

Topological Data Analysis

MATH 800

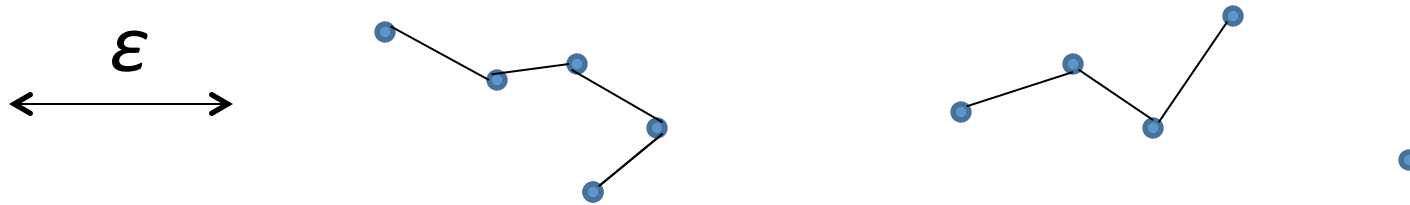
Fall 2011

Topological Data Analysis (TDA)

- An ε -chain is a finite sequence of points x_1, \dots, x_n such that $|x_i - x_{i+1}| < \varepsilon$ for all $i=1, \dots, n-1$
- x and y are ε -connected if there is an ε -chain joining them.
- A ε -connected set is a set that any two points can be linked by an ε -chain.
- A ε -connected component is a maximal ε -connected subset.

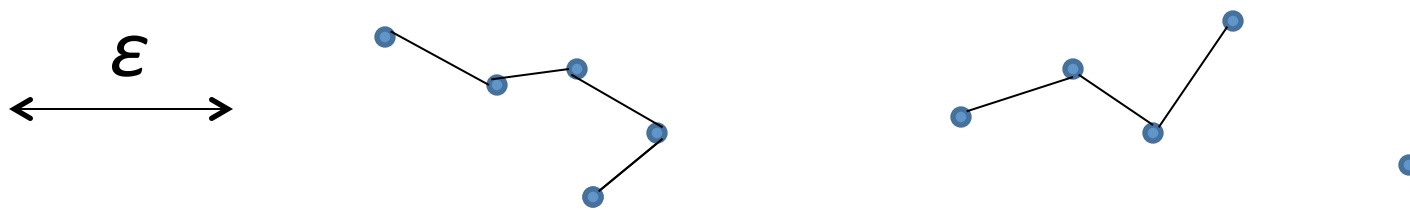
Topological Data Analysis (TDA)

- An ε -chain is a finite sequence of points x_1, \dots, x_n such that $|x_i - x_{i+1}| < \varepsilon$ for all $i=1, \dots, n-1$
- x and y are ε -connected if there is an ε -chain joining them.
- A ε -connected set is a set that any two points can be linked by an ε -chain.
- A ε -connected component is a maximal ε -connected subset.



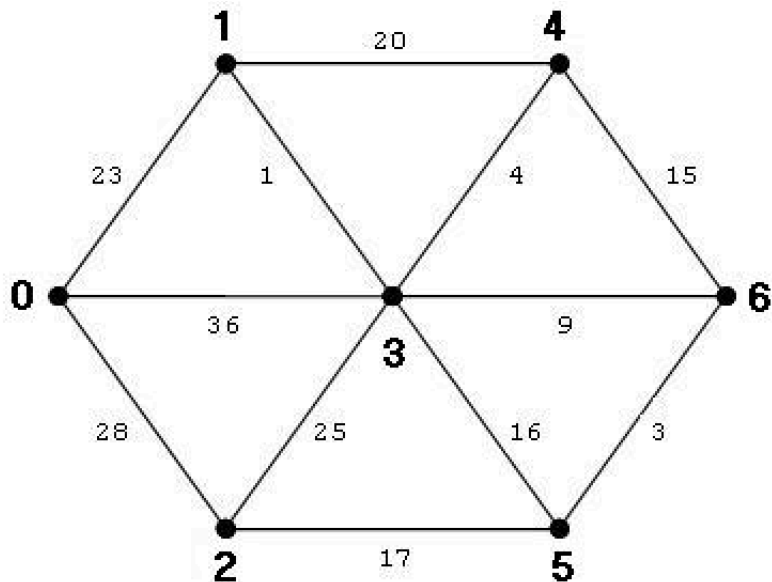
Topological Data Analysis (TDA)

- C_ε = the number of ε -connected components
- D_ε = the maximum distance of the points in ε -connected components
- I_ε = the number of component consisting of a single point (*isolated points*)



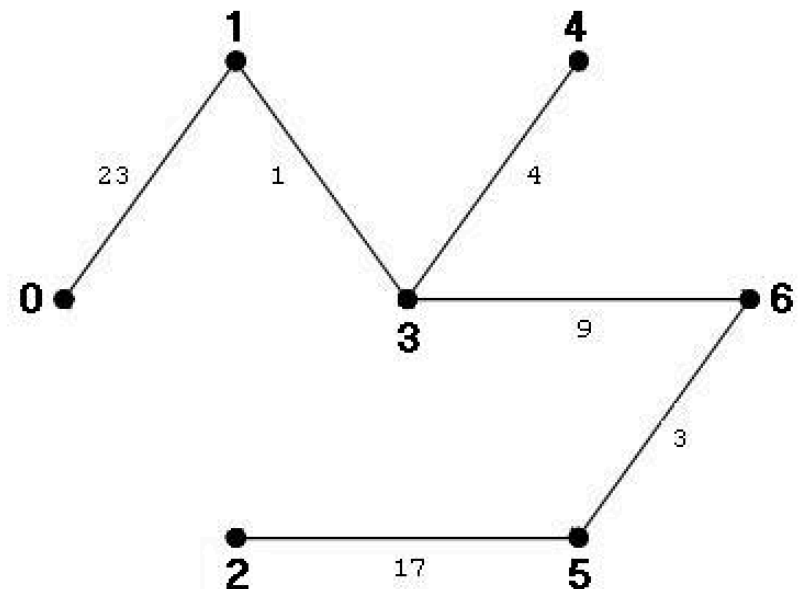
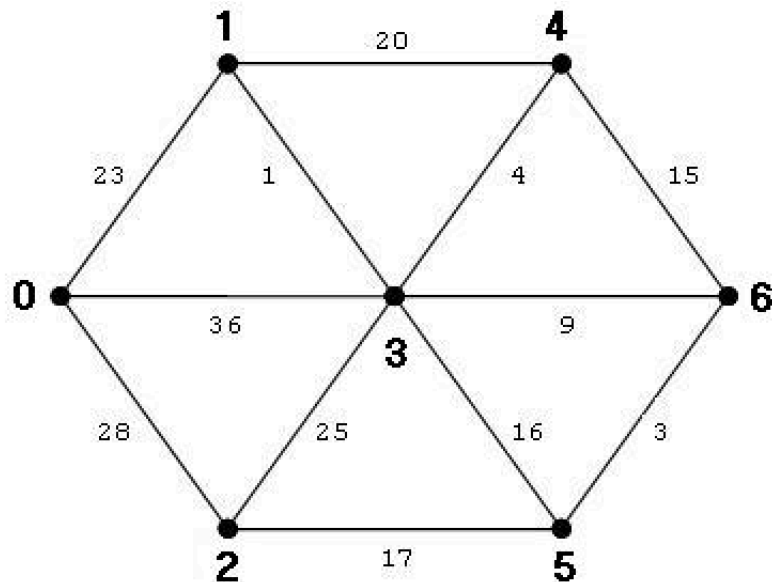
Minimal Spanning Tree (MST)

- MST is the tree of minimum total branch length that spans the data.
- To construct the MST (Prim's algorithm)
 - Start with any point and its nearest neighbor,
 - Add the closest point
 - Repeat until all points are in the tree



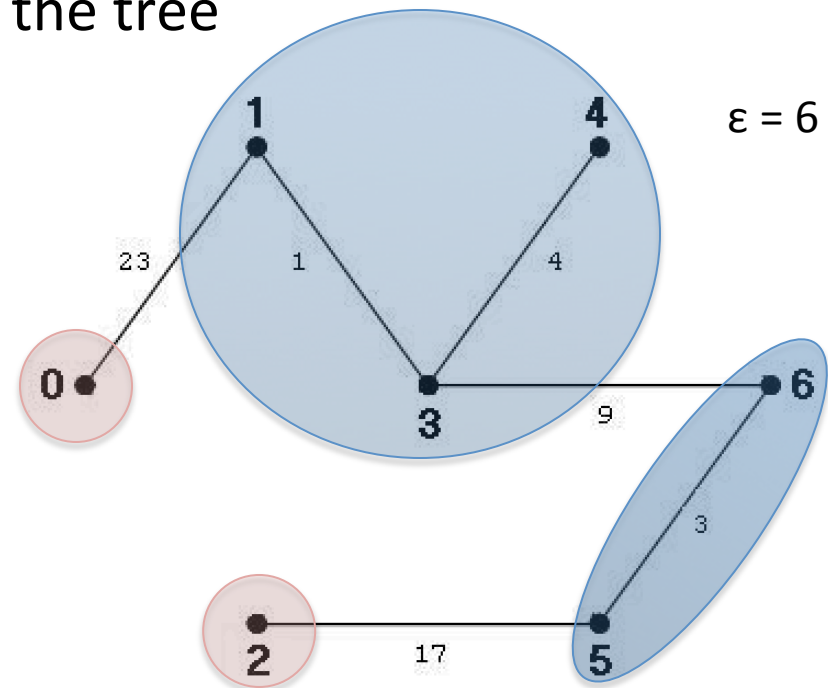
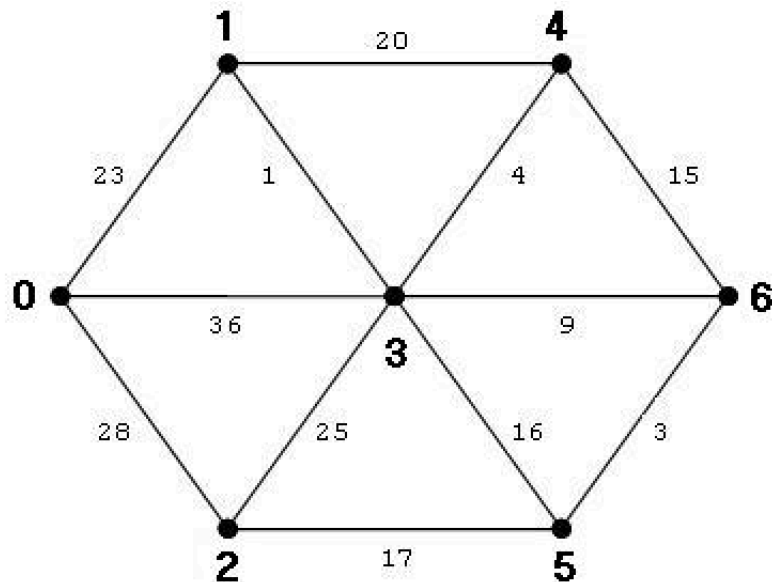
Minimal Spanning Tree (MST)

- MST is the tree of minimum total branch length that spans the data.
- To construct the MST (Prim's algorithm)
 - Start with any point and its nearest neighbor,
 - Add the closest point
 - Repeat until all points are in the tree

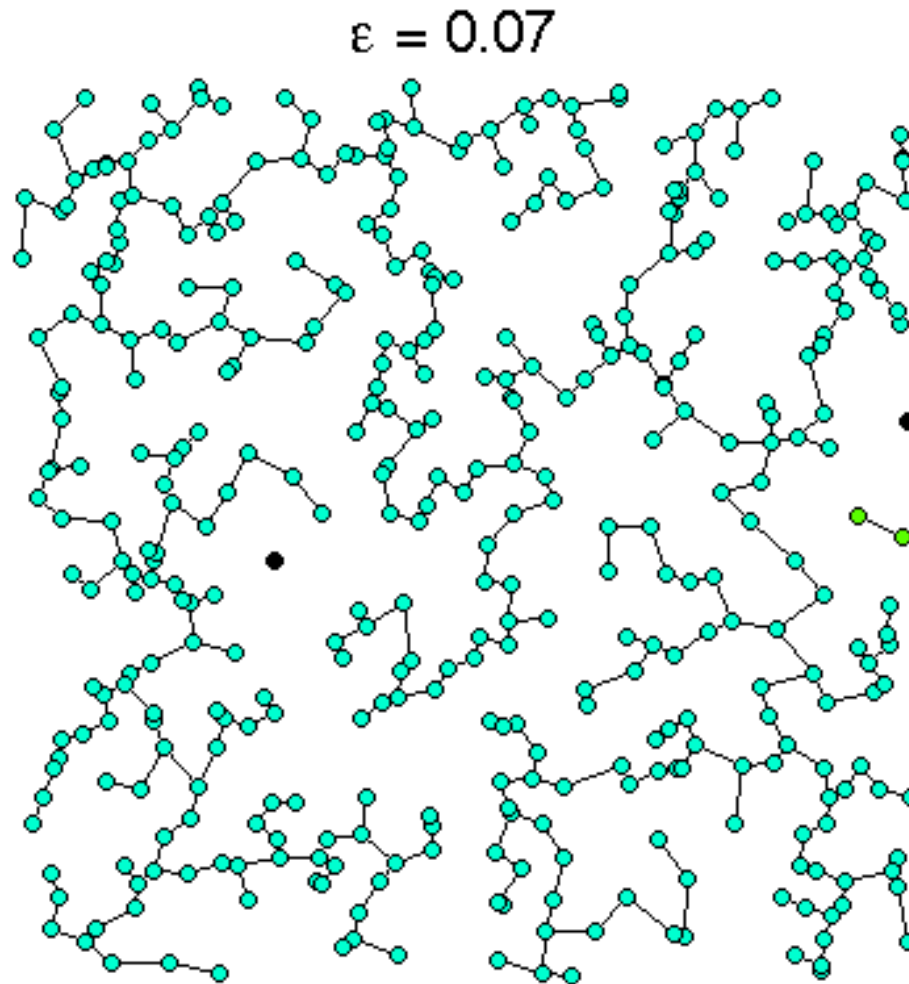


Minimal Spanning Tree (MST)

