Machine Learning in the Era of Multiwavelength Astronomy

Tsutomu T. TAKEUCHI

Suchetha COORAY, Kai T. KONO,
Wen E. SHI, Hai-Xia MA

Division of Particle and Astrophysical Science, Nagoya University
Institute of Statistical Mathematics

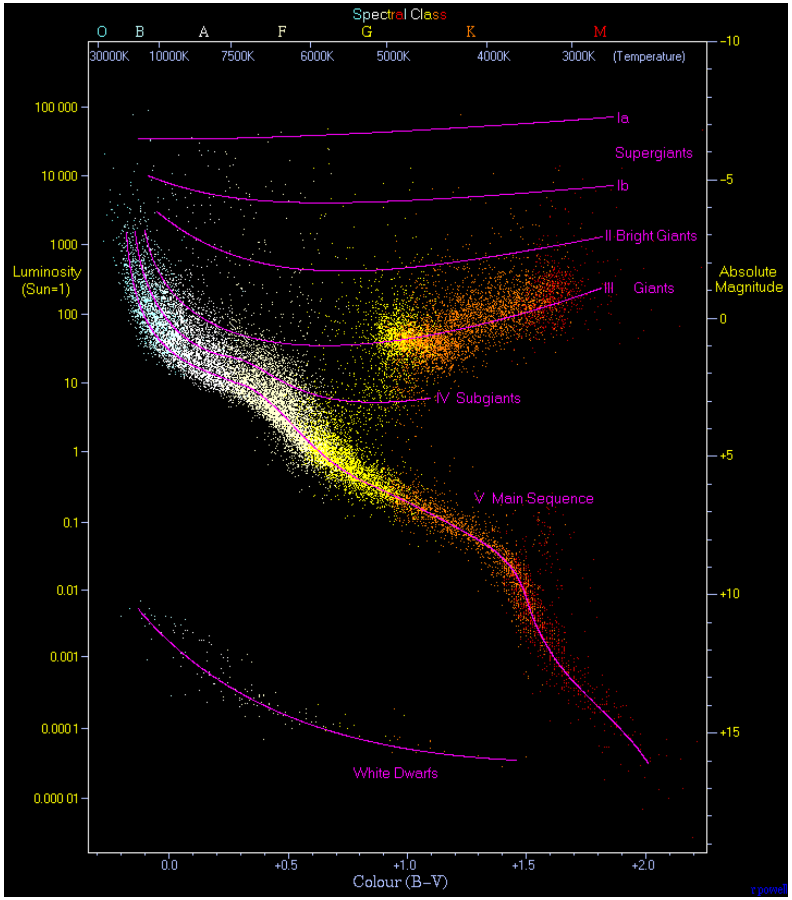
GOPIRA Symposium 2020, Online, 15 Sep., 2020

1. Introduction: Multiwavelength Astronomy

1.1 Traditional method

Hertzsprung-Russel diagram

The most well-known example of the color-magnitude diagram. Stellar evolution was quantified vastly through the H-R diagram.

