Spectral differences between galaxies in low- and high-density environments

Author: Josep Tous Mayol.
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.**

Advisor: Josep Maria Solanes Majúa

Abstract: We are interested in detecting the possible impact of the environment in which the galaxies reside on their spectra. With this aim, we have implemented a stacking technique for the calculation of average spectra of galaxies with high signal-to-noise ratio (S/N) and resolution, as well as an extinction-corrected local density estimator. We use the huge dataset of $\sim 1,500,000$ homogeneous optical spectra reduced and published in the Sloan Digital Sky Survey Data Release 12 inside the limits of the Legacy Survey. Our methodology is tested against the class of galaxies with measured Hα emission, finding important differences in the continuum and in the heights of the principal emission lines.

I. INTRODUCTION

Although galaxies in the universe come in many different forms and sizes, they can be broadly divided into two main categories. Late-type objects (which can be essentially identified with Hubble's S and Irr types), with a flattened disk-like shape, blue colours, and widespread star formation activity, and early-type ones (mainly Hubble's E and S0), having mostly spheroidal or lenticular shapes, red colours, and little or null star formation activity. In the first case, the stellar light is dominated by young massive stars, while in the latter it basically arises from old red-giant and supergiant stars. At the same time, it is well known that the density of galaxies in the universe is not at all constant, but spans from $\sim 0.2\rho_0$ in voids to $\sim 5\rho_0$ in superclusters and filaments, $\sim 100\rho_0$ in the cores of rich clusters, up to $\sim 1000\rho_0$ in compact groups, where ρ_0 is the mean cosmic galaxy density [5].

Galaxy morphology and density are not, however, independent quantities. In his analysis of 55 nearby clusters, Dressler [3] demonstrated that the fraction of spiral galaxies decreases from $\sim$ 60–70% in the field to virtually zero in the cores of rich clusters, a trend that is mirrored, but in the opposite sense, by the lenticular population. This phenomenon, known today as the morphology-density relation, is currently considered one of the clearest observational signatures of the intimate linkage that exists between the properties of galaxies and their evolution with the environment in which they reside.

The aim of the present work is to contribute to the understanding of the origin of this phenomenon by devising a strategy to detect systematic differences between the typical spectra of galaxies as a function of the environment. While most of the studies of this kind make use of photometric data in one or more individual broad wavelength bands, it is quite rare to find works that analyse the wealth of information that is revealed when the light of an object in any of these windows is divided into its component wavelengths. The reason is that this sort of studies have been traditionally limited by the lack of large, systematic spectroscopic surveys of low-redshift galaxies. This situation has changed lately thanks to the advent of multi-object spectrographs, that allow multiple galaxies to be observed in each exposure. The primary example of these surveys is perhaps the Sloan Digital Sky Survey (SDSS) [4], which over the last fifteen years has obtained spectra for more than one million galaxies over approximately 8000 square degrees of the sky. Using this precious mine, we will develop a tool to investigate how the environment affects to Hα galaxies by the stacked spectra of those in low- and high density regions.

The next section is focused on distinguishing between galaxies in low- and high-density environments of a volume-limited subset of the main galaxy sample of the SDSS. Section III is dedicated to the spectral stacking process. In section IV we study the composite spectra of Hα galaxies in different environments. Section V summarizes the highlights and sets out some future prospects. Finally the Appendix shows how we will maximize the number of objects in the volume-limited subset.

II. CALCULATION OF LOCAL DENSITIES

A sample of clean photometric data[1] (equatorial coordinates, redshift, r-band magnitude and its Galactic extinction) from 284160 galaxies up to $z = 0.107$ is downloaded from the SDSS making use of *CasJobs*.

