UMAP + MAPPER

=

UMAPPER

Leland McInnes

June, 2020

UMAP and MAPPER are both data visualization methods

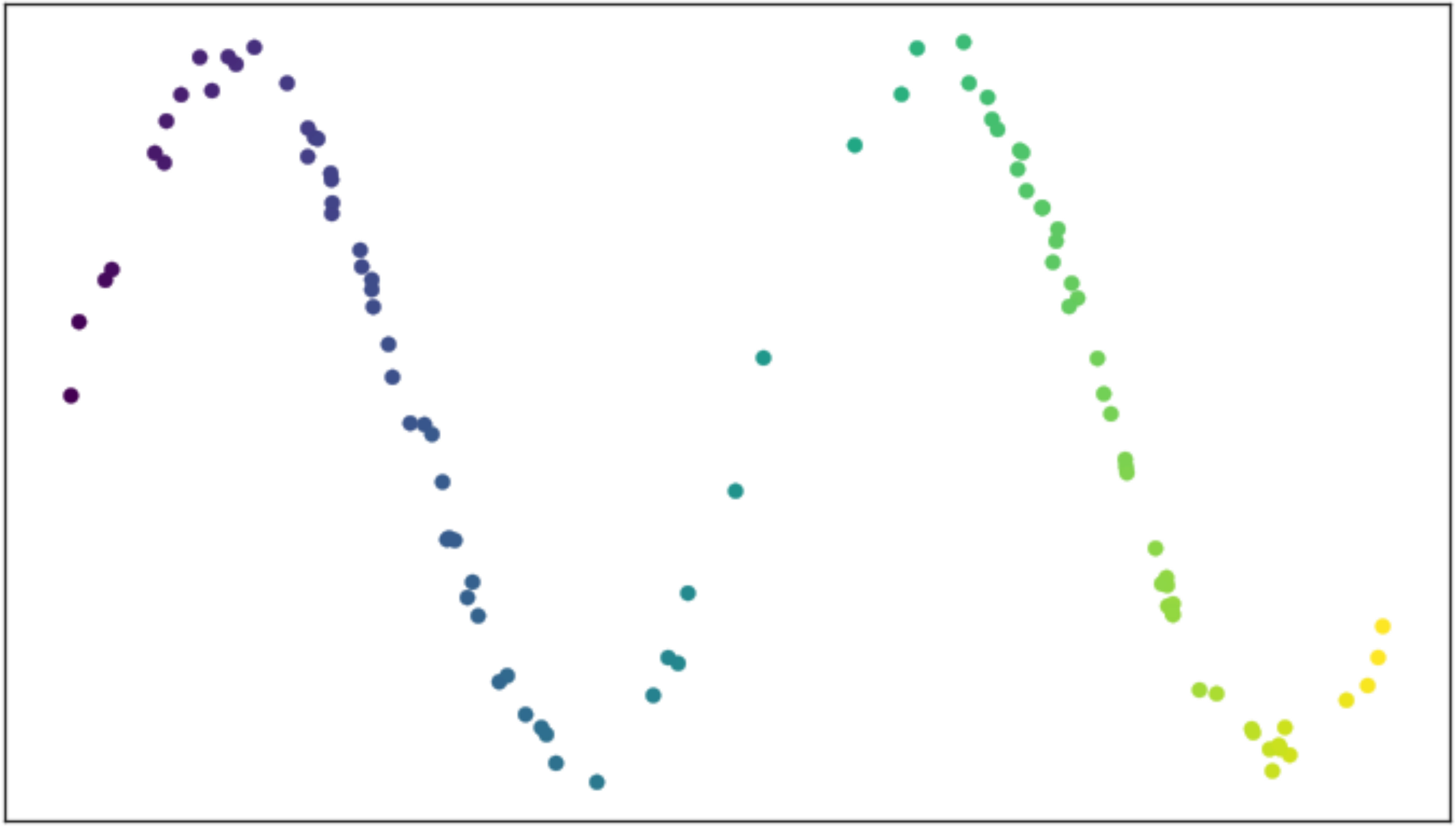
Can we combine them?

