the lower the Rossby number Ro. For large scale motions, typically $Ro << 1$ and the inertial and the acceleration terms are negligible. An important classical example of balance between the left and the

Rotation frequency about axis is 2Ω

$(0, 2\Omega\cos\theta, 0)$

$(0, 0, 2\Omega\sin\theta)$

In local Cartesian coordinates the rotation
vector has two nonzero components: $(0, f^*, f)$
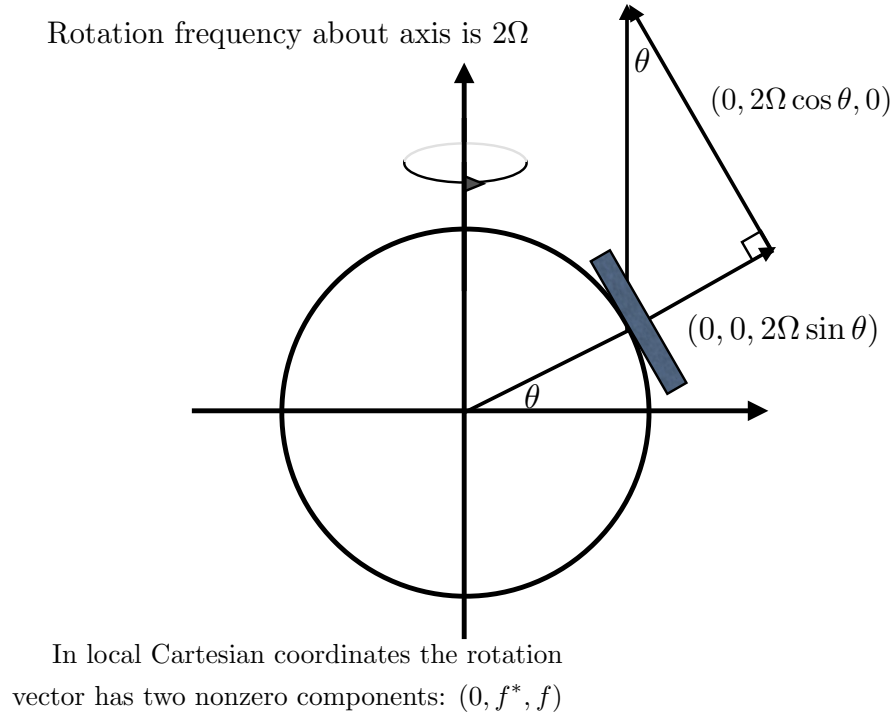
Figure 1.5: Figure showing the f-plane approximation for the rotating shallow water equations (M. Stastna *used with permission*)

.

right hand side terms in 1.11, is the geostrophic balance where the pressure gradient term is balanced by the Coriolis acceleration. In such a case, $g\frac{H}{fLU} = 1$ and the approximate equations dropping the tildes read as,

$$(-v, u) = -\nabla e \tag{1.12}$$

A more general way of balancing the equations is by introducing the Froude number $Fr = \frac{U}{c_0}$, where $c_0 = \sqrt{gh}$ is the non-rotating shallow water wave speed. This is used to rewrite the non-dimensional equations as follows,

$$Ro\frac{D\vec{u}}{Dt} + (-v, u) = -\frac{Ro}{Fr^2}\nabla e \tag{1.13}$$

9

1.5 Description of the Eulerian vs. the Lagrangian approach

The first step of looking at a fluid kinematics problem is to observe the changes in fluid properties like density, pressure, velocity, temperature, salinity, bacteria content etc. all over the flow contained within a specified two or three dimensional domain. Now there are two ways of doing this, either by tracking specific identifiable fluid material volumes moving with the flow, the Lagrangian method, or by observing the fluid properties from a fixed point in space usually from within a control volume, the Eulerian method, as shown in **Fig.**1.6. The arrows can be thought of as the velocity fields at these points. The rectangular shaped control volume is fixed in space and observes the change in the properties at that spatial locations as fluid parcels move through it with time. The small circles on the other hand, thought of as fluid particles, are made up of fluid parcels and have their observers sitting on top of them. Each observer sees the changes in properties of its own particle as it traverses through space with time. It's like a weather balloon measuring changes in temperature. Basically an Eulerian approach uses a control volume and a Lagrangian approach uses a material volume to describe the properties of a flow. Both control volumes and material volumes can be made as small as possible to resolve a flow effectively. Almost all of the theories in fluid mechanics have been developed using the Eulerian approach. However, the Lagrangian approach is useful for observational and measuring techniques.

It is often hard to find the path of a particle analytically. Usually we solve pdes for the Eulerian fields for e.g. the velocity field $\vec{u}(\vec{x}, t)$. To trace the path of a particle, we will need to solve the pathline equation, which is simply,

$$\frac{d\vec{x}}{dt} = \vec{u}(\vec{x}(t), t), \quad \vec{x}(0) = \vec{a} \tag{1.14}$$

Since the Lagrangian concept works with material volumes, in order to write Lagrangian equations using the Eulerian field variables, we need to refer to the material derivative. Let's look at the material derivative of $\vec{u}(\vec{x}, t)$,

$$\frac{D\vec{u}}{Dt} = \frac{\partial \vec{u}}{\partial t} + \vec{u} \cdot \nabla \vec{u}. \tag{1.15}$$

The left hand side of equation 1.15 represents the acceleration of a particle moving with velocity $\vec{u}(\vec{x}, t)$ in terms of the Eulerian acceleration ($\frac{\partial \vec{u}}{\partial t}$) and the convective acceleration ($\vec{u} \cdot \nabla \vec{u}$).
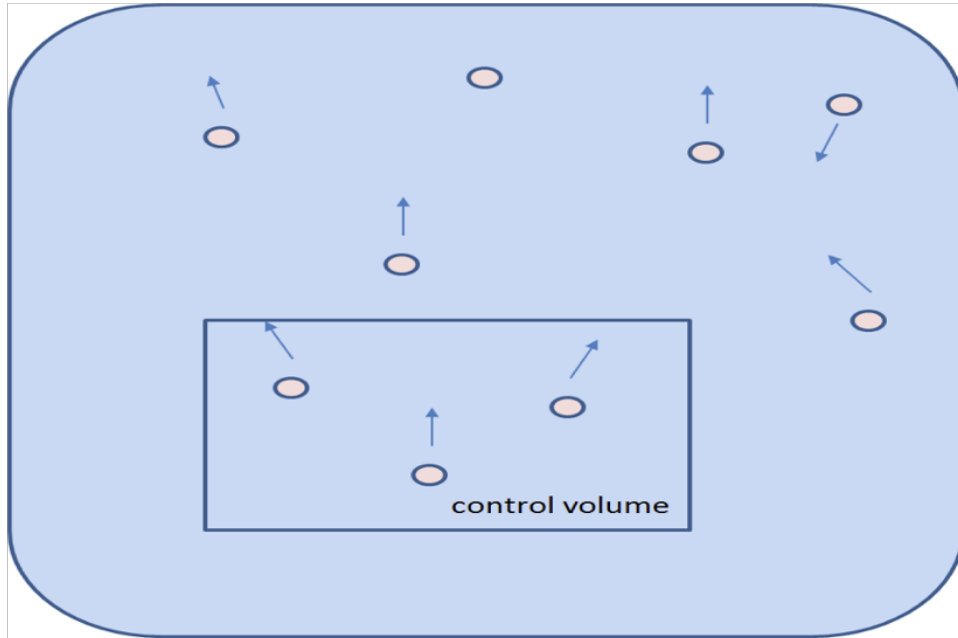
Figure 1.6: The rectangular box shows an invisible imaginary control volume fixed in space for observing the fluid properties at a particular time within it (Eulerian approach). The small circles, called particles represent material volumes made of fluid parcels with observers on the particles (Lagrangian approach).

1.6 Structure of remainder of the thesis

This thesis uses the Eulerian approach to simulate a double-jet flow, initially in geostrophic balance, eventually forming vortices, a common phenomenon in oceans. However, the more important part of this work rather uses the Lagrangian approach not for measuring purposes, but to build particle trajectories over the course of the flow. Our aim is to use graph theoretic methods and leverage this trajectory information to gain insights on localized mixing happening in the process. Chapter 2 begins by surveying the existing literature and building up the motivation for the techniques we use in the remaining sections of the chapter. In the following section of the chapter we describe the methods we implemented (i.e. dense subclustering) and the ones we used for comparison (i.e. spectral clustering). Next we present a section on the detailed results and their analysis, in which we show what we found using the subclustering technique, discuss its properties and details of its usage and compare its results to some approaches already used in existing literature. In the final
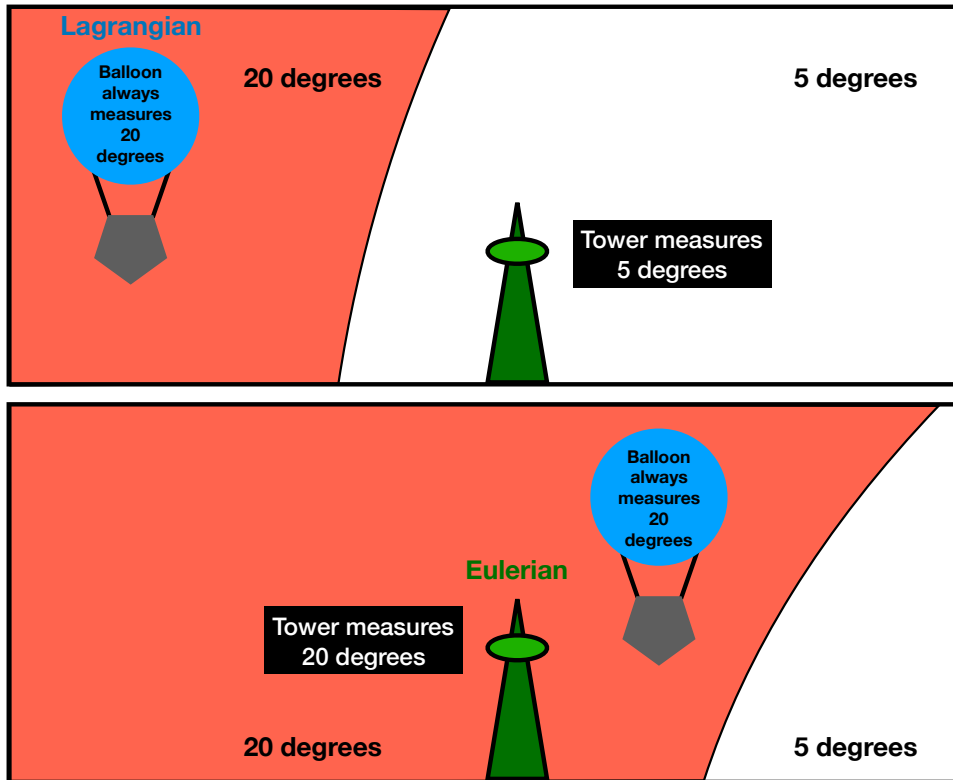
Figure 1.7: Figure showing the differences between the Lagrangian and the Eulerian techniques (M. Stastna *used with permission*)

.

chapter of the thesis, we summarize the work that we have done in context of the previous literature. We also discuss in details the areas where our technique has to be worked on in the future, to make it more widely implementable and cost efficient.

Chapter 2

Graph–based Methods For Identifying Mixing During Instability

2.1 Introduction and Literature Survey

A version of this chapter has been accepted by the *Nonlinear Processes in Geophysics* journal for publication [3]. R Chakraborty carried out majority of the work including software development and writing, M. Stastna carried out a close supervision and research role while A. Coutino contributed to the software development of the used numerical methods.

There are two different geometric approaches to fluid mechanics, the Eulerian and the Lagrangian approach. In the Eulerian approach, field values are obtained on a spatial grid, for example from numerical simulation output. In the Lagrangian approach measurement data is obtained following the fluid, as in the case of temperature measurements by a weather balloon. Many naturally occurring flows are complex, three–dimensional and at least to some extent, turbulent. Such flows are characterized by a richness of vorticity and the rapid mixing of passive tracers as discussed in [5], chapter 3. At the same time, satellite imagery suggests large scale flows exhibit prominent coherent patterns, and this is theoretically supported by the so-called inverse cascade of two dimensional turbulence in which energy moves to larger scales while enstrophy moves to smaller scales [5], chapter 10.

Even three dimensional turbulent flows are known to contain quasi-deterministic coherent structures [13]. Coherent structures can be thought of as turbulent fluid masses having temporal correlation in vorticity over some spatial extent (e.g. a shear layer in a flow).
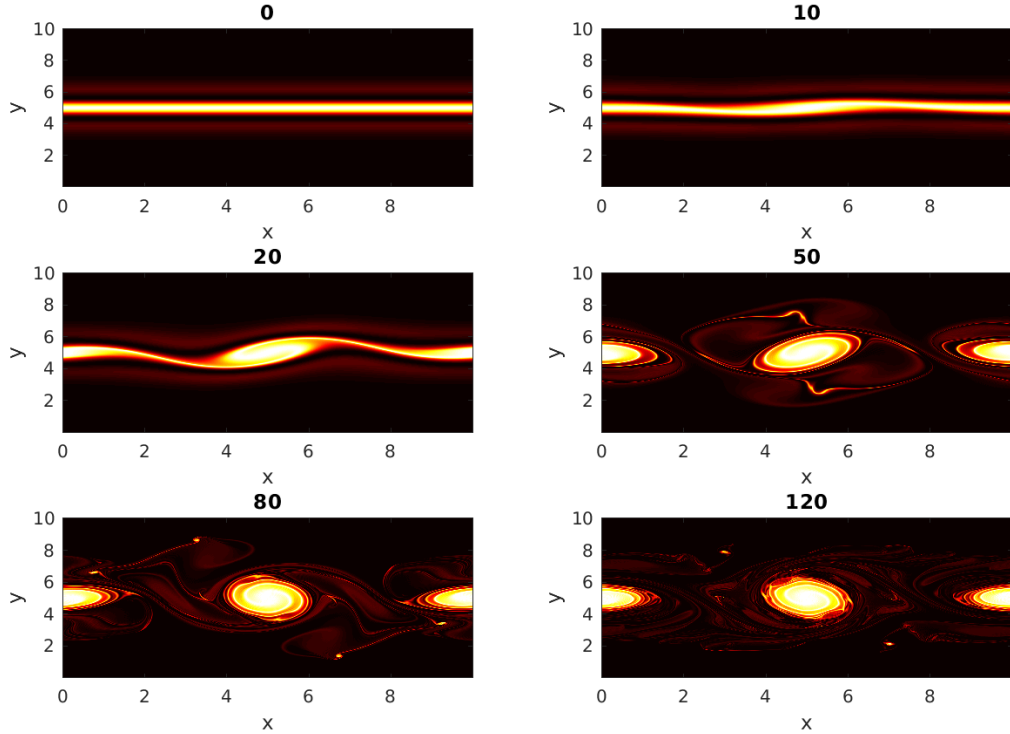
Figure 2.1: The enstrophy field showing the evolution of the unstable double jet with time. The bright areas indicate regions of high enstrophy which are found between the two jets at early times.

Fig.(2.1) shows the evolution of the enstrophy field of a two dimensional double jet initially in geostrophic balance, subjected to ageostrophic perturbations. The evolution depicts the formation of vortices due to instability of the geostrophic flow. Coherent structures like vortices and filaments, undergo frequent stretching and folding. The identification of coherent structures in turbulent flows gave the revolutionary notion in fluid mechanics that turbulent flows are not completely random but can contain orderly organized structures and these coherent structures in specific regions can influence mixing, transport and other physically relevant features [15].

