Investigation of the Effect of Bars on the Properties of Spiral Galaxies: A Multivariate Statistical Study

Prasenjit Banerjee[1*] , Tanuka Chattopadhyay[2†] , Asis Kumar Chattopadhyay[1‡]
[1]*Department of Statistics, University of Calcutta, 35, Ballygunge Circular Road, Kolkata-700019, India*
[2]*Department of Applied Mathematics, University of Calcutta, 92, A.P.C Road, Kolkata-700009, India*

Article Information

Article history:

Received XX YY ZZZZ
Received in revised form XX YY ZZZZ
Accepted XX YY ZZZZ
Available online XX YY ZZZZ

Keywords:

Galaxies: Fundamental parameters,
Galaxies: Spirals,
Method: Statistical,
Astronomical Databases: Miscellaneous

Abstract

Subjective classification of spiral galaxies is not sufficient for studying the effect of bars on their physical characteristics. In reality the problem is to comprehend the complex correlations in a multivariate parametric space. Multivariate tools are the best ones for understanding this complex correlation. In this work an objective classification of a large set (26,089) of spiral galaxies was compiled as a value added galaxy catalogue from sdss DR 15 virtual data archive.

Initially for dimensionality reduction, Independent Component Analysis is performed to determine a set of Independent Components that are linear combinations of 48 observed features (namely ionised lines, Lick indices, photometric and morphological properties). Subsequently a K-means cluster analysis is carried out on the basis of the 14 best chosen Independent Components to obtain 12 distinct homogeneous groups of spiral galaxies. Amongst these, 3 groups are the oldest ones (1.6 Gyr - 5.9 Gyr), while 5 groups fall in the medium aged category (1.4 Gyr - 1.6 Gyr), 2 groups consist of only unbarred spirals, 1 group is the youngest one and the remaining one is an outlier. In many groups there are clear indication of recurrent bar formation phenomena which is consistent with few previous simulation works. In order to study the robustness a second method of clustering by Gaussian Mixture Modeling Method (GMMBC) is applied.

1. Introduction

Effect of bars on the properties of spiral galaxies is an interesting topic as several studies have shown that many important properties like star formation activity, color, metallicity etc. may depend on the strength of bar (Vera et al. (2016); Athanassoula (1983); Buta and Combes (1996); Combes and Elmegreen (1993); Martin (1995); Ellison et al. (2011); Zhou et al. (2014)). Many other works (Debattista and Sellwood (1998); Weinberg (1985); Debattista and Sellwood (2000); Athanassoula (2003); Erwin (2019); Kim et al. (2020a); Garma-Oehmichen et al. (2020)) show by numerical simulations that bars can effectively transport gas from the outskirt towards the central regions of the barred galaxies. Subsequently this gas undergo interaction with the edges of the bar which produces shock waves. This shocked gas looses angular momentum which accentuates the flow of gas towards the central region and thus produces

*Email id: b.prasenjit1994@gmail.com , Telephone number: (+91)9433977094
†Email id: tchatappmath@caluniv.ac.in , Telephone number: (+91)9433648255
‡Email id: akcstat@gmail.com , Telephone number: (+91)9831985850

starburst. Some works show that bars can be destroyed by the presence of large central mass (Roberts Jr et al. (1979); Norman et al. (1996); Sellwood and Moore (1999); Athanassoula et al. (2005); Spinoso et al. (2016); Barbuy et al. (2018); Guo et al. (2020); Rosas-Guevara et al. (2020)). This theory indicates that many non-barred disk galaxies might had bars in the past and thus presence and absence of bars are nothing but a recurrent phenomenon of galaxy life (Bournaud and Combes (2002); Berentzen et al. (2004); Gadotti and de Souza (2006); Pettitt and Wadsley (2018); Katz et al. (2018); Hilmi et al. (2020)). Many authors have established that inflow of gas is an efficient mechanism for triggering active galactic nuclei (AGN) and they form bulges or pseudo bulges (Kormendy and Kennicutt Jr (2004); Debattista et al. (2005); Debattista et al. (2006); Martinez-Valpuesta et al. (2006); Aguerri and González-García (2009); de Lorenzo-Cáceres et al. (2019); Fragkoudi et al. (2020); Barbuy et al. (2018)).

