Nordic-Baltic Astronomy Days,
Turku, Finland

Galaxy Unsupervised Learning classification combining UMAP and clustering algorithms

Luis E. Suelves

Tartu Observatory, University of Tartu, Estonia

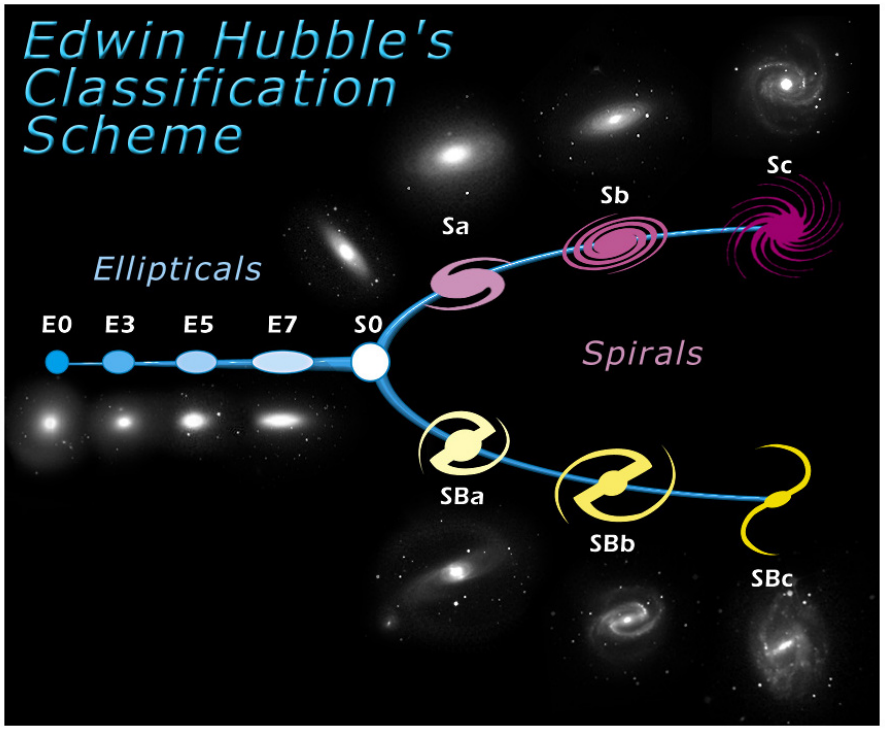
27th May, 2026