The study of coherent flow structures has received significant interest in the recent past. The existing methods for detecting coherent behaviour mathematically are either geometric or probabilistic; [1] discusses and compares the different methods. Geometric

14

methods aim to find distinct boundaries between the coherent structures, whereas probabilistic methods use the concept of sets with minimal dispersion moving in a flow to identify coherent structures. [26] in their Introduction, however, note that existing methods for finding coherent structures require the full knowledge of the flow-field and the underlying dynamical system. This, in turn, requires high resolution trajectory data. This can be numerically expensive, as well as challenging to find in applications. [11], in their review of various Lagrangian techniques for finding coherent structures, say that the Lagrangian diagnostic scalar field methods are incapable of providing a strict definition of coherent flow structures and are also not effective in establishing a precise mathematical connection between the geometric features and the flow structures. Such diagnostic methods include: Finite time Lyapunov exponents (FTLE), Finite-Size Lyapunov Exponent (FSLE), Mesochronic analysis, Trajectory length, Trajectory complexity and Shape coherence. [11] also describes the various methods of applying mathematical coherence principles to locate coherent structures. However, these principles only apply for finite time intervals from the beginning of the flow evolution, it is not guaranteed that the coherence principles comply with observed coherent patterns at later times. Examples of mathematical coherence principles include transfer operator methods like the probabilistic transfer operator [7] and the dynamic Laplace operator [8]. These methods identify maximally coherent or minimally dispersive (not dispersive in the sense of wave theory) regions over a finite time interval. Such regions are expected to minimally mix with the surrounding phase space and are named "almost-invariant sets" for autonomous systems and "coherent sets" for non-autonomous systems. A different mathematical approach is the hierarchical coherent pairs method [10], which initially splits a given domain into a pair of coherent sets using the transfer operator method, and then subsequently refines the coherent sets iteratively. This is accomplished using the probabilistic transfer operator. The iteration is carried out until a reference measure of the probability, μ, falls below a user defined cut-off. A third category of mathematical approaches for finding coherent structures based on Lagrangian data is clustering. [11] reviews the Fuzzy C-means clustering of trajectories by [9] which uses the traditional fuzzy C-means clustering to identify finite-time coherent structures and mixing in a flow. This method uses trajectories of Lagrangian particles, over discrete time-intervals, and applies the Fuzzy C-means algorithm to locate coherent sets as clusters of trajectories according to the dynamic distances between trajectories. Another similar method for locating coherent structures is the spectral clustering of trajectories as proposed by [12] and implemented by [26]. [19] discusses algorithms to compute hyperbolic trajectories from data sets on oceanographic flows and how to locate their stable and unstable manifolds. [21] also discusses how phase portraits obtained using Lagrangian descriptors can provide a representation of the interconnected features of the underlying dynamical system. [29] uses a coupled implementation of a mix of Eulerian and Lagrangian models for

simulating the full life cycles of fish species anchovy and sardine in the California Current Systems. The Lagrangian model used is an individual fish based model which tracks each fish of every species. [26] used a generalized graph Laplacian eigenvalue problem to extract coherent sets from several fabricated examples (e.g. Bickley jet) as well as measured data. The authors also highlighted regions of strong mixing in flow, using local network measures like node degree and the local clustering coefficient. These local network measures provide information for each Lagrangian particle.

Inspired by these, we wish to extract regions of dense mixing in flow using a graph theoretic network approach and compare the results with those obtained from spectral clustering. We also wish to use an evolving simulation for which coherent regions evolve dynamically through stretching and folding and are not known *a priori*. [30]'s trajectory encounter volume idea is similar to our methodology, but the volume in which particles are pre-identified is chosen based on features that are assumed to be already present in the flow (i.e. eddies). Moreover, the authors state that the method breaks down for sparse grids since it is dependent on being able to define an effective density of particles. Detailed comparison with our method are thus left to future work.

From an Eulerian point of view, mixing can be characterized by studying the advection-diffusion equation for a passive tracer θ [31],

$$\frac{\partial \theta}{\partial t} + v \cdot \nabla \theta = \kappa \nabla^2 \theta \tag{2.1}$$

where v is the fluid velocity and κ is the diffusion coefficient. Mixing and stirring depends on the gradient of θ and the hence the extent of mixing and stirring in a given domain for a given flow can be measured by the spatial variability index

$$C = \frac{1}{2} \int \int \nabla \theta \cdot \nabla \theta dx. \tag{2.2}$$

Taking the time derivative of C, and following the simplification procedure in [31], we obtain,

$$\frac{dC}{dt} = \int \int [(v.\nabla\theta)\nabla^2\theta - \kappa(\nabla^2\theta)^2]dx \tag{2.3}$$

Fundamentally, mixing is a result of molecular diffusion, and hence the diffusive (second) term in equation 2.3 represents the effect of mixing, while the first term containing the gradient of θ represents the effect of stirring. This implies that an initial high value of $\nabla\theta$ will promote mixing and hence diffusion, which in turn will to lead to a decrease in $\nabla\theta$. This can also be verified from a dynamical systems point of view. [28] in his review paper

describes mixing as follows. Let us consider the basin A with a circulation where there is a domain B with a dye occupying at t = 0 the volume $V(B_0)$. Let us consider a domain C in A. The volume of the dye in the domain C at time t is $V(Bt \cap C)$, and its concentration in C is given by the ratio $V(B_t \cap C)/V(C)$. Full mixing is defined in the sense that in the course of time, for any domain $C \in A$, the concentration of the dye is the same as in every other region in A. However, calculating the true three-dimensional Eulerian flow field, and the distribution of θ, for an actual geophysical flow (e.g. a hurricane) is an impossible task. This is due to the immense range of scales that typifies naturally occurring fluid motions. If one considers a hurricane, active scales range from hundreds of kilometers to sub millimeter scales. Many models in geophysical fluid dynamics thus focus on representing the coherent scales of motion. In such cases the fundamentally three dimensional motions that would carry out efficient mixing are filtered out during the theoretical derivation of the governing equations. A Lagrangian approach to mixing, based on particle proximity, may thus be more profitable. This is because it allows for an idealized representation of the three-dimensional turbulence that is ignored by the governing equations .

[14] provides an example of this approach to describe mixing. His idea is stochastic, where each particle has a deterministic component of motion governed by the known flow field and a random walk component. The particles are assigned scalar properties which can change due to mixing. The random walk component depends on the joint probability distribution of the particle as functions of position and the scalar properties. In his equation (36) the author defines the intensity of mixing between two particles as proportional to the distance between the particles in physical space. Inspired by [14], we use a numerically inexpensive version of this idea, by loosely saying that, there is some non-zero probability of mixing with exchange of properties taking place between two particles that approach below a given threshold and a qualitative measure of mixing is given by interaction among particles. Interaction once occurred, is counted as a unit of mixing and our hypothesis says that, if we have three particles, say, A, B and C, and if particle A interacts with particle B and if particle B interacts with particle C, then indirectly, particle A has interacted with particle C, to some extent. We then extend this idea to the assumption that a region comprising of a higher number of interacting particles corresponds to one with higher probabilities of mixing. The technical details are discussed in section [2.2.3].

The remaining parts of the chapter are structured in the following manner. Section [2.2] discuses the methods used in our work including the governing equations and description of the numerical code used to solve them. This is followed by the methods for clustering particles (section 2.2.2), identifying regions of mixing (section 2.2.3) and the methods for spectral clustering (section 2.2.4). Section [2.3] presents a detailed discussion of the results obtained by implementing each of the methods above and also draws relevant comparisons

as needed. The final section [2.4] concludes the work and highlights the major findings.

2.2   Methods

2.2.1   Governing Equations and Numerical Methods

We consider the shallow water equations on the f–plane [16]. All simulations are carried out with a code developed in house using CUDA, called CUDA Shallow Water and Particles (cuSWAP), which provides numerical solutions to the Shallow Water equations. CUDA is a C/C++ based parallel computing platform developed by NVIDIA to harness the computational power of GPUs [24]. We choose to solve these equations using spectral methods to take advantage of the cuFFT library [25]. Spectral methods are those which convert a PDE to its frequency domain, usually by taking its Fourier transform. Here the spatial domain is transformed into the frequency domain by decomposing the variables η, u and v into their wave number components. Derivatives of Fourier transforms are linearly related to the original variable. This converts a differential equation into a algebraic equation after taking the transform. Solving algebraic equations (like Eigen value problems) is easier than solving a differential equation. Once we solve for the variable in its frequency domain, we take its inverse as permitted by the inverse Fourier transform theorem. For the shallow water equations, this returns an ODE (ordinary differential equation) in time, which is solved using low memory Heun's time stepping scheme [2]. Heun's method is a numerical method for solving ODEs. Heun's method is derived from Euler's method, which says the following. Given an ODE,

$$y'(t) = f(t, y(t)) \tag{2.4}$$

the solutions can be approximated as follows,

$$y^{n+1}(t) = y^n(t) + f(t^n, y^n)\Delta t \tag{2.5}$$

where the temporal discretization ($\Delta t = t^{n+1} - t^n$) is uniform. Also known as the two stage second order Runge-Kutta or the explicit trapezoidal method, the Heun's method calculates an intermediate step before obtaining the final solution as follows,

$$\tilde{y}^{n+1} = y^n(t) + f(t^n, y^n)\Delta t \tag{2.6}$$

$$y^{n+1}(t) = y^n(t) + [f(t^n, y^n) + f(t^{n+1}, \tilde{y}^{n+1}]\frac{\Delta t}{2} \tag{2.7}$$

It is called low memory, because we use the same variable in the code to store the intermediate variable and the final solution. For functions like ours (described below) which are smooth enough spectral methods are much more accurate and faster than its peers Finite Difference and Finite Volume Methods. It's well suited for problems with periodic boundary conditions which is easy to implement in a scientific setting like this. The reason spectral methods are more accurate is, the inverse Fourier transform interpolates the function between the grid points fairly accurately (given it's smooth) and hence the derivatives computed are very accurate at the grid points. Since this is a numerical study, we perform an inverse discrete Fourier transform (DFT), which in one dimension for a function $u(x)$, sampled at uniformly spaced discrete points reads,

$$u(x_j) = \frac{1}{N} \sum_{k=0}^{N-1} e^{\frac{2\pi i k j}{N}} \hat{u}(k), \qquad j \in \{0, 1, ..N-1\} \tag{2.8}$$

The above computation has a time complexity of $\mathcal{O}(N^2)$, which can be undesirably long for high resolution simulations. Hence we use the famous fast Fourier transform algorithm (FFT) by [4]. FFT algorithm performs DFT with $\mathcal{O}(NlogN)$ time complexity. For FFT to work best we choose $N = 2^m$ for some m. FFT gains speed by using a divide and conquer approach. It splits the N data points into even and odd indexed sets recursively, until there is nothing to split, thereby gaining some speed at each split. It also uses the periodicity or symmetry of the complex exponential which is essentially powers of Nth root of unity to calculate only half of the values needed from each split. This code solves the governing equations in a doubly periodic domain with variable topography. The high wavenumbers in the spectra are suppressed, to filter out the noisy effects of very small length scales. This is done by choosing a cut-off wavenumber equal to 0.6 times the Nyquist frequency (k_{nq}). Amplitudes of wavenumbers greater than the cut off wavenumber (k_0) are multiplied by $f = e^{-\alpha \frac{(k-k_0)^\beta}{(k_{nq}-k_0)^\beta}}$, choosing $\alpha = 20$ and $\beta = 2$. The I/O is handled using NETCDF.

This code also has a Lagrangian attribute which performs particle tracking using cubic interpolation and symplectic Euler time-stepping [22]. The symplectic Euler is a modification of the standard Euler method for approximating the dicrete solution for equations of motion. When solving a system of equations, the symplectic Euler uses the computed values of variables at the next time step to compute the values of remaining variables at the next time step. Consider the position and velocity of particles in one dimension given

19

by,

$$\frac{dv}{dt} = g(t, x, v)$$
$$\frac{dx}{dt} = f(t, x, v)$$

(2.9)

In symplectic Euler, we perform the following update rules,

$$v^{n+1} = v^n + g(t^n, x^n, v^n)\Delta t$$
$$x^{n+1} = x^n + f(t^n, x^n, v^{n+1})\Delta t$$

(2.10)

The symplectic Euler conserves energy better than the standard Euler method and hence is more accurate.

Additionally this code dynamically calculates and outputs neighbours of a particle based on inter-particle distance. This is done in CUDA for every particle in parallel and heuristic bucket based comparison technique saving us a lot of computational time. This data represents particle interactions and is used to construct adjacency matrices relevant to our work as described in section [2.2.2].

The shallow water equations, written out in the form amenable to numerical solution with the FFT-based method, express the conservation of mass

$$\frac{\partial \eta}{\partial t} + (H + \eta)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\left(\frac{\partial H}{\partial x} + \frac{\partial \eta}{\partial x}\right) + v\left(\frac{\partial H}{\partial y} + \frac{\partial \eta}{\partial y}\right) = 0,$$

and the conservation of linear momentum,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -g\frac{\partial \eta}{\partial x},$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -g\frac{\partial \eta}{\partial y},$$

where $\eta(x, y, t)$ is the perturbation height field, $H(x, y)$ is the bottom topography (taken as constant throughout the present work), $(u(x, y, t), v(x, y, t))$ are the velocity field, f is the rotation rate taken as constant (i.e. the f–plane), and g is the acceleration due to gravity. The pressure field is hydrostatic.

The initial conditions consist of a geostrophically balanced jet and an ageostrophic perturbation with a radially symmetric form. The exact functional form of the perturbation

was not found to be important for triggering the instability of the jet. The functional form of the initial conditions is given by,

$$u(x, y, 0) = 2ga_0 \frac{\tanh(y)}{\cosh^2(y)}$$

$$v(x, y, 0) = 0$$

$$\eta(x, y, 0) = a_0 \left(\frac{1}{\cosh^2(y)} + \frac{1}{\cosh^8(\sqrt{x^2 + y^2}/2)} \right)$$

where $a_0 = 0.1H_0$. The two relevant dimensionless numbers are the Froude number and Rossby number,

$$Fr = \frac{U}{\sqrt{gH}} \approx 0.17,$$

$$Ro = \frac{U}{fL} \approx 0.3775.$$

Results will be reported in dimensionless form. The simulation is thus carried out in a square domain with side dimension 10. The resolution used is 2048×2048 and the number of particles tracked is 400×400, initially distributed uniformly in a grid pattern. The resolution is fine enough to represent both the primary, vortex generating instability, and the filaments formed from the interaction between vortices. We have carried out a number of resolution checks and indeed the 2048×2048 grid over resolves the relevant phenomena. A factor of four decrease leaves the results essentially unchanged. While mixing is a small scale phenomenon, it is not believed the results reported below are affected by the numerical discretization. Moreover, on a grid of fixed side, the spectral method employed is very close to the optimal numerical method available. Indeed a far more serious question down the line is how to represent the transition from large scale, nearly two-dimensional flow to three-dimensional flow; a change that would require a fundamental change in the software used.

2.2.2   Clustering particles

Clustering the particles in a flow means we group the particles based on some form of particle behaviour we wish to identify. In this project we target the phenomenon of mixing in a flow by measuring instances of particle-particle proximity below a threshold. The

inter-particle interactions we employ fall under the category of binary classification, i.e. two particles have either interacted or they have not. We set a threshold inter-particle distance ϵ such that at some given time, if the distance between any two particles becomes less than ϵ, those two particles will be said to have interacted with each other at that time. For mixing, it is natural to demand that the value of ϵ is less than particle particle grid spacing (though note that [26] in fact demand ϵ to be greater than the particle grid spacing for spectral clustering). Thus, for every time step we search for particles which are within a radial distance of ϵ from every particle. A natural mathematical way to represent this information is to build a matrix. These matrices are known as adjacency matrices which are symmetric square matrices with dimensions *(number of particles^2)*. Each row in an adjacency matrix corresponds to a particle and the columns correspond to all the particles this particle may interact with. If particle 'i' is said to have interacted with particle 'j', then the adjacency matrix, an initially zero matrix, will have 1 in cells (i, j) and (j, i). **Fig.** (2.2) demonstrates a tutorial example of how to construct an adjacency matrix from particle interactions. There are two ways in which we create an adjacency matrix in our work:

- *Cumulative adjacency matrix:* One interaction between two particles in the entire time span will yield a permanent 1 in the corresponding cells of the particles in the matrix.