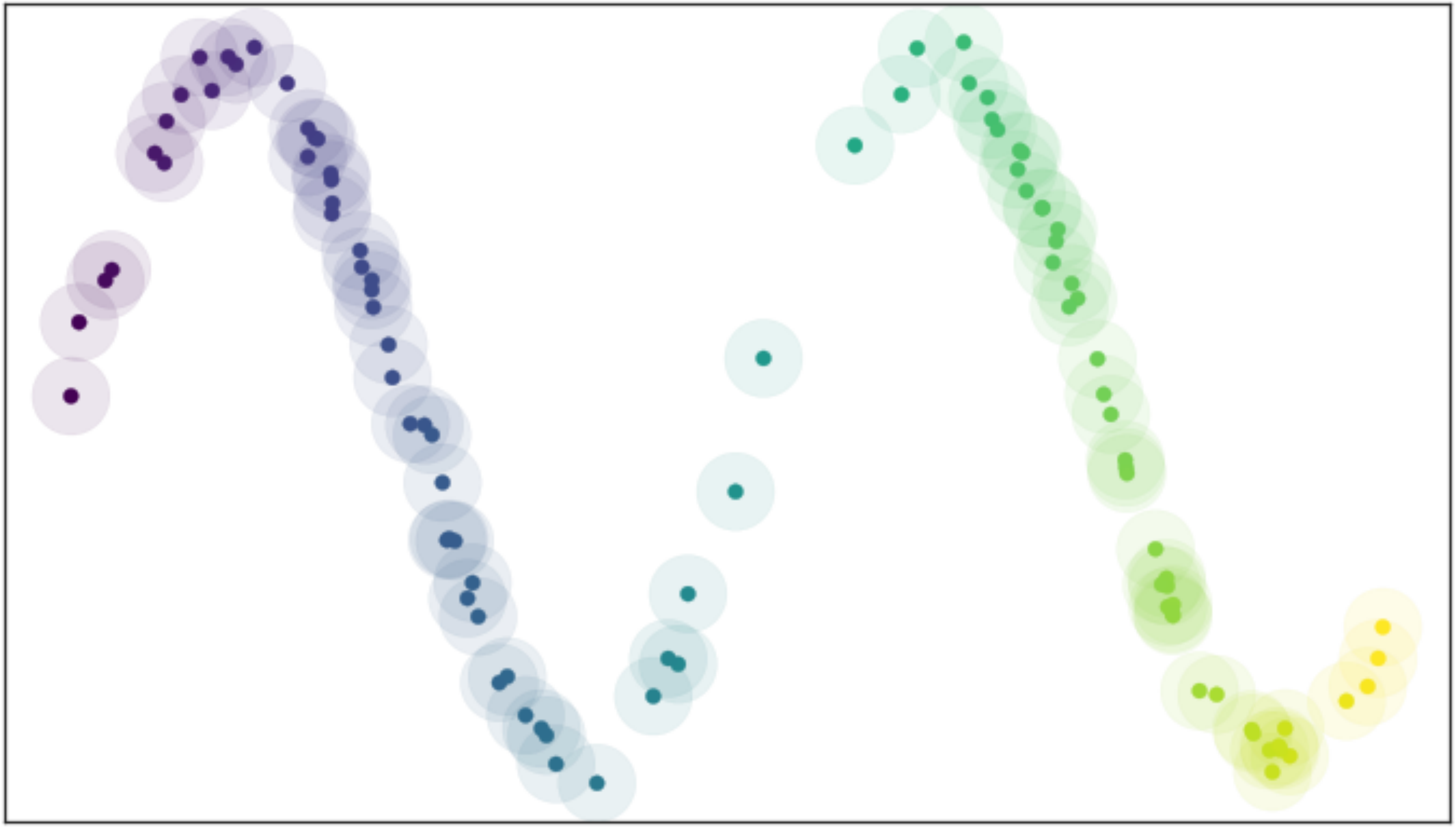
INSTITUT TUTTE INSTITUTE

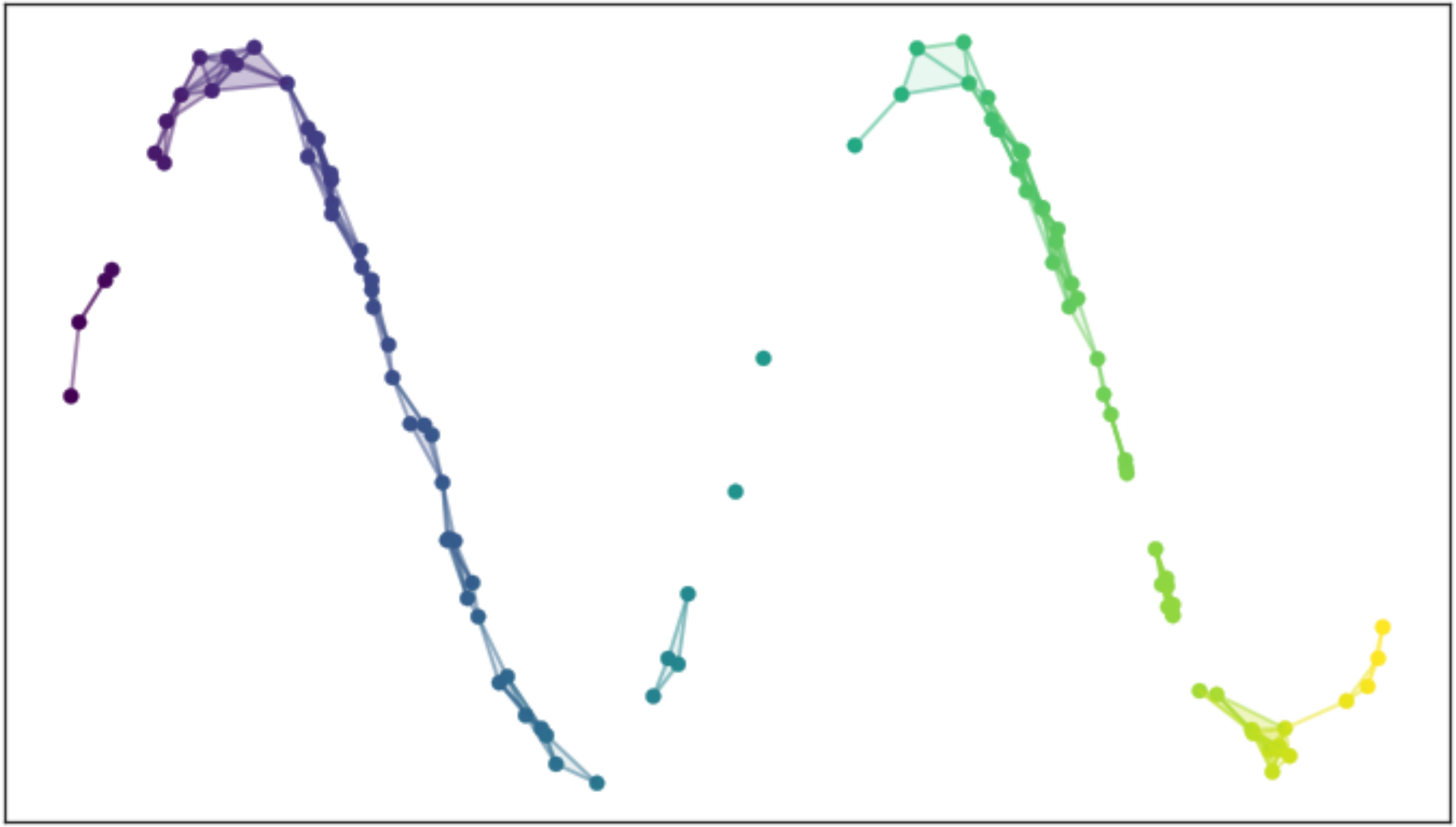
UMAP

McInnes, Healy and Melville 2018

UMAP is a dimension reduction technique using basic topology

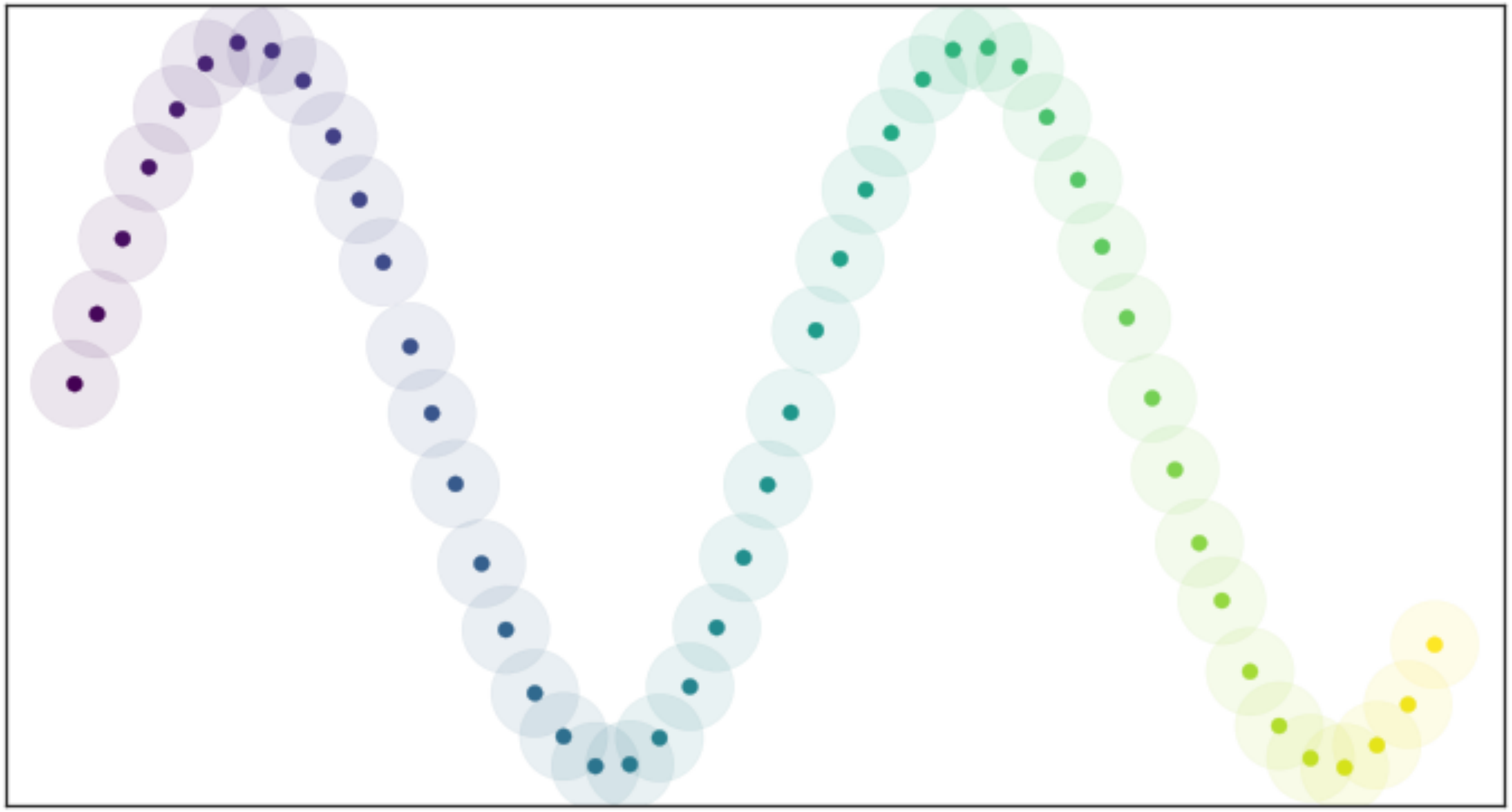




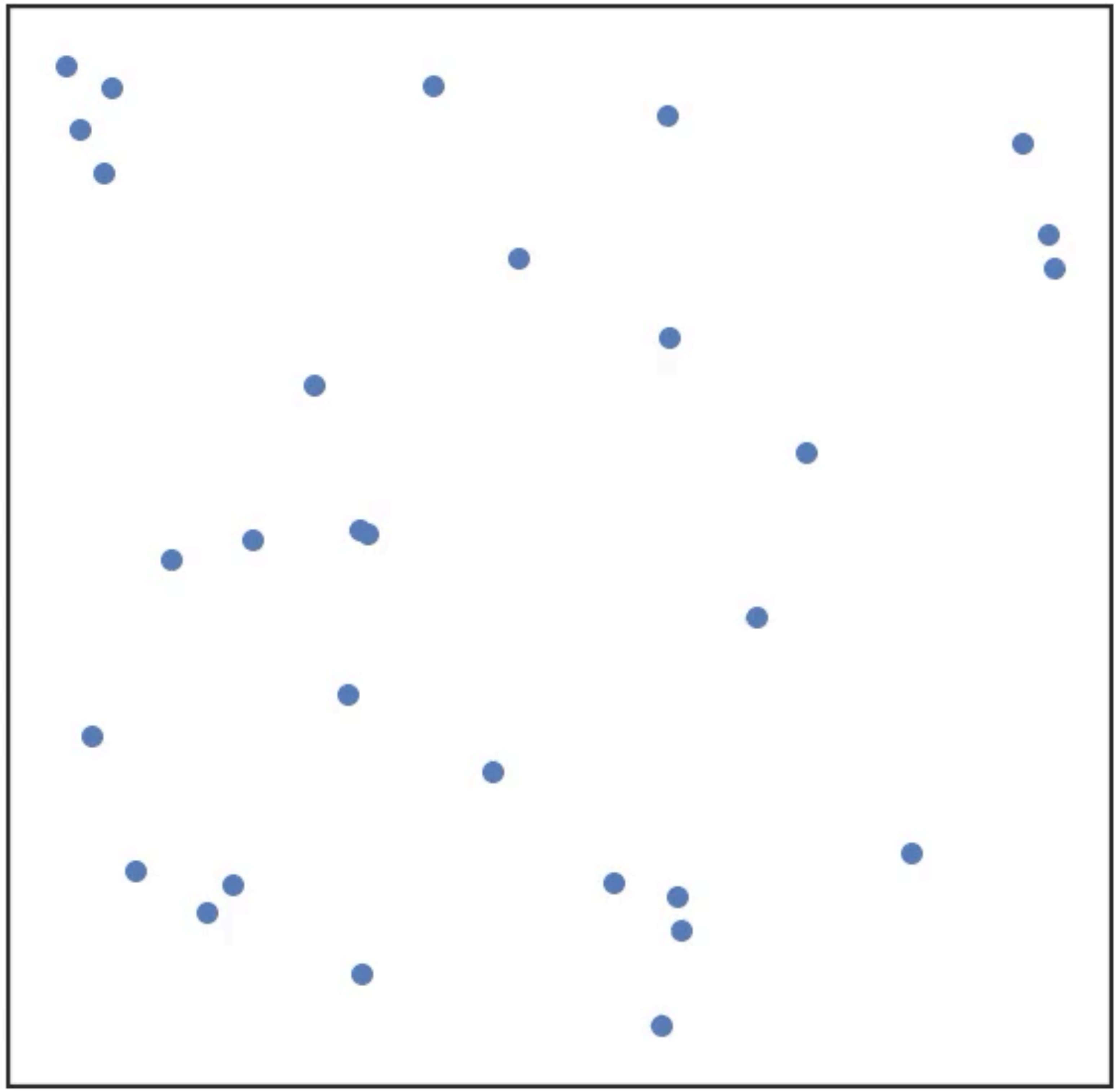


INSTITUT TUTTE INSTITUTE

If the data is uniformly distributed on the manifold then the cover will be "good"



Vary the notion of distance according to the density



We can derive probabilistic weights to simplices based on when they form

Suppose we were given a low dimensional representation

We can apply the same process to get a probabilistic graph!

Now measure the distance between the graphs using cross-entropy and optimize

INSTITUT TUTTE INSTITUTE

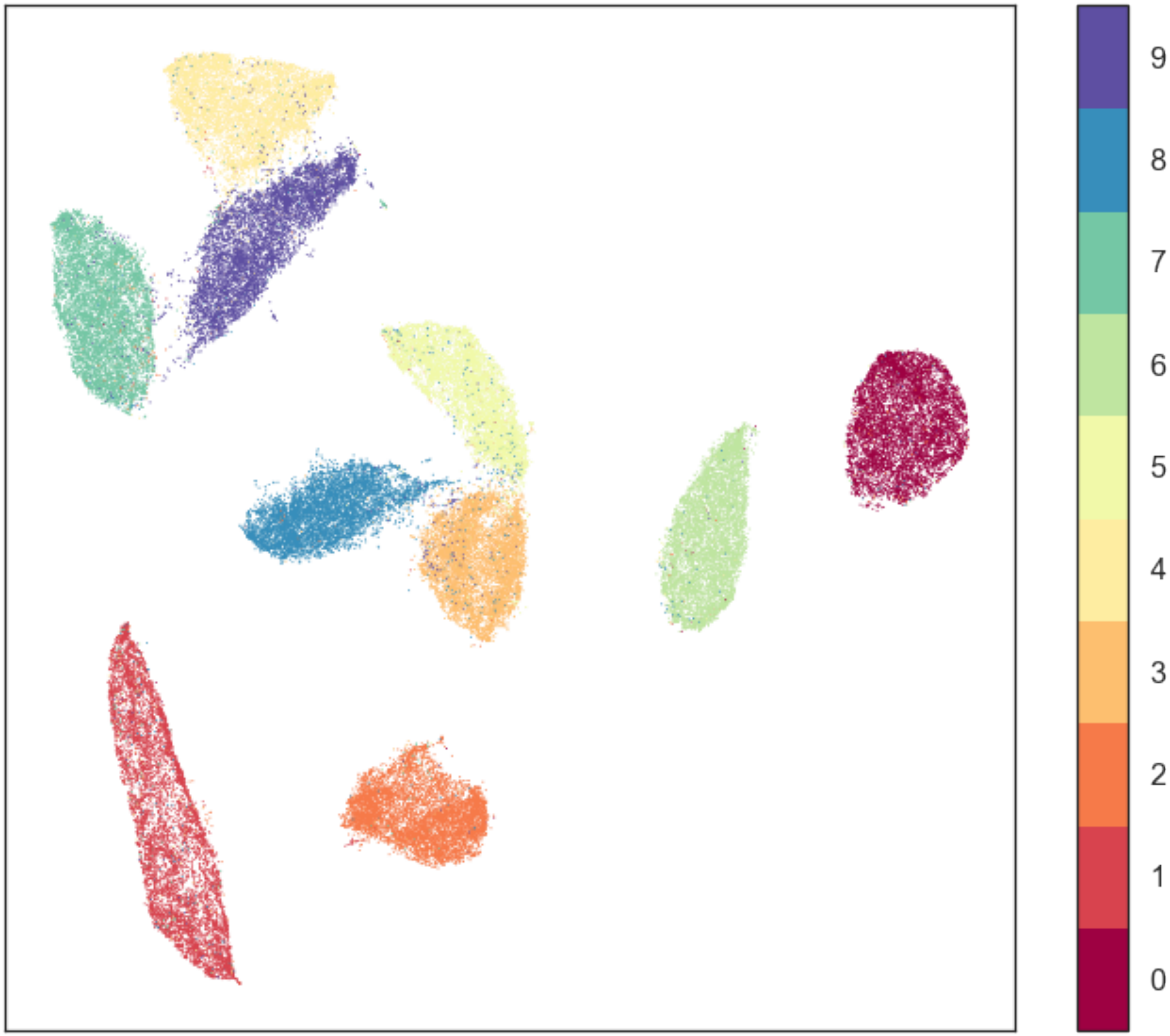
$$\sum_{a \in A} \mu(a) \log\left(\frac{\mu(a)}{\nu(a)}\right) + (1 - \mu(a)) \log\left(\frac{1 - \mu(a)}{1 - \nu(a)}\right)$$

INSTITUT TUTTE INSTITUTE

We are just embedding the graph

$$\overbrace{\sum_{a \in A} \mu(a) \log\left(\frac{\mu(a)}{\nu(a)}\right)}^{\text{Get the clumps right}} + \underbrace{(1-\mu(a)) \log\left(\frac{1-\mu(a)}{1-\nu(a)}\right)}_{\text{Get the gaps right}}$$

