Synergies between low- and intermediate-redshift galaxy classifications via unsupervised machine learning

Seb Turner, Malgorzata Siudek,

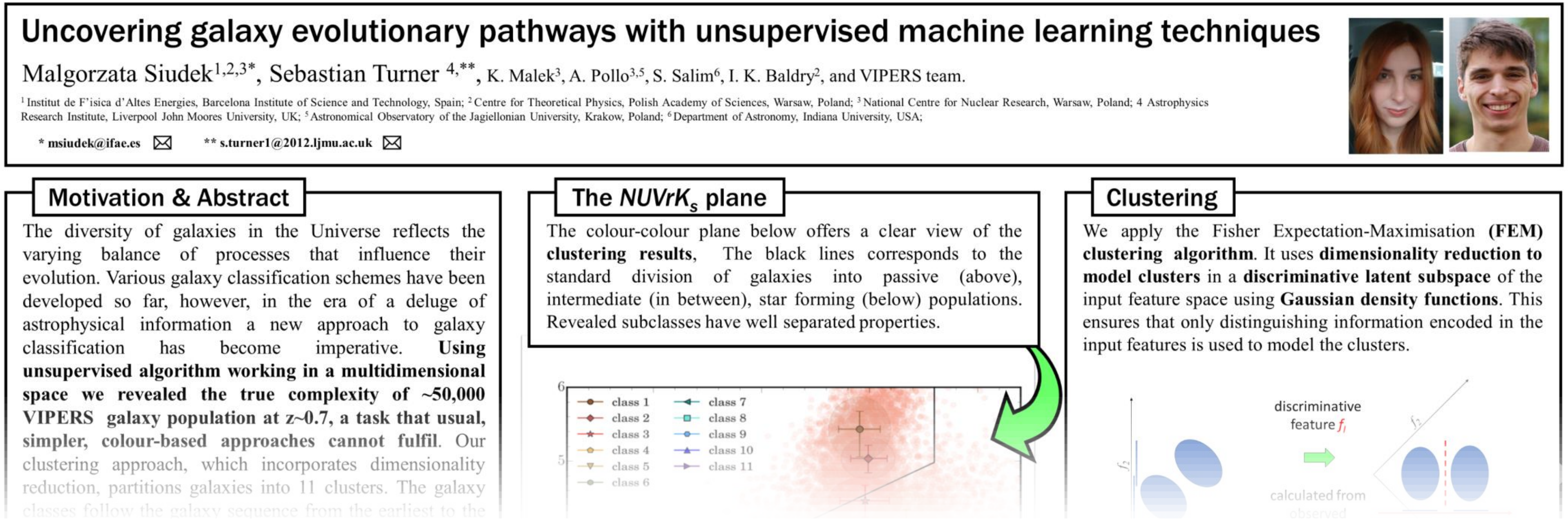
Agnieszka Pollo, Samir Salim, Kasia Malek, Ivan Baldry, et al.

LIVERPOOL JOHN MOORES UNIVERSITY

s.turner1@2012.ljmu.ac.uk

@sebturne

Poster



Uncovering galaxy evolutionary pathways with unsupervised machine learning techniques

Malgorzata Siudek[1,2,3*], Sebastian Turner [4,**], K. Malek[3], A. Pollo[3,5], S. Salim[6], I. K. Baldry[2], and VIPERS team.

[1] Institut de F'isica d'Altes Energies, Barcelona Institute of Science and Technology, Spain; [2] Centre for Theoretical Physics, Polish Academy of Sciences, Warsaw, Poland; [3] National Centre for Nuclear Research, Warsaw, Poland; 4 Astrophysics Research Institute, Liverpool John Moores University, UK; [5] Astronomical Observatory of the Jagiellonian University, Krakow, Poland; [6] Department of Astronomy, Indiana University, USA;

* msiudek@ifae.es ** s.turner1@2012.ljmu.ac.uk

Motivation & Abstract

The diversity of galaxies in the Universe reflects the varying balance of processes that influence their evolution. Various galaxy classification schemes have been developed so far, however, in the era of a deluge of astrophysical information a new approach to galaxy classification has become imperative. **Using unsupervised algorithm working in a multidimensional space we revealed the true complexity of ~50,000 VIPERS galaxy population at z~0.7, a task that usual, simpler, colour-based approaches cannot fulfil.** Our clustering approach, which incorporates dimensionality reduction, partitions galaxies into 11 clusters. The galaxy classes follow the galaxy sequence from the earliest to the

The *NUVrK*$_s$ plane

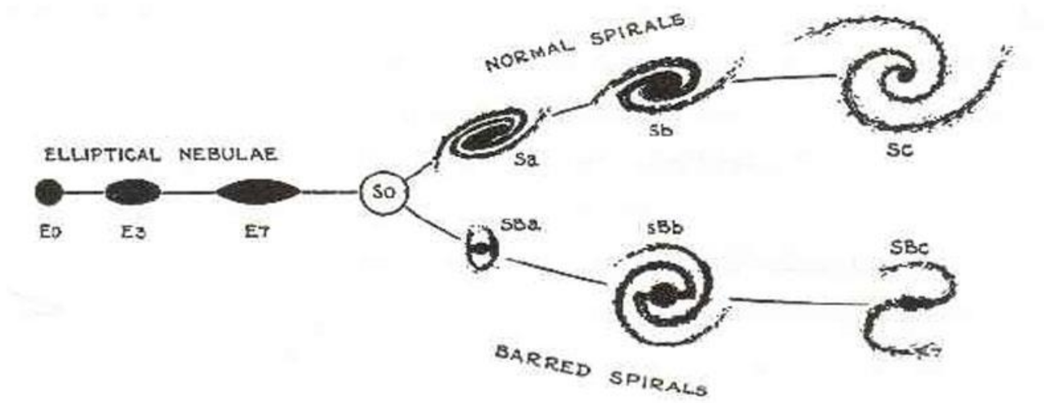
The colour-colour plane below offers a clear view of the **clustering results**. The black lines corresponds to the standard division of galaxies into passive (above), intermediate (in between), star forming (below) populations. Revealed subclasses have well separated properties.

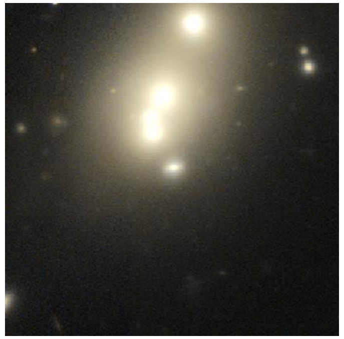
Clustering

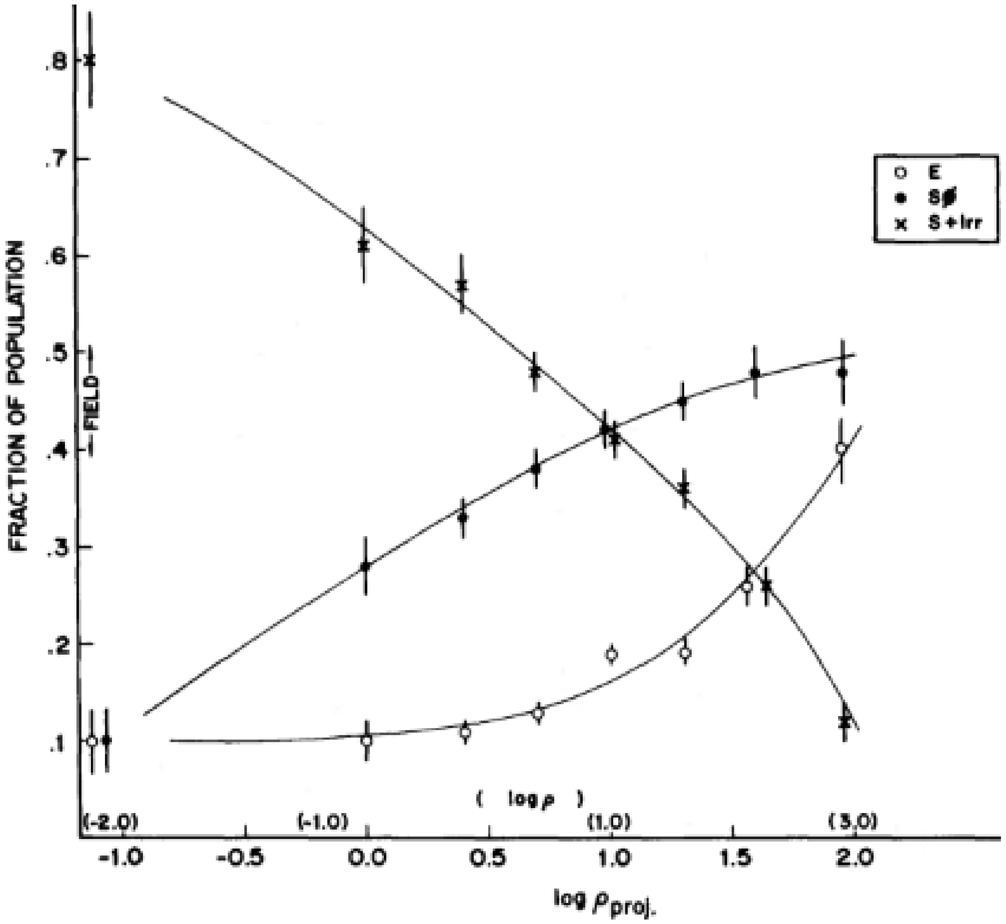
We apply the Fisher Expectation-Maximisation **(FEM) clustering algorithm**. It uses **dimensionality reduction to model clusters** in a **discriminative latent subspace** of the input feature space using **Gaussian density functions**. This ensures that only distinguishing information encoded in the input features is used to model the clusters.

astro.ljmu.ac.uk/~aststurn/poster.pdf

Galaxy classifications: background

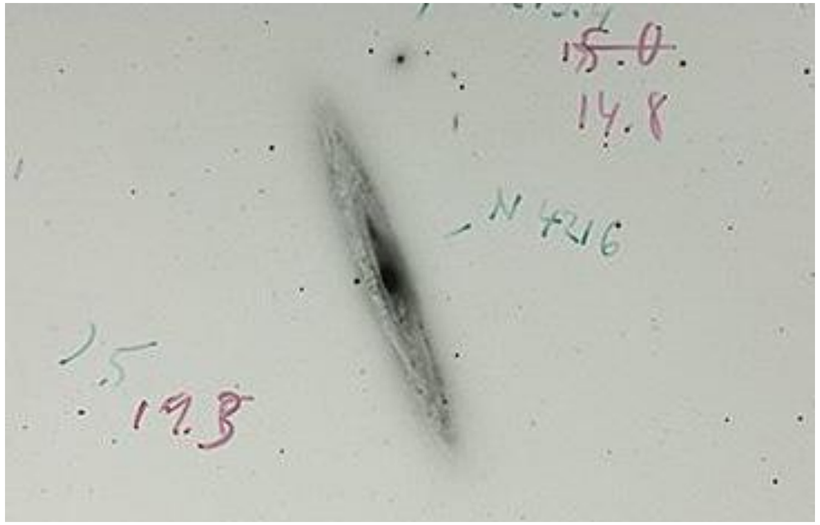




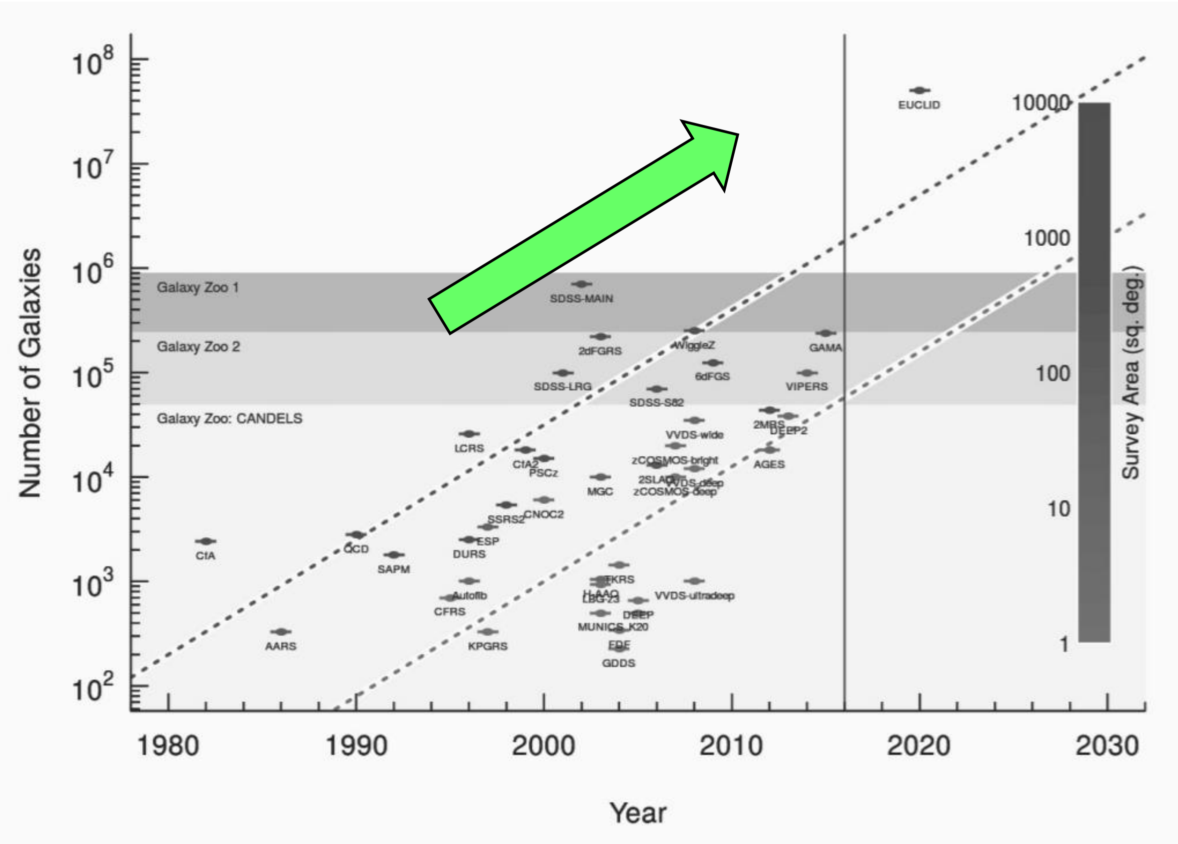


Hubble-1936; Dressler-1980

Galaxy samples: bigger and bigger



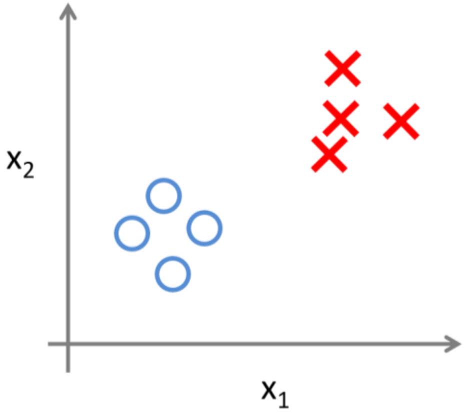
- Traditional: ~$10^{2\text{-}4}$
- Crowdsourced: ~$10^{5\text{-}6}$
- Automated (future): >10^{7}!!