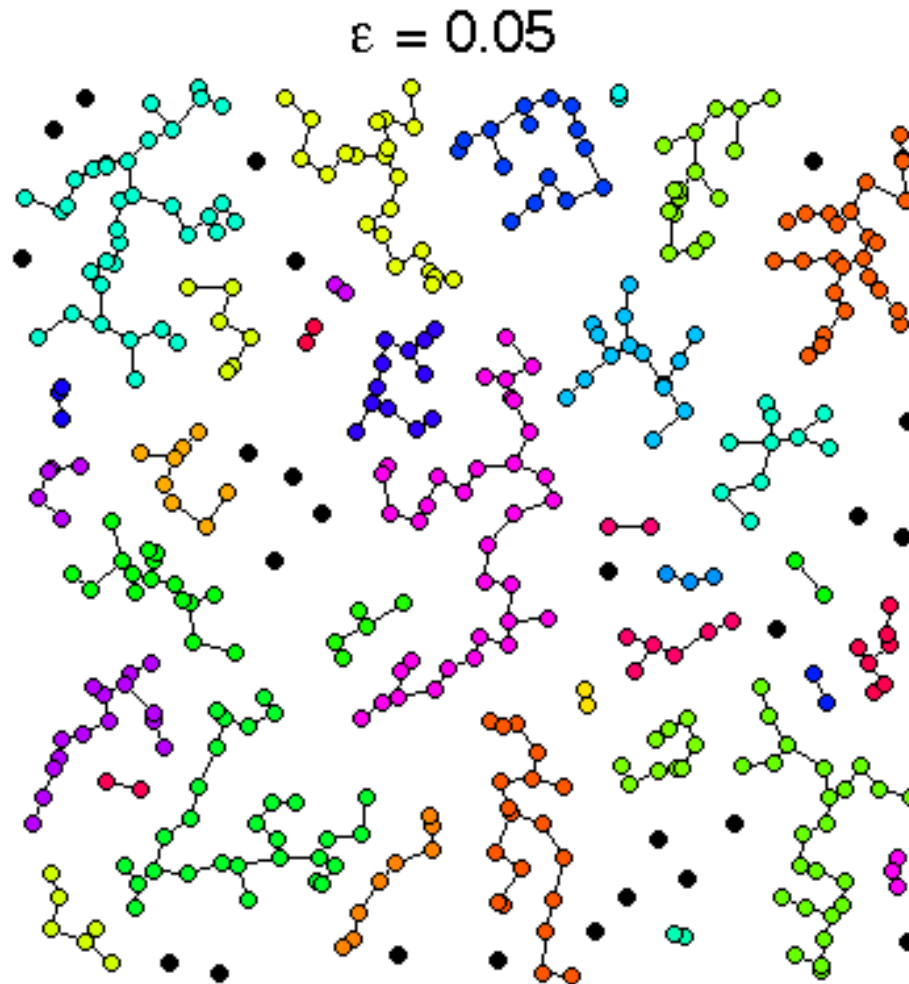
- MST is the tree of minimum total branch length that spans the data.
- To construct the MST (Prim's algorithm)
 - Start with any point and its nearest neighbor,
 - Add the closest point
 - Repeat until all points are in the tree



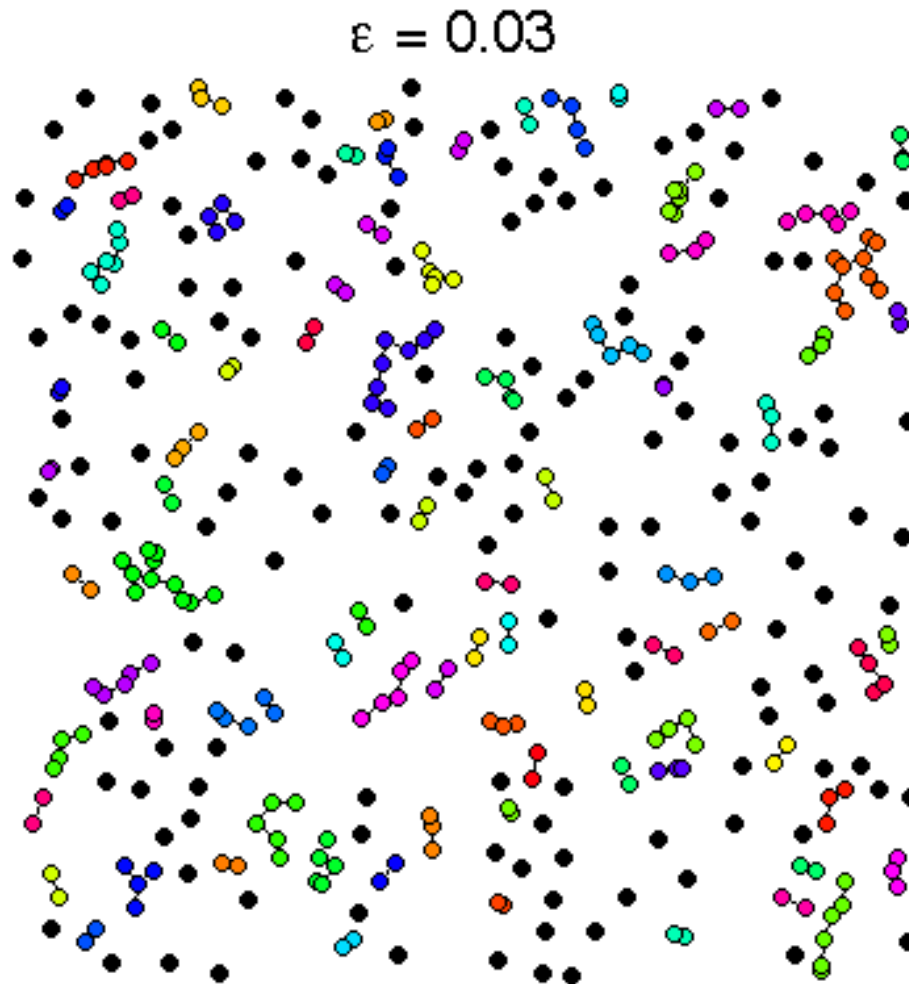
Topological Data Analysis (TDA)



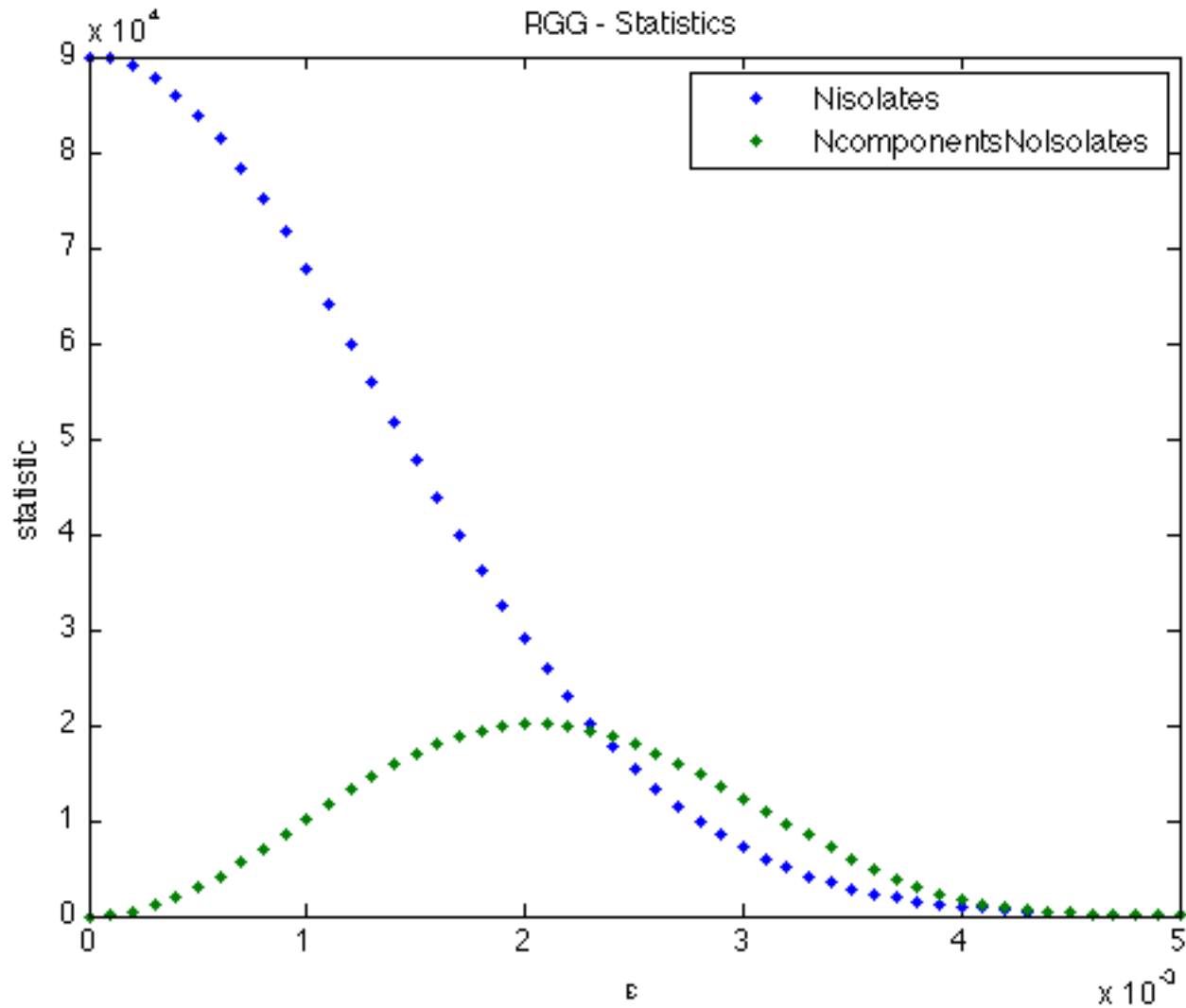
Topological Data Analysis (TDA)



Topological Data Analysis (TDA)

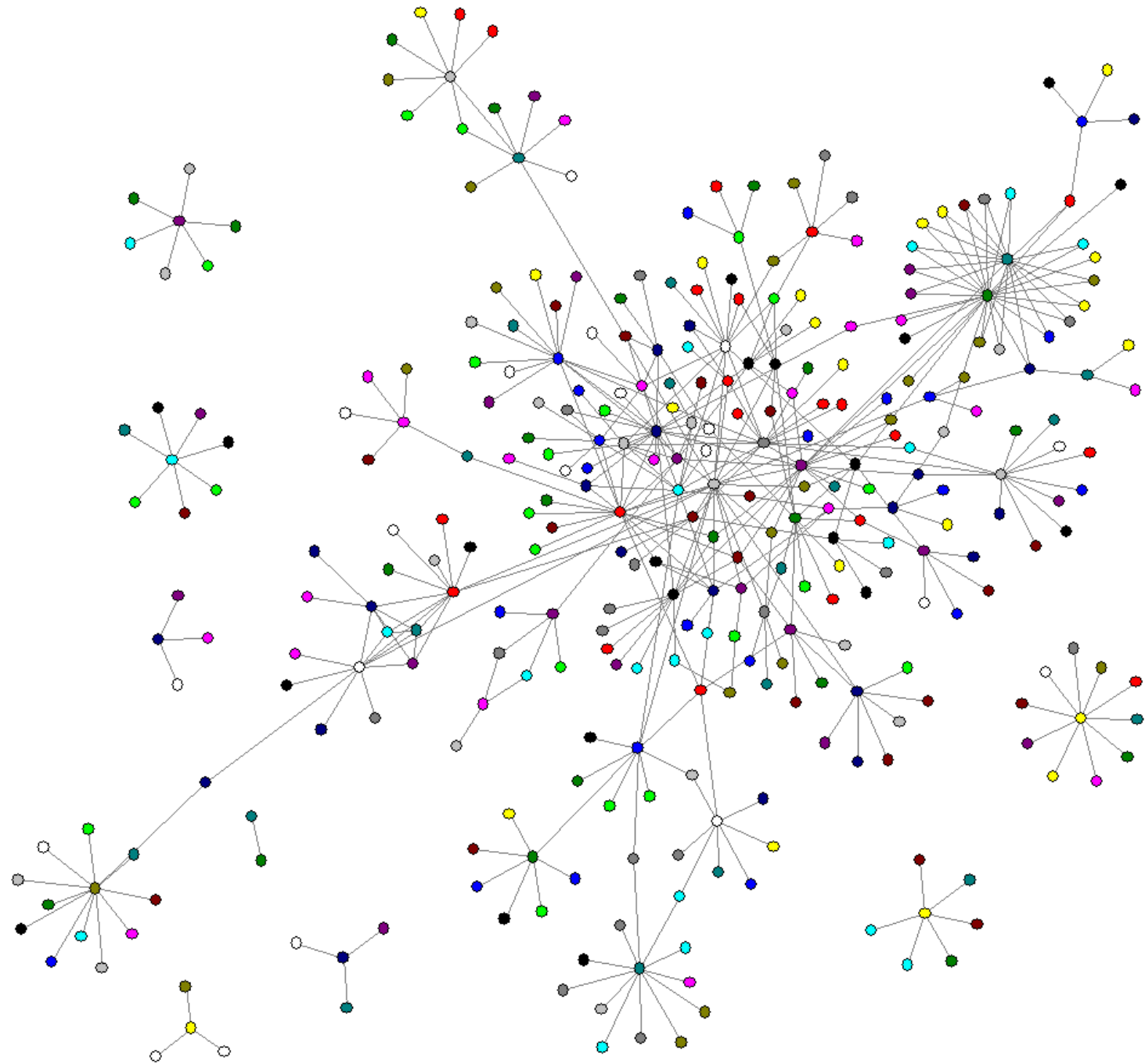


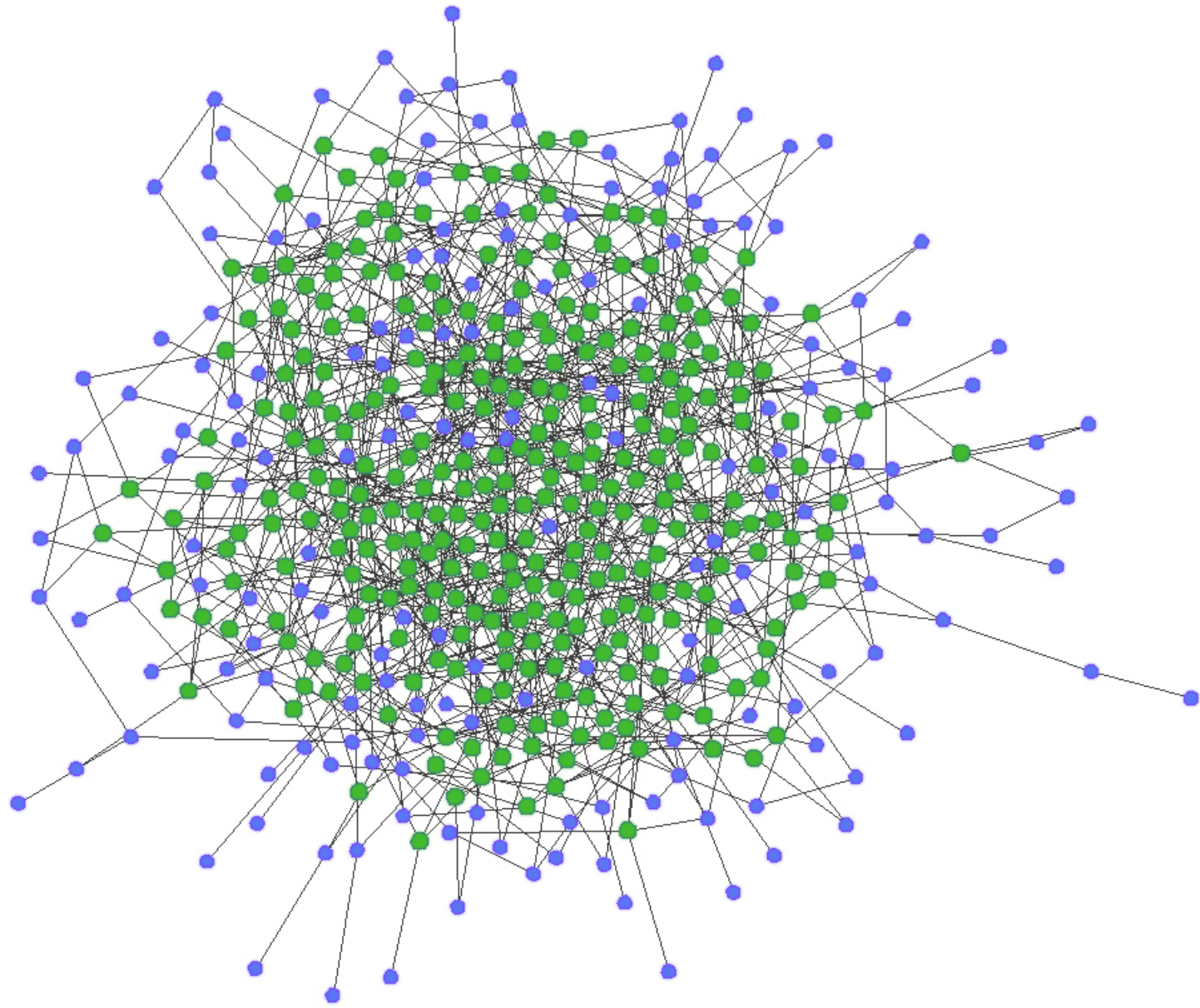
Topological Data Analysis (TDA)