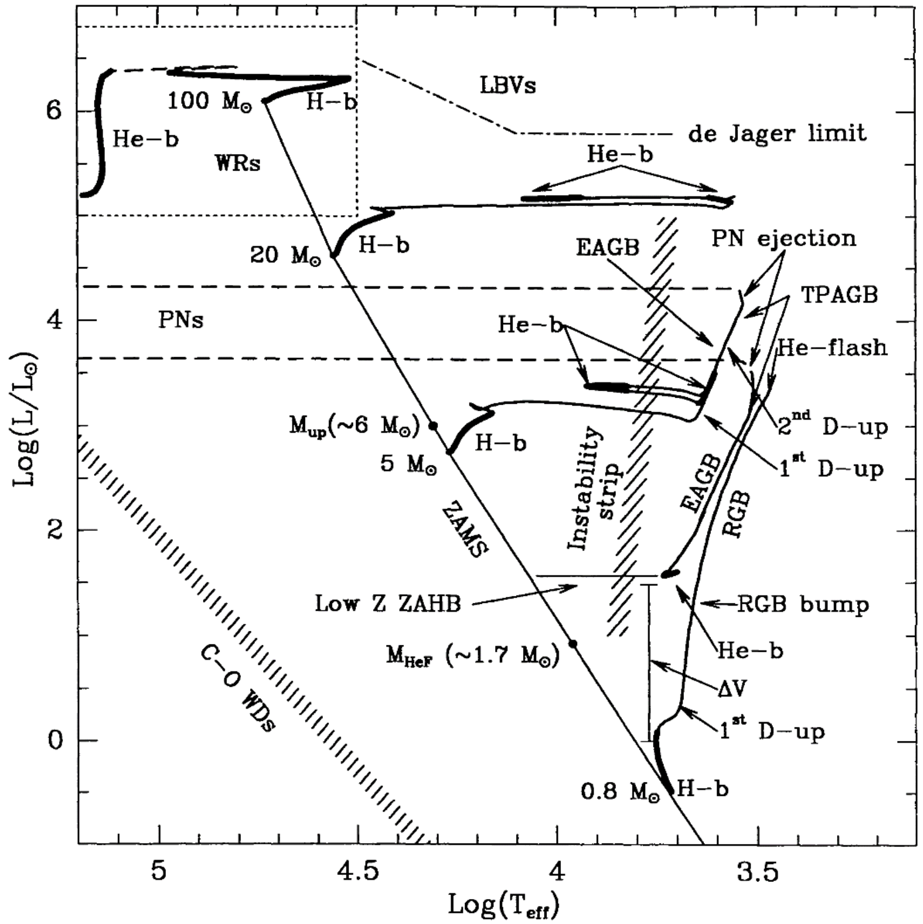


https://en.wikipedia.org/wiki/Hertzsprung%E2%80%93Russell_diagram

H-R diagram and stellar evolutionary track

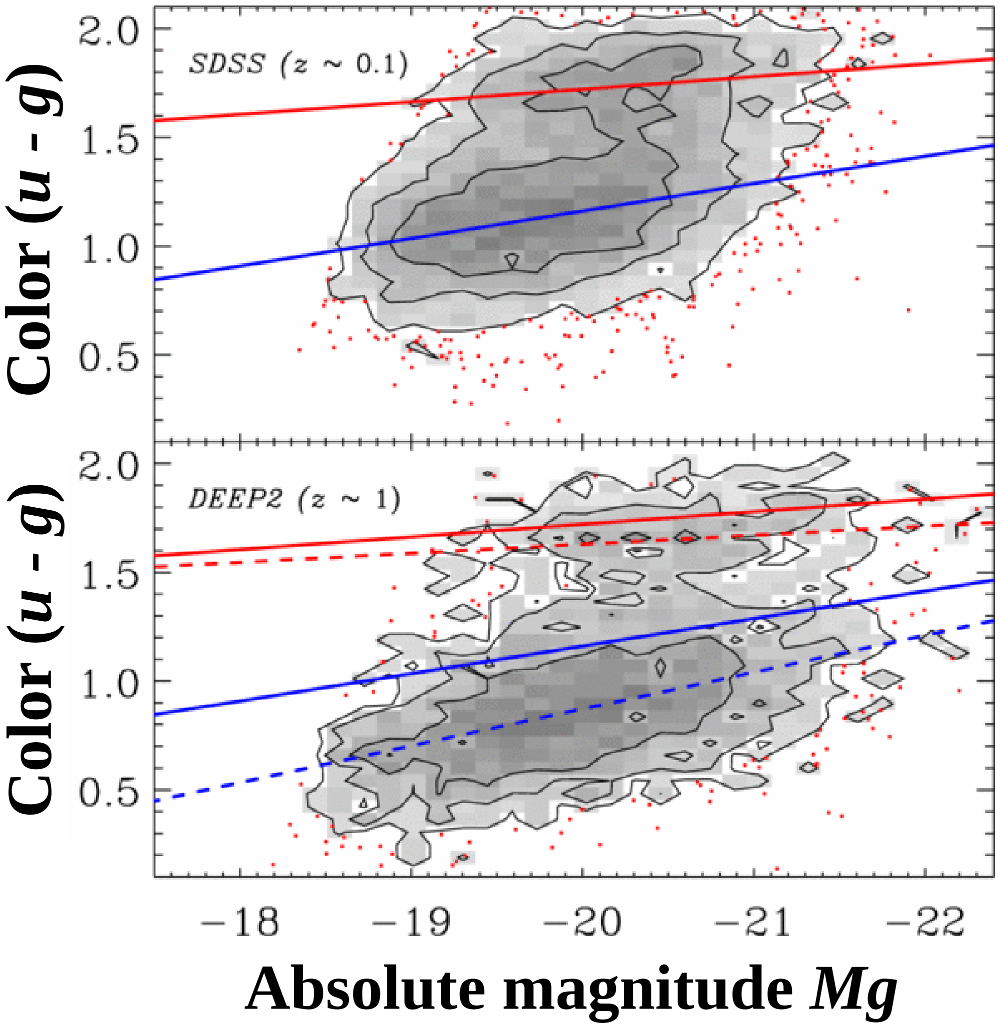
The theoretical evolutionary track of stars has been developed by a detailed comparison with observed H-R diagrams.

The H-R diagram has been a tremendously important tool in astronomy.



Chiosi et al (1992)

Color-magnitude relation of galaxies



If we plot galaxy luminosity (absolute magnitude) vs. color, a clear dichotomy is found: the color bimodality.

Redder galaxies:

red sequence

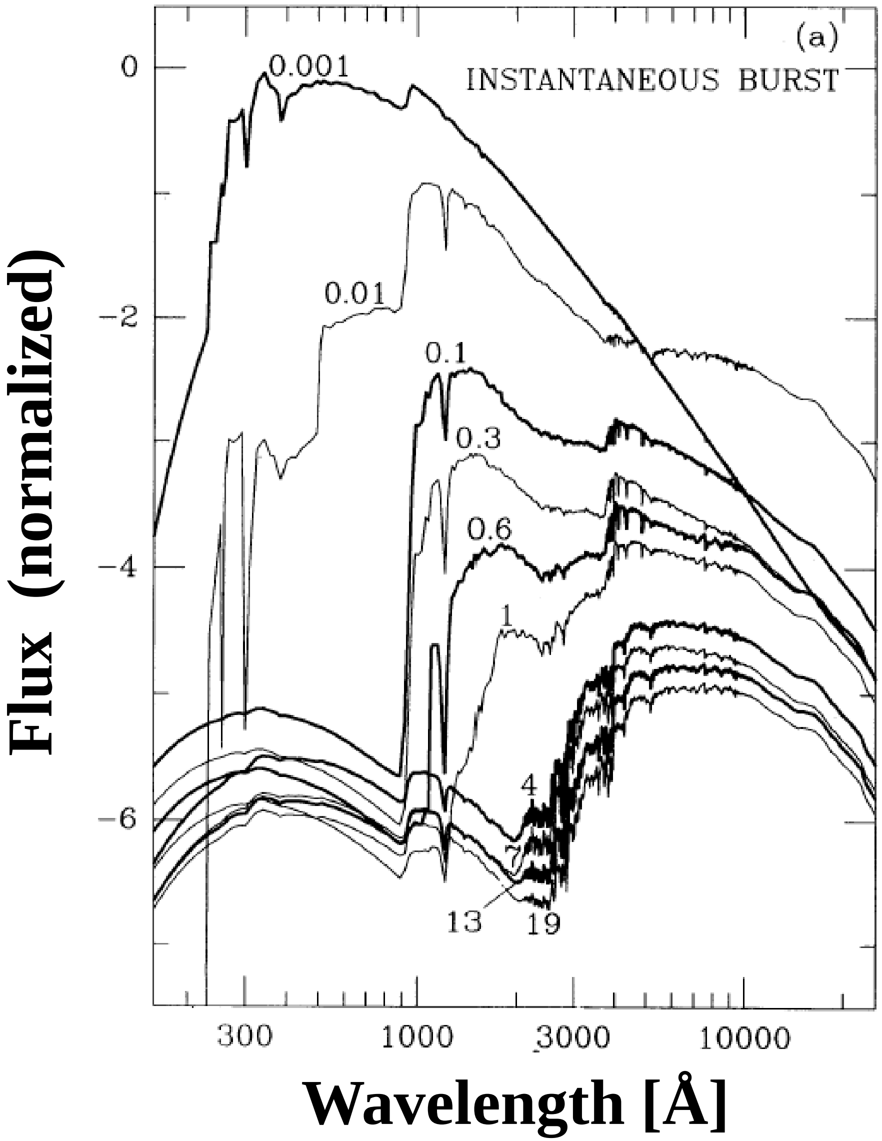
Bluer galaxies:

blue cloud

Boundary: green valley

Blanton (2006)