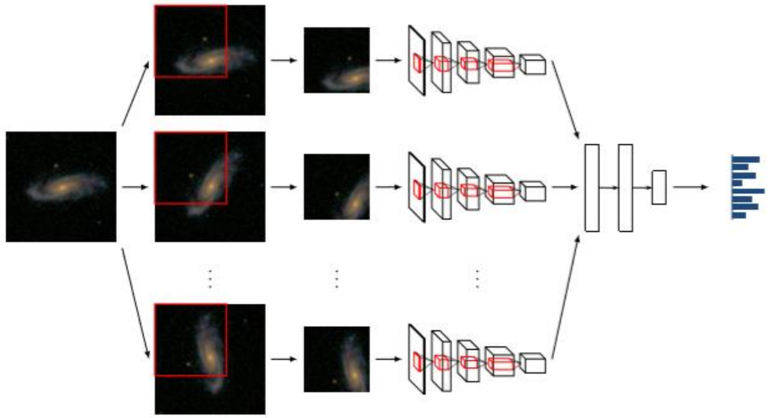
nautil.us/issue/32; galaxyzoo.org; ESA, 2009; Lee Kelvin

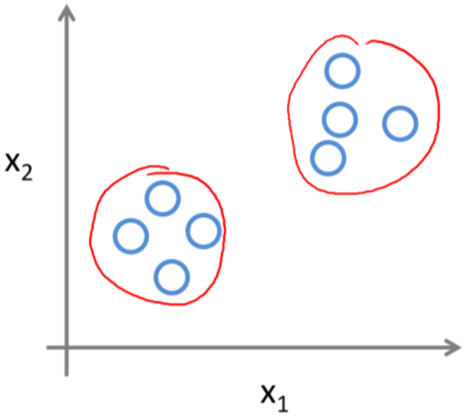
Seb Turner, LJMU

Machine learning

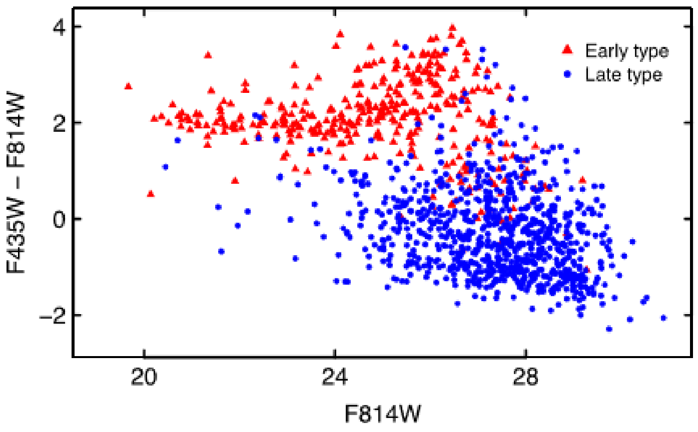


Supervised



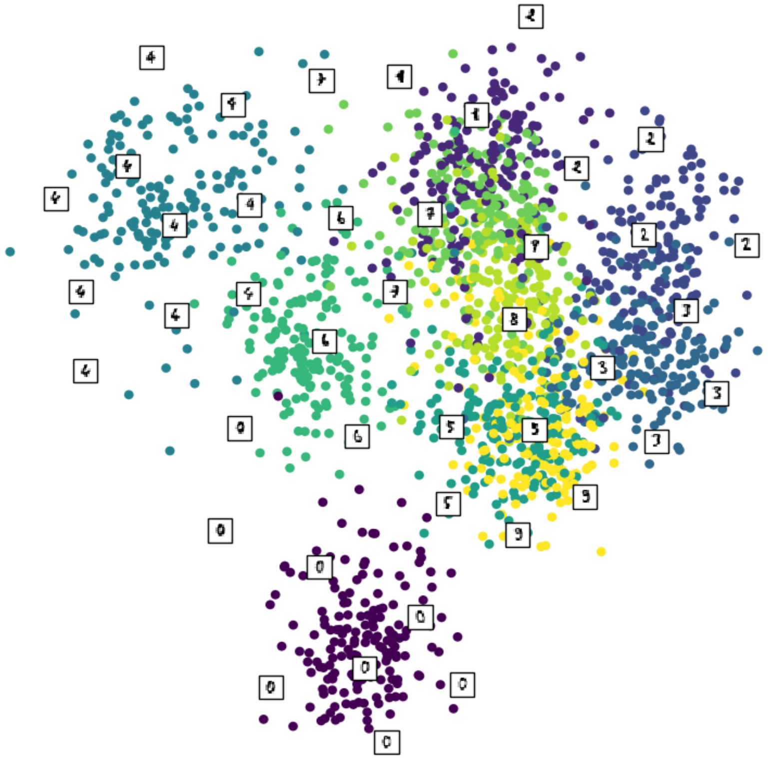


Unsupervised



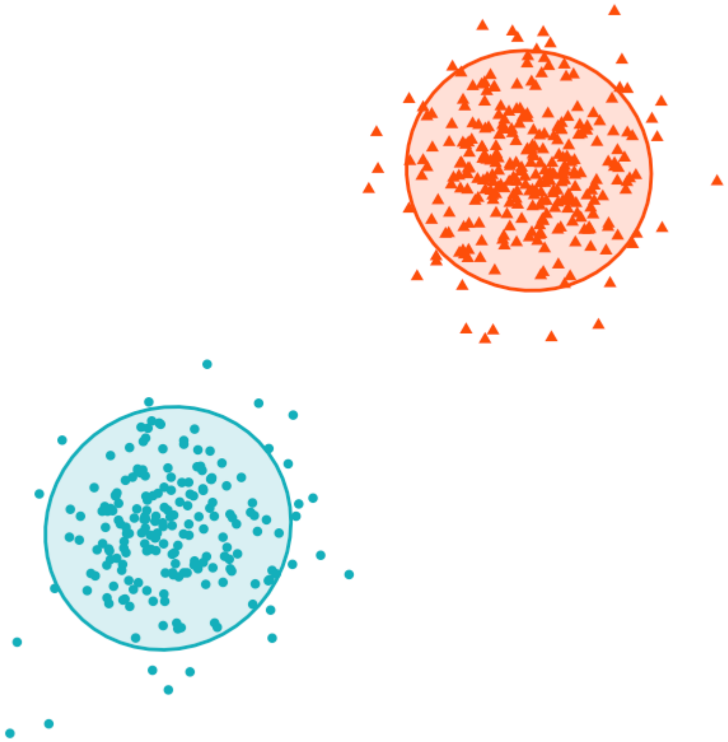
dosSantos-2017, gitbook.com; Dieleman+2015; Hocking+2018