To estimate the local density of a given galaxy, we use the Bayesian density estimator [1], corrected from Galactic extinction [9], required due to the big extension

*Electronic address: jtousmayol@gmail.com

[1] Following the criteria of SDSS about clean photometric data. See *http://www.sdss.org/dr13/algorithms/photo_flags_recommend/*

of the sample in the sky,

$$\mu^{c}_{2k,j} = (10^{0.6\Delta m_j}) \cdot C\Big[\sum_{i=0}^{k} d_i^2\Big]^{-1} \quad (1)$$

with Δm_j the extinction in apparent magnitude on the r-band provided by the survey associated to the target galaxy j, C is a constant set to one as we are interested only in ranking local densities (not in a realistic density estimate), and d_i is the projected distance to object i among those that have a recessional velocity within 1000 km s^{-1} from the target galaxy. We take the five nearest neighbours, so $k = 5$ with index $i = 0$ corresponding to the target itself. This estimate provides a better representation of locality than the traditional kth nearest neighbour metric, which only considers the distance to the kth object.

We note that the definition of locality is something delicate in local density analysis. The metric to use, the number of neighbours and the depth in the recessional velocity with respect to the target are factors that must be chosen with care. For instance, we also calculated the local densities using a recessional velocity of 500 km s^{-1} as many authors do, finding that sometimes this resulted into almost empty shells where the nearest neighbour was really far from the target, so thinner shells actually led to a less local density estimate.

A. Taking care of edge effects

The procedure just delineated in the previous section leads to an underestimate of the local density for the galaxies near the edges of the sample. To solve this problem, we performed a first computation of d_5 for every galaxy in the whole sample. The scale of the probability density function, σ (see Fig. 1), was used to determine the thickness of the band around the edges where density calculation should be avoided (the border region).

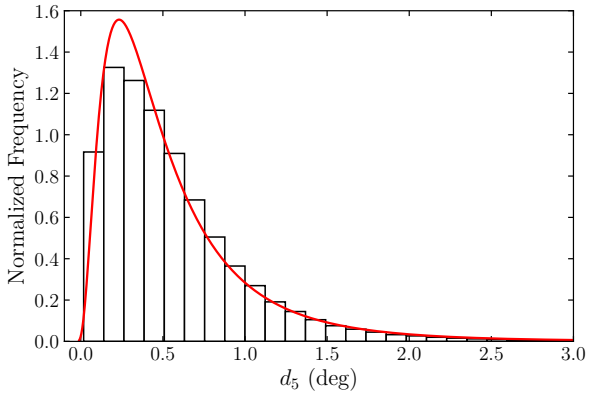

FIG. 1: Distribution of distances to the fifth neighbour. In red, a log-normal probability density function fitted with parameters $(\mu,\ \sigma) = (0.46,\ 0.52)$ deg.

To stablish the region, several cuts were applied to the sample to clean it from some outer isolated galaxies. The sample was divided into vertical and horizontal bands to find the two extreme galaxies of each band. All the galaxies that were found at a distance less than 3σ from the two galaxies in the extremes were flagged as belonging to the border region (see Fig. 2).

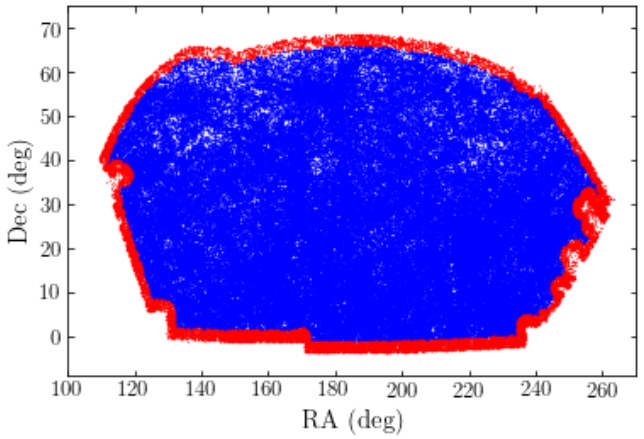

FIG. 2: Sky distribution of the objects in the sample. In red, the galaxies inside the border region, and in blue those whose density could be well determined.