Galaxy evolution in multiwavelength

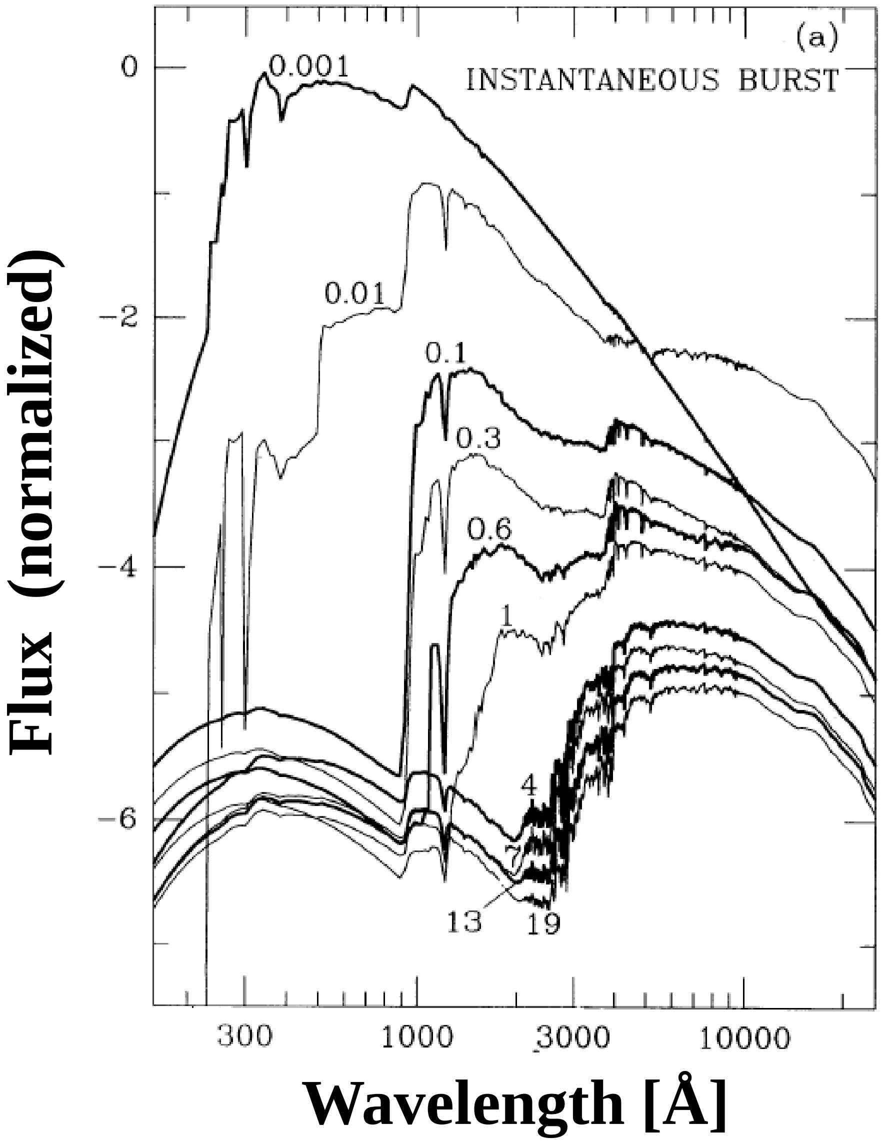


Star formation history (SFH) is one of the key factors of galaxy evolution.

SFH is directly reflected to the spectral luminosity of galaxies.

Bruzual & Charlot (1993)

Galaxy evolution in multiwavelength



Star formation history (SFH) is one of the key factors of galaxy evolution.

SFH is directly reflected to the spectral luminosity of galaxies.

Galaxy evolution related to the SFH will be well represented in the multiwavelength (band) luminosity space.

1.2 Problem in the Color-Based Methods

Colors are basically ratios of two luminosities.

⇒ Selection effect is always too entangled and messy.

⇒ Completeness test is almost impossible in a simple way.

1.2 Problem in the Color-Based Methods

Colors are basically ratios of two luminosities.

⇒ Selection effect is always too entangled and messy.

⇒ Completeness test is almost impossible in a simple way.

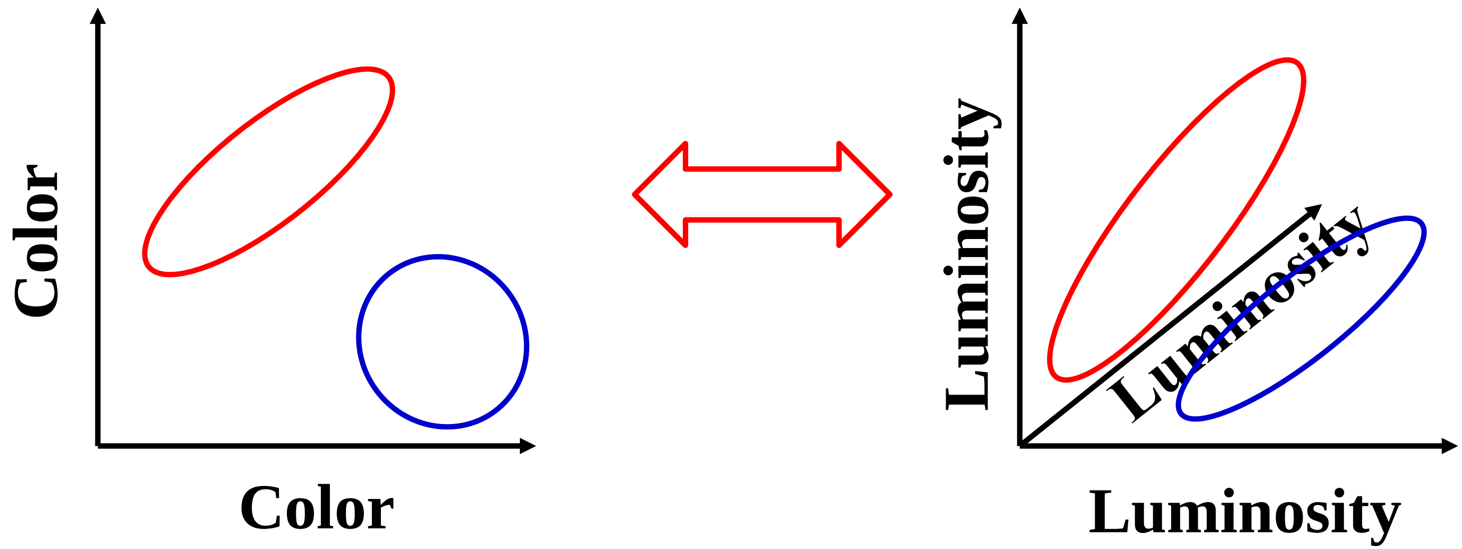
⇩

Suggestion: forget about colors!

