

Multivariate Classification of Galaxies: which Observables?

D. Fraix-Burnet

Univ. Grenoble-Alpes / CNRS / IPAG

<http://ipag.osug.fr/public/fraixbud/>

<http://astrocladistics.org/>



Outline

- 1 Unsupervised Classification of Galaxies
- 2 Which Parameters?
- 3 Phylogenetic Tools
- 4 Conclusion

1 Unsupervised Classification of Galaxies

2 Which Parameters?

3 Phylogenetic Tools

4 Conclusion

Unsupervised Classification of Galaxies

Galaxy classification needed for

Galaxy physics and diversification (formation and evolution)

As a probe of their environment – i.e. the Universe (cosmology)

Unsupervised multivariate classification (clustering)

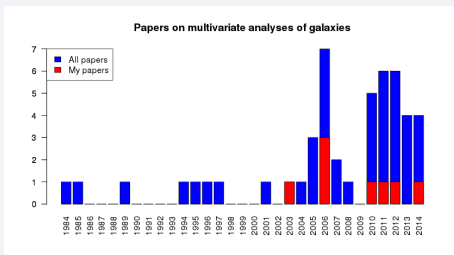
Currently no general, physical, complete and objective classification

Unsupervised Classification of Galaxies

Multivariate Approaches to Classification in Extragalactic Astronomy

(Fraix-Burnet, Thuillard, Chattopadhyay 2015, *Frontiers in Astronomy and Space Sciences*)

- Nearest Neighbor
- Support Vector Machine
- K-Means
- Fuzzy clustering
- Information Bottleneck
- Mixture Models
- Wavelet Analysis
- Hierarchical Classification
- Minimum Spanning Tree (FoF)
- Phylogenetic methods



Unsupervised Classification of Galaxies

<i>Method</i>	<i>p</i>	<i>Parameters</i>
kNN	16	colours
SVM	12	morpho, lumin.
k-means	1637	spectra
–	4	$\sigma, \mu_e, r_e, \text{Mg2}$
–	6 (23)	$\sigma, \mu_e, \text{NaD}, \text{D/B}, [\text{OIII}], [\text{MgbFe}]'$
–	3	M, r, M/L
fuzzy partitioning	3	PCA from 8 morph., photom., spectra
Information Bottleneck	738	spectra
Mixture Models	4	colours + redshift
–	10	red. dim. from spectra
Shapelets	?	images, morpho
Hierarchical clustering	3	PCA on morphological indicators
MST	1637	spectra
Maximum Parsimony	4/6(23)...	various

- 1 Unsupervised Classification of Galaxies
- 2 Which Parameters?
- 3 Phylogenetic Tools
- 4 Conclusion

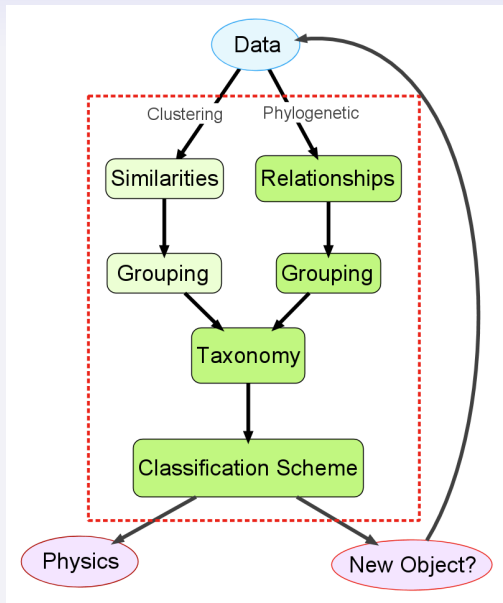
Which parameters?