B. Local densities in a volume-limited sample

The SDSS main galaxy sample is flux-limited with an r-band magnitude limit of $m_r = 17.77$ mag [10]. Note that if we compute the density of the whole magnitude limited sample, we would be introducing a Malmquist bias (faint galaxies would be under-represented at large distances). Therefore, we must cut the sample in the radial coordinate to transform it into a volume-limited one, establishing a minimum absolute magnitude detectable inside the volume. The calculation of the local density assumes that the luminosity function of galaxies is universal.

We calculate r-band absolute magnitudes approximating radial distances by $d \approx cz/H_0$ with $H_0 = 70$ km s^{-1} Mpc^{-1} as all our galaxies are nearby ($z < 0.1$). Then, we divide the sample into volume-limited subsets in steps of 0.5 mag. Each sub-sample is characterised by its limits in absolute magnitude and redshift. We select

N	37800	81837	**143526**	125450	83130	56642	38064
$\boldsymbol{M_r}$	-21.5	-21.0	**-20.5**	-20.0	-19.5	-19.0	-18.5
$\boldsymbol{z}$	0.165	0.131	**0.104**	0.083	0.066	0.052	0.042

TABLE I: Number of galaxies, upper limits in absolute magnitude and redshift for the 7 largest volume-limited subsamples.

the most populated sub-sample, i.e. the associated to

$M_r = -20.5$ (Table I). To avoid edge effects along the line of sight, we increase the volume of the sub-sample in 1000 km s^{-1} in the two extremes of the radial coordinate, flagging these extra galaxies as border objects.

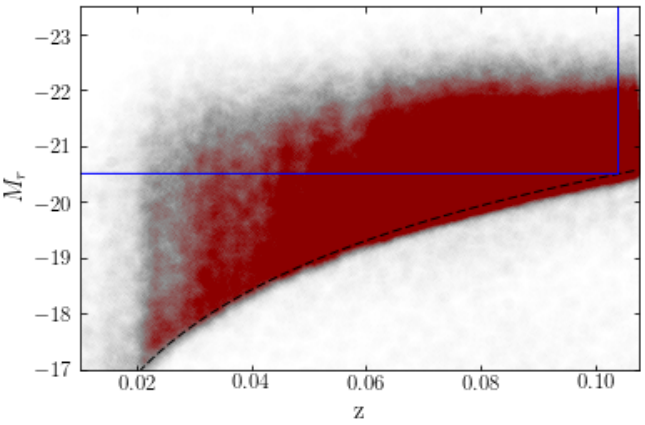

FIG. 3: Scatter plot of galaxy absolute magnitude in the r-band as a function of redshift. The dashed line shows the faint-end limit at $m_r = 17.77$ mag. The solid blue lines show the limits of the largest volume-limited sub-sample.

Finally, we calculate the local density of the galaxies in the sub-sample as explained in section II. The local density follows a log-normal distribution, with a first quartile ($Q1$) defined by $\mu^{c}_{2k} < 0.47$ deg^{-2} as the low-density environment subset and the last quartile ($Q4$) with $\mu^{c}_{2k} > 4.42$ deg^{-2} defining the high-density environment subset.

III. COMPOSITE SPECTRA

In this section the stacking technique is explained following the detailed steps in [7] adapted to our requirements.

Firstly, the spectra are read from the files downloaded from the SDSS going pixel by pixel and, for those that have the mask turned on, i.e. the ones with untrustworthy values, we interpolate their fluxes and errors from the surrounding pixels.

All the spectra have to be shifted to the same rest-frame so the wavelength is given by $\lambda_r = \lambda_{obs}/(1+z)$, where λ_r is the rest-frame wavelength and λ_{obs} is the observed wavelength. The flux and its error are re-binned into pixels of 0.5 Å in the wavelength range of (3400 − 9000) Å. The size of the re-binned pixels respects the mean resolution of the spectrograph in the visible [4]. This re-binning is done through interpolation.