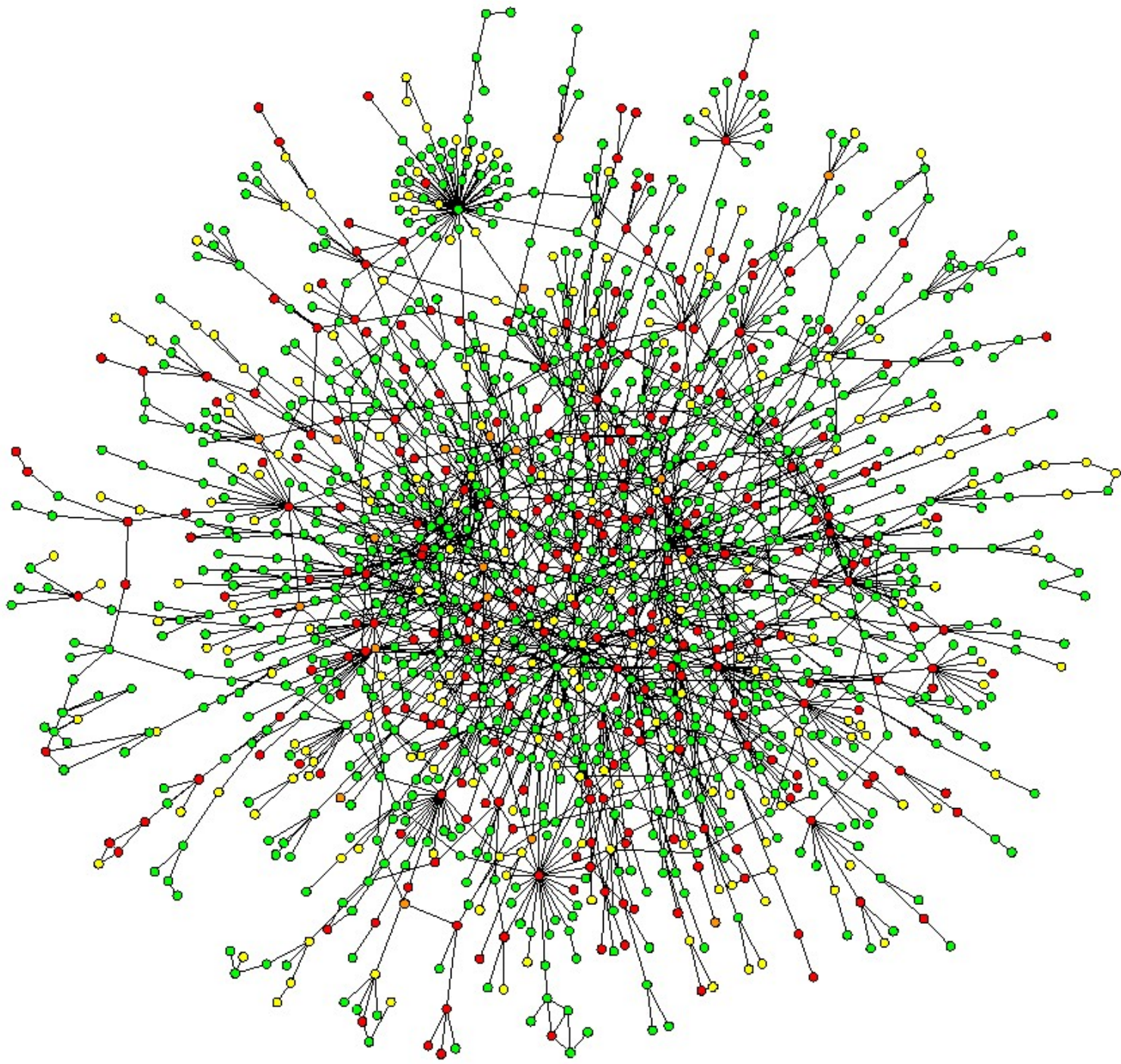


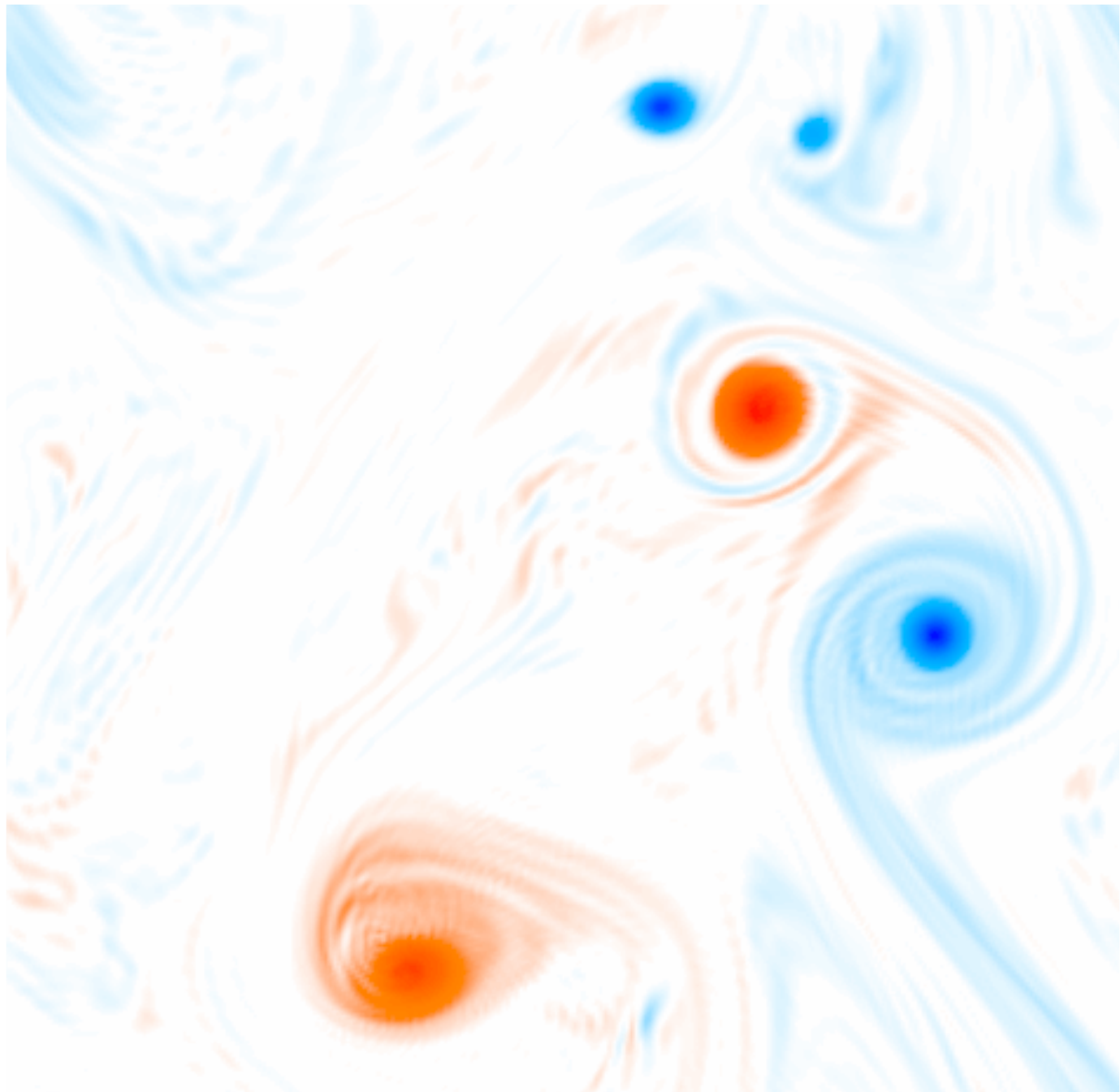
Applications of TDA

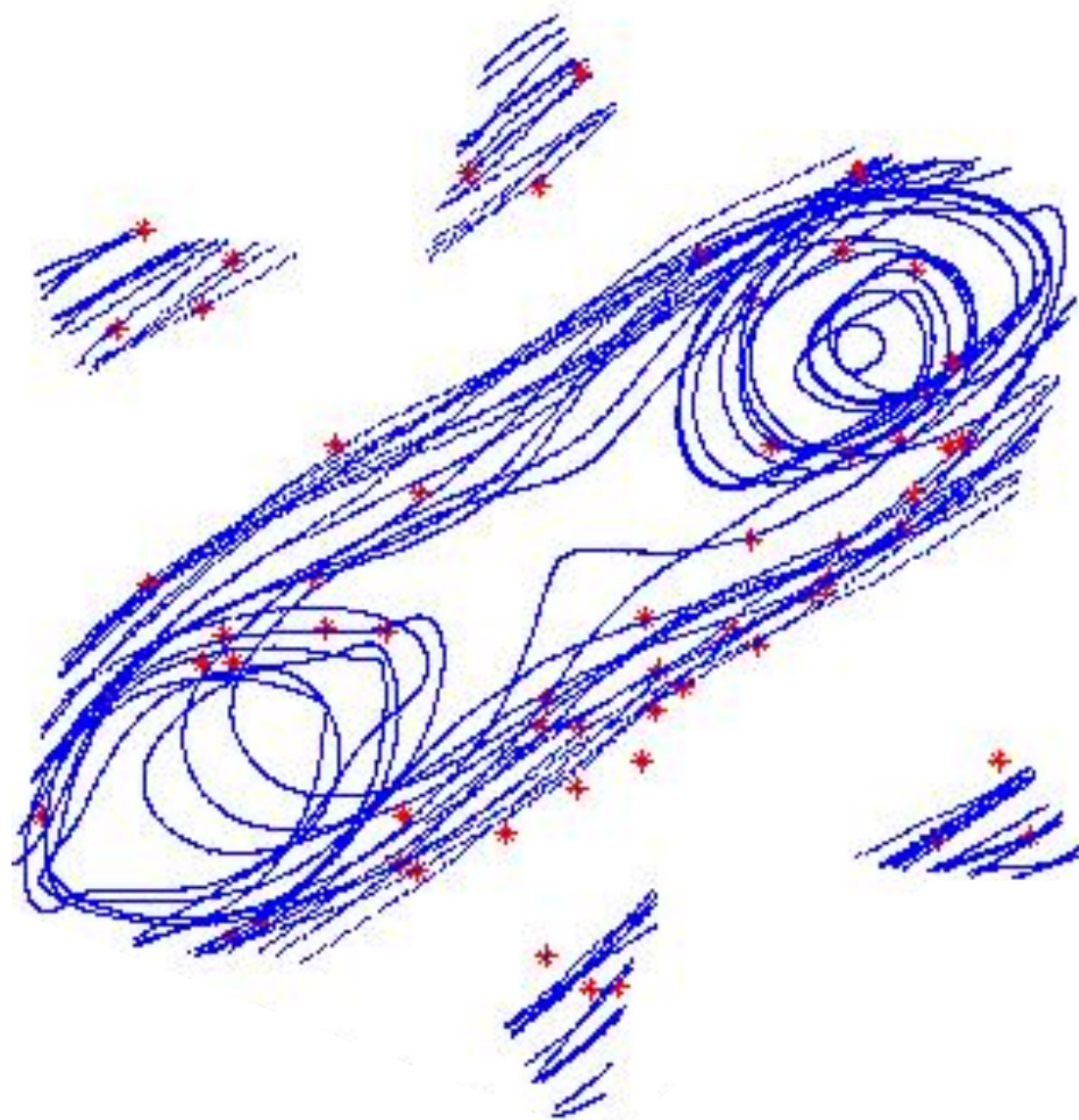
- **Cluster analysis:** is the task of assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters.
 - *connected components*
- **Outlier:** is an observation that is numerically distant from the rest of the data.
- **Noise:** is an unwanted value in a data
 - *isolated points*