Fisher Expectation-Maximisation



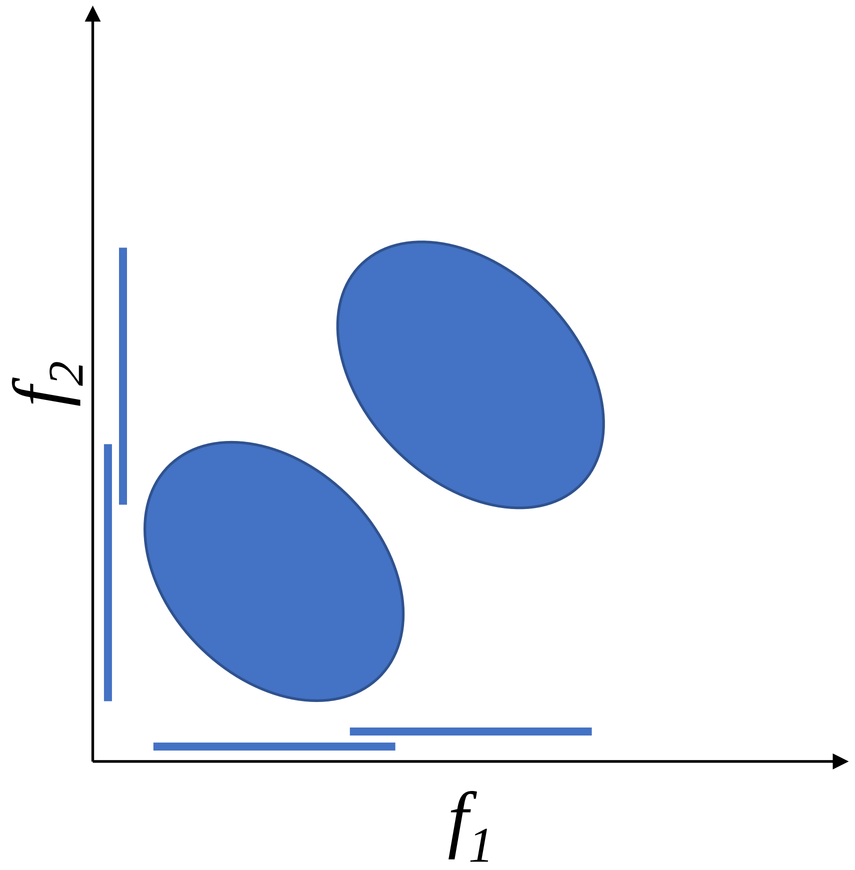
Dimensionality
Reduction

+



Clustering

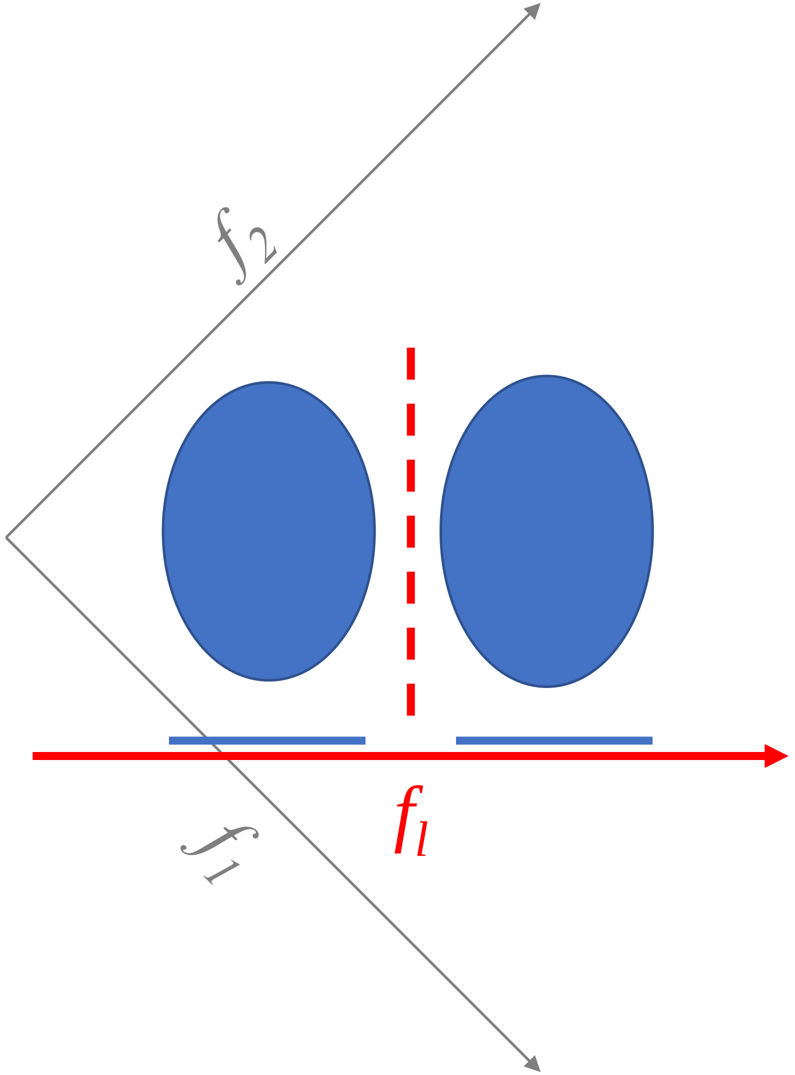
Fisher Expectation-Maximisation



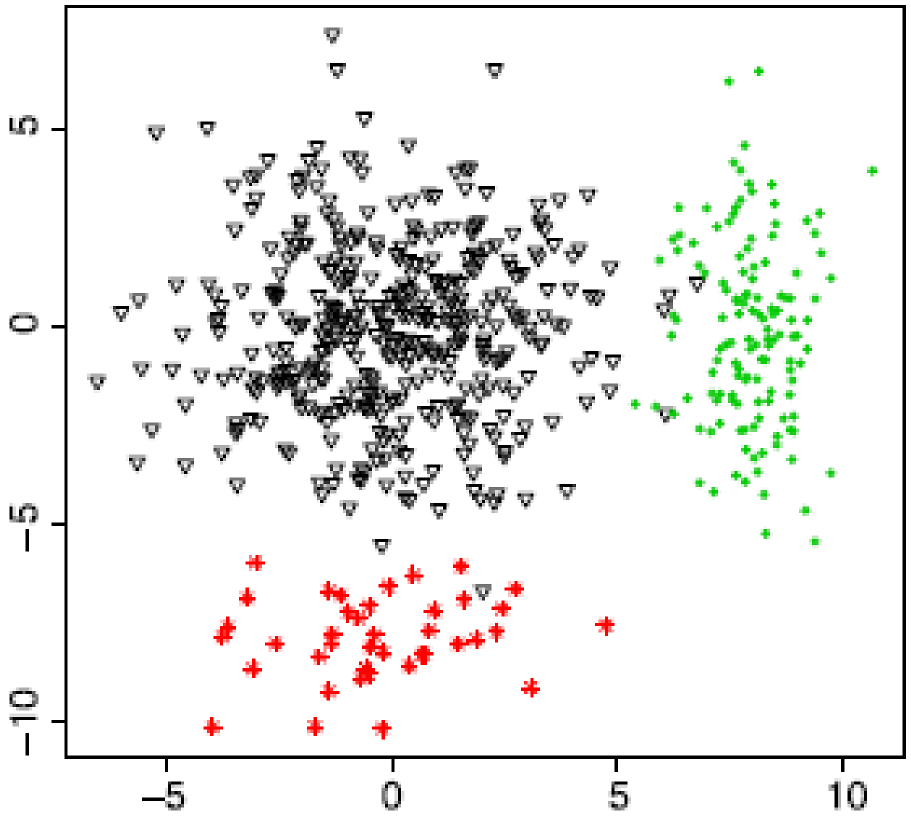
U

orientation matrix

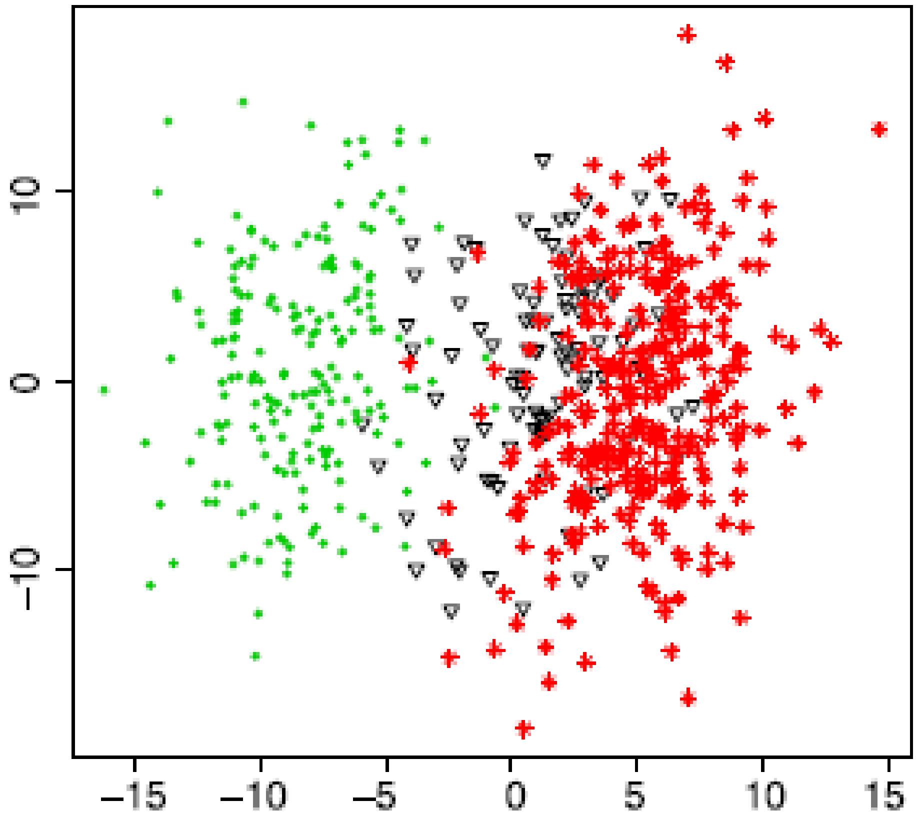
determines discriminative latent subspace



Subspaces

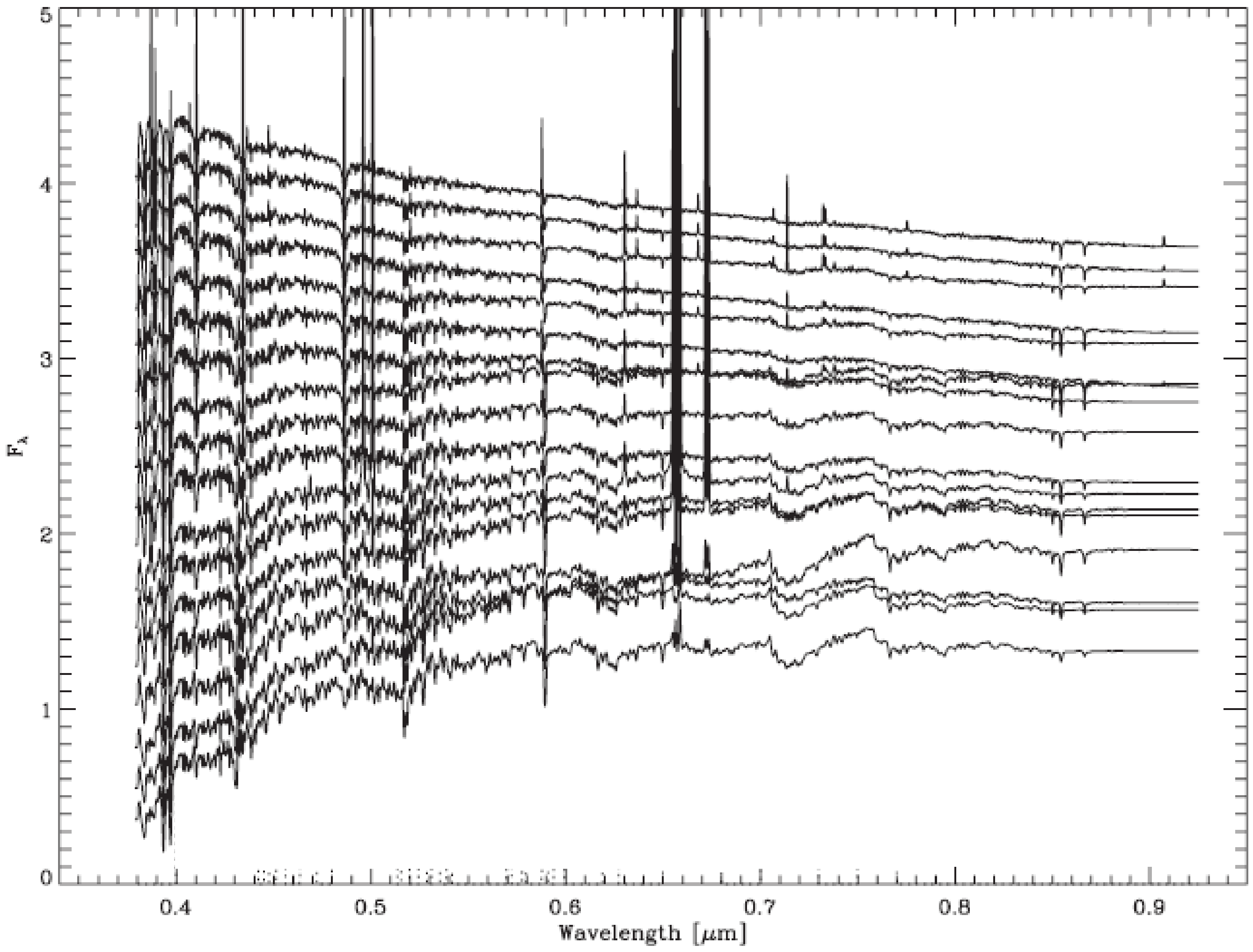


FEM subspace



PC subspace

Why do we care about subspaces?



SanchezAlmeida+2010

FEM steps

E

Map clusters onto data in original space (/ initialise randomly)

F

Determine subspace which best separates clusters

M

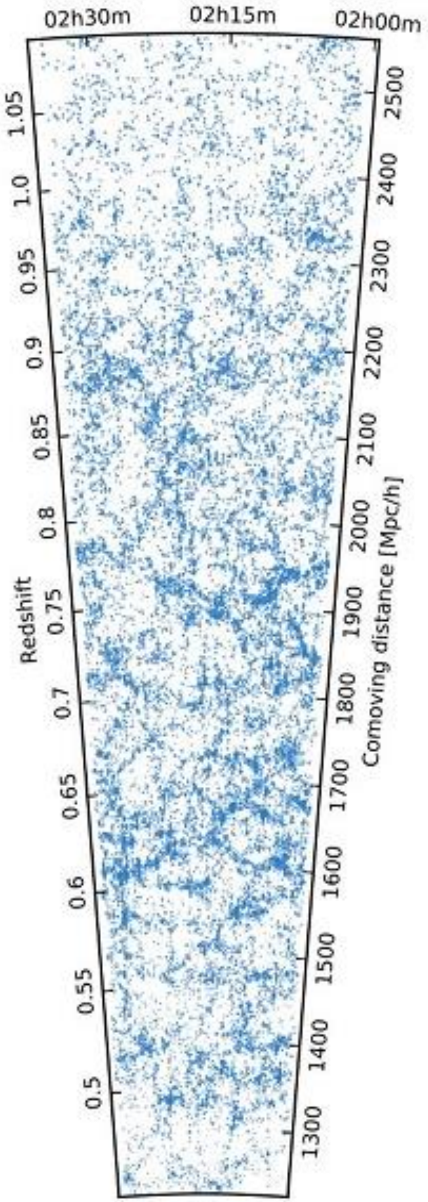
Find new clusters in subspace using Gaussians

Break cycle when algorithm converges: get final clusters!

VIPERS sample

- ~50000 galaxies (of ~100000 in PDR2)
- Magnitude limited i_{AB} < 22.5, redshifts 0.5 to 1.2
- Comparable volume to SDSS, but higher redshift

- Features for clustering:
  - FUV, NUV, u, g, r, i, z, B, V, J, H, and K_s band abs mags from LePhare SED fits to UV to IR photometry
  - Spectroscopic redshifts, to include potential cosmic evolution of galaxies

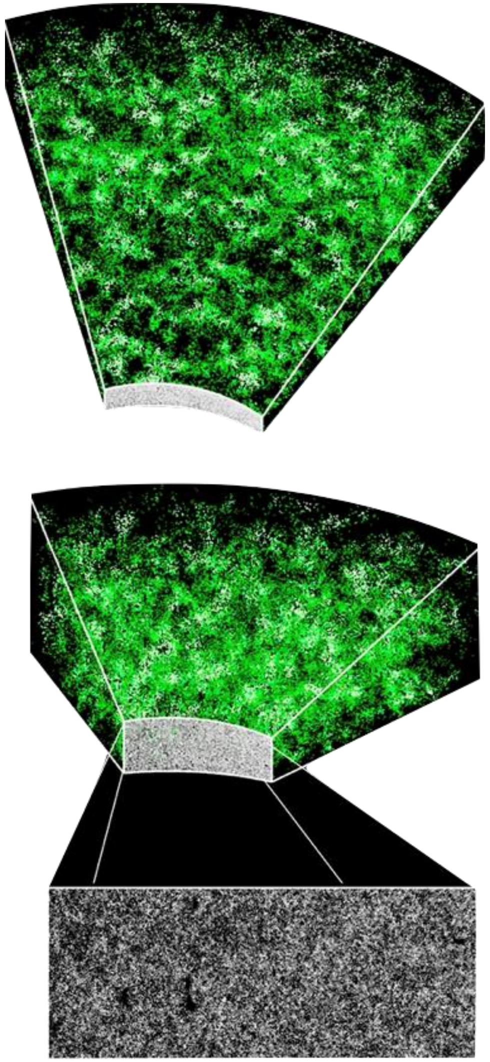


Scodeggio+2018; Guzzo+2014; Garilli+2014; Siudek+2018

GSWLC sample

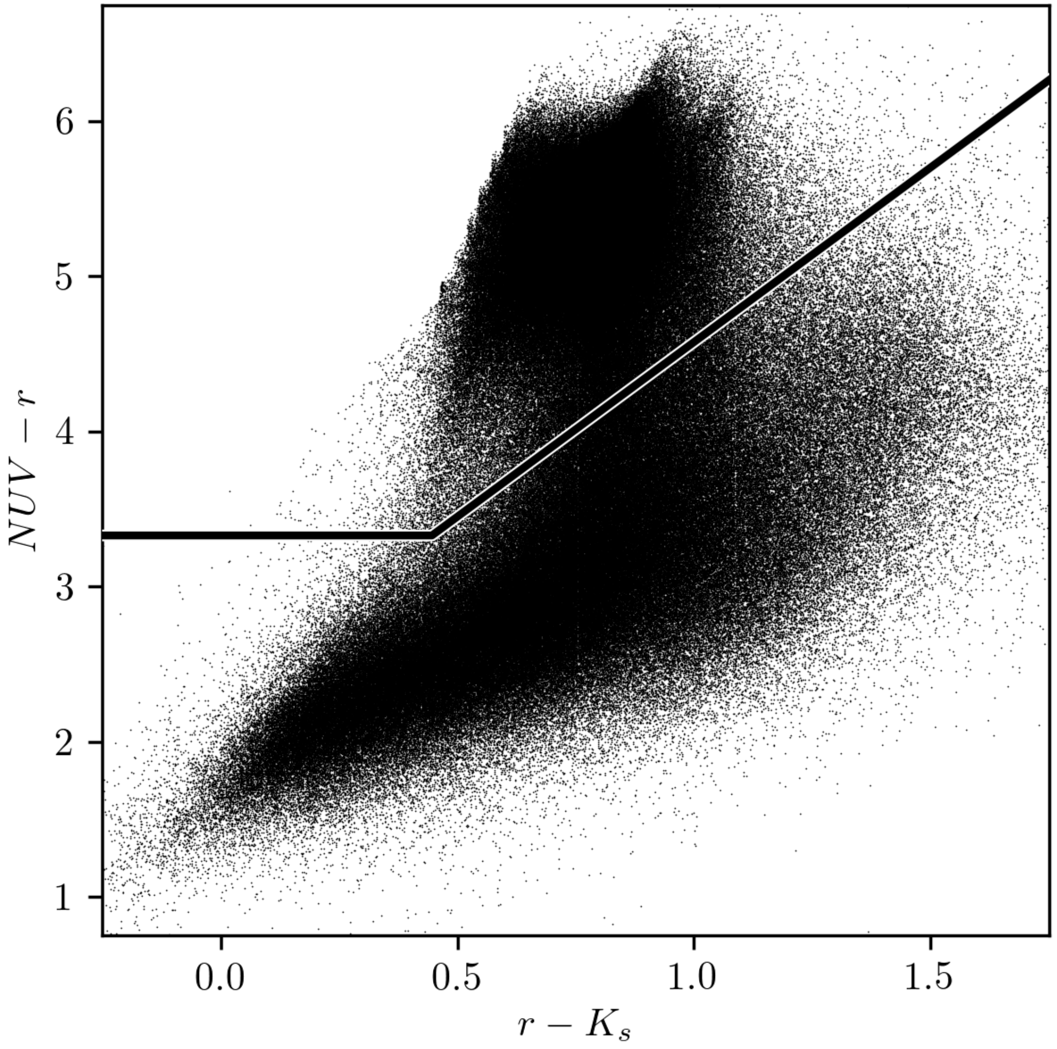
- ~600000 galaxies – full GSWLCX-2.1
- Magnitude limited r_{PETRO} < 19.8, redshifts < 0.3
- Contains 90% of SDSS