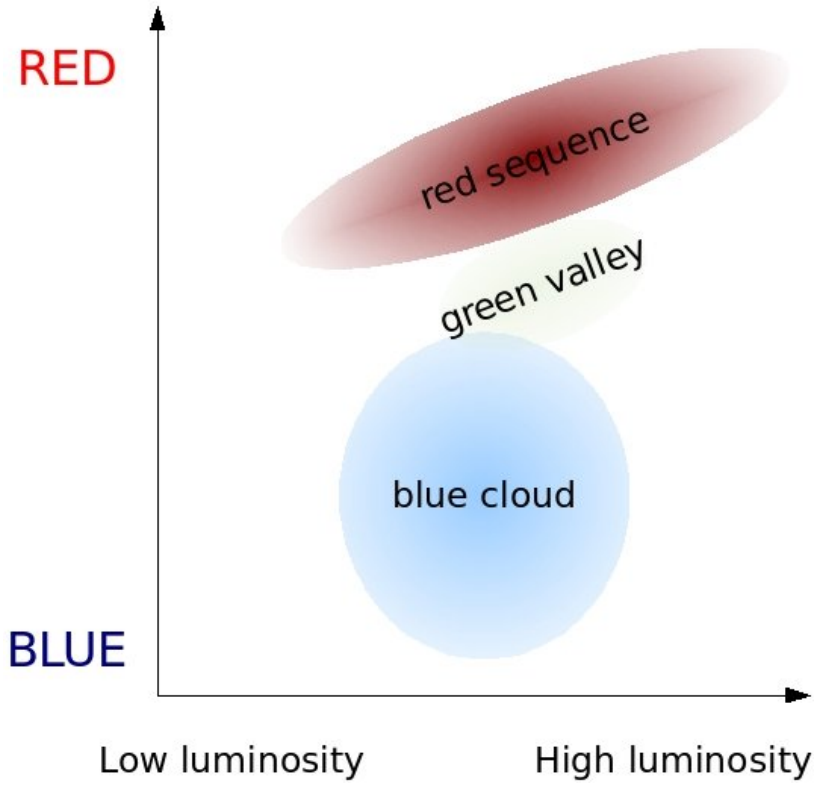
EXCOSM

Funded by the European Union

UNIVERSITY OF TARTU
Tartu Observatory
1632

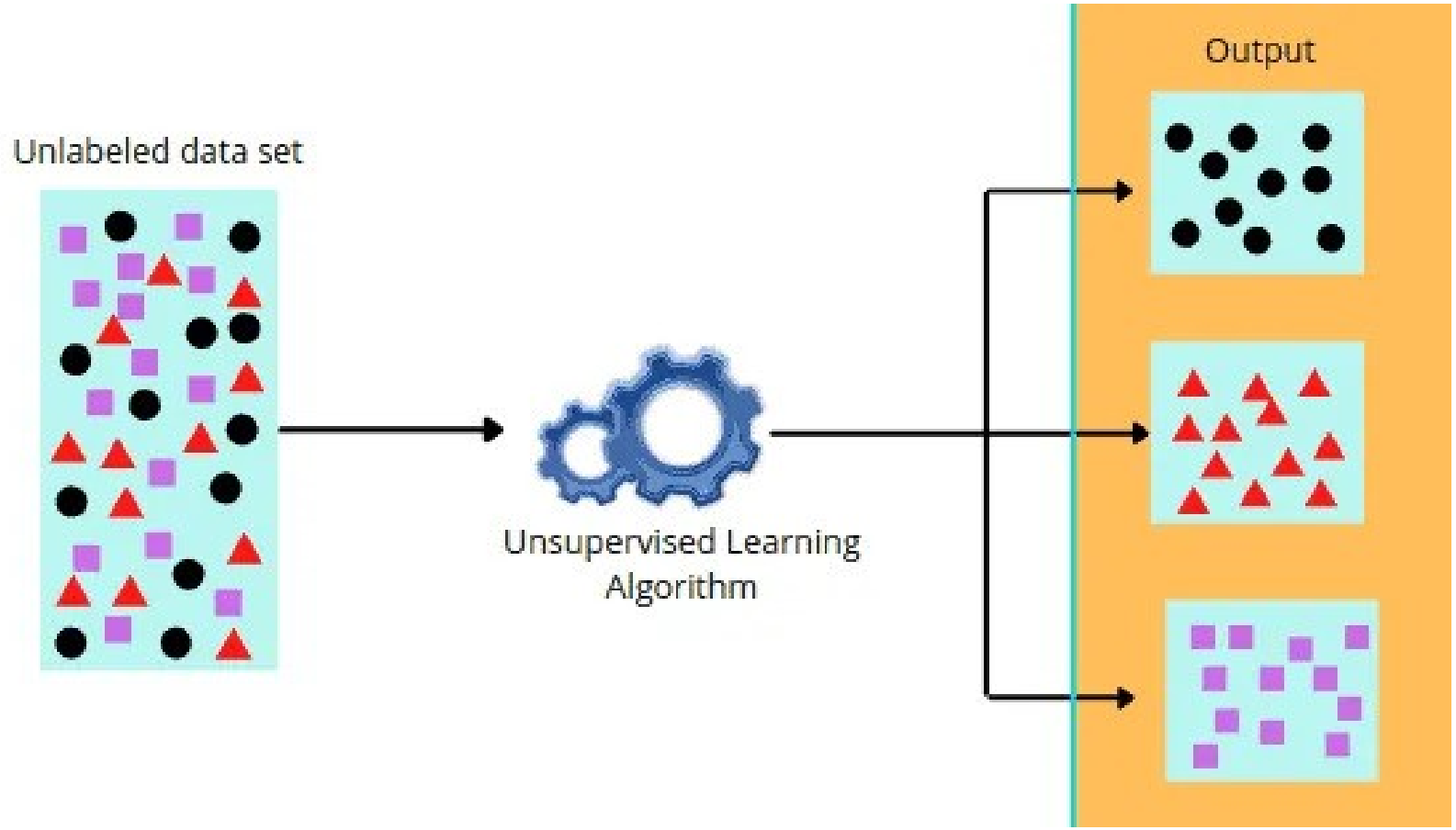
Galaxy Classification





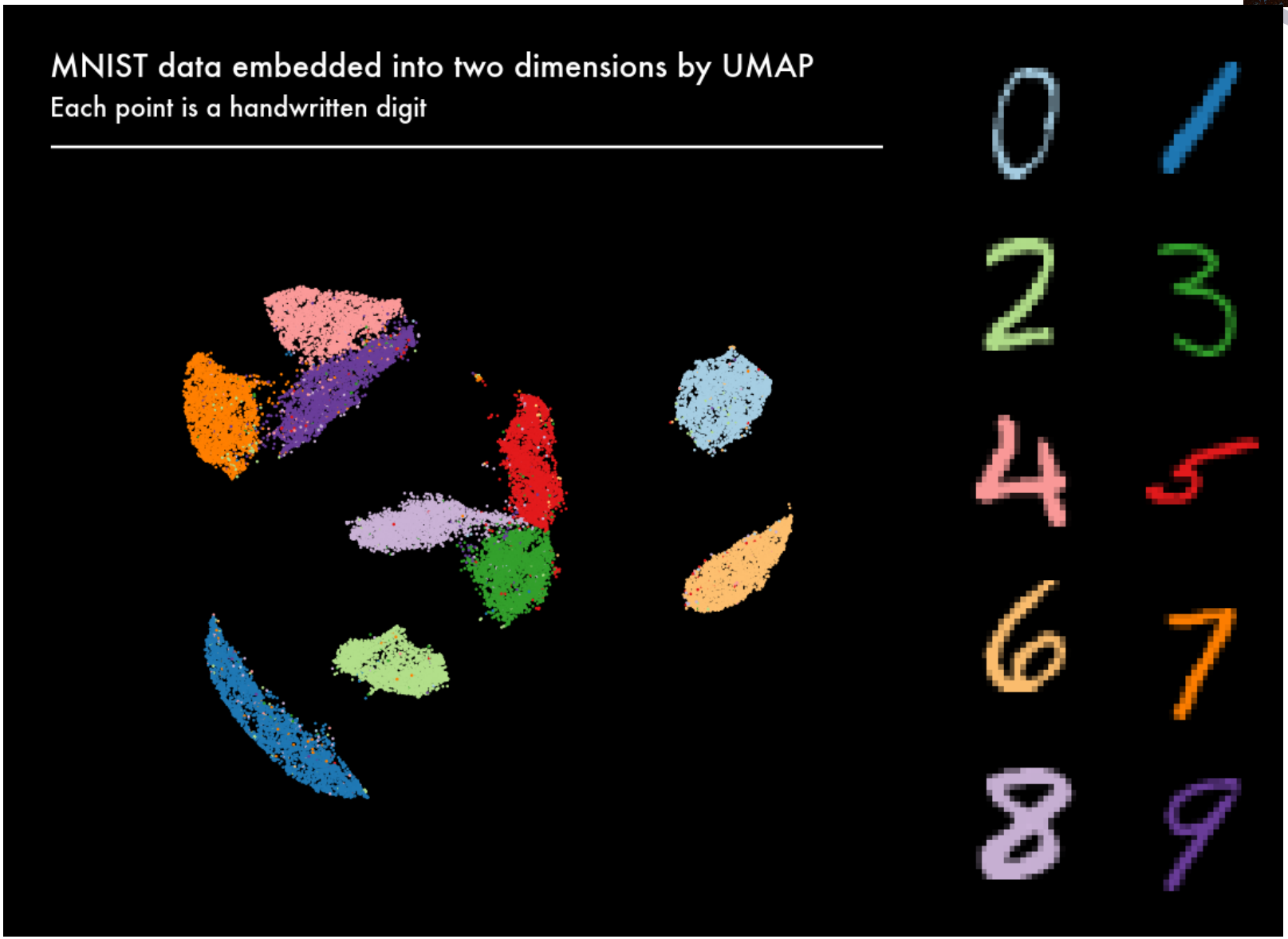
Unsupervised Machine Learning
What it is