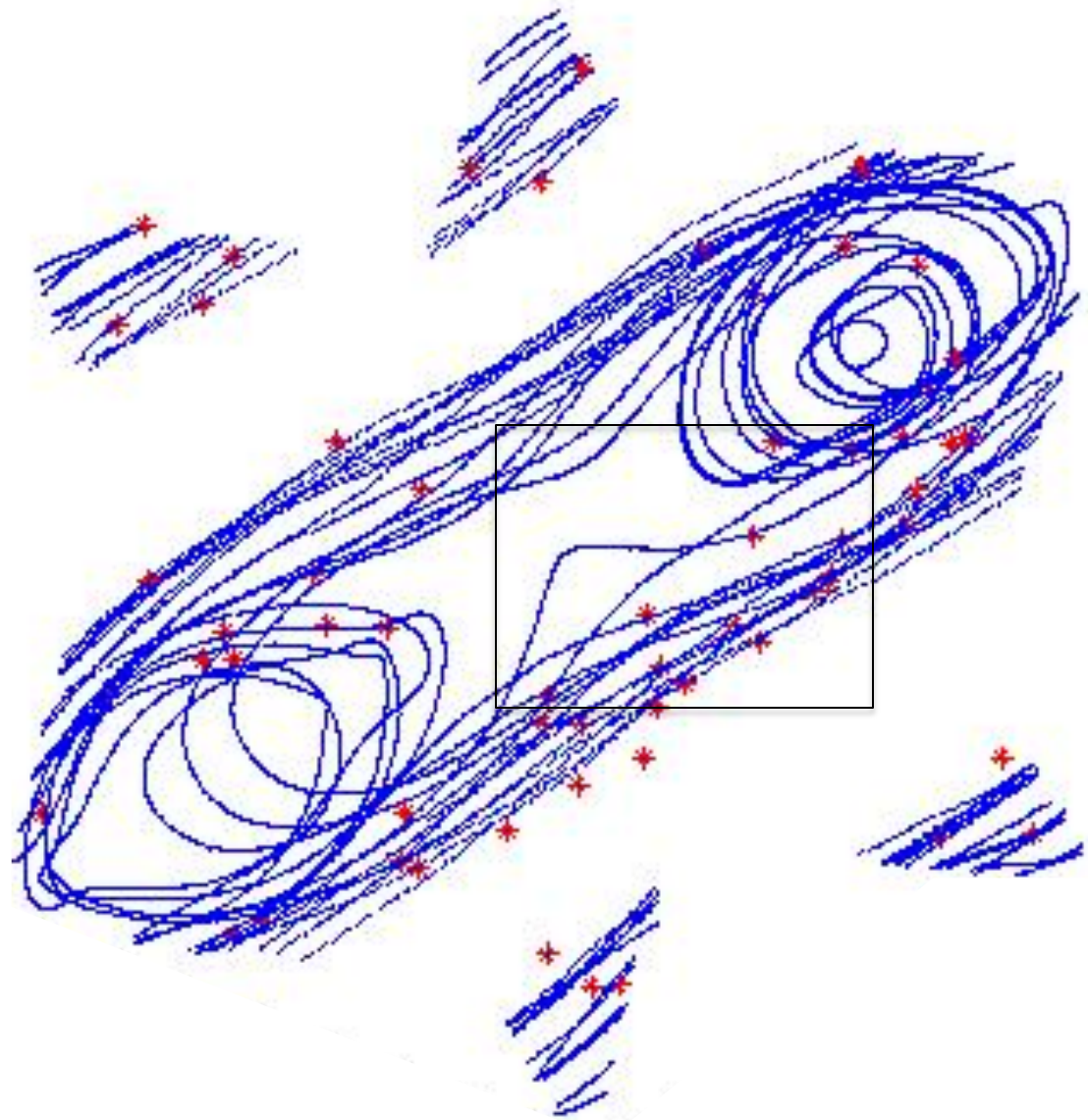


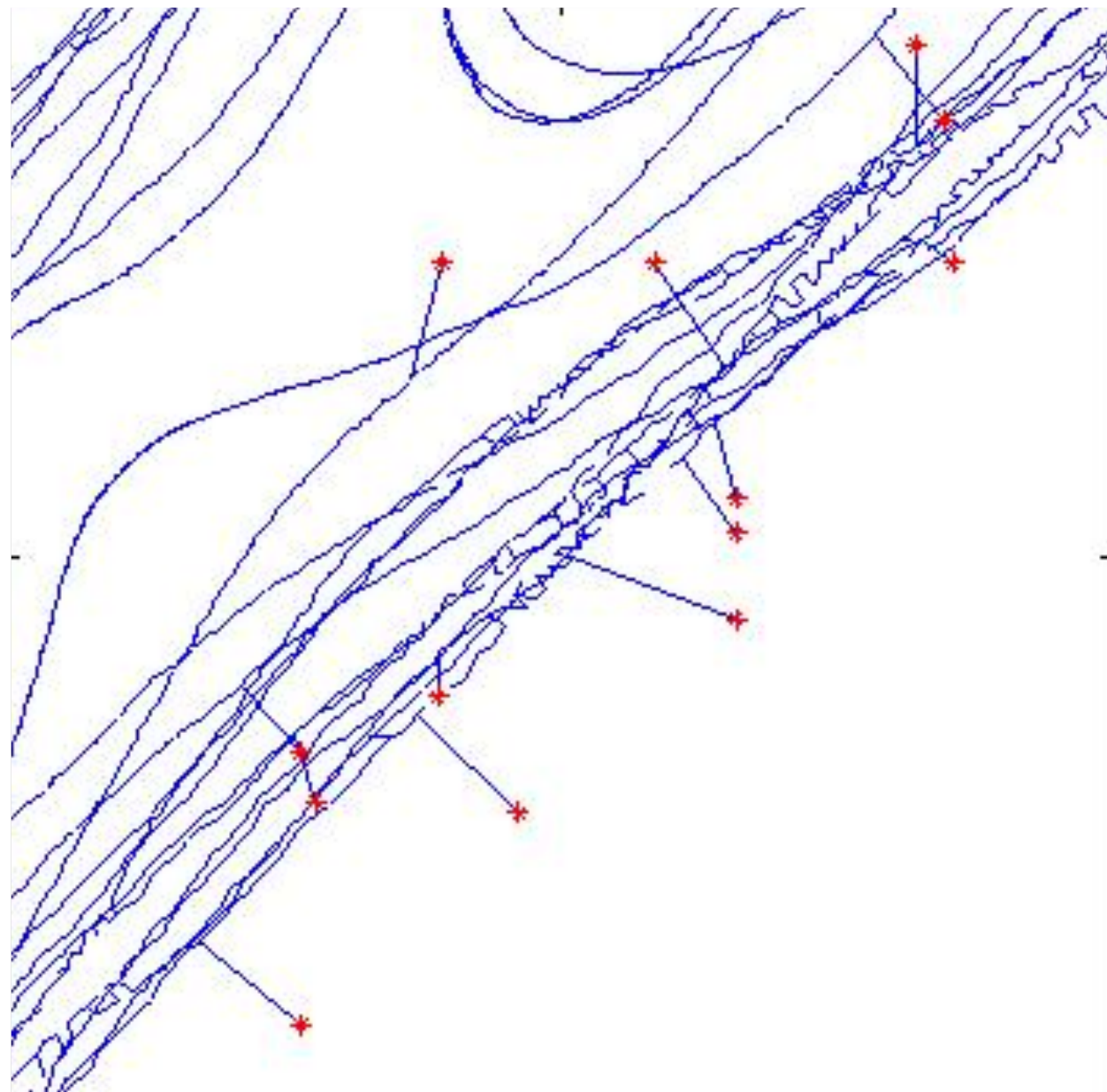




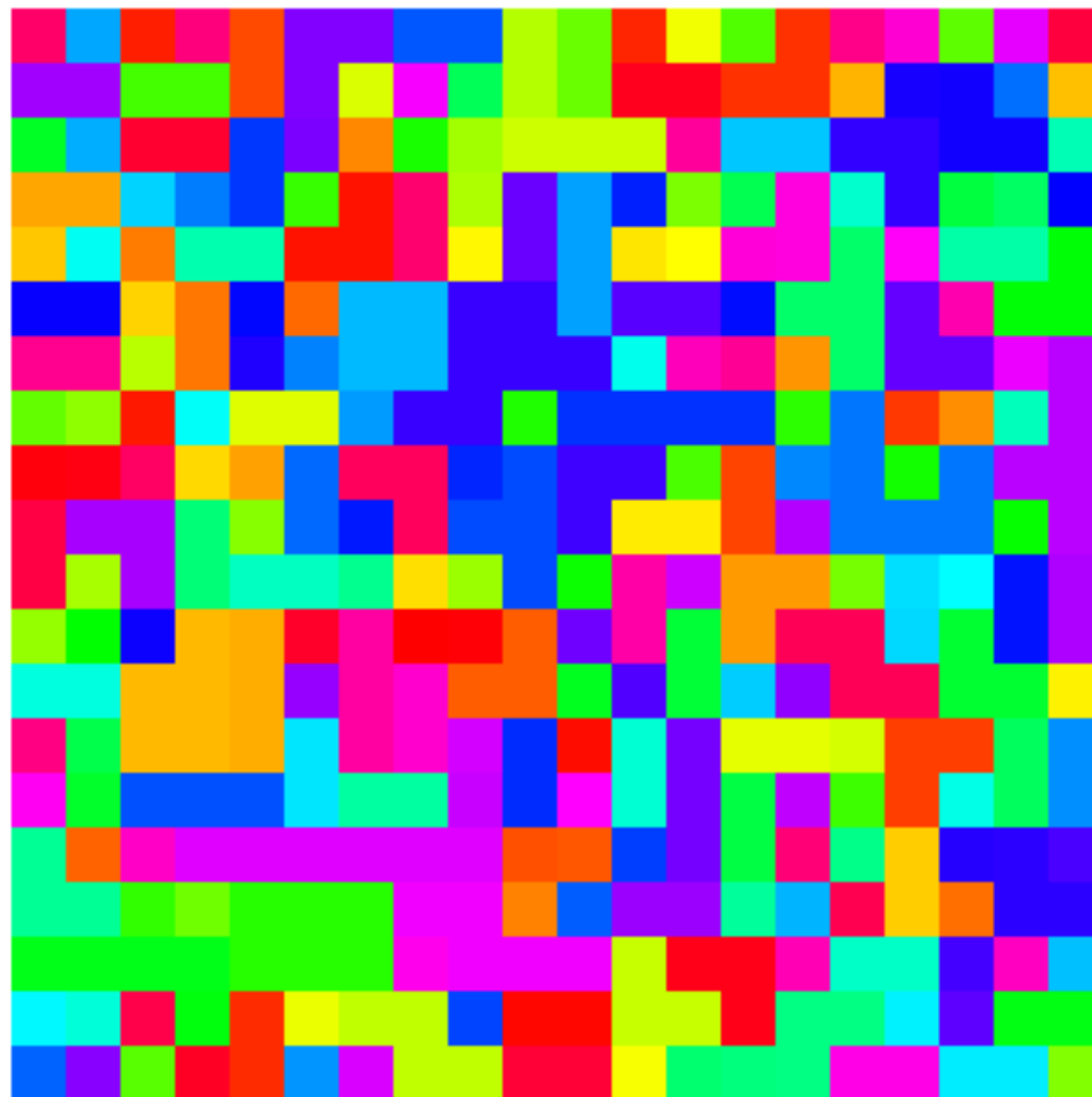




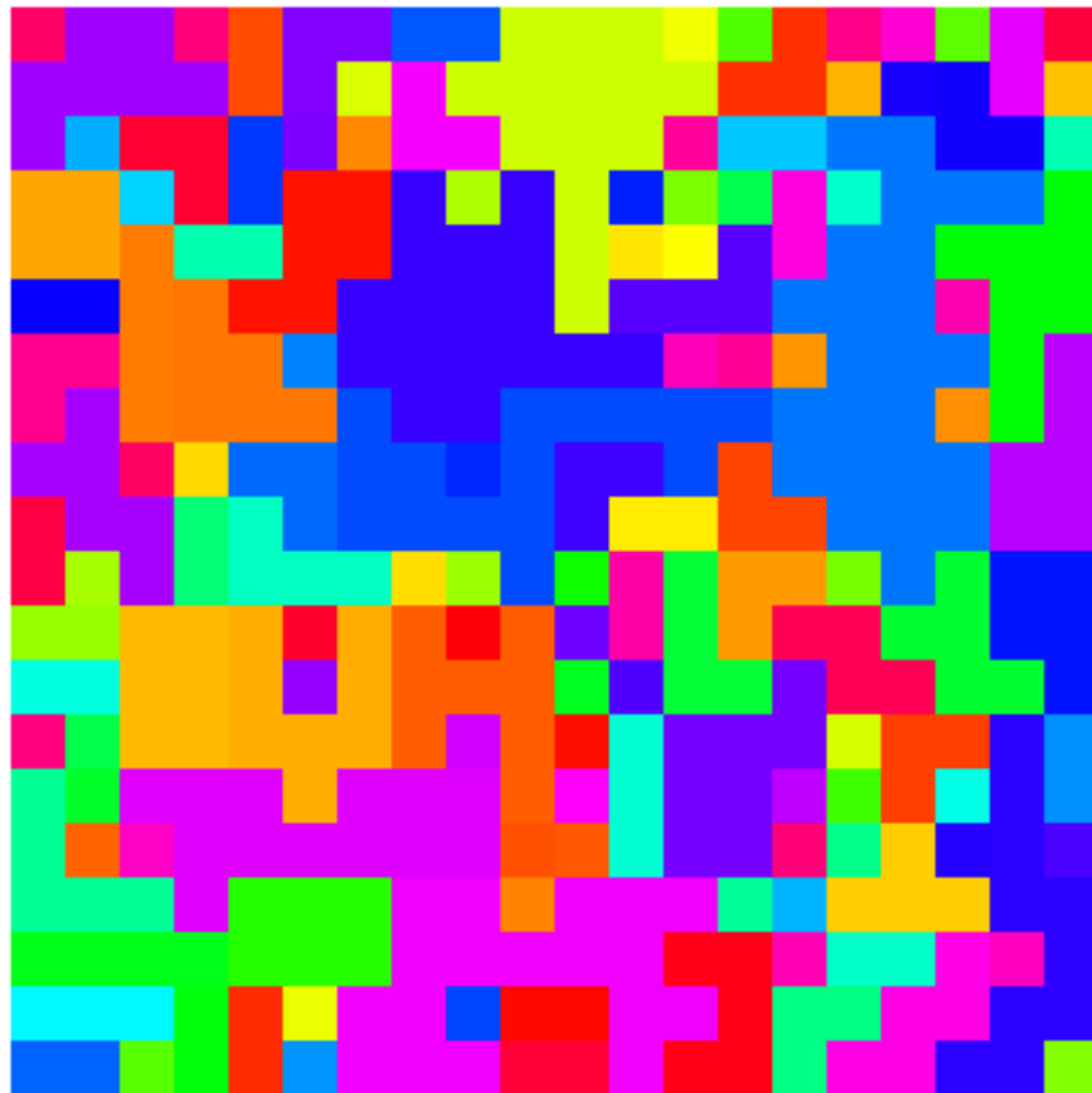




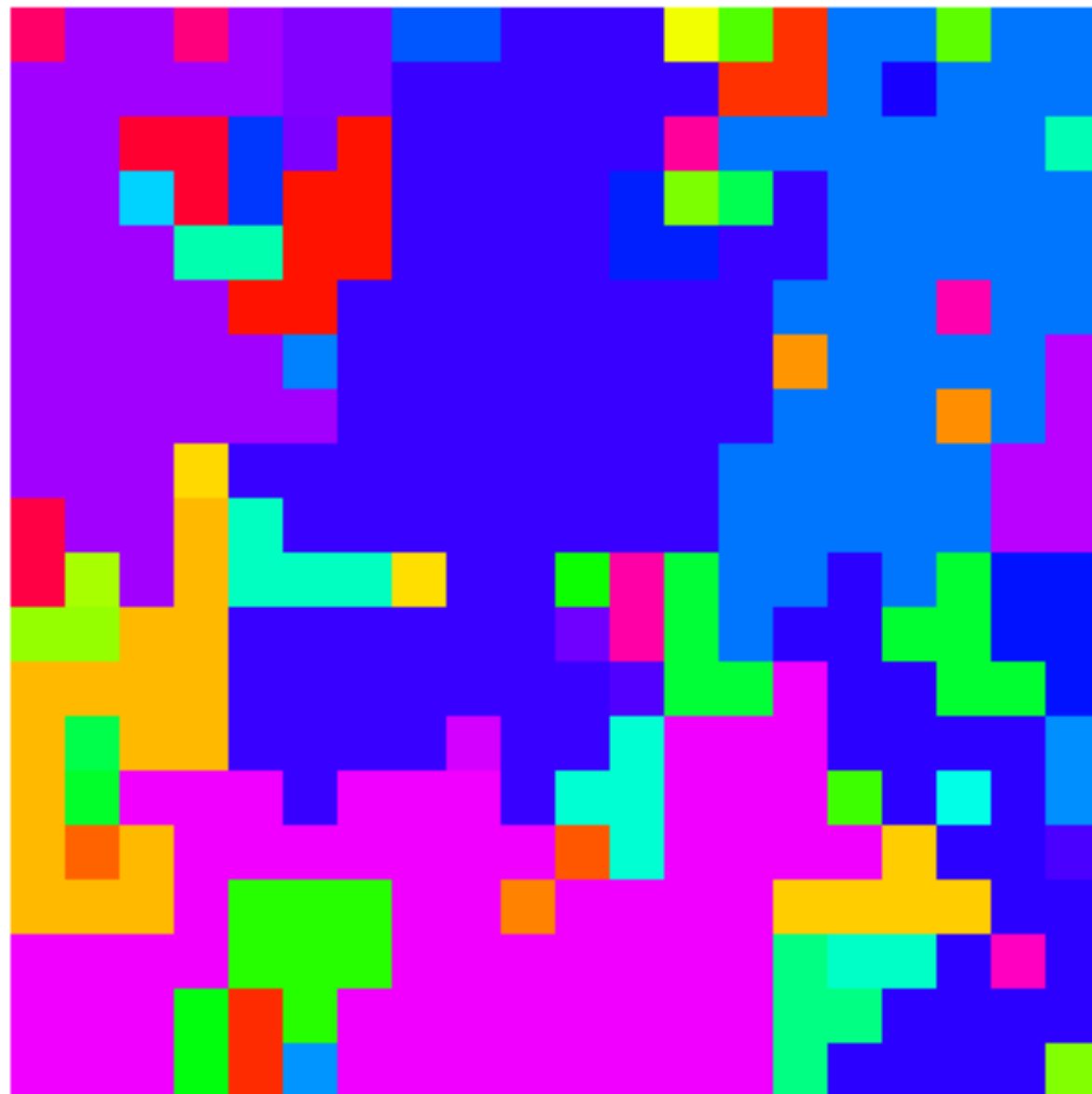
$\varepsilon = 0.10$

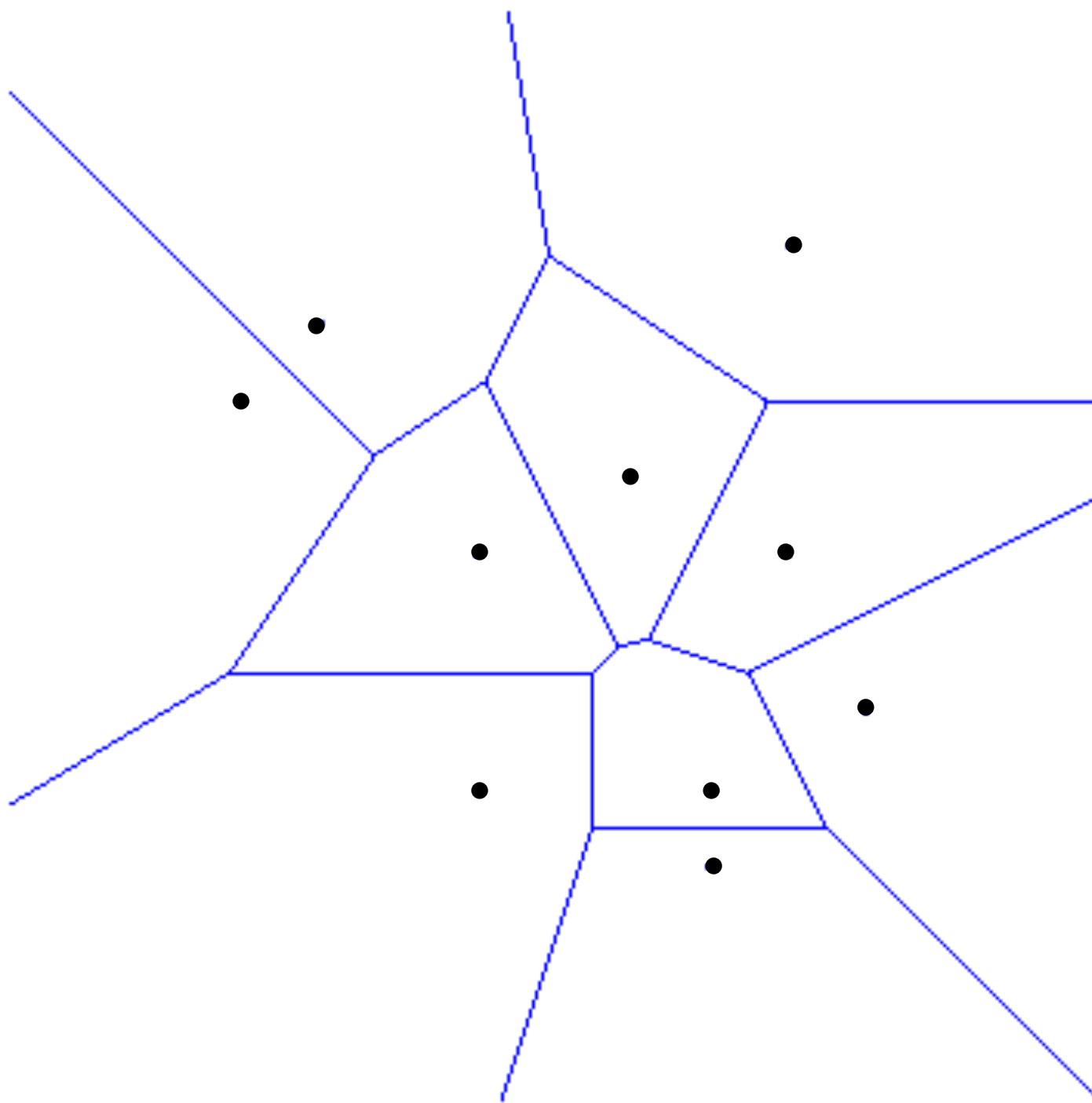


$\varepsilon = 0.20$

