TDA for medical data analysis, how we can help in the current pandemic?

Paweł Dłotko

Swansea $\rightarrow$ **Dioscuri Centre for TDA**.

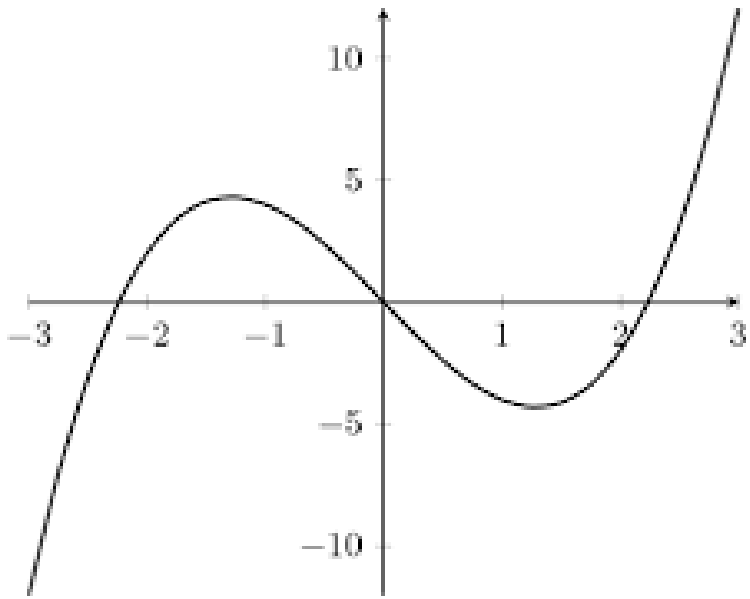
Wish I could be in Rome with you!

Plan for today

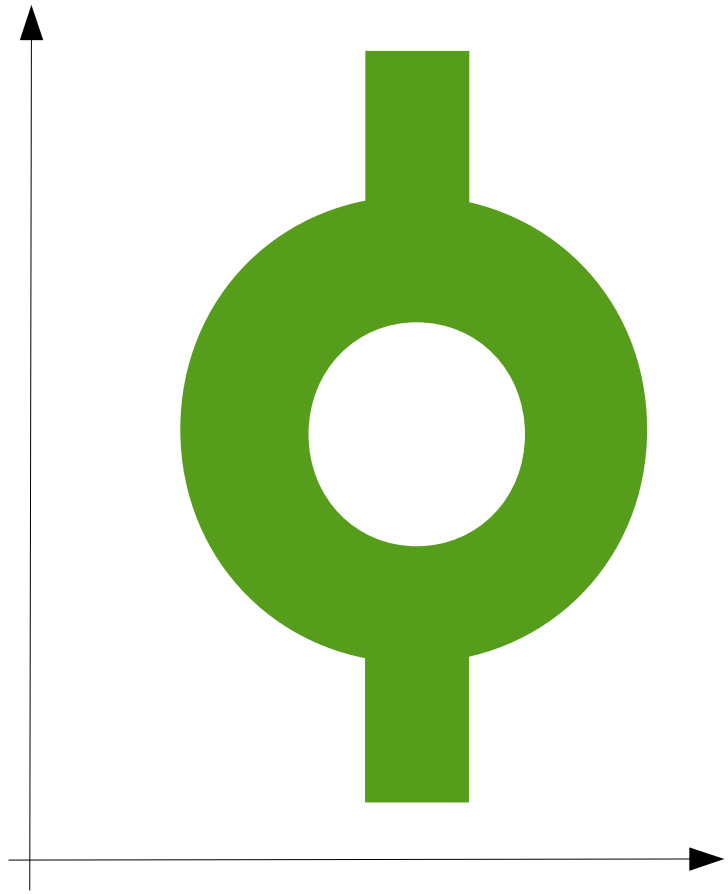
1. Mapper algorithm(s) – topological tools to visualize high dimensional data,
2. Clinical data, individual therapies through the lens of clinical outcomes,
3. Persistent homology and bones.
4. Knots and TDA?

Mapper(s) algorithms, what they are mend to do?

1. Building graph–based / simplicial complex models of data.
2. Plot function values on the top of them.

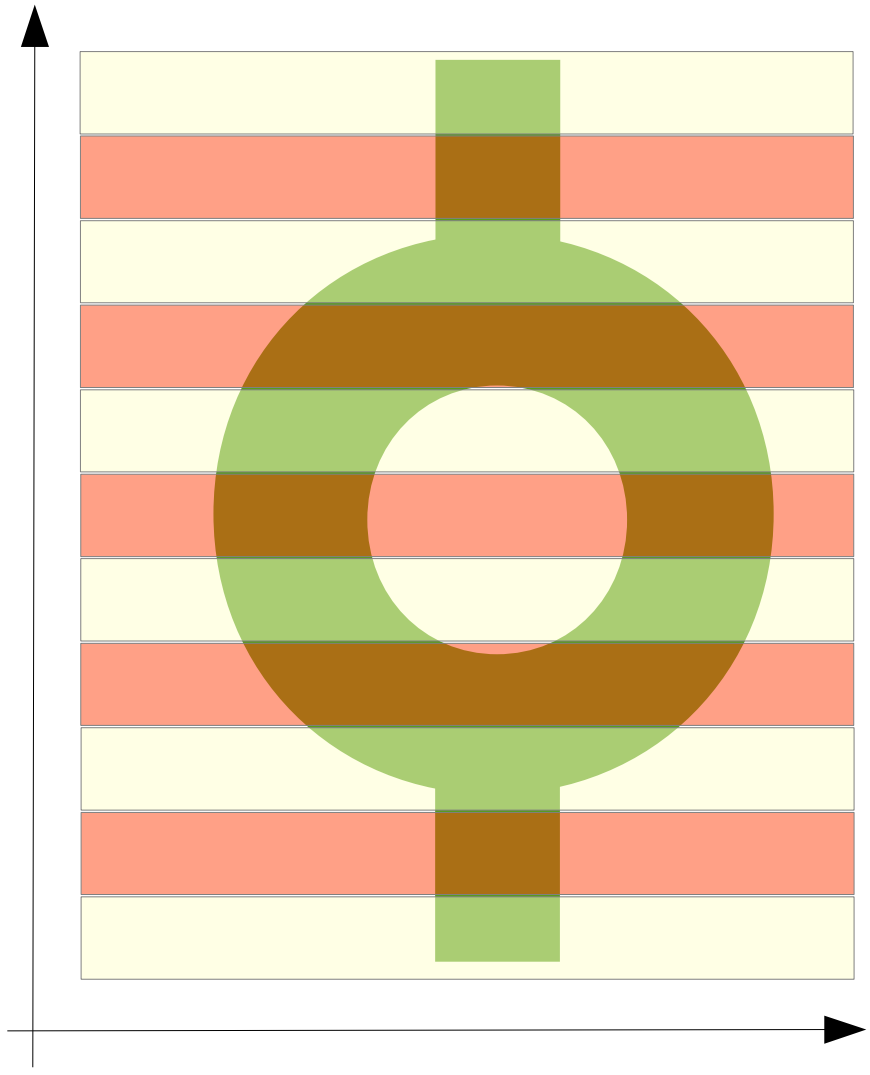


Conventional Mapper

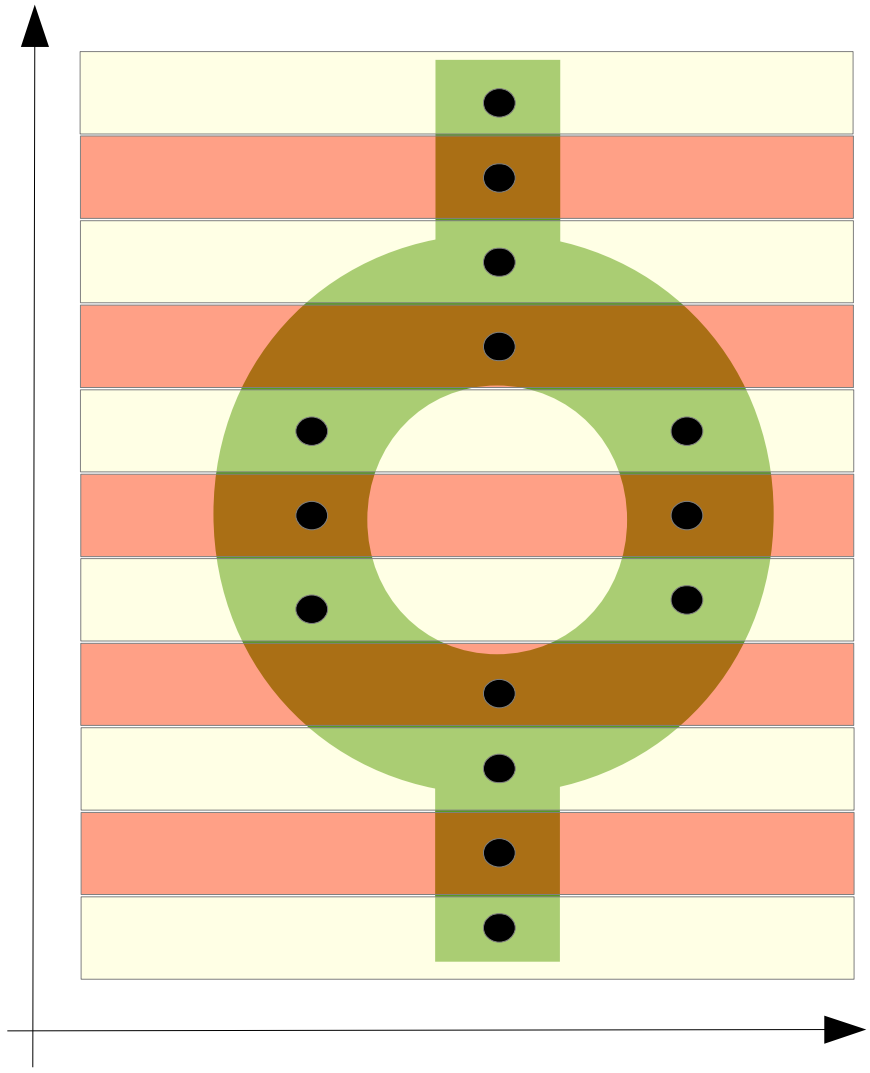


Manifold M, $f : M \to \mathbb{R}$

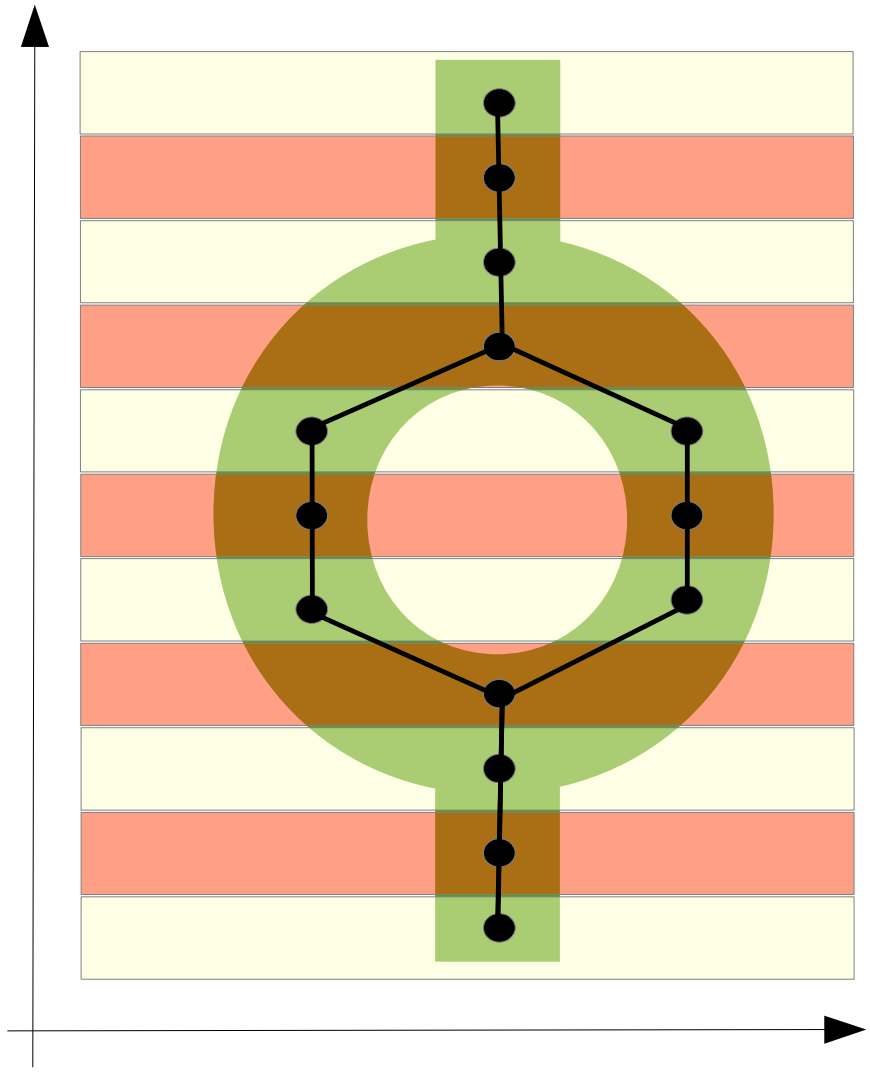
Fibers of lenses



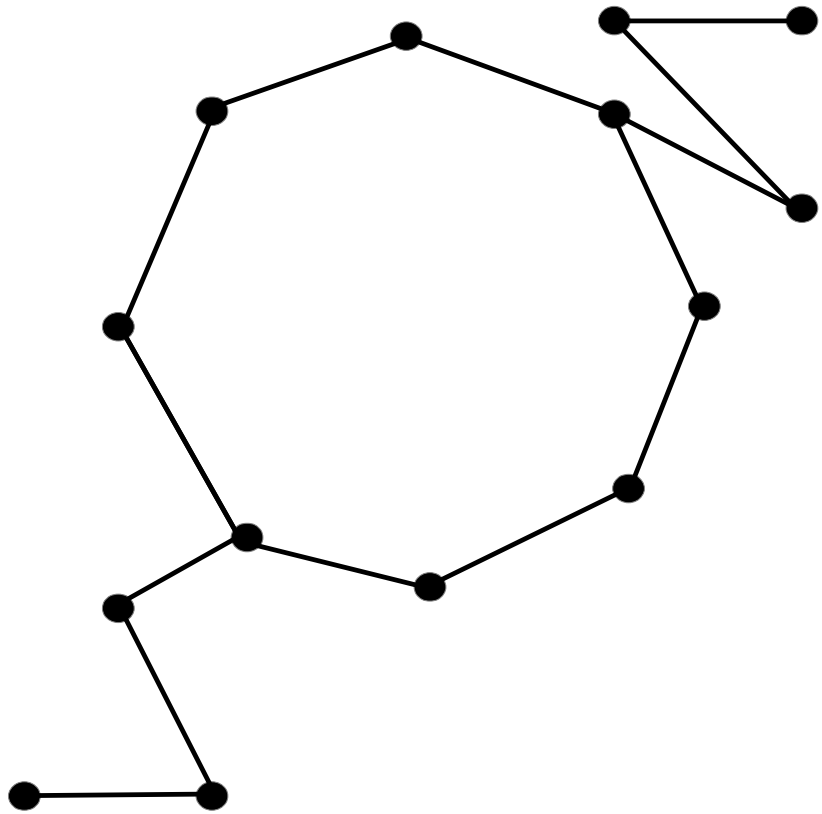
Vertices correspond to connected components in fibers



Edges between components with nonempty intersection



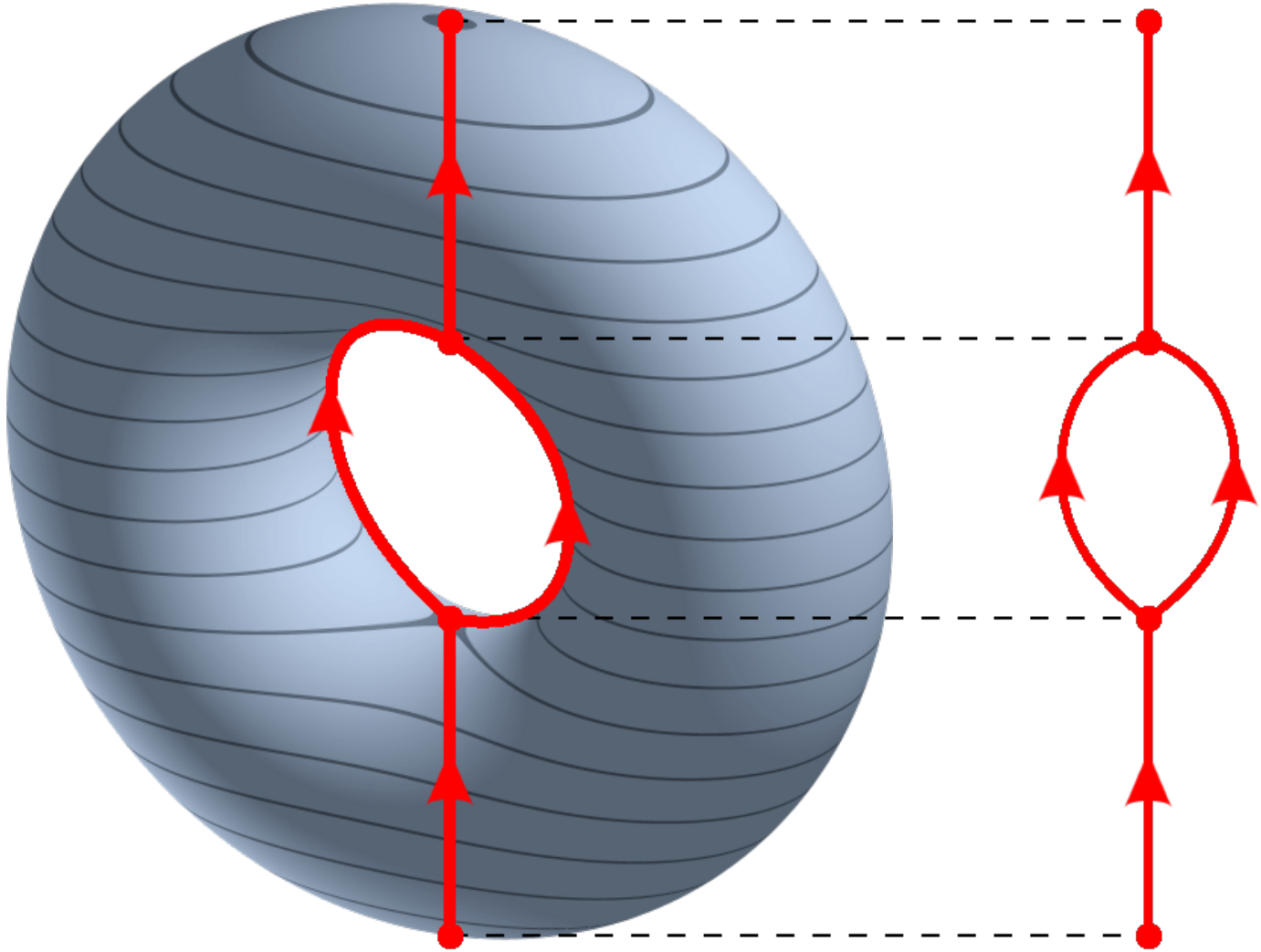
Obtained abstract graph is a Mapper graph



Theoretical motiviation – the Reeb graph

1. $\mathcal{M}$ - manifold, $f : \mathcal{M} \to \mathbb{R}^1$.
2. $x, y \in \mathcal{M}$, xRy iff $f(x) = f(y)$ and x and y are in the same connected component of $f^{-1}(f(x))$.
3. Reeb graph $R = M/_R$.

Reeb graph



Reeb graph

1. $\mathcal{M}$ - manifold, $f : \mathcal{M} \to \mathbb{R}^n$, typically for $n = 1$.
2. $x, y \in \mathcal{M}$, xRy iff $f(x) = f(y)$ and x and y are in the same connected component of $f^{-1}(f(x))$.
3. Reeb graph $R = M/_R$.
4. There are a few adjustments needed to make it work for discrete metric spaces.

Reeb graph adjustments for point clouds

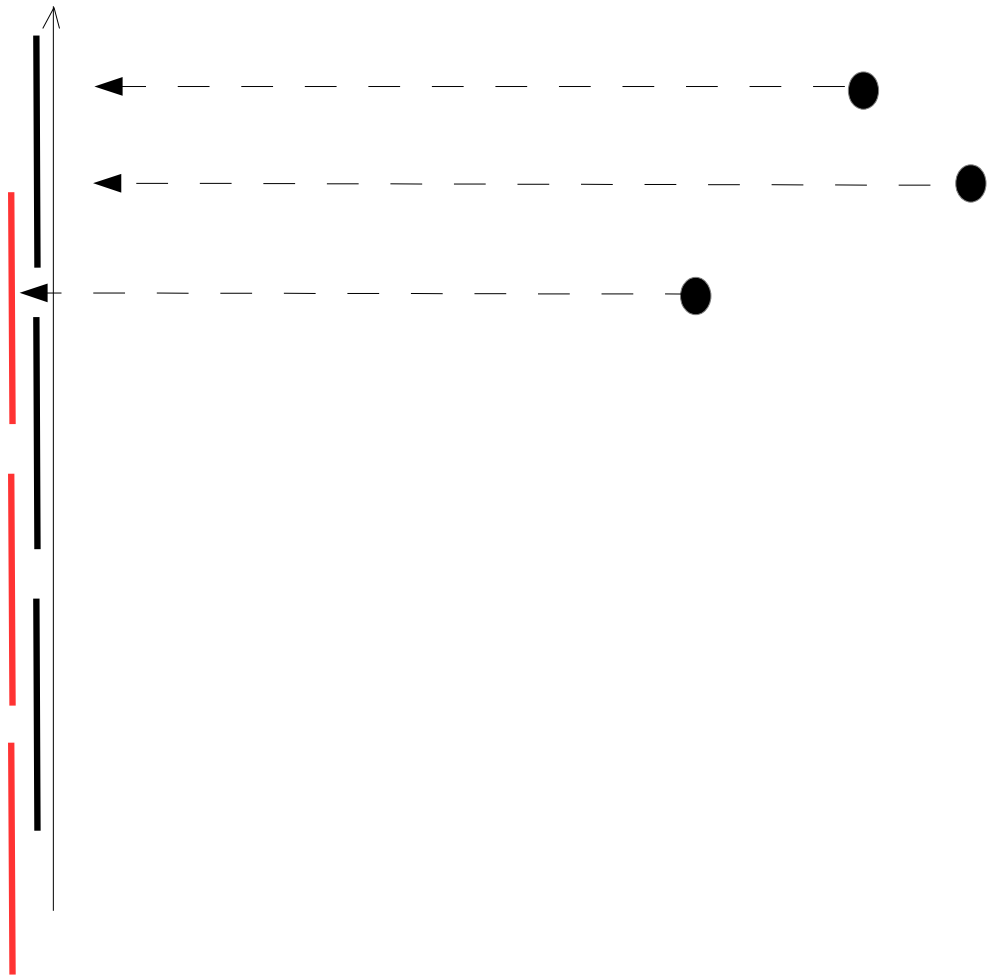
1. Take a cover of $\mathbb{R}$ with a collection of overlapping intervals.
2.
3.
4.

Cover of a line

Reeb graph adjustments for point clouds

1. Take a cover of $\mathbb{R}$ with a collection of overlapping intervals.
2. $f(x) = f(y)$ is replaced by statement that $f(x)$ is close to $f(y)$.
3. In our case, that $f(x)$ and $f(y)$ are mapped to the same interval in the cover.
4.

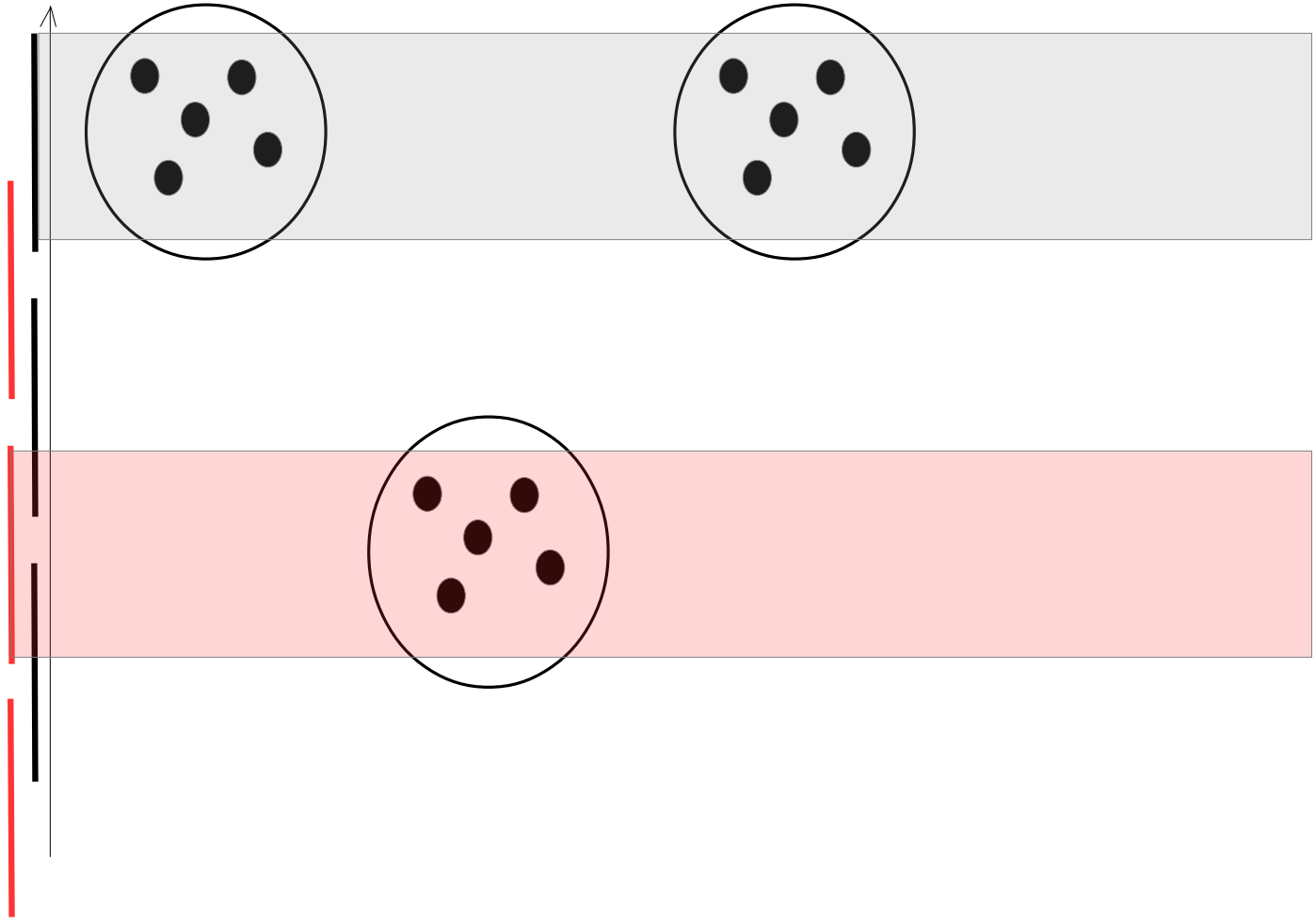
Proximity of points



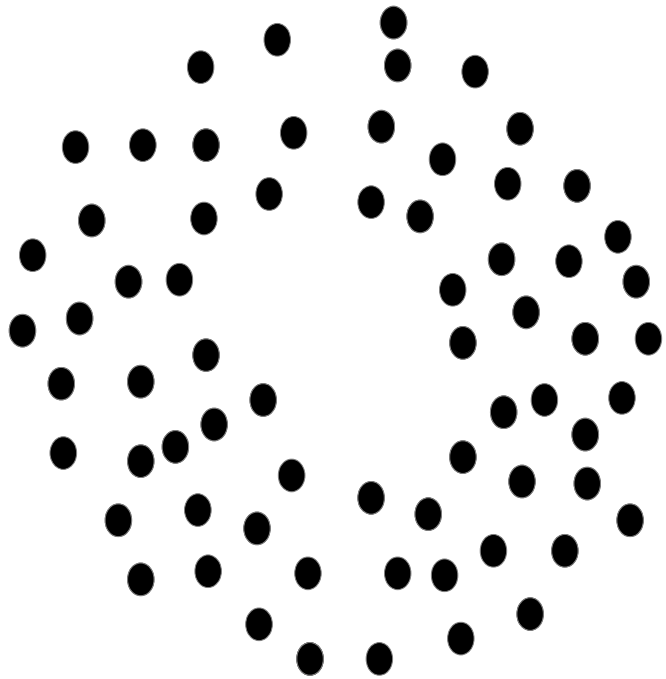
Reeb graph adjustments for point clouds

1. Take a cover of $\mathbb{R}$ with a collection of overlapping intervals.
2. $f(x) = f(y)$ is replaced by statement that $f(x)$ is close to $f(y)$.
3. In our case, that $f(x)$ and $f(y)$ are mapped to the same interval in the cover.
4. "The same connected component" replaced by requirement of belonging to the same cluster in the inverse image of a $\mathbb{R}$ cover element.