Machine Learning on **unlabelled data**



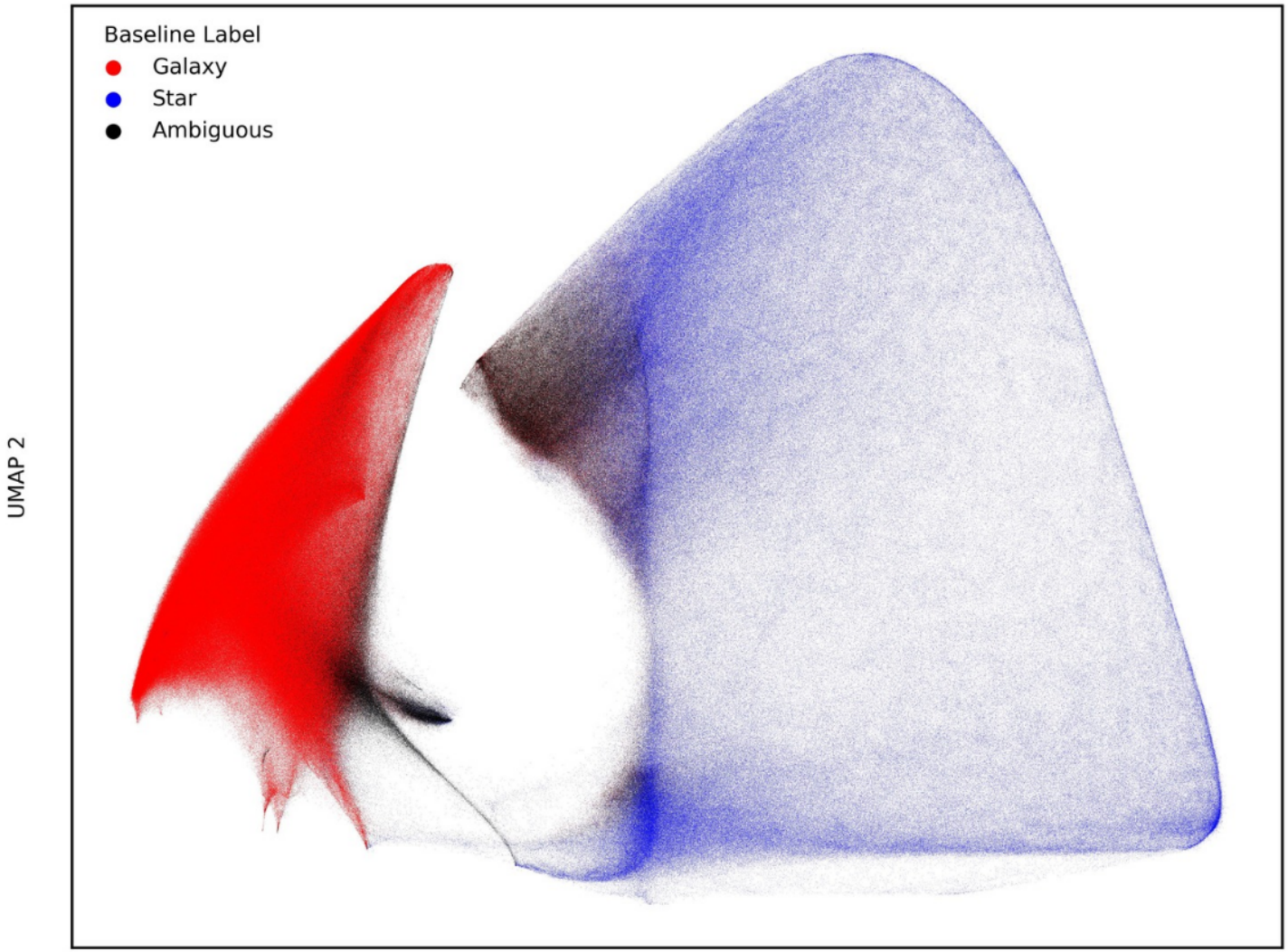
Credit: https://hands-on.cloud/ml-unsupervised-learning-guide/

Unsupervised Example I



Credit: Jeroen Janssens

Unsupervised Example II



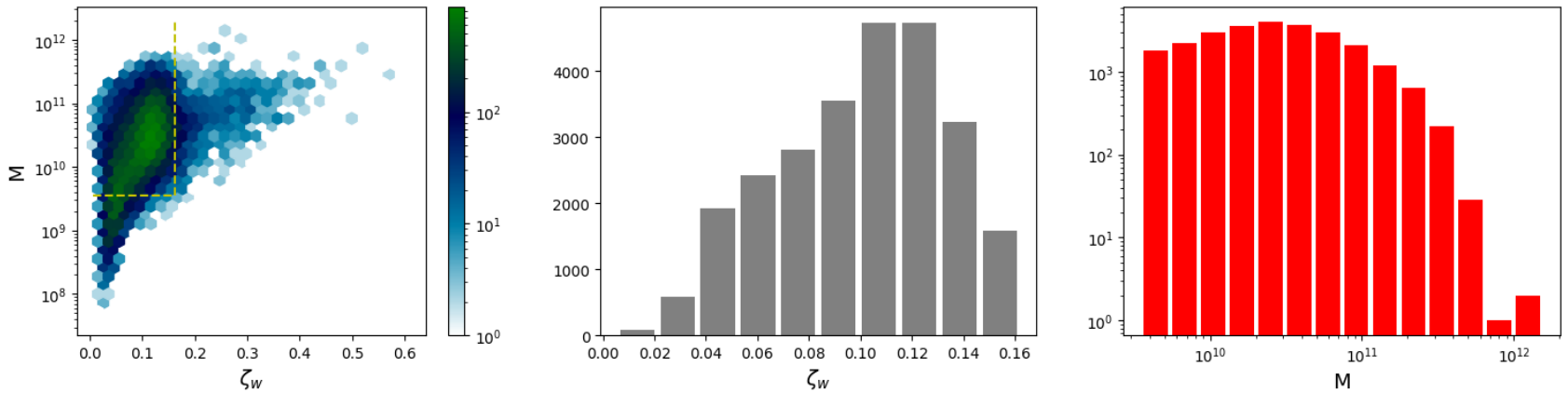
Credit: Cook +24

Can Unsupervised Machine Learning be useful for galaxy classification?!

What data, how, and why?

- GAMA spectroscopic sources (redshifts!) + Galaxy Zoo classifications

  mass limited sample



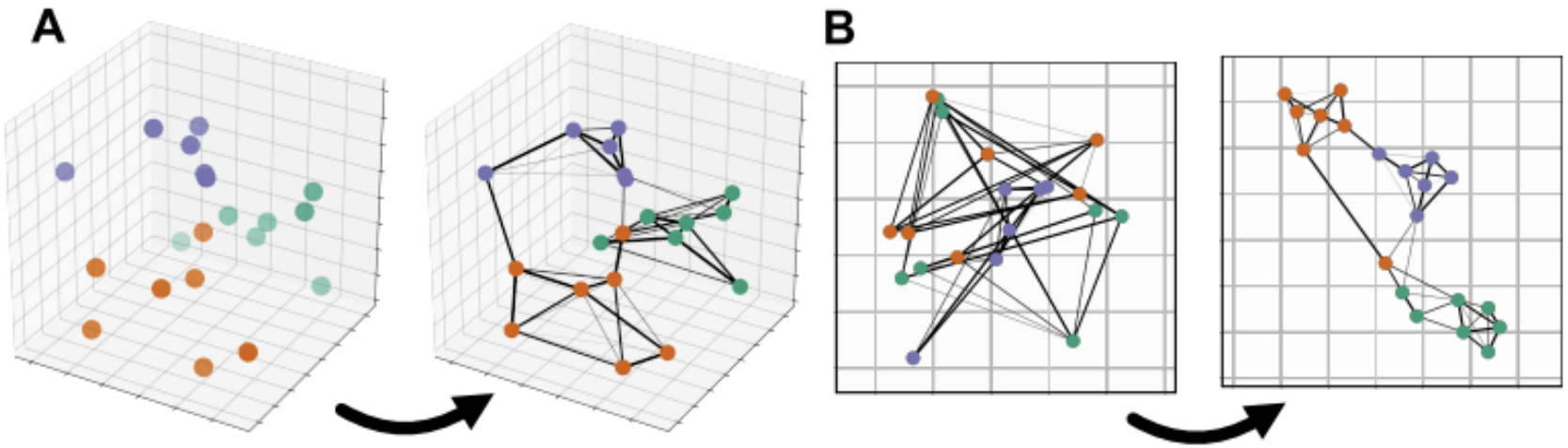
ζ = Ln (1 + z)