Criteria

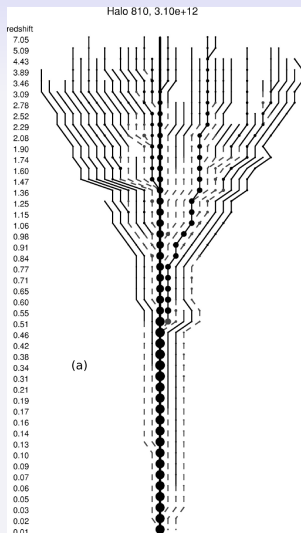
- **availability**
technology evolves with time
- **reliability**
uncertainties
- **completeness**
missing values, sample diversity
- **objectivity**
the parameters affects the physical meaning of the classification
- **relevance**
a given parameter set may not be adequate for all clustering tools

- 1 Unsupervised Classification of Galaxies
- 2 Which Parameters?
- 3 Phylogenetic Tools**
- 4 Conclusion

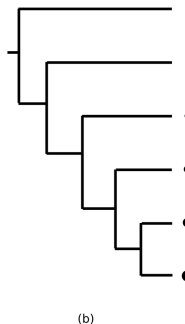
Clustering vs Phylogenetic



Merger Tree (Genealogy)



Cladogram (Phylogeny)



Diversification of Dark Matter Haloes

Stewart et al. 2008

Fraix-Burnet 2009

Minimum Spanning Tree (also Friends-Of-Friends)

Labelled tree
No internal node
Polynomial time

$$w(T) = \sum_e w(e)$$

Cladistics (Maximum Parsimony)

Semi-labelled tree
More general
NP-hard

$$S = \sum_k \sum_{i>j} d(f_k(i), f_k(j))$$

For a L1 norm

$$w(e) = \sum_k |f_k(i) - f_k(j)|$$

$$d(f_k(i), f_k(j)) = \sum_k |f_k(i) - f_k(j)|$$

$$w(T) = \sum_{i>j} \sum_k |f_k(i) - f_k(j)|$$

$$S = \sum_k \sum_{i>j} |f_k(i) - f_k(j)|$$

$$\frac{w_{min}}{2} < S_{min} < w_{min}$$

Redundancies

123 Virgo galaxies

48 parameters:

photometry wide and narrow band

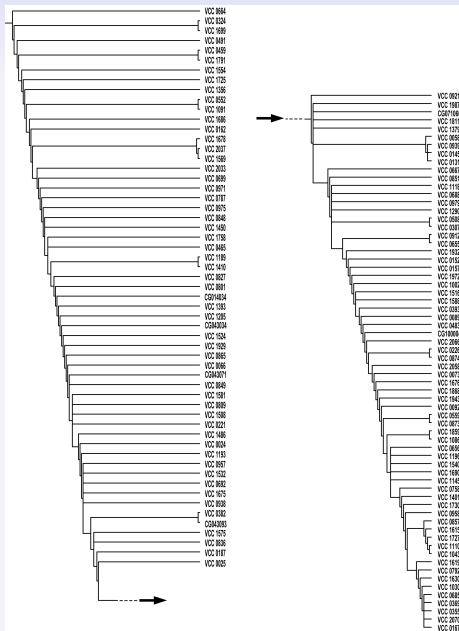
line ratios

kinematics

morphology

Linear tree

Dominated by stellar evolution



Incompatibilities, uninformative parameters creating noise,...

424 galaxies

23 parameters:

- 10 Lick indices: $H\beta$, Fe5015, Mg1, Mg2, Mgb, Fe5270, Fe5335, Fe5406, Fe5709, NaD.
- $[MgbFe]' = \sqrt{Mgb * (0.72 * Fe5270 + 0.28 * Fe5335)}$ indicator of metallicity.
- $Mgb/Fe = Mgb / \frac{(Fe5270+Fe5335)}{2}$ indicator of light-element abundance.
- 6 broad band observables: Brie, Mabs, Kabs, B-R, J-H, and H-K.
- 2 parameters from medium-resolution spectra: $\log \sigma$ and OIII.
- 3 geometrical parameters: $\log(\text{diam})$, $\log r_e$ and D/B.