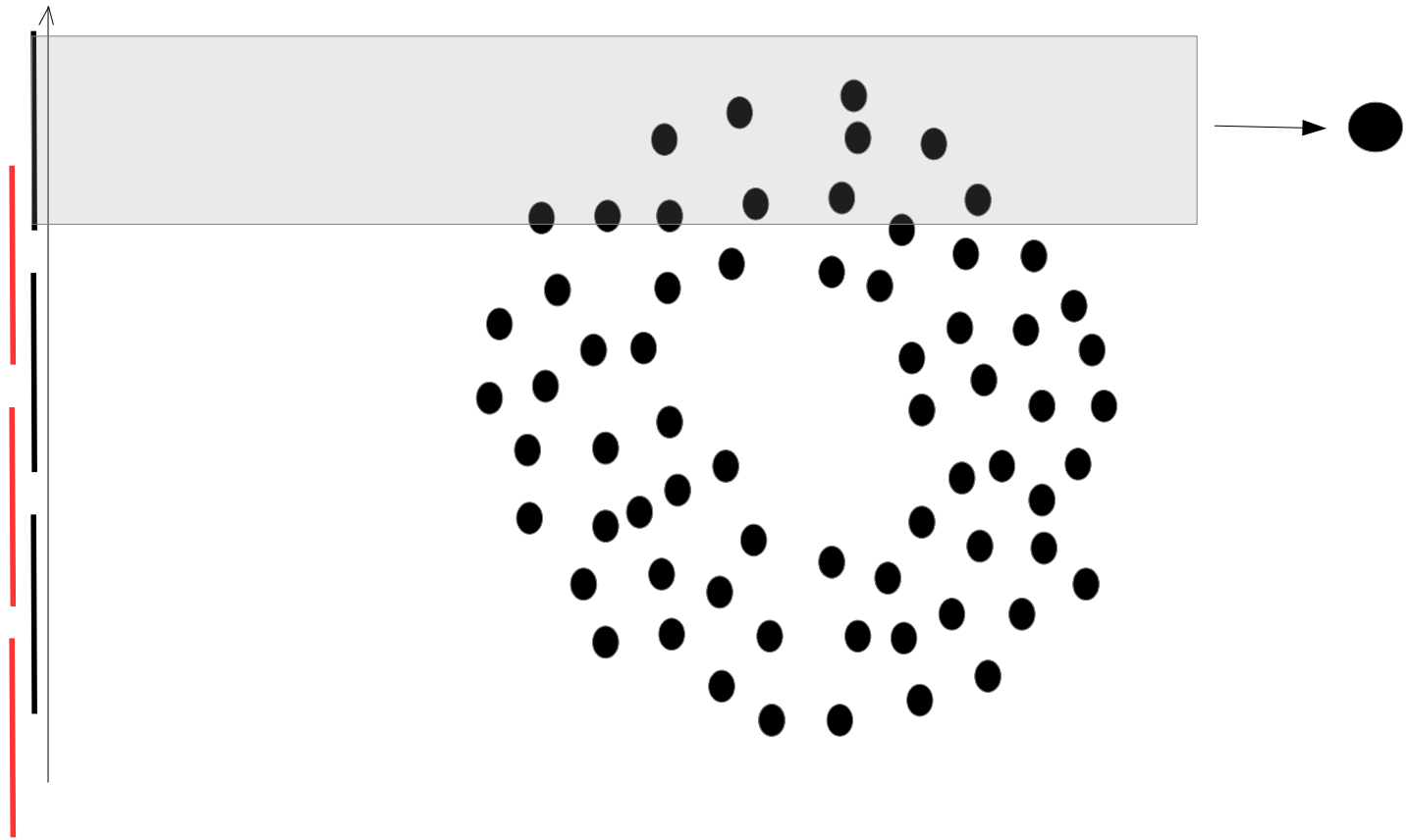
Connected components



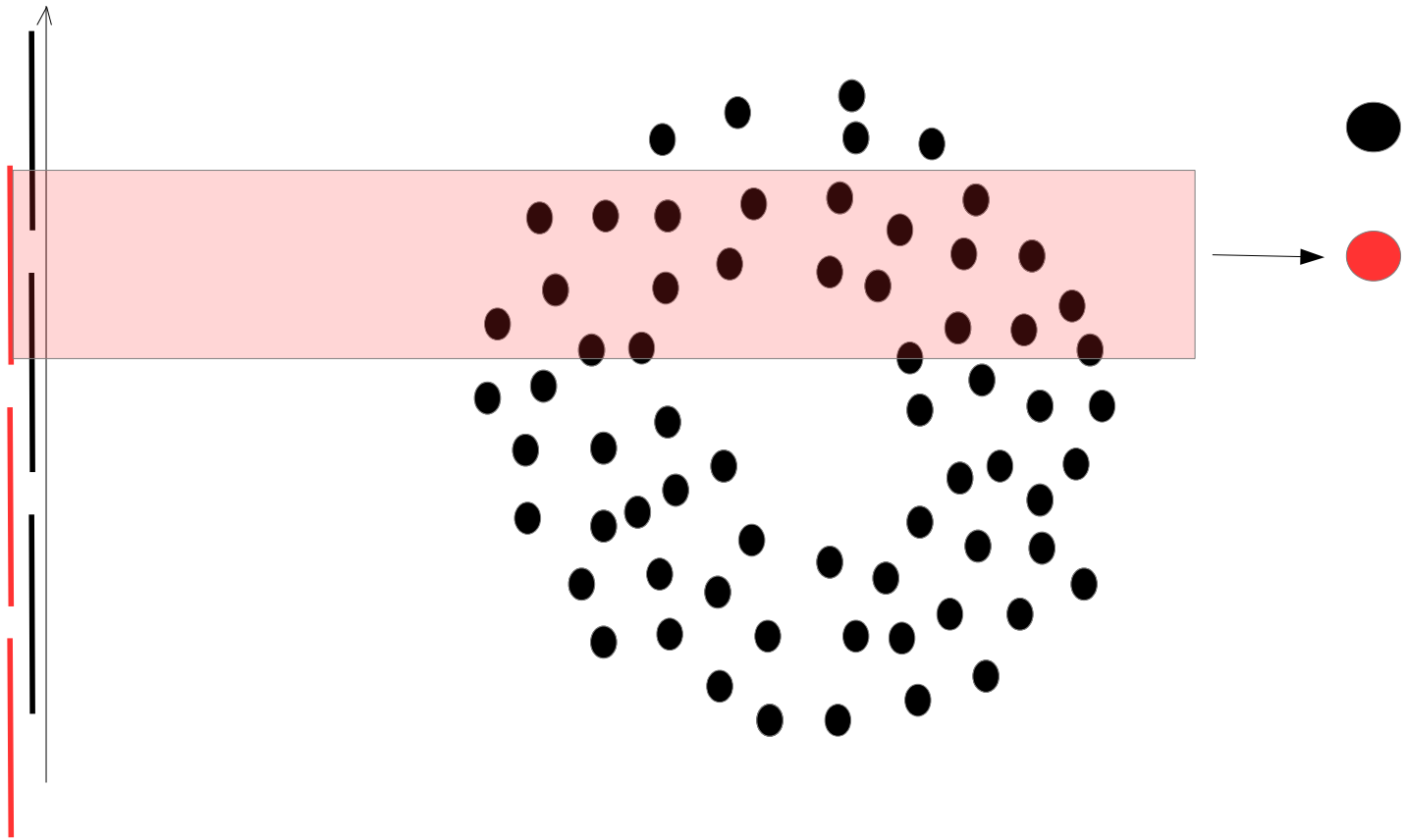
M – point cloud, $f : M \to \mathbb{R}$



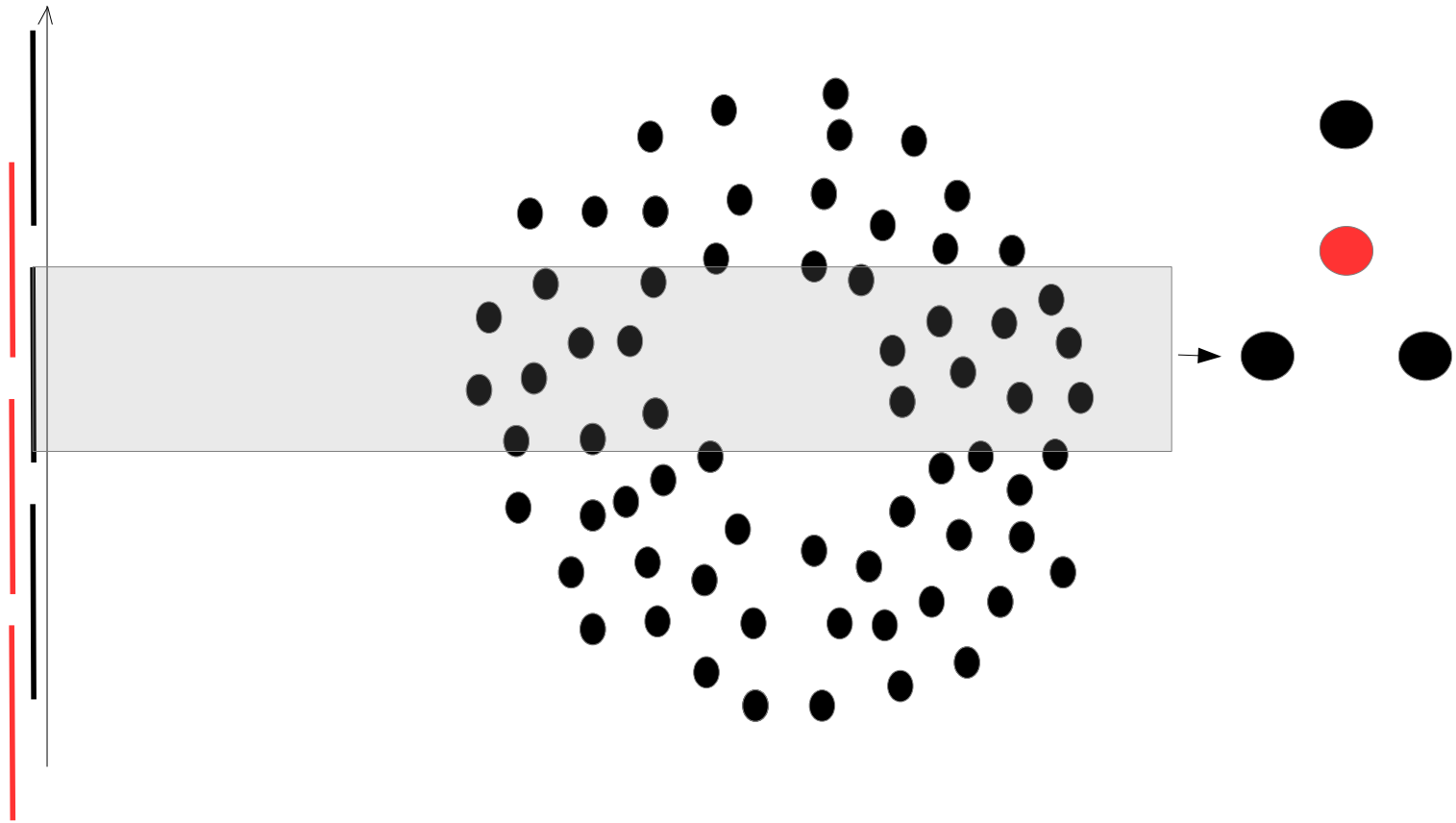
Conventional Mapper algorithm



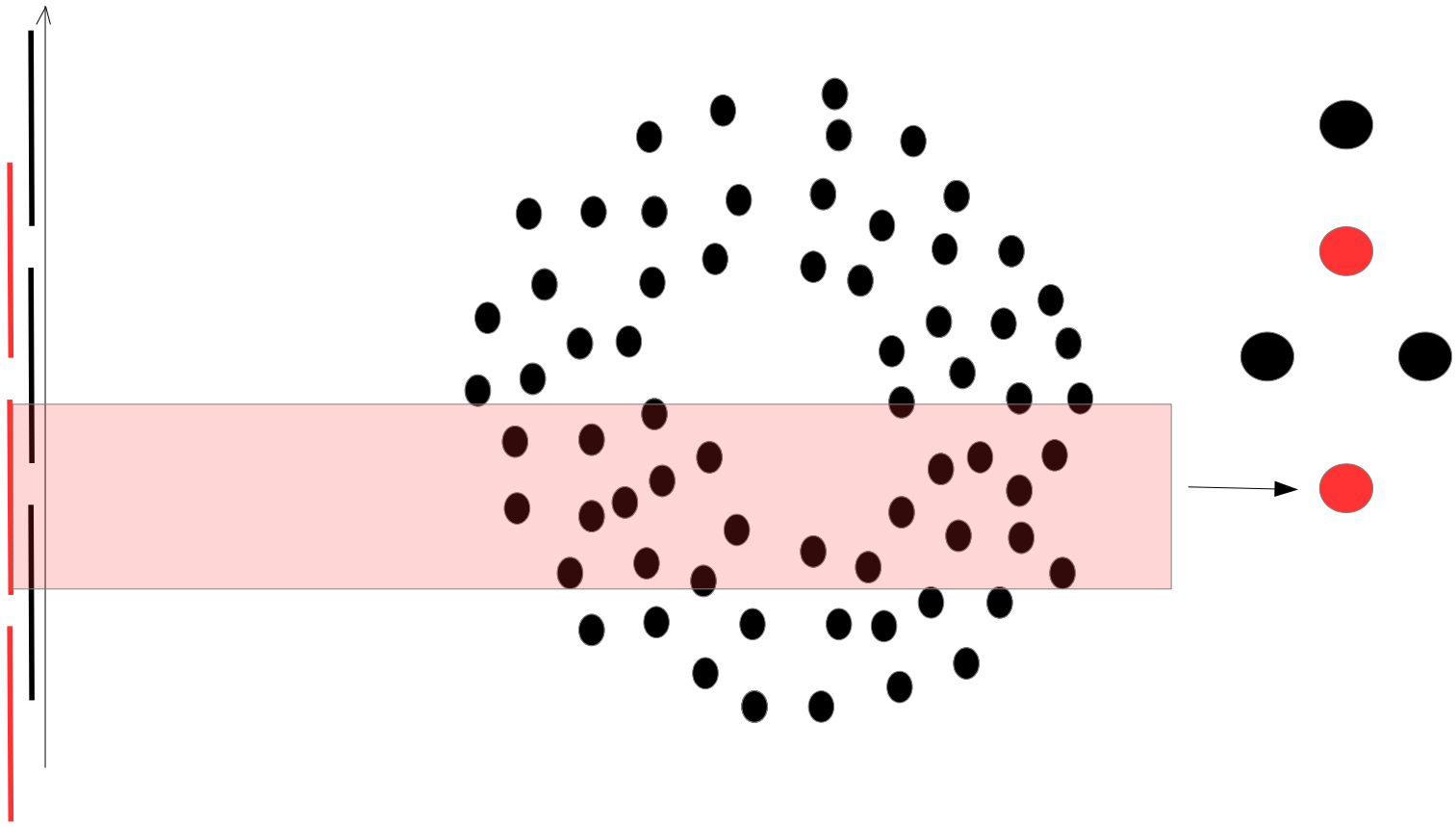
Conventional Mapper algorithm



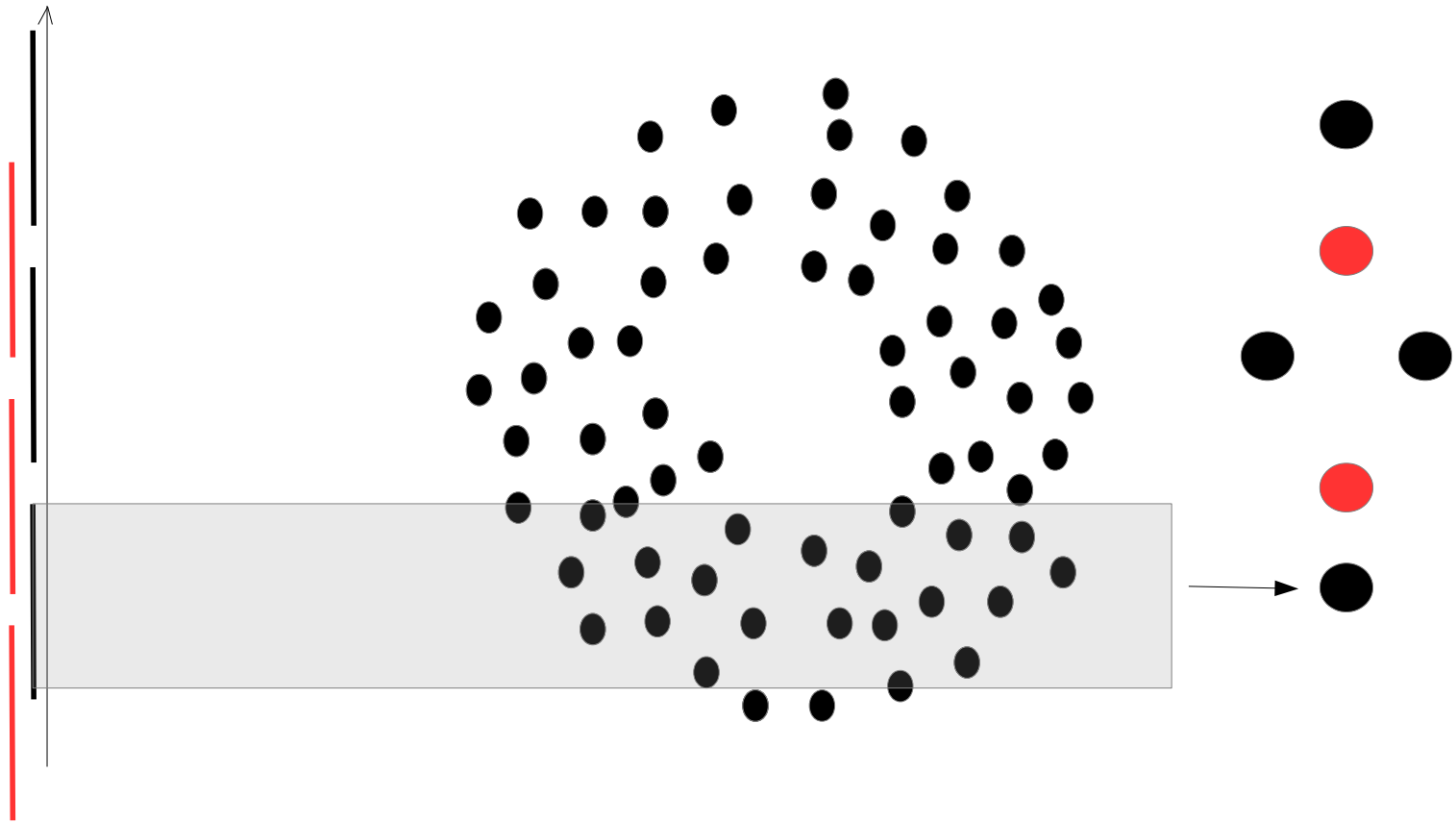
Conventional Mapper algorithm



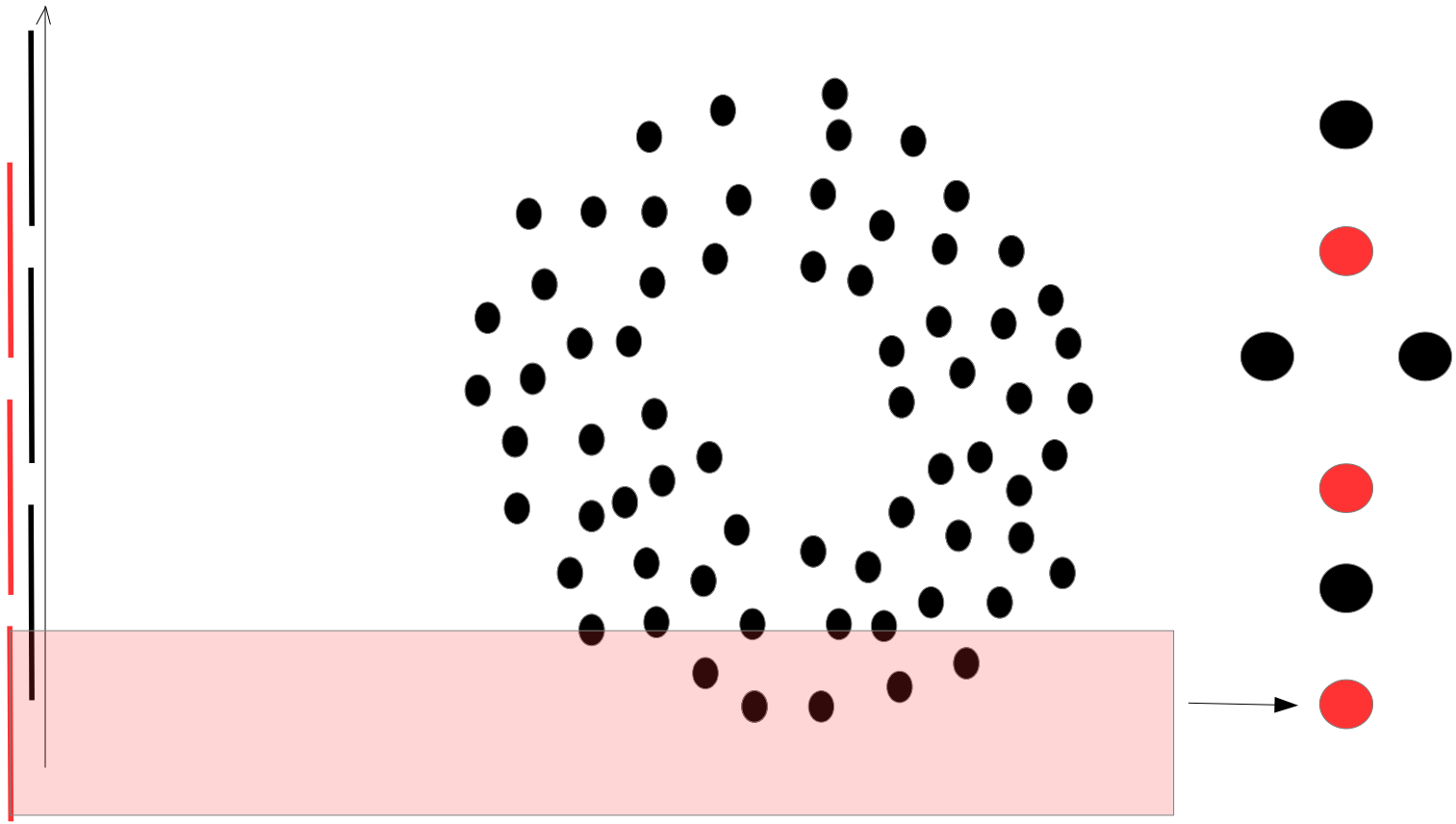
Conventional Mapper algorithm



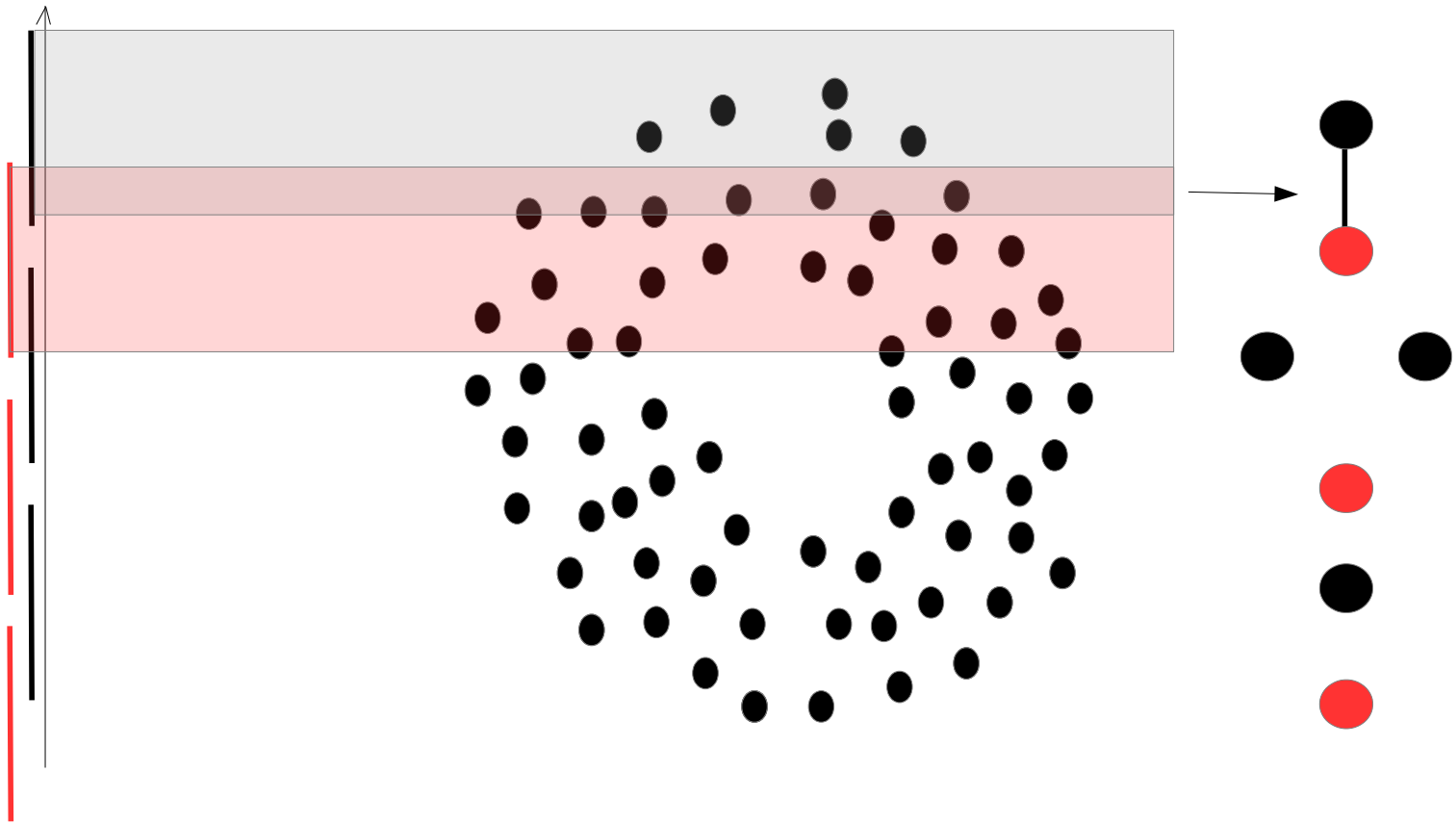
Conventional Mapper algorithm



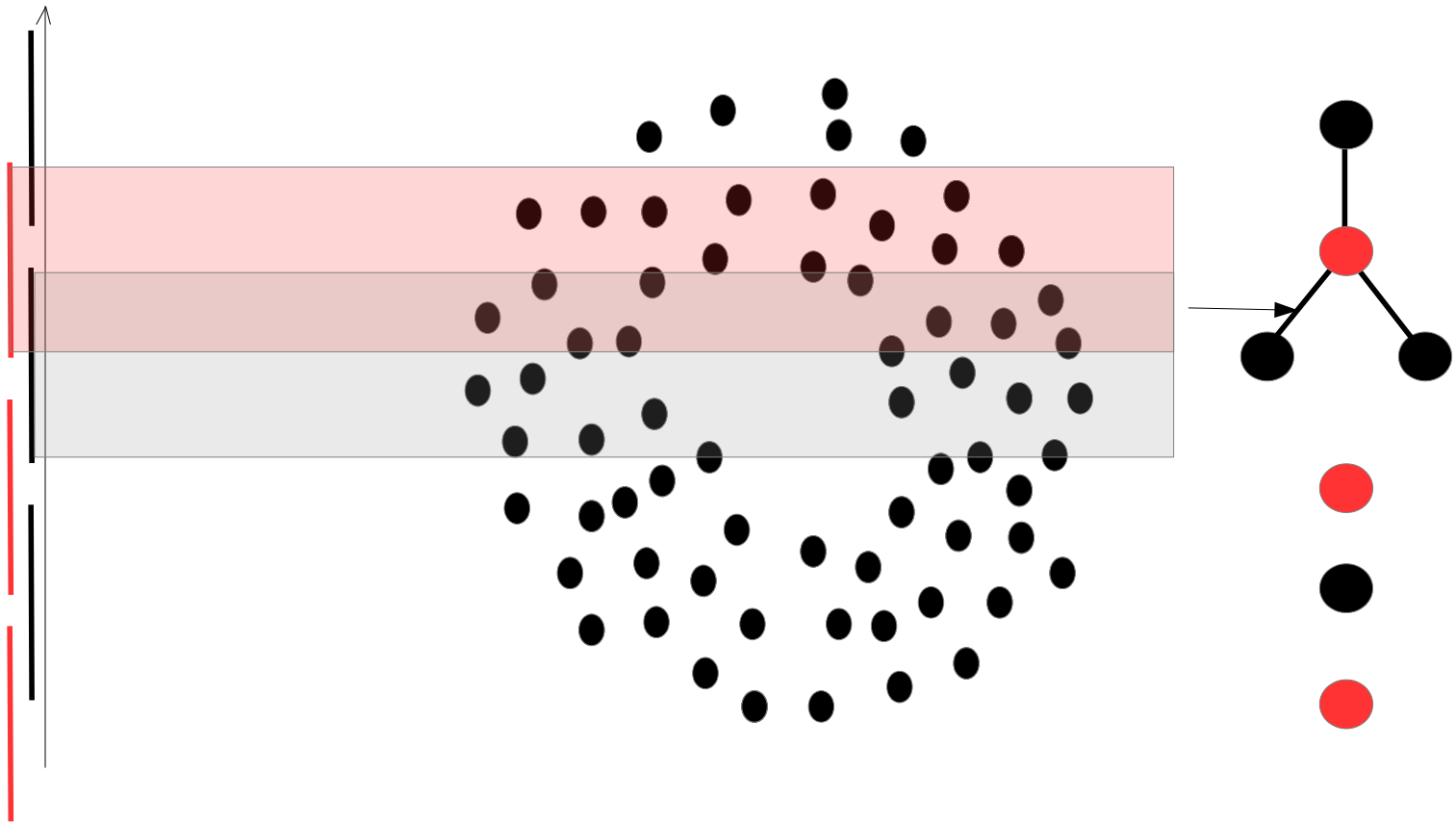
Conventional Mapper algorithm



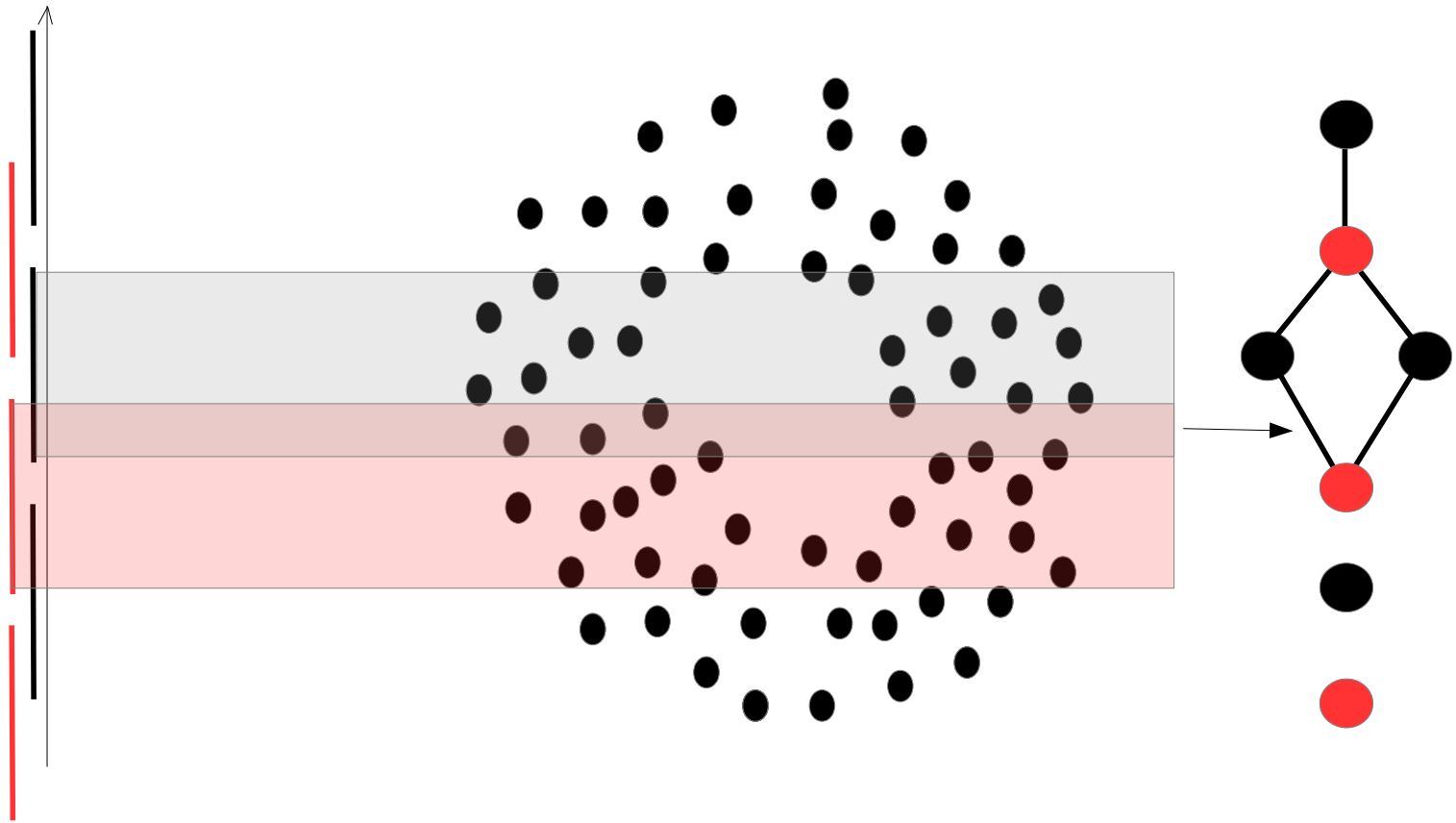
Conventional Mapper algorithm



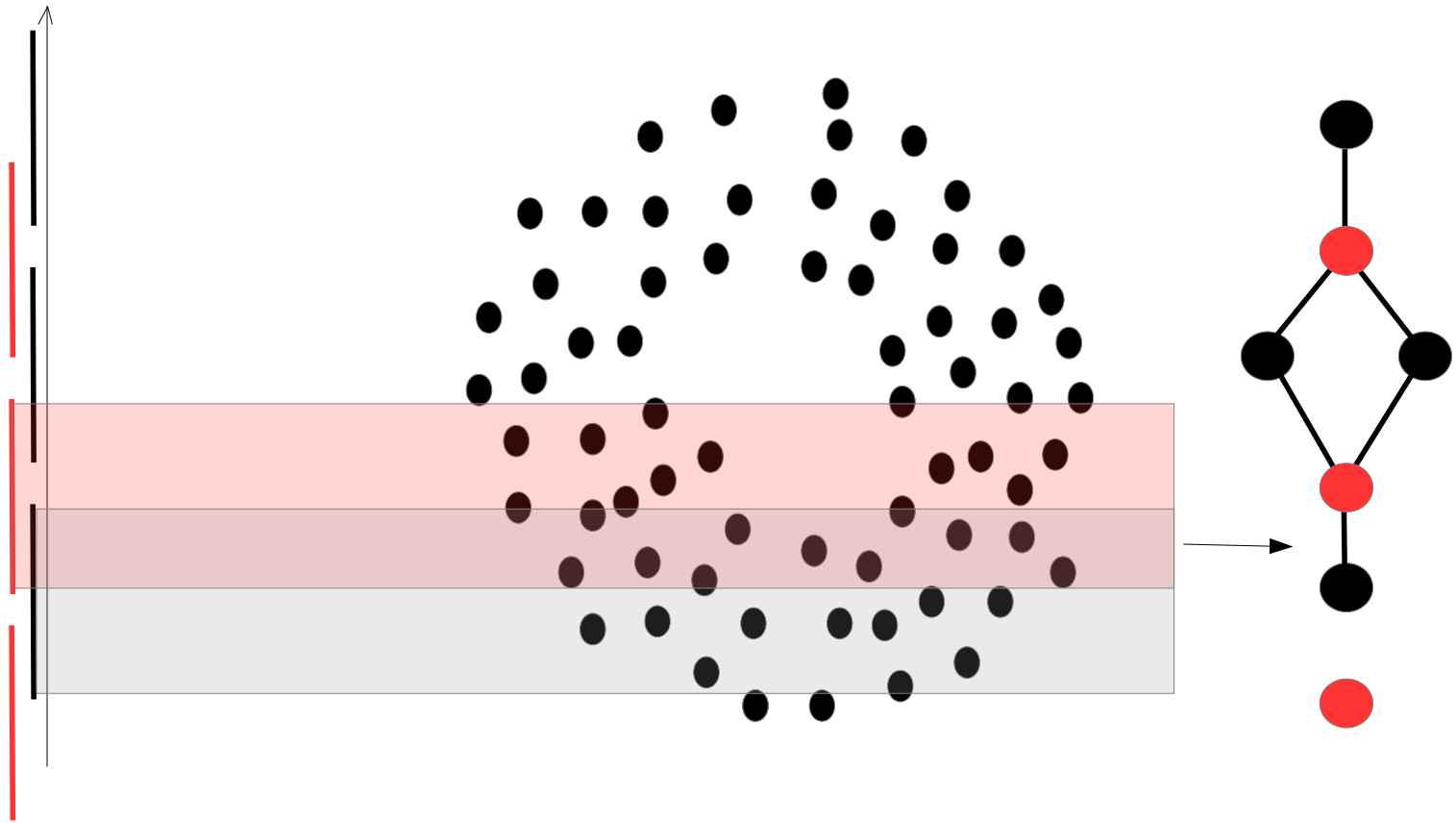
Conventional Mapper algorithm



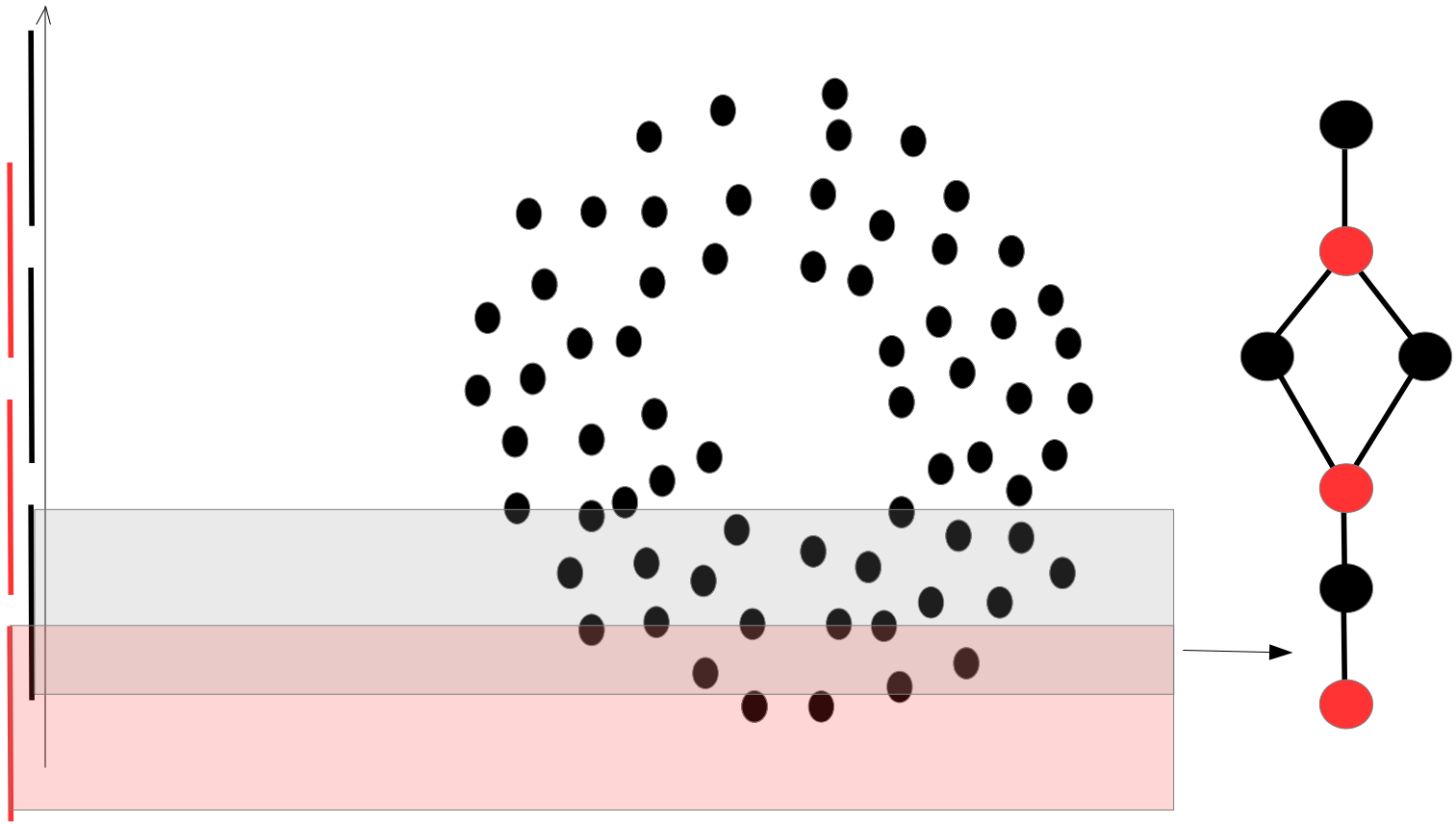
Conventional Mapper algorithm



Conventional Mapper algorithm



Conventional Mapper algorithm