- k-corrected KiDS+VIKING photometric colours

  median-normalized $\rightarrow$ 4+5 bands $\rightarrow$ 36 colours

Step 1. Dimensionality Reduction

UMAP

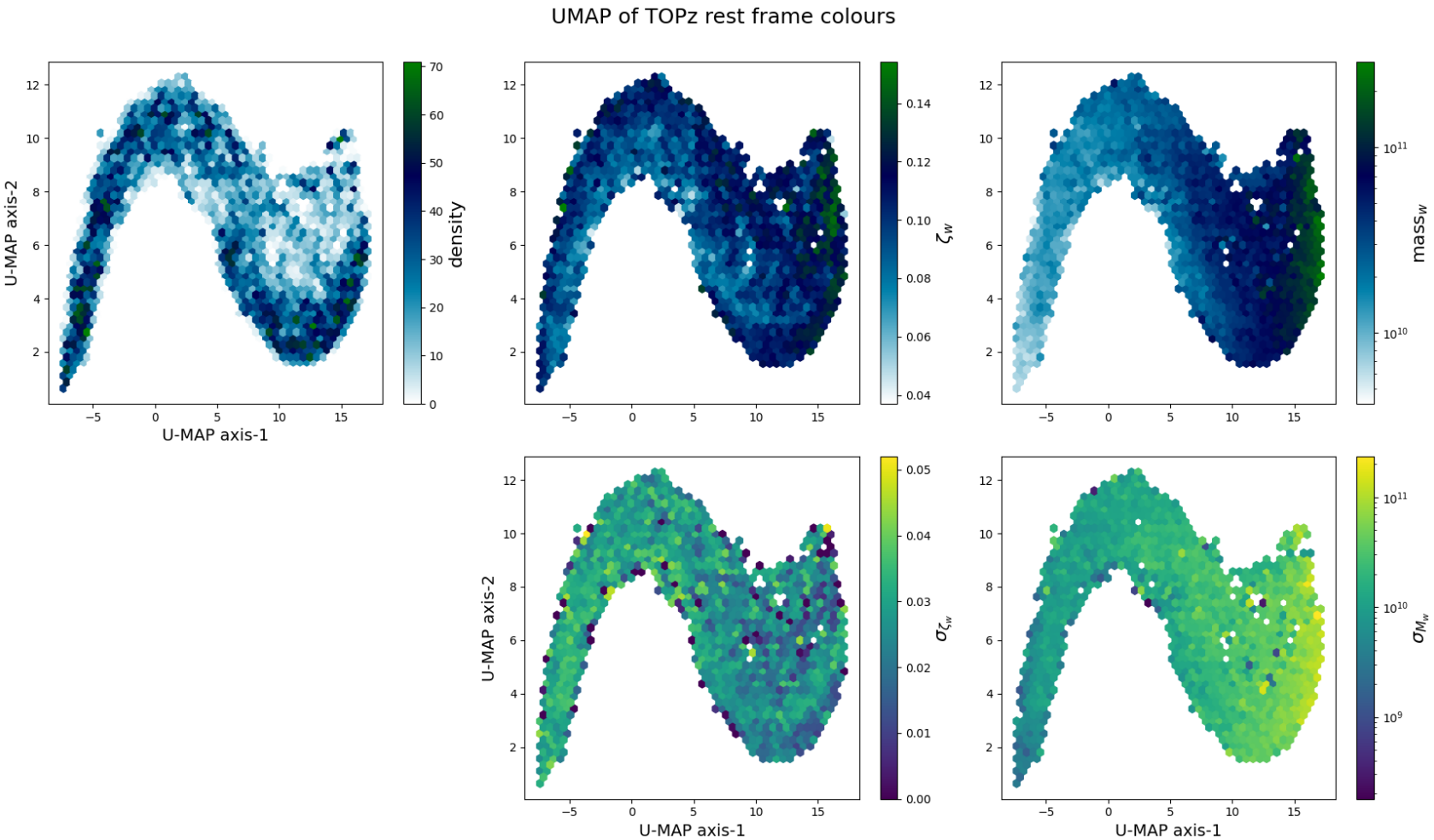


A **Manifold**: generated in the N-dimensional space

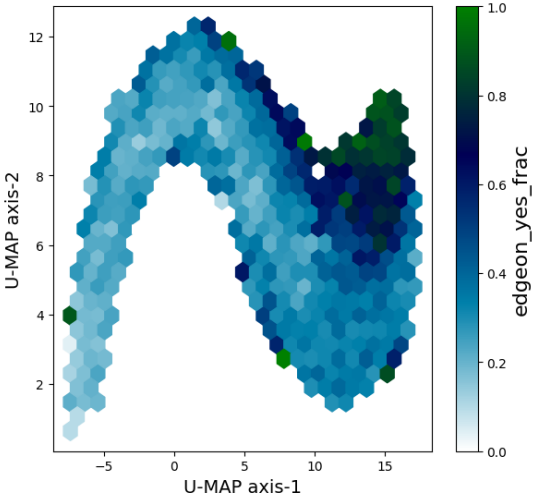
- described as a *fuzzy simplicial complex*

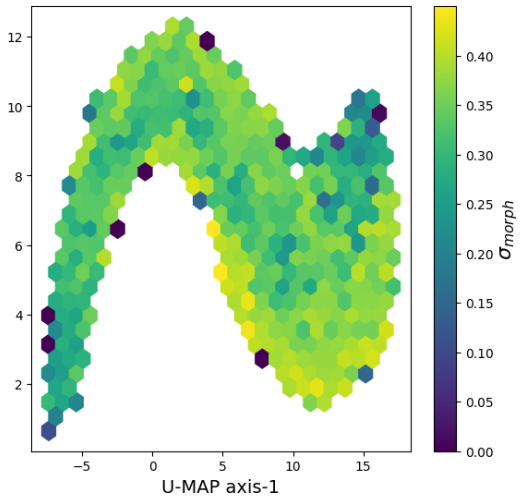
B **Embedding**: fit the 2-dim manifold (according to algorithm parameters)

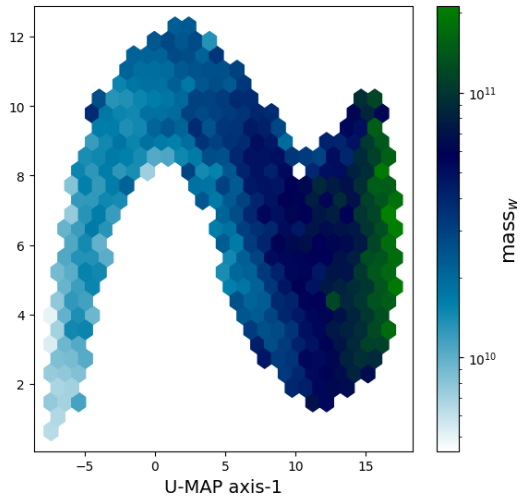
UMAP Rest Frame k-corrected colours for GAMA-Galaxy Zoo Mass-Limited sample