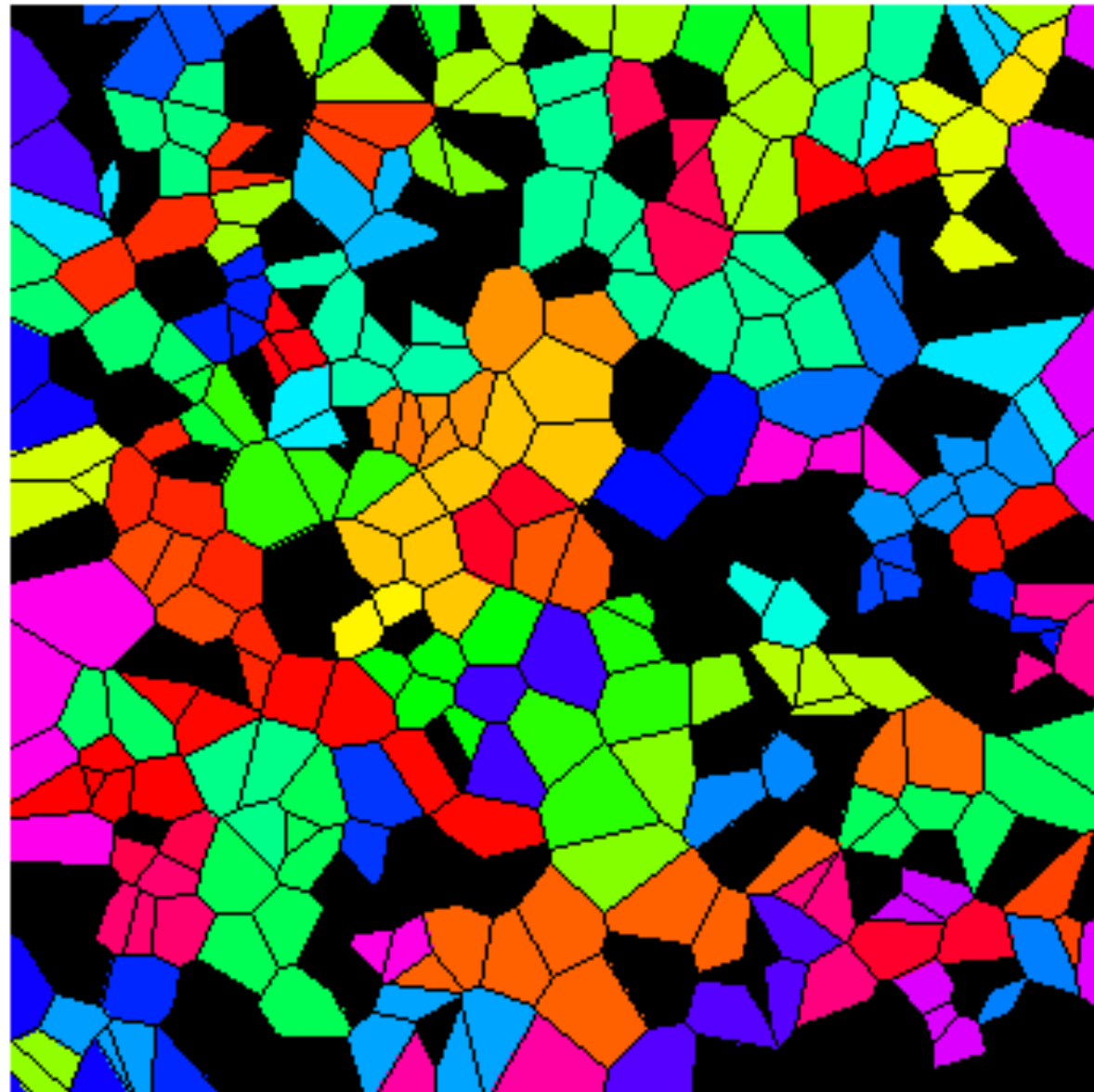


$\varepsilon = 0.30$

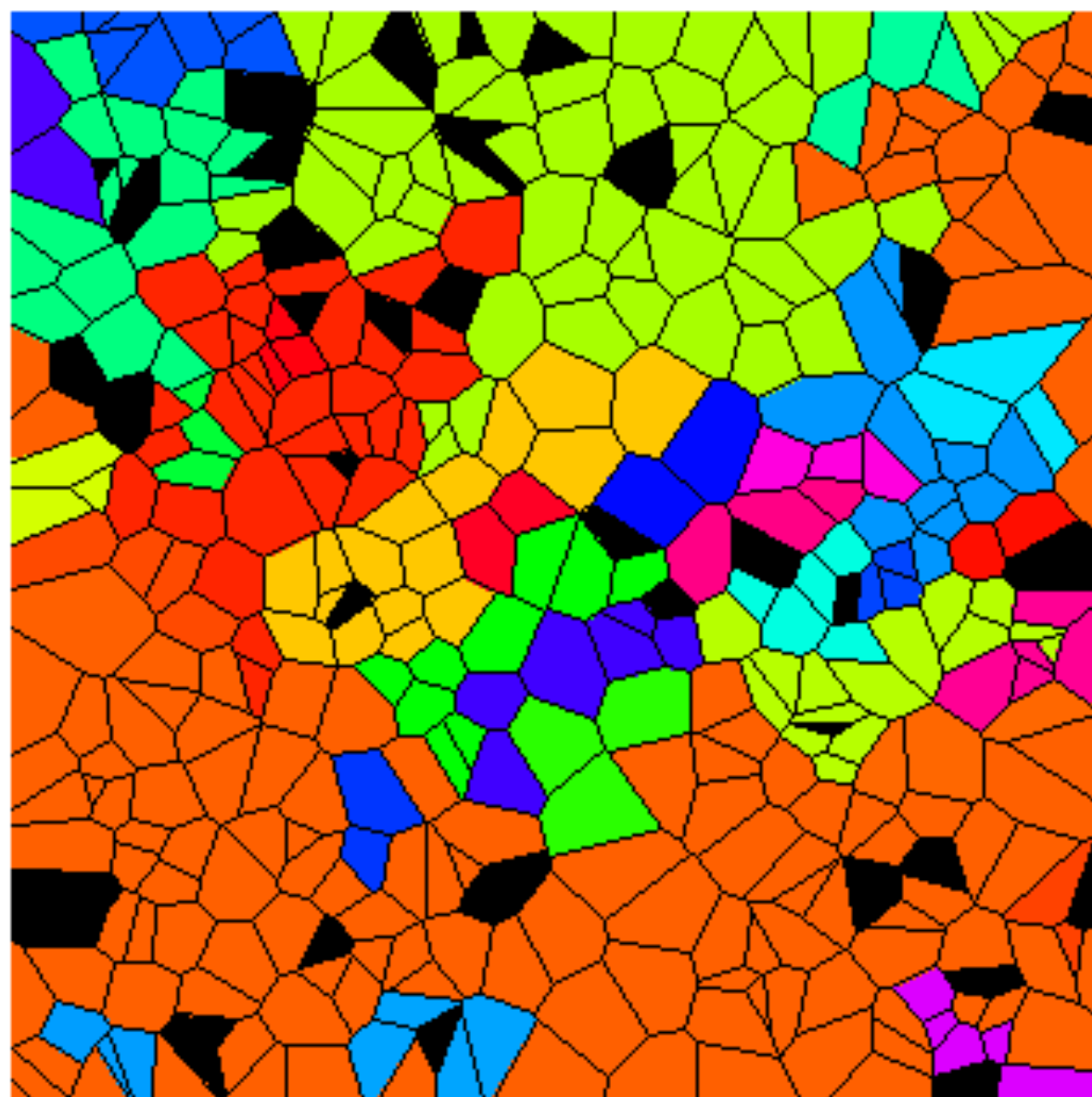




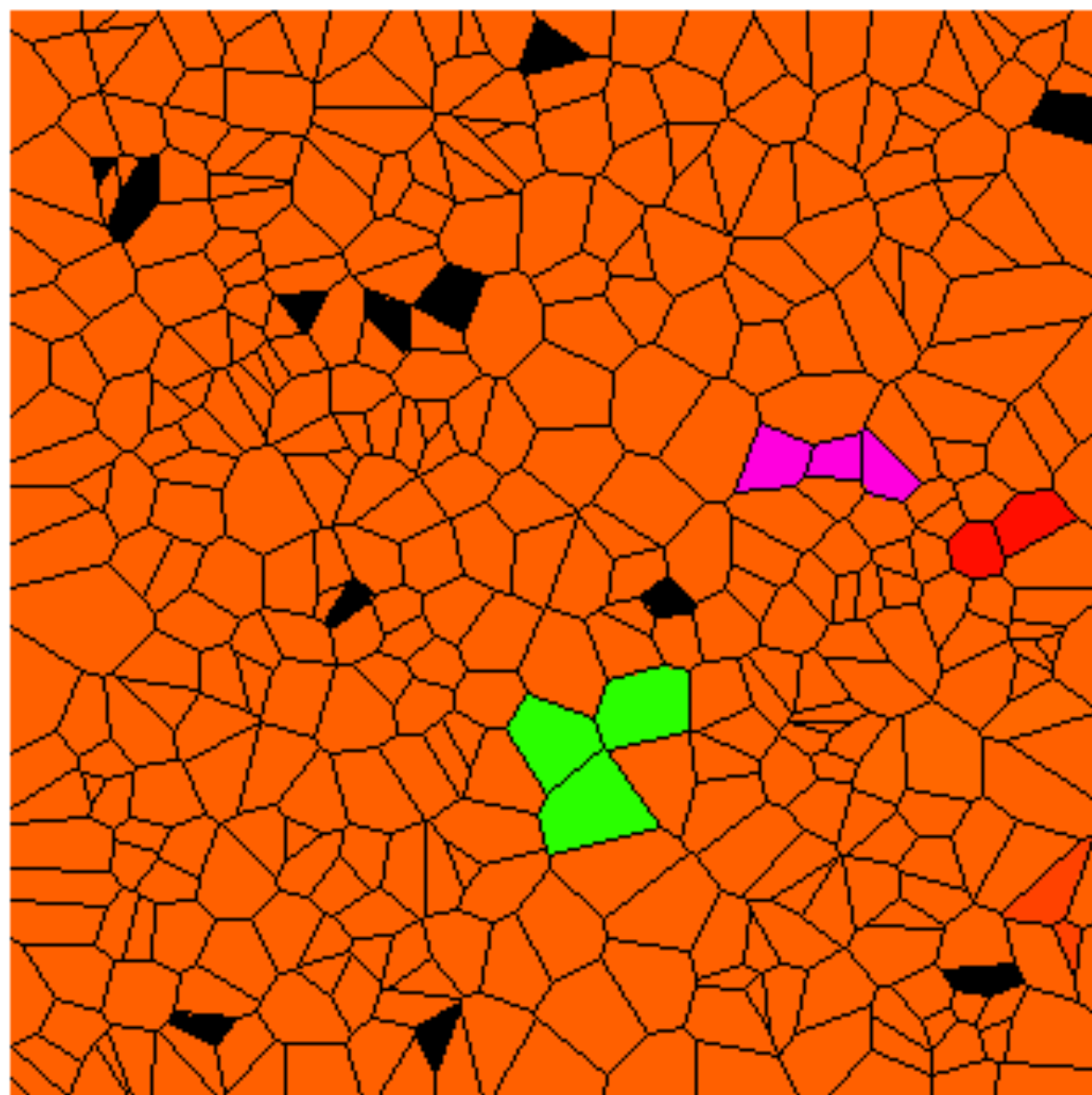
$\varepsilon = 0.10$

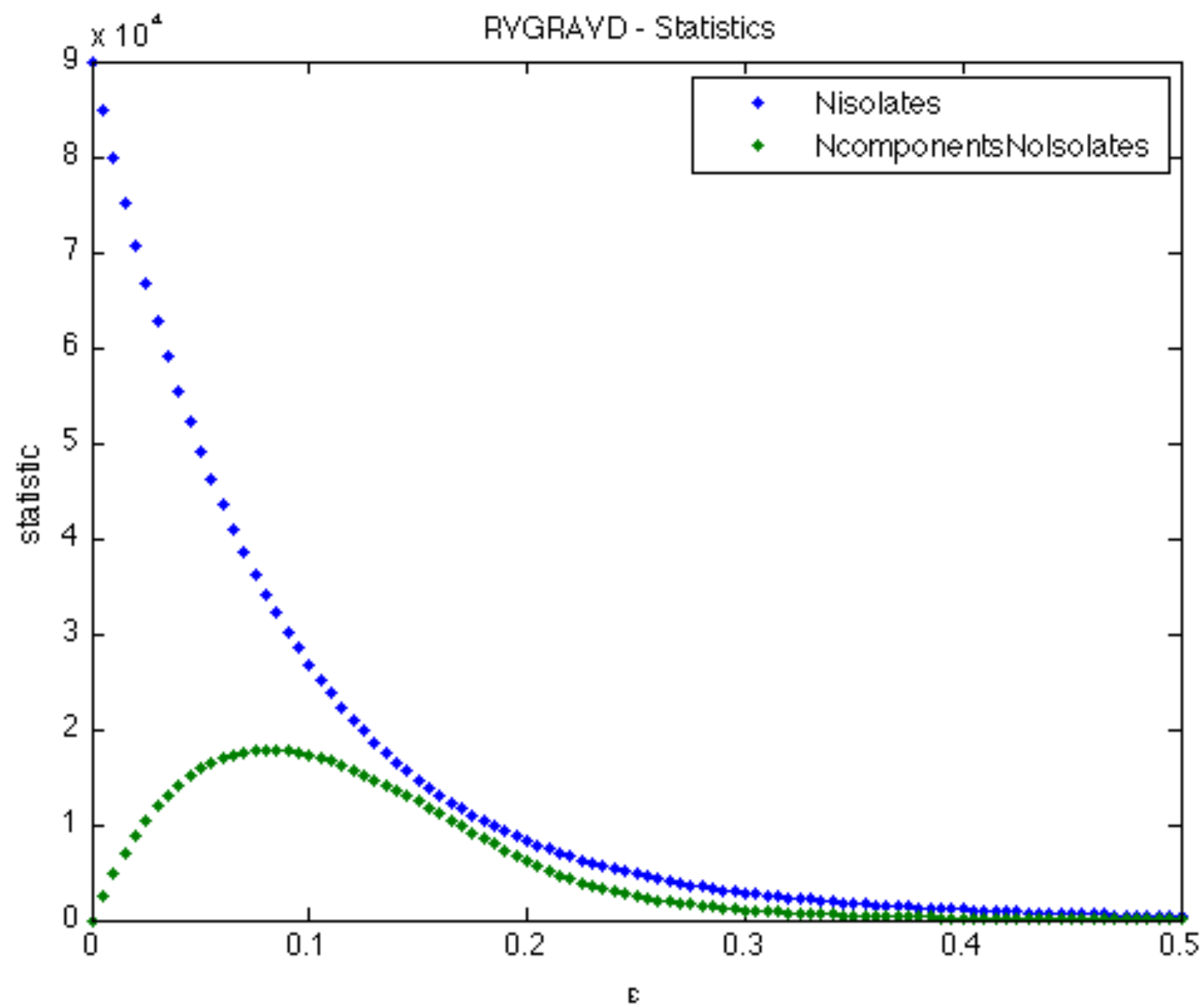


$$\varepsilon = 0.20$$



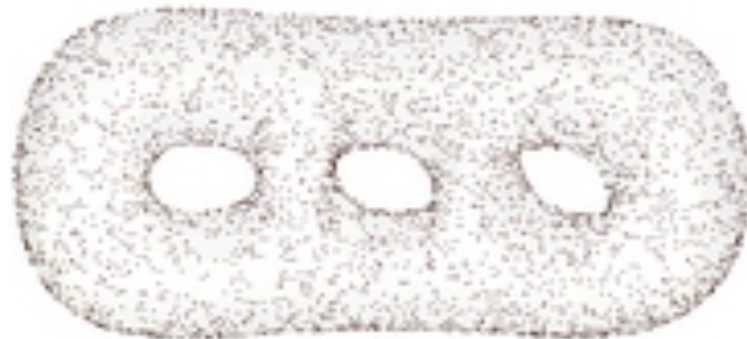