In order to investigate the relation between bars and host galaxy colors, different studies inferred that bars are frequently found in late type spiral galaxies those are bluer and less concentrated systems (Barazza et al. (2008); Aguerri and González-García (2009)). On the contrary other studies found an excess of barred galaxies with redder colors from different samples (Masters et al. (2010); Lintott et al. (2011); Oh et al. (2011); Alonso et al. (2013); Alonso et al. (2014); Vera et al. (2016); Kim et al. (2020b); Cuomo et al. (2019)).

In connection with star formation activity, many researchers indicate that presence of bars enhance star formation rate (SFR) (Hawarden et al. (1986); Devereux (1987); Hummel et al. (1990)), while several others show that bars do not guarantee increase in star formation activity (Pompea and Rieke (1990); Martinet and Friedli (1997); Chapelon et al. (1999); Donohoe-Keyes et al. (2019); Wang et al. (2020); Kim et al. (2017); Newnham et al. (2020)). Similar controversial results are obtained in case of metallicity also (Vila-Costas and Edmunds (1992); Martin and Roy (1994); Ellison et al. (2011); Sánchez-Blázquez et al. (2014)).

The above studies are mostly empirical and drawn conclusion either from simulation studies or from a control data sets as a whole, which are prepared in various ways. As the galaxy properties are inter dependent, the corresponding parameters constitute a multivariate set up and for data analysis multivariate techniques have to be used. Some authors studied the barred galaxy properties by means of bar properties but that is a subjective classification (Vera et al. (2016); Carles et al. (2016); Kruk et al. (2018); Seo et al. (2019); Cavanagh and Bekki (2020)) which is related to algorithm that is trained on label data. The labels arise out of pre-established classifications on very few features, or in other words the algorithm is trained to find results of human subjectivity (eg. Zooniverse *).

In the above context one is tempted to apply statistical (unsupervised) classification e.g. a multivariate partitioning analysis to explore the homogeneous groups of galaxies, not only focusing one or few particular aspects of the physics of galaxies but also by exploring the cluster structure of the data set (Chattopadhyay et al. (2019); Fraix-Burnet et al. (2015)). One basic tool is Principal Component Analysis (PCA). This is used by various authors (Whitmore and Forbes (2012); Cabanac et al. (2002); Chattopadhyay and Chattopadhyay (2006); Peth et al. (2016)). The main target is the reduction of the dimensionality (i.e. number of parameters, here). But use of PC to perform a Clustering (unsupervised) is not recommended since the components with largest eigenvalues are the axes of maximum variance and those are generally not the most discriminative ones to reveal the cluster structure (Chang (1983)).

For unsupervised classifications some attempts have been made by K-means cluster analysis on the basis of all the parameters (Ellis et al. (2005); Chattopadhyay and Chattopadhyay (2007); Chattopadhyay et al. (2007); Mondal et al. (2008); Chattopadhyay et al. (2009); Babu et al. (2009); Almeida et al. (2010); Fraix-Burnet et al. (2010); Fraix-Burnet et al. (2012); De et al. (2016);Modak et al. (2017); Modak et al. (2020)).

Partitioning of objects into robust groups will be more prominent when the features are independent. For this, Independent Component Analysis is used which is applicable to a non-Gaussian data set like the present one for dimensionality reduction. On the basis of the Independent Components (ICs), which are the linear combinations of various observable properties of the spiral galaxies e.g. broad-band line fluxes (magnitudes), slopes (colors), medium band line fluxes (Lick indices), the large data set can be classified (unsupervised) into various homogeneous groups. Then the groups are studied with the help of other estimated properties of the bars, SFR, metallicity, age along with the observed ones.

*https://www.zooniverse.org/

In this study we have prepared a large data set of spiral galaxies from sdss DR15 and cross-matched with it the corresponding bar properties retrieved from Zooniverse (Galaxy-Zoo). The work contains several novelties for unsupervised classification as follows:

- A large data set of spiral galaxies from sdss DR15, cross matched with zooniverse.
- A large number of observable features.
- A large number of estimated features.
- The application of ICA.
- The justification of ICA to the fact that the present data set is non Gaussian.
- Robustness checking by a widely applicable method like Gaussian Mixture Modelling Method (GMMBC).