UMAP change with Galaxy Zoo properties



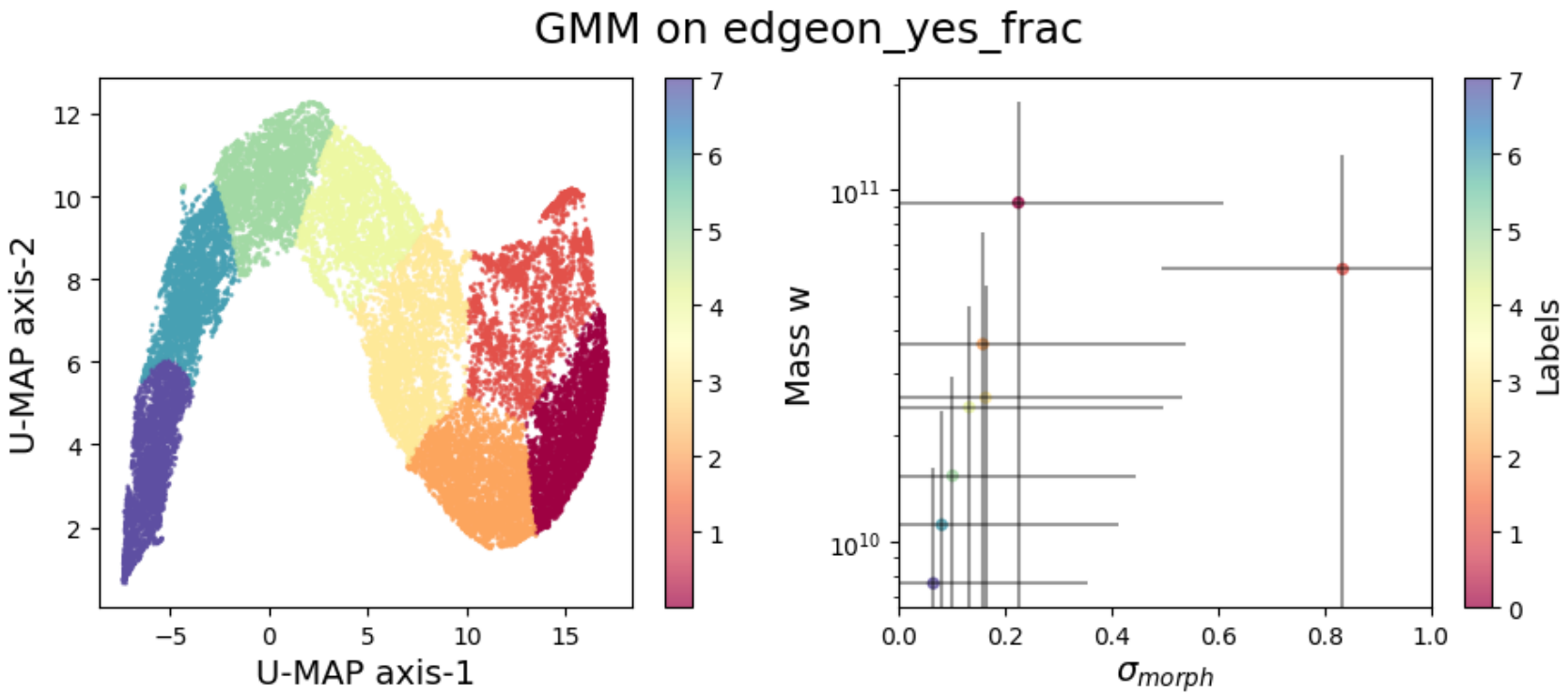




Step 2. Clustering

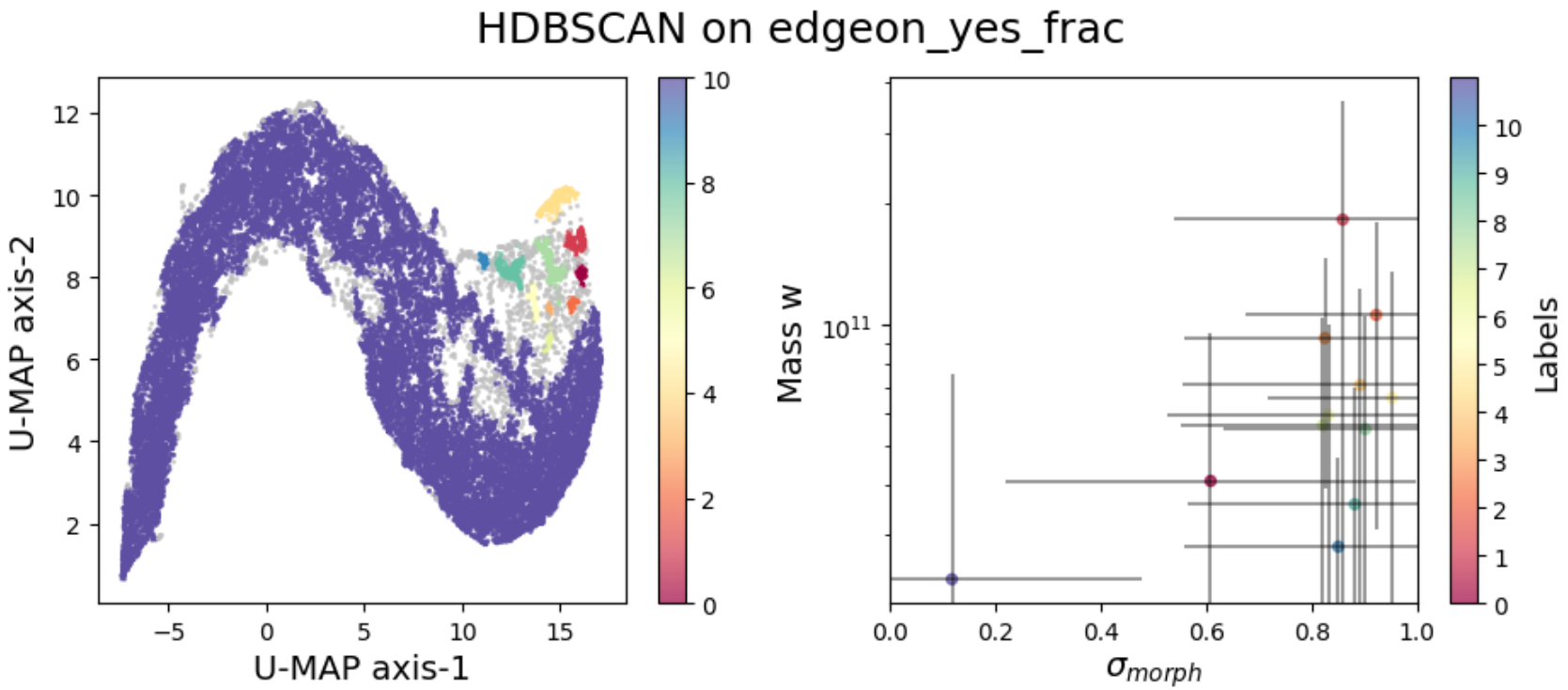
Gaussian Mixture Model (GMM)