- *Instantaneous adjacency matrix:* One interaction between two particles at a particular time will yield a temporary 1 in the corresponding cells of the particles in the matrix. This type of matrix is refreshed every output time and new 1s and 0s are registered for the new output time.

Before we describe how we cluster these particles based on their interactions, we quickly introduce graphs from discrete mathematics. A graph is a structure which has a set of objects and some objects may be related to each other in some way. The objects are called nodes, and if two nodes are related to each other in some way, they are connected by an edge. Mathematically, a graph is represented in the form of an ordered pair $G = (V, E)$ where V is a set of vertices or nodes and E is set of edges which consists of two element subsets of V. An adjacency matrix can be converted into a graph with the particles forming the nodes and the interactions forming the edges. Looking at **Fig.**2.2a, we construct a corresponding graph shown in **Fig.**2.2b

A graph formed from an adjacency matrix of particle interactions, can be used to cluster the particles by finding connected components in a graph. We demonstrate this concept
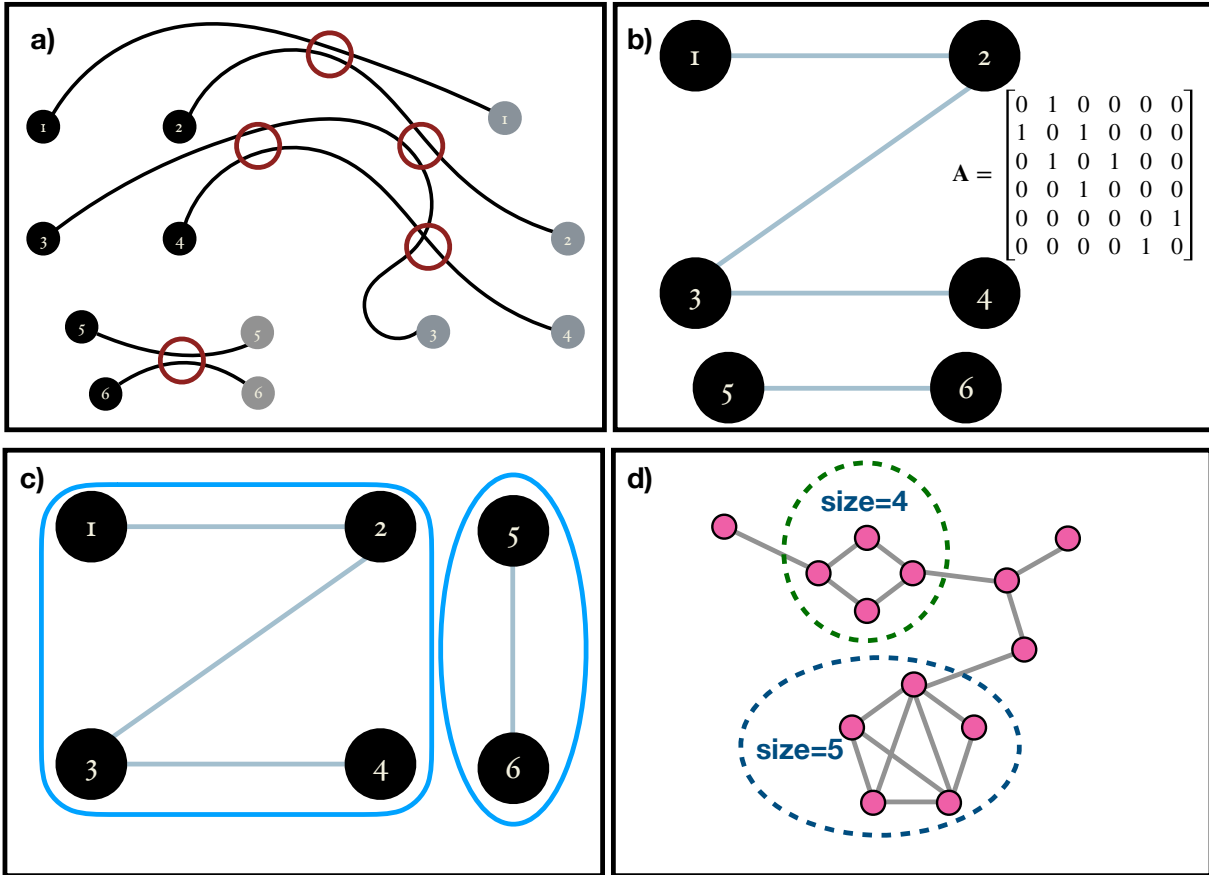
Figure 2.2: a) Idealized Lagrangian paths of six particles showing where they have interacted along the course of their paths. b) Adjacency matrix and graph corresponding to the particle interactions shown in part (a) c) Graph split into its connected components d) A connected graph symbolizing a scaled down version of a cumulative cluster; The black dotted circles denote the dense-sub graphs for an arbitrary $min\_size = 3$ and $\gamma = 0.4$

in **Fig.**2.2c. It is seen that the graph can be visually split into two parts. These are two separate, connected components in our imaginary graph. The connected components in a graph can be mined by using a standard depth first search algorithm. We carry out this procedure on the graph in our problem using MATLAB. The different connected components in the graph form the different clusters. In regards to our earlier point of mixing we see that each cluster has particles that have interacted with at least another particle inside the cluster and thus odds are high that some mixing may be happening

among particles within these clusters. This gives us a level one classification of particles which will later help us track down regions of mixing.

2.2.3   Mining dense sub-clusters from a cluster

Until this point, clusters have been based on inter-particle interactions. Though, these clusters tell us about which particles interacted, they do not tell us anything about the degree or intensity of interaction. We want to find regions in the flow where there are higher intensities of mutual interactions among particles compared to rest of the flow. We consider a cumulative cluster, which is a connected graph and use the pruning algorithm *Quick* described by [17] to look for dense sub-clusters within this cluster.

A clique is a graph whose nodes are all connected to each other, hence a clique is 100% dense. The minimum degree of a graph is the minimum number of neighbors that a node has in the graph. Let the minimum degree be denoted by deg_{min} and N be the size of the graph. A γ-quasi clique is a graph which satisfies:

$$deg_{min} >= \gamma[N-1] \tag{2.11}$$

where $\gamma \in (0,1)$. The density of a sub-graph is based on the following parameters:

- The density parameter γ, such that (2.11) is satisfied.

- Minimum size of a subgraph. The algorithm will only look for solutions whose sizes are greater than or equal to the specified minimum size parameter, min_size.

All subgraphs mined, hence, have a minimum degree greater than or equal to $\gamma(min\_size - 1)$. These two parameters drive how many minimum particles we want from a dense sub-cluster to have interacted with a particle in the same dense sub-cluster. We search for sub-clusters throughout the entire flow with a minimum size of 20 and $\gamma = 0.25$, so that the minimum degree is at-least 5 at $t = 50$. There are cases where subsets of a bigger γ-quasi clique are also γ-quasi cliques. The algorithm *Quick* makes sure that it mines only the maximal γ-quasi cliques for a specified γ. The algorithm is described in the next subsection.

Fig. 2.2d shows an example of how dense sub-clusters are mined. The connected graph in **Fig.** 2.2d can be a considered as a small illustration of an actual cumulative cluster of particles. For an arbitrary $\gamma = 0.4$ and minimum size of the sub-graphs equal to 3, the algorithm shows that the nodes inside the dotted circles are dense sub-graphs inside the graph. In the context of Lagrangian fluid mechanics, interactions among particles in these sub-clusters are much denser than other regions in the flow.

Description of the Quick algorithm

We will now introduce graph theoretic terminology that we will be required in the following section. This work is based on [17].

A *graph G* is an ordered pair of sets (V, E), where V is a set of vertices and E is a set of edges joining the vertices.

Neighbours of a vertex v in G are denoted by $N^G(v)$ which are the nodes adjacent to v in G.

The *degree* of a vertex v in G, denoted by $deg^G(v)$, is the number of neighbours of v, $|N^G(v)|$.

The *distance* between two vertices u and v in G, denoted by $dist^G(u, v)$, is the number of edges on the shortest path from u to v.

For a vertex v in V, $N_k^G(v) = \{u | dist^G(u, v) \leq k\}$ denote the k-nearest neighbours of v.

The *diameter* of a graph G, denoted by $diam(G)$ is defined as $max_{u,v \in V} dist^G(u, v)$.

For any vertex set $\{X | X \subset V\}$, $cand\_exts(X)$ represents the set which contains vertices that can be used to extend the set X in order to form a γ-quasi clique.

For a vertex u in a vertex set X, $indeg^X(u)$ represents the number of neighbours of u in X and $exdeg^X(u)$ represents the number of neighbours of u in the set $cand\_exts(X)$.

The minimal degree of vertices in X, denoted by $deg_{min}(X)$, is $min\{indeg^X(v) + exdeg^X(v) | v \in X\}$.

It follows from the definition of a γ-quasi clique that the maximal number of vertices in $cand\_exts(X)$ that can be added to X concurrently, is less than $U_X^{min} = \lfloor deg_{min}(X)/\gamma \rfloor + 1 - |X|$.

In another case where, vertex $u \in X$ and $indeg^X(u) < \lceil \gamma(|X|-1) \rceil$, it becomes apparent that at-least some vertices must be added to X so it can be extended to form a γ-quasi clique. This lower bound is denoted by L_X^{min}. Let $indeg_{min}(X) = min\{indeg^X(v) | v \in X\}$, then L_{min}^X is defined as $min\{t | indeg_{min}(X) + t \geq \lceil \gamma(|X| + t - 1) \rceil\}$

Quick uses several effective pruning techniques to eliminate vertices from $cand\_exts(X)$ of a vertex set X. Valid extensions are added to X, to check if the new vertex set ($X \cup cand\_exts(X)$) satisfies the γ-quasi clique criterion. The following pruning techniques form an essential part of *Quick* algorithm. The proof of the Lemmas used by these techniques can be found in [17].

- Depending on γ, we find a k such that vertices not in $\bigcap_{v \in X} N_k^G(v)$ are removed from $cand\_exts(X)$. This is called pruning based on diameter.

- We use the Cocain algorithm [34] to eliminate all such vertices u from $cand\_exts(X)$ who satisfy $indeg^X(u) + exdeg^X(u) < \lceil \gamma(|X| + exdeg^X(u)) \rceil$. This is because, neither such a vertex u nor any of its neighbours in $cand\_exts(X)$, if added, will satisfy the γ-quasi clique criterion.

- We set an upper bound U_x based on U_X^{min}, such that, $U_X = max\{t | \sum_{v \in X} indeg^X(v) + \sum_{1 \leq i \leq t} indeg^X(v_i) \geq |X| \lceil \gamma(|X| + t - 1) \rceil, 1 \leq t \leq U_X^{min}\}$, where v_i are vertices in $cand\_exts(X)$ sorted in descending order of their $indeg^X$ value. If vertex $u \in cand\_exts(X)$ and $indeg^X(u) + U_X - 1 < \lceil \gamma(|X| + U_X - 1) \rceil$, such a vertex u can be pruned from $cand\_exts(X)$. Otherwise, if $u \in X$ and $indeg^X(u) + U_X < \lceil \gamma(|X| + U_X - 1) \rceil$, then γ-quasi cliques cannot be generated by extending X.

- We set a lower bound L_X based on L_X^{min}, such that, $L_X = min\{t | \sum_{v \in X} indeg^X(v) + \sum_{1 \leq i \leq t} indeg^X(v_i) \geq |X| \lceil \gamma(|X| + t - 1) \rceil, L_X^{min} \leq t \leq n\}$, if such t exists, else $L_x = |cand\_exts(X)| + 1$. If vertex $u \in cand\_exts(X)$ and $indeg^X(u) + exdeg^X(u) < \lceil \gamma(|X| + L_X - 1) \rceil$, such a vertex u can be pruned from $cand\_exts(X)$. Otherwise, if $u \in X$ and $indeg^X(u) + exdeg^X(u) < \lceil \gamma(|X| + L_X - 1) \rceil$, then γ-quasi cliques cannot be generated by extending X. Before performing the above checks, we also check if $L_X > U_X$, and if true there is no need to extend X further.

- In a vertex set X, if we have a vertex $v \in X$ such that $indeg^X(v) + exdeg^X(v) = \lceil \gamma(|X| + L_X - 1) \rceil$, then v is called a critical vertex of X. If $G(Y)$ is a γ-quasi-clique and v is a critical vertex, we have $\{u | (u, v) \in E \wedge u \in cand\_exts(X)\} \subseteq Y$. Hence, whenever we encounter a critical vertex in our vertex set X, we instantly add its neighbours present in $cand\_exts(X)$ to X.

- We are mining exclusively maximal γ-quasi-cliques and it can be proved that if u is a vertex in $cand\_exts(X)$ such that $indeg^X(u) \geq \lceil \gamma |X| \rceil$ and if for any $v \in X$ such that $(u, v) \notin E$, we have $indeg^X(v) \geq \lceil \gamma |X| \rceil$, then for any vertex set Y such that $G(Y)$ is a γ-quasi-clique and $Y \subseteq (X \cup (cand\_exts(X) \cap N^G(u) \cap (\bigcap_{v \in X \wedge (u,v) \notin E} N^G(v))))$, $G(Y)$ cannot be a maximal γ-quasi-clique. So we use $C_X(u) = (cand\_exts(X) \cap N^G(u) \cap (\bigcap_{v \in X \wedge (u,v) \notin E} N^G(v)))$ to denote the vertices covered by u and u is called the cover vertex of X. We find u such that it maximizes $C_X(u)$, put the vertices in $C_X(u)$ at the end of $cand\_exts(X)$ and then use the vertices in $cand\_exts(X) - C_X(u)$ to extend X.

2.2.4 Spectral Clustering

Spectral clustering is based on the normalized cut criterion of solving a graph segmentation problem [32]. Here we explore a different method of sub-clustering a cumulative cluster that does not require the threshold spacing ϵ to be greater than the particle grid spacing. Once we identify a cumulative cluster, we extract the portion of the adjacency matrix corresponding to particles exclusively within it. Let's suppose we name this adjacency matrix A. We find the degree matrix, D which is a diagonal matrix with $D_{ii} = d_i$, where d_i is the degree of the node x_i, i.e., $D_{ii} = \sum_{j=1}^{n} A_{ij}$, the number of neighbours of node i. The non-normalized graph Laplacian is given by $L = D - A$, and the normalized graph Laplacian is given by $\mathcal{L} = I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. The eigenvalues of $\mathcal{L}$ are real and non-negative and are in the order $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq ... \leq \lambda_n$. The second smallest eigenvalue λ_2 is called the algebraic connectivity [6] of a graph and can only be non-zero if the graph is connected. We expect that to be true in our case as the cumulative cluster corresponds to a connected graph. Spectral clustering is expected to help find coherent structures in fluid transport, which in lay-man's terms means particles whose trajectories stay close to each other or interact more often. The mathematics in this section is the outcome of solving a balanced cut problem in a network [12]. So the idea is if λ_2 is the only eigenvalue close to zero then the graph is nearly decoupled into two communities. Similarly if all λ_i, $i = 2, 3, ...k$ for some $k < n$ are close to zero and there is a spectral gap between λ_k and λ_{k+1}, then the cluster is nearly separated into k communities. The corresponding eigenvectors carry information about the division of these particles. Hence, we capture these eigenvectors, performing a dimensional reduction on our data, and apply unsupervised clustering on them. We employ the standard *k-means clustering algorithm* [18] on the reduced data to identify the different communities. Since we are already in a cumulative cluster, and the further clustering is supposed to reveal the coherent structures in the flow, we expect to find the regions with a comparatively higher intensity of interaction. However, since we use *k-means clustering*, we do not expect it to identify precise locations of solely high intensity interactions because *k-means* will produce communities whose union is exhaustive.