In most of the previous studies authors retrieved data from sdss and they performed supervised classification. ICA has been used widely for source separation (Pires et al. (2006); Pike et al. (2017); Martins-Filho et al. (2018); Sheldon and Richards (2018)) and dimensionality reduction (Richardson et al. (2016); Sarro et al. (2018)) but rarely for unsupervised classification (Mu (2007); Das et al. (2015); Modak et al. (2017); Modak et al. (2020)).

The paper is organised as follows: A brief description of the data set is given in Section 2. The methods are descried in Section 3. The results and discussion are included in Section 4. Finally, Section 5 concludes the summary.

2. Data description

The present data is a cross matched collection of galaxy catalogues, used for the study of galaxy formation and evolution. It is based on the sdss Data Release 15 (sdss DR15) [†].

2.1. Data Preparataion

The present value added galaxy catalogue of spiral galaxies has been compiled in the following manner:

1) The Baldwin, Phillips & Telervich diagram (Baldwin et al. (1981); Veilleux and Osterbrock (1987); Kauffmann et al. (2003); Kewley et al. (2006); Kewley et al. (2013); Here after, BPT diagram) (include reference) parameters of the galaxies (viz. H_β Flux, H_α Flux, Oiii 5007 Flux, Nii 6548 Flux, etc.) are collected by joining the catalogues GalSpecLine on SpecObj through the spectroscopic object id.
2) The Spectral Lick Indices of the galaxies (viz. Lick_Nad, Lick_Ca 4227, Lick_g 4300, Lick_Fe 4383 etc.) are retrieved by joining the catalogues GalSpecIndx on SpecObj through the spectroscopic object id.
3) The Continuum substracted emission EW parameters of the galaxies (viz. σ-forbidden, Oii 3729 Reqw, H_β Eqw, H_α Reqw etc.) are collected by joining the catalogues GalSpecLine on SpecObj through the spectroscopic object id.
4) The Star formation rates and the specific star formation rates of the galaxies are collected by joining the catalogues GalSpecExtra on SpecObj through the spectroscopic object id.
5) The Photometric properties of the galaxies (viz. u, g, r, i, z, petroR90_r, petroR50_r etc.) are collected from the Galaxy catalogue.
6) Parameters such as metallicity, logmass, age etc. are collected by joining the catalogues stellarMassStarforming-Port on SpecObj through the spectroscopic object id.
7) The Velocity dispersion the galaxies are collected by joining the catalogues GalSpecInfo on SpecObj through the spectroscopic object id.
8) The other parameters of the galaxies such as sersic indices, U magnitude etc. are collected from the nsatlas catalogue.
9) The Spiral properties of the galaxies (viz. p_cw, p_acw, p_edge, p_cs etc.) are obtatined by joining the catalogues zooSpec on SpecObj through the spectroscopic object id.

Now based on the celestial coordinates (RA,DEC) of the galaxies these datasets are cross-matched [‡] amongst

[†] https://skyserver.sdss.org/dr15/en/tools/search/sql.aspx
[‡] http://cdsxmatch.u-strasbg.fr/

themselves to obtain our master catalogue. For cross matching we have used the 'By position' cross match criteria and the radius is taken as 1 arcsec [§]. The cross match is done throughout the sky, not on the basis of any cone or healpix cells [¶]. Thus the present data set contains only spiral galaxies of different spectroscopic sub classes (viz, UNDEFINED, AGN, AGN_BROADLINE, BROADLINE, STARBURST, STARBURST_BROADLINE, STARFORMING, STARFORMING_BROADLINE etc.). The number of entries are 26,089. Further these master catalogue is crossed with the Hoyle Bar length Catalogue [‖], containing 3150 galaxies. After cross matching the number of unbarred spiral galaxies are 24,320 and the number of barred spirals are 1769. In the present data set the barred spirals are denoted by '1's and the unbarred are denoted by '0's.

We have limited the number of parameters to keep the computation tractable and to reduce noise. Redundant properties such as Sersic profile in different bands have been eliminated since they more or less bear the same information and only a few photometric bands and colors have been selected. Our data set consists of low redshift i.e. $z \leq 0.06$. We have kept most of the important physical information in our data set, so that it did not greatly impact our analysis based on dimentionality reduction through ICA. Now, our final data set consists of 48 parameters which covered spectroscopy, photometry, chemical composition, morphology, and kinematics. All these attributes are described in Table 1 below and details are available on the sdss website [**].