UMAP on MNIST digits



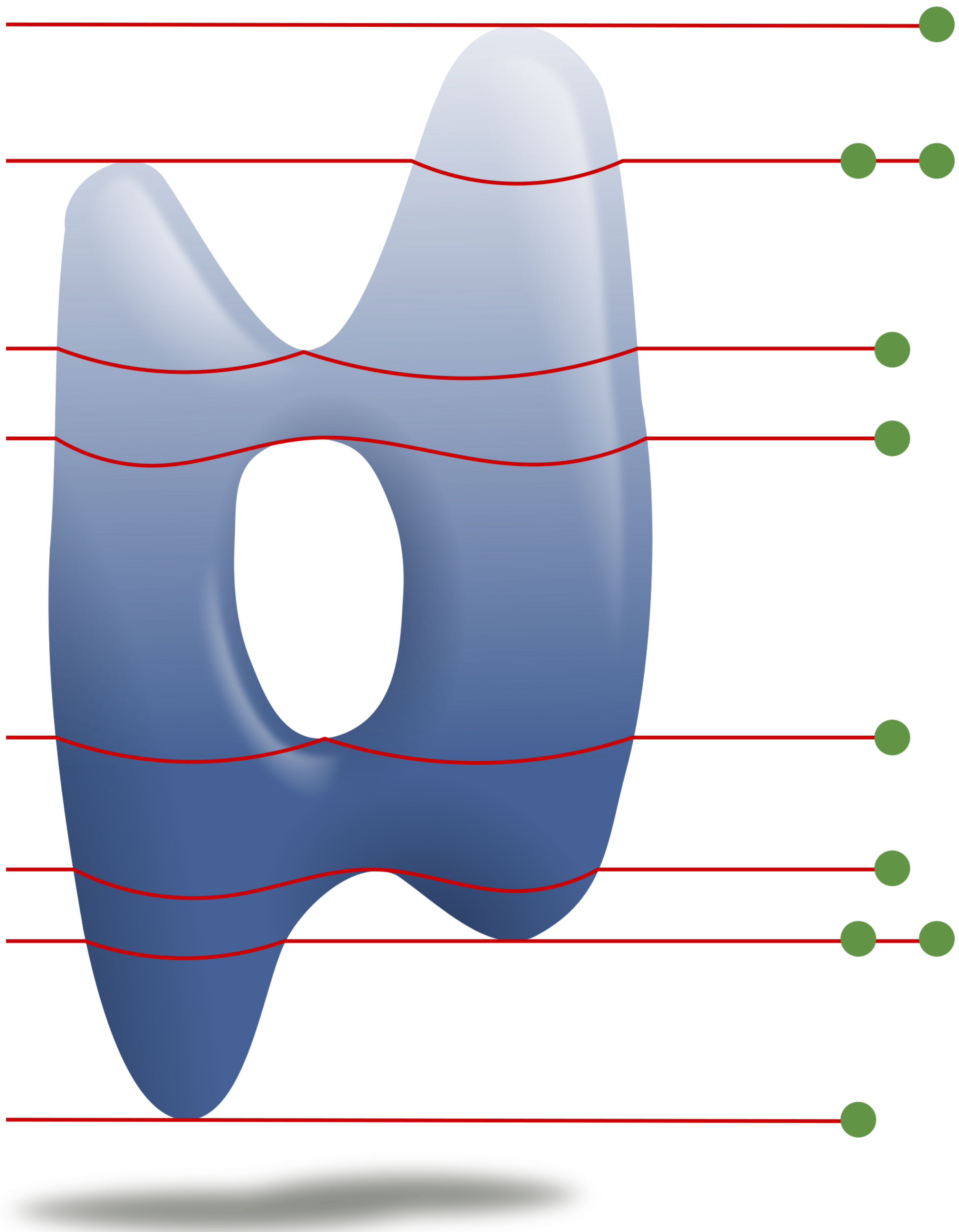
MAPPER

Singh, Memoli and Carlsson, 2007

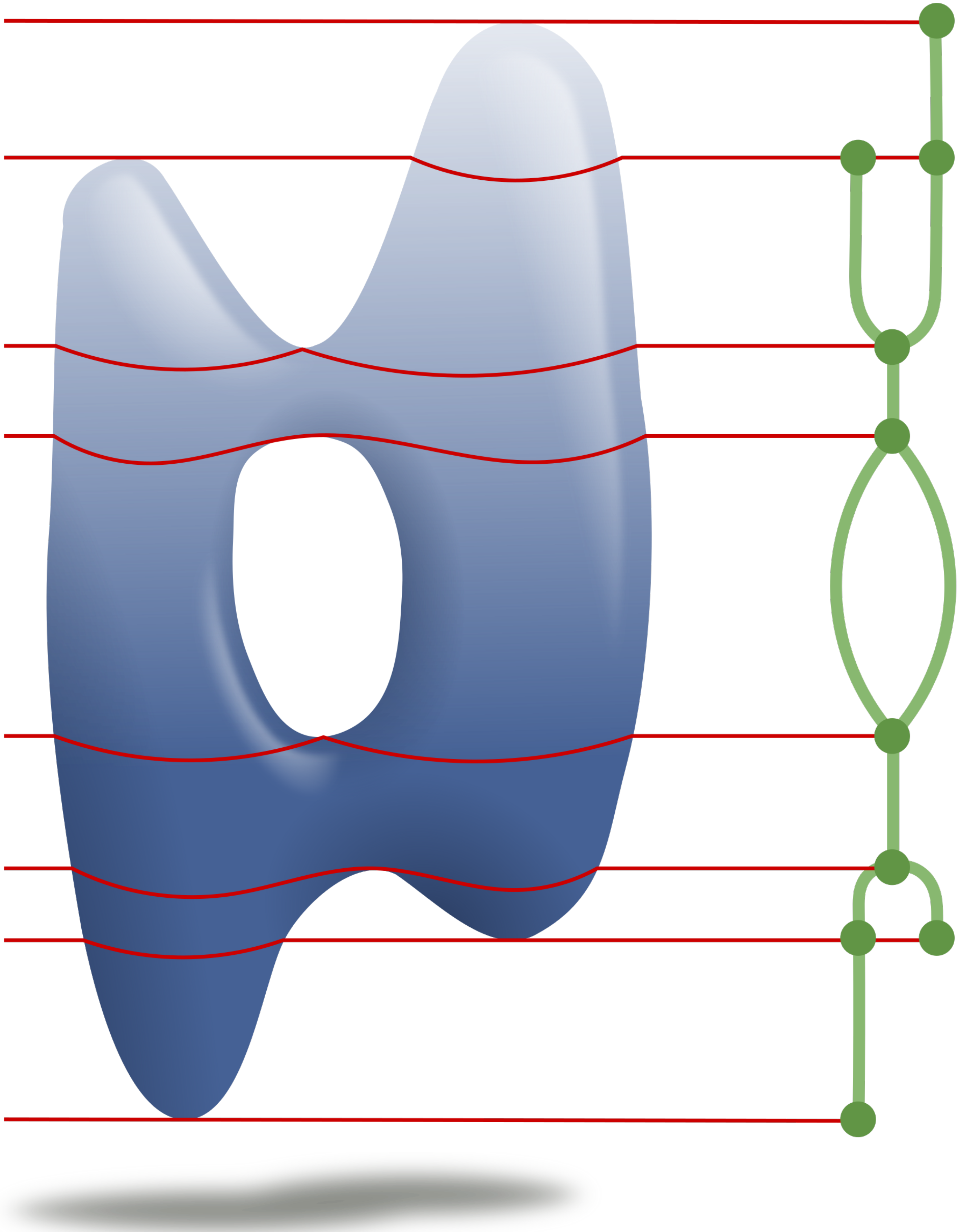
Reeb Graphs

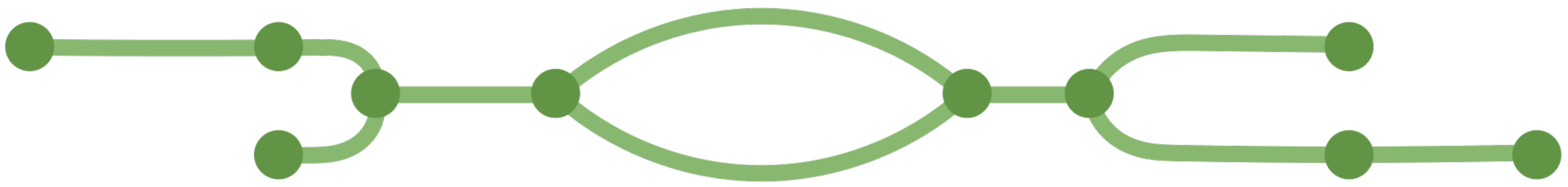
INSTITUT
TUTTE
INSTITUTE

INSTITUT TUTTE INSTITUTE

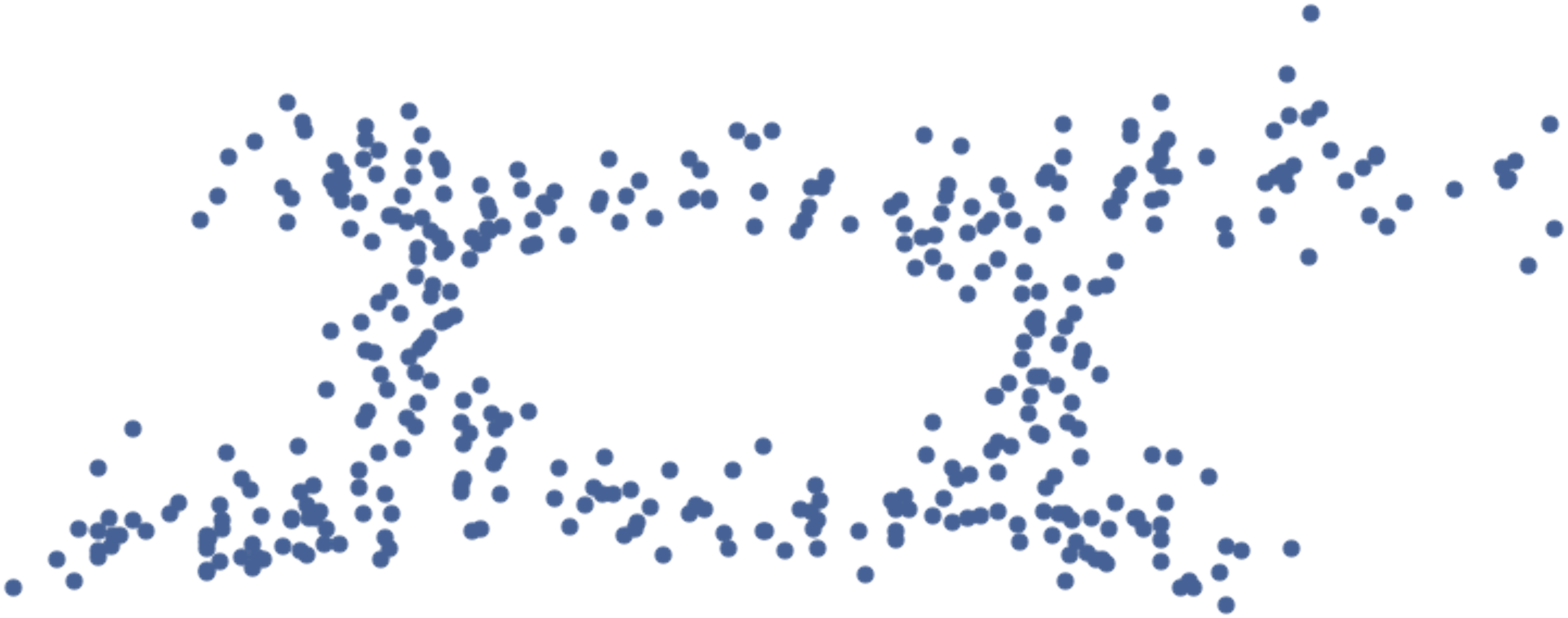


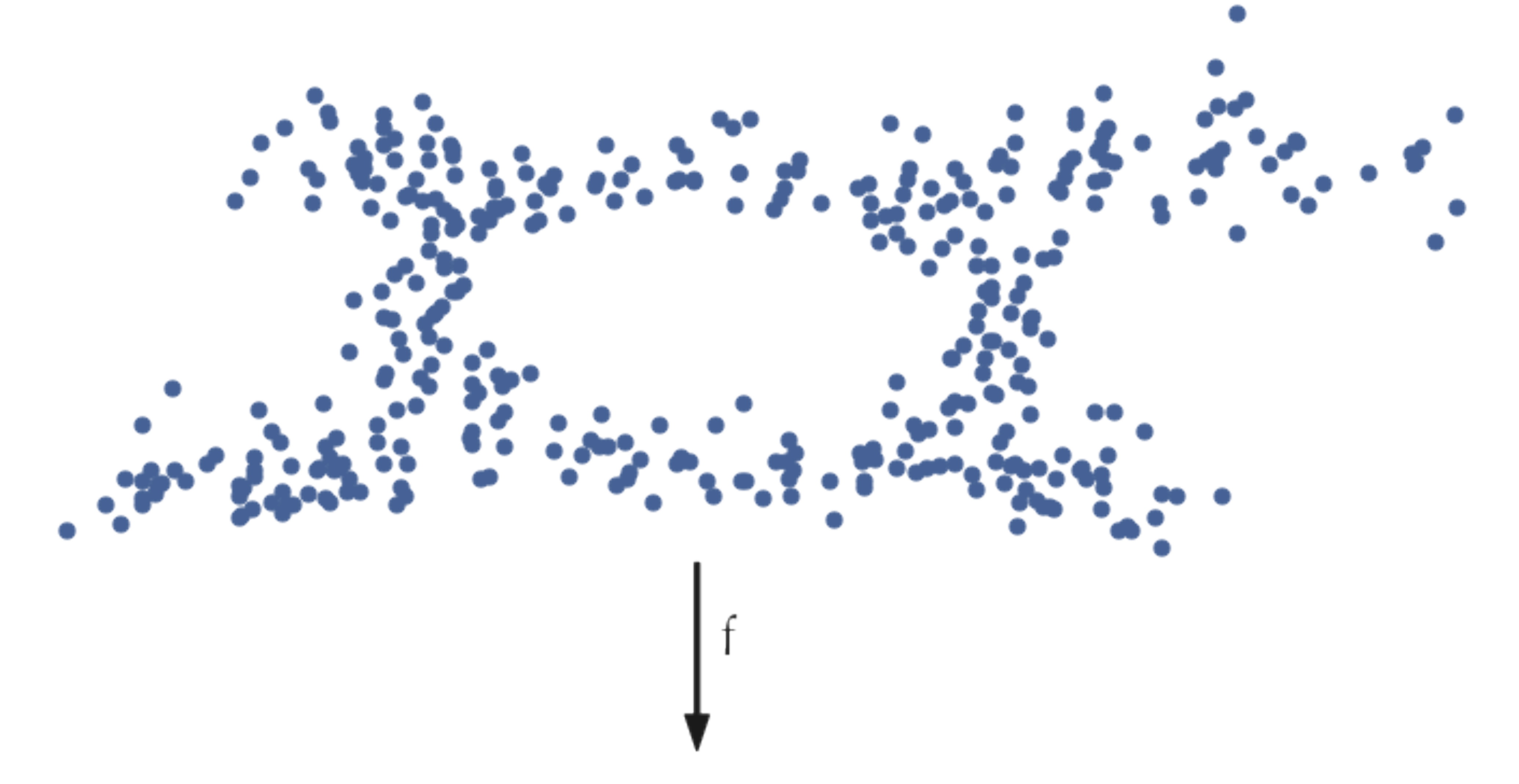