Result on UMAP input



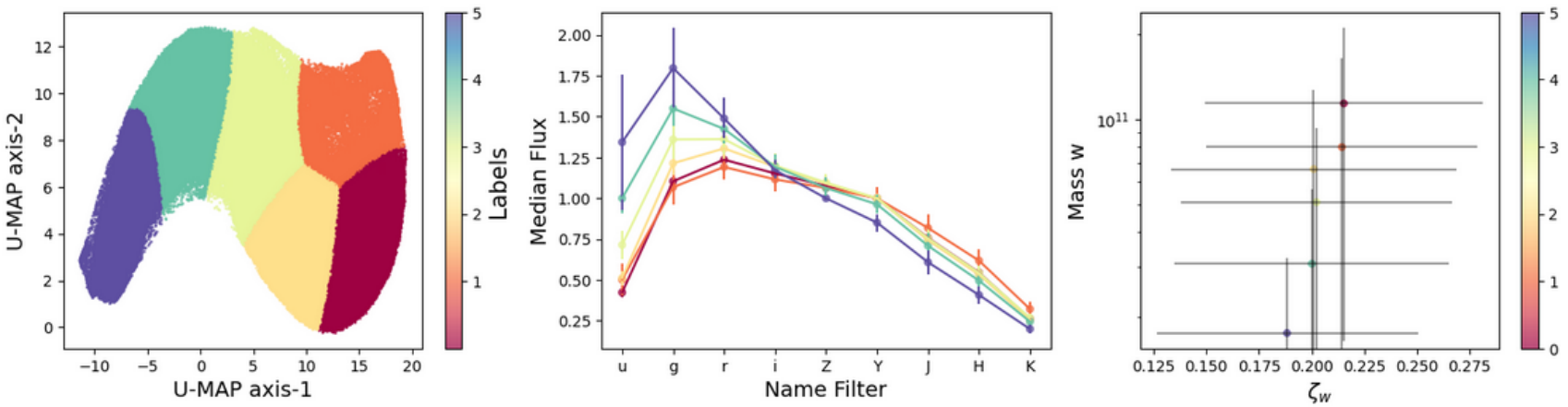
- Number of Clusters: input
- Internal Structure: attempts representing it through gaussian fitting

HDBSCAN Results



- Number of Clusters: according to the **Density**
- Presence of **Noise**: cluster -1
- Internal Structure: according to the **Density**

GMM physical meaning



bigger GAMA sample (without GZ values)

Conclusions

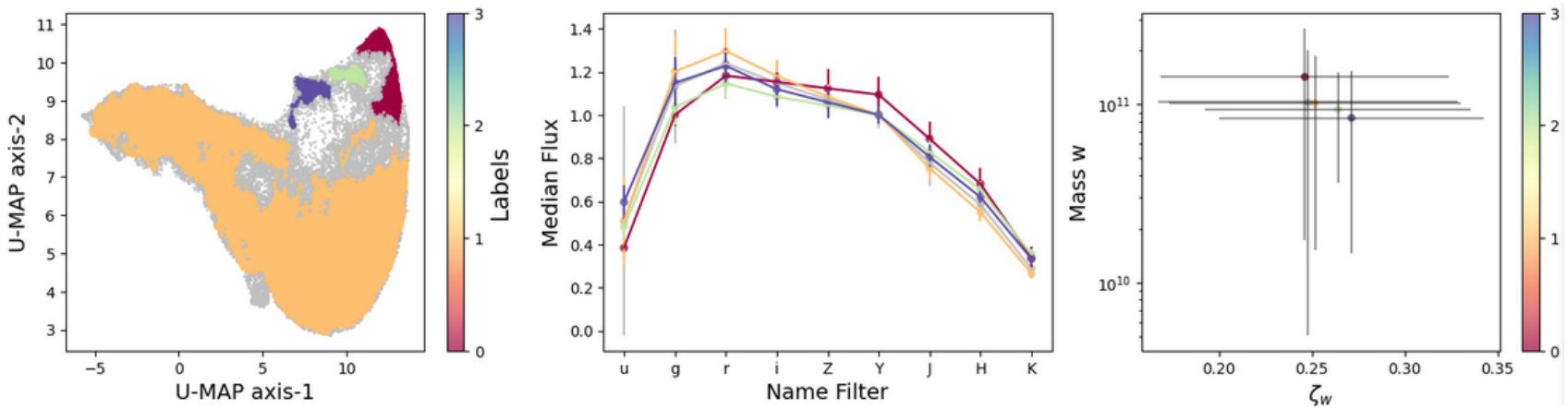
Future Work

- Find consistent cluster generation
- Control the **inherent stochasticity**
- Is it possible to reproduce known classifications?
- or to identify new classifications?

Thanks!

Take-home message

- Clustering on k-corrected colours is independent of redshift
- Mass and Morphology patterns "evolve" across UMAP embedding
- Clustering separates SEDs shapes



Large-scale structure of the Universe: from galaxies to cosmology

30 August – 3 September
2027

Estonian National Museum
Tartu, Estonia

THE COSMIC WEB:

theory
observations
simulations
galaxies
mathematical modelling

SCIENTIFIC ORGANISING COMMITTEE:

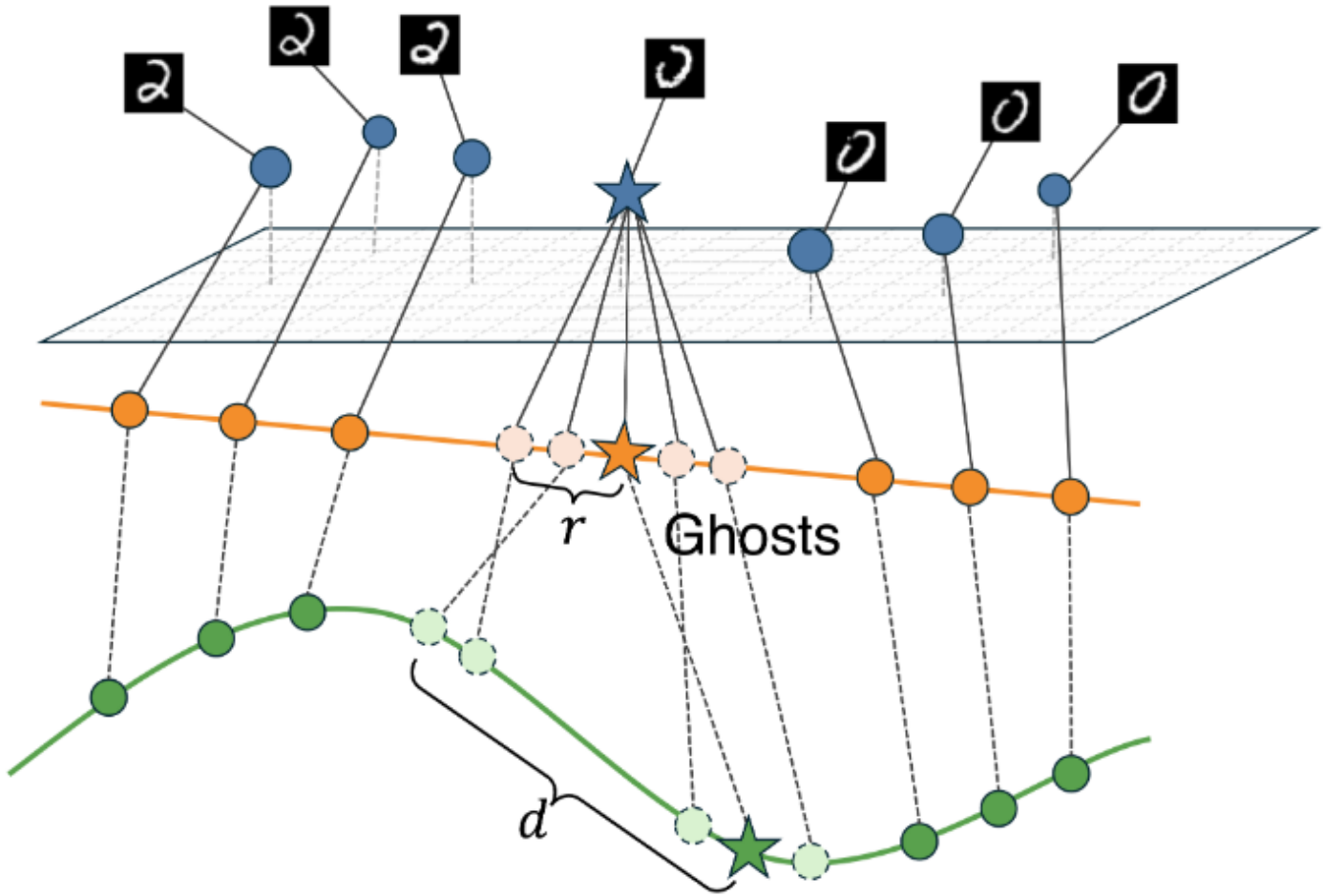
Elmo Tempel
Noam Libeskind
Rien van de Weijgaert
Radu Stoica
Jaan Einasto
Carlos Frenk
Jingjing Shi
Michelle Cluver
Dmitri Pogosyan
Nabila Aghanim