Table 1

Detail description of the parameters of present data set collected for study.

Parameters	Description	Parameters	Description
Lick_nad	Stellar absorption line (Lick) index lines	h_delta_eqw	The equivalent width for absorption
Lick_cn2	”	h_gamma_eqw	”
Lick_ca4227	”	h_beta_eqw	”
Lick_g4300	”	h_alpha_eqw	”
Lick_fe4383	”	U	Absolute magnitude (log of intensity)
Lick_ca4455	”	G	”
Lick_fe4531	”	R	”
Lick_c4668	”	I	”
Lick_hb	”	Z	”
Lick_fe5015	”	J	”
Lick_mgb	”	H	”
Lick_fe5270	”	K	”
Lick_fe5335	”	$u - g$	Apparent magnitude(u) minus Apparent magnitude (g)
Lick_fe5406	”	$g - r$	Apparent magnitude(g) minus Apparent magnitude (r)
Lick_fe5709	”	$r - i$	Apparent magnitude(r) minus

[§] http://cdsxmatch.u-strasbg.fr/xmatch/doc/CDSXMatchDoc.pdf
[¶] http://adass2010.cfa.harvard.edu/ADASS2010/incl/presentations/O01_2.pdf
[‖] https://data.galaxyzoo.org/
[**] http://skyserver.sdss.org/dr15/en/help/browser/browser.aspx#&&history=shortdescr+Tables+U

			Apparent magnitude (i)
Lick_fe5782	”	$i-z$	Apparent magnitude(i) minus Apparent magnitude (z)
Lick_hd_a	”	$u-z$	Apparent magnitude(u) minus Apparent magnitude (z)
$D_n(4000)$	Break in the spectrum at 4000Å	$J-H$	Magnitude(J) minus Magnitude (H)
sigma_balmer	Velocity dispersion (σ not FWHM) measured simultaneously in all of the Balmer lines in Km s^{-1}	$H-K$	Magnitude(H) minus Magnitude (K)
sigma_forbidden	Velocity dispersion (σ not FWHM) measured simultaneously in all the forbidden lines in Km s^{-1}	**The following parameters have been used for further study after finding the homogeneous groups**	
oii_3729_reqw	The equivalent width of the continuum-subtracted emission line with the other emission lines subtracted off (EW_stellar = REQW - EQW)	Metallicity	Metallicity of best fit template (5 categories : 0.004, 0.01, 0.02, 0.04, or "composite")
neiii_3869_reqw	”	z	Redshift
oiii_5007_reqw	”	SFR ($M_\odot yr^{-1}$)	Star-formation rate of best fit
hei_5876_reqw	”	$\log(M_*)$ ($M_\odot$)	Best-fit stellar mass of galaxy
oi_6300_reqw	”	Age (Gyr)	Age of best fit
h_alpha_reqw	”	un(0)barred(1)	Indicator variable for barred(1) and unbarred(0) galaxies
nii_6584_reqw	”	**For barred galaxies**	
sii_6731_reqw	”	length_scaled	The ratio of bar length to galaxy size
c	Concentration index: Sersic_r90_R / Sersic_r50_R	length_avg	The average bar length scatter per observer, averaged over galaxies being observed

The initial 48 parameters in the above table are used for the data analysis and the rest of the parameters, mentioned above, have been used for further study.

3. Statistical Methods

Thus, we are left with a data set containing 26,089 spiral galaxies with 48 different parameters for statistical analysis. The data set is quite large for computational tractability. Therefore we use the dimensionality reduction (ICA) technique to reduce the size of the data without loosing any vital information from it and subsequently we have clustered the data set to observe any coherent groups present in it.