INSTITUT TUTTE INSTITUTE



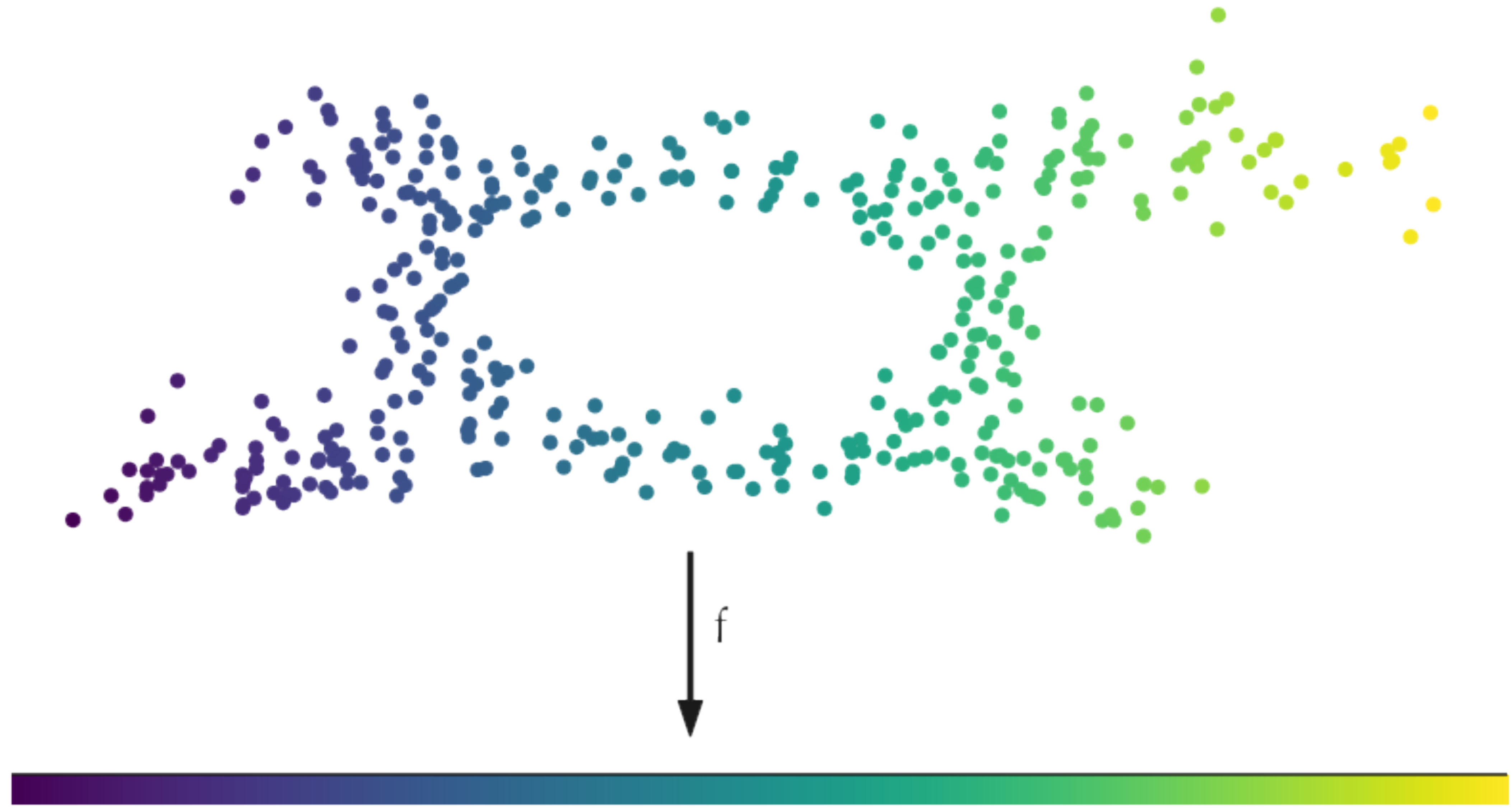


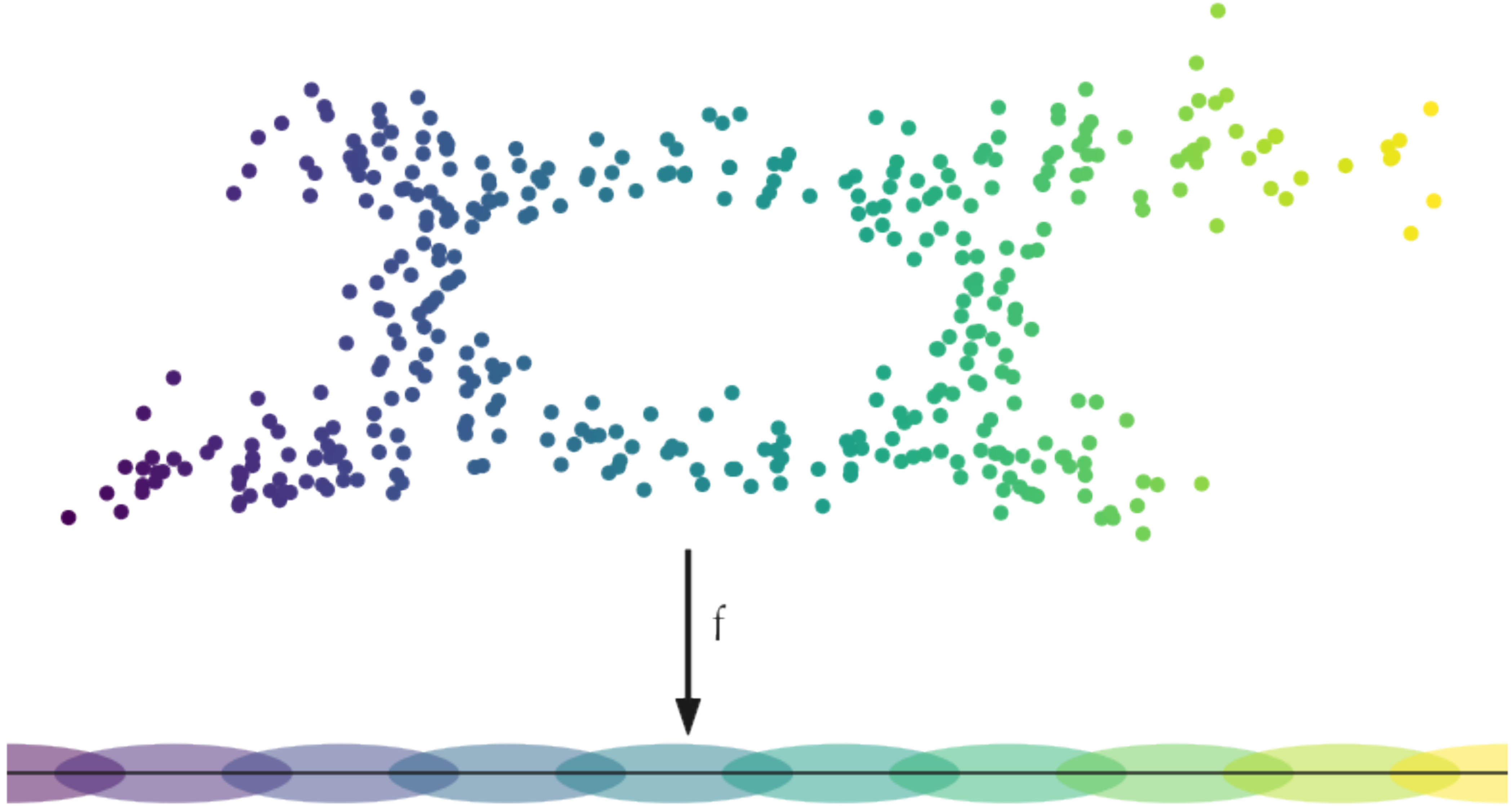
Approximating Reeb Graphs

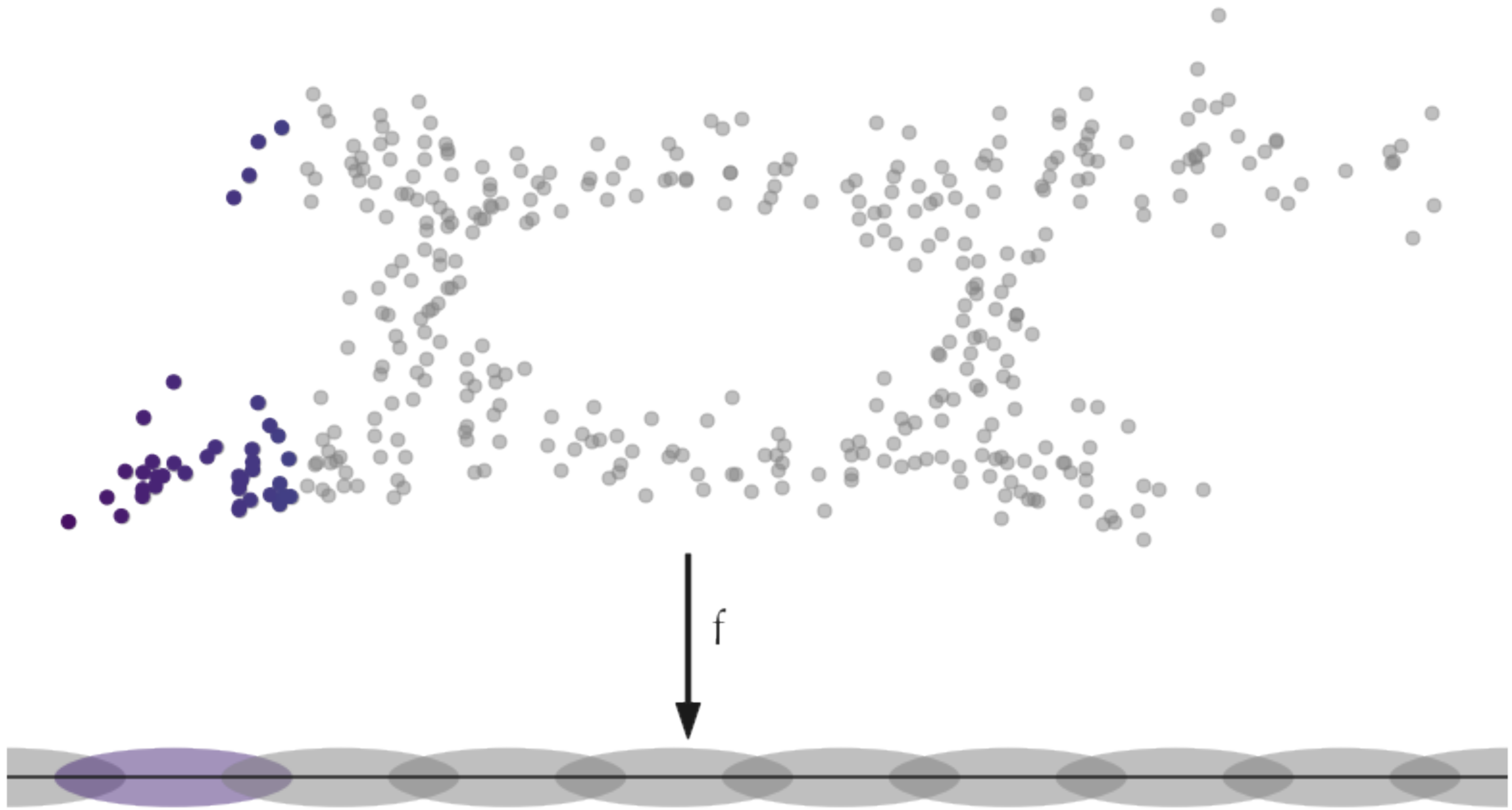



f

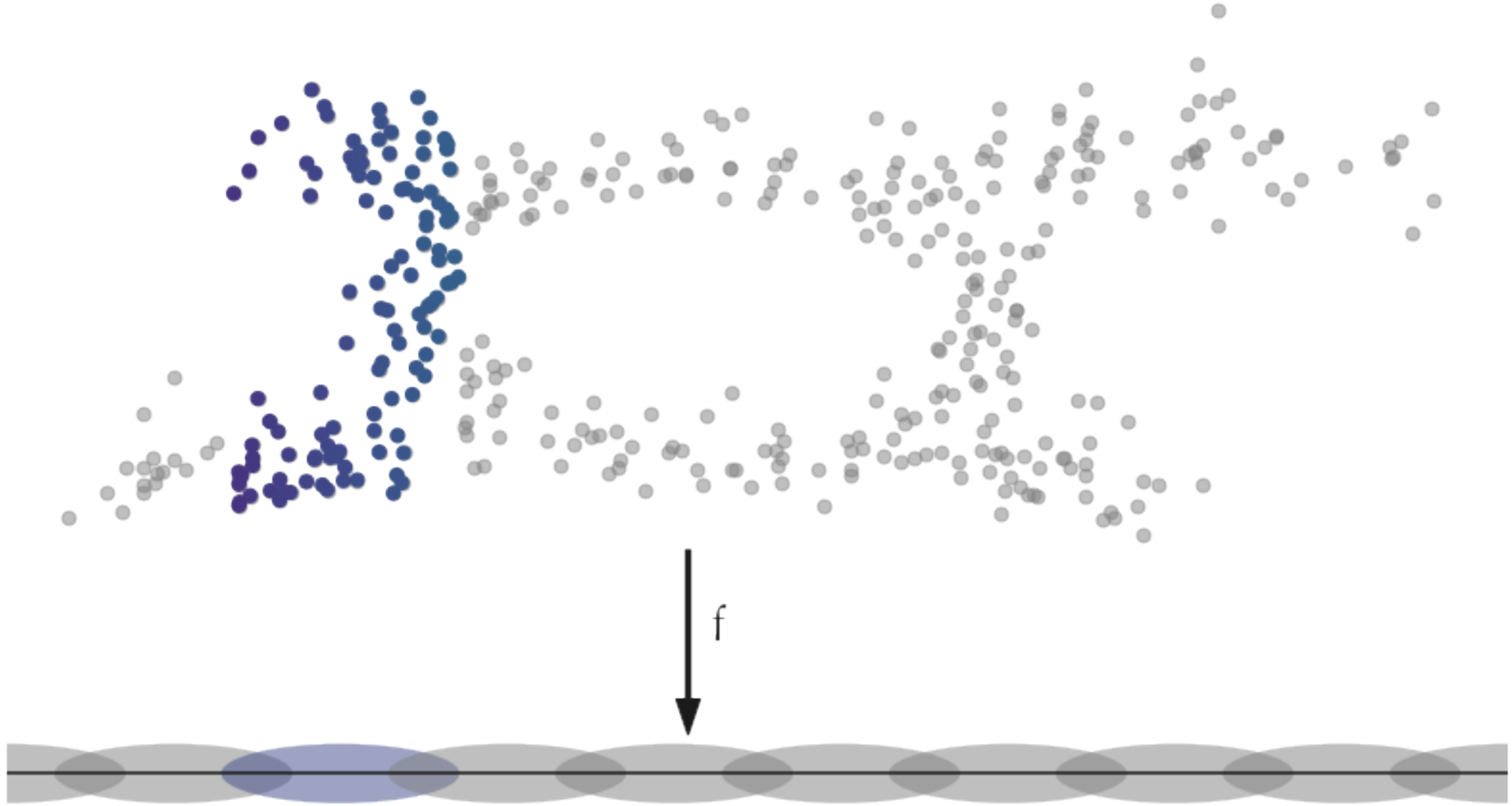
INSTITUT TUTTE INSTITUTE



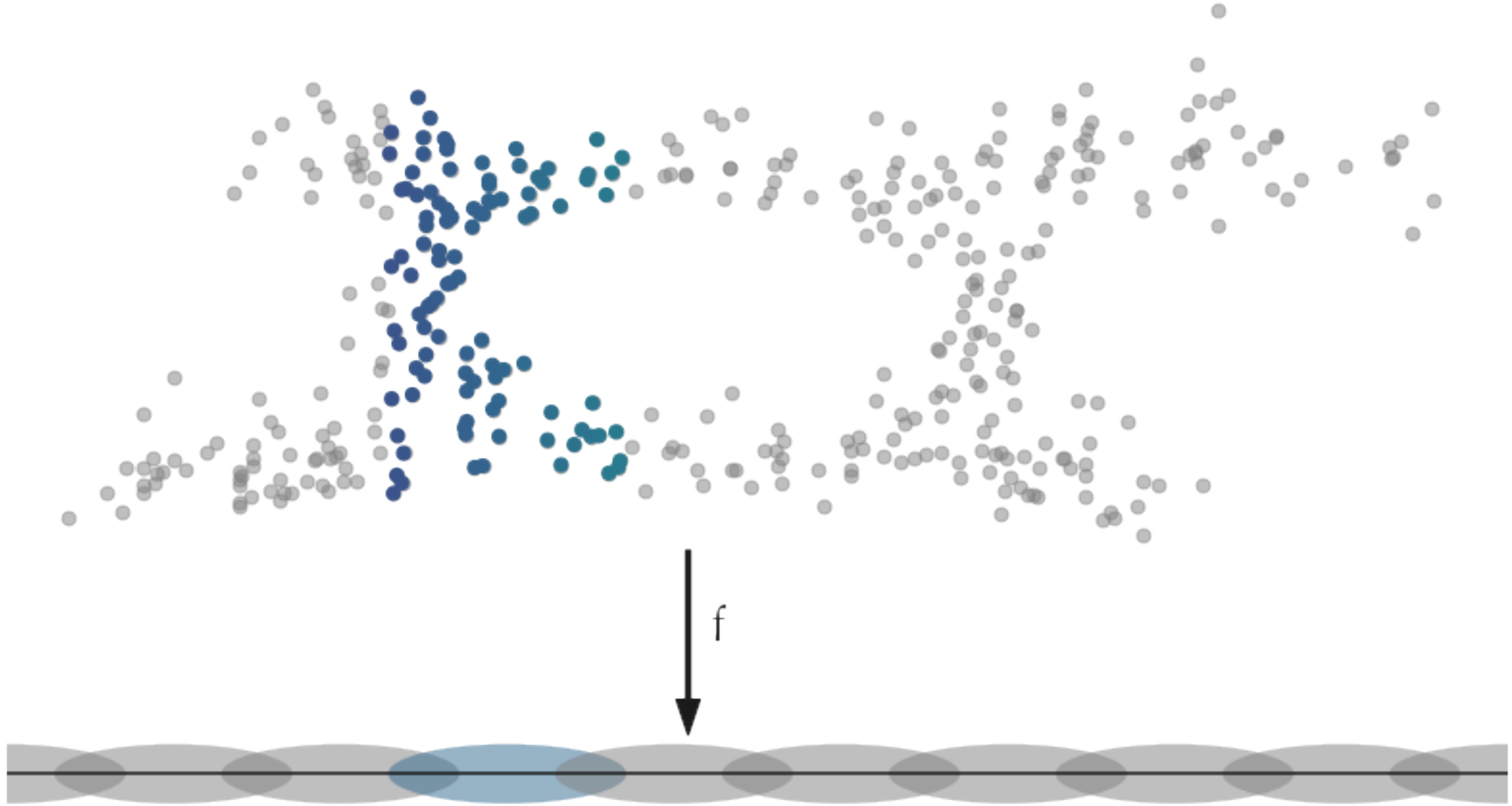

f

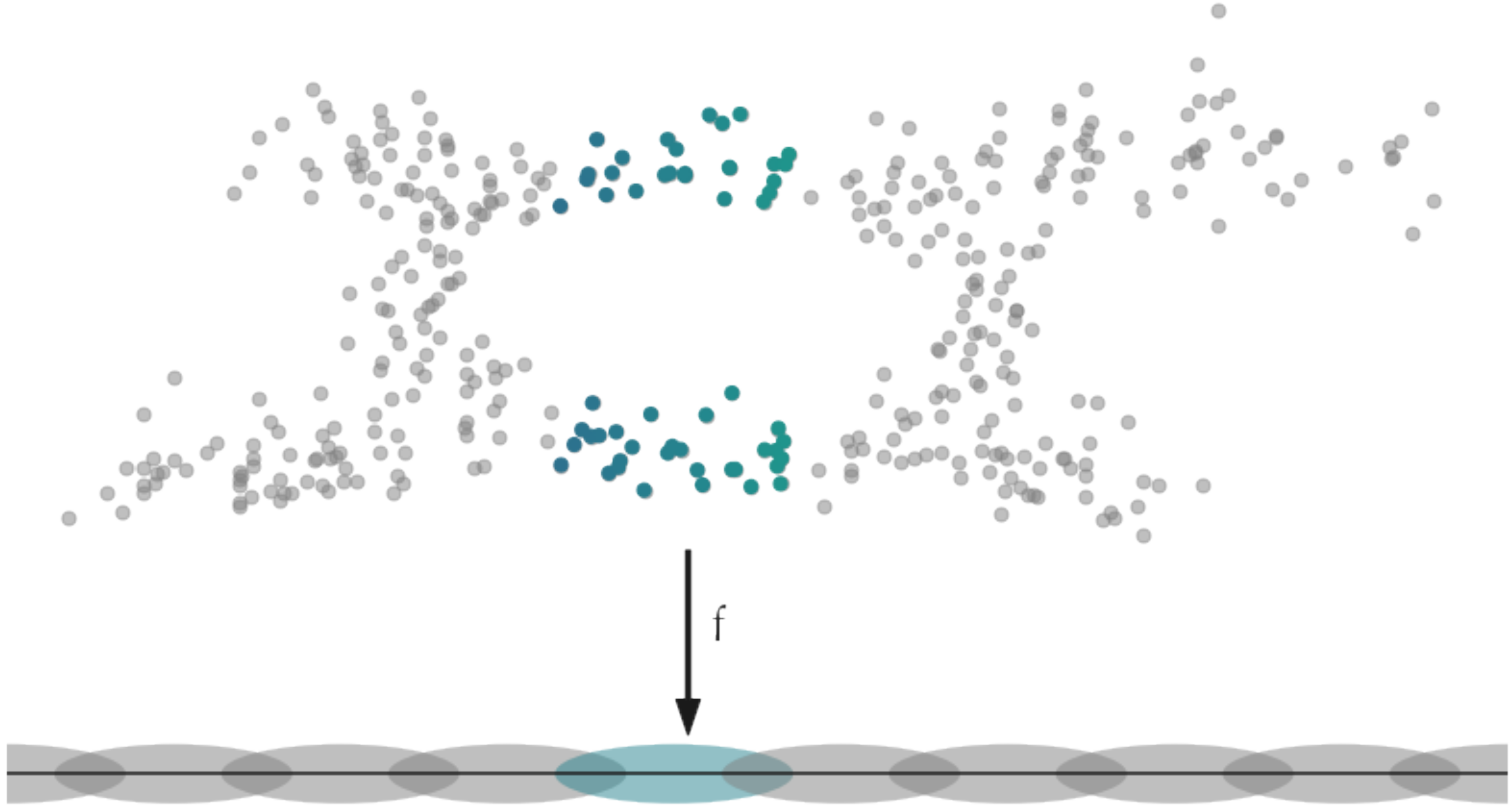