2.3 Results

2.3.1 Cumulative clusters

Fig.(2.3) shows the different cumulative clusters, found at time $50 - 58$ in the simulation, in different colors. By this time the double jet has undergone instability and coherent vortices,

as well as vorticity filaments, are formed **Fig.(2.1)**. As explained earlier, cumulative clusters are formed by particle-particle interactions that occur up to a particular time. The threshold separation ϵ for interaction between two particles is 40% of the particle grid spacing in this case. We can see in this figure how different clusters merge during their evolutions. An example for this is the transition from time 52 to 54 in **Fig.(2.3)**, where the green and magenta clusters merge into one magenta cluster. Two clusters merge into one when a particle from one cluster interacts with a particle from another cluster. A question that follows is "Can new clusters take the place of old clusters when they merge?" The answer is yes, we can easily show the formation of new clusters having size of the same order. We create another figure, **Fig(2.4)**, which is identical to **Fig.(2.3)**, except for the threshold interaction distance ϵ set to equal 20% of the initial spatial particle grid spacing now. Comparing **Fig.(2.3)** and **Fig.(2.4)**, we see that the clusters in the later are smaller than those in the first. This is obvious because fewer particles interact with a threshold distance equal to 20% of the particle grid spacing. In particular, particles in the clusters shown in **Fig.(2.4)** interact more strongly than those in **Fig.(2.3)**, and hence the clusters do not evolve the same way in the two cases. Specifically the clusters in the smaller 20% case, do not change size or merge, and their paths are more or less periodic moving around the coherent vortex.

2.3.2   Dense sub-clusters

Fig(2.5) shows the four largest cumulative clusters with $\epsilon = 40\%$ of the particle grid spacing, found at time 50 (particles in black) and also plots the dense subclusters mined from within these clusters (particles in blue). We number these clusters as cluster **1**, **2**, **3** and **4** in descending order of their sizes. Recalling the graph theoretic terminology from section 2.2.3, we know each of these subclusters is a graph with a minimum degree of 5. Dense subclusters locate the regions in a cluster where there are many interactions among particles, significantly more than regions which are not blue. In simpler words these are places where particle interactions are at their peak. Particles in a dense cluster, if from sources with varying properties, are an example of localized mixing. Else, if they are from the same source, the properties of that source remain preserved in that dense cluster. Mining $\gamma-$quasi cliques is thus useful for studying the traits of mixing specific to a problem. Interestingly, the blue regions in this figure have many similarities with the clusters in **Fig(2.4)**, which represents the stronger interactions. This tells us that the regions of stronger interactions are not very different from the regions of denser interactions in our double-jet flow. In **Fig.(2.6)**, we show the local clustering co-efficient and the node degree for the top four cumulative clusters at output time 50. Comparing with **Fig.(2.5)**,
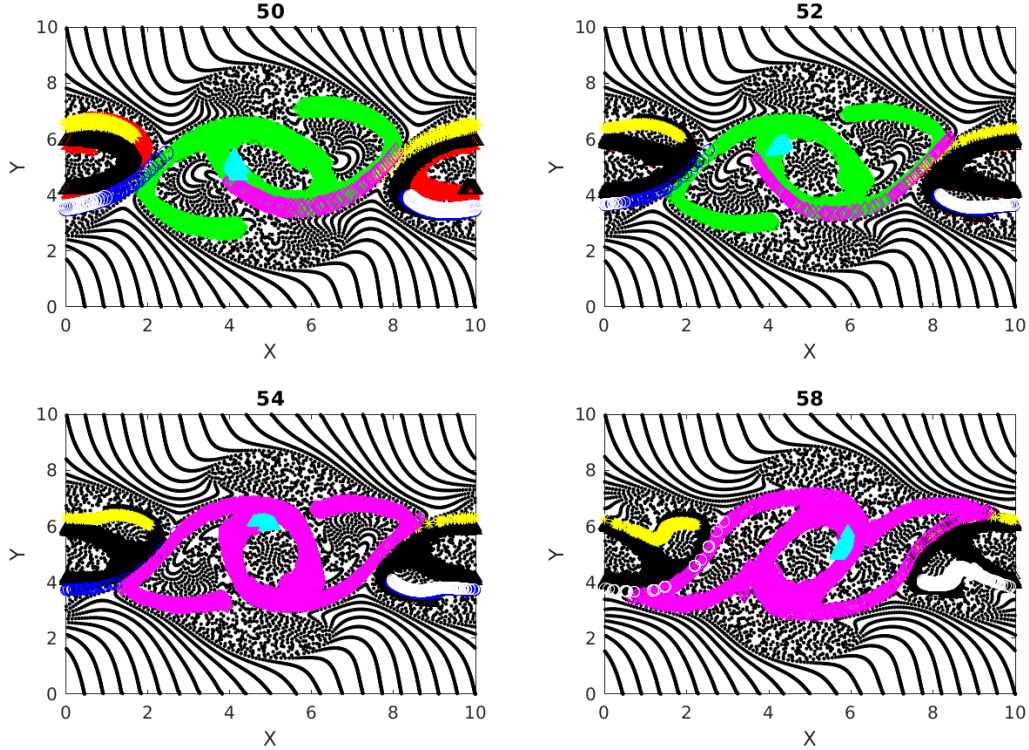
Figure 2.3: Cumulative clusters identified at time 50 with threshold distance for inter-action, $\epsilon = 40\%$ of initial separation of particles on uniform rectangular grid and their evolution tracked at later time steps (52, 54 and 58). Changing colors denote the merging of two clusters when particles from two clusters interact.

it is not surprising to find that some particles from the dense sub-clusters have large node degree and clustering co-efficient, meaning that they have potential to form local clusters.

Fig.(**2.7**), (**2.8**) and (**2.9**) show the temporal evolution of cumulative clusters **1, 2** and **3** respectively and the temporal evolution of the particles in the dense-clusters. **Fig.(2.8)** is different from **Fig.(2.7)** and **Fig (2.9)** in the sense that some particles forming the dense sub-clusters in this figure appear to split from other particles in the dense subgroups. This means that particles from these regions of dense interactions move out of their more or less periodic paths and mix with particles in other regions of the flow. We measure the displacement of the particles in dense clusters within clusters $1, 2$ and 3 from their
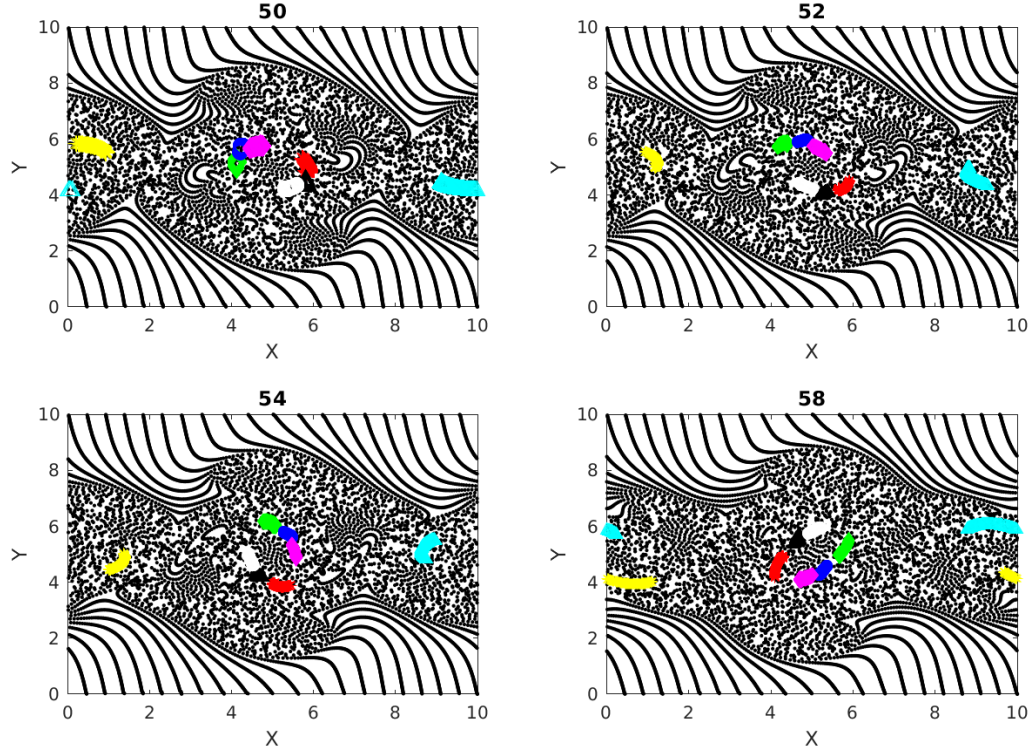
29

Figure 2.4: Cumulative clusters found at time 50 with threshold distance for interaction $\epsilon = 20\%$, of initial separation of particles on uniform rectangular grid and tracked at later time steps(52, 54 and 58). Changing colors denote the merging of two clusters when particles from two clusters interact.

positions at $t = 50$ and plot them vs output times in **Fig.(2.10)**. It is seen the paths are periodic with decreasing amplitude but same mean for clusters 1 and 2, meaning that the mean position of the particles slowly spirals toward the centre of the vortex. For the second cluster as mentioned earlier, the mean displacement increases implying that some of the particles have escaped from their original vortex. In this particular case, this is an indication that these particles that have undergone dense and strong interactions have exchanged physical properties among themselves, and when they move out of their periodic paths to mix with outside particles in the flow, there is a chance that they transfer their properties in this foreign part of the flow by interaction.
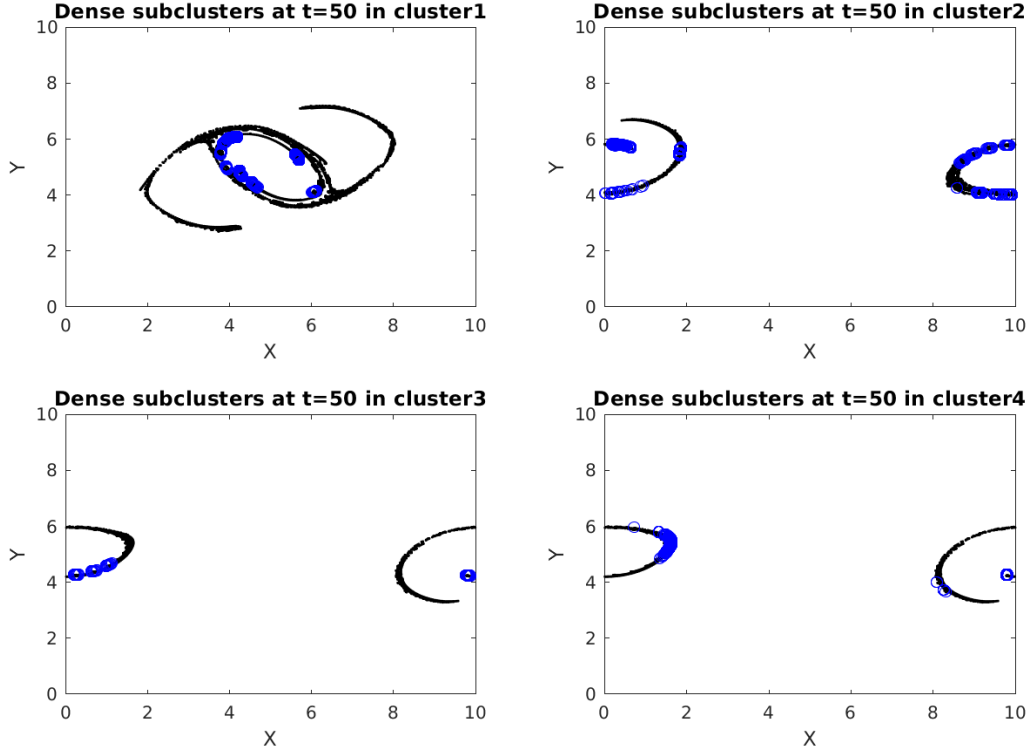
Figure 2.5: Top four (1 being the largest) cumulative clusters (black) with their dense sub clusters (blue) found at time 50. Spatially separated blue regions are distinct sub clusters with each of them having a minimum degree of 5 within themselves and hence called dense.

2.3.3 Characteristics of dense sub-clusters

In this section we explore a few characteristics of the dense sub-clustering technique. The run time of the *Quick* algorithm depends on the number of vertices V in the graph, the average degree d of the vertices, the minimum degree threshold γ, the size of quasi cliques present and the number of quasi cliques present. The data mining problem in this context doesn't have an *a priori* estimate. Hence the user has no control over the size and the number of quasi cliques present. [17] studies the effect of changing parameters on the run time of the algorithm. The run-time, t_{run} varies exponentially w.r.t the parameters as $t_{run} \sim 10^{k_v V} 10^{k_d d} 10^{-k_\gamma \gamma}$ for some constants k_v, k_d, k_γ depending on the graph.

We wish to report the effects of changing ϵ and how to determine 'the' ϵ for a problem.
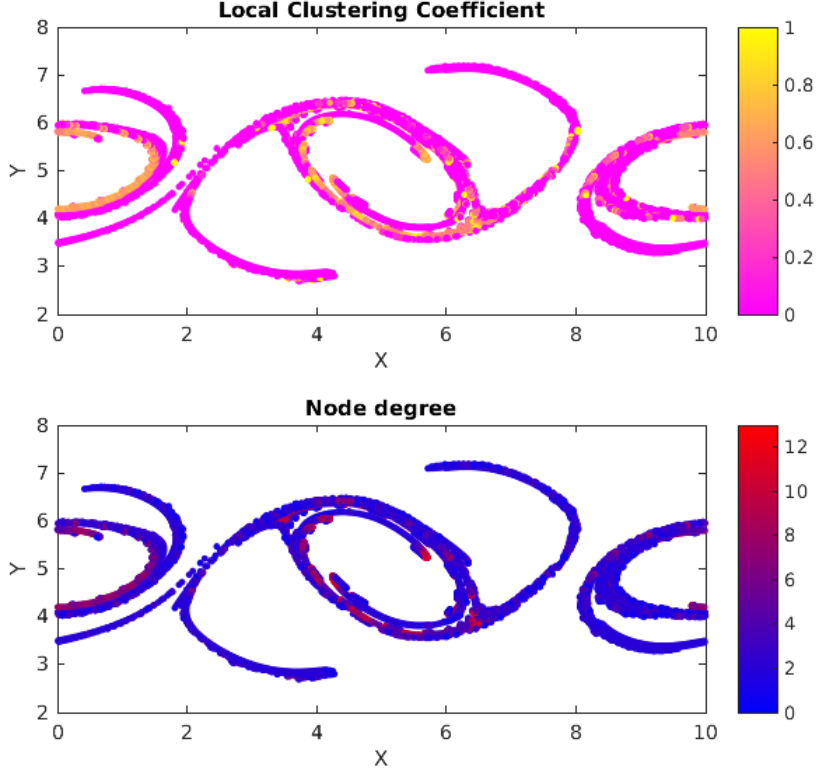
Figure 2.6: Local clustering coefficient (top panel) and node degree (bottom panel) for the top four cumulative clusters at output time 50