Registration opens in January 2027

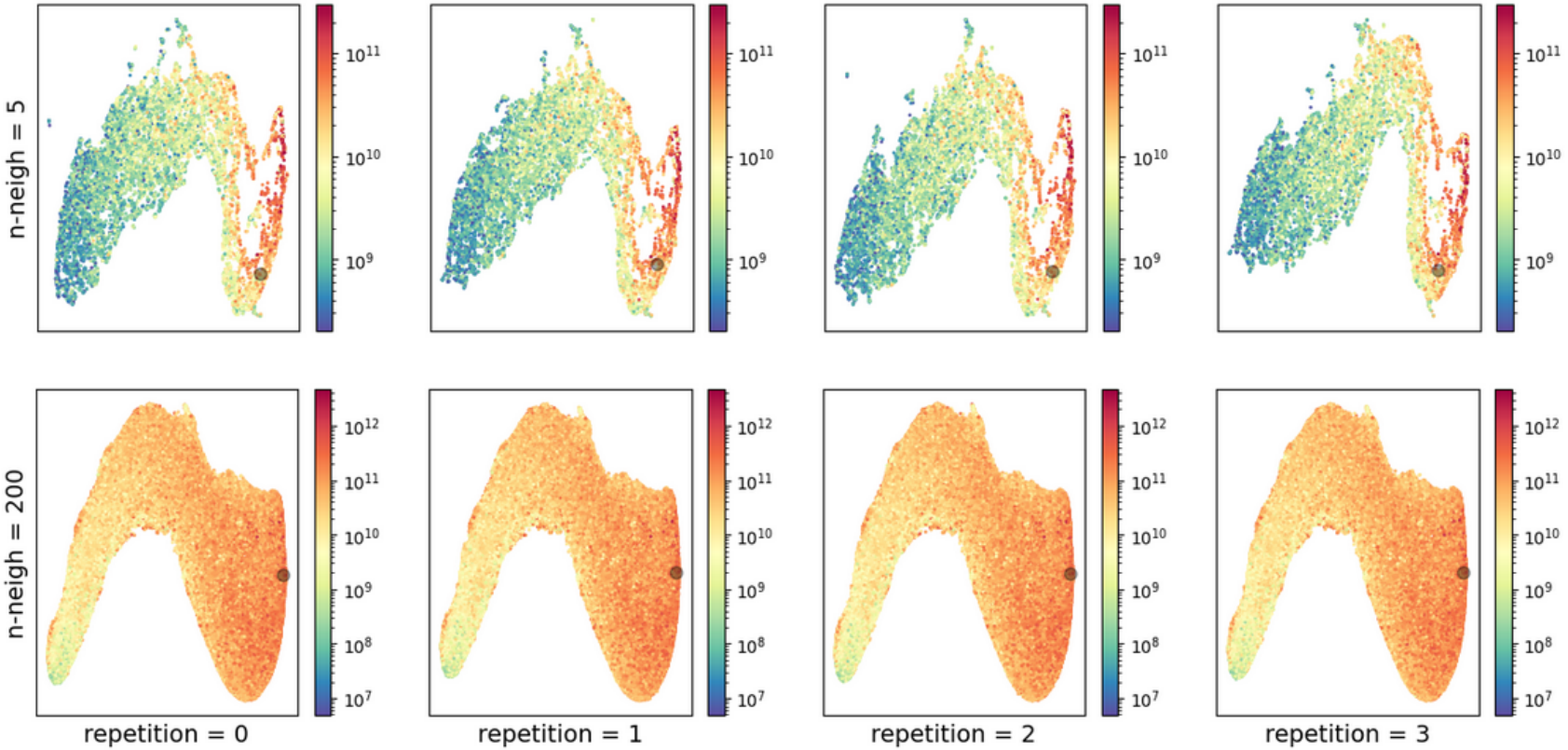
More info: https://excosm2027.ut.ee/

EXCOSM
University of Tartu Tartu Observatory
university of groningen
AIP
Université de Lorraine
Funded by the European Union