Conventional Mapper algorithm, parameters

1. Cover of the line (number of cover elements, percentage of overlap).
2. **Lens function.**
3. Clustering algorithm.

Conventional Mapper algorithm

1. Mapper is used in hundreds of papers as mathematically rigorous visualization tool.
2. It is commercialized by Ayasdi Inc., a Menlo Park (CA) based company employing 150 people, founded, among others, by Gunnar Carlson.
3. The most recognized trademark and a working horse of TDA.
4. But, it also have a number of issues...

Instability of Mapper

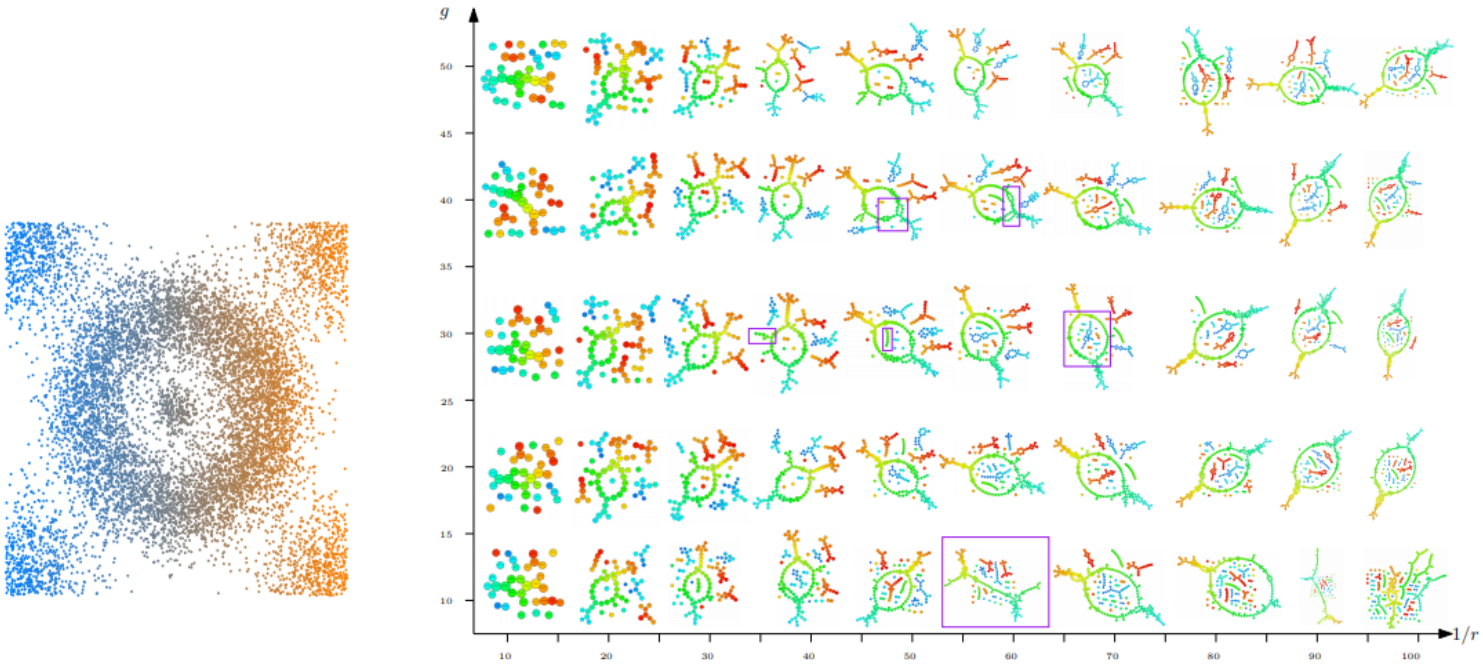
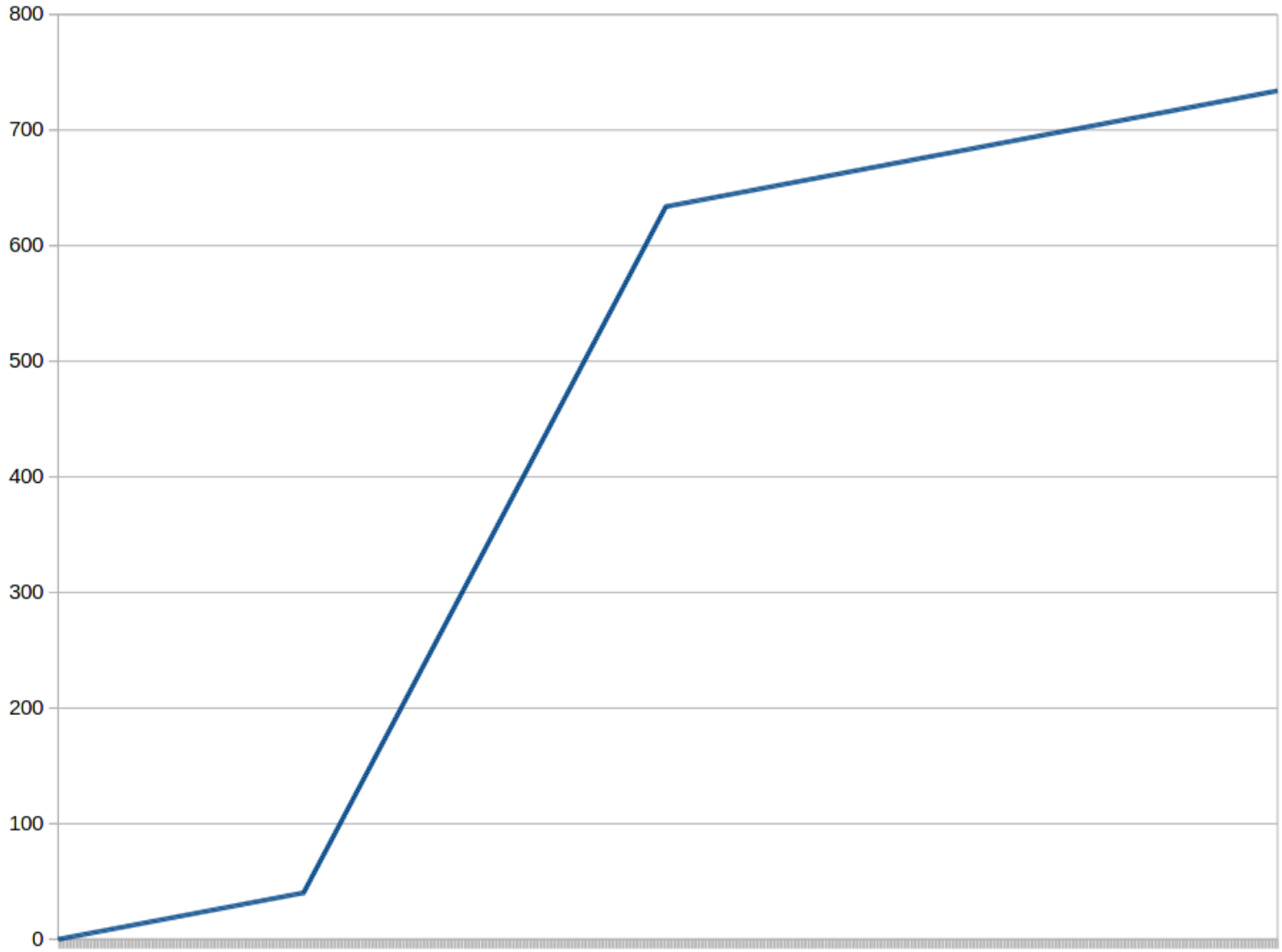
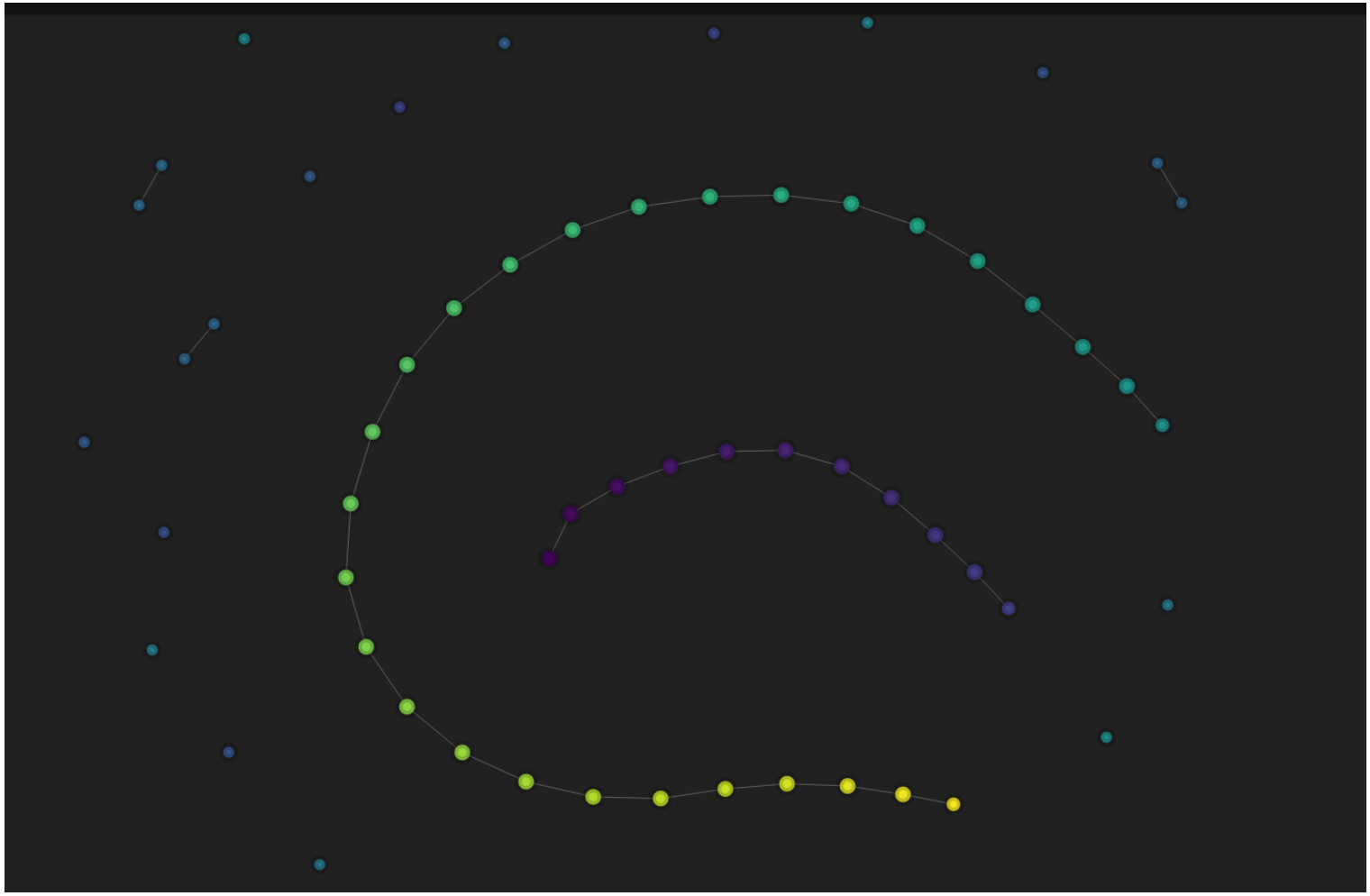


Figure 1, Statistical Analysis and Parameter Selection for Mapper by Carriere, Michel and Oudot.

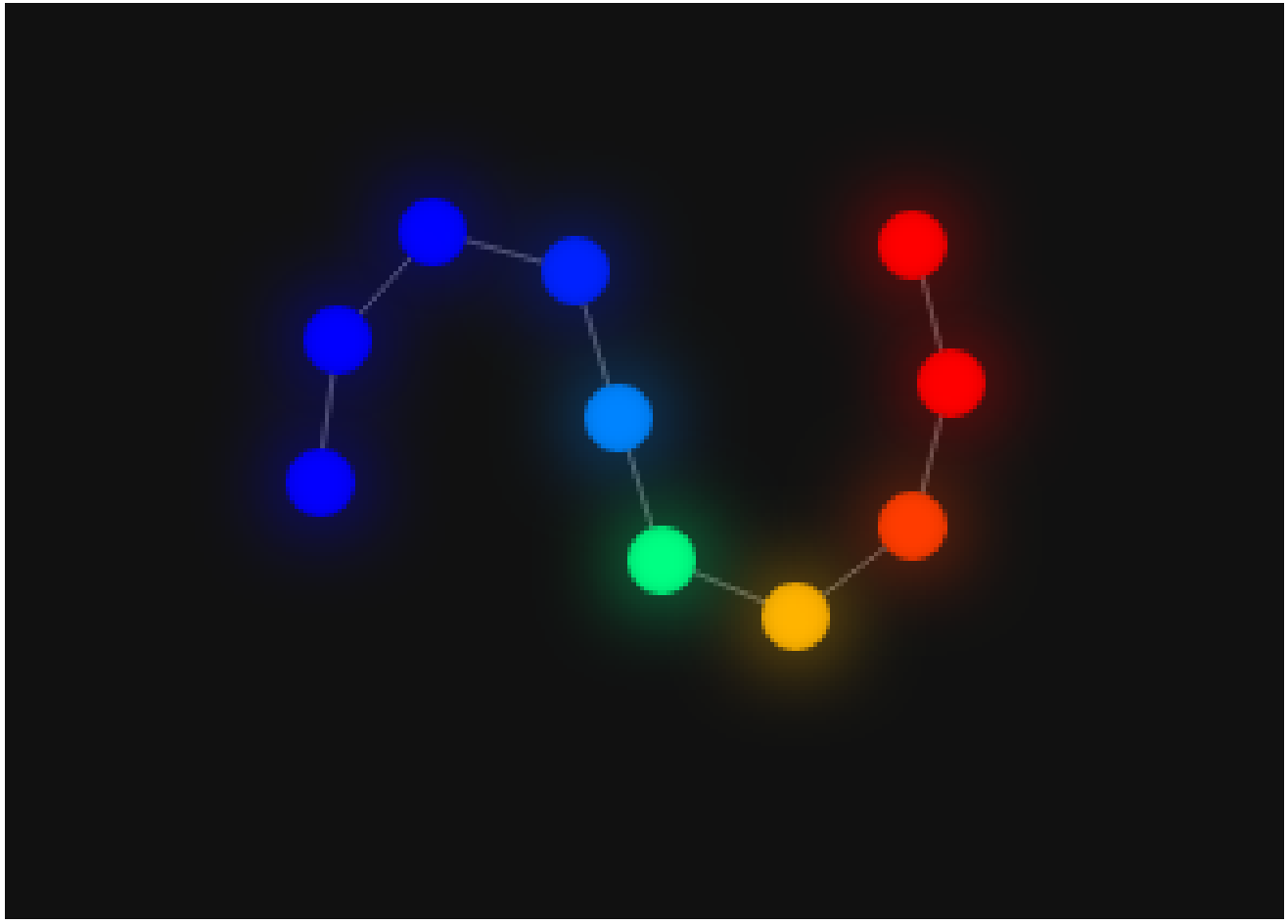
Do Mapper always capture the shape of the data?