For the double-jet problem, increasing ϵ increases the size of the cumulative clusters considerably when compared at a fixed output time. An increase in the size of a cluster increases the computational complexity for *Quick* to mine the quasi-cliques exponentially. Let N be the total number of particles, and let C_{40} and C_{60} denote the particles in the biggest cumulative clusters for $\epsilon = 40\%$ and $\epsilon = 60\%$ respectively. Since N is fixed, $C_{40} \subset C_{60}$. To avoid excessive computational time and to draw comparisons on the same grounds we look at the induced subgraph $C_{60}[C_{40}]$. The density of connections in $C_{60}[C_{40}]$ is more than C_{40}, specifically, the average degree of nodes rise to 8.1 from 5.0. Again, to compare sets of the same class, we propose, $\frac{\gamma(min\_size-1)}{average\_degree} = constant$. Thus parameter min_size is kept constant and γ is increased from 0.25 to 0.4. However, changing ϵ essentially changes the network and the connections do not scale linearly. In Fig.(2.11), we look at dense clusters in cumulative clusters 1 and 2 with $\epsilon = 60\%$. The top left panel, shows that the dense
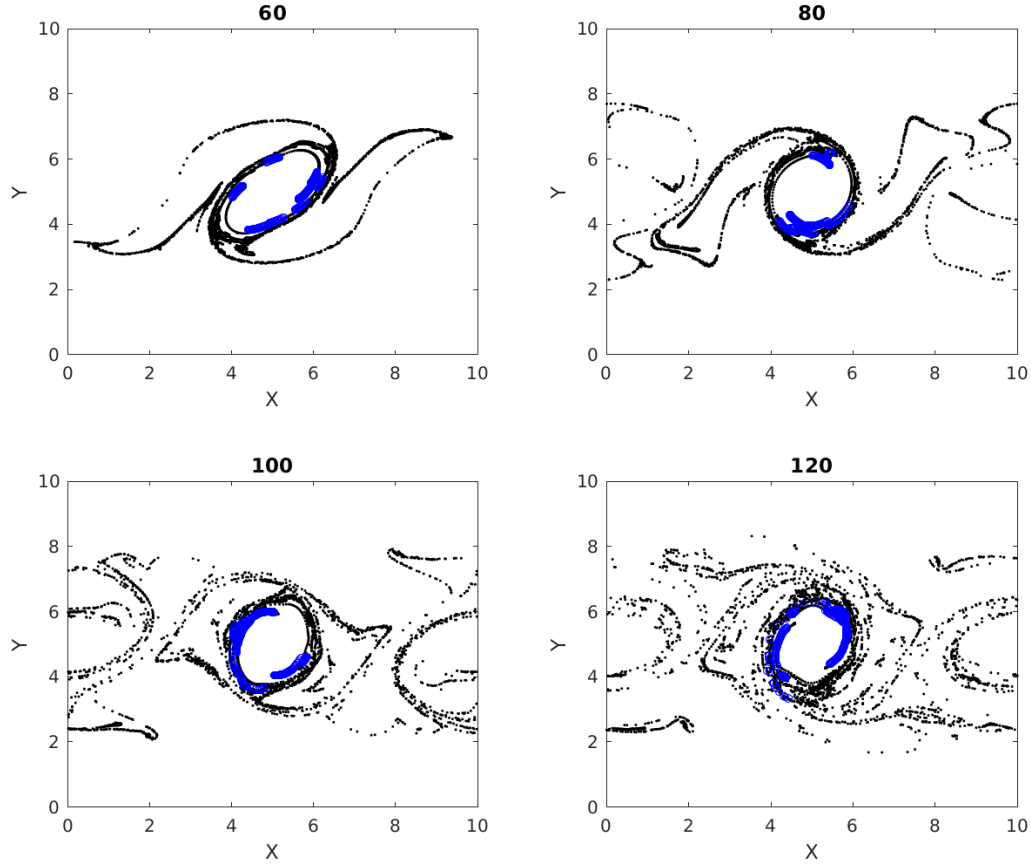
Figure 2.7: Multiple time images of cumulative cluster 1 (black) with its dense sub clusters (blue). Blue 'o's at later times represent particles that were parts of a dense sub cluster at time 50.

clusters mined with $\gamma = 0.4$ and $\epsilon = 60\%$ are a subset of those with $\gamma = 0.25$ and $\epsilon = 40\%$. The remaining particles in the $\epsilon = 60\%$ clusters cannot meet the tighter threshold criteria of the $\epsilon = 40\%$ case. The bottom left panel, shows the results with $\gamma = 0.3$. Relaxing the minimum degree criteria, yields more dense clusters, but some of them like those at the bottom of the vortex belong to a different class. This is because $\gamma = 0.3$ doesn't scale properly with $\epsilon = 60\%$. This helps us understand the scenarios of increasing ϵ further i.e. scaling up γ to make sure we remain consistent with our dense clusters. Otherwise, we are just mining densely connected graphs without physical meaning, and taking a very long
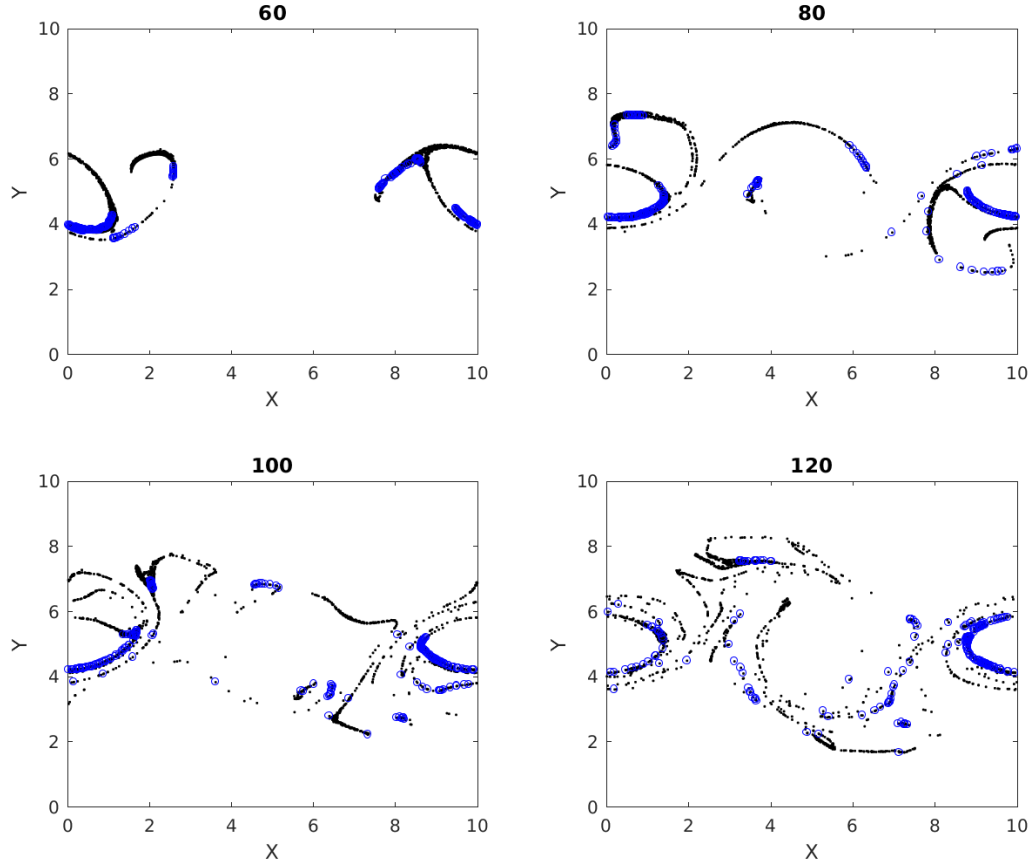
Figure 2.8: Multiple time images of cumulative cluster 2 (black) with its dense sub clusters (blue). Blue 'o's at later times represent particles that were parts of a dense sub cluster at time 50.

computational time to do so. The top and bottom right panels in the figure show the same results but for cumulative cluster 2 obtained with $\epsilon = 60\%$. It is interesting to observe in this case that improper scaling of γ might lead to re positioning of some of the maximal quasi cliques e.g. the dense cluster particles present in the left vortex of the $\gamma = 0.4$ case are absent from the $\gamma = 0.3$ case. This is because relaxing the threshold criteria caused the corresponding dense cluster to get bigger and exclude some of its previous residents. We also performed dense cluster analysis on $\epsilon = 20\%$, where the cumulative clusters are so small that almost all of them belong to the dense clusters. Hence, we suggest that the ideal
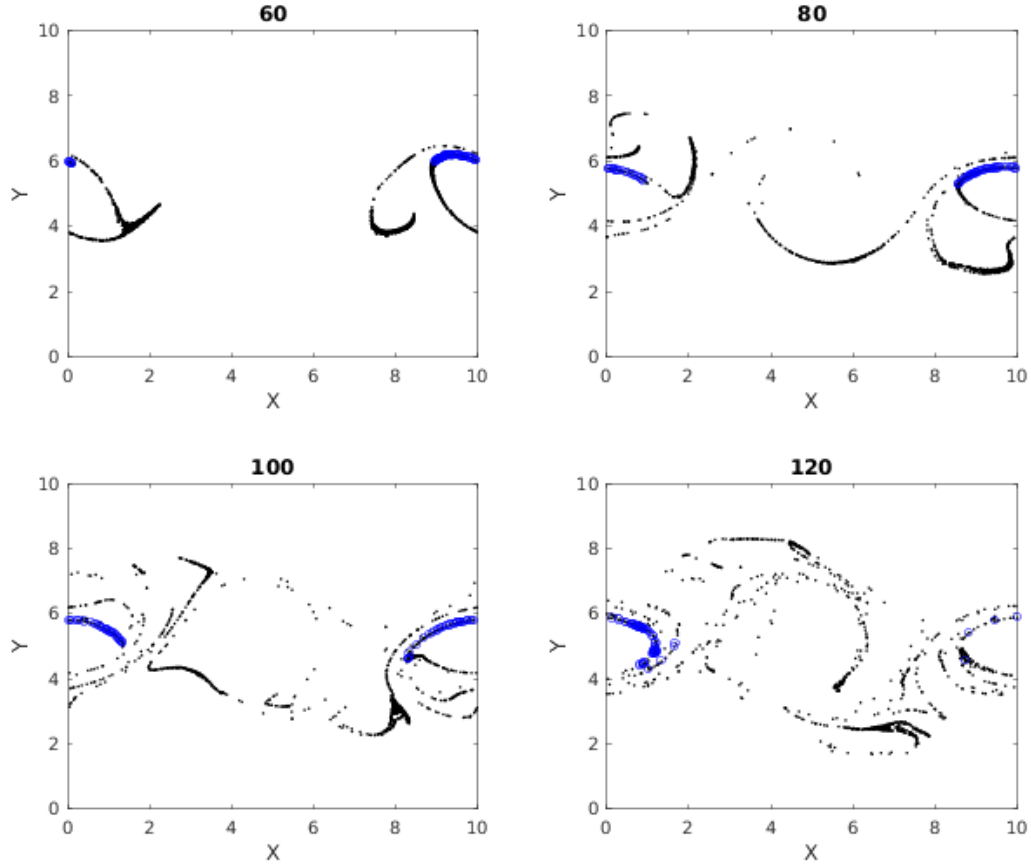
Figure 2.9: Multiple time images of cumulative cluster 3 (black) with its dense sub clusters(blue). Blue 'o's at later times represent particles that were parts of a dense sub cluster at time 50.

ϵ be kept around half of the particle grid spacing and the ideal γ as high as sufficient to obtain satisfactory quantity and quality of the dense clusters in a reasonable computational time. This requires some intuition on the part of the user, but leads to the most robust results.

Increasing min_size would simply eliminate the dense clusters which no longer meet the necessary criteria. However, it is important to note that it is necessary to tweak the min_size parameter for different cumulative clusters for best results. We show results

35

Figure 2.10: Displacement averaged over particles in dense clusters from clusters $1, 2, 3$ (DC 1,DC 2, DC 3) measured from positions at output time 50 vs output time.

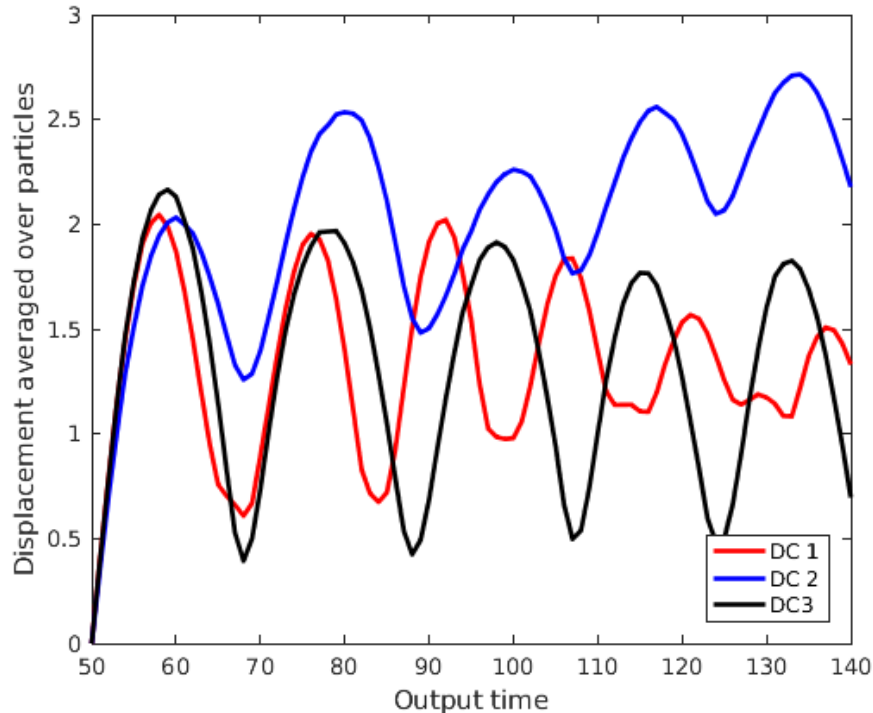of varying γ keeping min_size constant in Fig.(2.12). Increasing γ beyond 0.4 doesn't yield any dense clusters in this case. The results themselves are intuitive in the sense that decreasing γ includes returns a much larger number of sub-clusters and increasing the same beyond a point doesn't return any sub-cluster.

We tested to what extent our dense clusters are sensitive to perturbations of initial particle distribution. Fig.(2.13) shows the evolution of the dense clusters with uniformly distributed, random perturbations to the initial position of the particles. These had a maximum extent of 15% of the particle grid spacing in each direction and $\epsilon = 40\%$ in this case. The resulting dense clusters and their evolution are shown in Fig.(2.14). Comparing these two figures, we see that perturbing the particle positions changes the network and the location of the dense clusters, which is somewhat trivial. However, considering that this study is purely Lagrangian, the dense clusters from the perturbed case consistently convey qualitatively unchanged information about regions of potentially dense localized

mixing (e.g. the ring of dense subclusters around the central vortex which can be traced backwards in time to the flanks of the geostorphically balanced jet).

2.3.4 Instantaneous clusters

Fig.(2.15) shows the temporal behaviour of several of the largest instantaneous clusters found at output time 50. The instantaneous clusters at time 50 seem to be aligned along the boundary of the central vortex, showing that a large group of mutually interacting particles is concentrated in this region. Instead of finding new clusters at later times, we track the position of these clusters through later times. The instantaneous clusters act as a partial check if the dense sub-clusters make sense. It turns out that, these clusters keep moving around inside the central vortex. This means, the already highly interactive particles undergo more interactions with particles in the same region implying higher chances of mixing in the region. We have used this figure to act as a validation for our dense sub-clusters. For this problem, the flow evolves to be simple, with fluid parcels being very coherent around the outskirts of the quasi coherent vortices. This naturally means that these areas undergo a lot of property exchange through diffusion. This is validated by the fact that particles in the largest instantaneous clusters don't deviate from their paths. Although the reason for that is that the flow is periodic in that region, but this would also mean that the particles there stick to each other and thereby interact a lot. We intentionally choose the double-jet flow for this reason where the behaviour of strongest mixing can be verified easily by a simple concept like instantaneous clusters. In order to validate our method on an actual mixing problem we need a flow with properties that can be computed, which is left for future work.