$\varepsilon = 0.30$



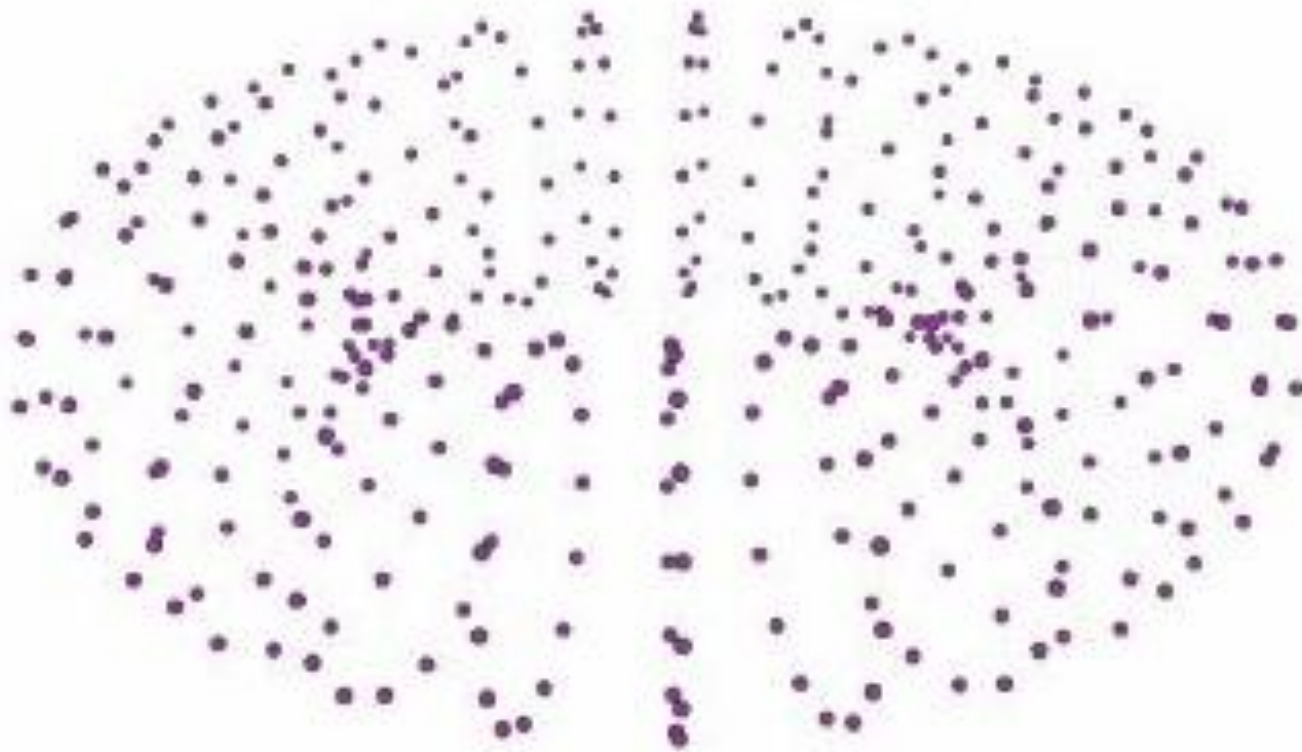


Geometry of Data

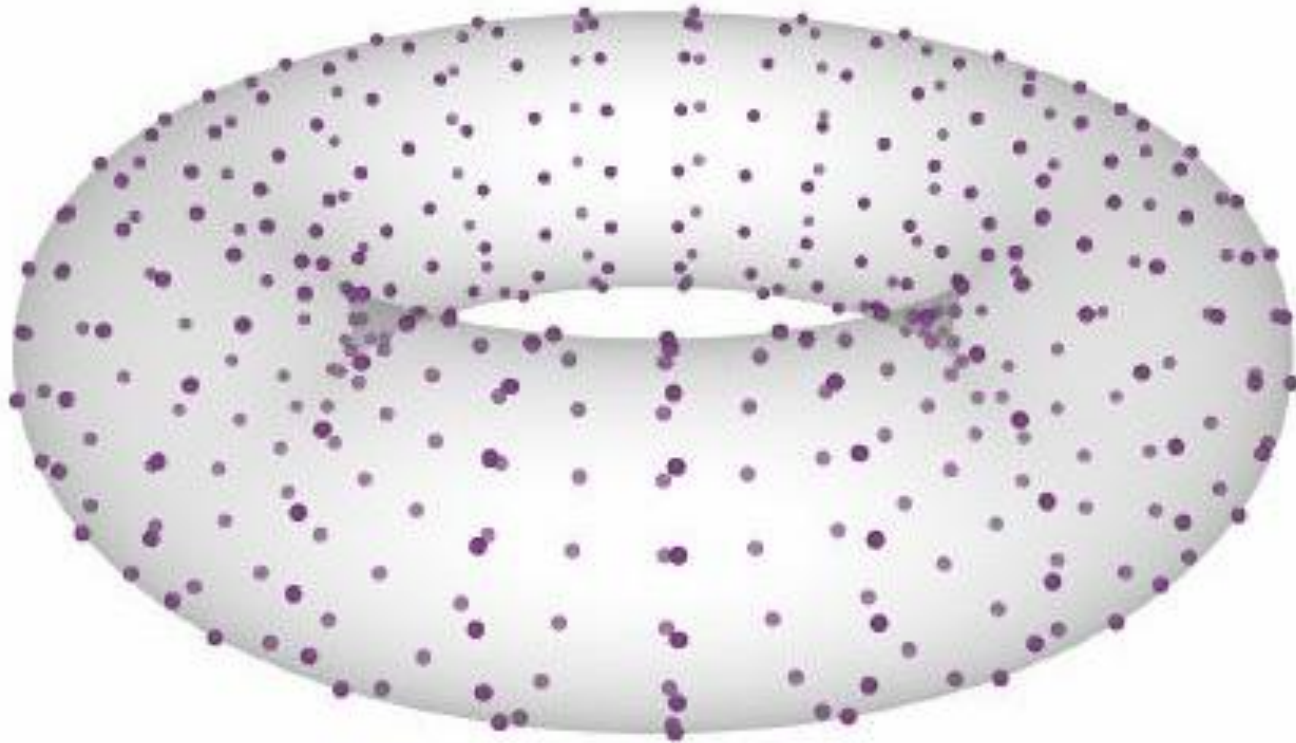
- Consider 10^6 points and each point is a string of 100 numbers.
 - Is it one piece or more?
 - Is there a tunnel? Or a void?
- Point clouds in 100-D Euclidean space.