Do Mapper always capture the shape of the data?



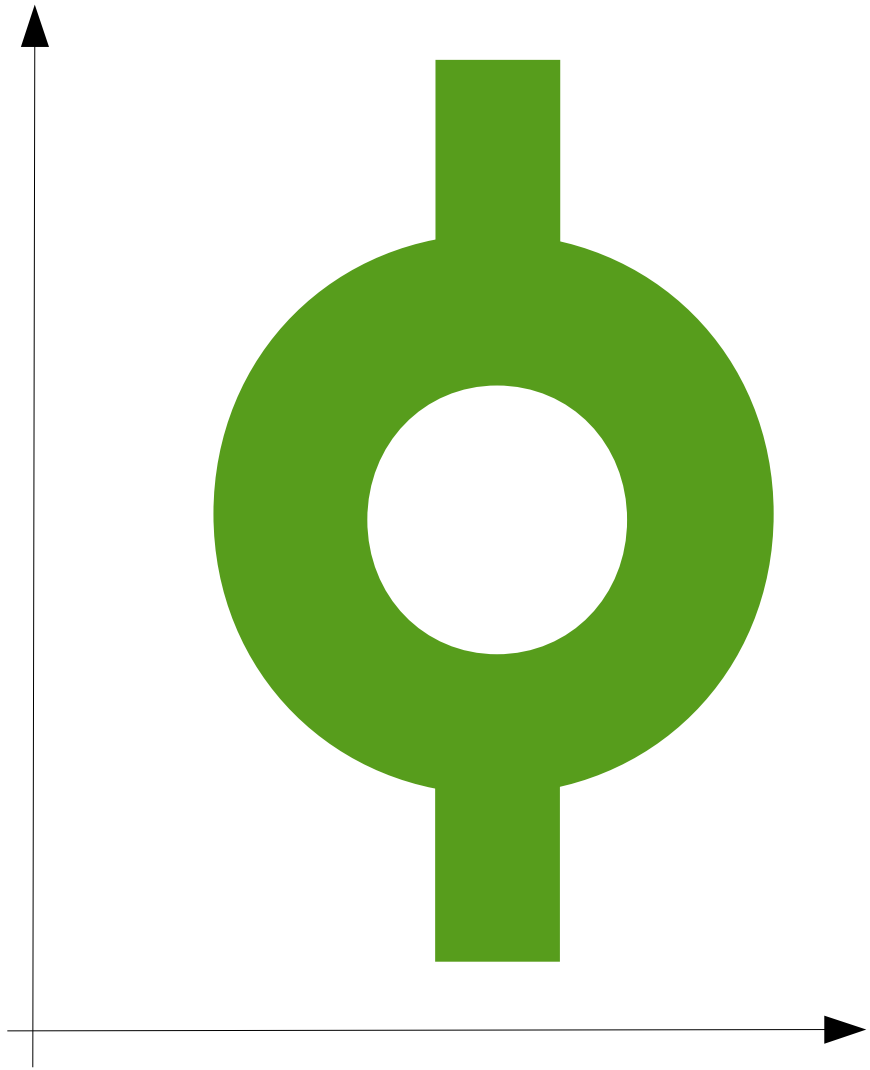
We would like to get...



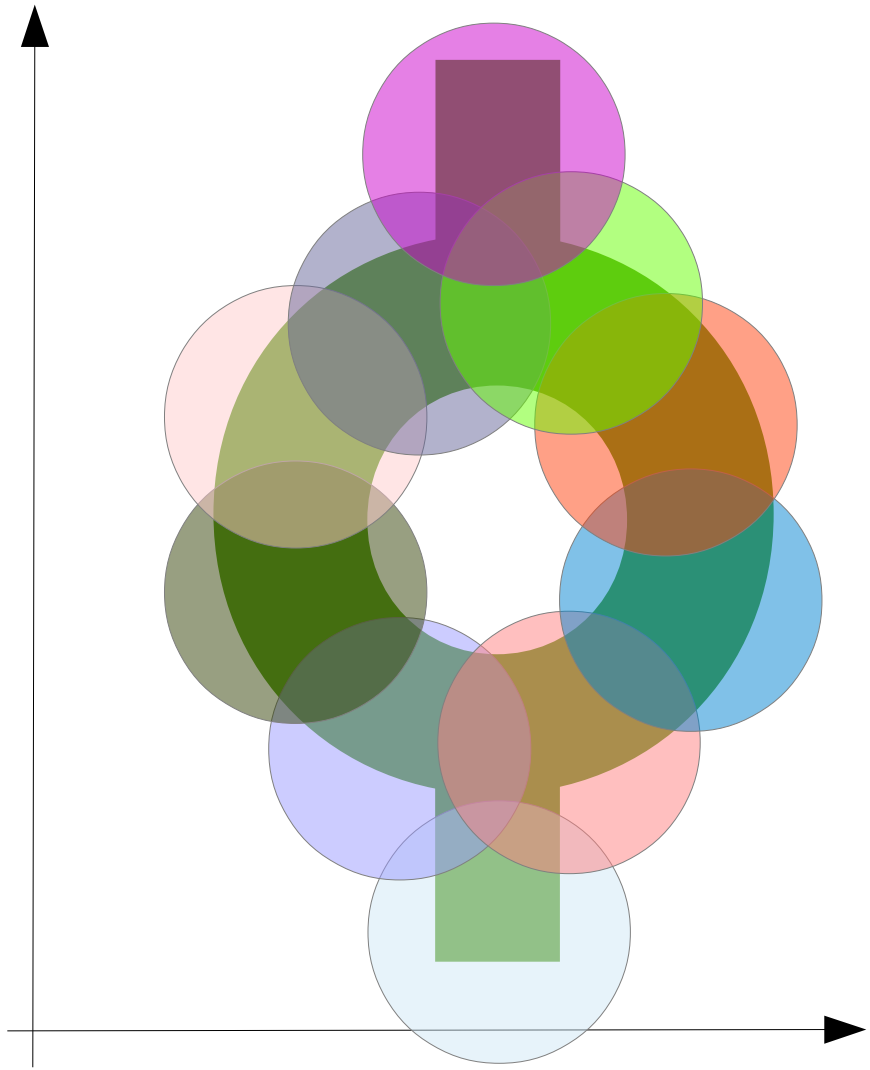
Ball Mapper

1. The main difficulty of building conventional Mapper is to obtain overlapping cover of M.
2. Once it is obtained, Mapper graph is extracted as a one dimensional nerve.
3. Ball Mapper gives a way of building such a overlapping cover in an alternative way.

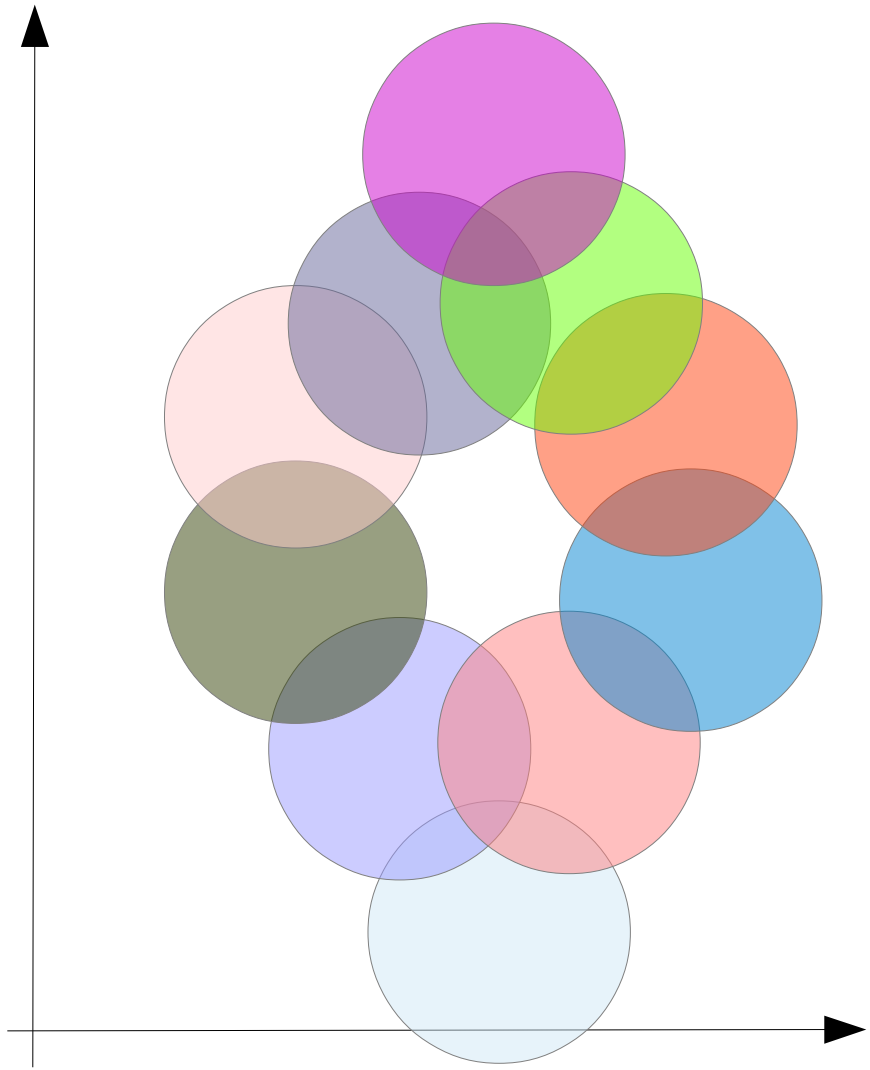
Ball Mapper



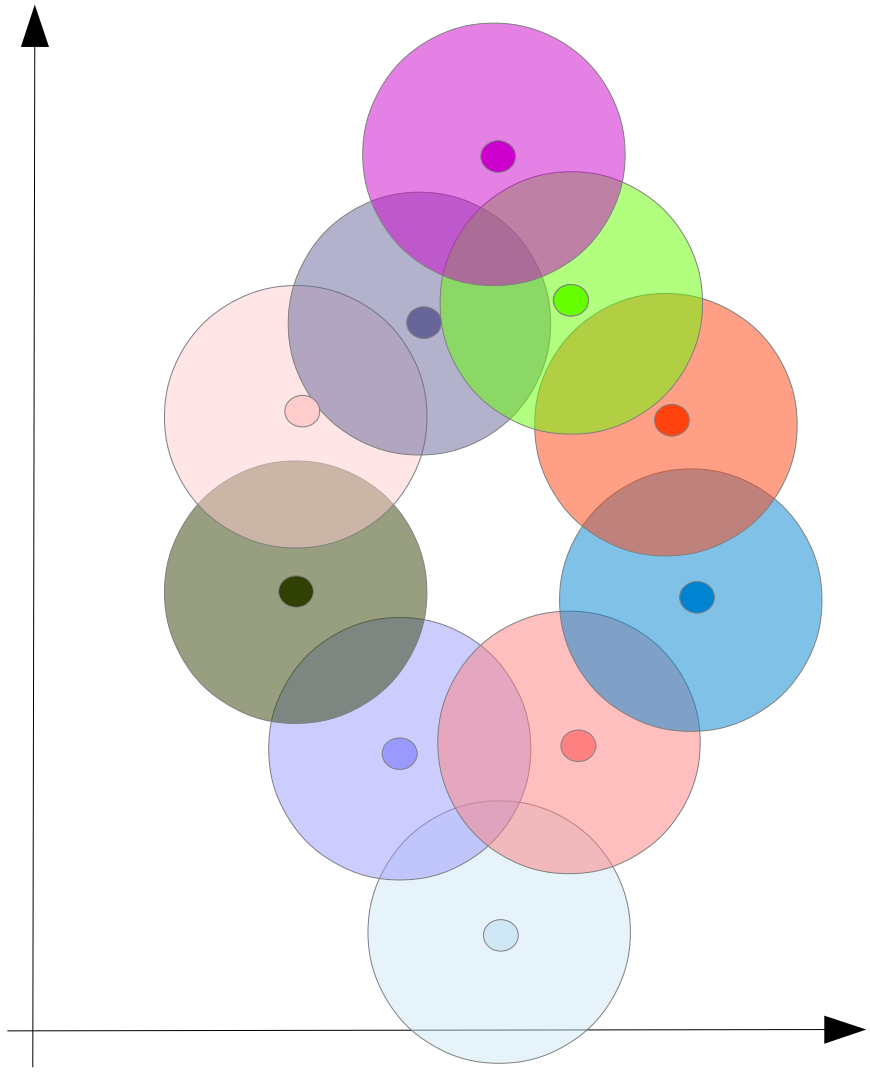
Cover the space of interest with balls



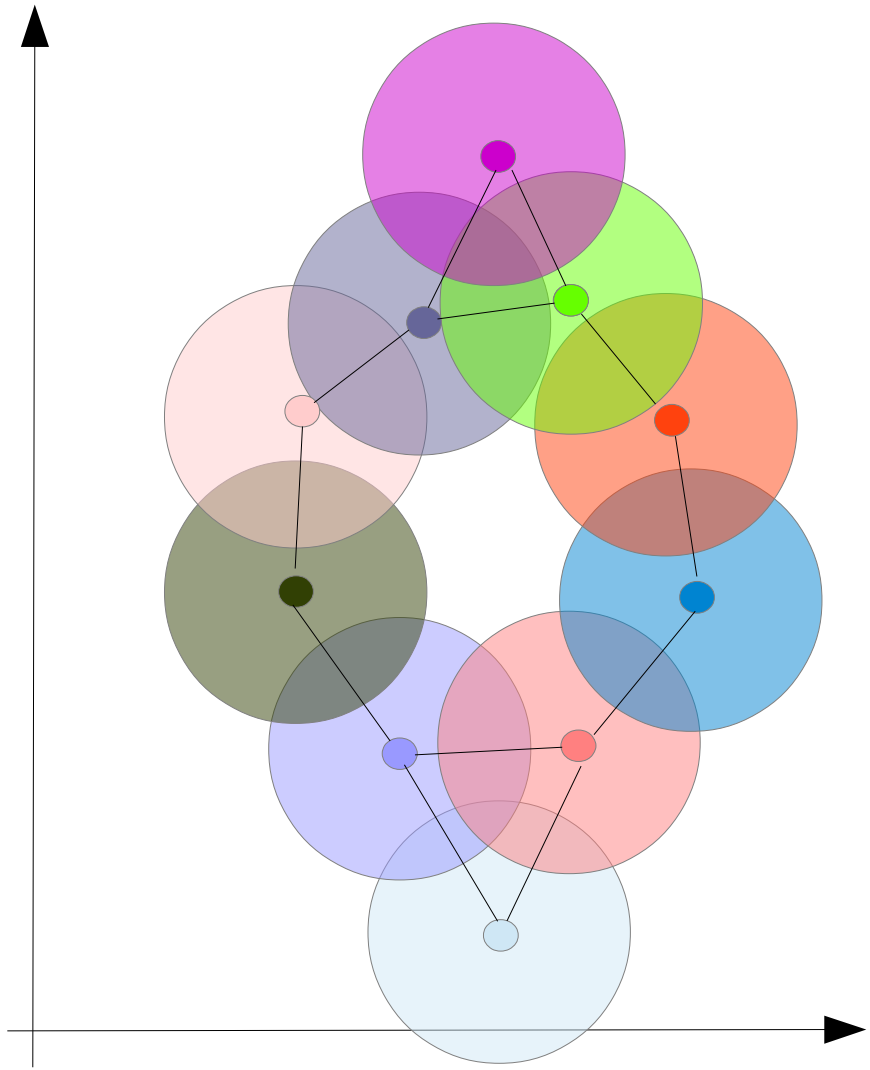
Restrict to the information from covering



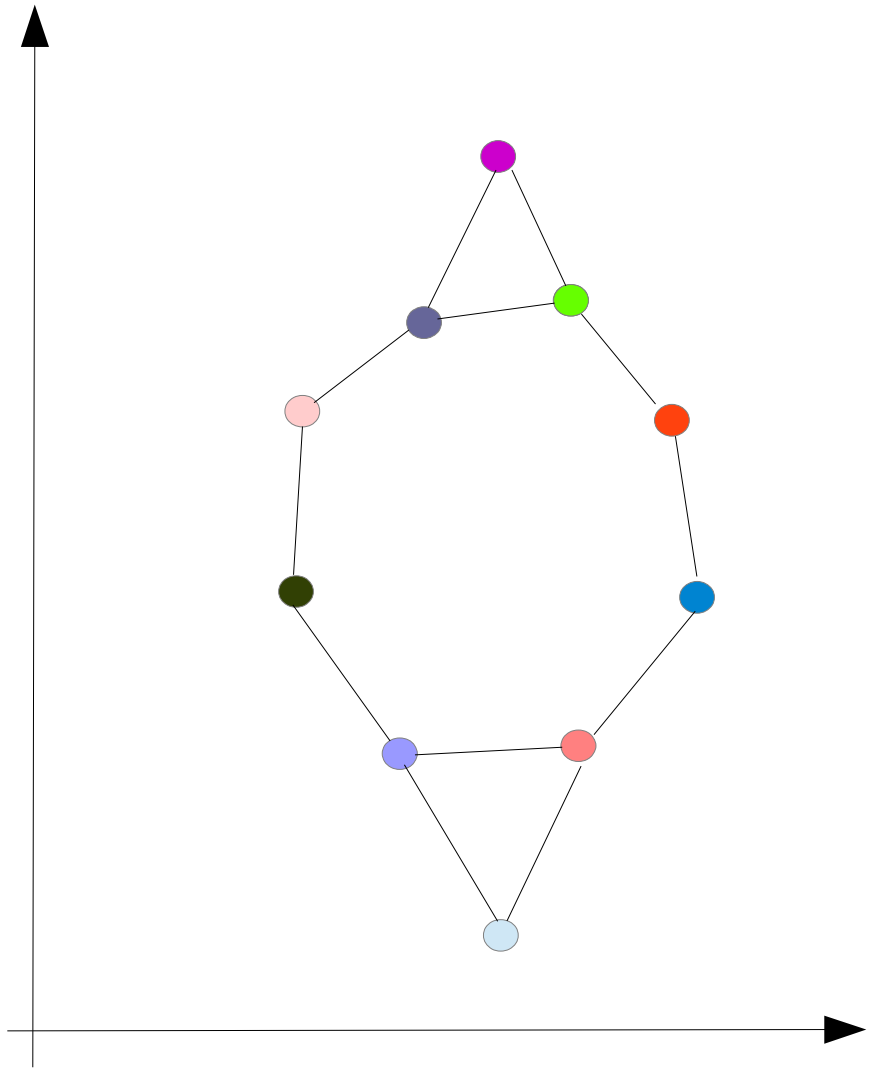
Centres correspond to vertices



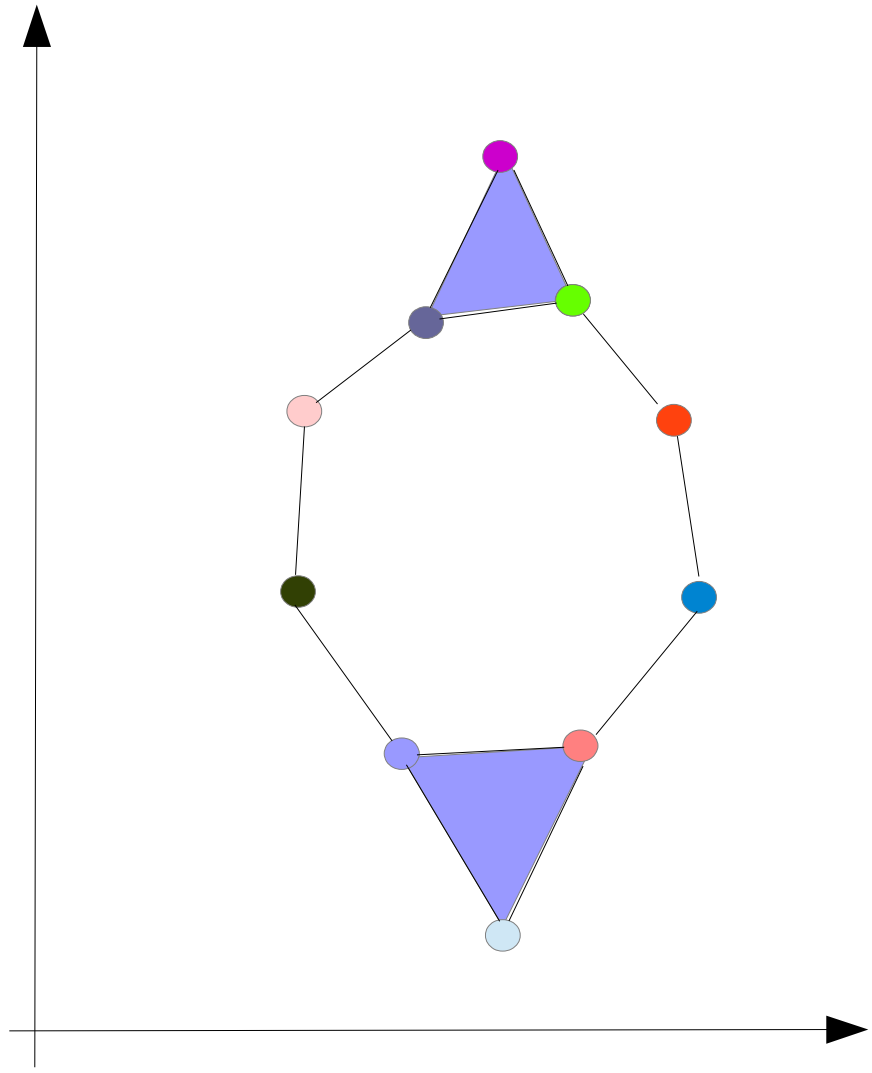
Intersections corresponds to edges



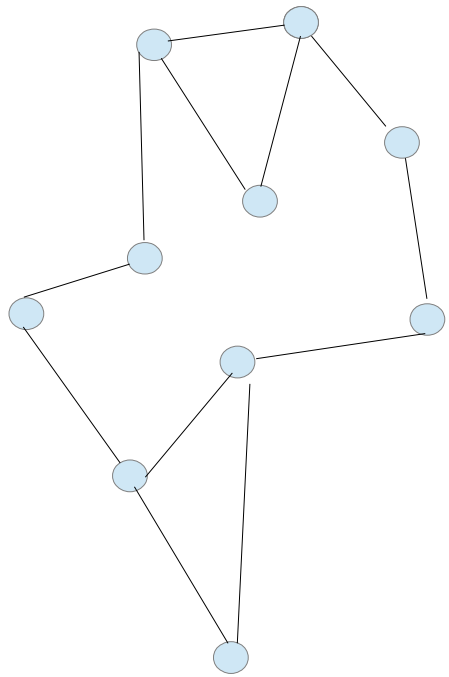
One dimensional nerve



Two dimensional nerve



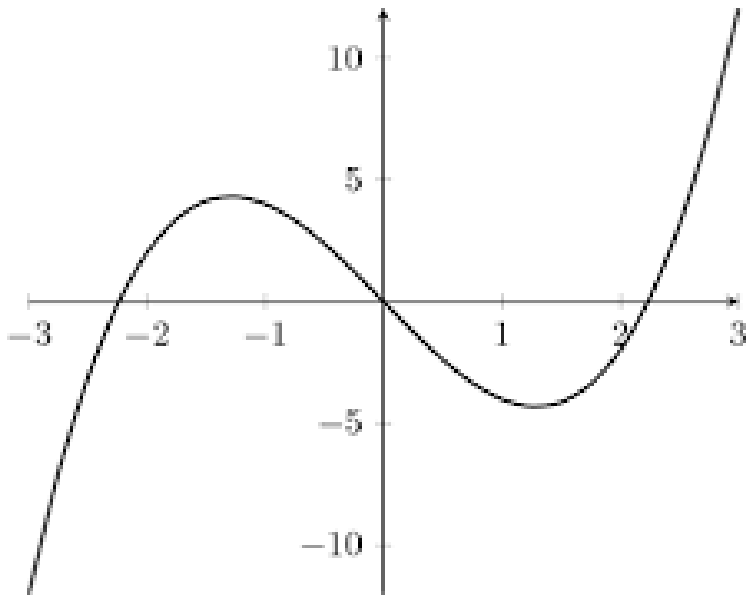
Ball Mapper graph



Ball Mapper graph

1. Centres of balls selected so that they form an ϵ net.
2. ϵ is the radius of balls.
3. The same strategy works for continuous and discrete spaces.
4. Only one parameter to set up.

How to plot functions on a representation of space?



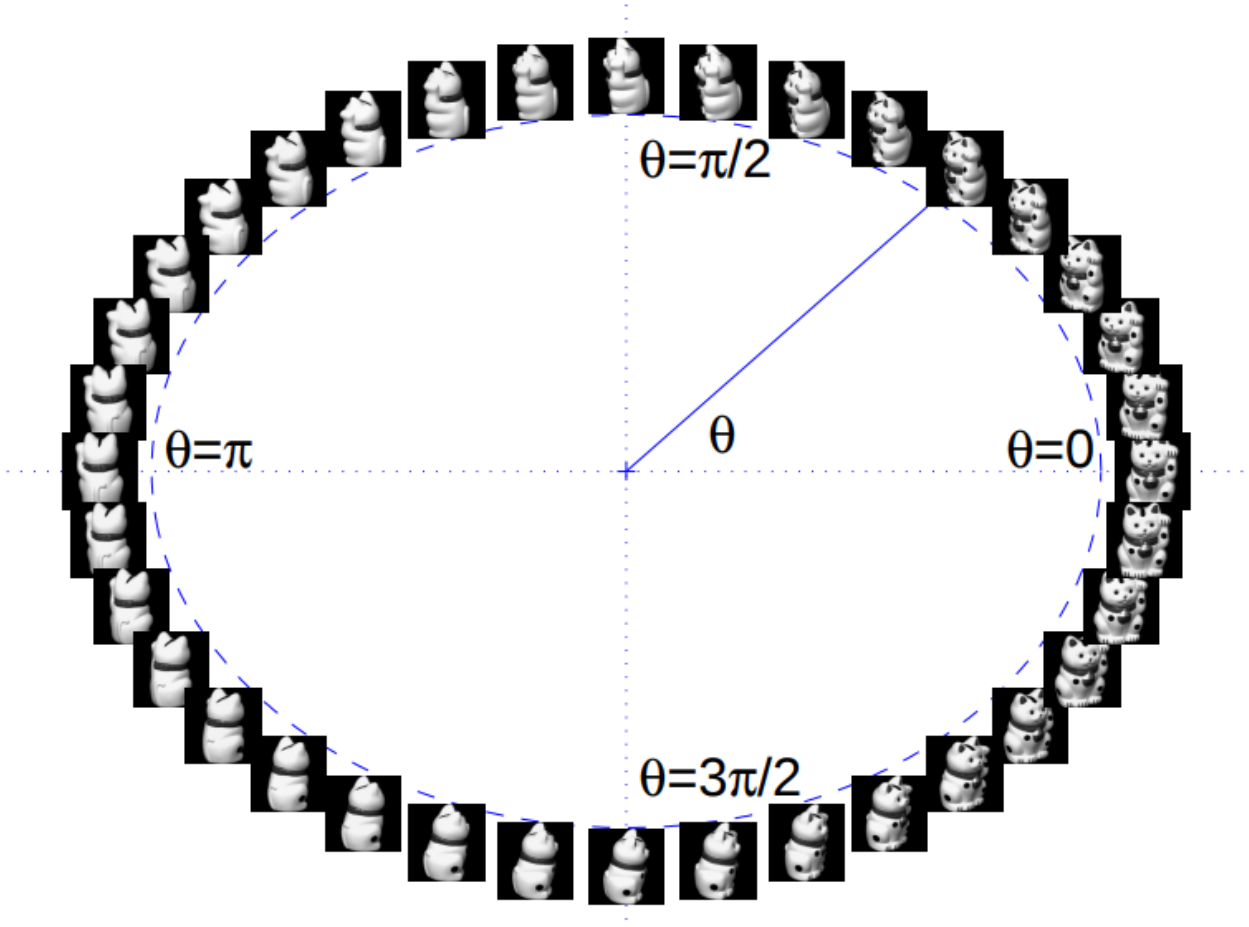
How to plot functions on a representation of space?