2.3.5 Spectral Clusters

In this sub-section we show the results of spectral clustering described in section 2.2.4. **Fig.(2.16)** shows the different spectral sub-clusters that this algorithm splits the largest cumulative cluster (cluster 1) into. **Fig.(2.17)** shows the temporal evolution of the spectral clusters of cluster 1 found at time 50. Giving a quick recap, the spectral clustering technique is responsible for dividing the set of particles into k communities, k being 5 in the results shown. A spectral sub-cluster is expected to have more inter-particle interactions inside itself than outside because the clustering is applied on the adjacency matrix of particle interactions. The spectral sub-clusters are exhaustive and hence unlike the dense sub-clusters, all of them are not equivalently rich in particles with high degrees of interaction. This can be seen from **Fig.(2.17)** where most of the particles in the sub-clusters
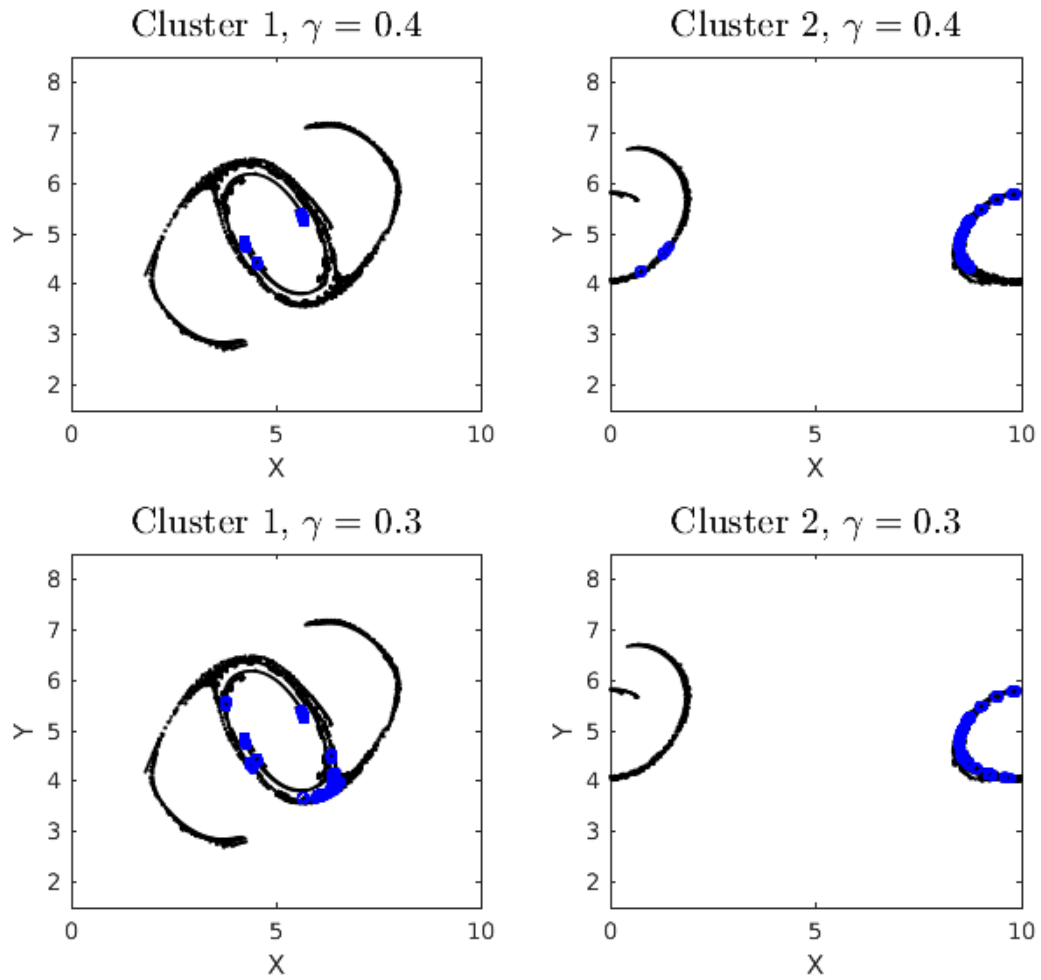
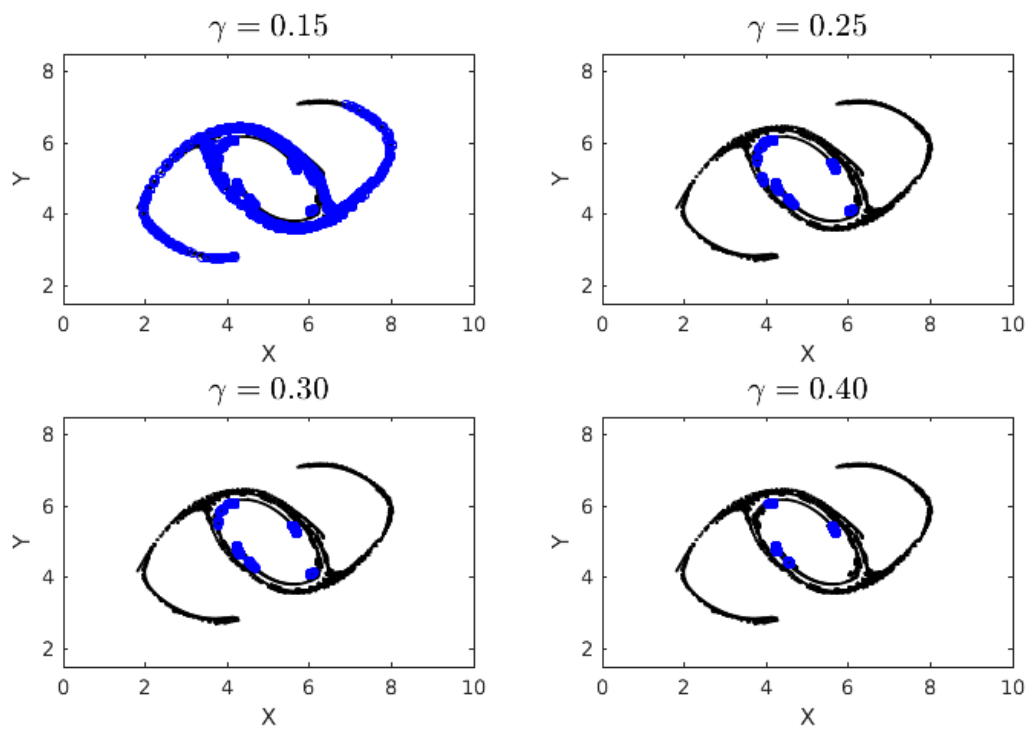Figure 2.11: Dense clusters with $\epsilon = 60\%$ in cumulative clusters 1 and 2 at $t = 50$.

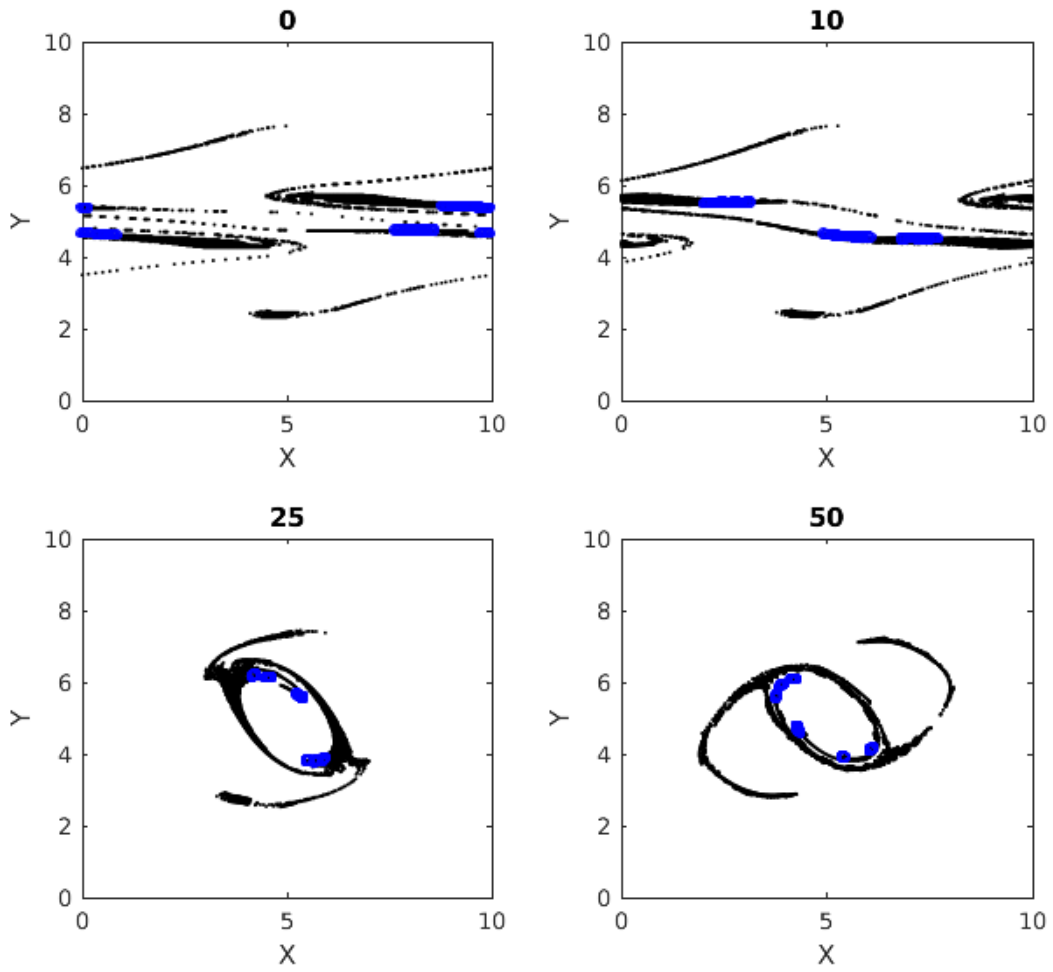Figure 2.12: Dense clusters with $\epsilon = 40\%$ for varying γ at $t = 50$

Figure 2.13: Dense clusters with $\epsilon = 40\%$ and particles on uniform rectangular grid.
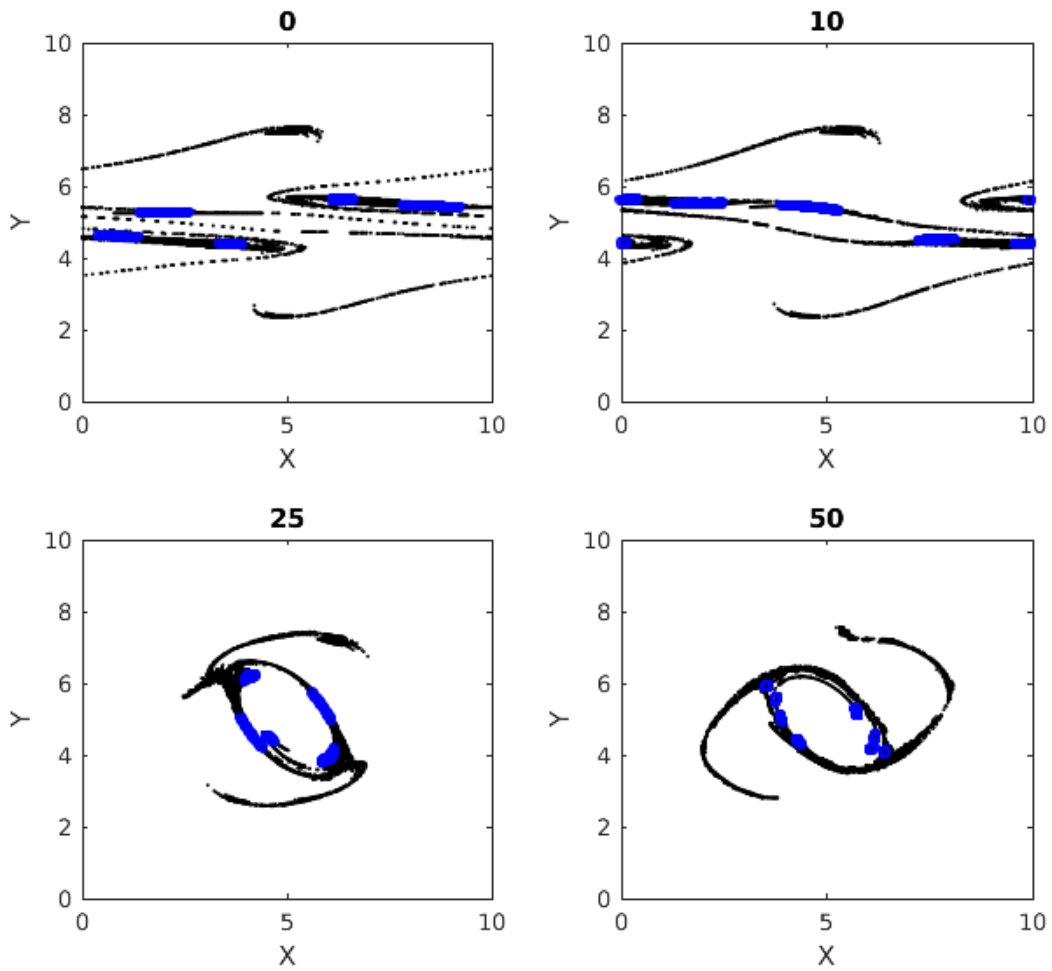
Figure 2.14: Dense clusters with $\epsilon = 40\%$ and particles on rectangular grid with perturbations.
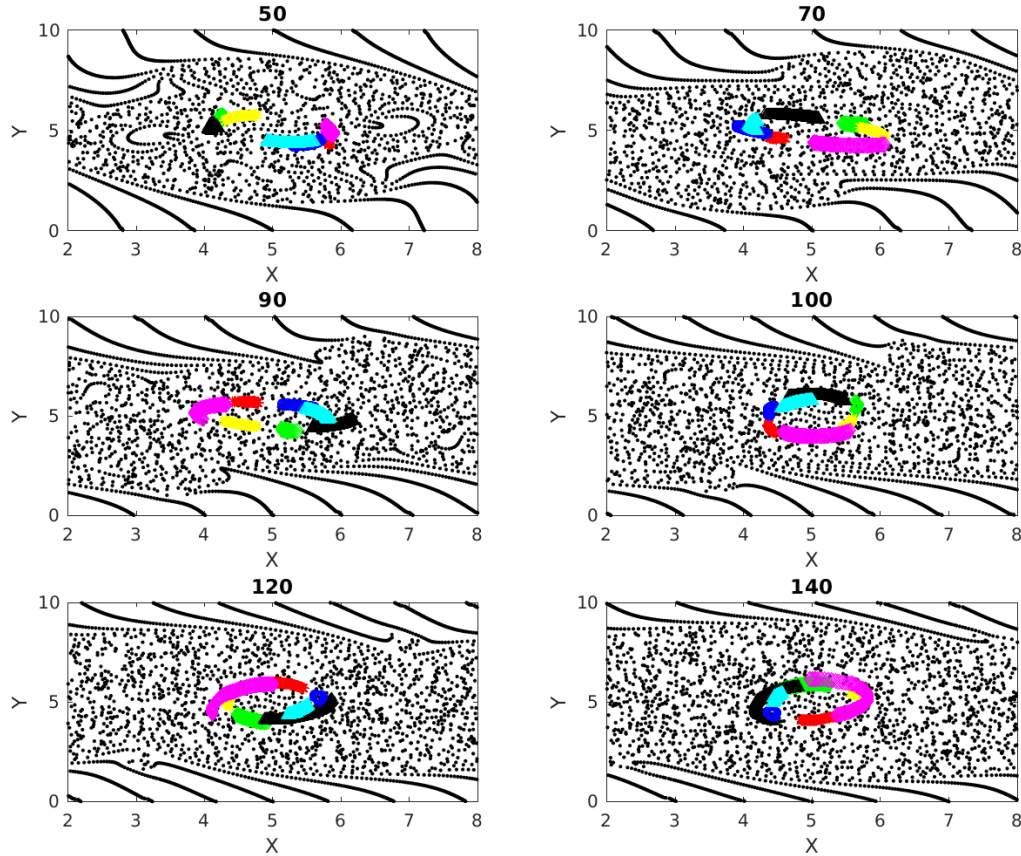
Figure 2.15: Multiple time images of top few instantaneous clusters found at time 50. Once found, particles in these clusters are tracked through later time steps.