As every galaxy is at a different distance, their fluxes cannot be directly compared before normalizing each individual spectrum. Hence, a normalisation factor is assigned to each spectrum, choosing an arbitrary value considered the continuum. This selection must be done after observing a reduced sample of the studied objects, looking for smooth wavelengths ranges where no emission nor absorption characteristic line are observed. We take the values from [2], (4200−4300, 4600−4800, 5400−5500 & 5600 − 5800) Å. The factor is

$$n_j = \sum_i \frac{f_{ij}}{N_j} \tag{2}$$

where f_{ij} is the flux in pixel i of the j spectrum and the sum runs for the N_j pixels in the four intervals conforming the continuum.

Then, S/N, computed for the same four wavelength ranges, is assigned to each spectrum,

$$s_j = \frac{\sum_i f_{ij}/N_j}{\left(\sum_i e^2_{ij}/N_j\right)^{1/2}} \tag{3}$$

here e_{ij} is the error linked to f_{ij}. A weight is given to each spectrum depending on its quality,

$$\omega_j = \frac{1}{s_j^{-2} + \sigma^2} \tag{4}$$

with σ a correction factor introduced to ensure a more trustworthy representation of the whole galaxy sample taking into account a certain variability between its members. For the time being, we take $\sigma = 0.1$, the same arbitrary constant value used by [7] (but see section V).

Finally, the flux for each re-binned pixel i is:

$$\bar{f}_i = \frac{\sum_j \omega_j (f_{ij}/n_j)}{\sum_j \omega_j} \tag{5}$$

IV. STUDY OF Hα EMISSION GALAXIES

Here we elaborate the composite spectra of galaxies in the two extreme quartiles, at $z < 0.1$, whose flux in Hα is greater than five times its error. These objects are expected to be a sensitive probe for differences on the environment.

From the main sample in SDSS, we find that 22835 galaxies with clean spectra[2] of this class belong to $Q1$ and 15007 belong to $Q4$. We compute the stacked spectra of each quartile as explained in the previous section. The result shown in Fig. 4. We note that both spectra are very similar to the one obtained in [2] for the same galaxy class.

The first remarkable difference between the two spectra is on the emission lines. The emissions are caused by clouds of ionized gas where star formation takes place. In fact, the Hα Balmer emission

[2] We use the flag *zwarning* set to 0 meaning that spectra have no known problems. See http://www.sdss.org/dr12/algorithms/bitmasks/#ZWARNING

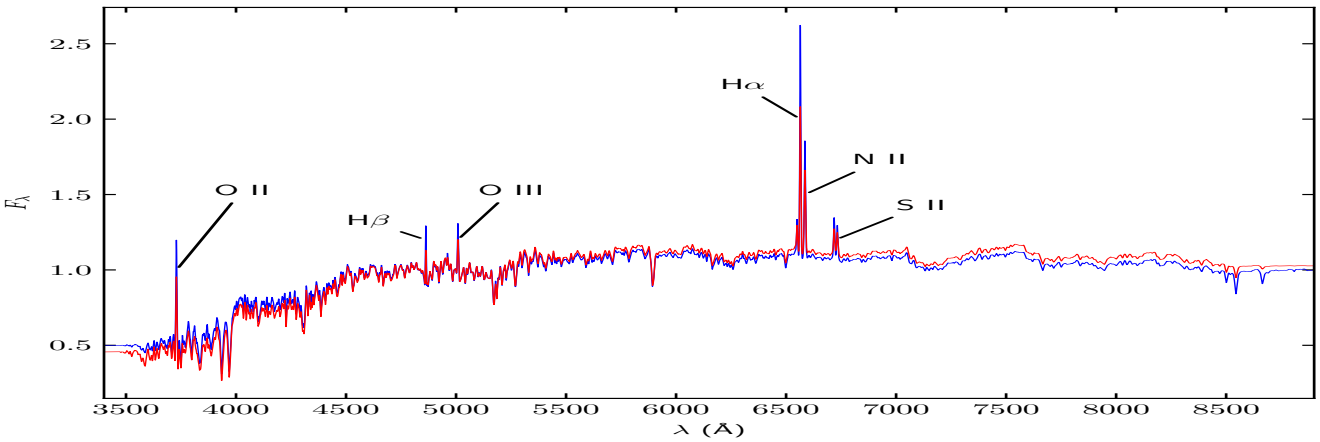

FIG. 4: Comparison between the composite spectra from the two galaxy subsets with detected Hα emission. The red line is for galaxies in high-density regions ($Q4$) and the blue line is for field galaxies ($Q1$). The most visible emission lines have been labeled.