1. M – point cloud, $f : M \to \mathbb{R}$.
2. G – mapper or ball mapper graph.
3. Every $v \in G$ correspond to a cluster $C_v \subset M$.
4. We set $\tilde{f} : G \to \mathbb{R}$ so that $\tilde{f}(v) = avg(f(C_v))$.

Let us do some exercises. Let us see if we can see in high dimensions.

Meet the Lucky Cat

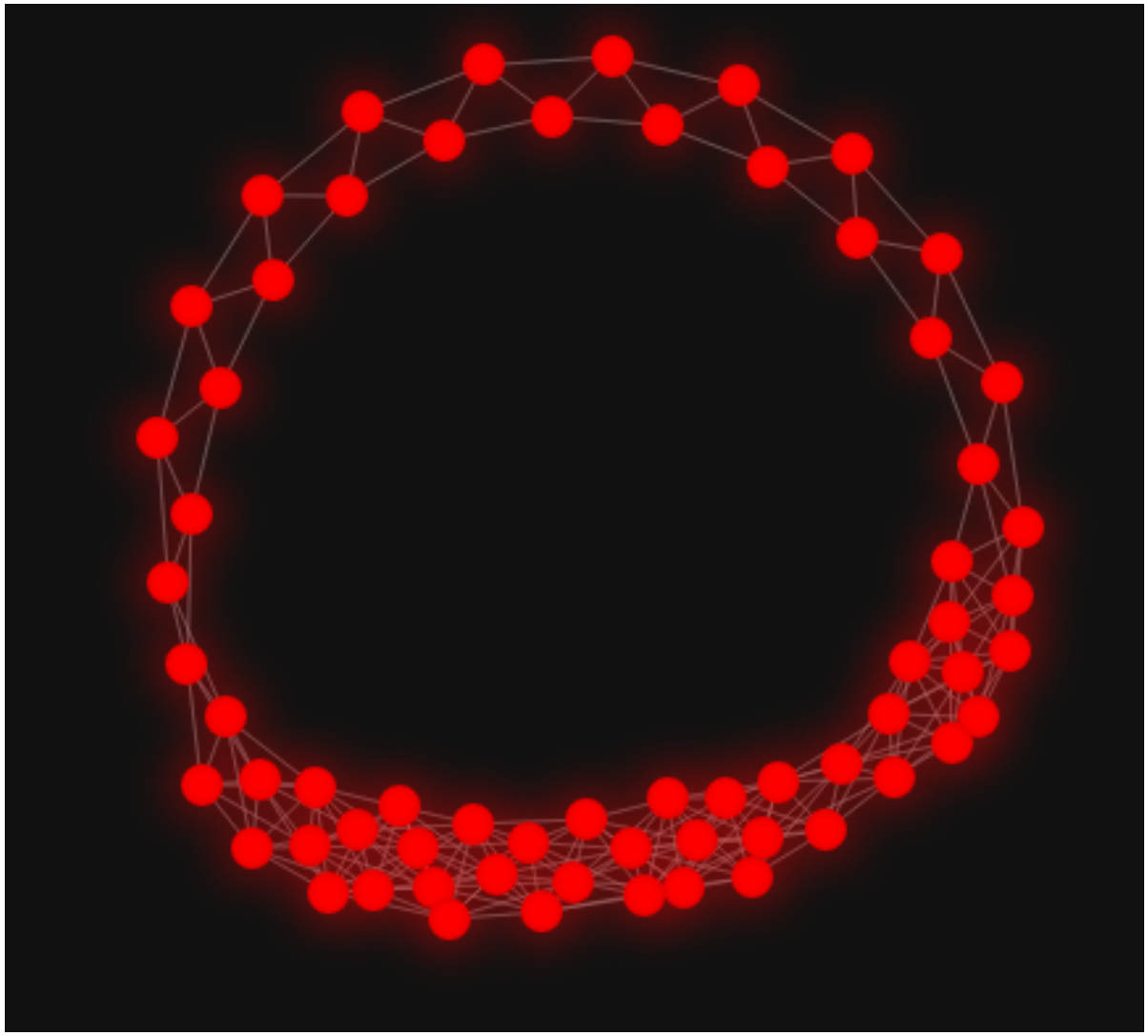
Take a picture of it from different angles



After flattening each 128×128 image correspond to point in 16384 dimensional space.

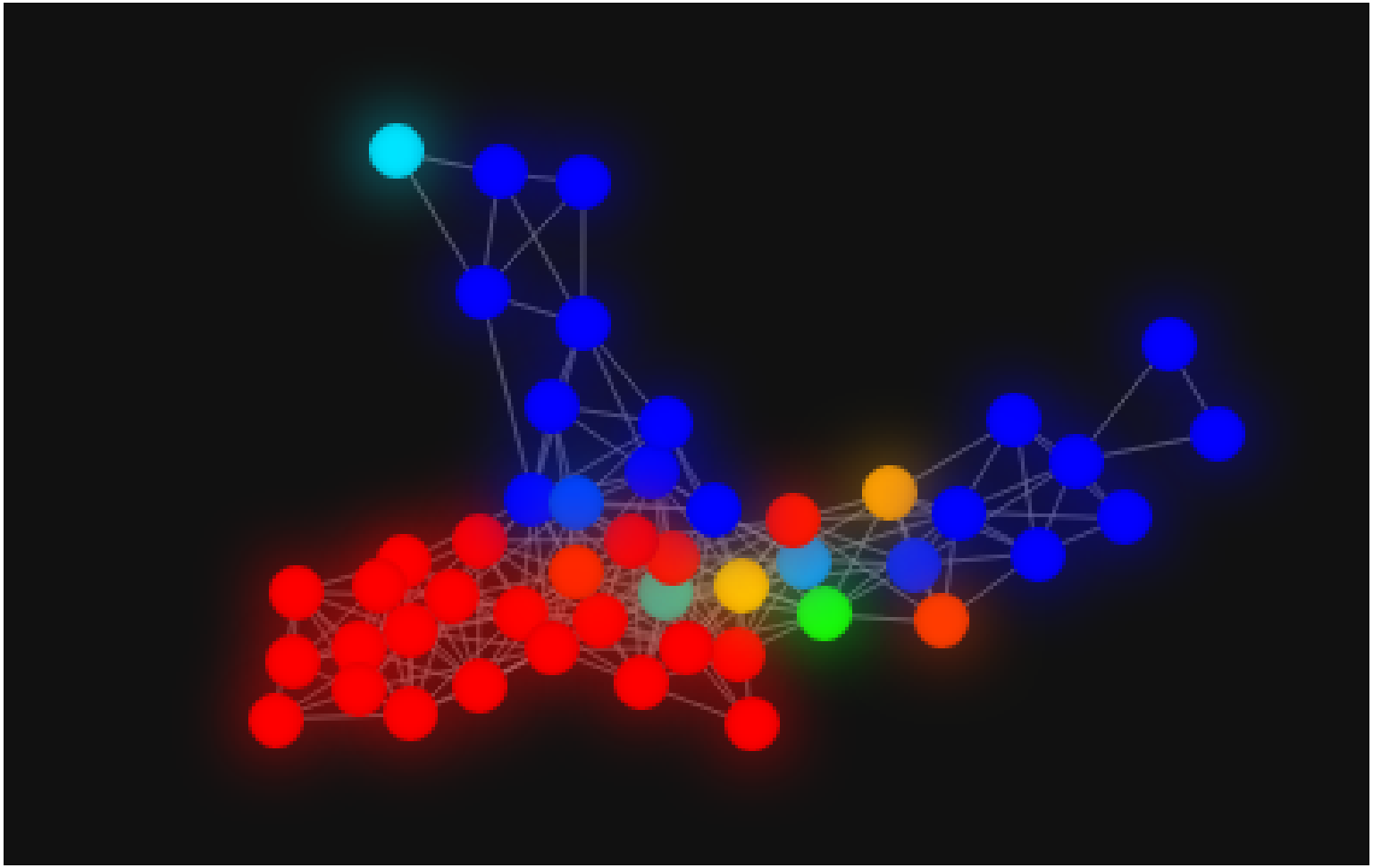
What is a shape of the obtained point cloud?

Network based landscapes of data



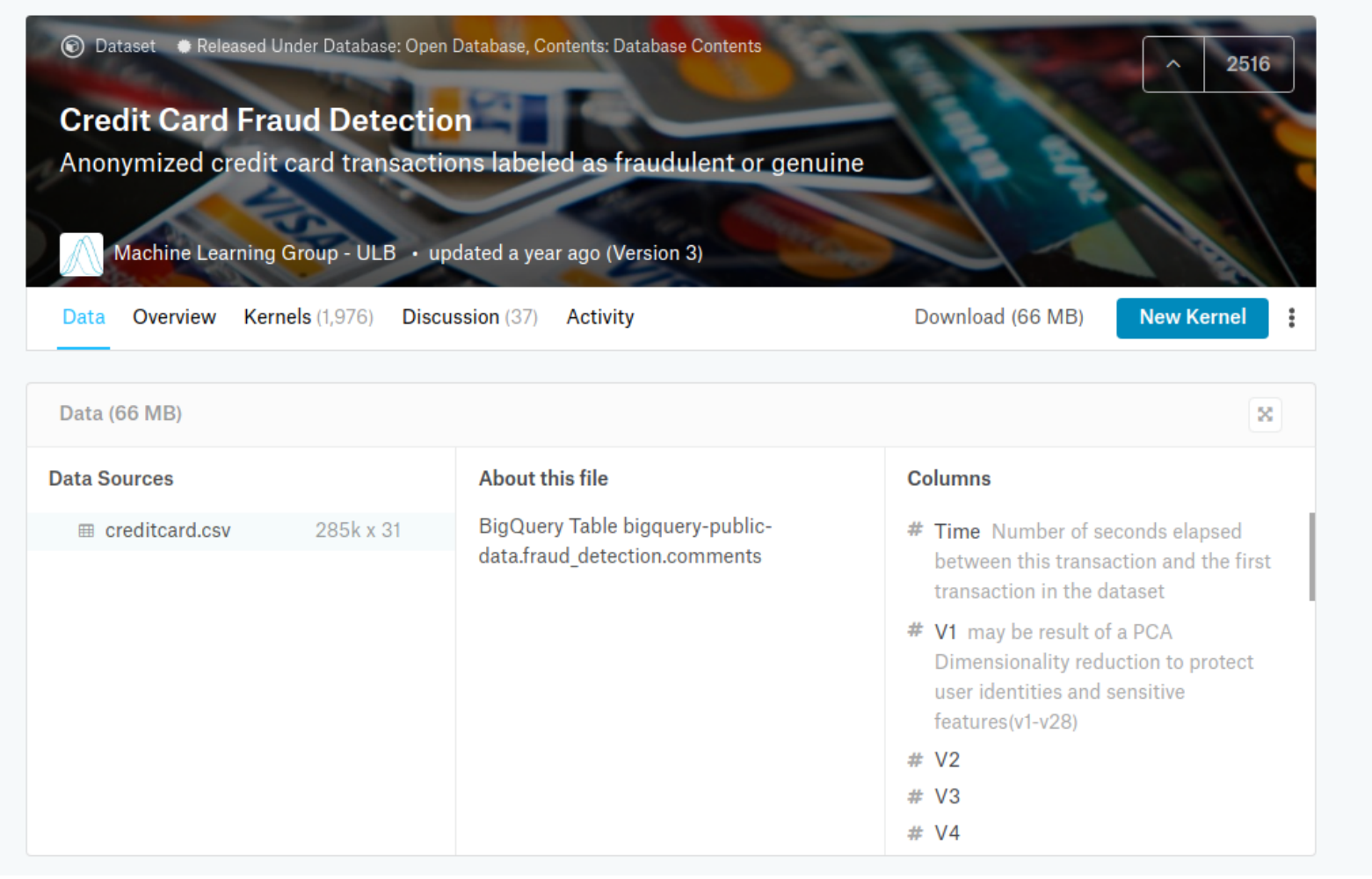
$128 \times 128 = 16384$ dimensional space

Banknote authentication data set



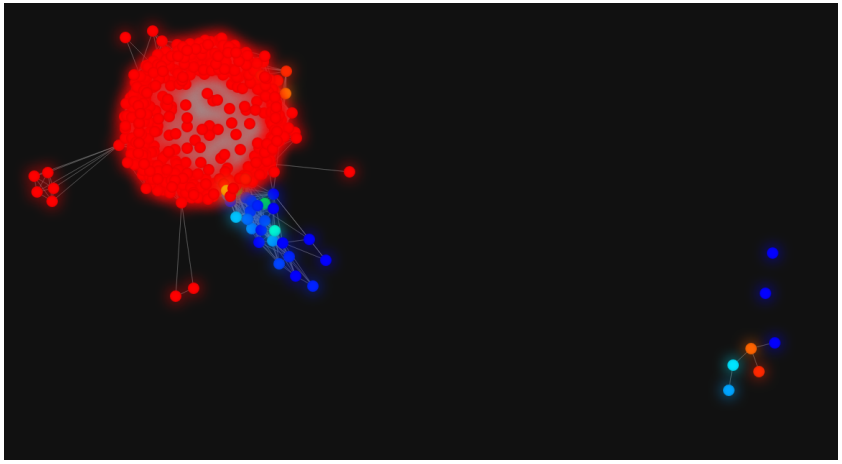
http://archive.ics.uci.edu/ml/datasets/banknote+authentication

Credit card fraud landscape

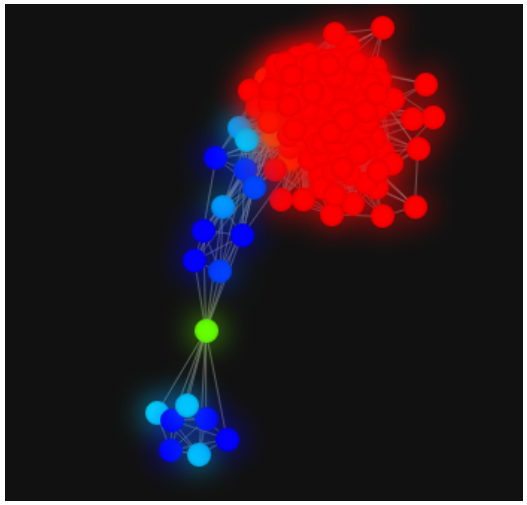


https://www.kaggle.com/mlg-ulb/creditcardfraud

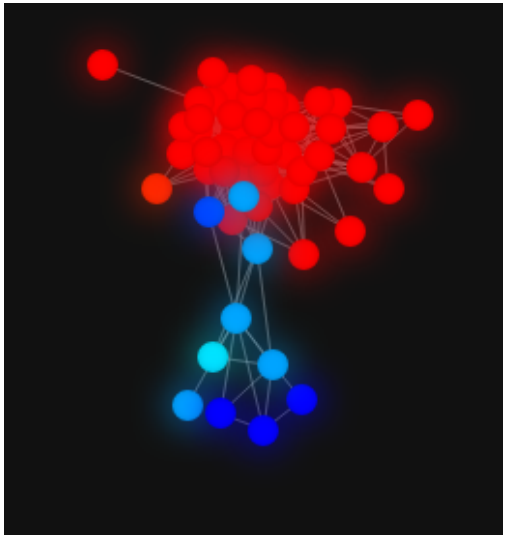
Credit card fraud landscape



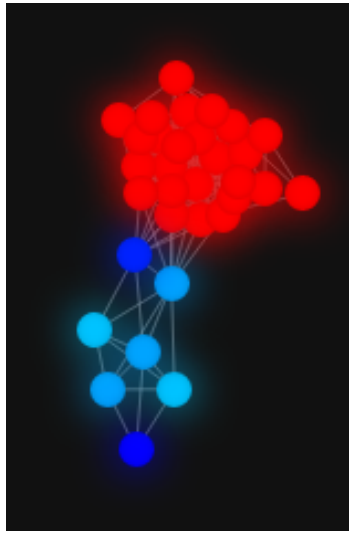
0.3



0.4



0.5



0.6

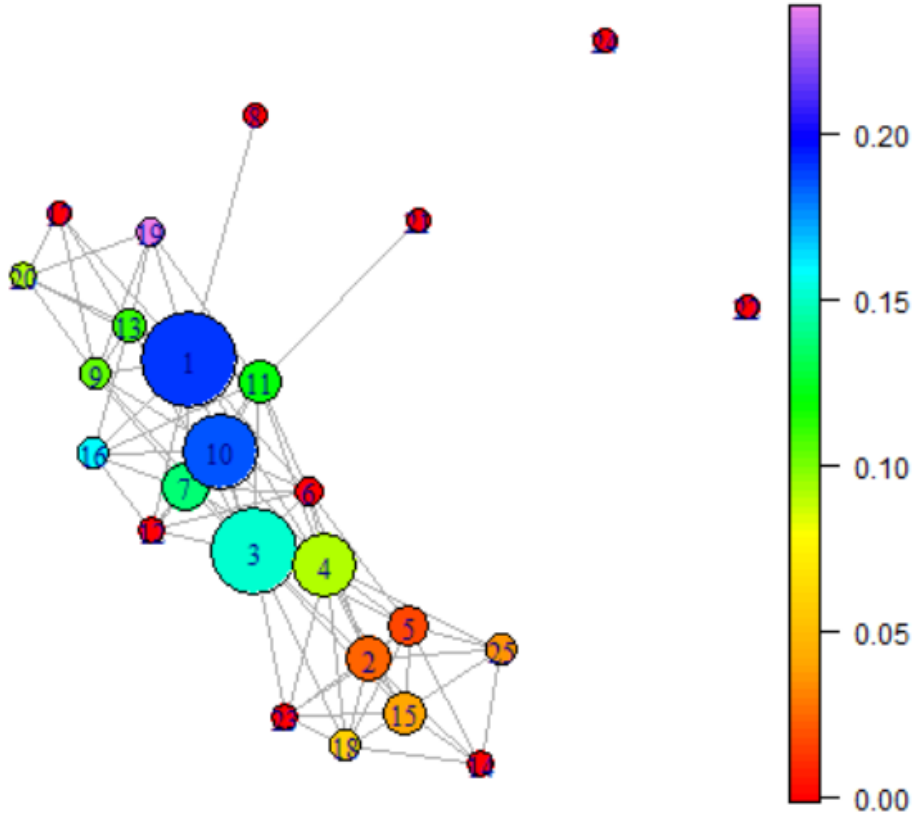
Einstein hospital dataset

1. Only Covid-19 related blood biomarkers dataset I was able to find.
2. Basic blood parameters of the patients (Age, Hematocrit, Hemoglobin,Platelets, Mean Platelet volume, Red blood Cells, Lymphocytes, MCHC, Leukocytes, Basophils, MCH, Eosinophils, MCV, Monocytes, RDW) - 15 parameters.
3. Circa 500 patients in total.
4. 81 Sars-Cov-2 positive.
5. 8 of then end up at ICU. No information about deaths.
6. Data normalized to have average zero and stdev 1.

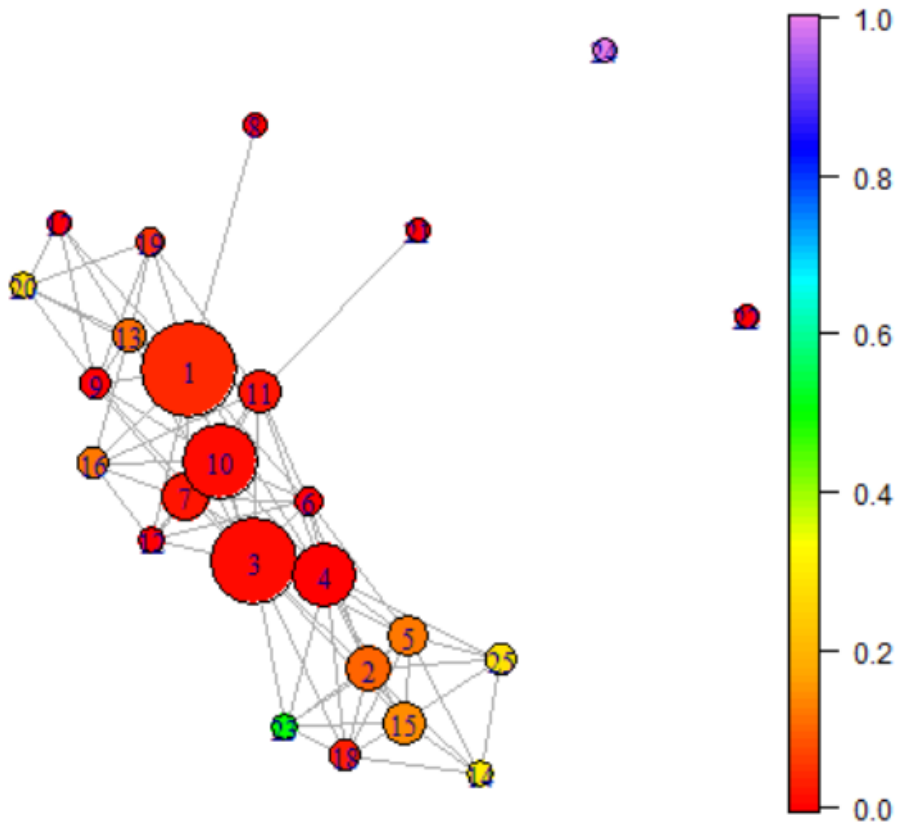
What we want to predict?

1. Are there parameters of the blood test characteristic for patients with Sars-Cov-2?
2. Among them, which blood characteristics makes patient likely to require ICU / of a high risk of death?

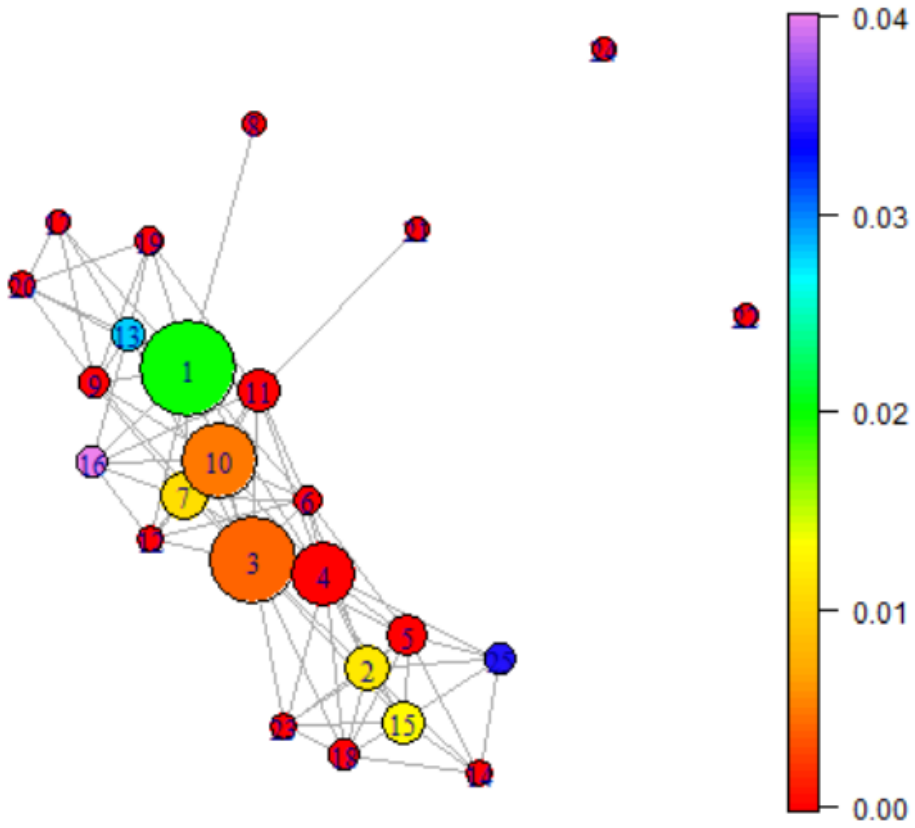
Sars-Cov-2 positive patients



All ICU instances



Covid-19 ICU instances



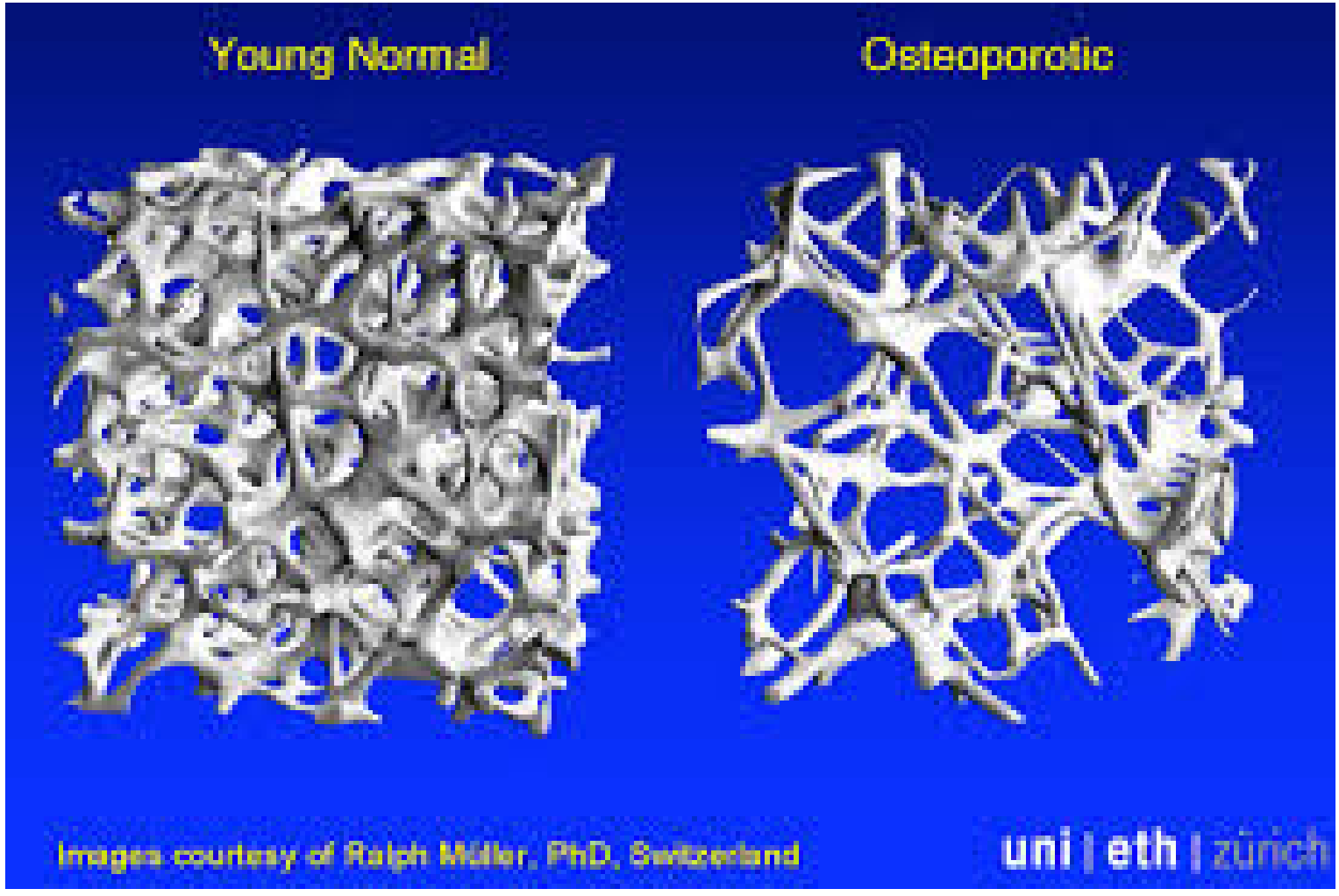
Not clinically valid, 8 patients only in the sample!!

Mapper graphs are not classifiers

1. They allow to show concentrations of some variables,
2. Motivate and allows clear explanation why simple classifier, like k-nearest neighbors, can be used in clinical practice.
3. The analysis presented here is a **proof of concept**, require more data to turn it to efficient tool.
4. More data may come but they are not easy to get.

Spongy bone and persistent homology.

Osteoporosis vs Bone Structure.