Instead, we can simply use the distribution of galaxies in a multidimensional luminosity (absolute magnitude) space.

From colors to multiwavelength luminosities

Since we have a bimodality in color-color space, we must have an equivalent peaks in the multidimensional luminosity space. Color-color plots only show reduced information.



2. What is the Galaxy Evolution?

In those days

Beatrice Tinsley

Galaxies evolve.

In galactic astronomy of 70-80's, the meaning of galaxy evolution seemed to be unambiguous.

Galaxy evolution

= evolution of stellar population and chemical composition

Today

Galaxies evolve in various aspects:

Today

Galaxies evolve in various aspects:

$$
\begin{aligned}
\mathrm{SFR}(t) &= f_1(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_*(t) &= f_2(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{mol}}(t) &= f_3(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{HI}}(t) &= f_4(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{dust}}(t) &= f_5(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{halo}}(t) &= f_6(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
\delta_{\mathrm{gal}}(t) &= f_7(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
&\vdots
\end{aligned}
$$

$$x = x(T|T > t)$$

Today

Galaxies evolve in various aspects:

$$
\begin{aligned}
\mathrm{SFR}(t) &= f_1(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_*(t) &= f_2(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{mol}}(t) &= f_3(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{HI}}(t) &= f_4(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{dust}}(t) &= f_5(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
M_{\mathrm{halo}}(t) &= f_6(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
\delta_{\mathrm{gal}}(t) &= f_7(\mathrm{SFR}, M_*, M_{\mathrm{mol}}, M_{\mathrm{HI}}, M_{\mathrm{dust}}, M_{\mathrm{halo}}, \delta_{\mathrm{gal}}, \ldots) \\
&\vdots
\end{aligned}
$$

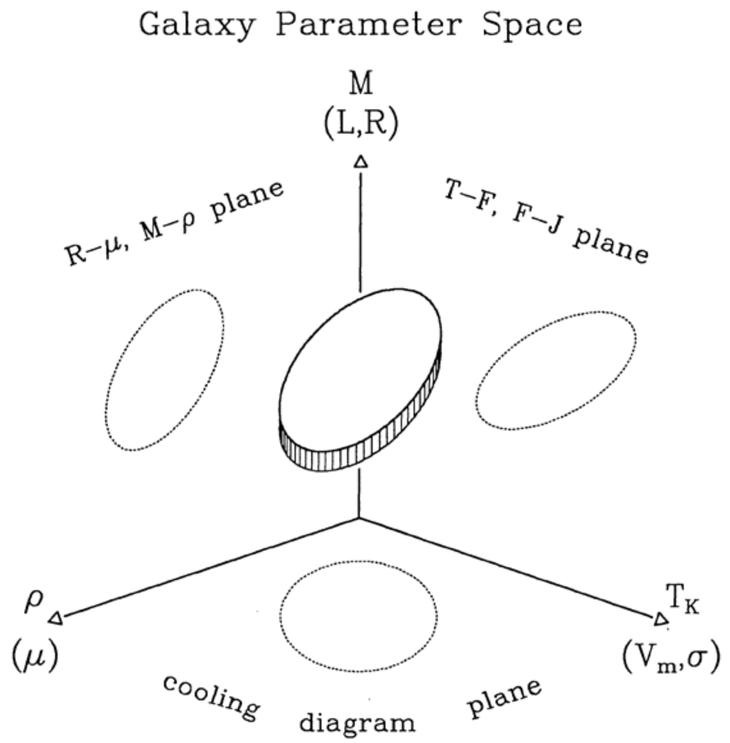
This is the formal and ultimate goal of the studies on galaxy evolution, but clearly it is a substantially complicated problem. It is time to define the evolution of galaxies with more objective point of view.

3. Galaxy Manifold

3.1 Galaxy manifold in the early days

Historically, in 80's, astronomers introduced a method of classical multivariate analysis such as PCA to find and unify various scaling relations (e.g., Djorgovski 1992).

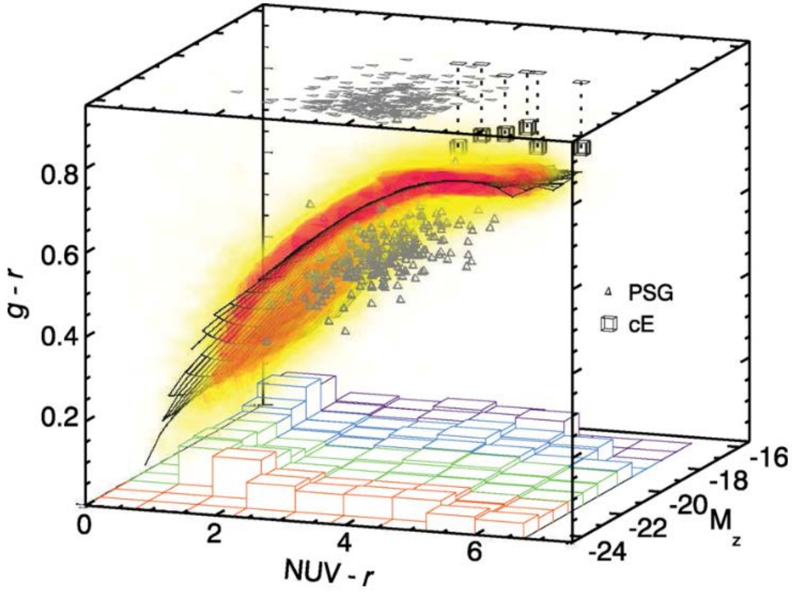
However, since classical PCA-type analysis could only find a linear structure in the feature space, the idea worked only to a limited problems and have been once forgotten.



Djorgovski (1992)

3.2 Galaxy manifold: reboot

Some preceding studies have suggested the existence of a smooth relation of galaxies in the 3D color–color–magnitude space smoothly continuing from the blue cloud to the red sequence (e.g. Chilingarian et al. 2012).



Chilingarian et al. (2012)

⇒ general idea of a low dimensional submanifold existing in a higher dimensional feature space: revival of the galaxy manifold!