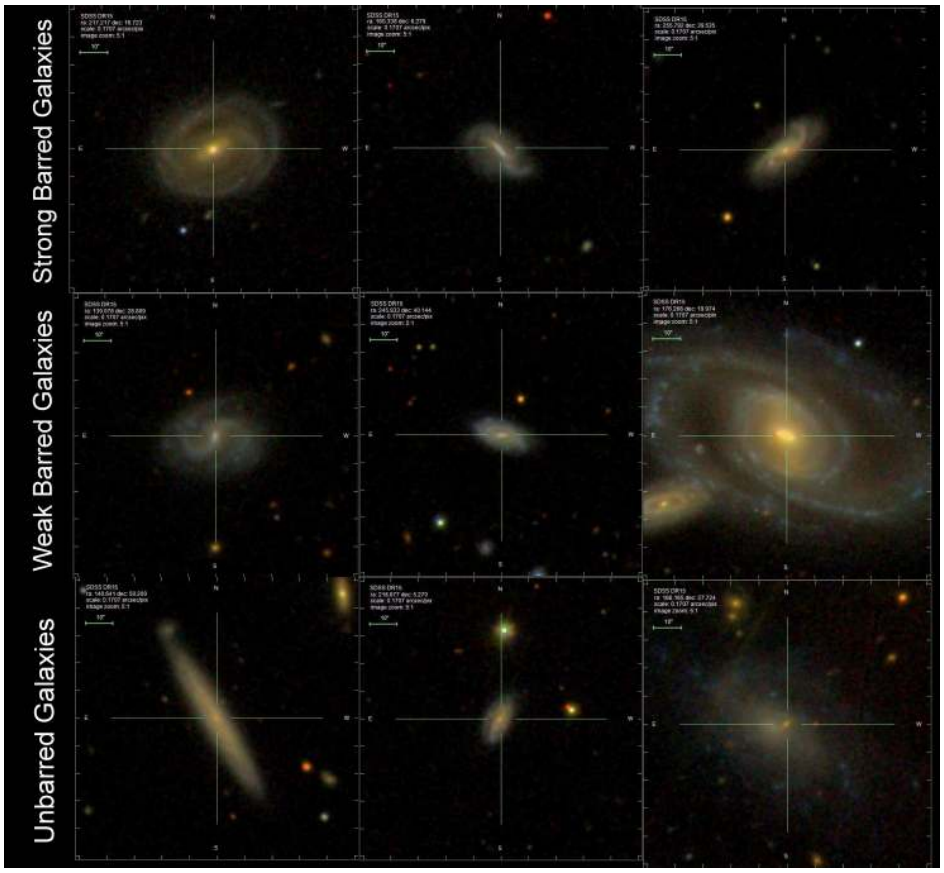

Fig. 1: Images of typical examples of galaxies used in our data set classified as strong barred, weak barred and unbarred galaxies.

3.1. Test for Gaussianity

There are several way for testing the Gaussianity of the data set, such as Shapiro-Wilk test (Shapiro and Wilk (1965)), Anderson Darling Test (Stephens (1974)), Kolmogorov-Smirnoff test (Lilliefors (1967)) etc. Here we have used the Shapiro-Wilk Test for checking the Gausianity of our data set, any other method could have easily opted for. In this technique the test statistics is defined as

$$\mathrm{W} = \frac{\sum_{i=1}^{n} a_i x_i^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

n being the number of observations, x_i's are the ordered sample values and a_i's are the constants generated from the order statistics of a sample from a normal distribution. As our data set is multivariate in nature, a multivariate extension of the Shapiro-Wilk test (Villasenor Alva and Estrada (2009)) have been used. The p-value of the test comes out to be 2.2×10^{-16}, which is quite small. Hence we are more inclined in rejecting the null hypothesis, that our data is from a Gaussian setup i.e. the data set is found out to be Non-Gaussian in nature.

3.2. Independent Component Anlysis

PCA has been applied by many authors (Brosche (1973); Whitmore and Forbes (2012); Murtagh and Heck (2012), etc) for several purposes but it is not appropriate for clustering and classification (Chang (1983)). Moreover, one of the inherent feature in PCA is that, the data set should be a Gaussian data, but our data set is a non-Gaussian one. ICA is a dimension reduction technique, i.e., it reduces the number of observed parameters p to a pre-defined number m (where, $\mathrm{m} \ll \mathrm{p}$) of new variables (here the significant IC components). This technique is mainly applicable to non-Gaussian

setup (Hyvärinen (1998); Hyvärinen (1999a); Hyvarinen (1999b); Pfister et al. (2019)). Another basic difference between ICA and PCA is that, in PCA the components are assumed to be uncorrelated but not independent where as in case of ICA the components are assumed to be mutually independent amongst each other. For further details regarding the comparison between these two one can consult Section 3 of Chattopadhyay et al. (2013a), and references therein.