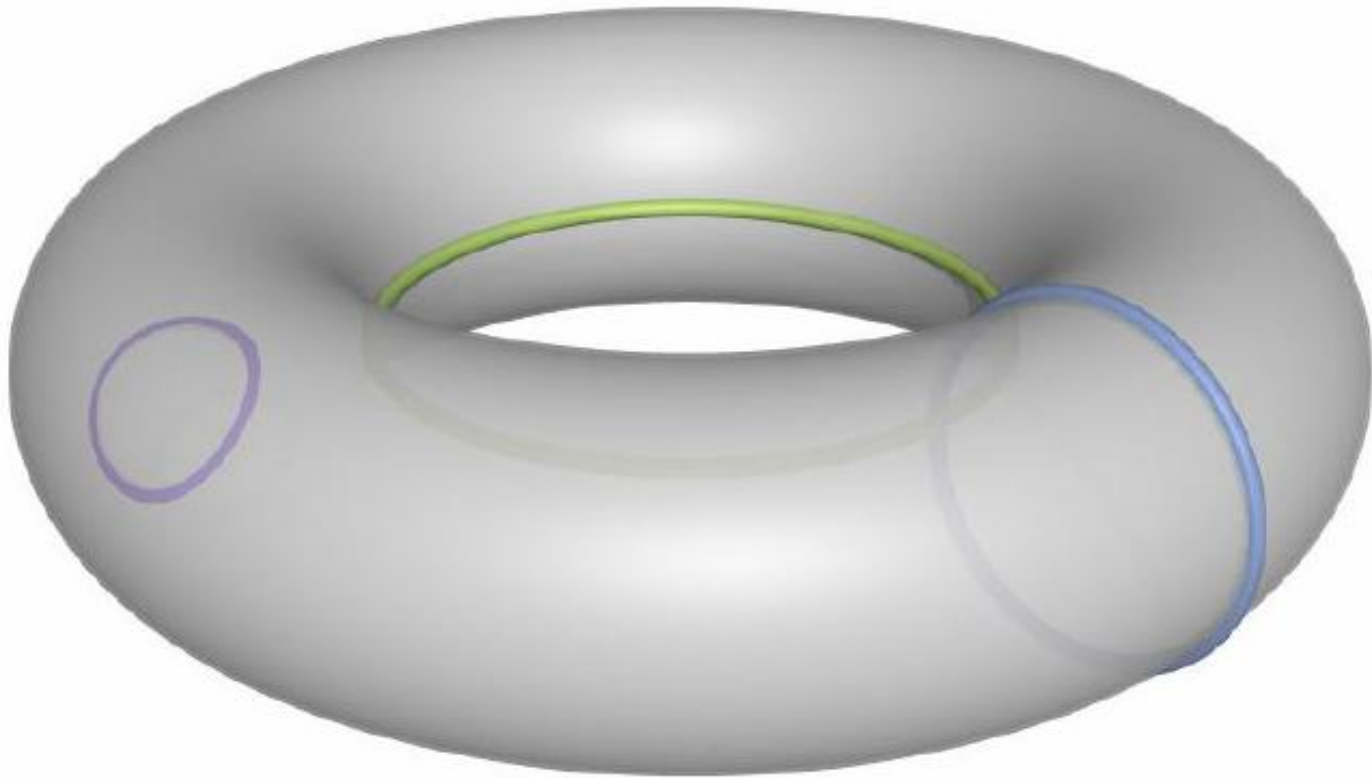
Point Cloud



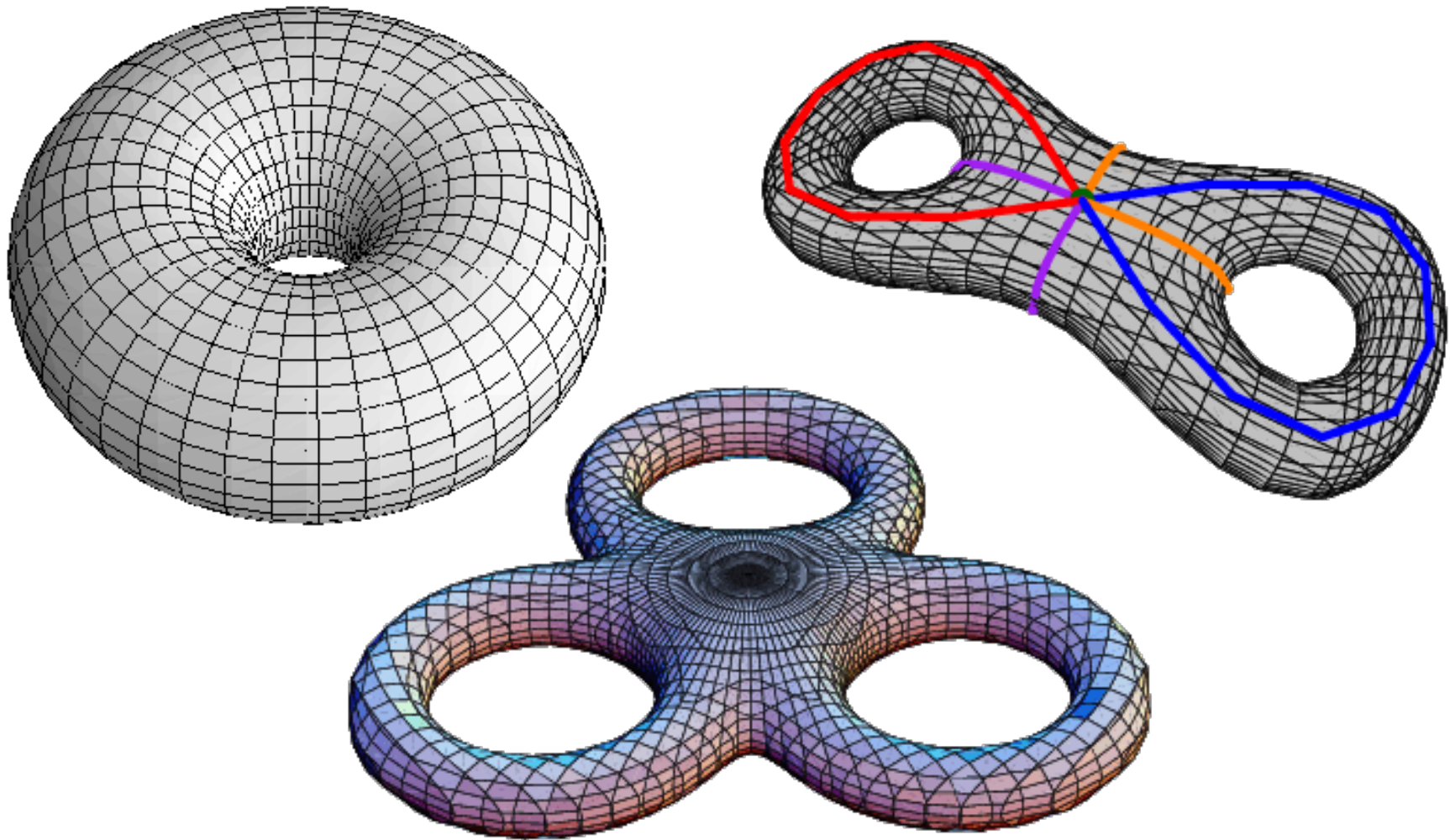
Point Cloud



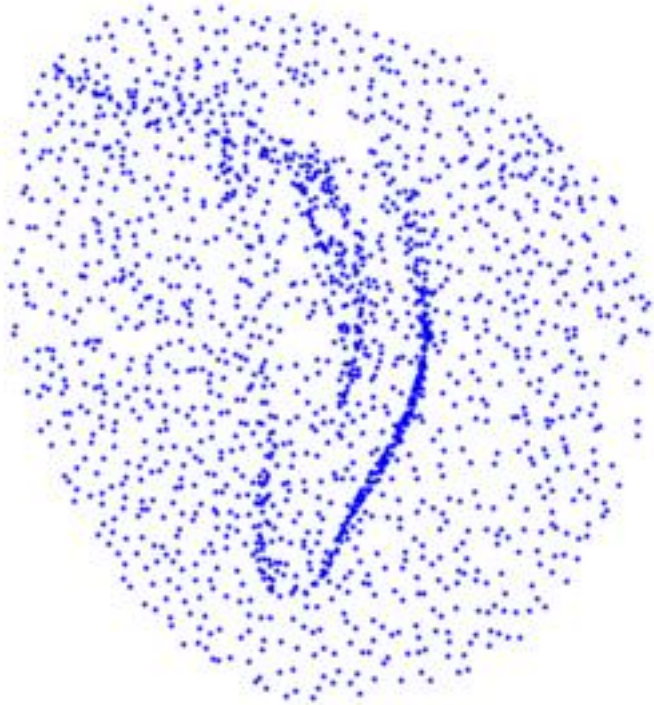
Point Cloud



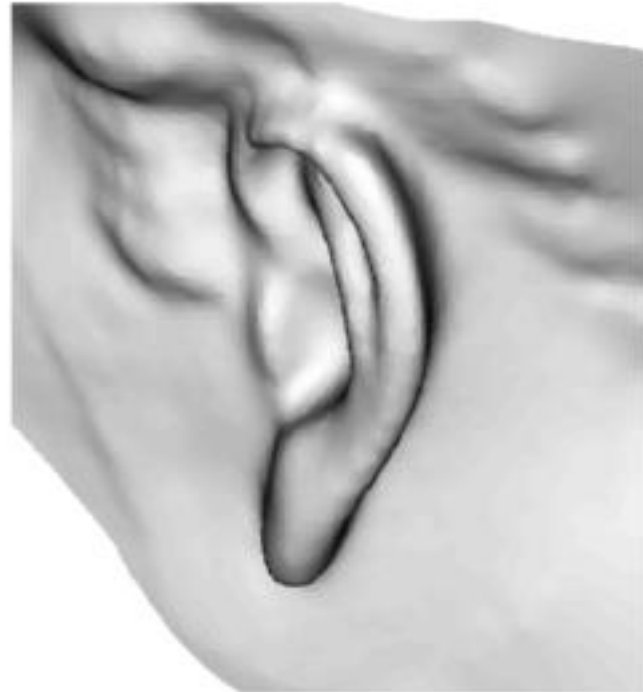
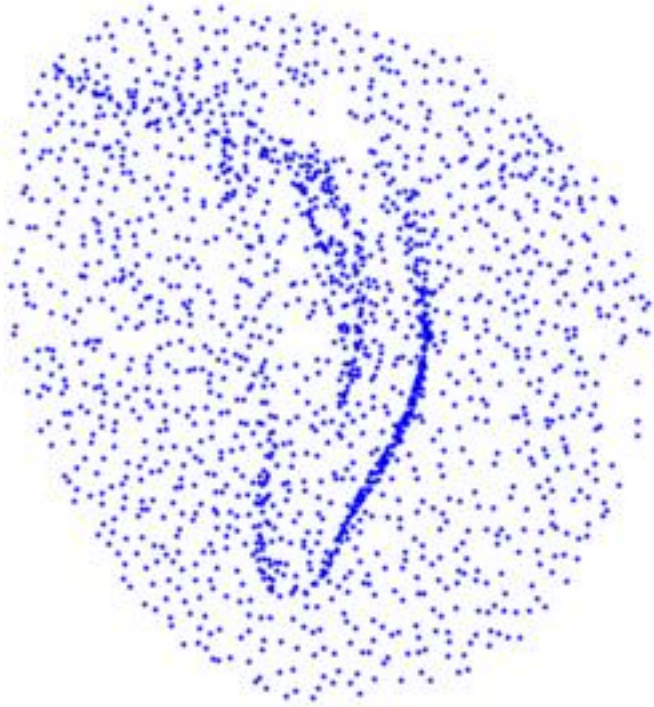
Point Cloud