Fraix-Burnet, Chattopadhyay, Chattopadhyay, Davoust, Thuillard AA, 2012 (hal-00708735)

Objective selection of the parameters

Which parameter subset that can best discriminate different groups?

Principal Component Analysis

axes of greatest variance and correlations between variables

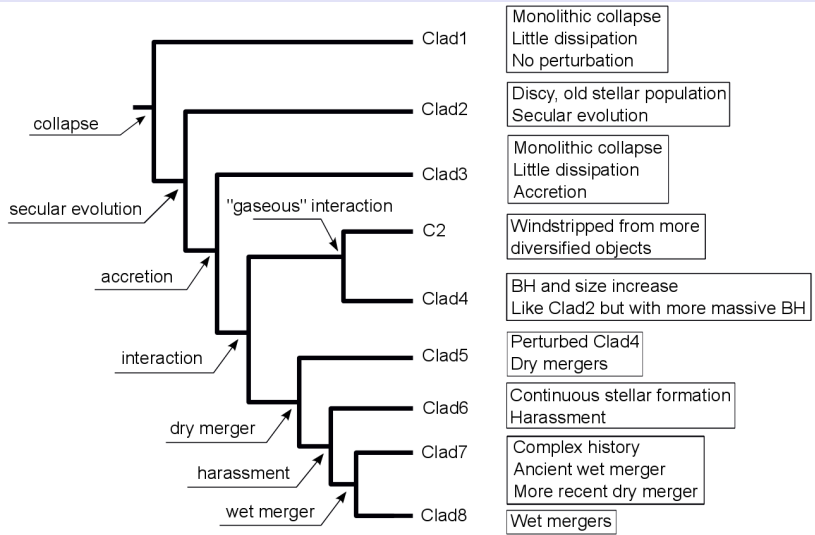
Minimum Contradiction Analysis

best order to arrange the sample on a network or a tree

Maximum Parsimony (Cladistics)

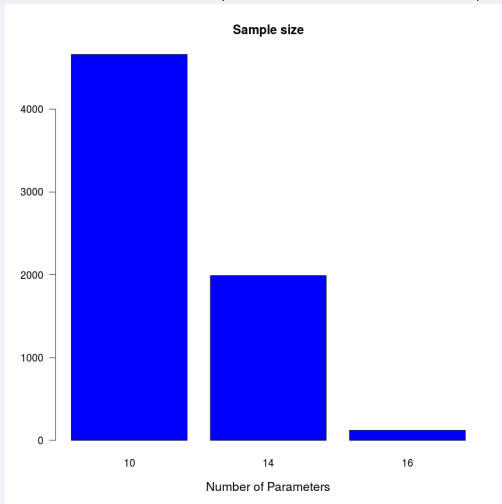
most parsimonious tree representing relationships between groups and objects

Winners: $\log \sigma$ D/B NaD [MgbFe]' Brie OIII

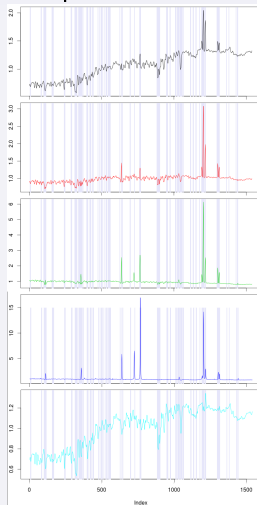


Undocumented parameters ... or too many

WINGS data set (DFB, M. d'Onofrio, P. Marziani)



SDSS spectra (DFB, Bouveyron)



- 1 Unsupervised Classification of Galaxies
- 2 Which Parameters?
- 3 Phylogenetic Tools
- 4 Conclusion

Conclusion

Unsupervised classification of galaxies (and others)

- Exploratory tool
- Insights into data sets and physics
- Needs more investments