($\lambda = 6562.8$ Å) is the most direct indicator of the current ($< 4\times10^6$ yrs), massive ($> 8M_\odot$) star formation activity in galaxies [6]. However, these clouds predominate in the galaxies in low-density regions populated by young stars of the type O and B that contribute to the Balmer series in absorption. Therefore, Balmer emission appear superimposed to the stellar absorption lines. This effect grows in importance towards the higher order Balmer lines, so that while Hβ emission ($\lambda = 4862.7$ Å) is moderately affected by the absorption, all of the Hδ emission (4102.9 Å) is usually lost into the absorption [8]. Of the two spectra in Fig. 4, the blue one has higher emission lines due to a greater presence of star formation in such a way that is less affected by the phenomenon explained above. In Fig. 5, one observes that the small emission Hδ only survives for the galaxies in $Q1$.

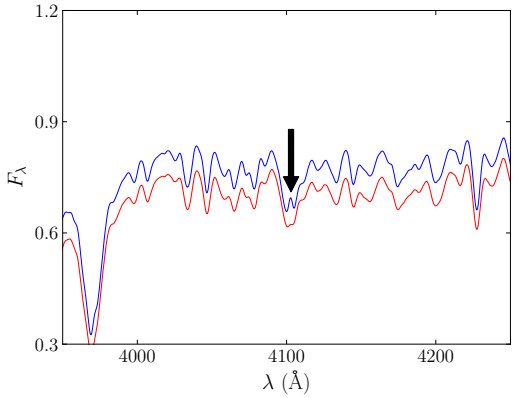

FIG. 5: Detail of the Hδ emission from Fig. 4.

The second important difference lies on the continuum. A greatest presence of red giants and a lesser abundance of hot blue stars in the galaxies of $Q4$ makes the break at 4000 Å to be slightly stronger for the high-density class. The same explanation justifies why for large wavelengths the continuum of $Q4$ is placed above the continuum of $Q1$.

V. SUMMARY AND FUTURE WORK

- We have implemented a methodology that allows us to study the imprint left by the environment on the optical spectrum of galaxies. The method, which is particularly suitable for the new generation of surveys based in MOS spectroscopy that contain a vast quantity of spectral data, requires the calculation of extinction-corrected local volume densities in a volume-limited subset of galaxies free of edge effects, as well as the creation of high S/N composite spectra after the shifting, re-binning, normalizing, weighting and stacking of the individual spectra of galaxies belonging to a given class.
- Our technique have been applied to galaxies with Hα emission in the SDSS Legacy Volume. We have found noticeable differences on both the two extremes of the continuum in the optical window and on the heights of the most important emission lines. A first analysis of these differences indicates that they are consistent with a reduction in the star formation activity for the galaxies that inhabit the densest environments with respect to the Hα galaxies in the lowest density regions.
- In addition, we have included an Appendix describing the basic steps of a procedure that we have devised to maximize the size of the volume-limited samples that can be extracted from magnitude-limited catalogues. It requires the comparison of the local densities of galaxies in the overlapping volumes of two adjacent regions large enough to be representative of the universe.

All these processes have been implemented through Python scripts. In particular, the libraries for astron-

omy *Astropy* and *pydl* have been very useful to handle the spectra stored in the *.fits* files downloaded from the SDSS Sky Server.

Some future work would be to improve the stacking process by looking for a weighting factor (4) specific for our sample and to add an error estimate associated to the stacked spectra, which would allow us to determine the significance of the differences between both spectra. We also intend to perform a quantitative study of the composite spectra through Principal Component Analysis. The ultimate goal, will be to apply this methodology to the study of lenticular galaxies for the purpose of trying to constrain their formation path.