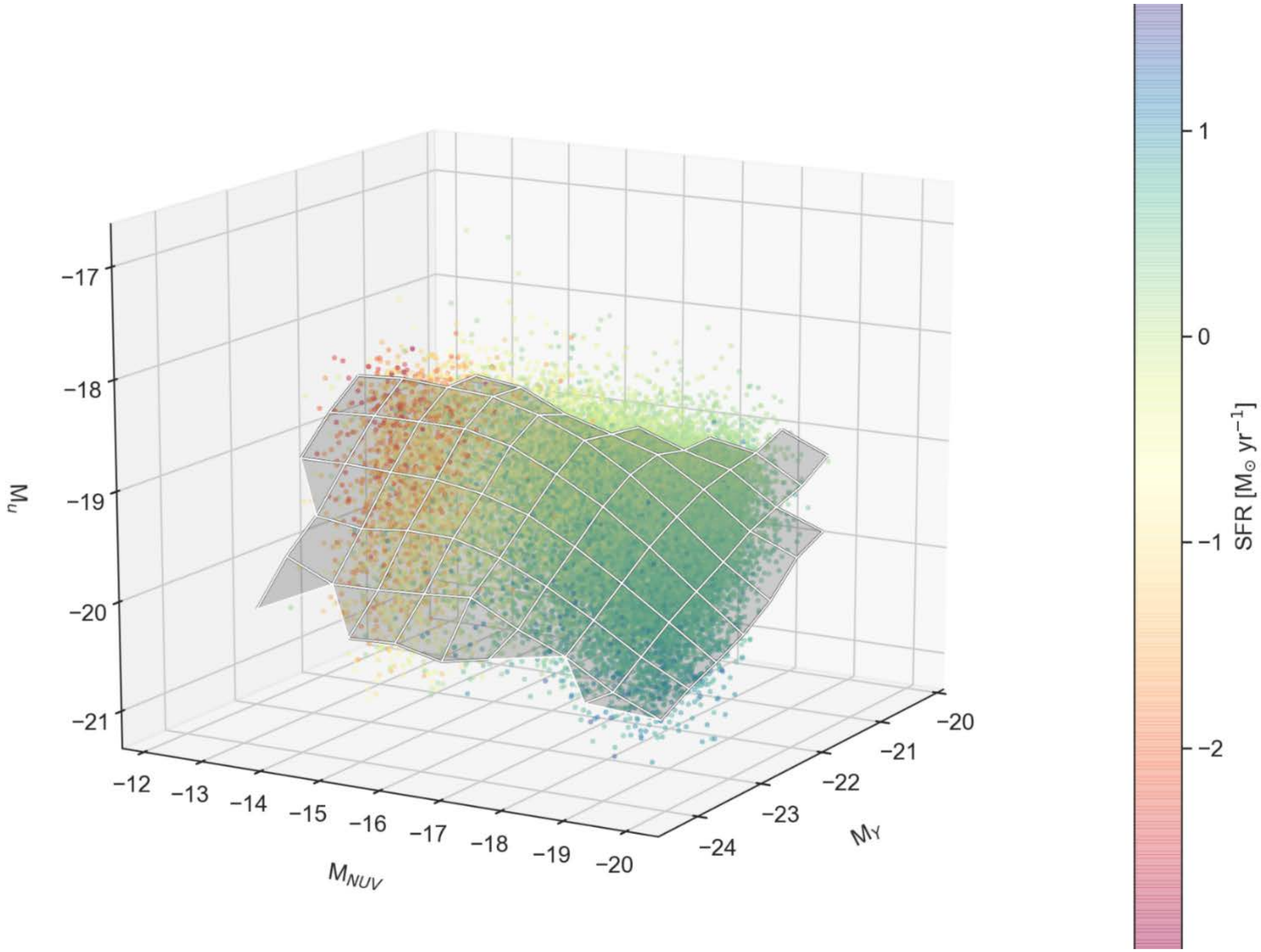
Galaxy manifold from RCSED

- **Reference Catalog of galaxy Spectral Energy Distributions (RCSED) (Chilingarian et al. 2016)**
- **Catalog of galaxies produced as join between *GALEX*, SDSS, and UKIDSS catalogs, and processed with state-of-the-art spectral analysis methods**
- **Covers approximately 25% of the sky and contains *k*-corrected ultraviolet-to-near-infrared photometry (11 bands of FUV, NUV, *u*, *g*, *r*, *i*, *z*, *Y*, *J*, *H*, *K*) of some 1 million galaxies, as well as some of their physical properties**

RCSED

http://rcsed.sai.msu.ru

Galaxy manifold from RCSED



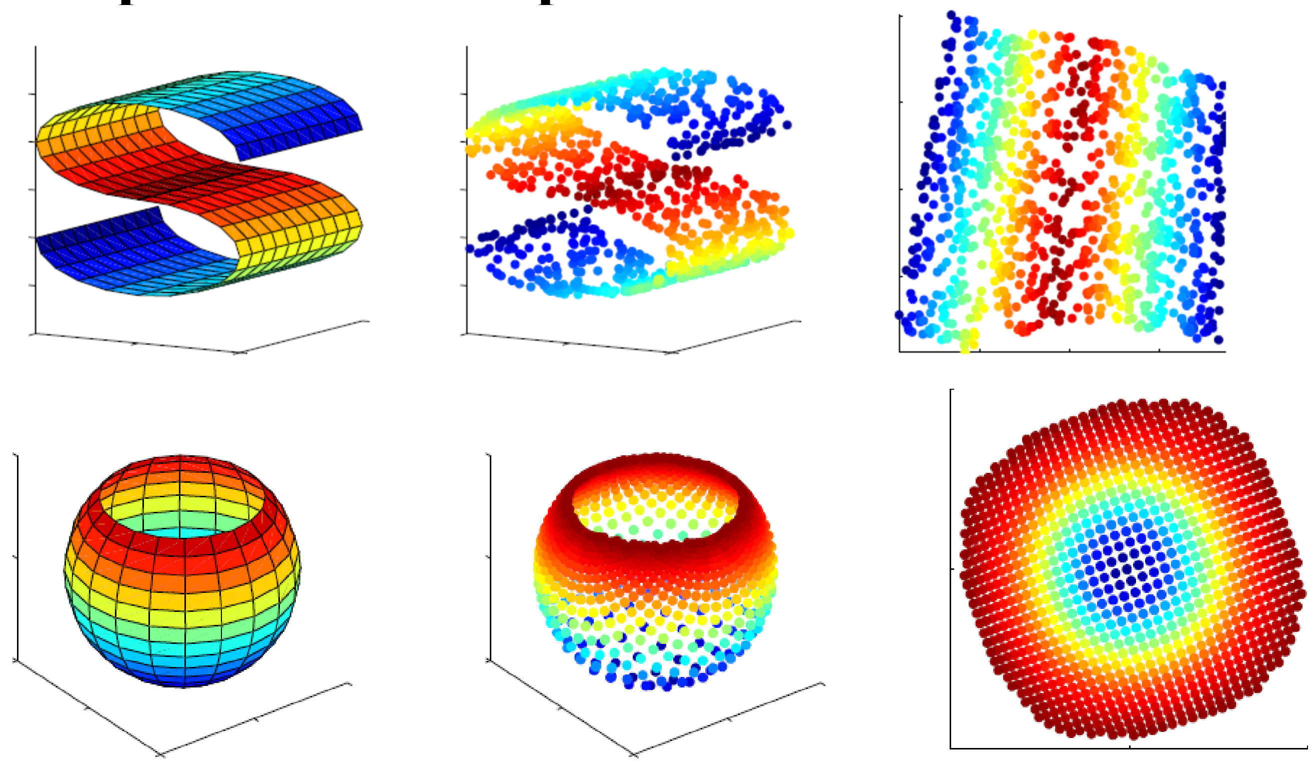
Cooray et al. (2020)