of cluster **1** stay within the central vortex, while some others take different paths over the course of the flow's evolution. This can be explained by our hypothesis that the paths of the densely interactive particles in cluster **1** tend to stay nearly periodic with time. Examining **Fig.(2.16)**, we realize that the spatial distribution of these clusters share similarities to some extent with the dense sub-clusters from the last sub-section, especially around the coherent central vortex. This validates that these coherent structures are home to all the blue regions around the central vortex in **Fig.(2.7)** representing dense interactions and thereby strong mixing. Spectral clustering relies on k-means clustering and hence is highly sensitive to change in data distribution e.g. different output times or small perturbations

to initial particle distribution. Spectral clustering also returns sub-clusters of incomparable sizes, leaving us no way to compare the degree of mixing among the sub-clusters mined. The dense sub-clustering method on the other hand controls the density of connections and hence all sub-clusters mined belong to the same class of mixing.
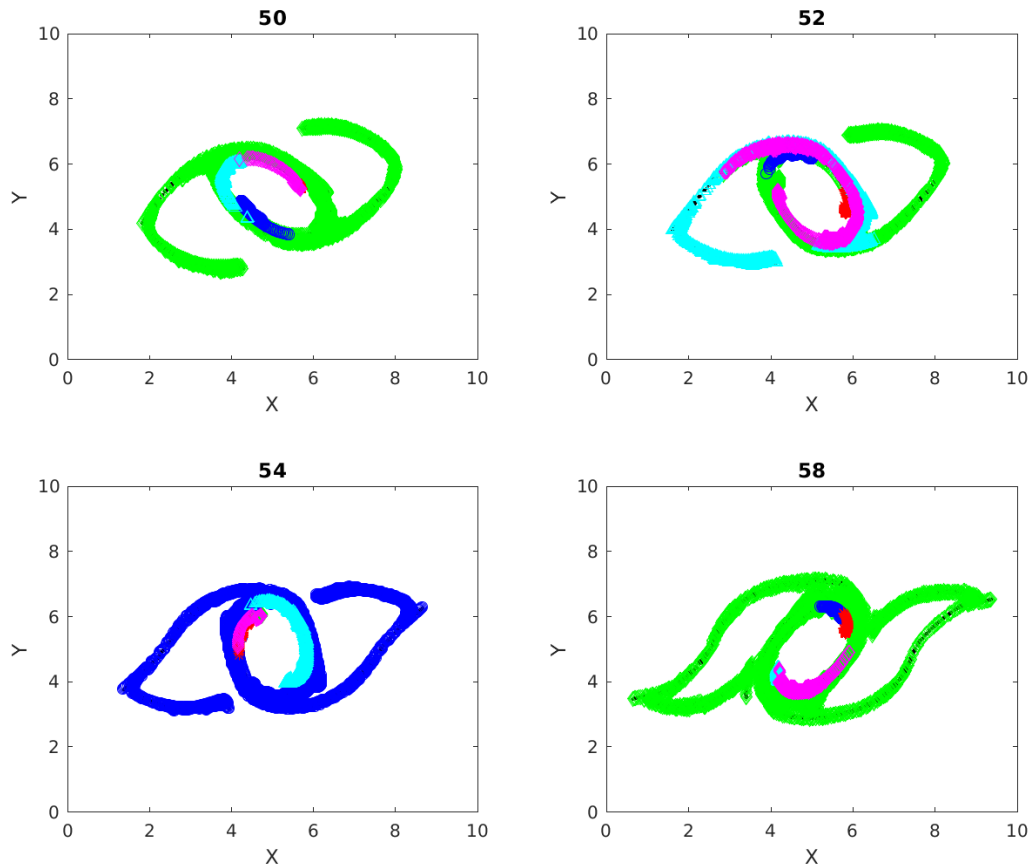
Figure 2.16: Spectral clusters found at multiple times from within cluster 1.

We also show the spectral clusters in cluster **2** identified at time 50 in **Fig.(2.18)** and look at the behaviour of the particles at different times. Comparing with **Fig.(2.8)**, we see that the particles in the dense clusters show a lot of similarity with the particles in the spectral sub-clusters, especially in the way they deviate from their initial paths and mix into other regions of the flow.
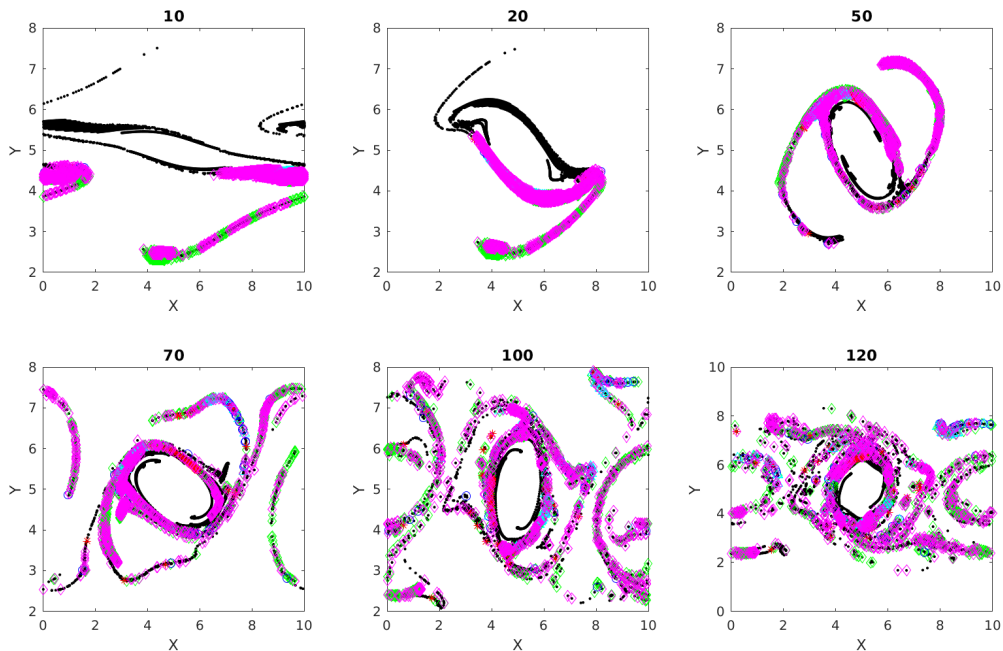
Figure 2.17: Spectral clusters in cluster 1 found at time 50 and tracked forward and backward

2.4 Conclusions

In this chapter we have outlined a Lagrangian-particle based technique to gain insight into mixing in non-linear geophysical flows. Our literature survey showed that clustering of particles based on inter-particle distances has been used to characterize mixing from a Lagrangian point of view. Local network measures like node degree and the local clustering coefficient of a particle, employed by previous researchers e.g. [26], gives an idea about the number of other particles a chosen particle has interacted with, or 'neighbours'. We have taken this approach one step further, by finding sub-clusters representing regions of dense interactions. The findings of our work can be partly summarized by **Fig.(2.19)**. In this figure we examine the output time 80, at which the double jet has broken up into a number of quasi-coherent vortices, as well as filaments of vorticity. The enstrophy field, scaled by its maximum, is shown shaded in the Figure, with green dots superimposed to show particles from a few of the largest cumulative clusters. This gives us an indication of particles that have passed through regions where mixing has taken place. The algorithm *Quick* is used

Figure 2.18: Spectral clusters in cluster 2 found at time 50 and shown at other times.

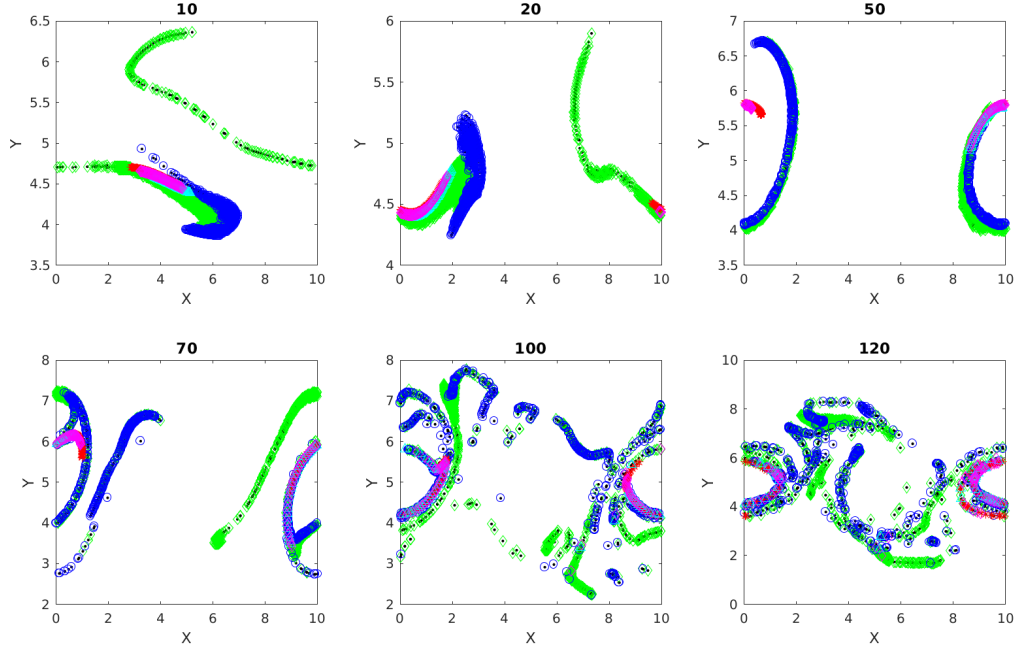to identify subclusters of particles with dense mutual interactions (i.e. strongest mixing). These particles are plotted in blue. These particles, and their path history, identify regions where the degree of mixing is relatively higher (regulated by a density parameter γ) than other portions of the cumulative clusters. In summary, this figure tells us that the outskirts of the large, coherent vortices involve the strongest mixing. The vorticity filaments away from the quasi-coherent vortices are marked as belonging to regions of mixing, but not the strongest mixing. The subclustering method thus provides a way to gain further detail on mixing intensity from a Lagrangian point of view.

We have compared our results with the coherent structures identified by spectral clustering. Spectral clustering shows that the location of the coherent structures is around the vortices, but fails to point out the regions of strong mixing. As discussed in section [2.2.4], the method of finding dense clusters is more precise and robust. We have also computed instantaneous clusters, which as opposed to the cumulative clusters represent regions of interaction for each output time. Instantaneous clusters proved useful in showing that they do not change their paths much during the course of the flow evolution and keep interacting with particles in the same region multiple times, implying dense mixing. This helped us
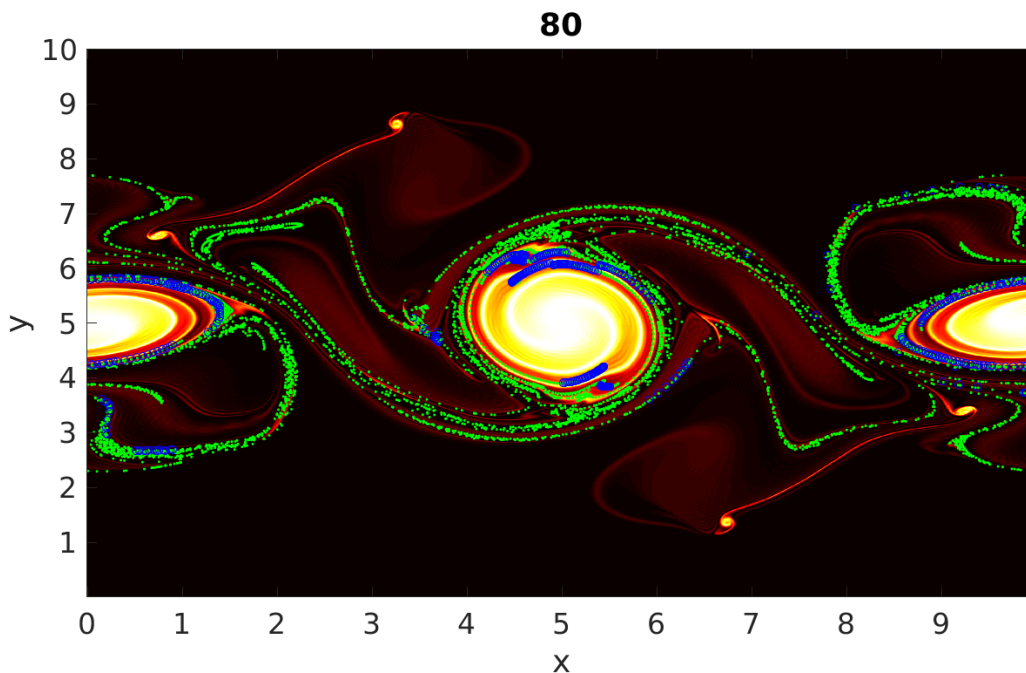
Figure 2.19: Enstrophy field with particles at output time 80. The green dots represent particles from the three largest cumulative clusters and the blue regions represent particles having dense interactions within these cumulative clusters.

validate our method for finding dense subclusters.

Summarizing the major findings in our work, we have seen that the size of cumulative clusters depend on the threshold interaction distance ϵ. In fact previous works like [26] have only used values of ϵ larger than the particle grid spacing, in order to make the entire graph connected and then apply techniques like spectral clustering to extract coherent sets. Our approach, has allowed us to regulate ϵ smaller than the particle grid spacing and observe the differences. We have inferred that, cluster merging is possible beyond a threshold ϵ. Decreasing ϵ less than the threshold corresponds to stronger interactions and hence stronger mixing. Regions of strong and dense mixing show a lot of similarity, which mostly are concentrated along the outskirts of the quasi-coherent vortices implying that coherent behavior can involve a lot of mixing as demonstrated in **Fig.(2.19)**. The highly interactive particles from the dense sub-clusters usually stay as a part of their original coherent vortex. However, interesting dynamics seem to be present when some of these particles deviate out of their usual paths and mix with other regions in the flow as discussed in section [2.2.3]. Even results from spectral clustering show that some particles showing

coherent behaviour may become incoherent over time. The striking similarities between the behaviour of the coherent spectral clusters and the dense subclusters indicate that dense interaction and thereby mixing is a characteristic of coherent structures. A study of the effects of parameters on the dense sub-clustering technique surfaces that ϵ should be chosen as small as sufficient to produce a satisfactory amount of information content about the regions of mixing. Smaller the minimum degree of interaction, stronger is the mixing represented by the mined blue regions. The minimum degree is controlled by parameters min_size and γ, where min_size is really a choice of the user based on the application and γ can be tuned to hit the sweet minimum degree value.

Chapter 3

Conclusions and Future Directions

While Lagrangian methods have a considerable literature, there is limited work on characterizing mixing in geophysical flows using Lagrangian particles. The literature survey in the Introduction of chapter 2 discusses existing methods that try to characterize mixing by detection of coherent structures. Multiple authors have proposed Lagrangian diagnostic scalar fields like FTLE, FSLE, mesochronic analysis, shape coherence etc. for detecting material coherence [11]. The reason these fields are classified as Lagrangian is because their values at a point in space solely depend on the trajectory segment starting at that point measured until some specific later time. Lagrangian clustering techniques also uses trajectory information but create a network of interactions between the trajectories and leverage the network to detect coherent sets. So technically, most existing work represents mixing by studying some characteristics of coherent sets, with details dependent on the mathematical technique used to extract those coherent sets. This thesis also uses trajectory information and a network of interactions like existing works, but claims to provide a new aspect of inspecting the problem of mixing in geophysical flows. The idea of modelling property exchange between material volumes by the introduction of particles and measuring their inter-particle distance is motivated from [14] and used by [26] and others before. The technique borrowed in this work is essentially an unsupervised data mining algorithm without any *a priori* estimate. We try to represent localized mixing by mining densely connected subgraphs (i.e. dense sub-clusters) using the cumulative adjacency matrix, where the probability of property transfer (degree of mixing) in a subgraph is expected to be higher than other regions. This degree of mixing is parameter controlled and has to be tuned by the user specifically for a given application and its requirements. The results of this method have been shown to generally agree with the coherent structures extracted using spectral clustering (a more established method). However, our method is

more specific and consistent about the details of mixing. The dense sub-clusters are also part of the coherent sets because the particles in a sub cluster are mutually interactive (to some extent), and this is only possible if the trajectories of particles in a sub-cluster are coherent.