f

INSTITUT TUTTE INSTITUTE

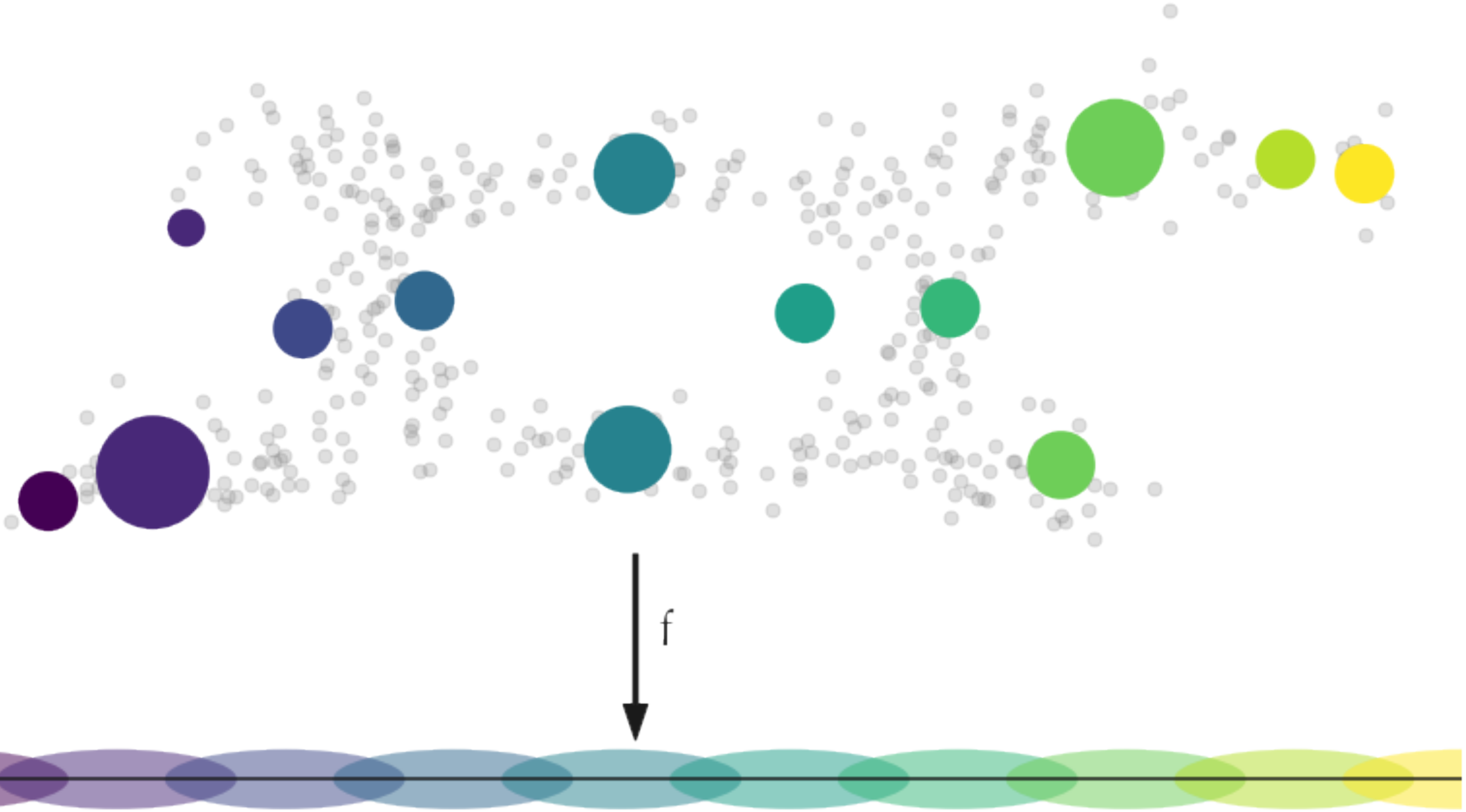


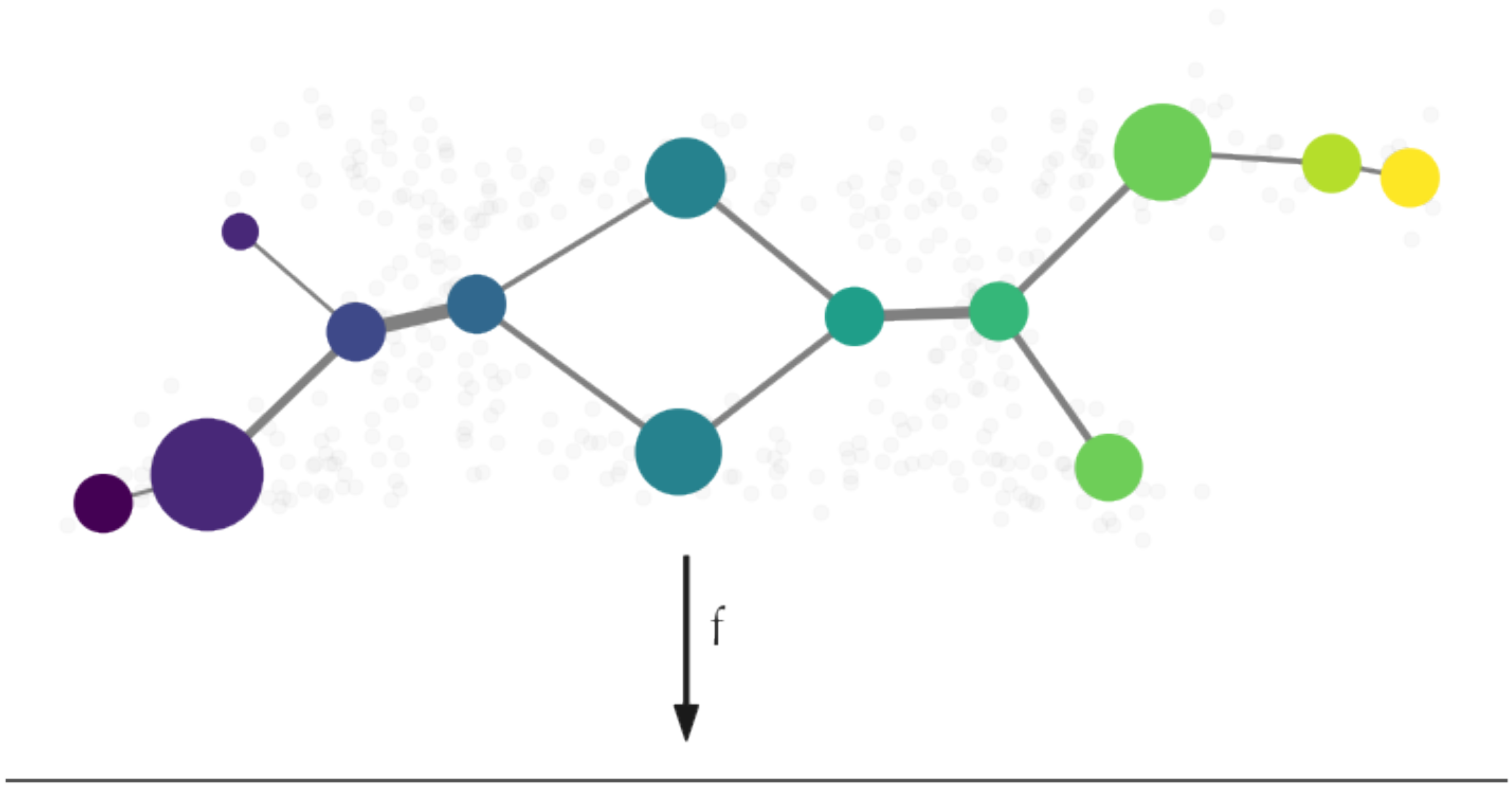
INSTITUT
TUTTE
INSTITUTE



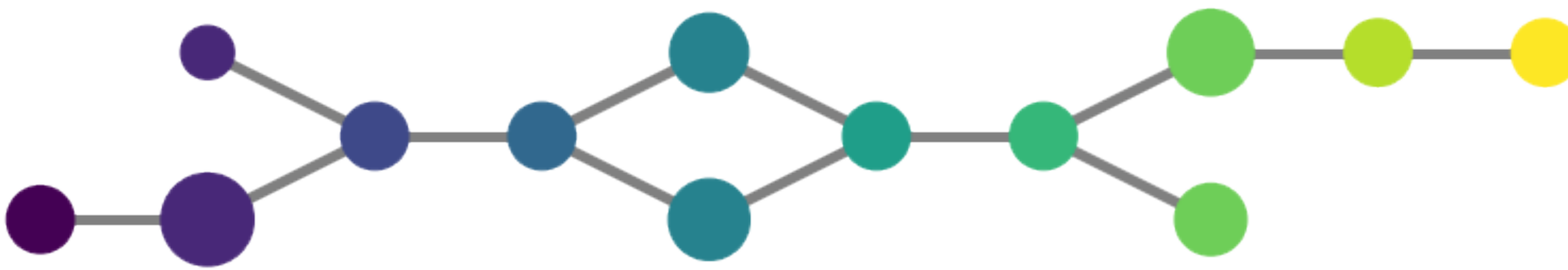

f

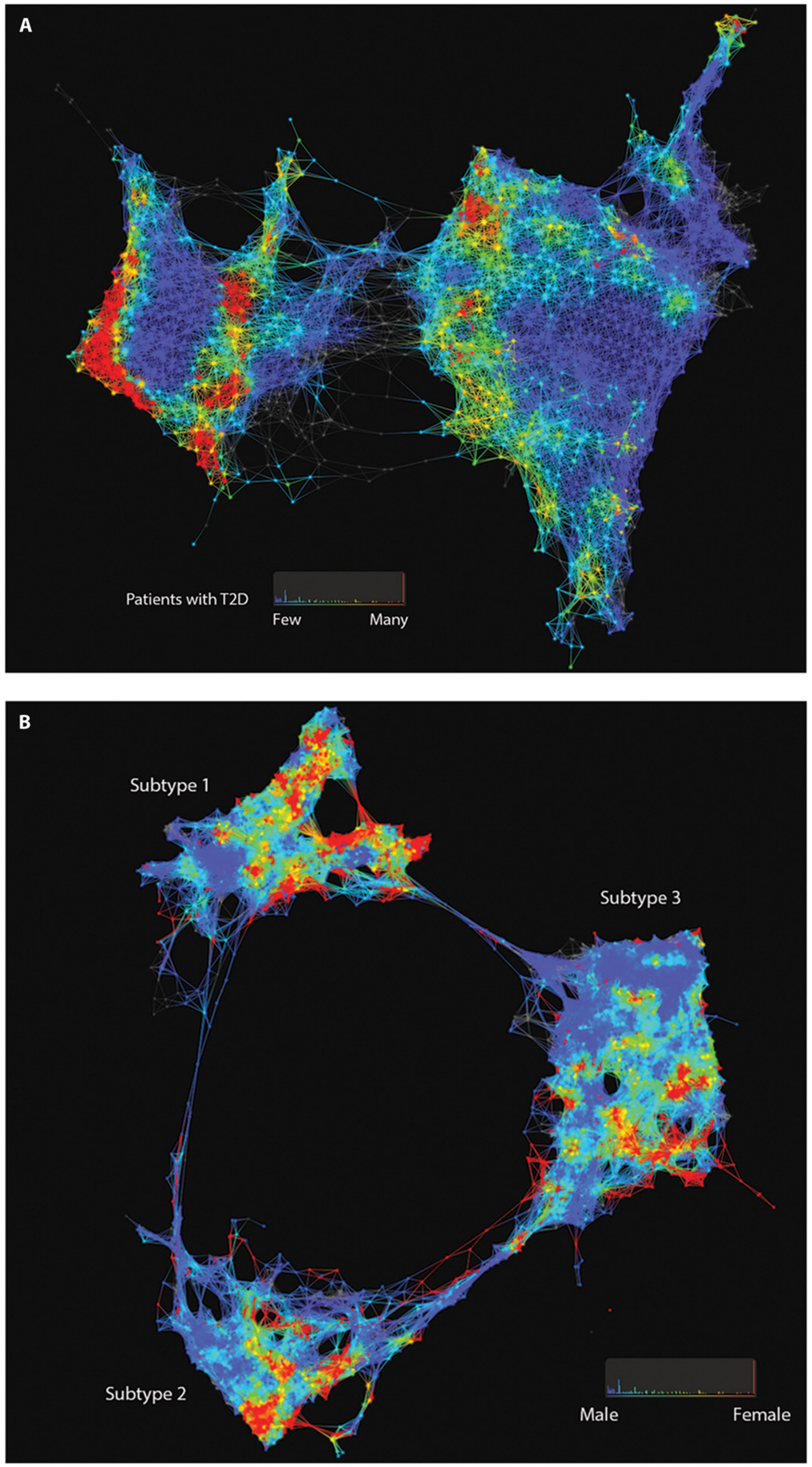
INSTITUT TUTTE INSTITUTE




f

INSTITUT TUTTE INSTITUTE