- Same features used as for VIPERS sample:
  - Abs mags estimated from CIGALE SED fits to UV to optical photometry
  - IR photometry only partially available: extrapolated IR abs mags in order to retain SDSS selection

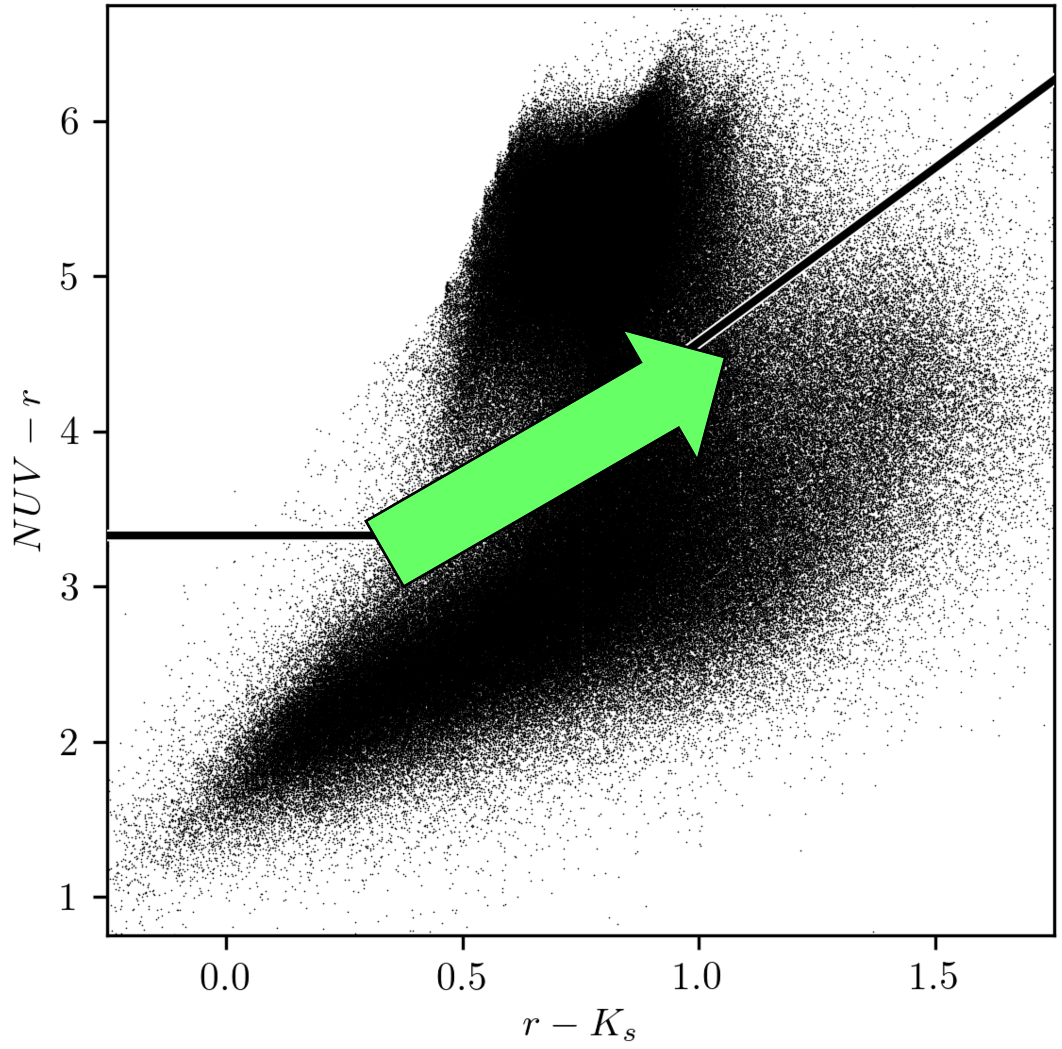


Salim+2016; Salim+2018

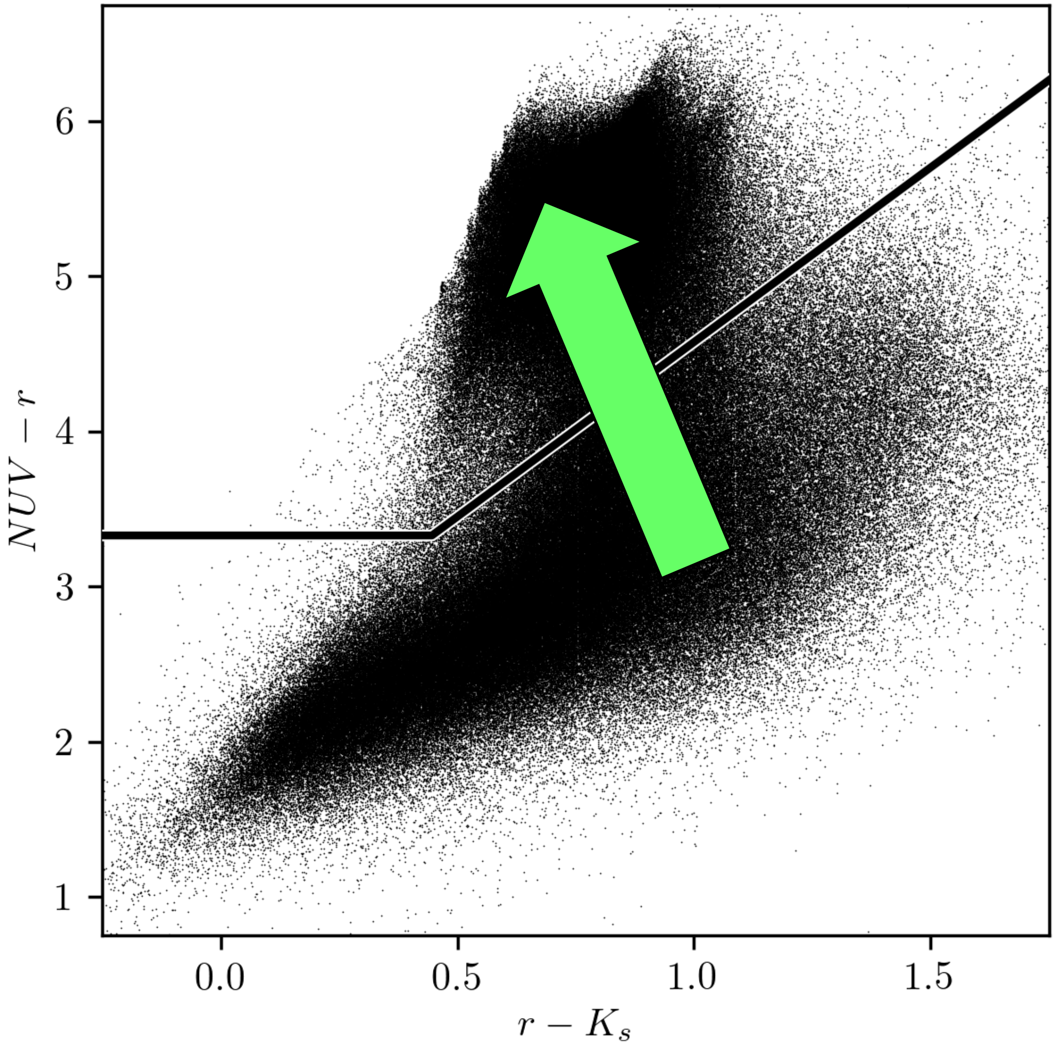
GSWLC: NUVrK diagram



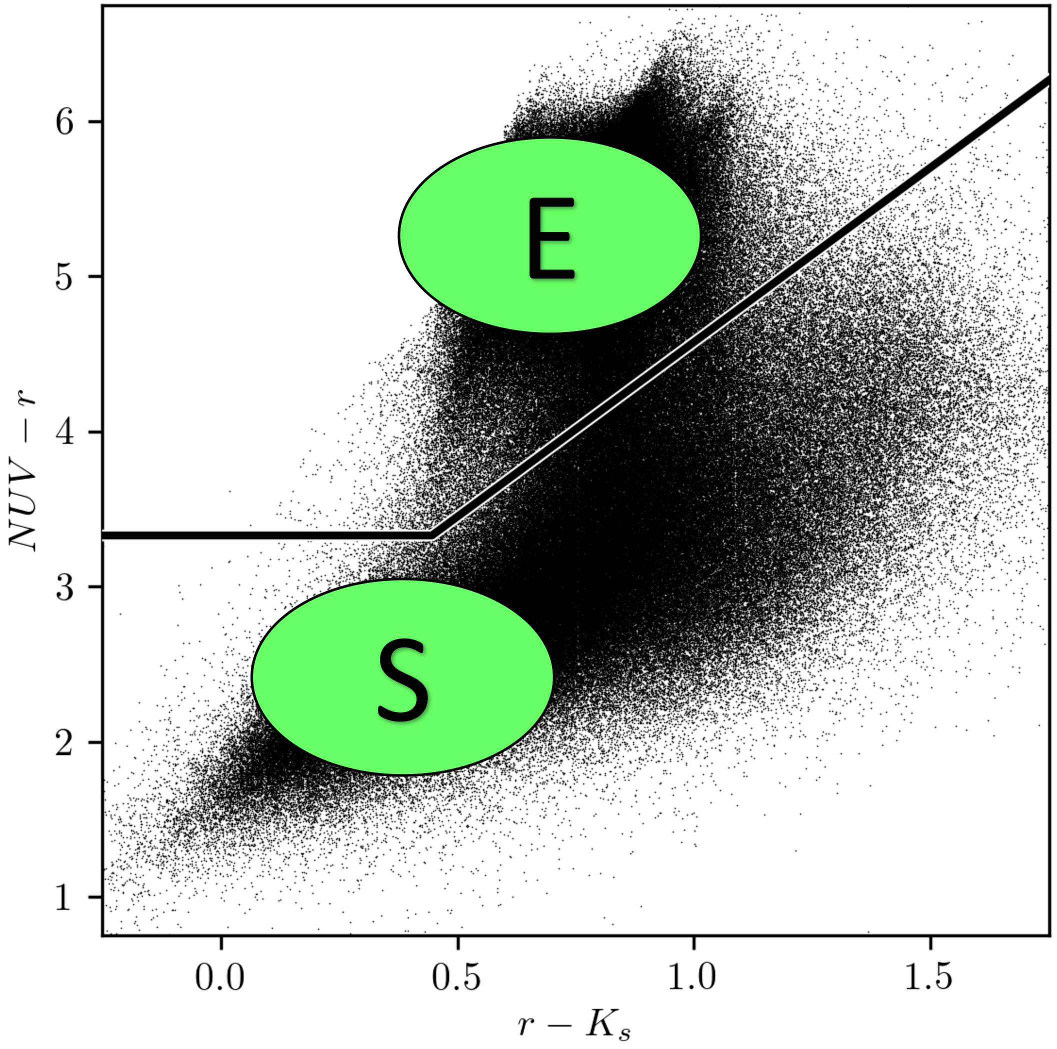
GSWLC: NUVrK diagram