We want to characterize and parametrize this nonlinear manifold, and hopefully interpret physically!

3.3 Galaxy manifold from manifold learning

Manifold Learning (or non-linear dimensionality reduction) embeds data that originally lies in a high dimensional space in a lower dimensional space, while preserving characteristic properties.

A manifold is a topological space that locally resembles Euclidean space near each point.



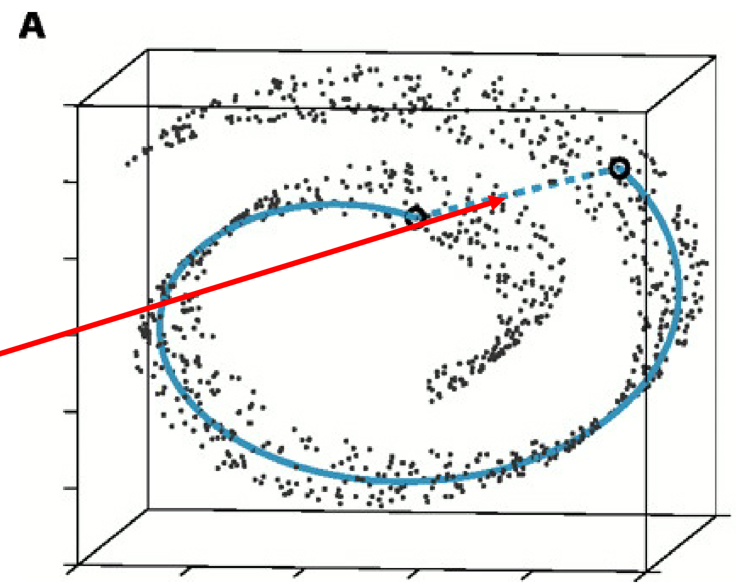
Isomap

Tenenbaum et al. (2000)

Neighboring points: **input-space distance**

Distant points: **a sequence of "short hops" between neighboring points**

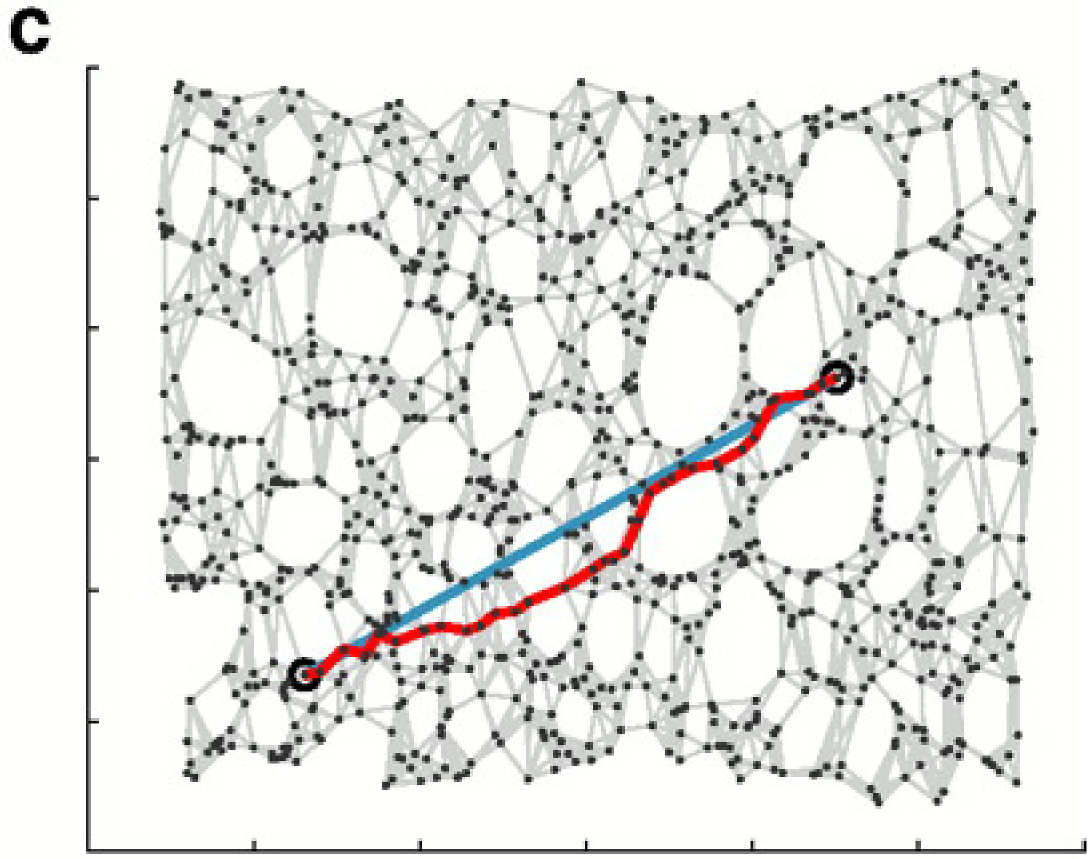
Method: **Finding shortest paths in a graph with edges connecting neighboring data points**



N.B. **Unlike the geodesic distance, the Euclidean distance cannot reflect the geometric structure of the data points.**

Isomap

The 2-D embedding recovered by Isomap.