Identification of type 2 diabetes subgroups through topological analysis of patient similarity, Li et al, 2015

UMAPPER

INSTITUT TUTTE INSTITUTE

UMAP provides exact positioning and relationships among all points

UMAP doesn't require a filter function to operate

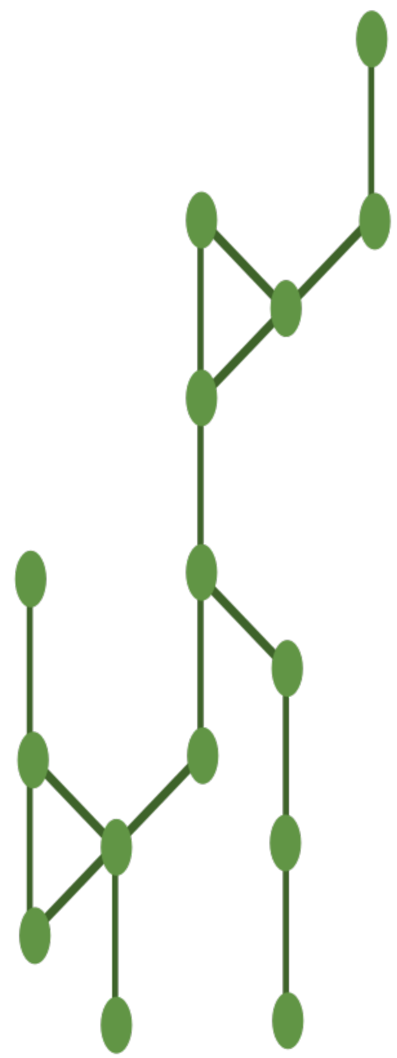
MAPPER is very expressive for suitable filter functions