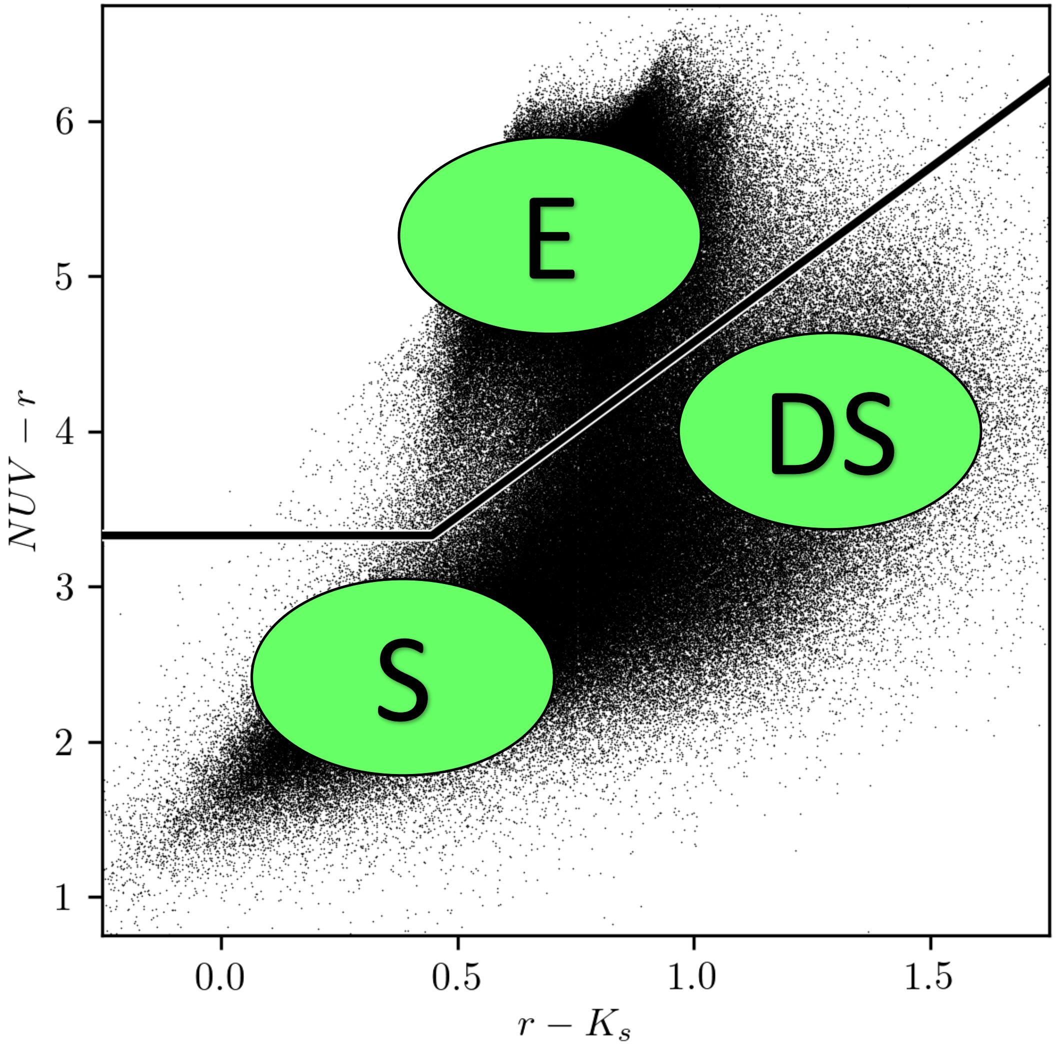
GSWLC: NUVrK diagram



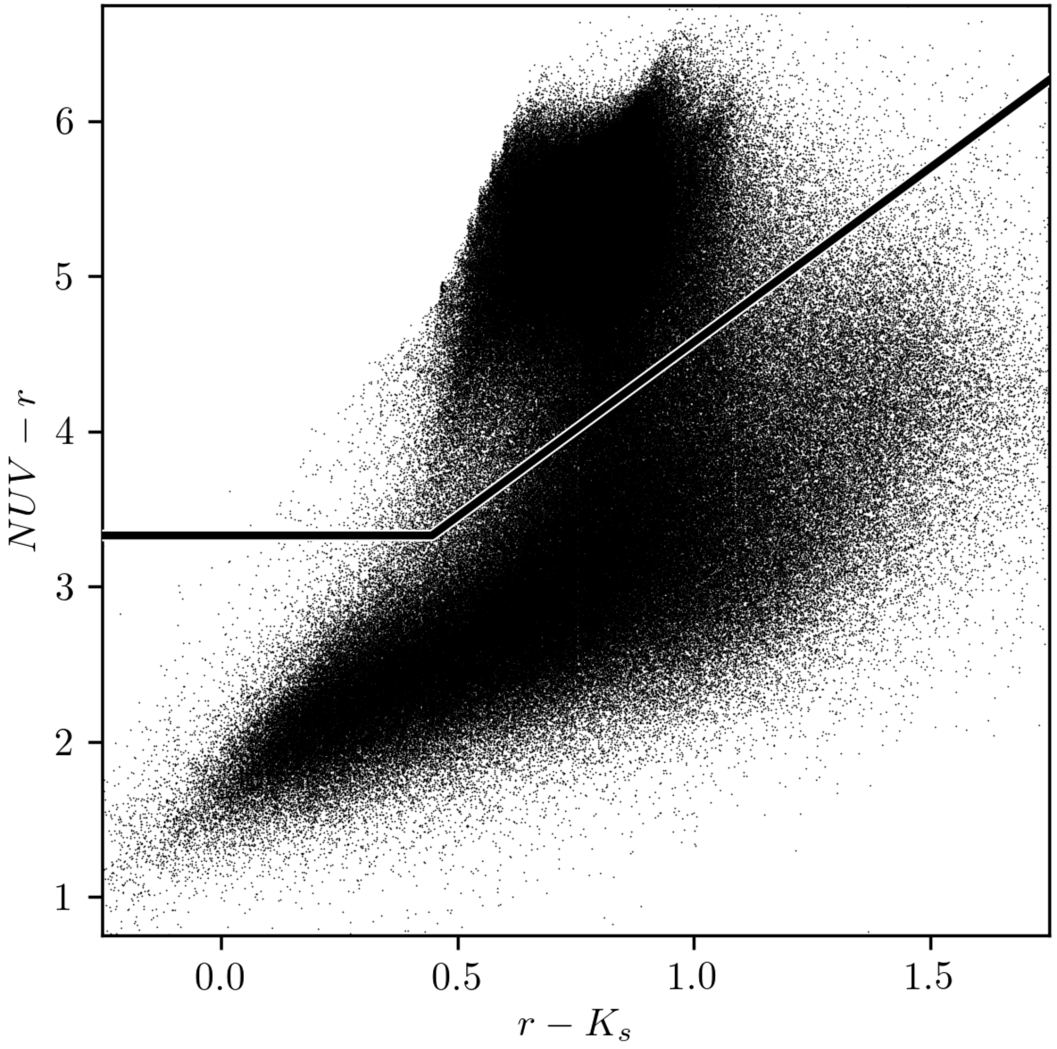
GSWLC: NUVrK diagram



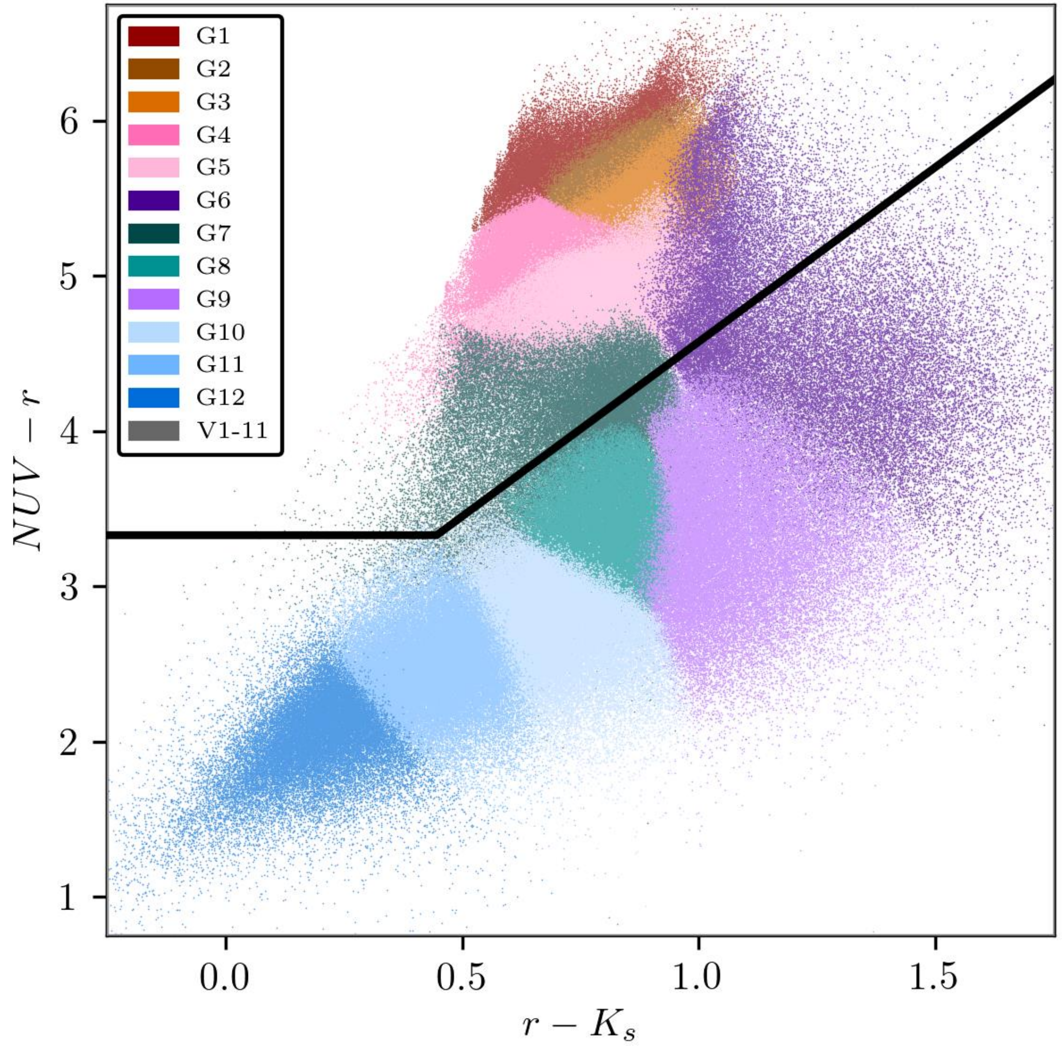
GSWLC: NUVrK diagram



GSWLC: NUVrK diagram

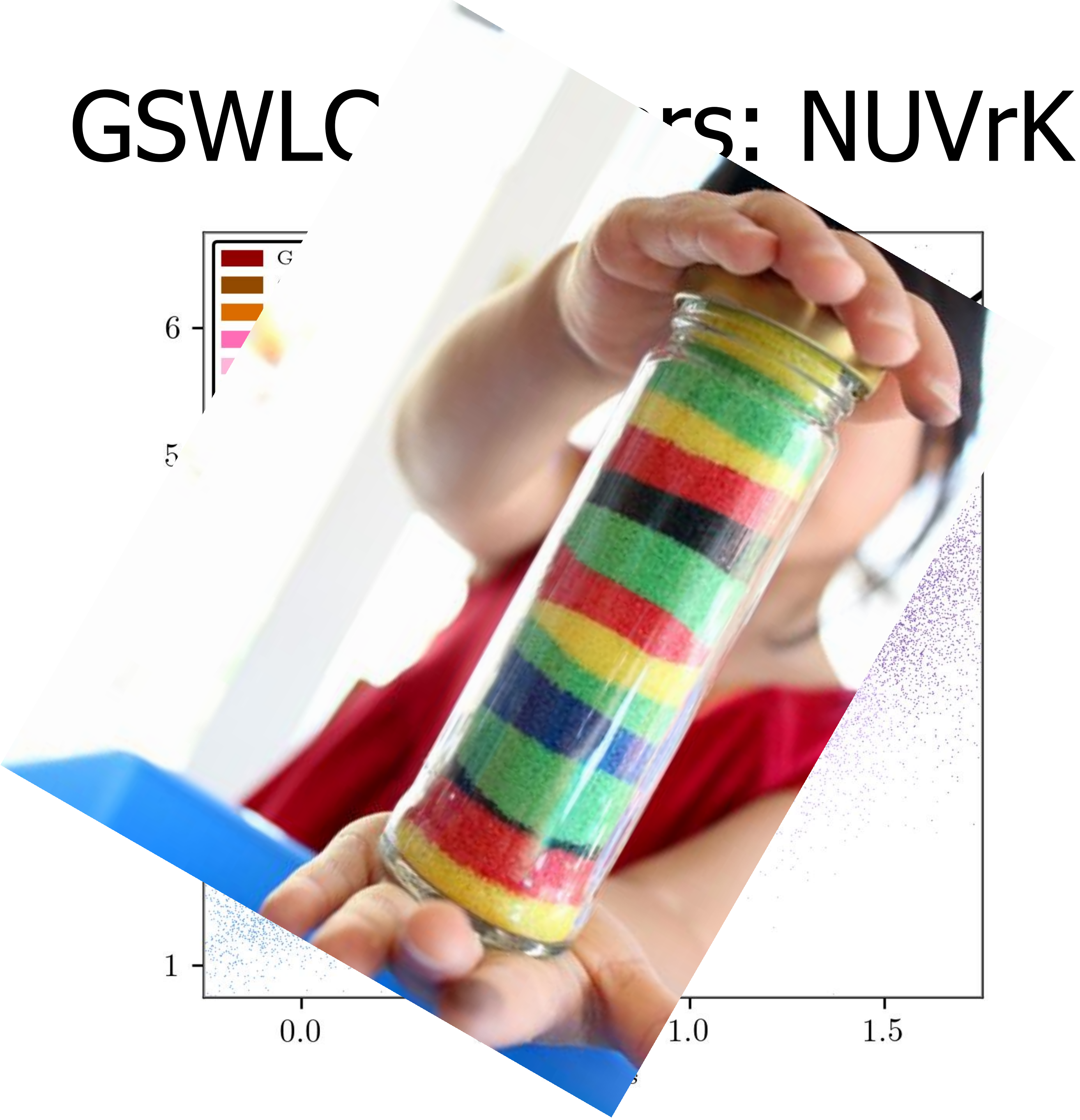


GSWLC clusters: NUVrK

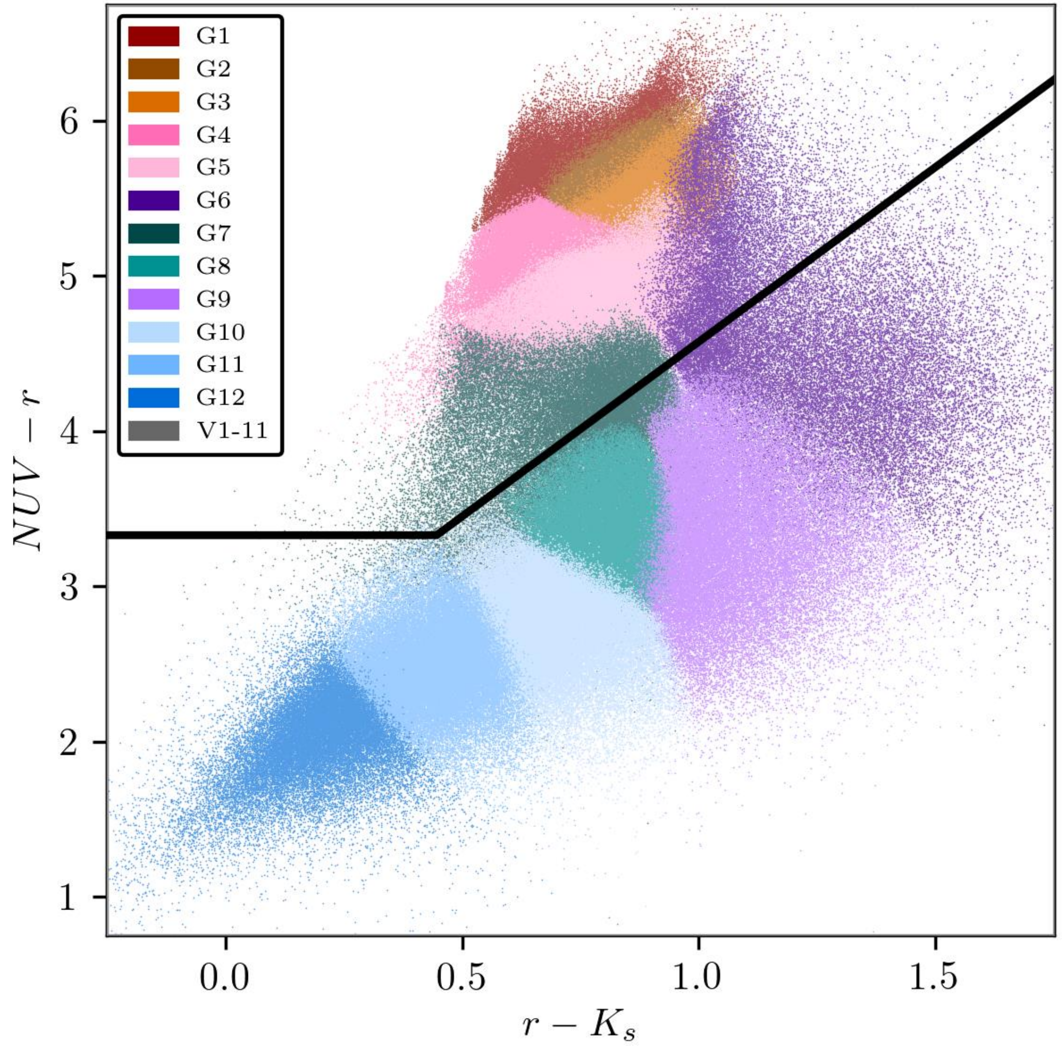


Seb Turner, LJMU

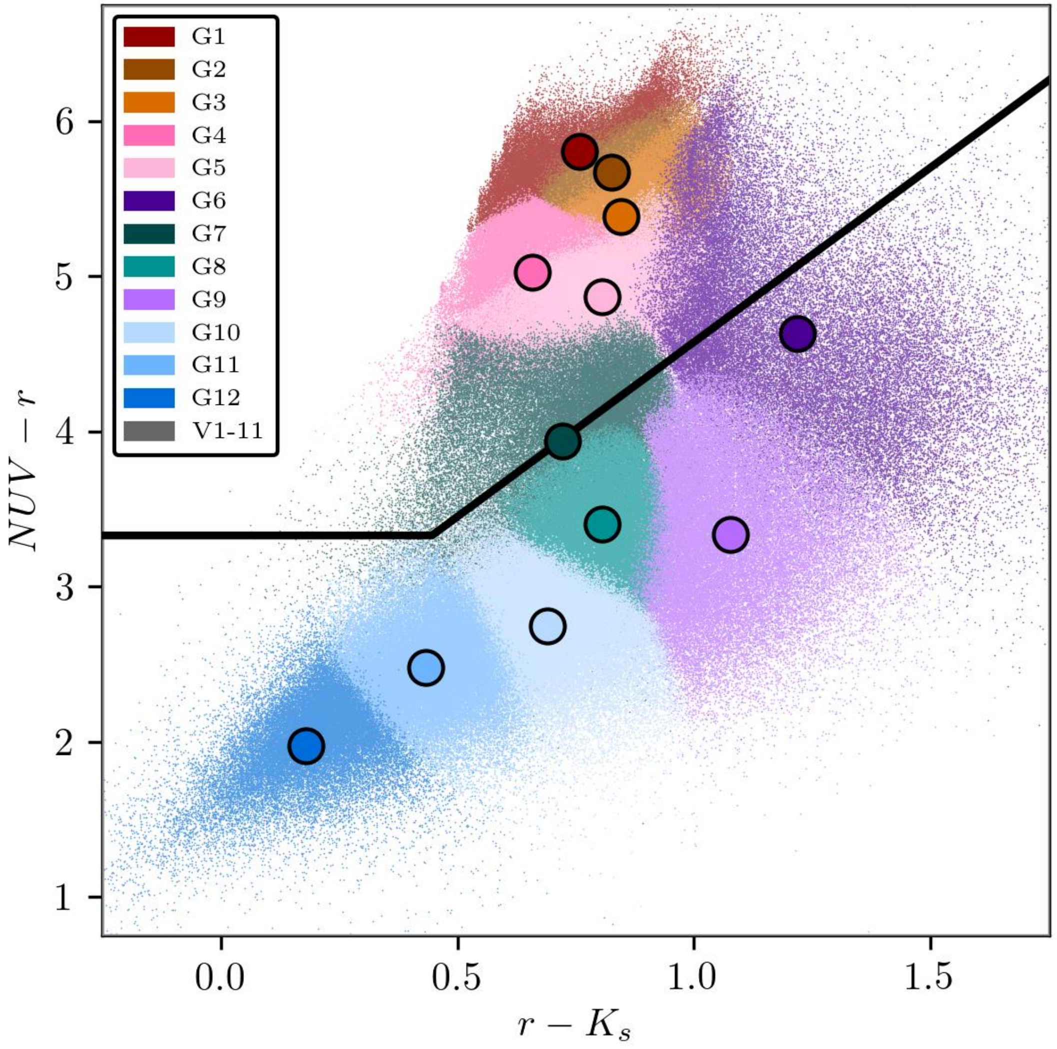
GSWLC ... rs: NUVrK