Tenenbaum et al. (2011)

UMAP

McInnes et al. (2018)

UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction

The algorithm is founded on three :

1. **The data are uniformly distributed on a Riemannian manifold.**

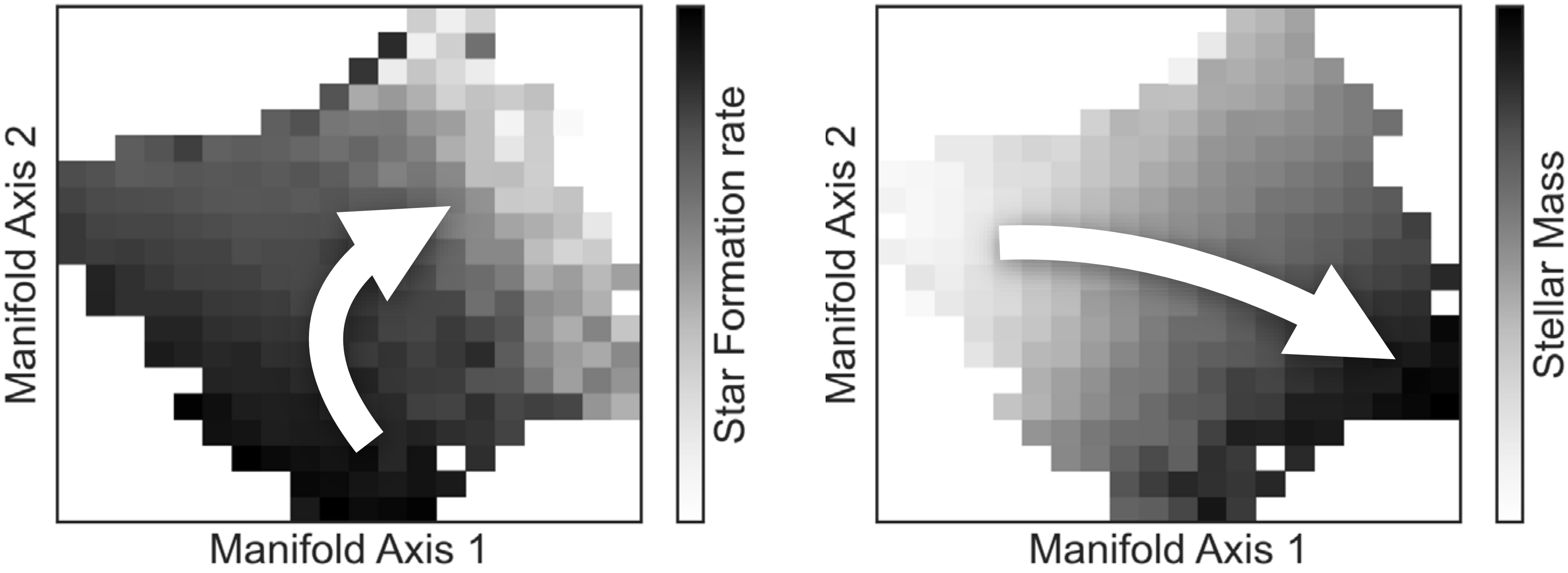
2. **The Riemannian metric is locally constant (or can be approximated as such).**

3. **The manifold is locally connected.**

From these assumptions it is possible to model the manifold with a fuzzy topological structure.

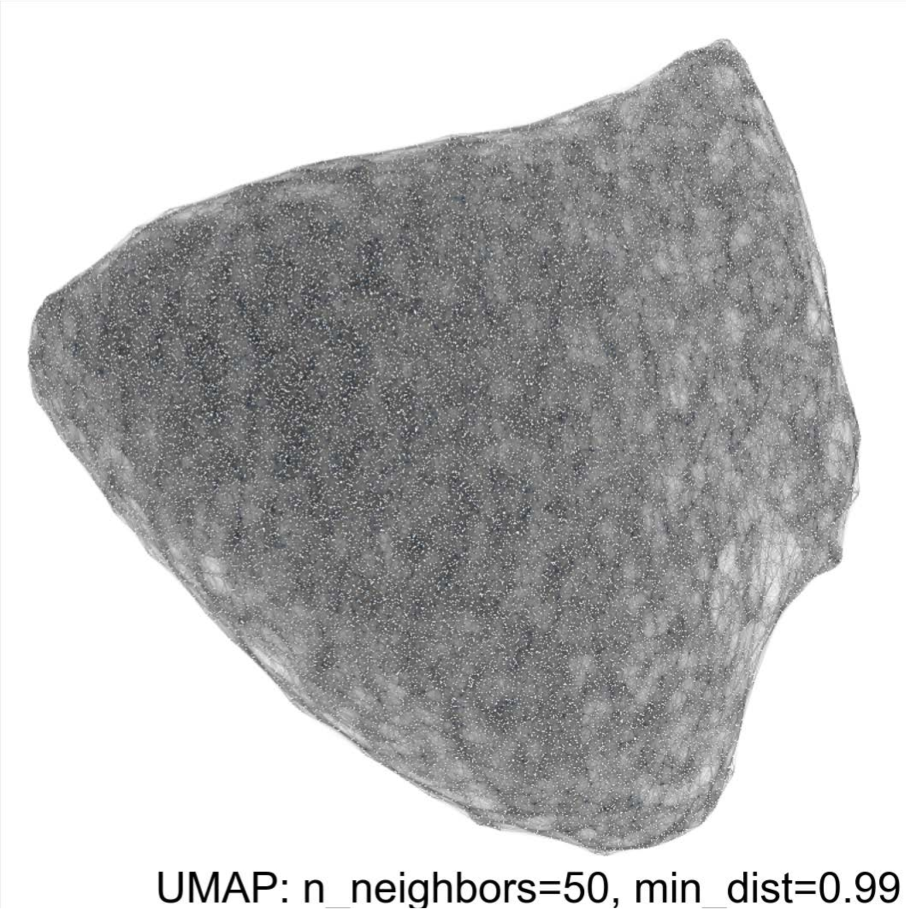
Result: galaxy manifold from Isomap

Isomap preserves the density of the data point cloud (i.e., distribution of galaxies) on the manifold (i.e. dense regions remain dense).