3.2.1. Method: Independent Component Analysis

let $X_1, X_2, X_3, ..., X_p$ be p random vectors (here, p = 48) and n (here, n = 26,089) be the number of observations for each X_i, (i = 1, 2, 3, ..., p).

Let $X = AS$, where, S=$\{S_1, S_2, S_3, ..., S_p\}'$ is a random vector of hidden components S_i's, (i = 1, 2, 3, ..., p). A is a non-singular matrix, also known as the mixing matrix. S_i's are mutually independent amongst themselves. The objective of ICA is to find S by inverting A, i.e., $S = A^{-1}X = WX$, where, $A^{-1} = W$. W is called the unmixing matrix as it is the inverse of A, the mixing matrix. ICA separates the Independent Components (ICs) (sources) present in a mixture (Comon (1994); Chattopadhyay et al. (2013b)). To obtain independence, the non-Gaussianity of the data is maximised using negentropy. There are several techniques for ICA such as FastICA, ProDenICA (Hastie and Tibshirani (2003)), KernelICA etc. One of them is FastICA algorithm (Hyvärinen and Oja (2000)). In this method the ICs are estimated one by one. This algorithm converges very fast and is very reliable. It is the most commonly used algorithm and is also very easy to use.

There is no good method available for the determination of the optimum number of ICs. We generally choose it by using the optimum number of PCs (irrespective of the data being non-Gaussian) (Albazzaz and Wang (2004); Chattopadhyay et al. (2013b); Eloyan and Ghosh (2013)), to find m (m$\ll$p) (Chattopadhyay and Chattopadhyay (2007); Babu et al. (2009); Fraix-Burnet et al. (2010); Chattopadhyay et al. (2010); Chattopadhyay et al. (2013b)). Another novel criterion of choosing the optimal number of ICs is Maximally Stable Transcriptome Dimension (MSTD) (Kairov et al. (2017)).

This technique depends on a fundamental parameter M (effective dimension of the data, here 48), as well as the number of ICs computed) whose effects are being investigated on the stability of the ICs. The range of M values are from M_{min} (here, 2) to M_{max} (here, 40). For each M ranging from M_{min} to M_{max}, the data dimension is being reduced to M by PCA and then data has been whitened. Afterwards, in the whitened space the actual signal decomposition is applied by defining M new axes. Each of them maximize the non-Gaussianity of data point projections distribution (Fig. 2).

The algorithm for determining the MSTD:

1) Define two numbers M_{min} and M_{max} which denote the maximal and minimal possible numbers of computed independent components, respectively.

2) Define a number T (here, T=100). It denotes the number of ICA runs for estimating the components stability.

3) For each M ranging between M_{min} and M_{max} :

   3.a) Find out the M, ICs using the fastICA algorithm and iterate it for T times. Thus we will get a data set of $M \times T$, ICs.

   3.b) Now, cluster the newly formed $M \times T$, components into M clusters using the agglomerative hierarchical clustering algorithm, where the measure of dissimilarity being 1- $|r_{ij}|$. r_{ij} is the Pearson's correlation coefficient between the components.

   3.c) For each cluster C_k out of M clusters (C_1, C_2,. . . , C_M), the stability index is obtained, using the formula mentioned below:

$$I_q(C_k) = \frac{1}{|C_k|^2} \sum_{i,j \in C_k} |r_{ij}| - \frac{1}{|C_k| \sum_{l \neq k} |C_l|} \sum_{i \in C_k} \sum_{j \notin C_k} |r_{ij}|$$

   where, $|C_k|$ denotes the size of the k^{th} cluster.

3.d) The Average Stability Index (ASI) for M clusters is now given by:

$$S(M)=\frac{1}{M}\sum_{k} I_q(C_k)$$

4) MSTD is given as the point of intersection of the two lines (magenta and the green lines in Fig. 3) approximating the distribution of stability profiles. The lines are computed using a simple k-lines clustering algorithm (Agarwal et al. (2005)) for k = 2. Here MSTD comes out to be 14 (black line in Fig. 2).

The clustering quality index as mentioned under 3.c) is used here. It measures the quality of the clustering of ICs after multiple runs with random initial conditions by taking the difference between the average intra-cluster similarity and the average inter-cluster similarity (Himberg et al. (2004)).