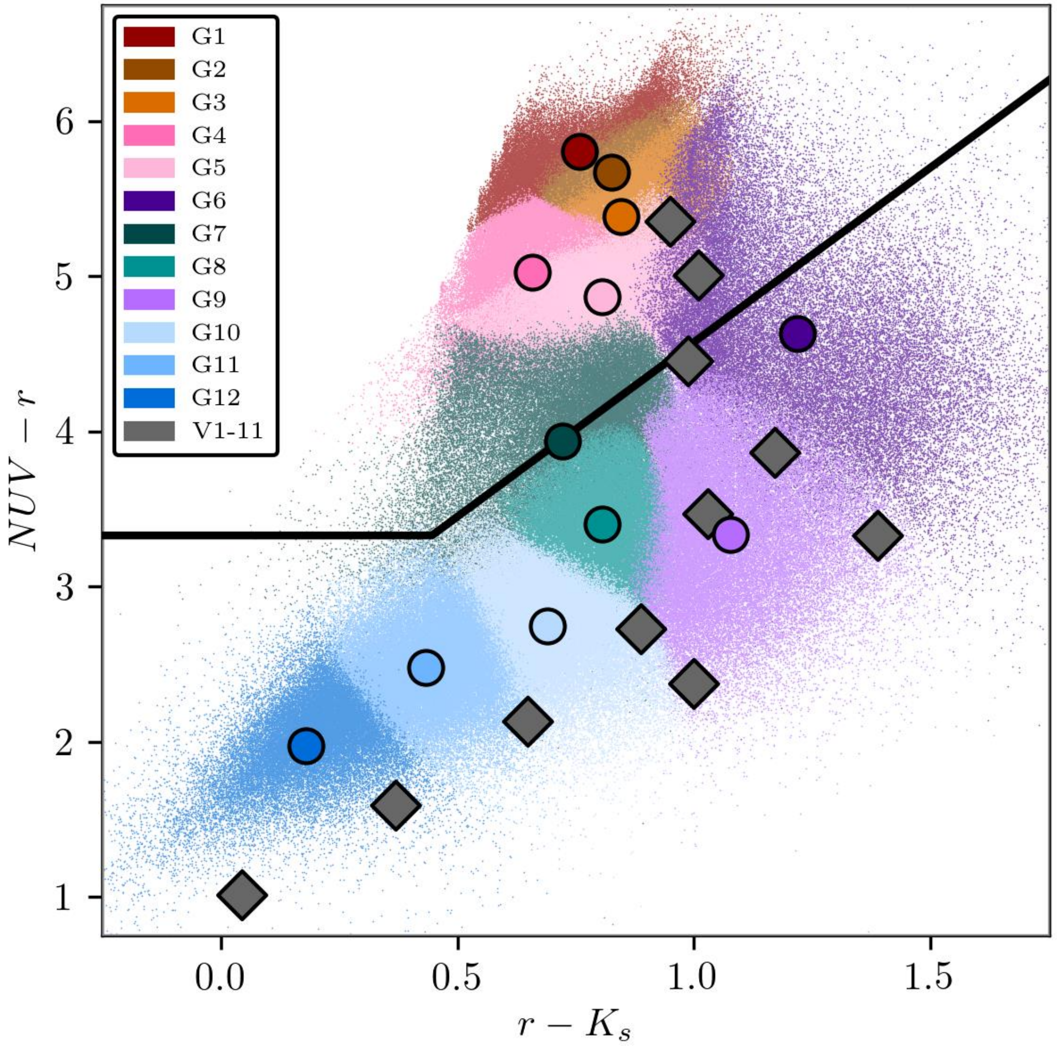
GSWLC clusters: NUVrK



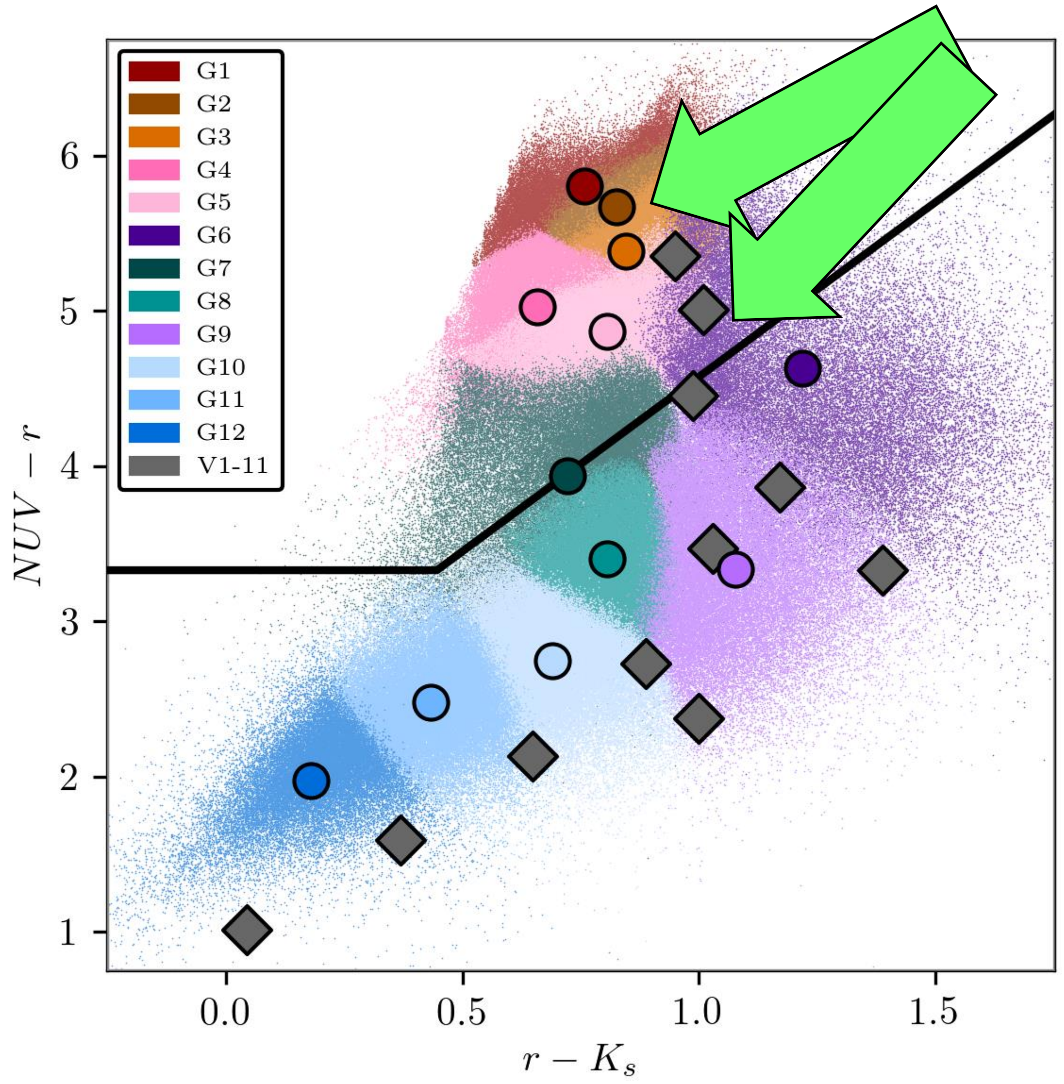
GSWLC clusters: NUVrK



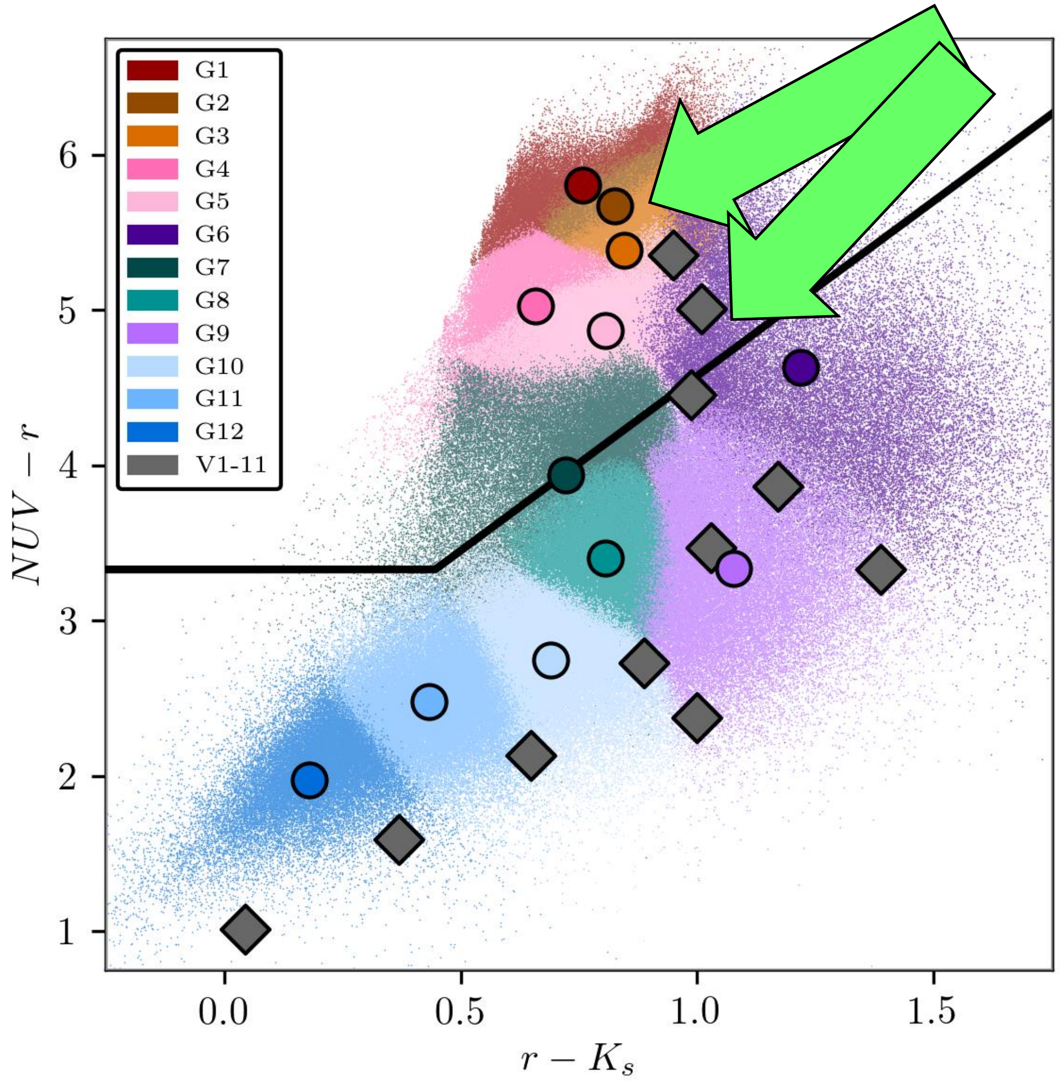
GSWLC + VIPERS clusters: NUVrK

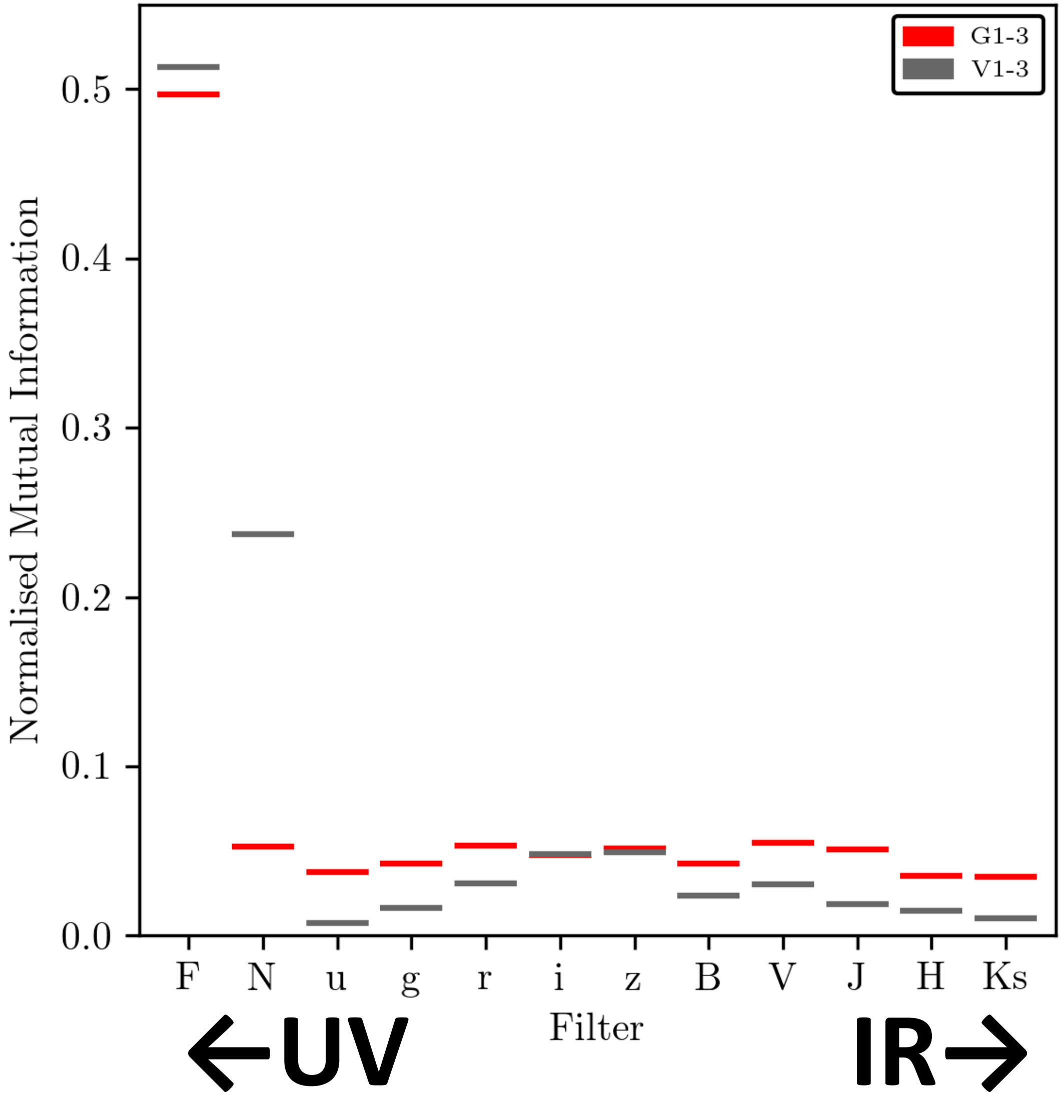


GSWLC + VIPERS clusters: NUVrK

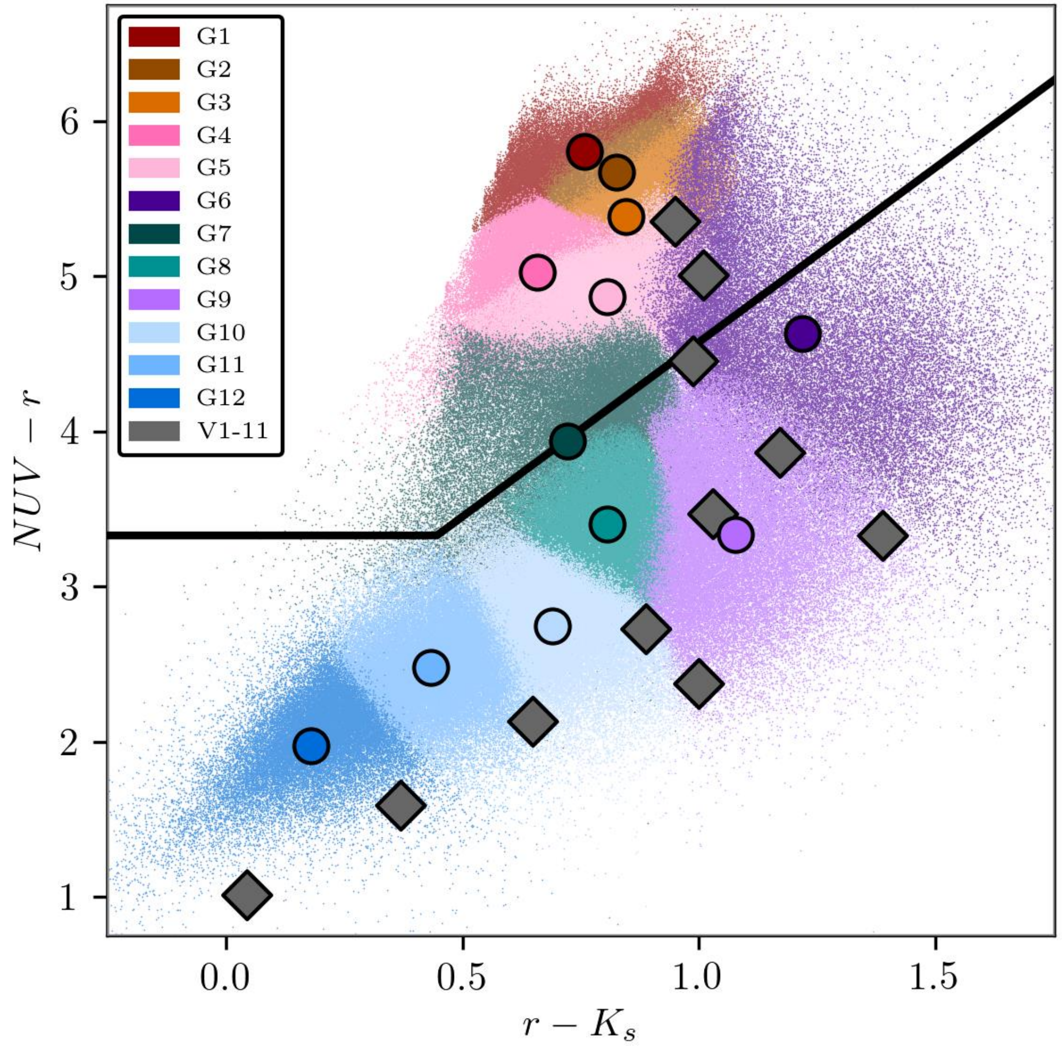


GSWLC + VIPERS clusters: NUVrK

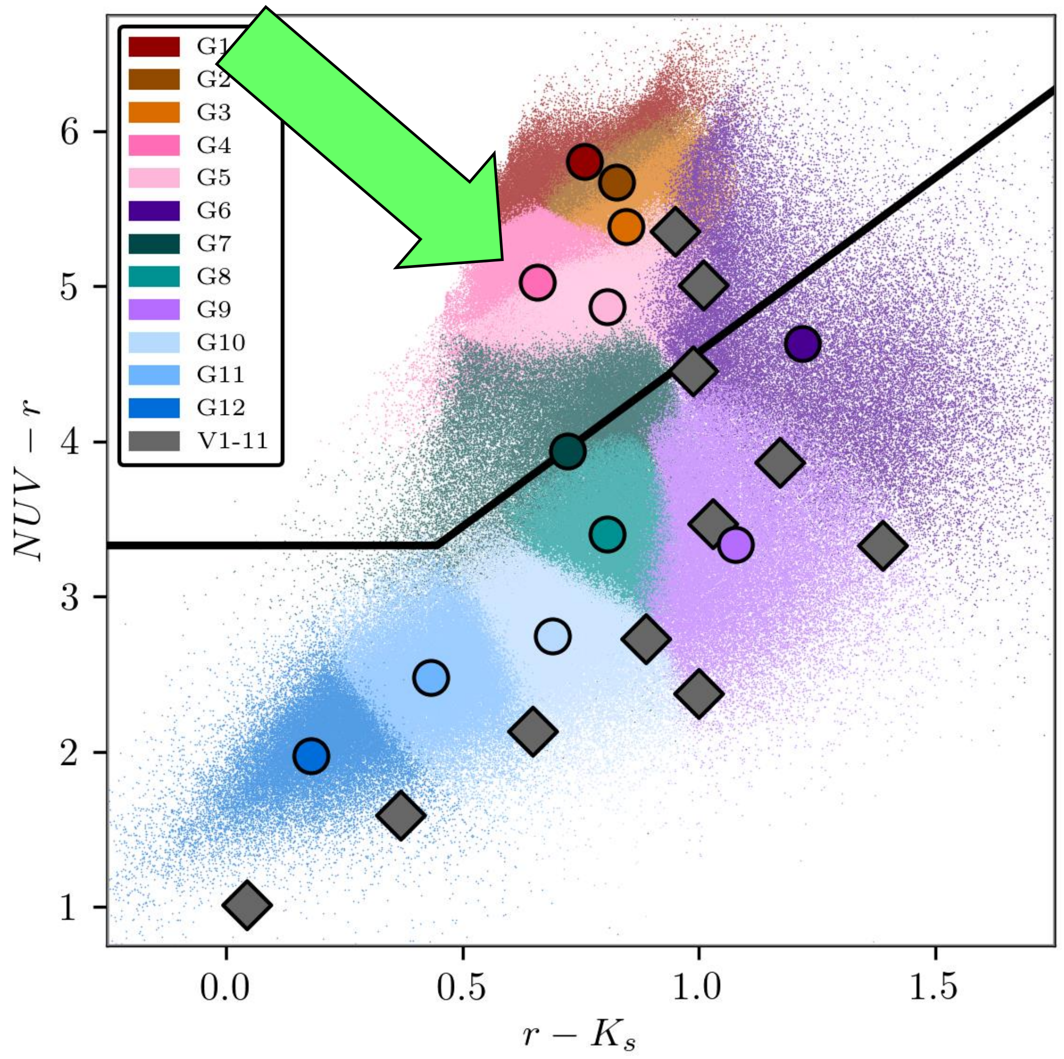




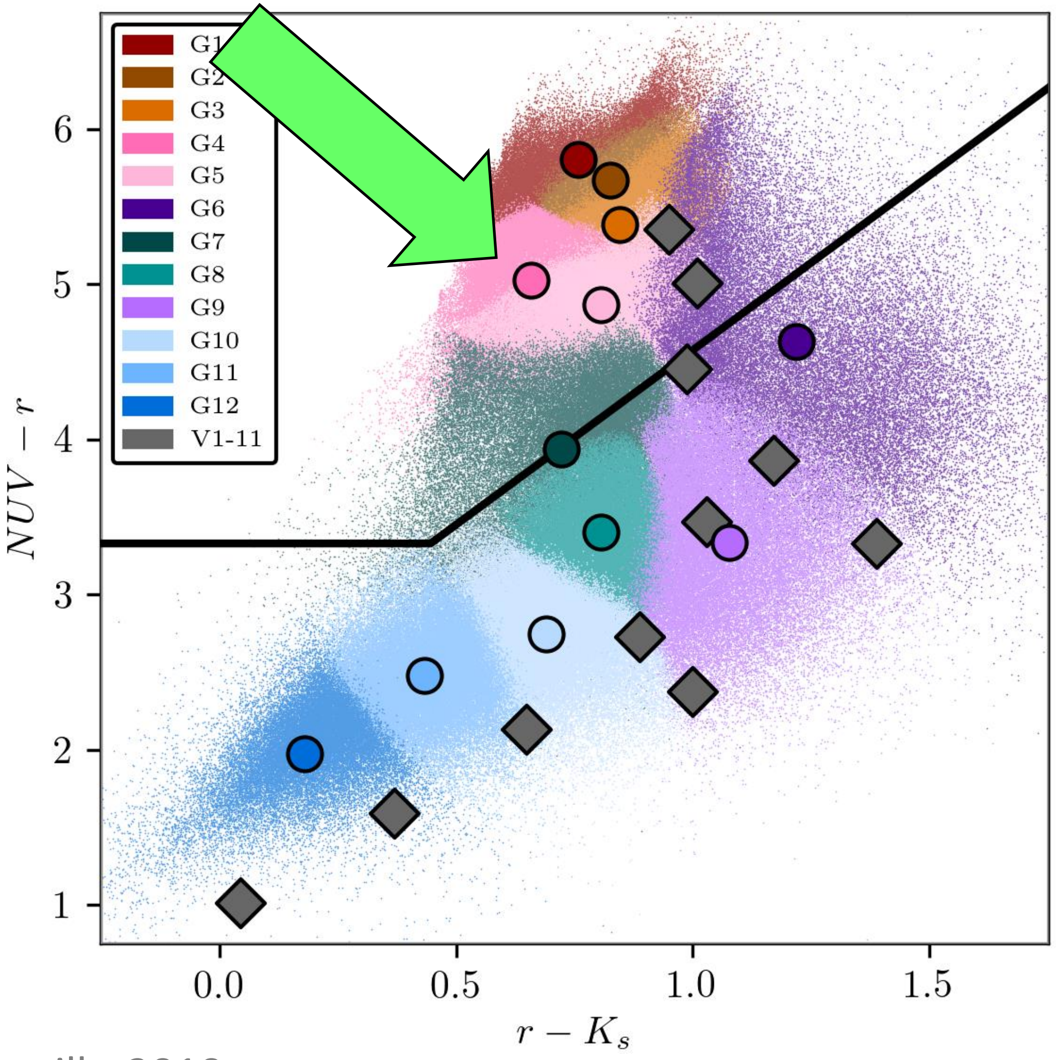