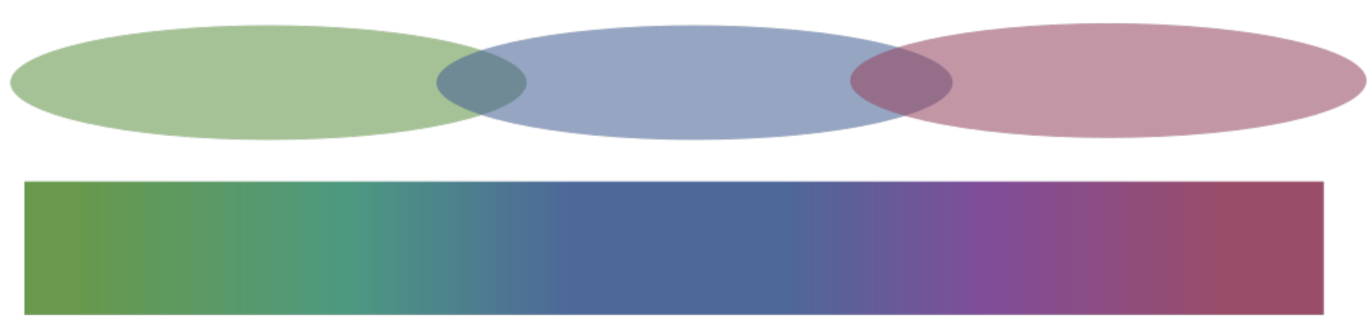
UMAP can't make use of a filter function

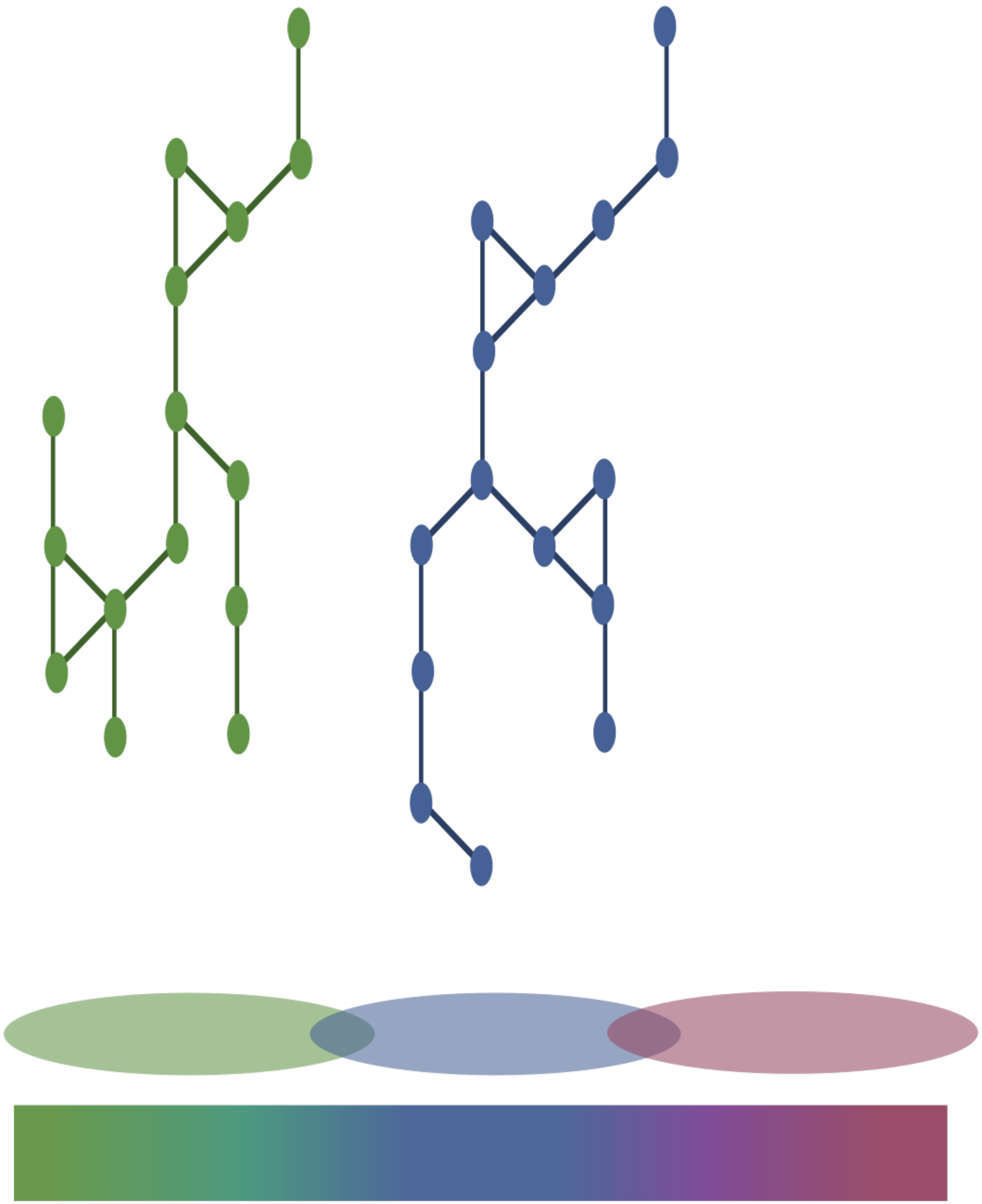
Can we get the benefits of both?

What if we apply UMAP instead of clustering?