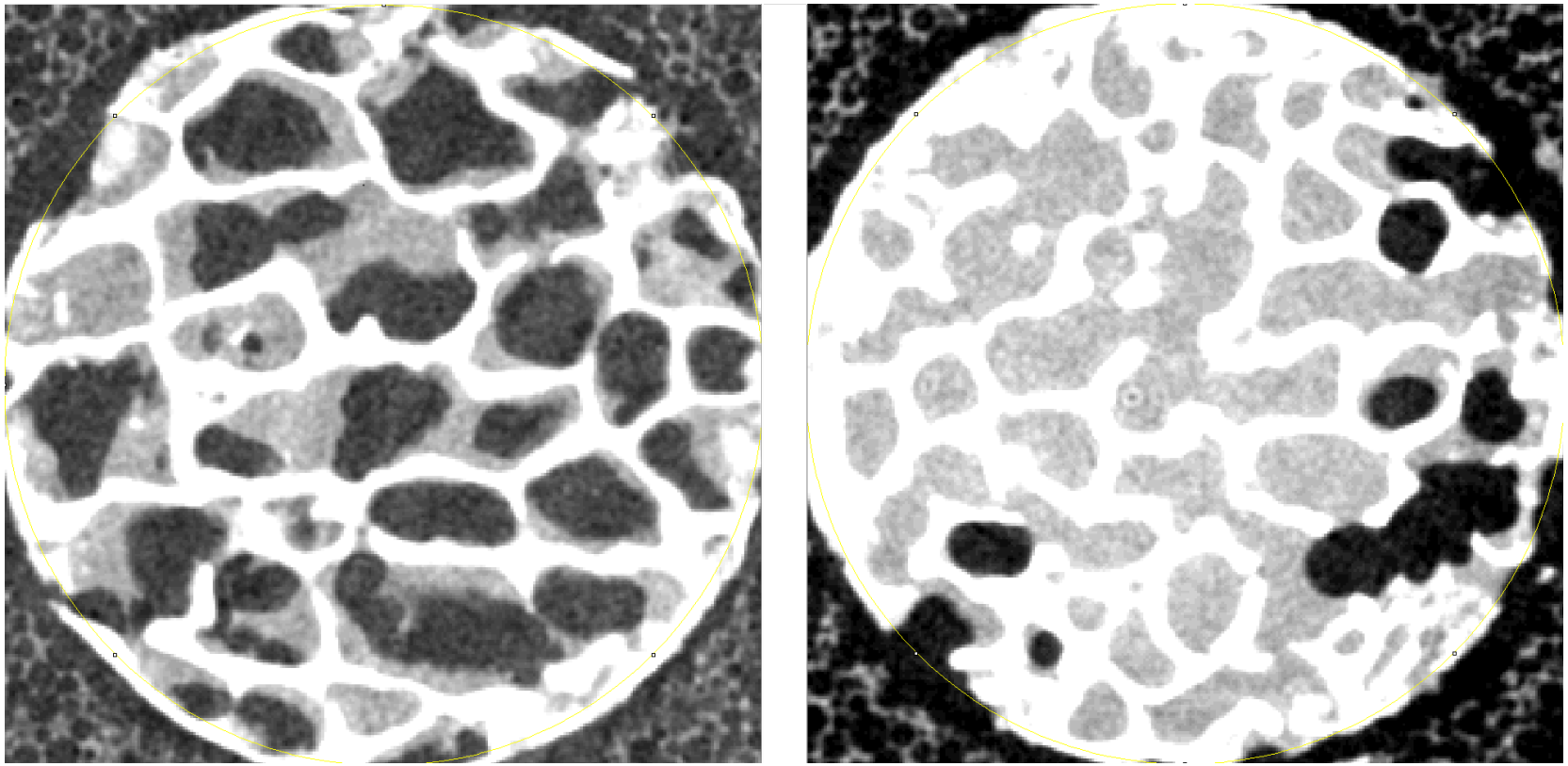
Osteoporosis vs Bone Structure.

1. Density (DXA) is a standard measure.
2. Instances of patients with low density and resilient bone are known.
3. Structure seems to play an important role.
4. 3 dimensional images can be obtained from CT / MRI scans.
5. Initial study based on micro-CT high resolution scans from Richard Abel (Imperial).

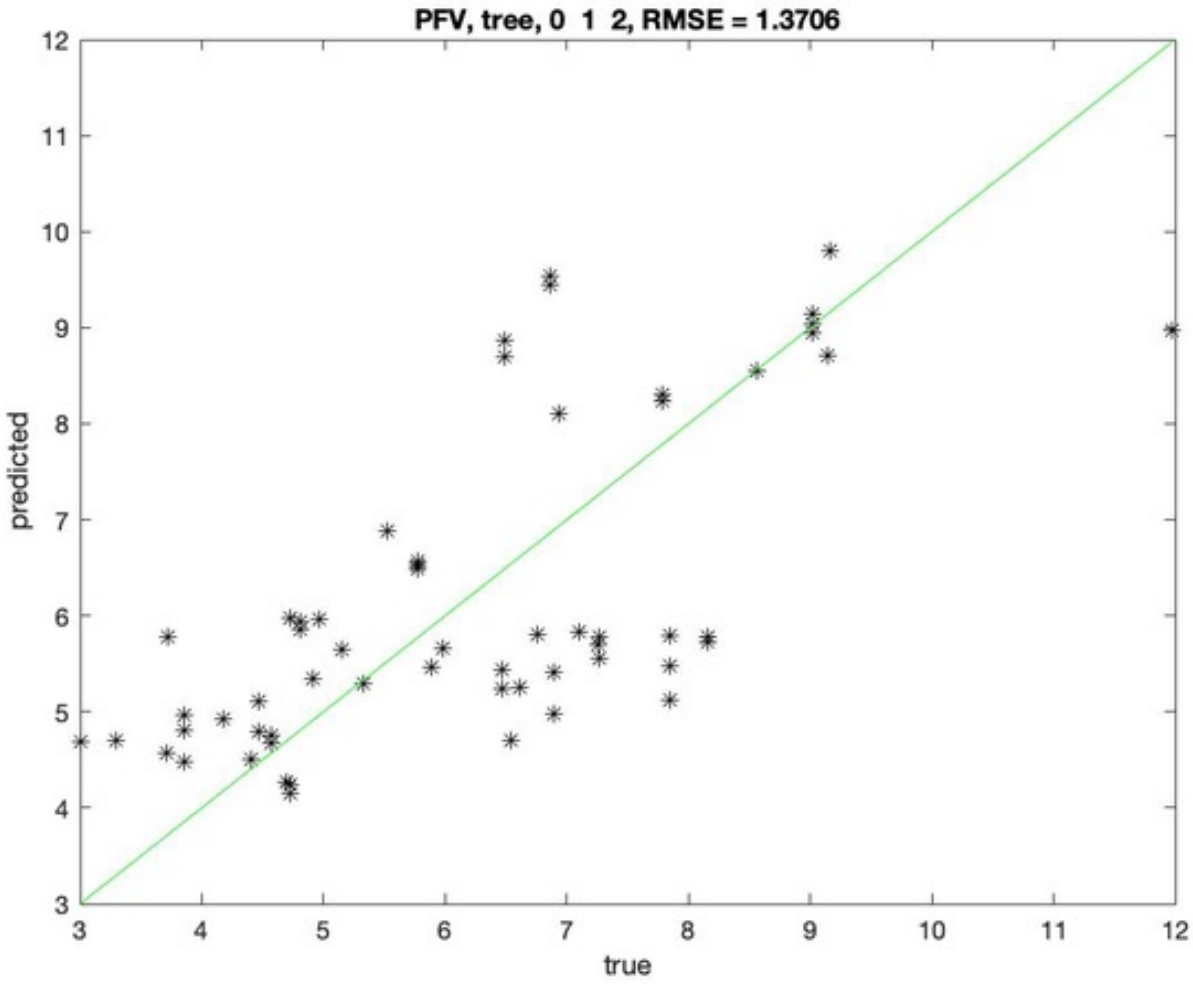
Richard's data.

1. Post mortem vertebrae from humans.
2. Micro-CT scan done.
3. Vertebrae is places in a vise and a force required to crash it is recorded.
4. Persistent homology of the image is computer, bag of words representation is extracted.
5. Does the information about the structure correlate with the force required to crash the vertebrae?

Richard's data.

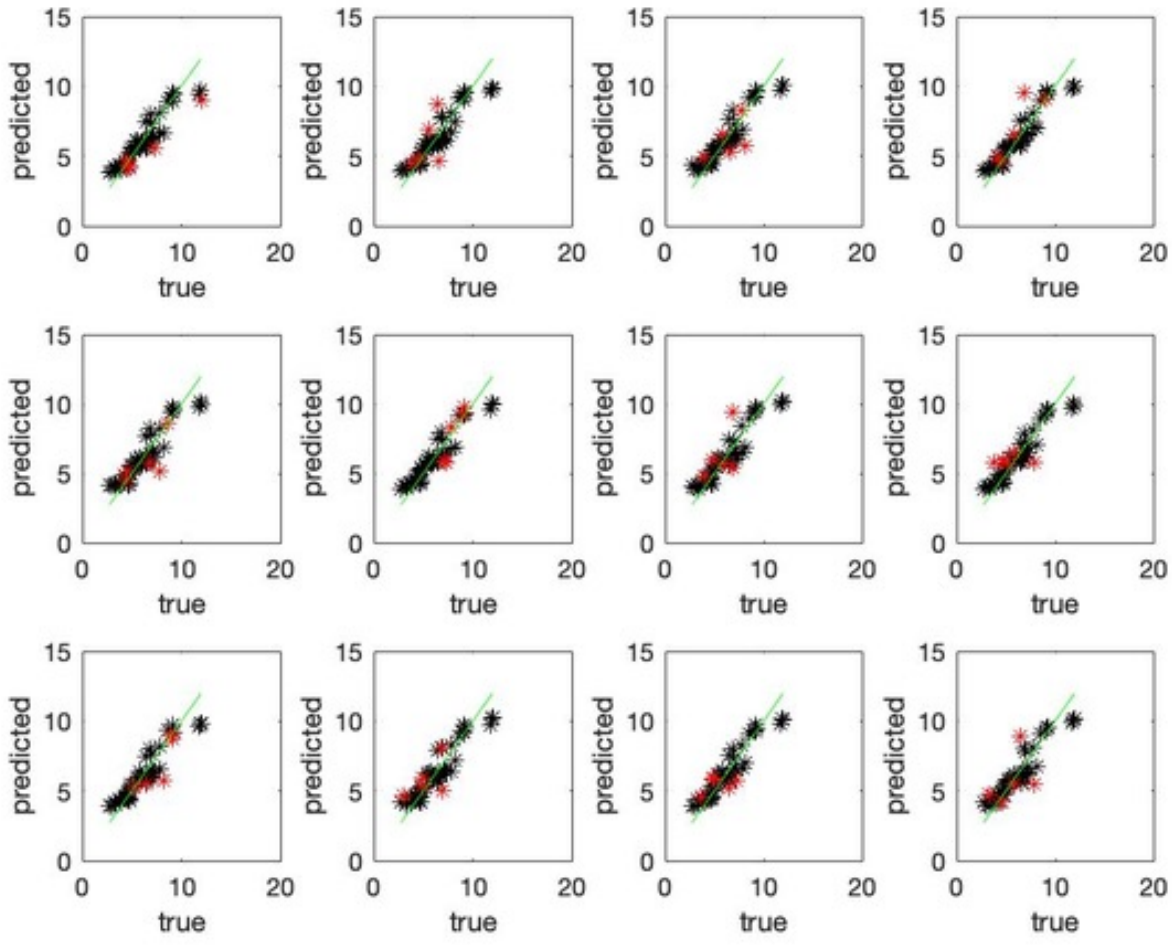


Bone strength vs persistence



Corelation close to 70%

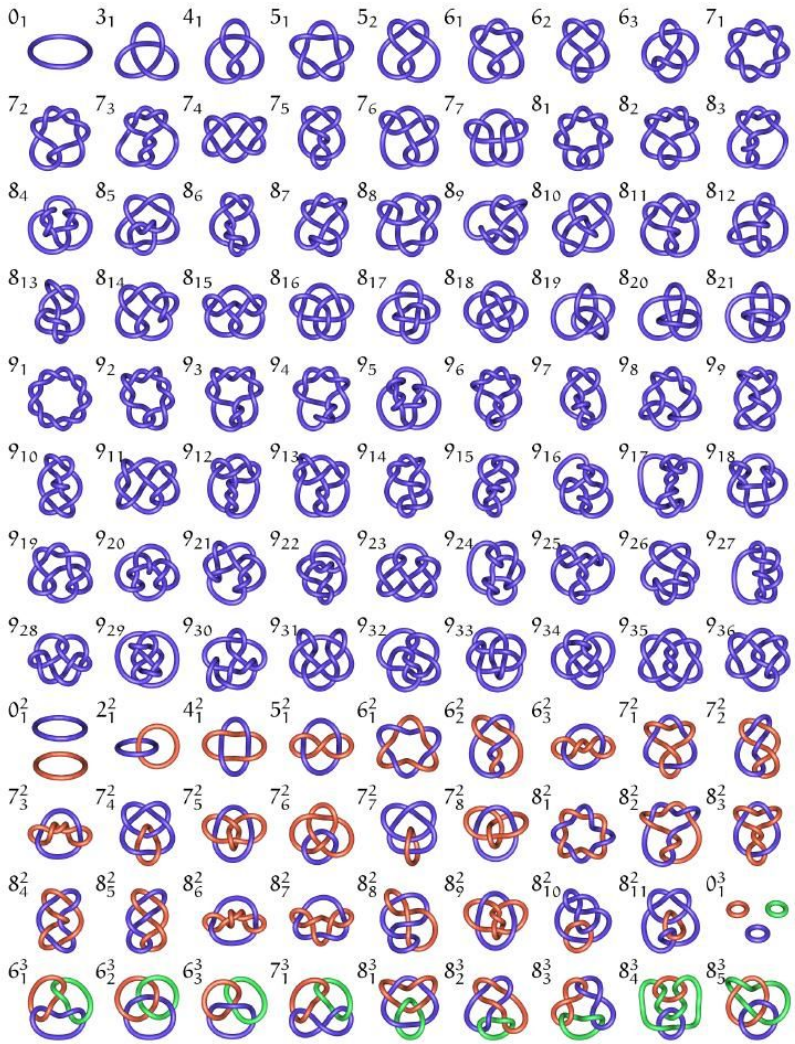
Cross validation



Questions

1. We have clear correlation between the compressing force and the persistence of the bone.
2. Better than a correlation with a bone density (measured as averaged value of a pixel).
3. Can we get similar information in vivo, from a standard CT?
4. Can we push it to clinical practice?

Knots and their properties

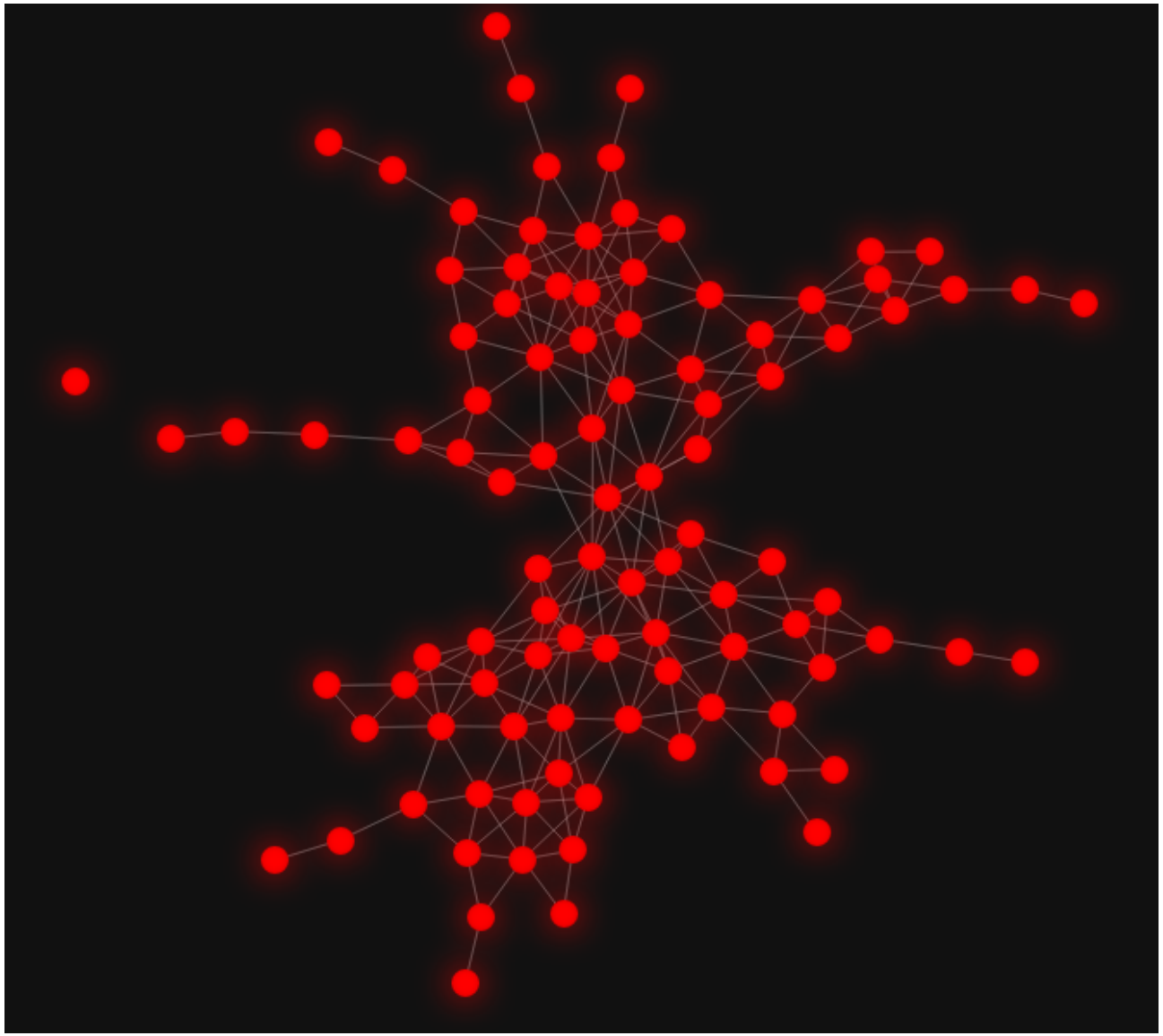


With Radmila Sazdanovic

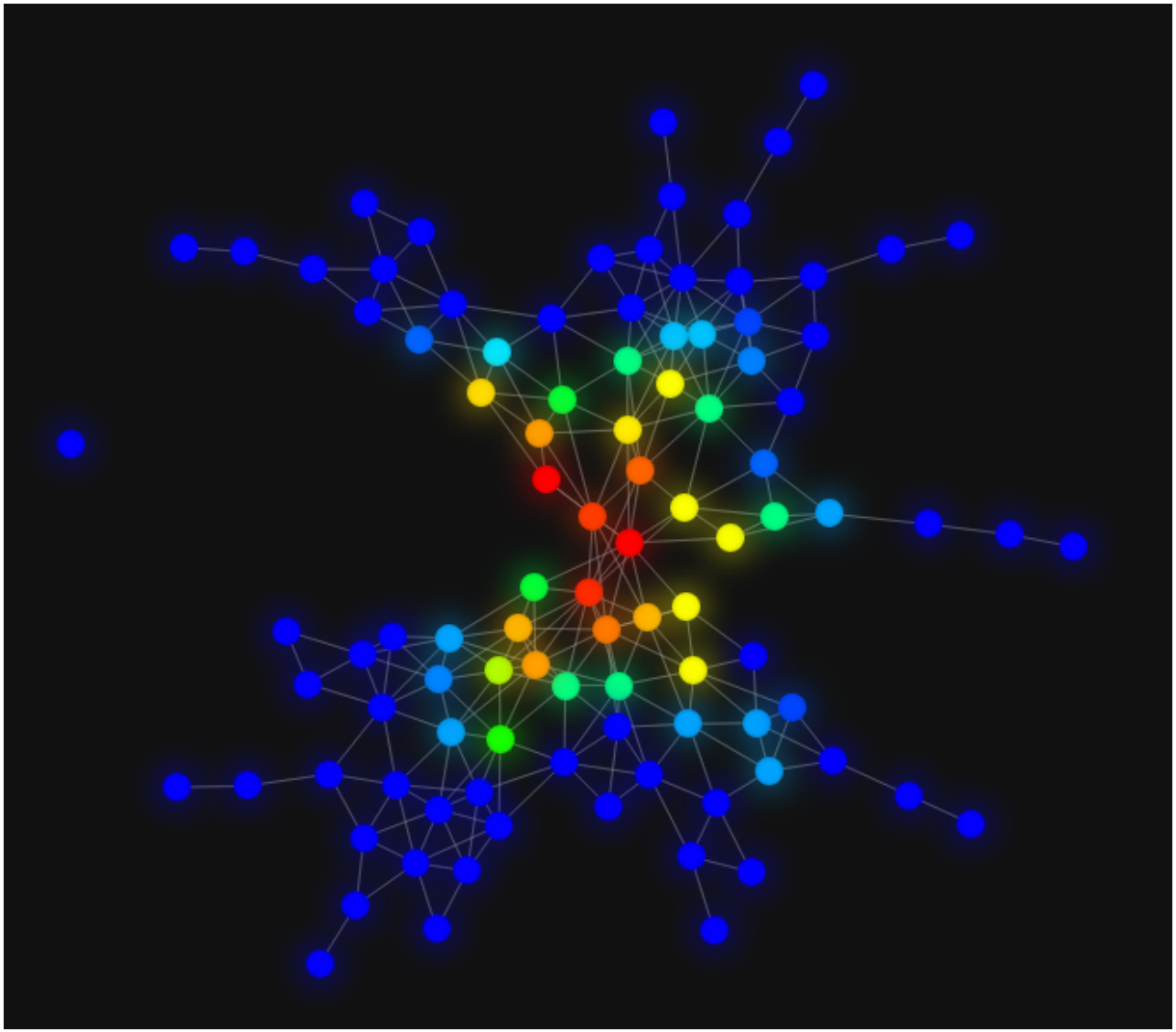
Knots and their properties, functional Ball Mapper.

- A knot is an embedding of S^1 to $\mathbb{R}^3$ up to continuous deformations (isotopies).
- A number of so called knot polynomials (Alexander, Jones, HOMPFLY-PT) have been introduced to describe knots.
- Let us consider the Jones polynomials of all knots up to 15 crossings.
- They corresponds to point cloud in 32 dimensional space.

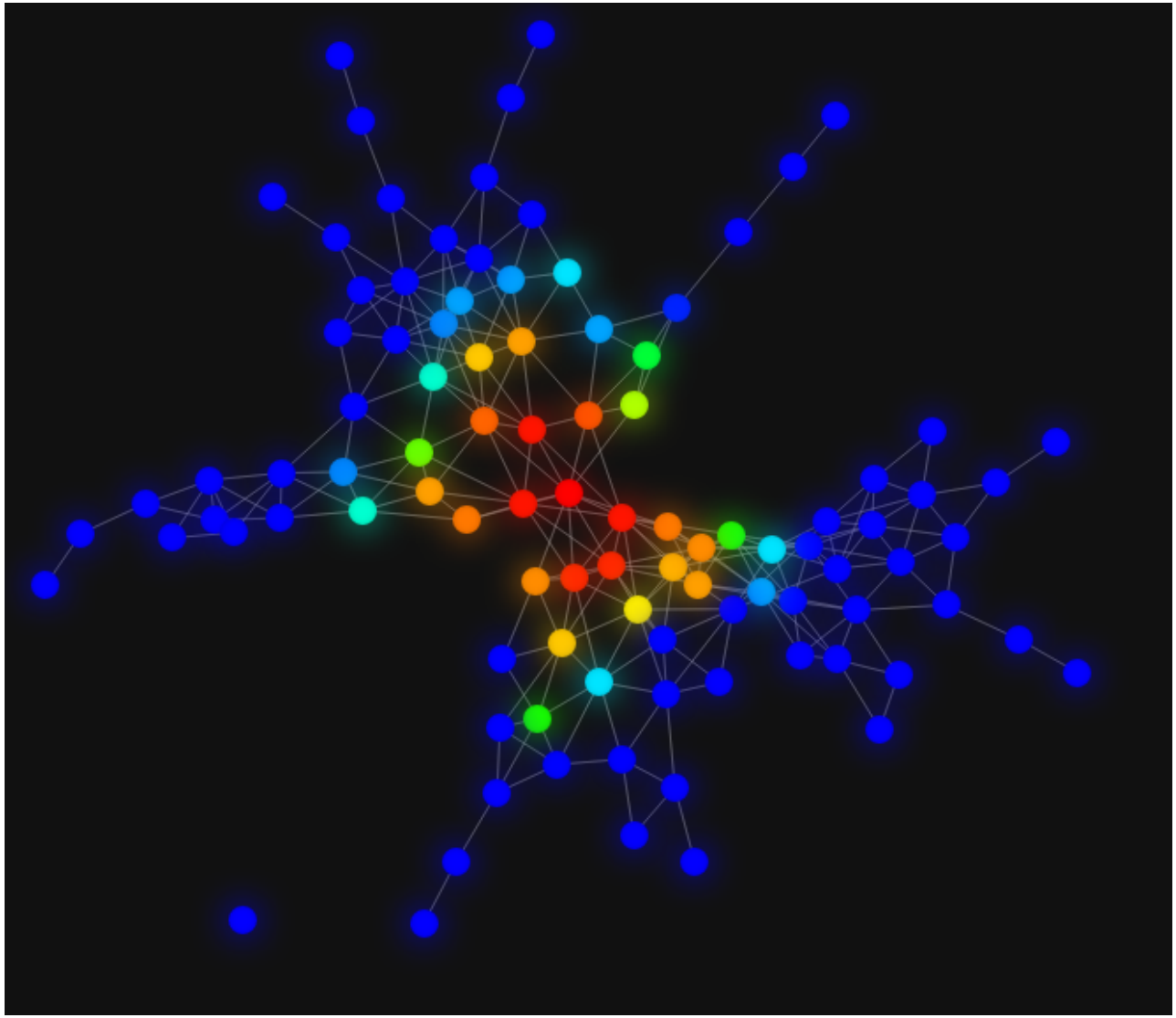
The shape of Jones polynomials of knots.



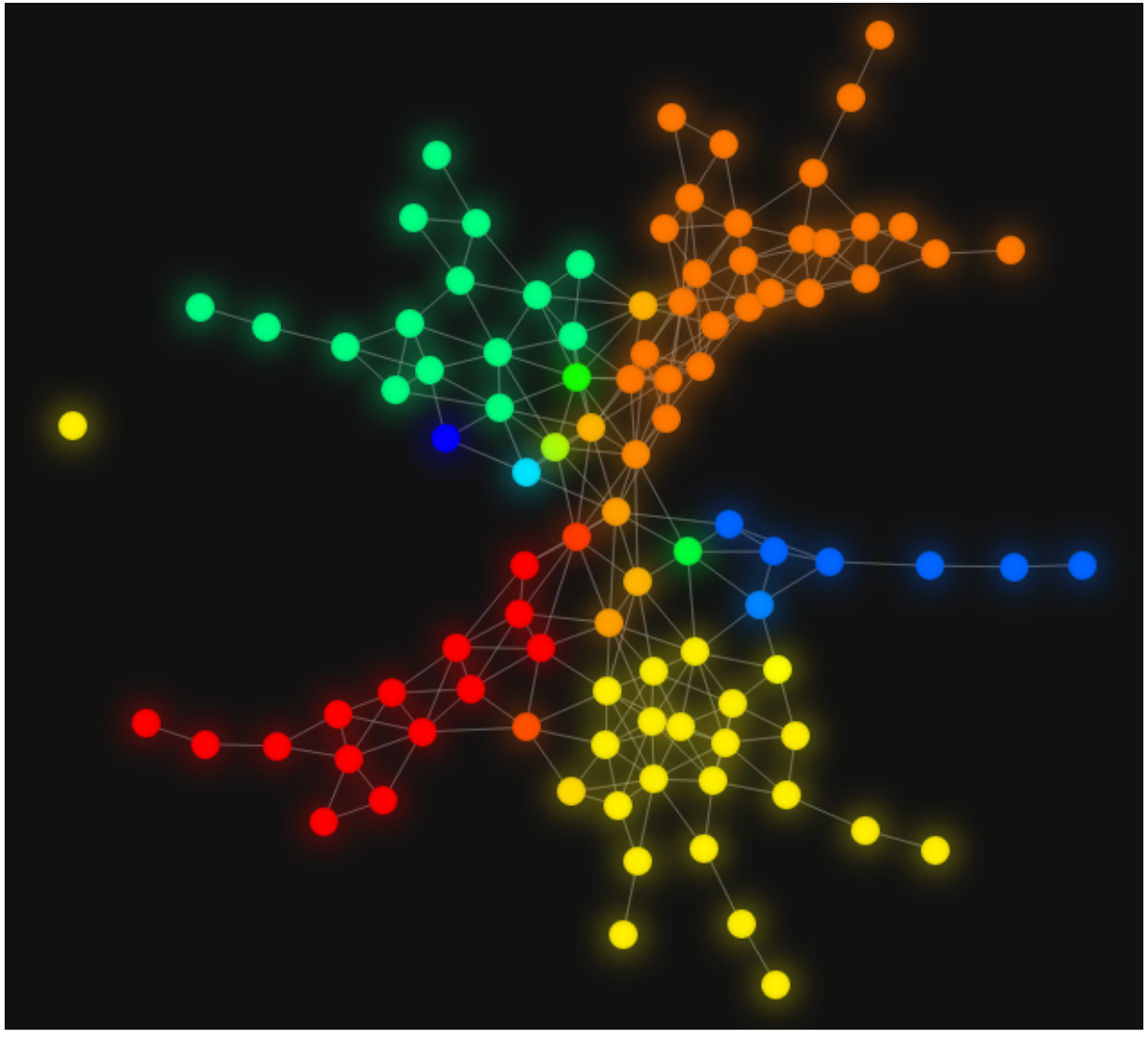
Jones polynomials, coloured by the number of crossings.