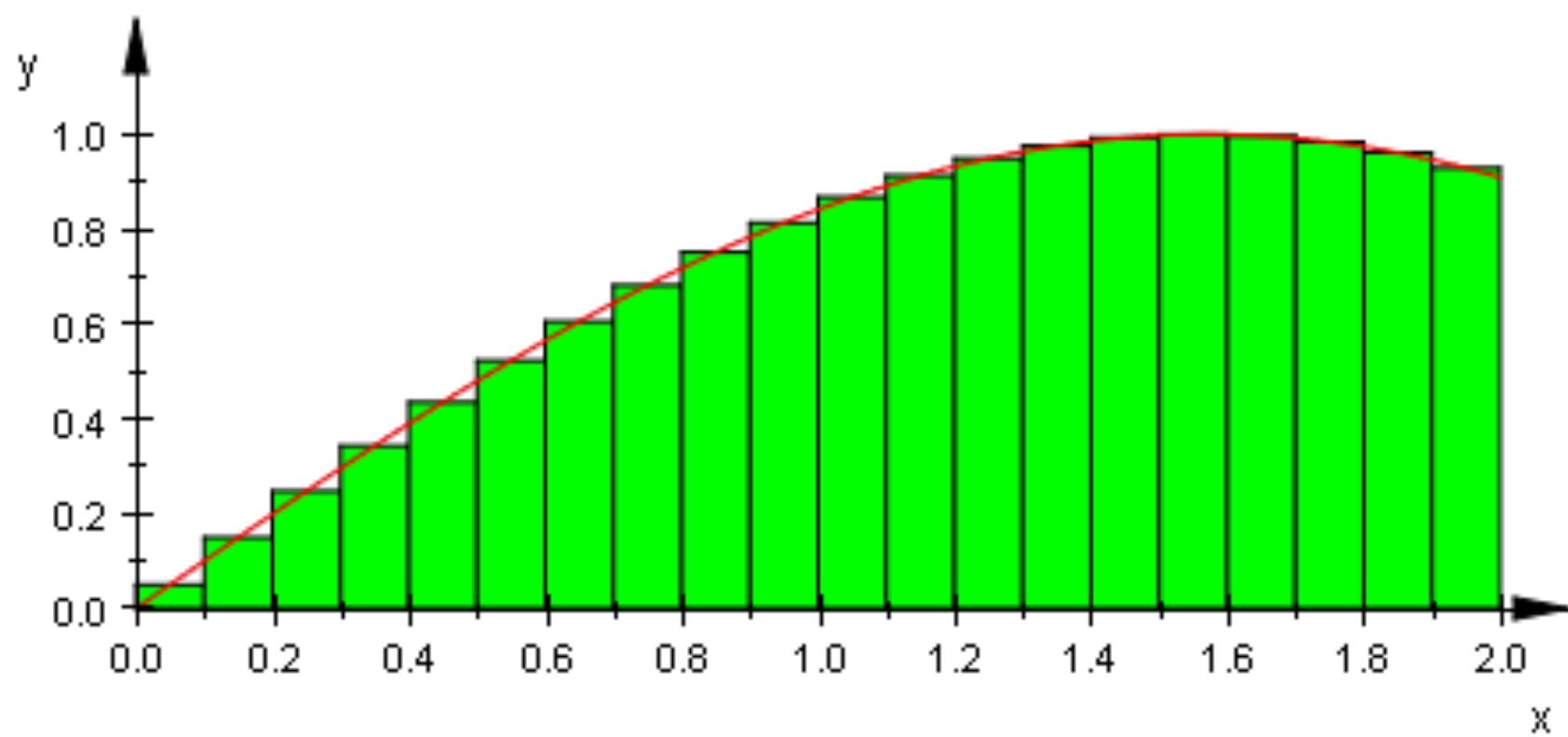


Point Cloud

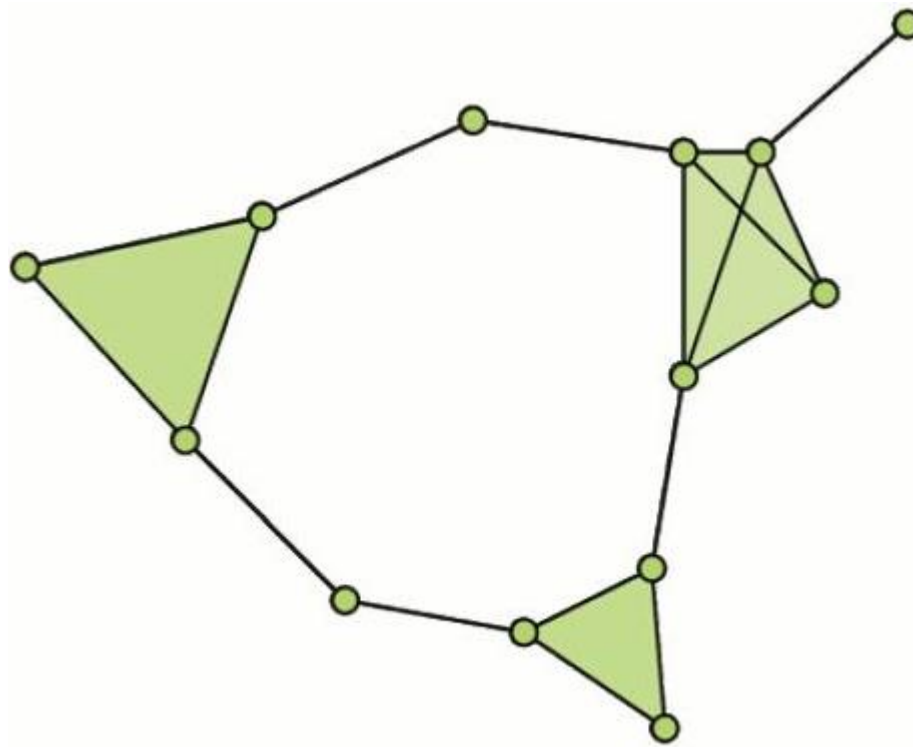


Point Cloud

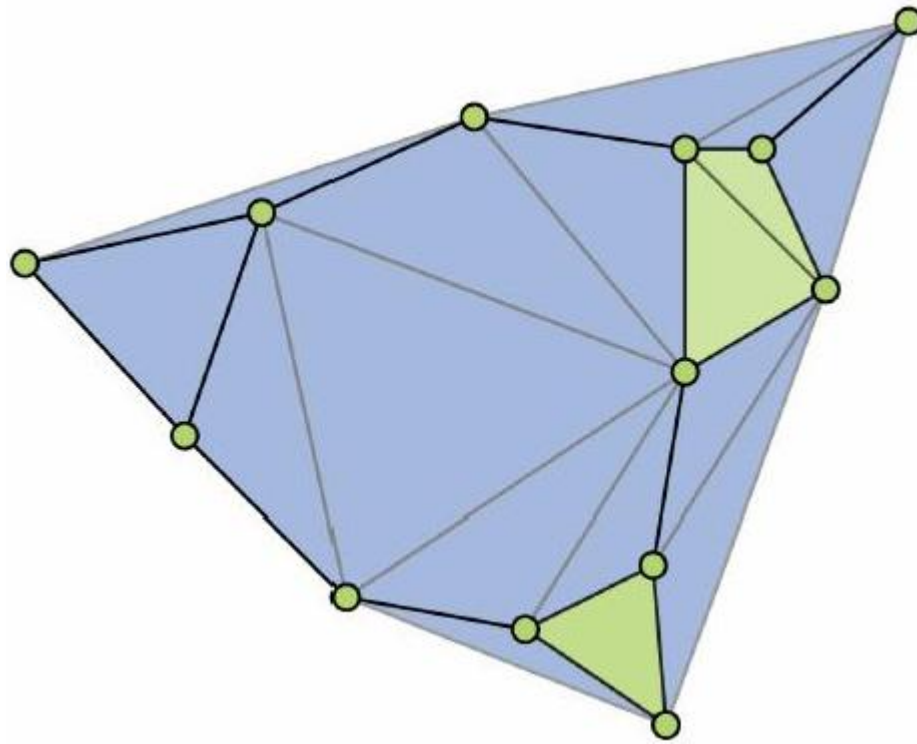




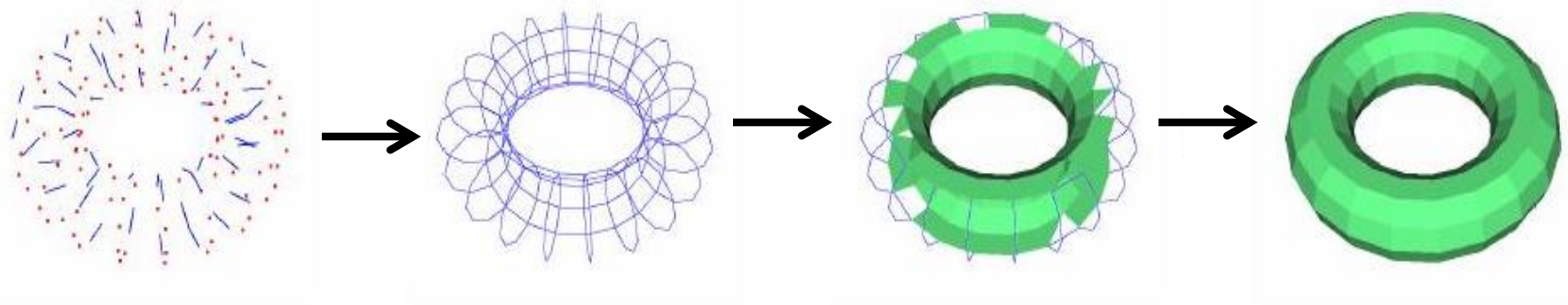
Triangulation



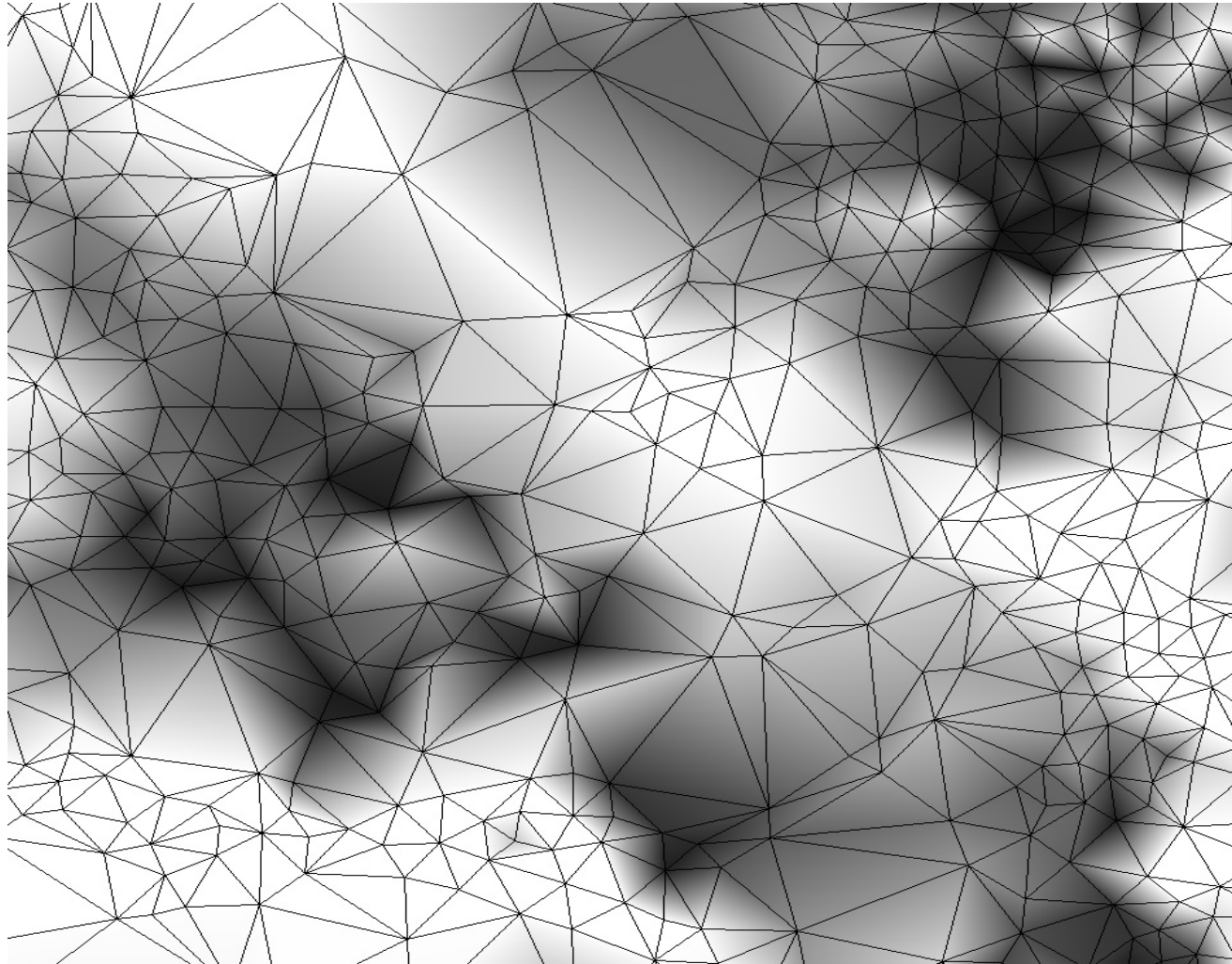
Triangulation



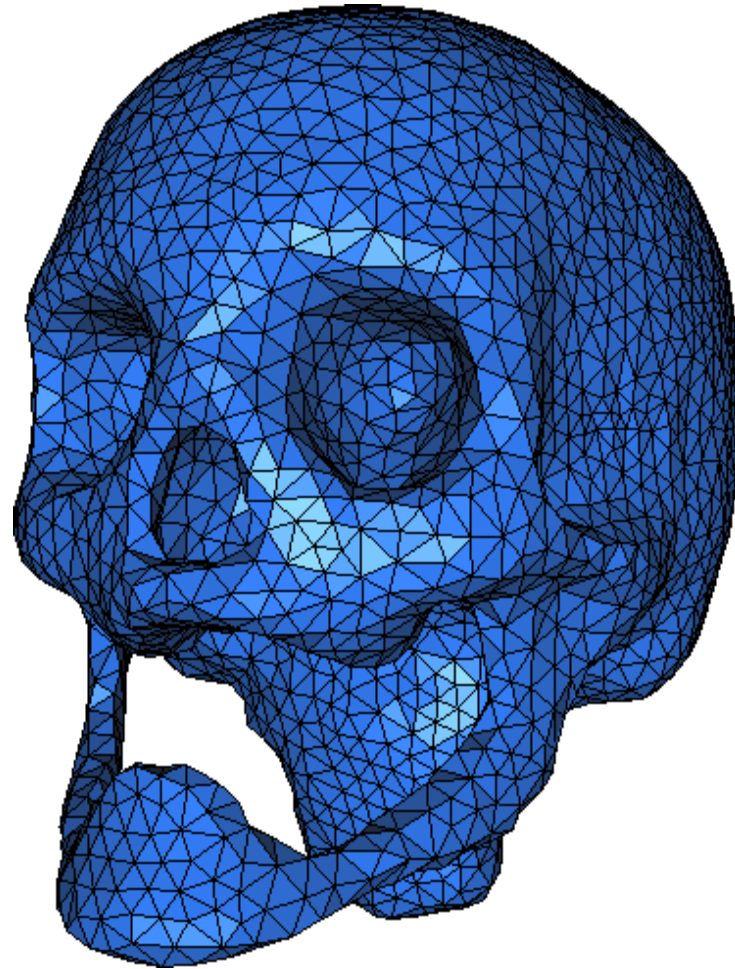
Triangulation



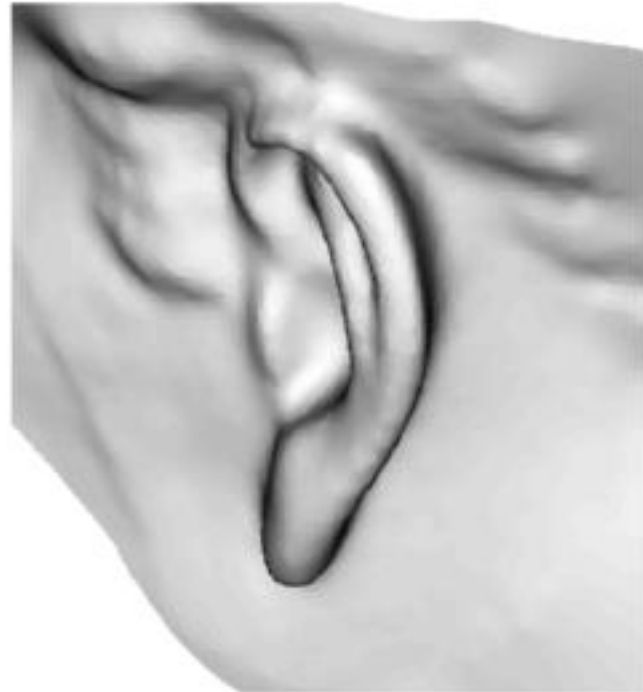
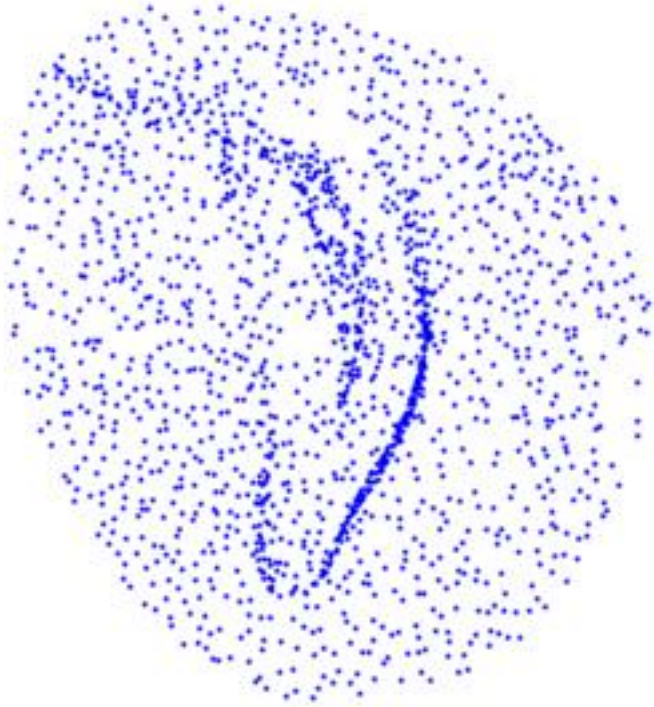
Triangulation



Triangulation



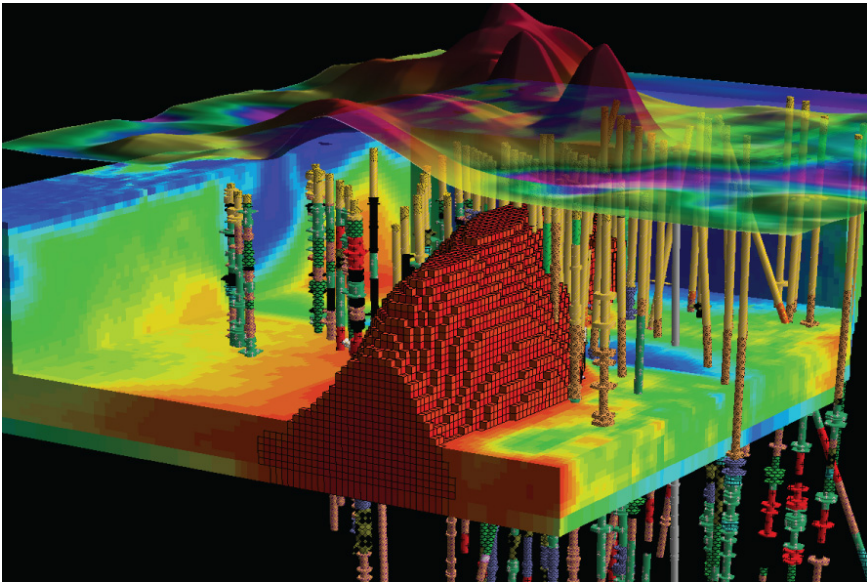
Point Cloud



MRI



GIS – Elevation Model



GIS – Urban Environment