VI. APPENDIX: MAXIMIZATION OF VOLUME-LIMITED SAMPLES

Here we describe the basic idea behind the process to maximise the available galaxies in a volume-limited sub-sample. It might result necessary if one studies peculiar morphologies like for example S0 galaxies which are not very numerous on the literature. The main idea is to use as many volume-limited sub-samples as possible shifting each one to the largest one which we call the reference (S_0), i.e. the one with $M_r = -20.5$, scaling the density of every galaxy.

For example, the very next sub-sample to S_0 is $M_r = -20.0$ (S'). We find that some galaxies are shared by the two sub-samples and therefore each of these has associated two values of density given by (1): $\mu_{0,j}$ from S_0 and μ'_j from S' (now j refers to every galaxy in S'). For some objects the two values of the density may be different but for some others may not. Hence, plotting $\mu_{0,j}/\mu'_j$ vs. μ'_j we should be able to fit a law which leads to the desired scaling function to shift the galaxies in S' to S_0.

In Fig. A1, we represent the relationship of the densities between the two subsets where the points are binned along the horizontal logarithmic axis. Each bin is represented by the red dots corresponding to the medians of the ordinates and the abscissae as the distribution seems to be very asymmetric. The error bars are given by the interquartile distances of each bin. This yields the following linear regression with $r^2 = 0.95$:

$$\frac{\mu_{0,j}}{\mu'_j} = -0.06 \log_{10} \mu'_j + 0.68 \qquad \text{(A1)}$$

where the right side of the equality is the factor for which multiplying the density of an object in S' one can obtain its scaled density in S_0. Then, we would repeat the same steps for the subsequent volume-limited subsets, always between consecutive. If one pretends to bring the sub-sample with $M_r = -19.5$ (S'') to the reference S_0, the described process would be applied between S'' and S' so, the second one, would include the galaxies of both sub-samples and then between the actualised S' and S_0.

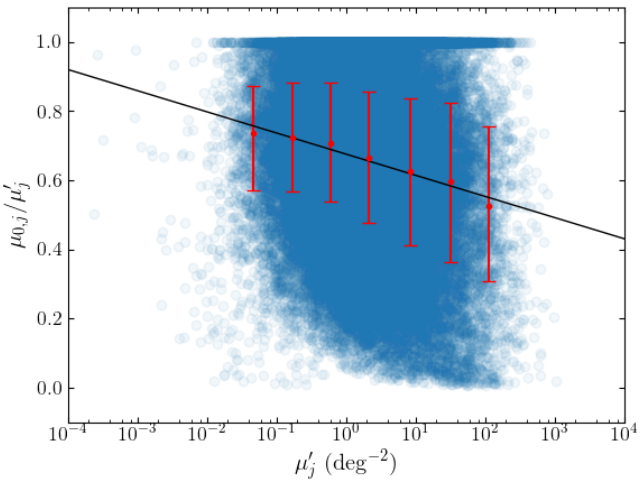

FIG. A1: Relationship between the local densities in the two consecutive volume-limited subsets corresponding to $M_r = -20.5$ and $M_r = -20.0$.

Acknowledgments

I would like to thank Josep Maria Solanes for his endless patience in my moments of panic. Jaime D. Perea for introducing me to Python and for his precious knowledge about spectral analysis. Also Lluís Mas for advising me with the stacking process and for his detailed answers to my e-mails. Finally, my family for backing me up.

[1] Cowan, N. B., & Ivezic, Z. 2008, ApJL, 674, L13

[2] Dobos, L., Csabai, I., Yip, C.-W., et al. 2012, MNRAS, 420, 1217

[3] Dressler, A. 1980, ApJ, 236, 351

[4] Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

[5] Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

[6] Kennicutt, R. C., Jr. 1998, ARA&A , 36, 189

[7] Mas-Ribas, L., Miralda-Escudé, J., Pérez-Ràfols, I., et al. 2016, arXiv:1610.02711

[8] Olofsson, K. 1995, A&AS , 111, 57

[9] Solanes, J. M., Giovanelli, R., & Haynes, M. P. 1996, ApJ, 461, 609

[10] Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810