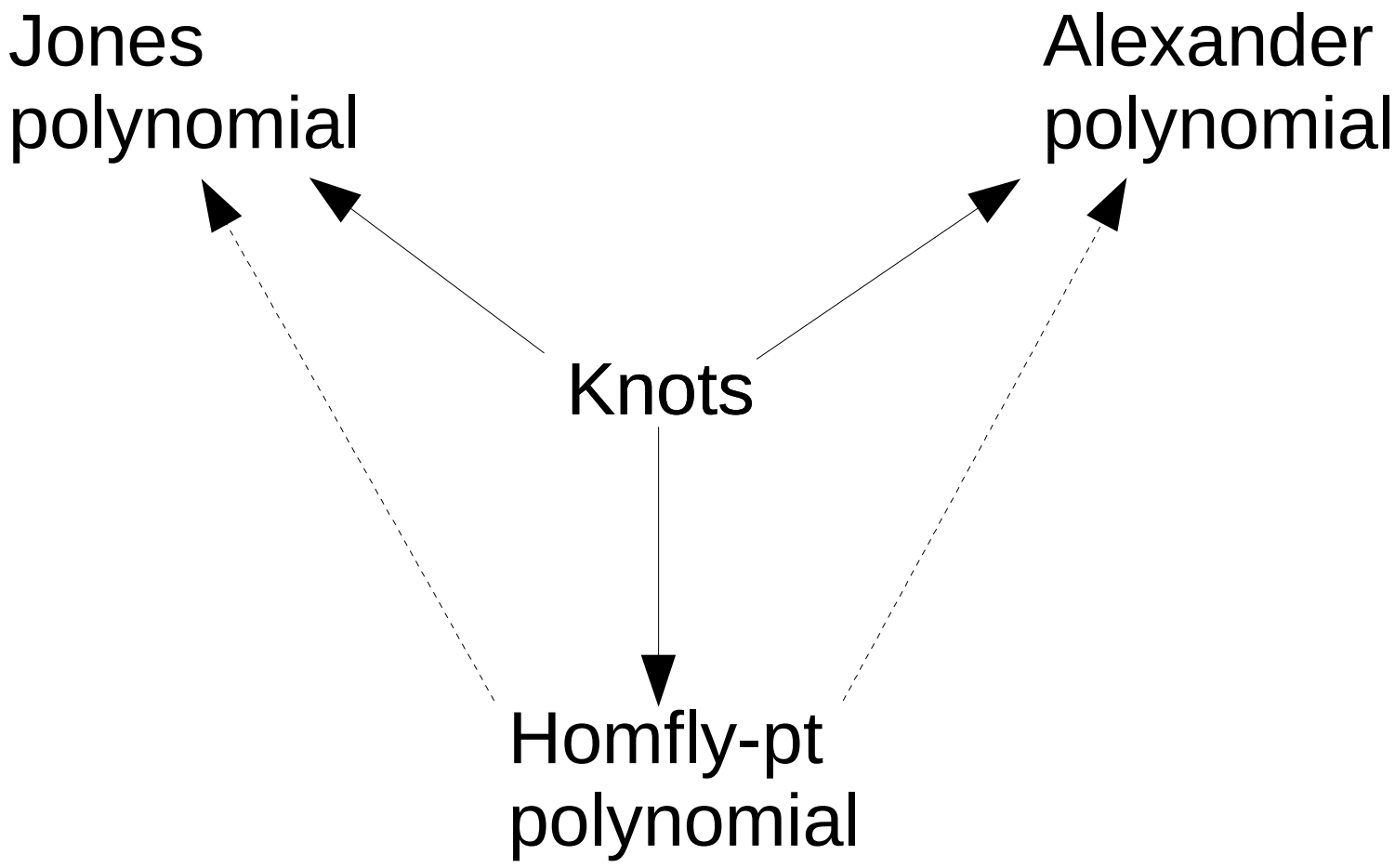
Jones polynomials, alternating vs non–alternating.



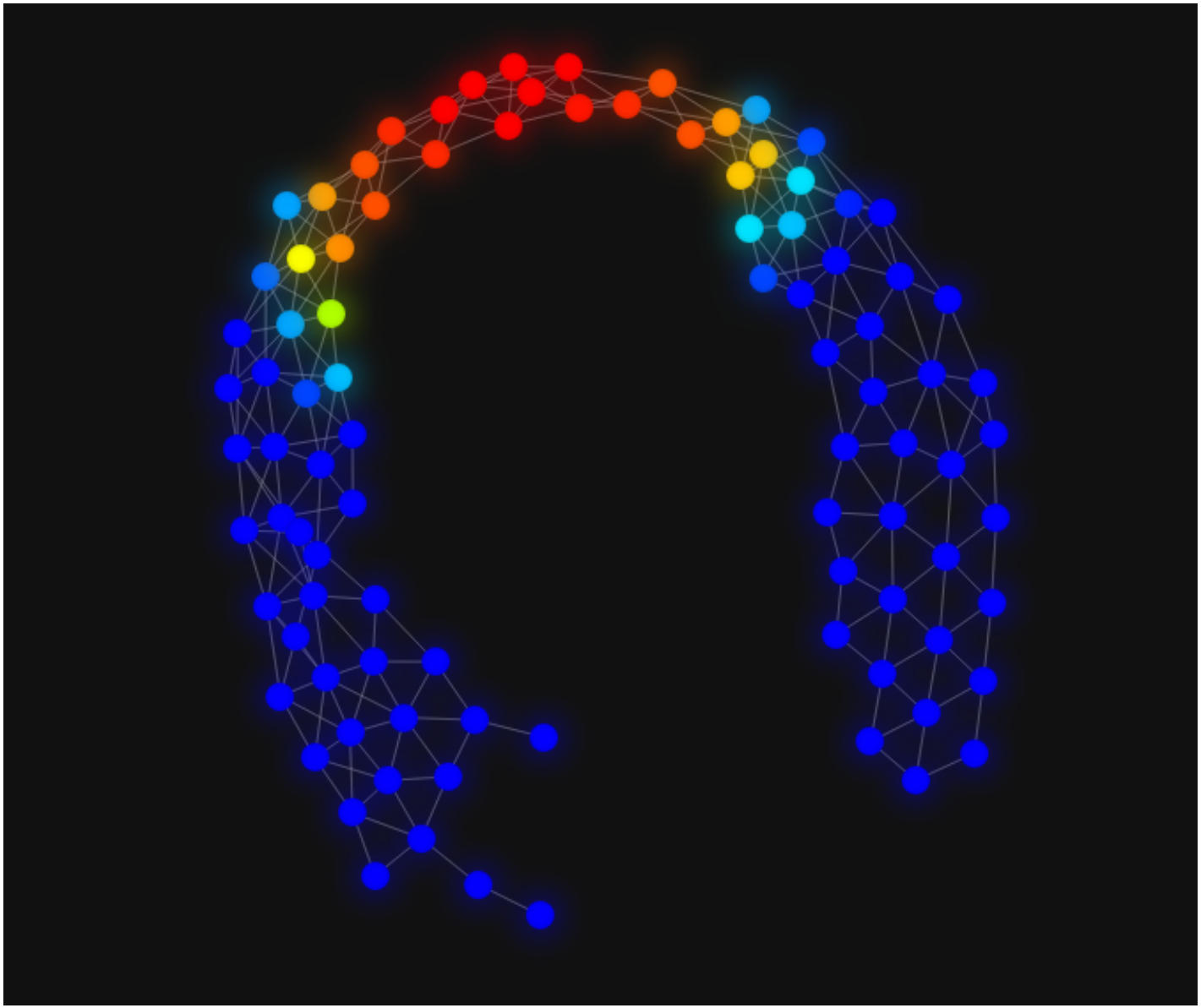
Jones polynomials, knots signature.



Other polynomials.

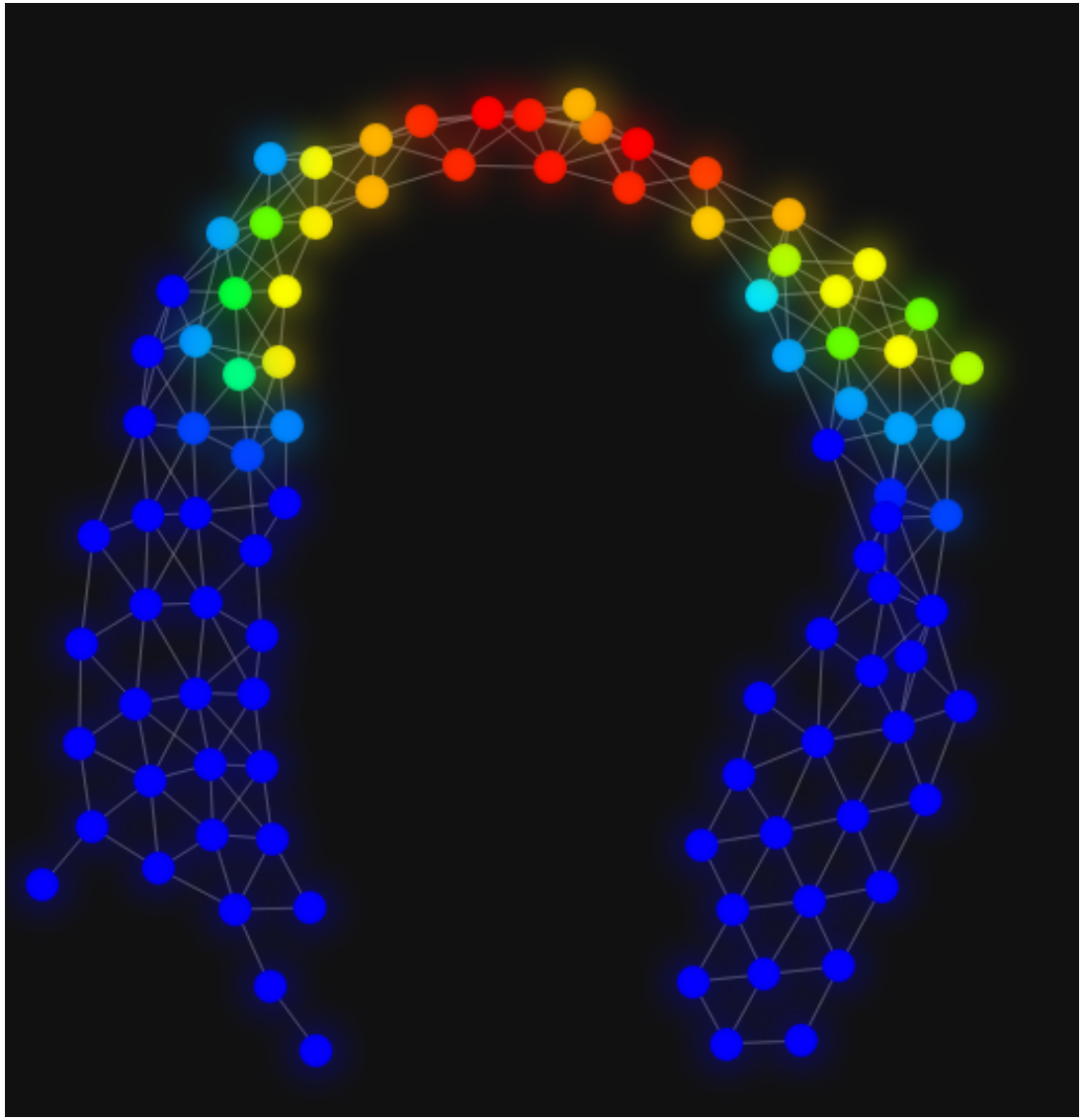


Alexander polynomial, $r = 45$.



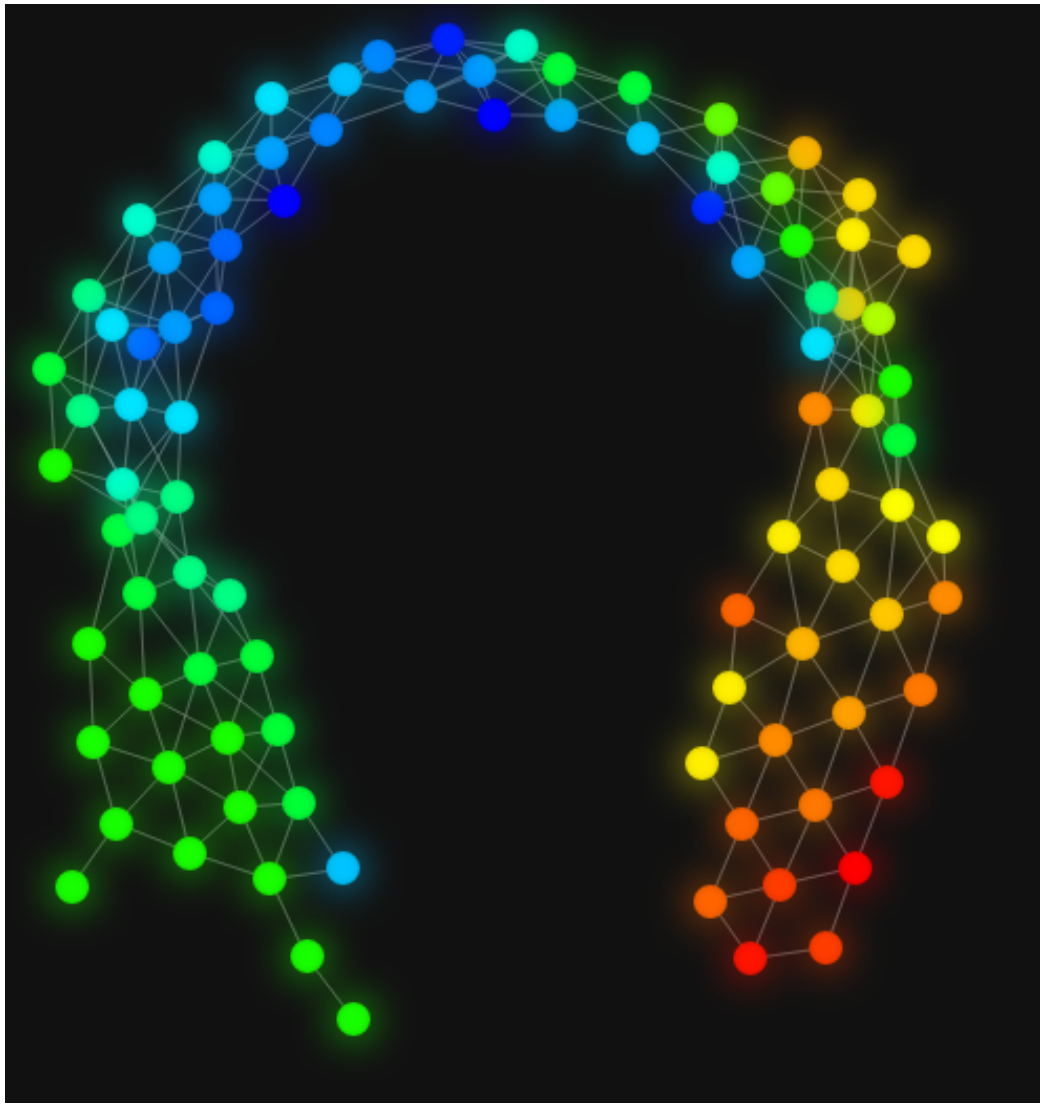
Alternating vs non alternating.

Alexander polynomial, $r = 45$.



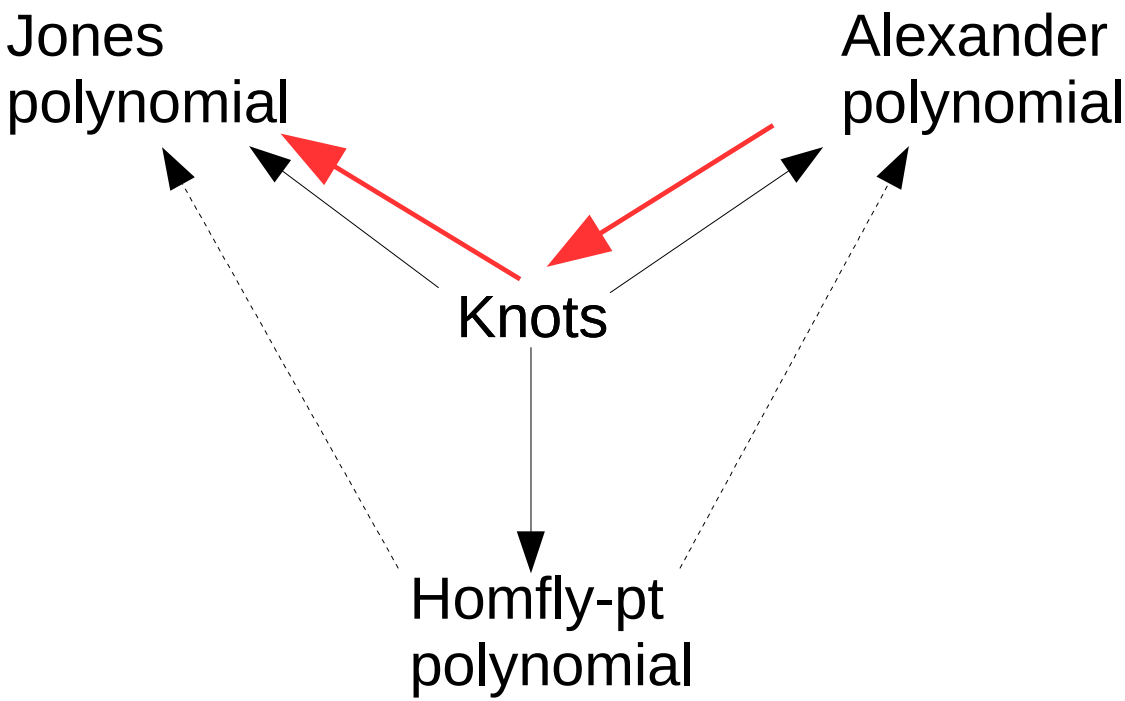
Number of crossings.

Alexander polynomial, $r = 45$.

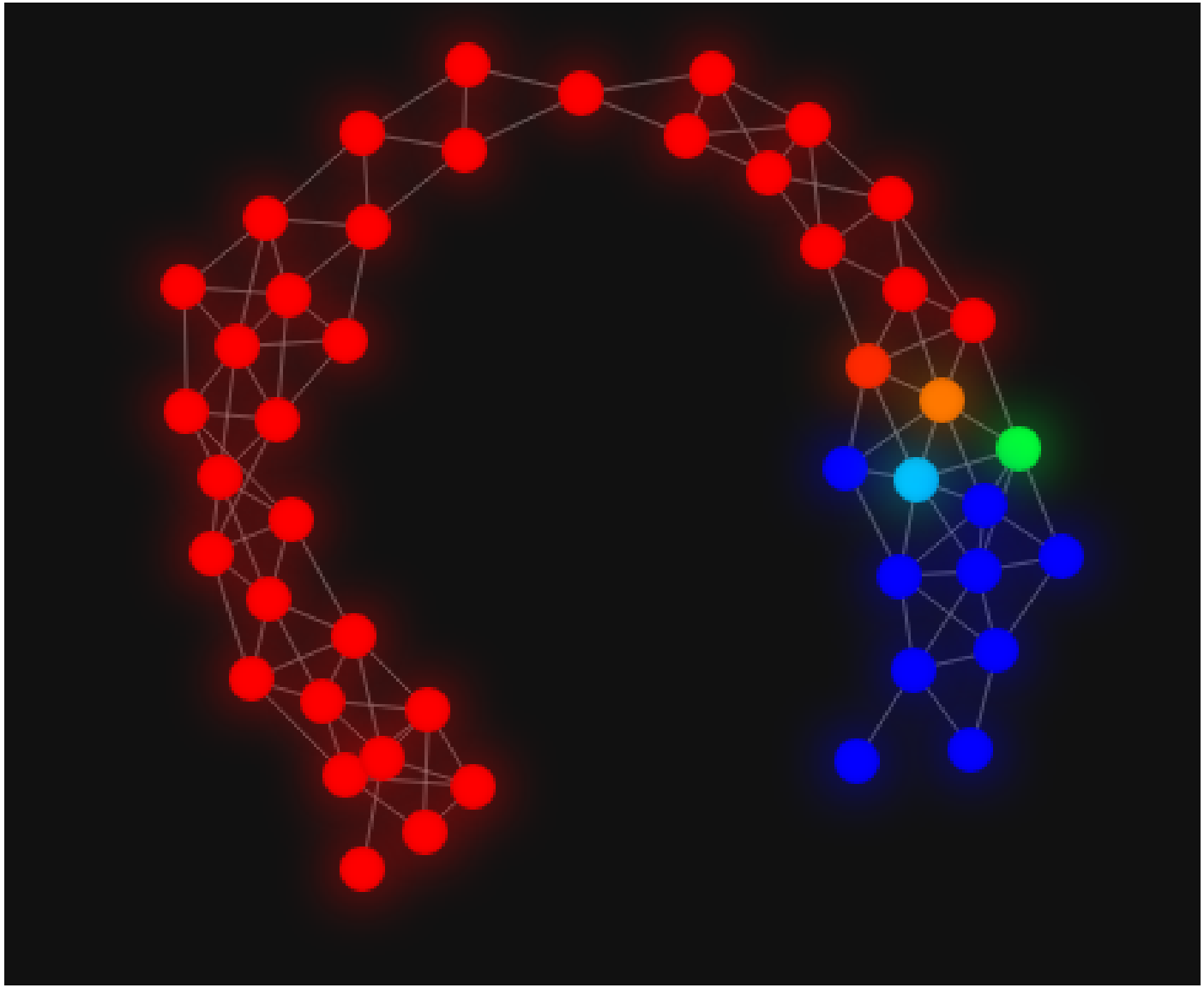


Knot's signature.

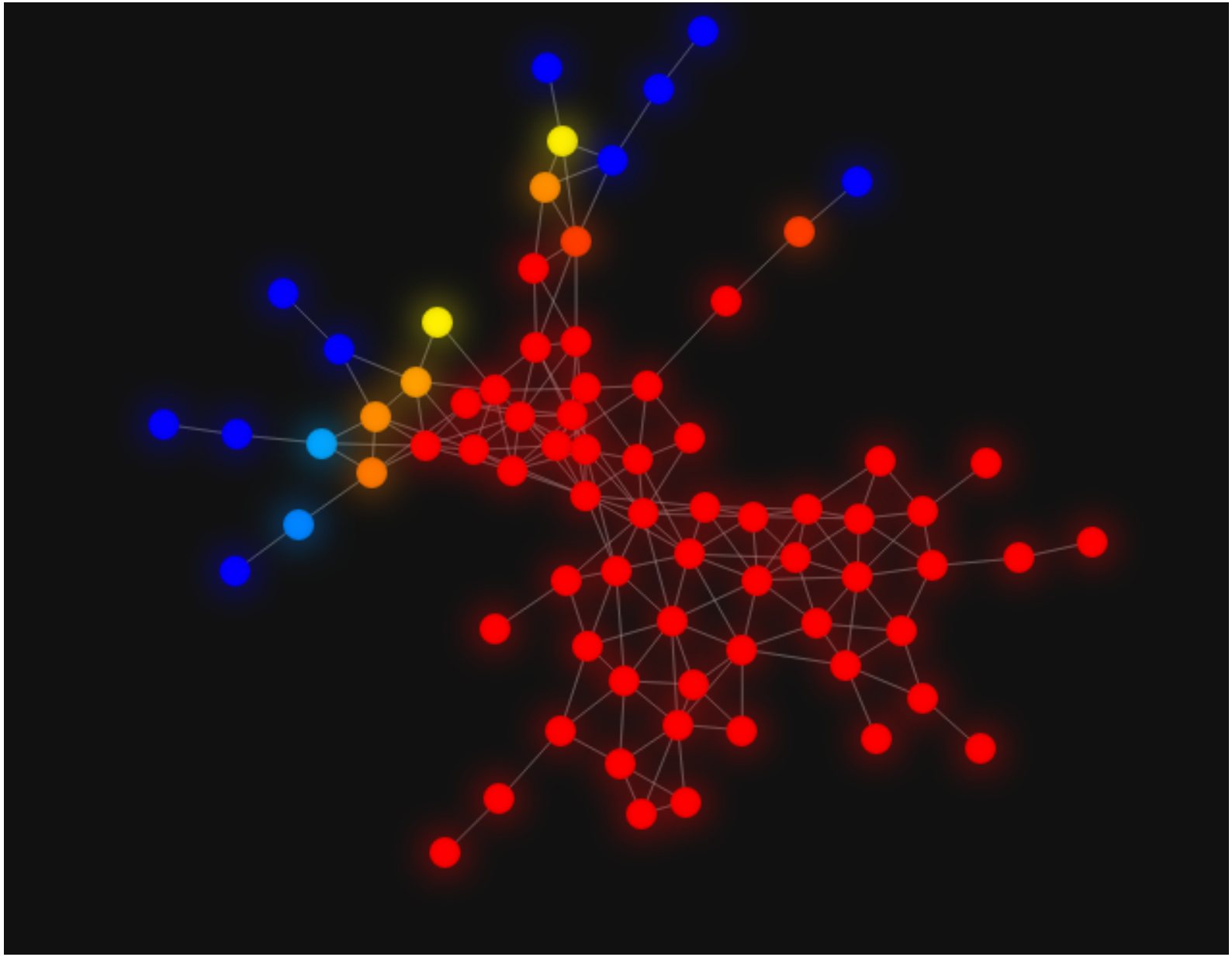
How Alexander map to Jones?



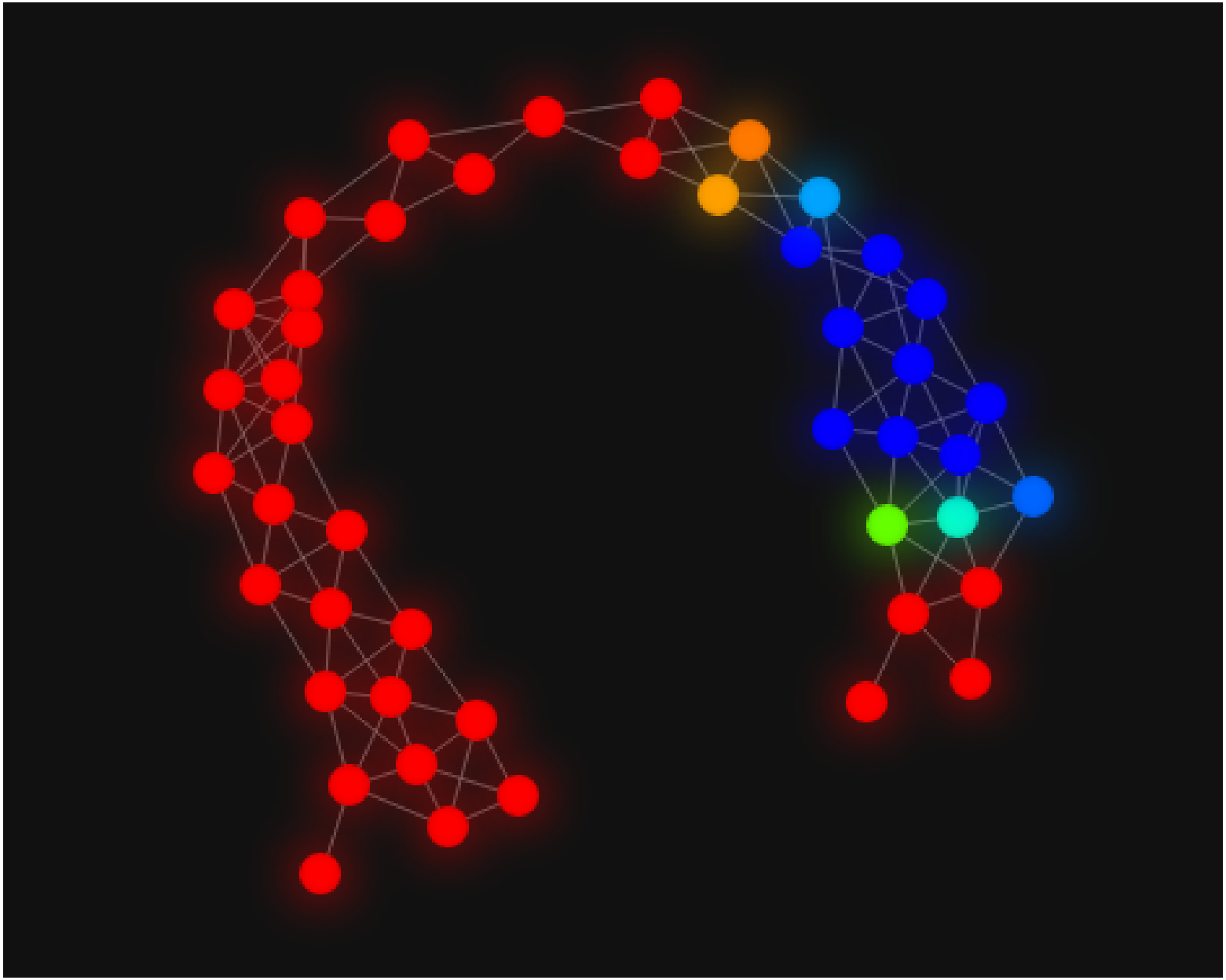
How Alexander map to Jones?