GSWLC + VIPERS clusters: NUVrK

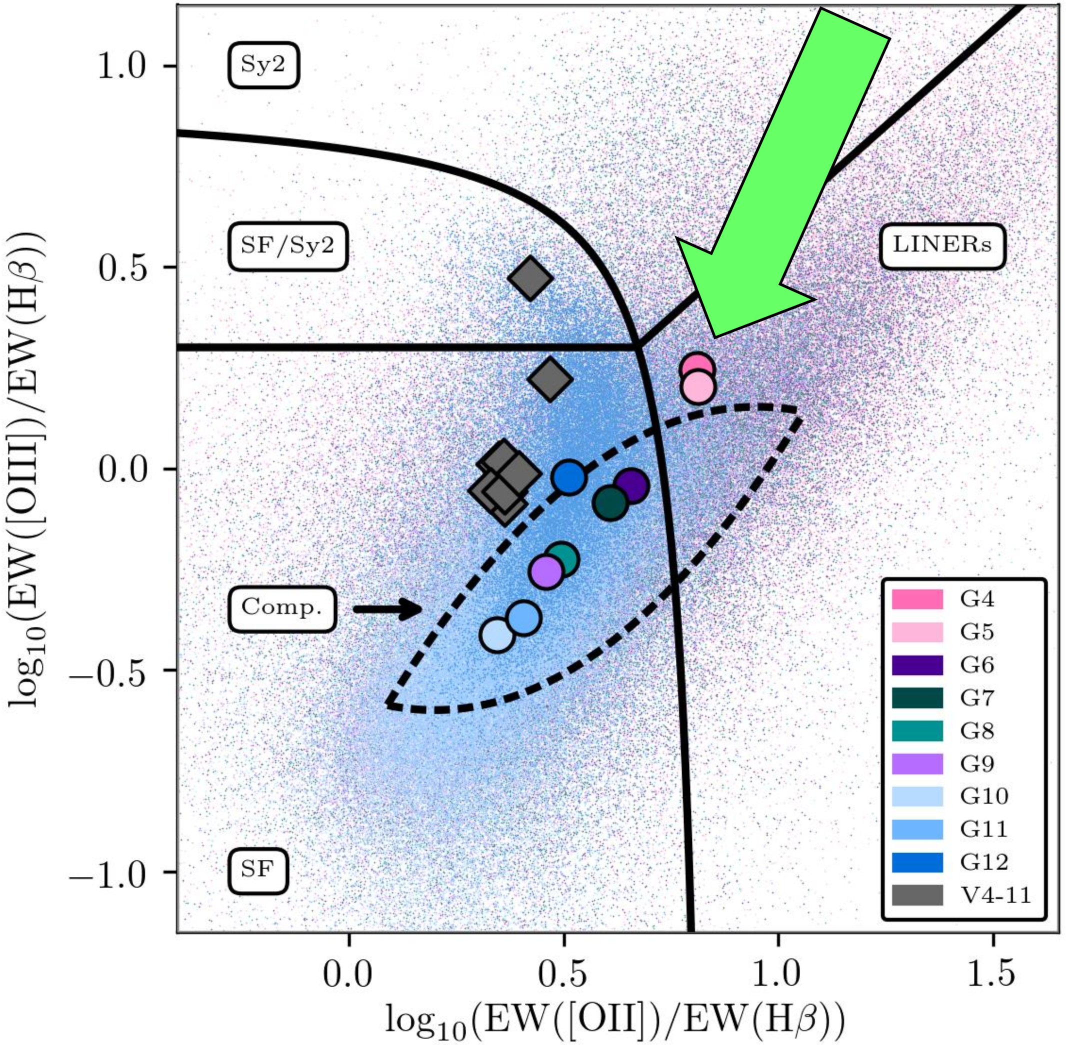


GSWLC + VIPERS clusters: new class?



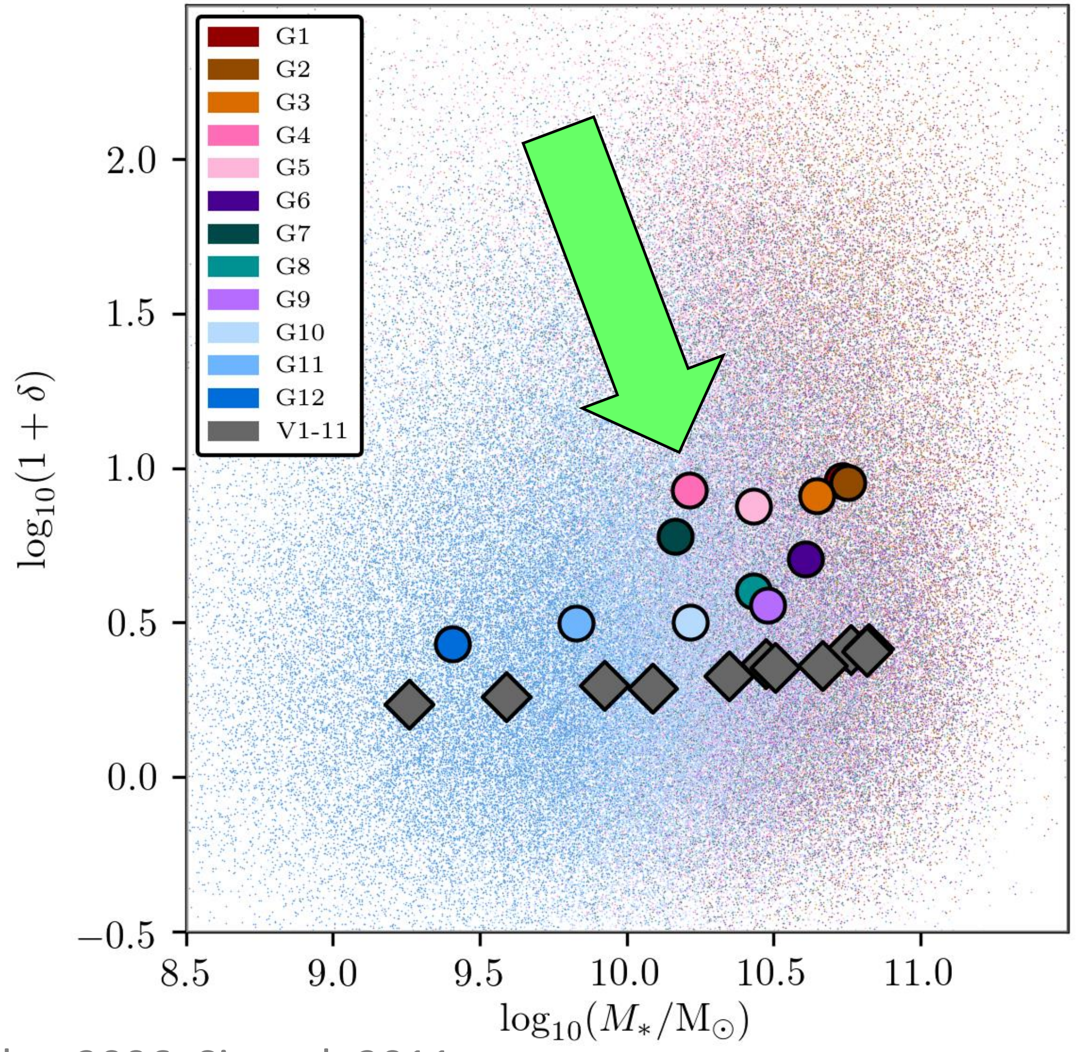
New class?: emission lines

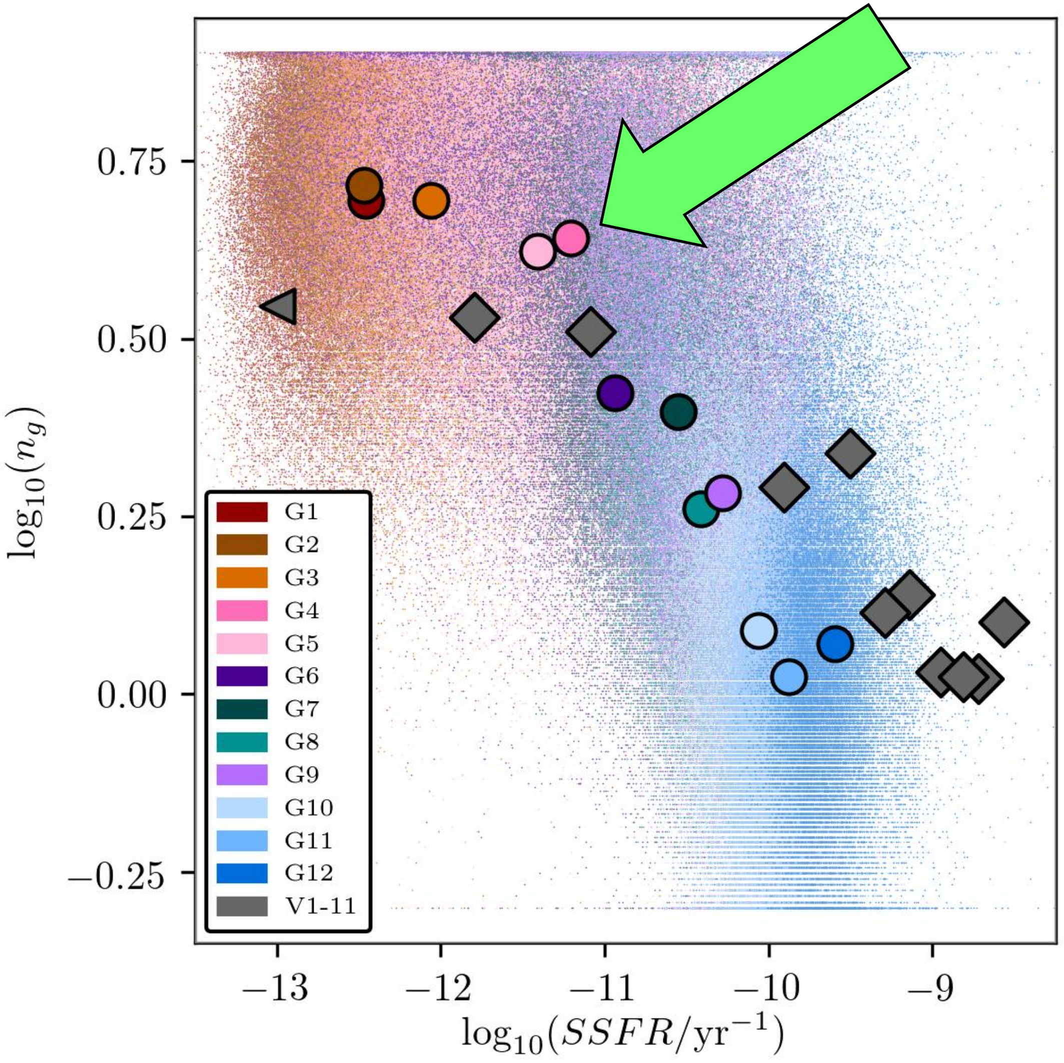




Lamareille-2010

New class?: environment and morphology





Baldry+2006; Simard+2011

Summary

- Comparing low- and intermediate-redshift galaxies with unsupervised machine learning.
- Want to understand cosmological evolution of subpopulations derived by clustering.
- Broad similarities: clear evolutionary sequence, diversity of red galaxies driven mostly by fall-off in FUV luminosity.
- Potential new cluster at low redshift, extra interpretation needed.
- Also: constrain influence of SED fitting on results.

LIVERPOOL JOHN MOORES UNIVERSITY

s.turner1@2012.ljmu.ac.uk

@sebturne