Even though mixing and transport is an inherently Lagrangian process involving particles, historically mixing has been actively researched using an Eulerian framework. This is because in the past, it was difficult to perform simulations with Lagrangian particles due to scarcity of computational resources. [27] studies mixing efficiency of stratified shear flows in the vertical plane as opposed our study in the horizontal plane. [33] develop Eulerian indicators to quantify events like streamline crossing and shear rate and multiply them together and call the resulting measure the transverse shear rate. They use blinking flows or consider for example two time snapshots of a flow. Streamline crossing, does not refer to streamlines at a a single time, and is instead related to the change in angle of the velocity field along the streamline at a point in space across the time snapshots and shear rate is technically the difference between velocity vectors around the point in consideration. The authors use these quantities and average them over the entire domain to obtain a single value for the flow. They then measure the quality of mixing by measuring the variance in concentration over a given domain. The lower the variance in concentration, the better is the mixing ([20]).The authors claim that the transverse shear rate quantity shows a strong correlation with good/strong mixing. However, measuring variance of concentration can give us only an overall idea about the quality of mixing in the flow. Moreover, the authors try to skip the logistic of measuring concentration and guess the quality of mixing from the Eulerian indicators proposed by them. With the increasing ease of access to computational power, we show in this work that studying Lagrangian particles is a promising way to gain much more detailed and specific information about mixing. However we can definitely learn from the past. For e.g. while dealing with a real world flow where we study the concentration of certain properties, we can record the changes in variance of concentration in a γ-quasi clique to quantify the extent of property exchange in it.

There are improvements that can be made to this algorithm to decode more information about localized mixing. Firstly, in the cumulative adjacency matrix, we take into account only one interaction between the particles. Technically we wish to model localized mixing by diffusion through our γ-quasi cliques. Hence it is useful to include the information that how many times two particles come close to or interact with each other. This number is a discrete way of representing how long two particles stay with each other. This can be incorporated by adding weights to the edges between the nodes in the graph i.e. allowing the adjacency matrix to record the number of interactions, as opposed to just a zero or a one as shown in **Fig.(3.1)**
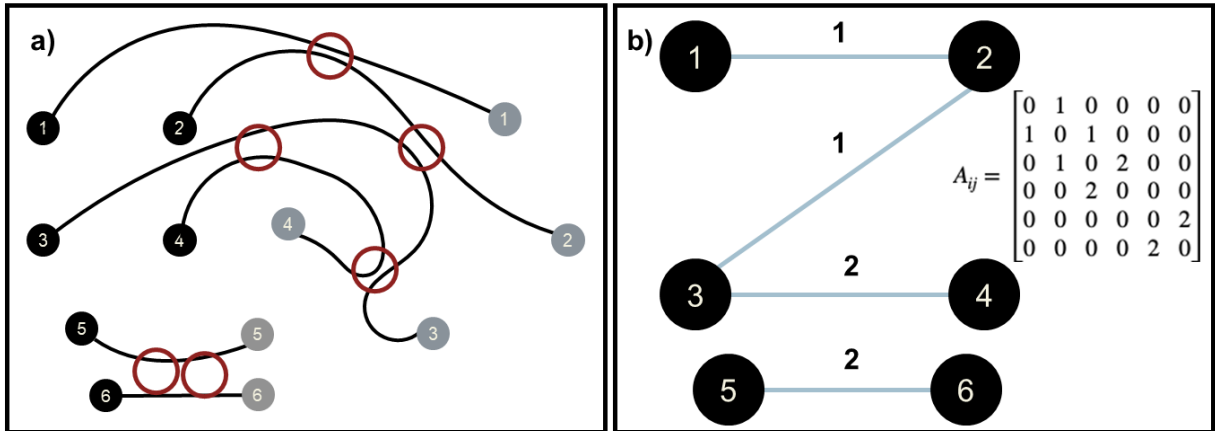
Figure 3.1: a) Illustrative particle trajectories with red circles denoting particle inter-actions. b) Corresponding graph and adjacency matrix with edge weights and entries, respectively, equal to number of particle interactions.

The higher the number of interactions, the larger the amount of properties (e.g. colour) exchanged between the particles. Hence we decide to add the weights of the edges in a dense sub-cluster. This sum, may be normalized by the maximum sum of weights in the simulation, and can be used to convey the degree of mixing in a dense sub-cluster. We hypothesize that mixing form a Lagrangian perspective is proportional to the number of interactions and variance in concentration as mentioned before. This thesis uses a prototype simulation, where we do not consider any concentration, hence the next step is to allocate the dense sub-clusters their sum of edge weights. Once particles belong to a dense cluster, they lose their individual identity and share the identity of the dense cluster. For this

reason, the sum of edge weights becomes an attribute of a particular dense cluster. We can then rank these dense sub-clusters based on sum of the weights. Subsequently, we can plot these dense clusters for all times to see how their sum of weights and spatial positions evolve with time. The sub-clusters with higher sum of weights will naturally have particles with more and longer coherent trajectories and vice versa. We show an illustration of this potential future result in **Fig.(3.2)**. Here we show two imaginary cumulative clusters and then their dense sub-clusters contained in dotted ellipses resulting from some flow. The yellow particles represent those which are not part of the dense sub-clusters. The size of the dense sub-clusters has been kept same for the sake of comparison and illustration. As mentioned before, during implementation, we need to normalize the sum of weights so that it is on the same scale for all the dense sub-clusters. The colorbar shows how we can demonstrate the extent of mixing, inferred from the sum of edge weights, with gradient colors. We assume while modelling we know nothing about the properties we wish to study the distribution of. Hence, it is also possible to model property exchanges between the nodes based on the edge weights and some stochastic process.

We would also like to be able to reuse the code *Quick*, e.g. at multiple output times. In the long run, we would even like to automate the subclustering to work in real time as part of the simulation. A possible avenue of future work is thus to investigate how to parallelize the *Quick* algorithm using CUDA, MPI or OPenMP. The aim is to tackle the exponentially increasing time complexity with increase in the size of the network. Our group has already used CUDA to perform numerical simulations of the flow in consideration here. Additionally, computing the inter-particle interactions for every time-step is an expensive step and becomes difficult to handle with increasing numbers of particles. In the future, it is desirable to perform simulations with an increased number of particles to capture the property exchange phenomenon (diffusion) with more detail. This portion of the code has been parallelized in CUDA as part of a SHARCNET Dedicated Programming Time award (Pawel Pomoroski from Sharcent worked with Aaron Coutino in our group). The paralelization strategy puts the particles in square buckets of size equal to the threshold distance ϵ. For every particle, we thus calculate its distance from particles in the same and neighbouring buckets and we do this for every particle in parallel. This improvement gives us about a 100 time improvement in speed over the serial brute force approach, and hence saves us a lot of time. We envision applying the *Quick* algorithm in parallel to every particle in the future in a similar fashion.

Another possibility for future work is to draw mathematical relationships between the traditional coherent sets (like spectral clusters) and the dense sub-graphs. We have seen in the results section that coherent sets from spectral clusters often also contains particles which are part of the dense sub-clusters. This is due to the coherent nature of the dense
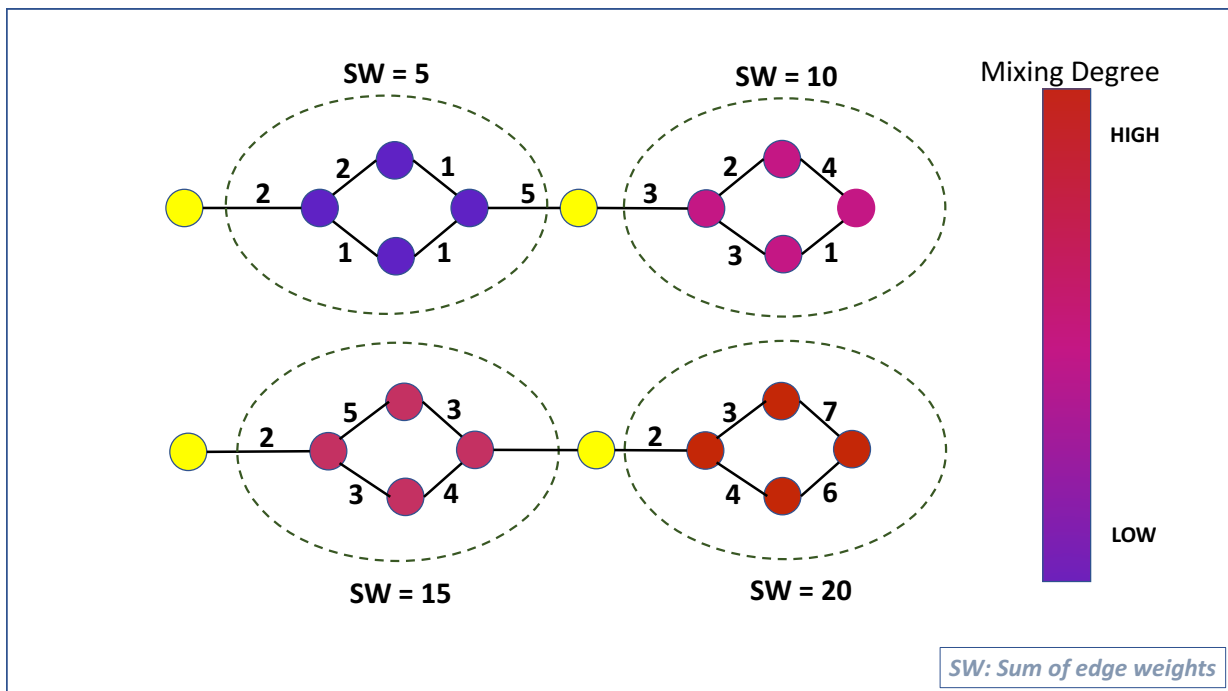
51

Figure 3.2: Illustration showing sum of edge weights concept to measure degree or extent of mixing in a dense sub-cluster. The colormap shows clusters with highest degree or sum of weights are colored red while those with lowest are colored blue.

clusters as well. We want to establish if this is consistent across flows and coherent set extraction techniques computationally and mathematically in the future.

On the application side, the central future challenge is how to appropriately think of particles, and hence Lagrangian based mixing ideas, in more complex models. For example should particles migrate across isopycnal layer boundaries in multi-layer models? This question opens the door to more philosophical inquiries, such as the inherent resolution limitations of large scale models, and what "mixing" means in a context other than that of three-dimensional, direct numerical simulation.

References

[1] Michael R Allshouse and Thomas Peacock. Lagrangian based methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097617, 2015.

[2] Uri M Ascher and Linda R Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations*, volume 61. Siam, 1998.

[3] Rishiraj Chakraborty, Aaron Coutino, and Marek Stastna. Particle clustering and subclustering as a proxy for mixing in geophysical flows.

[4] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[5] Peter Davidson. *Turbulence: an introduction for scientists and engineers*. Oxford University Press, 2015.

[6] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2):298–305, 1973.

[7] Gary Froyland. An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. *Physica D: Nonlinear Phenomena*, 250:1–19, 2013.

[8] Gary Froyland. Dynamic isoperimetry and the geometry of lagrangian coherent structures. *Nonlinearity*, 28(10):3587, 2015.

[9] Gary Froyland and Kathrin Padberg-Gehle. A rough-and-ready cluster-based approach for extracting finite-time coherent sets from sparse and incomplete trajectory data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(8):087406, 2015.

[10] Gary Froyland, Naratip Santitissadeekorn, and Adam Monahan. Transport in time-dependent dynamical systems: Finite-time coherent sets. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):043116, 2010.

[11] Alireza Hadjighasem, Mohammad Farazmand, Daniel Blazevski, Gary Froyland, and George Haller. A critical comparison of lagrangian methods for coherent structure detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(5):053104, 2017.

[12] Alireza Hadjighasem, Daniel Karrasch, Hiroshi Teramoto, and George Haller. Spectral-clustering approach to lagrangian vortex detection. *Physical Review E*, 93(6):063107, 2016.

[13] A. K. M. F. Hussain. Coherent structures?reality and myth. *Physics of Fluids*, 26(10):2816, 6 1983.

[14] AY Klimenko. Lagrangian particles with mixing. i. simulating scalar transport. *Physics of Fluids*, 21(6):065101, 2009.

[15] Stephen J Kline, William C Reynolds, FA Schraub, and PW Runstadler. The structure of turbulent boundary layers. *Journal of Fluid Mechanics*, 30(4):741–773, 1967.

[16] Pijush K Kundu, IM Cohen, and HH Hu. Fluid mechanics. 2004. *Elsevier Academic Press, San Diego). Two-and three-dimensional self-sustained flow oscillations*, 307:471–476, 2008.

[17] Guimei Liu and Limsoon Wong. Effective pruning techniques for mining quasi-cliques. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 33–49. Springer, 2008.

[18] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[19] AM Mancho, D Small, and S Wiggins. Computation of hyperbolic trajectories and their stable and unstable manifolds for oceanographic flows represented as data sets. *Nonlinear Processes in Geophysics*, 11(1):17–33, 2004.

[20] Kevin L McIlhany, Stephen Guth, and Stephen Wiggins. Lagrangian and eulerian analysis of transport and mixing in the three dimensional, time dependent hills spherical vortex. *Physics of Fluids*, 27(6):063603, 2015.

[21] Carolina Mendoza and Ana M Mancho. The lagrangian description of aperiodic flows: a case study of the kuroshio current. *arXiv preprint arXiv:1006.3496*, 2010.

[22] Ronald E Mickens. *Applications of nonstandard finite difference schemes*. World Scientific, 2000.

[23] Ruty Mundel, Erick Fredj, Hezi Gildor, and Vered Rom-Kedar. New lagrangian diagnostics for characterizing fluid flow mixing. *Physics of Fluids*, 26(12):126602, 2014.

[24] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with cuda. In *ACM SIGGRAPH 2008 classes*, page 16. ACM, 2008.

[25] CUDA Nvidia. Cufft library, 2010.

[26] Kathrin Padberg-Gehle and Christiane Schneide. Network-based study of Lagrangian transport and mixing. *Nonlinear Processes in Geophysics*, 24(4):661, 2017.

[27] WR Peltier and CP Caulfield. Mixing efficiency in stratified shear flows. *Annual review of fluid mechanics*, 35(1):135–167, 2003.

[28] SV Prants. Chaotic lagrangian transport and mixing in the ocean. *The European Physical Journal Special Topics*, 223(13):2723–2743, 2014.

[29] Kenneth A Rose, Jerome Fiechter, Enrique N Curchitser, Kate Hedstrom, Miguel Bernal, Sean Creekmore, Alan Haynie, Shin-ichi Ito, Salvador Lluch-Cota, Bernard A Megrey, et al. Demonstration of a fully-coupled end-to-end model for small pelagic fish using sardine and anchovy in the california current. *Progress in Oceanography*, 138:348–380, 2015.

[30] Irina I Rypina and Lawrence J Pratt. Trajectory encounter volume as a diagnostic of mixing potential in fluid flows. 2017.

[31] Rick Salmon. *Lectures on geophysical fluid dynamics*. Oxford University Press, 1998.

[32] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Departmental Papers (CIS)*, page 107, 2000.

[33] Rob Sturman and Stephen Wiggins. Eulerian indicators for predicting and optimizing mixing quality. *New Journal of Physics*, 11(7):075031, 2009.

[34] Zhiping Zeng, Jianyong Wang, Lizhu Zhou, and George Karypis. Coherent closed quasi-clique discovery from large dense graph databases. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 797–802. ACM, 2006.