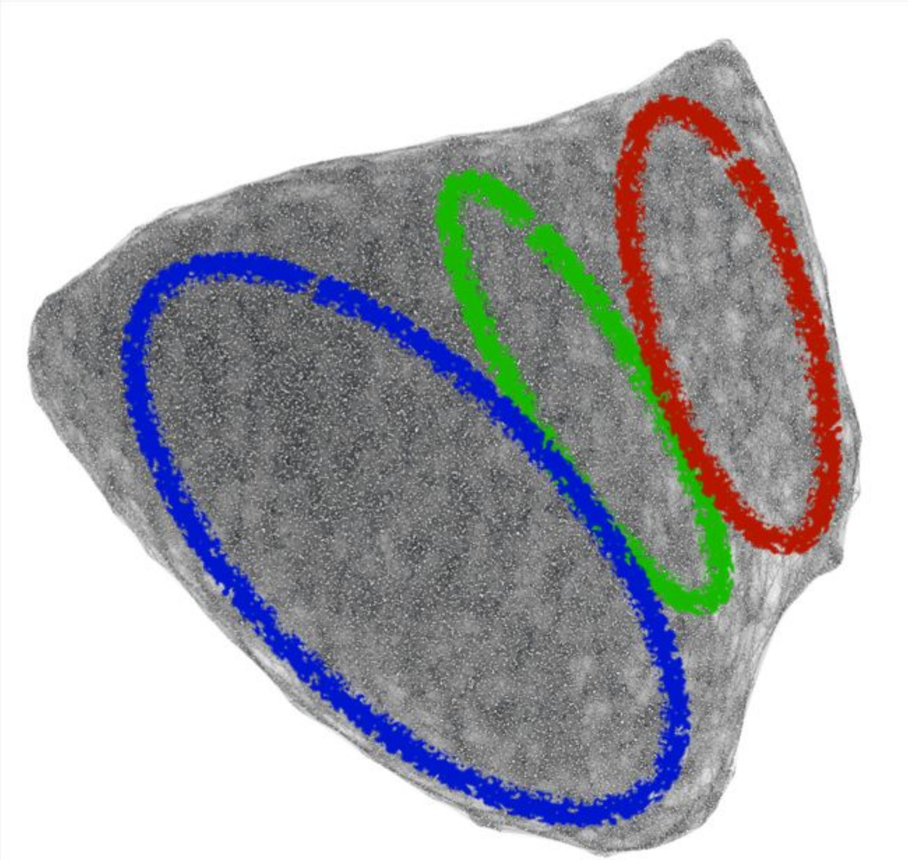
The arrows show the continuous evolution of less-massive, actively star-forming galaxies to massive, quiescent galaxies.

Result: galaxy manifold from UMAP



Cooray et al. (2020)

Result: galaxy manifold from UMAP



Blue cloud: actively star forming

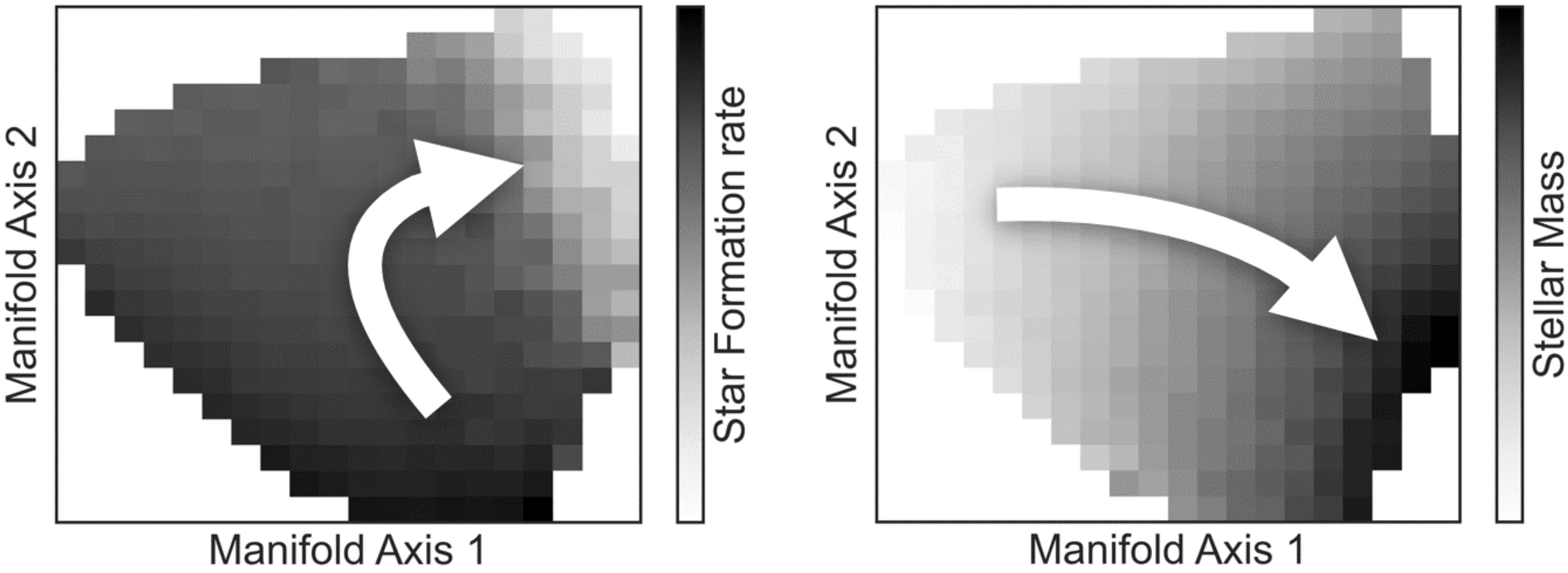
Red sequence: quiescent

Green valley: transient

Cooray et al. (2020)

Result: galaxy manifold from UMAP

UMAP expands dense regions and contracts sparse regions, i.e., uniform representation of the observational galaxy parameters.



The arrows show the continuous evolution of less-massive, actively star-forming galaxies to massive, quiescent galaxies.

4. Conclusions

1. **We demonstrated how machine learning methods can aid our understanding of galaxy evolution.**

2. **We found highly nonlinear continuous structure in the multidimensional luminosity space: the galaxy manifold. This is a revived, improved version of the concept discussed in 80's.**

3. **All the known empirical relations are projection of the manifold.**

4. **The galaxy manifold represents the evolutionary sequence of galaxies. Possibly we can parametrize the galaxy evolution by a few parameters.**

4. Conclusions

Take-home message

Forget about colors, and formulate your problem in luminosity (and other feature) space!

Machine learning can serve as a powerful and revolutionary method to develop fundamental part of astrophysics!