Stochasticity I
Appendix 1



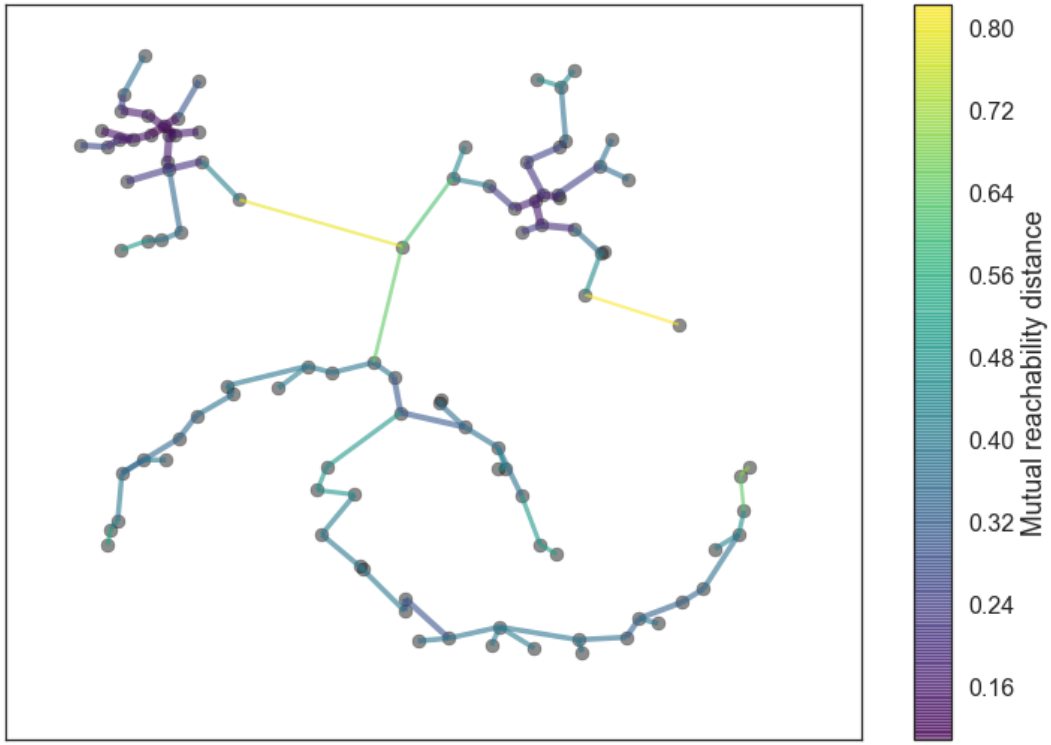
Stochasticity II
Appendix 1



the colour map is the galaxy mass

HDBSCAN I

Appendix 2

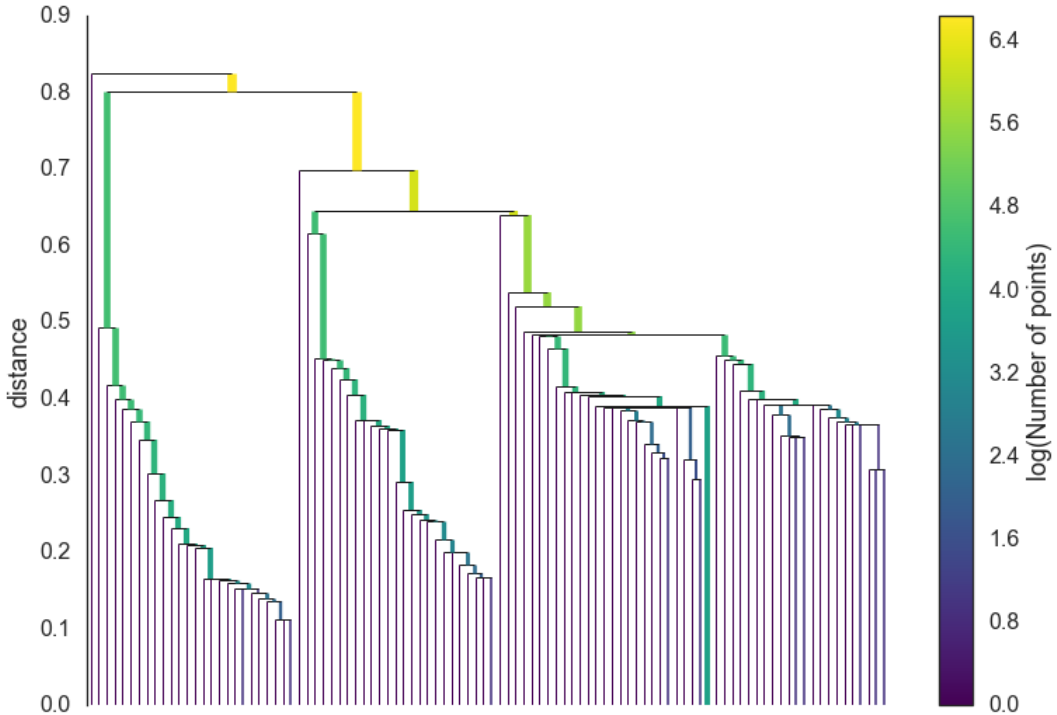


hdbscan.readthedocs.io

A (Mutual Reachability) Distance for each galaxy + Single linkage "skeleton" (Minimum Spanning Tree)

HDBSCAN II

Appendix 2

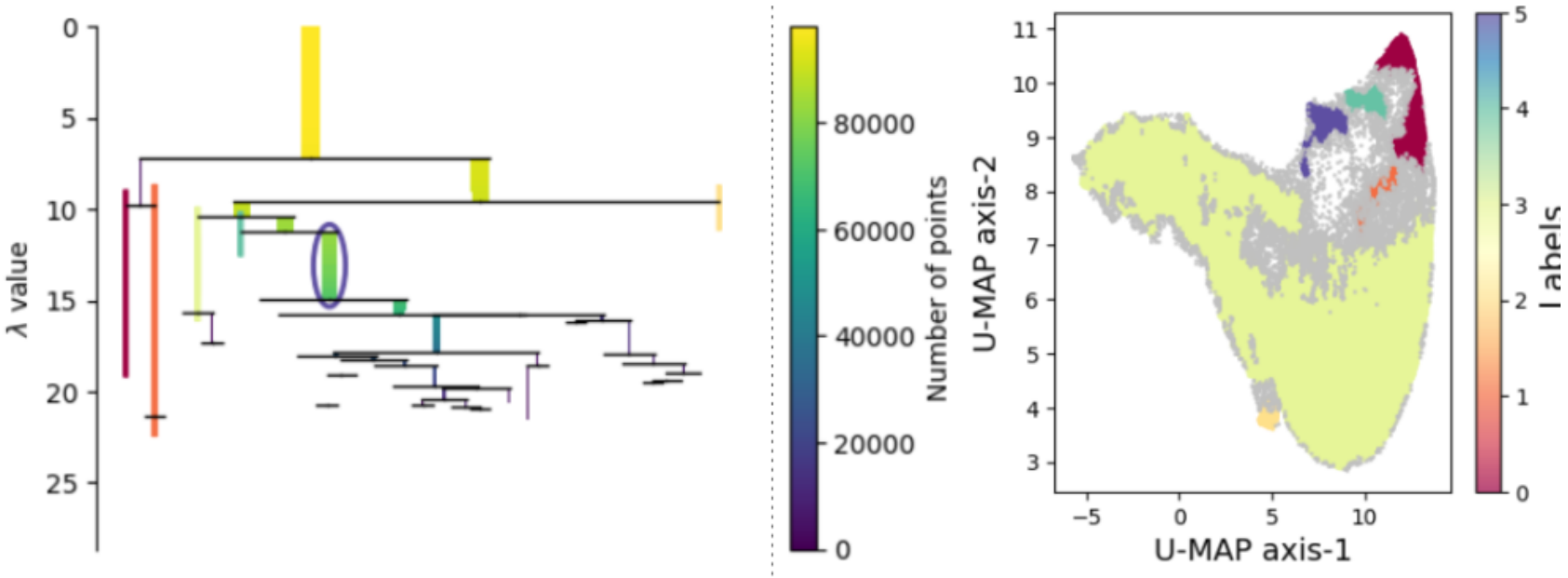


hdbscan.readthedocs.io

B Generate the cluster tree, cutting links in decreasing order

HDBSCAN III

Appendix 2



C Condense and select stable clusters

Noise = single-point clusters

HDBSCAN Results 2

Appendix 3