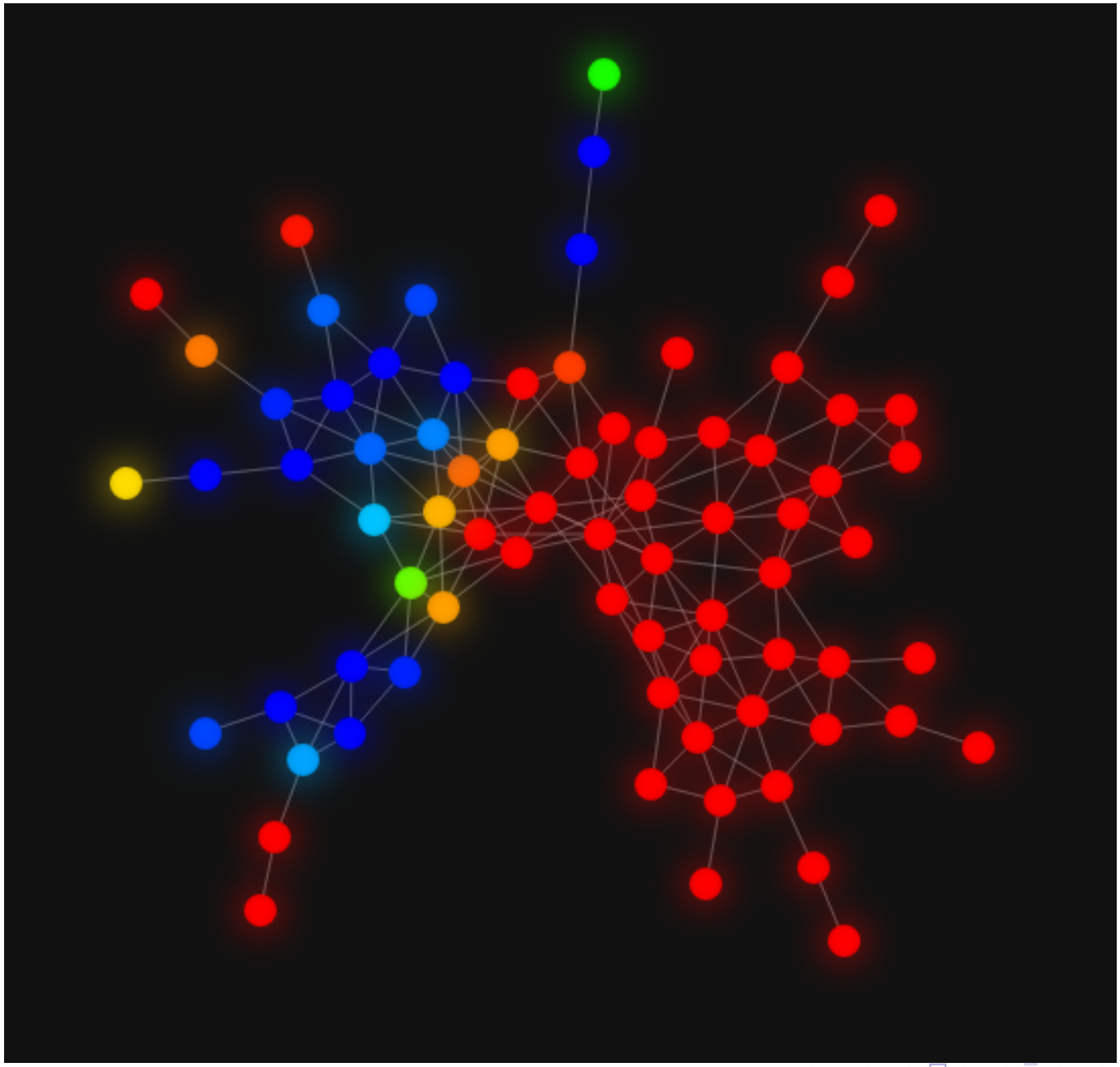
How Alexander map to Jones?



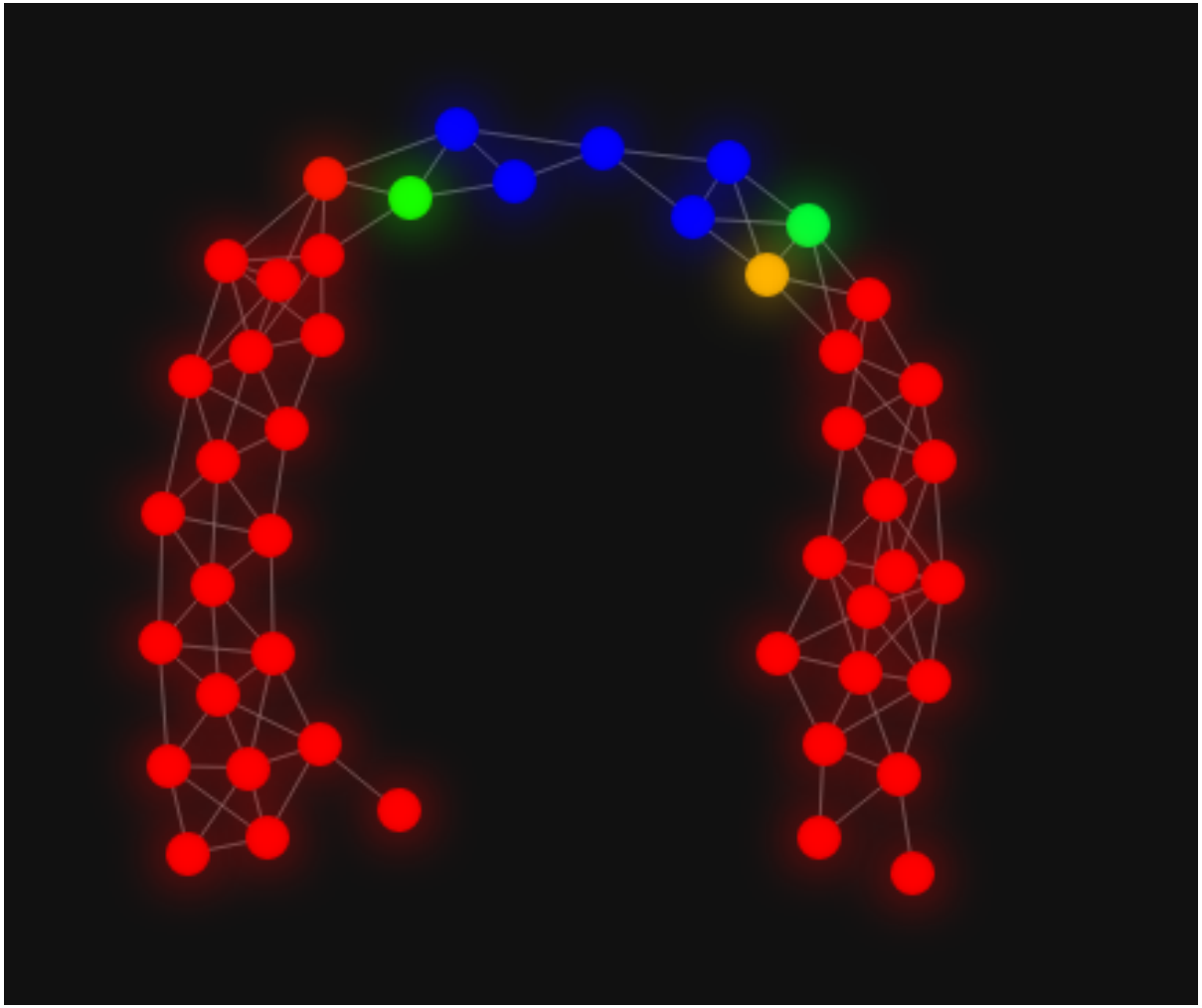
How Alexander map to Jones?



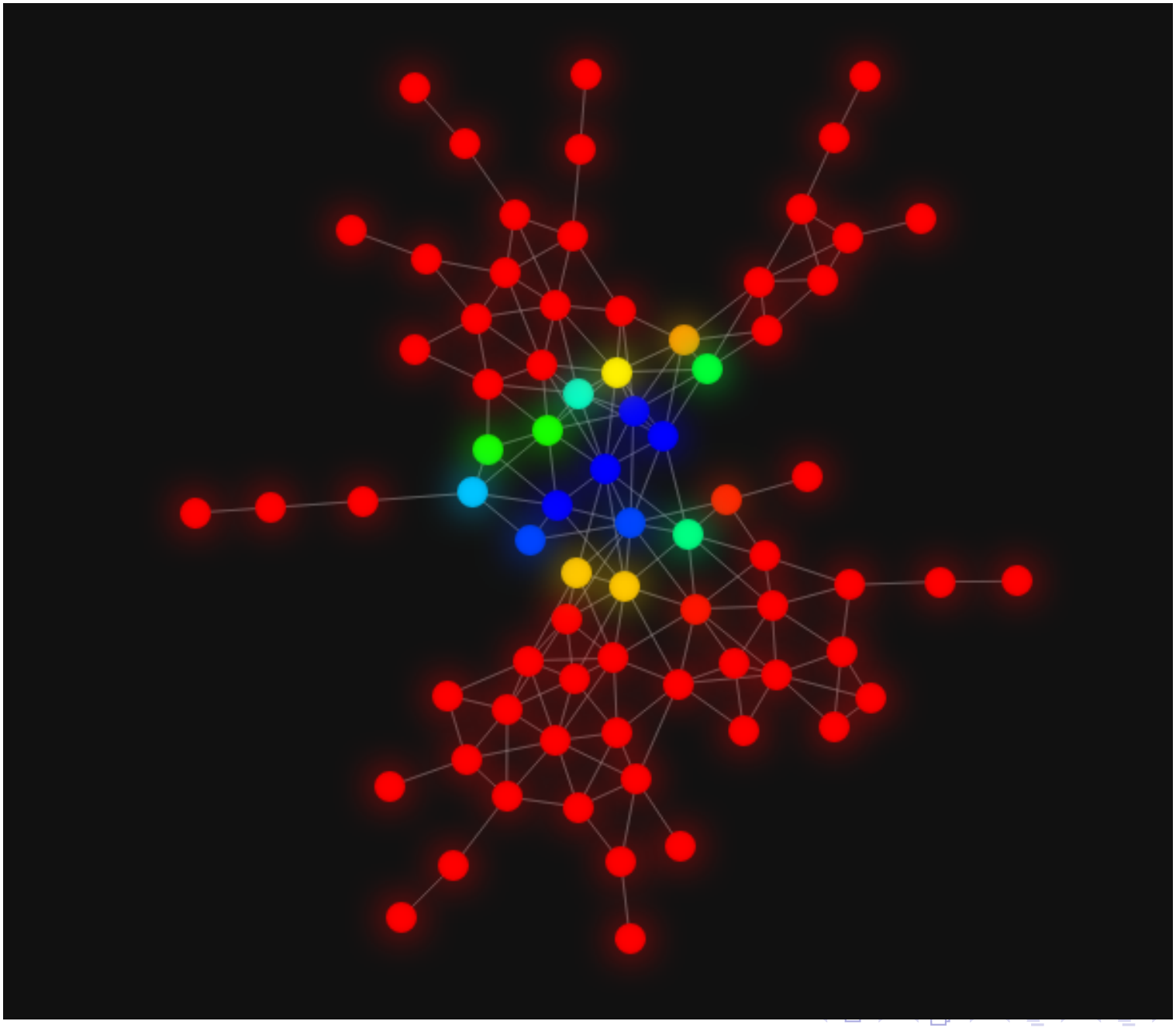
How Alexander map to Jones?



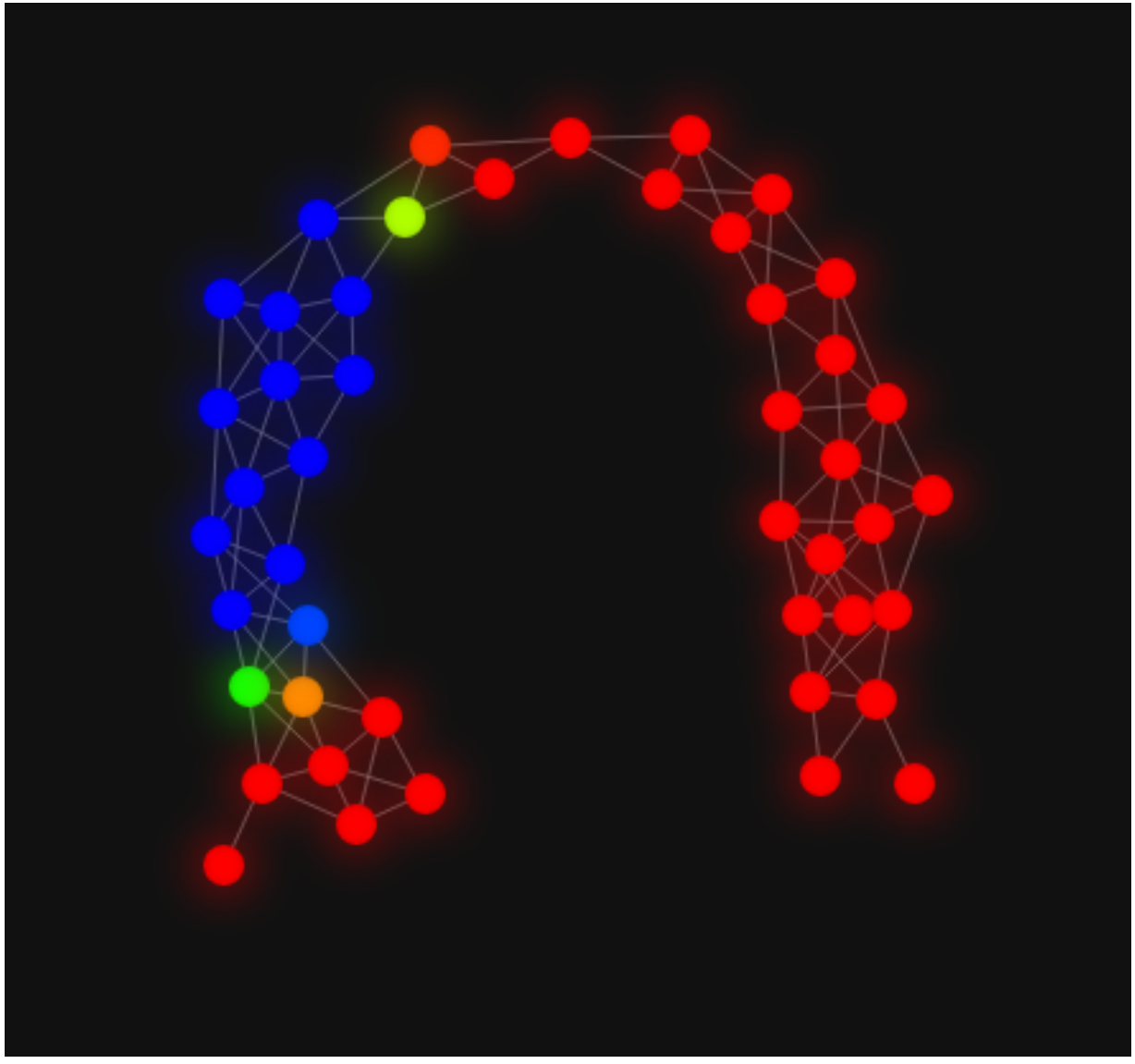
How Alexander map to Jones?



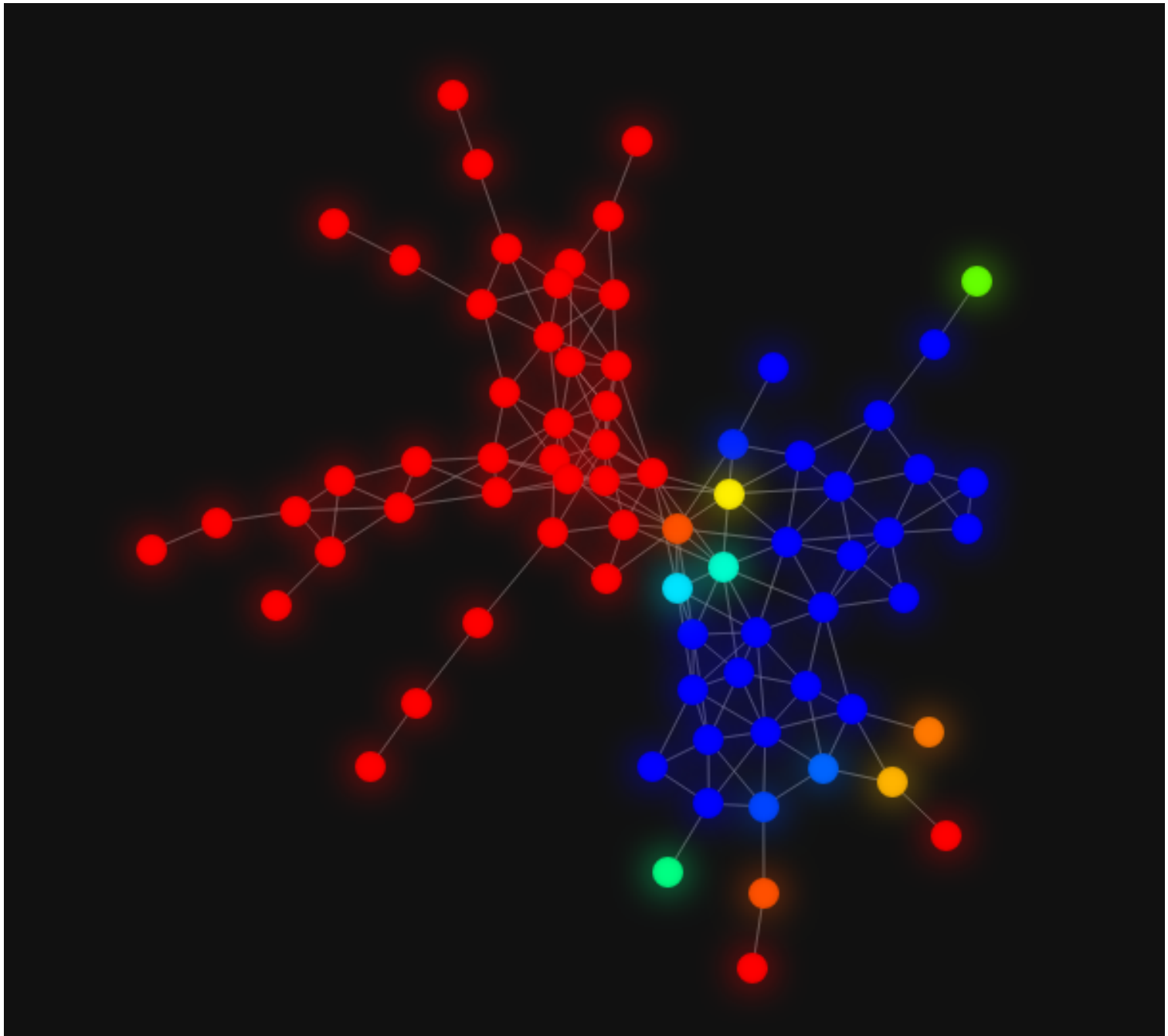
How Alexander map to Jones?



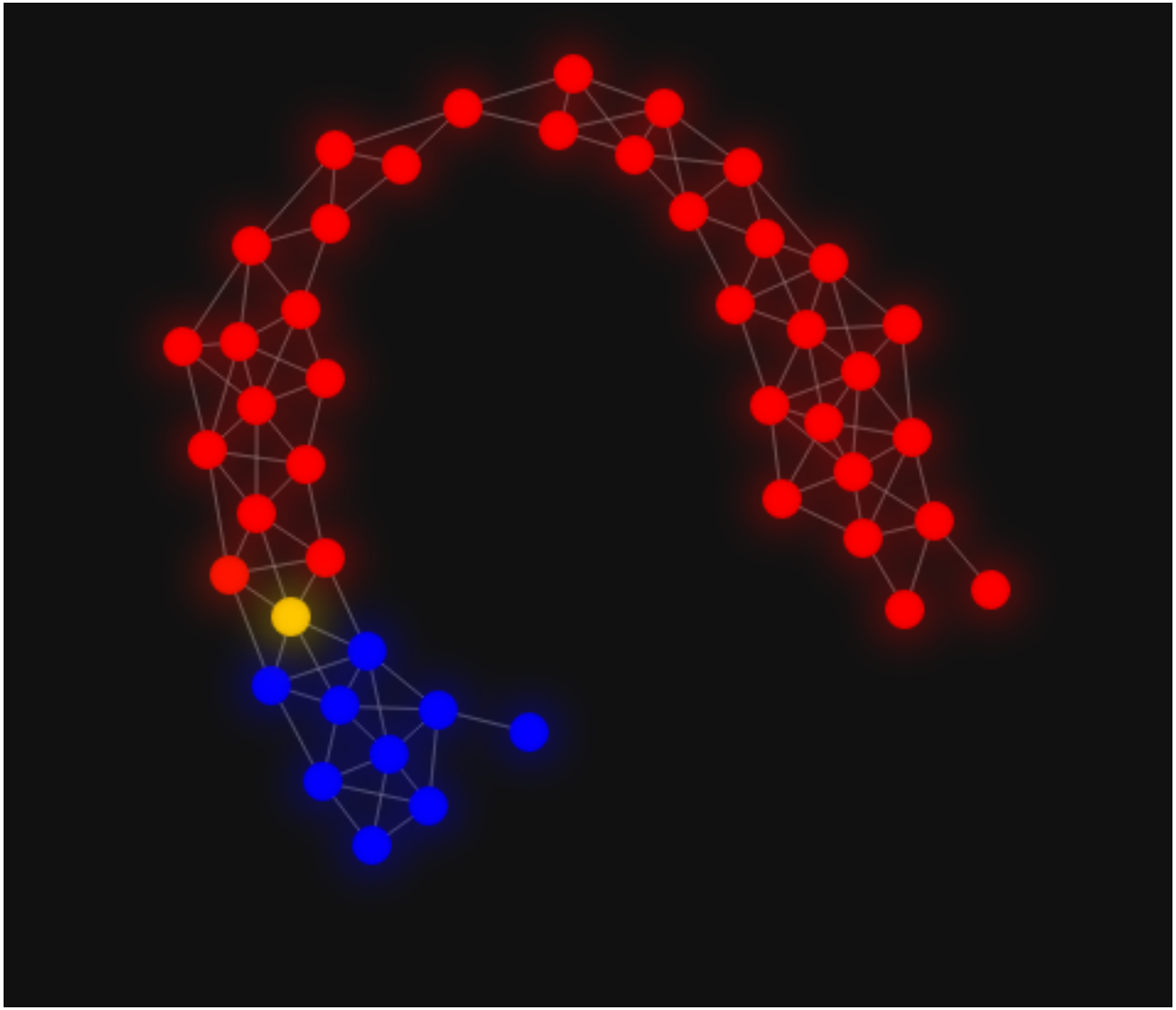
How Alexander map to Jones?



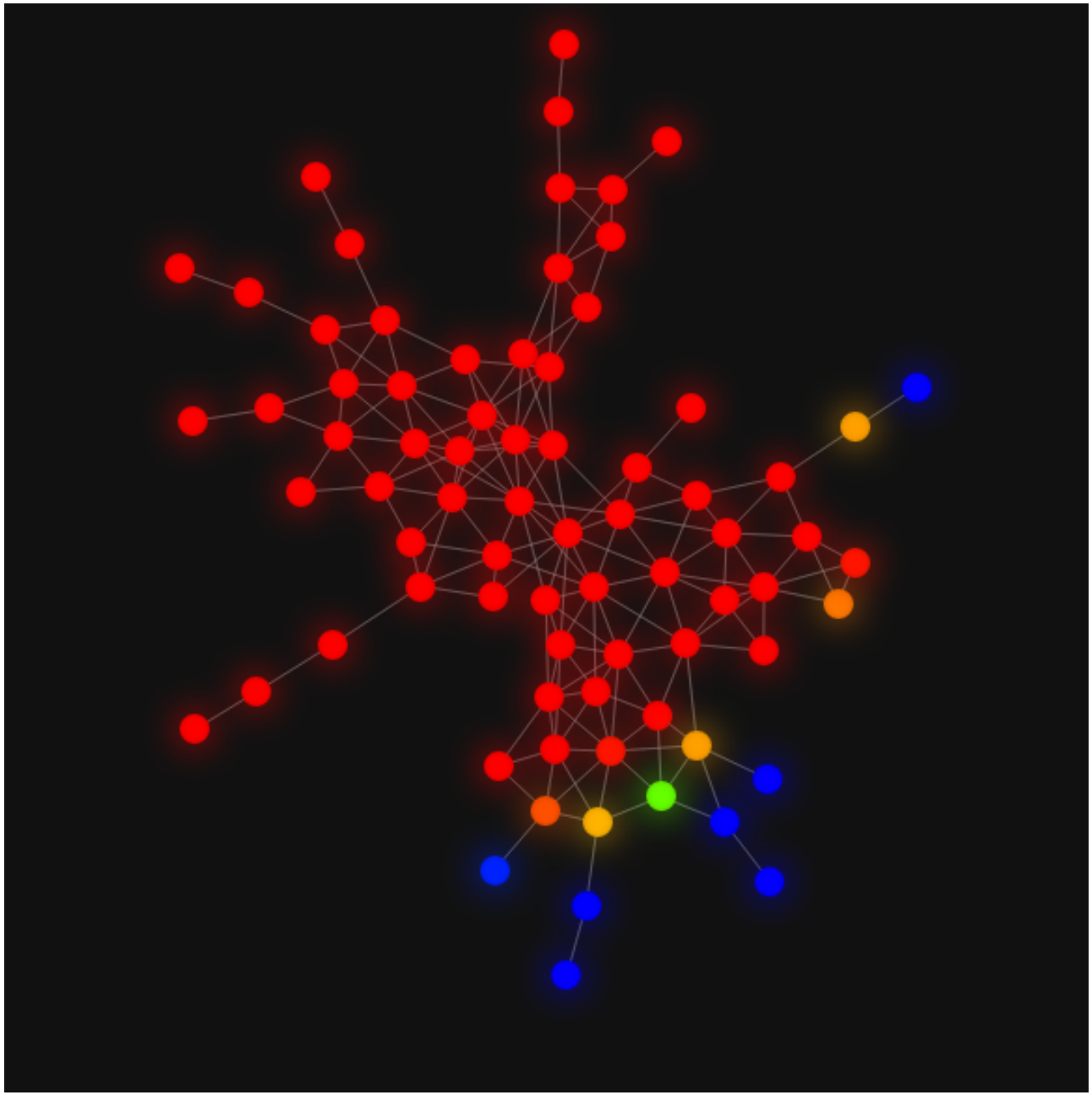
How Alexander map to Jones?



How Alexander map to Jones?



How Alexander map to Jones?



Some conclusions

1. We are living in the Data Revolution age.
2. Methods of Topology and Geometry are of a vital importance.
3. They can provide stable, well defined and interpretable descriptors that are of common interest.
4. More people and idea are needed.

Thank you for your time

`dioscuri-tda.org`

Postdocs, PhD and Visiting Researcher positions available!

Hope we will stay in touch.