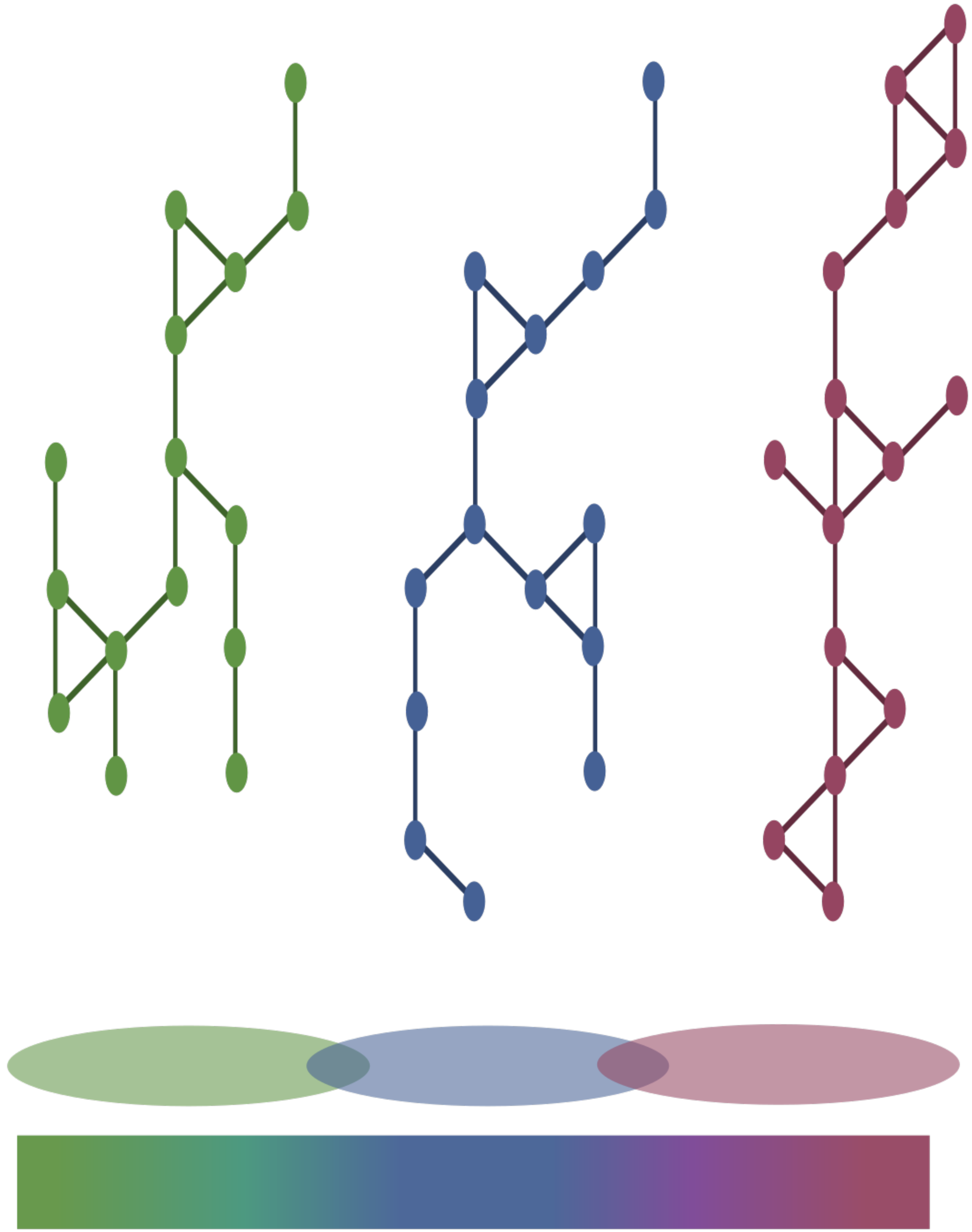
INSTITUT TUTTE INSTITUTE



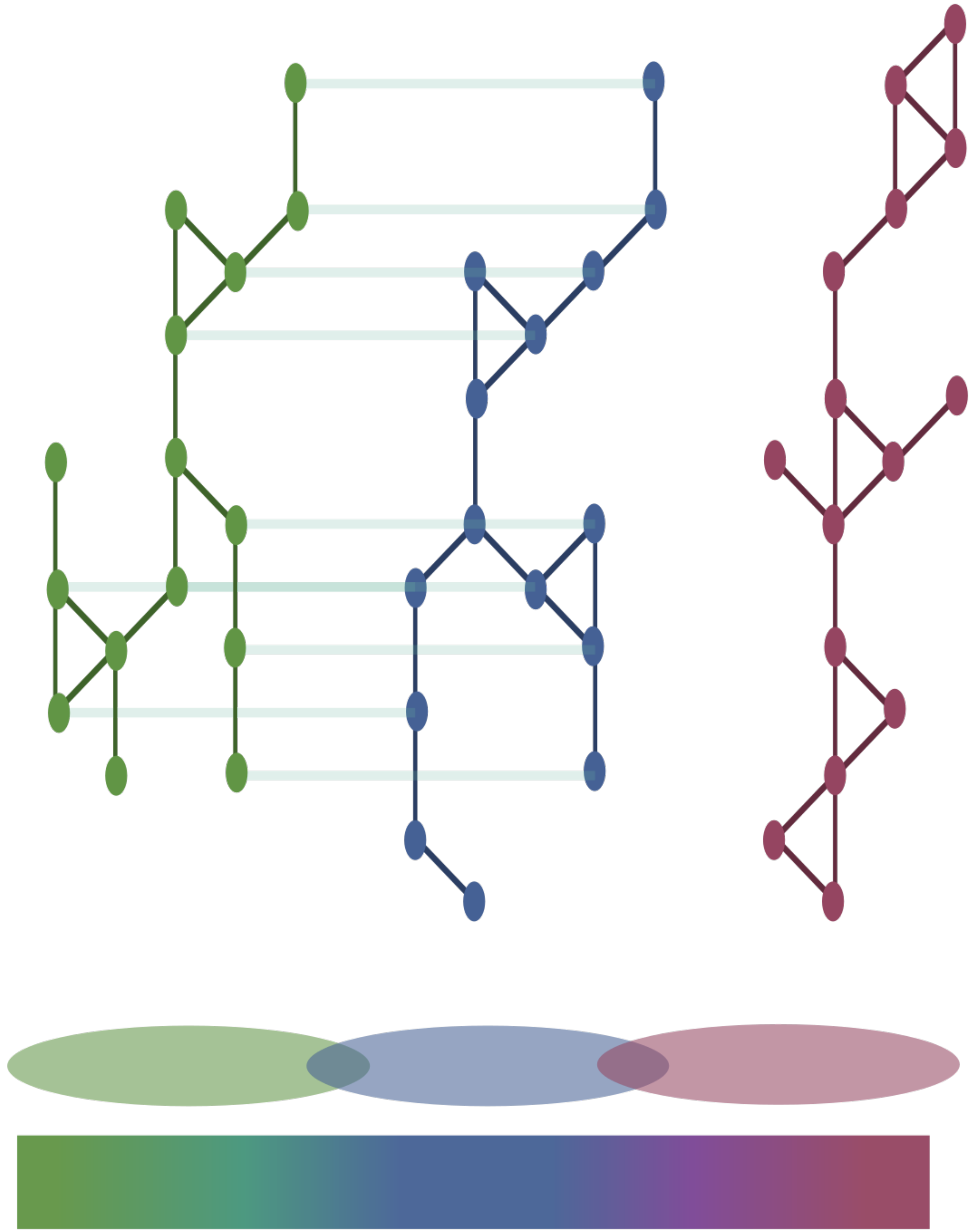


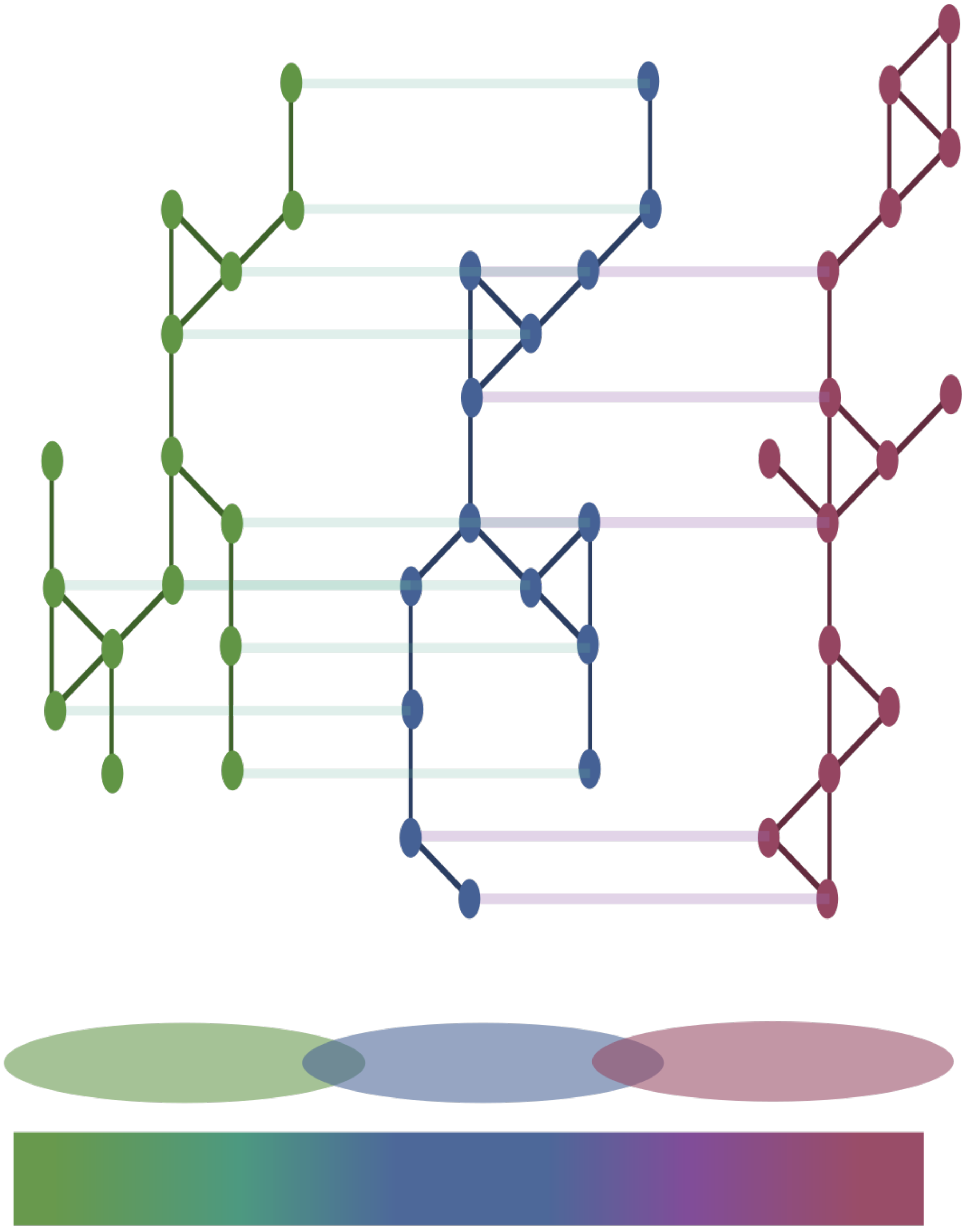


INSTITUT TUTTE INSTITUTE



Can we build a MAPPER simplicial structure?





A bi-simplicial structure?

Optimize the UMAP representations simultaneously, using MAPPER simplices for alignment

U_i be the 1-skeleton in the i^{th} UMAP complex

M be the 1-simplices in the MAPPER complex

Let

$$C_i = \sum_{u \in U_i} \mu(u) \log\left(\frac{\mu(u)}{\nu(u)}\right) + (1 - \mu(u)) \log\left(\frac{1 - \mu(u)}{1 - \nu(u)}\right)$$

and

$\ell oc(\partial_i(m))$ be the current location of the zero simplex $\partial_i(m)$

Optimize

$$\sum_i C_i + \lambda \sum_{m \in M} \| \ell oc(\partial_0(m)) - \ell oc(\partial_1(m)) \|_2^2$$

Can we have a notion of membership strength on MAPPER edges?

In standard MAPPER we might use Jaccard Index between clusters

Apply Jaccard Index to "neighborhoods"

For $x \in U_i$ let $N_{\rightarrow}(x)$ denote the neighbours of x in U_i that are faces of simplices in M spanning U_i and U_{i+1}

For $x \in U_i$ let $N_{\leftarrow}(x)$ denote the neighbours of x in U_i that are faces of simplices in M spanning U_i and U_{i-1}

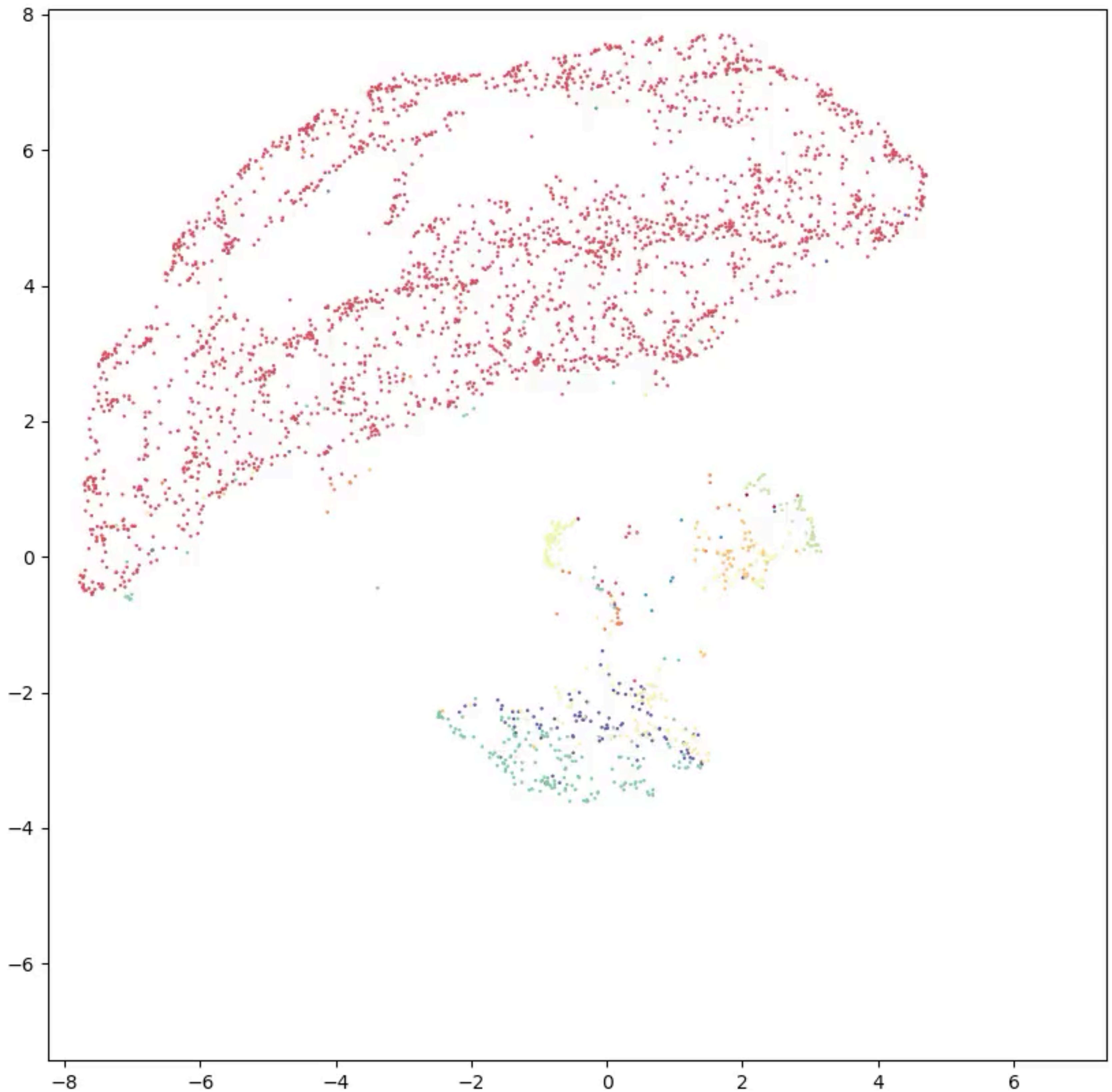
INSTITUT TUTTE INSTITUTE

$$\mathrm{Jacc}(m) := \frac{\left| N_{\rightarrow}(\partial_0(m)) \cap N_{\leftarrow}(\partial_1(m)) \right|}{\left| N_{\rightarrow}(\partial_0(m)) \cup N_{\leftarrow}(\partial_1(m)) \right|}$$

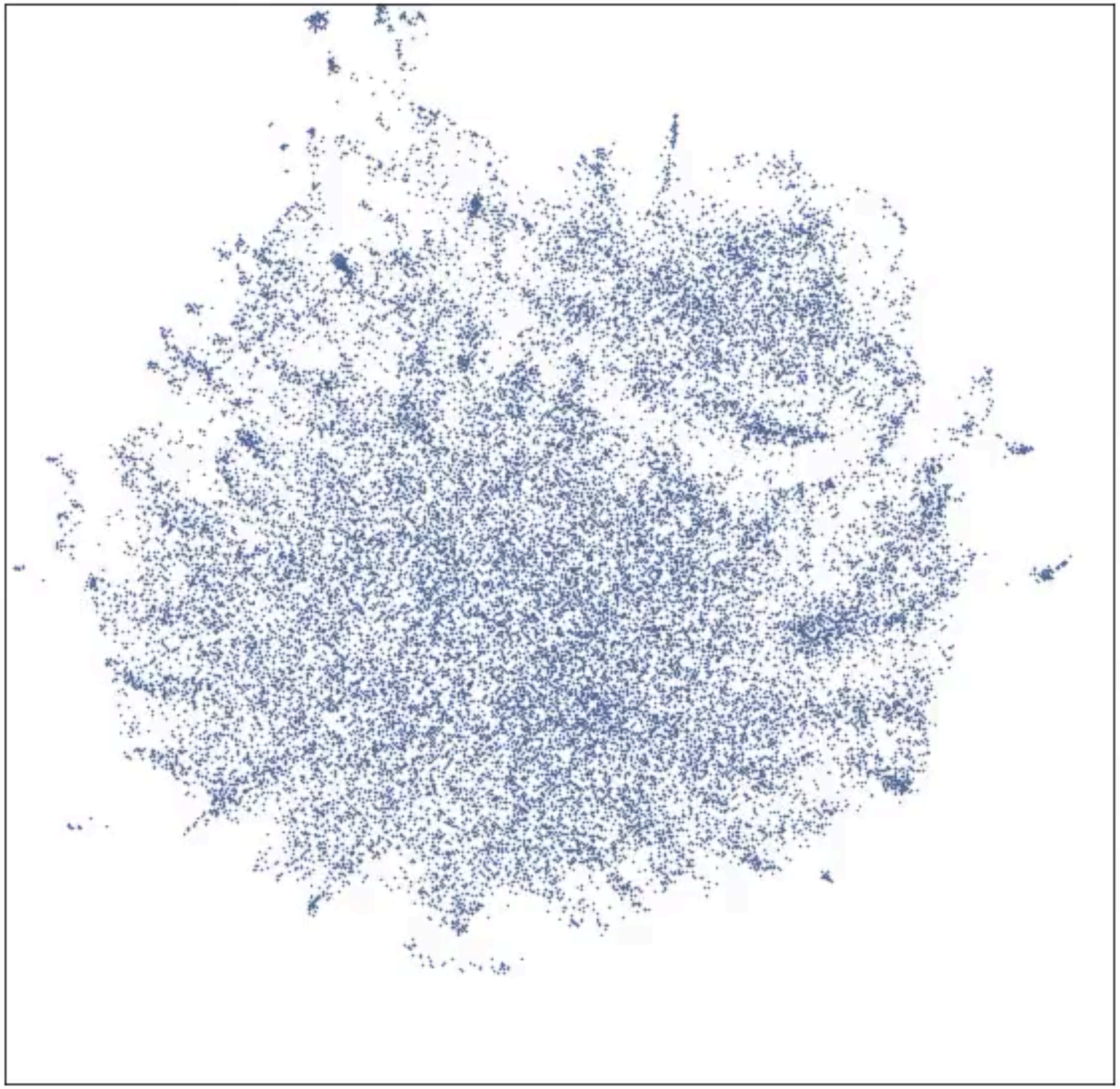
Optimize

$$\sum_i C_i + \lambda \sum_{m \in M} \mathrm{Jacc}(m) \| \ell oc(\partial_0(m)) - \ell oc(\partial_1(m)) \|_2^2$$

We can still do this with stochastic gradient descent!



8
6
4
2
0
−2
−4
−6
−8
−6
−4
−2
0
2
4
6



Conclusions

INSTITUT TUTTE INSTITUTE

It is possible to combine the strengths of UMAP and MAPPER and provide rich data visualizations

Richer topological structures and more advanced approaches are possible

INSTITUT TUTTE INSTITUTE

https://github.com/lmcinnes/umap/tree/0.5dev

demo: https://bit.ly/2zxcWsZ