It is further hypothesized that the point of inflection in the distribution of the stability profiles indicates the optimal number of ICs (Fig. 2). To find that point, the stability measures are clustered along the two lines (similar to 2-means clustering, but here the lines are taken as centroids instead of points) (Feldman (2003)). In this technique, the line (Fig. 2, red line) with positive slope grouped the stability profiles with lower values of M, while another line (Fig. 2, blue line) matched the stability components for the rest. This intersection of these lines (Fig. 2, black line) provided a consistent estimate of the effective number of ICs. This estimate is known as Maximally Stable Transcriptome Dimension (MSTD). This estimate is free of parameters (thresholds) unlike various information theory based criteria (BIC, AIC). It exploits the qualitative change in the character of the stability profile in higher dimensional data.

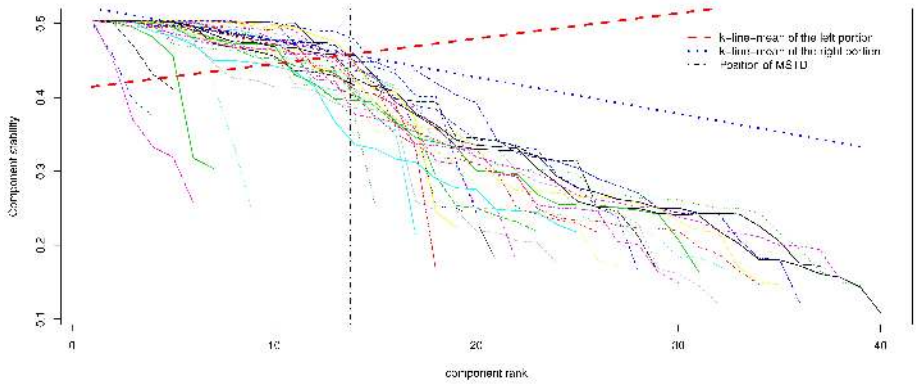

Fig. 2: Stability Index for ICA decompositions in various dimensions (from 2 to 40).

Three major conclusions can be made from the figures:

1) With the increase of M, the average stability of the computed components (S_M^{Total}) decreases (Fig. 3).

2) S_M^{Total} is characterized by the presence of local maxima, defining certain distinguished values of M that correspond to the (locally) maximally stable set of components (Fig. 3).

3) The stability profiles for various values of M can be classified into two, viz. (a) Stability values for which the value of M is low (M upto 9), in this case the stability values are unstable, it shows an irregularity in the trend of the stability profiles over M and (b) the stability values for those forming a large proportion of the components with higher values of M, in this case the stability values are stable and gradually decreasing with M (Fig. 2).

3.3. Cluster Analysis

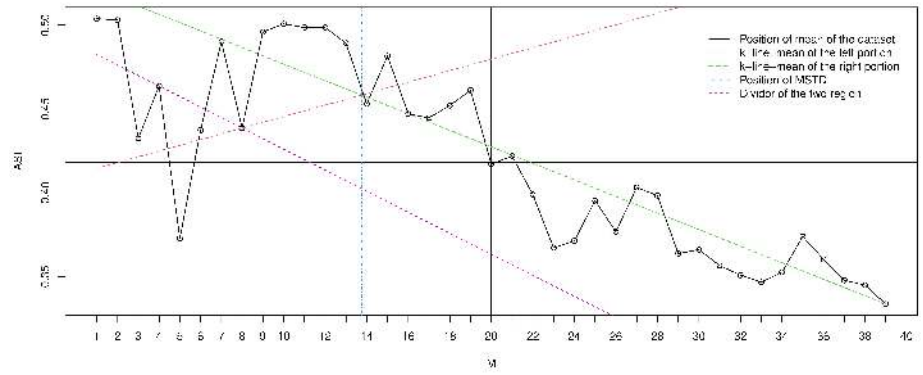

Fig. 3: Average Stability Index along with the two-line clustering result is shown by green and red dashed lines, with MSTD determined as the point of their intersection (vertical turquoise line).

In our data set we have applied the K-means clustering technique and its robustness in further examined by another clustering technique, viz. Gaussian Mixture Model Based Clustering (GMMBC). In this work we have done clustering on the basis of the 14 Independent Components, which includes spectroscopic, photometric, chemical composition, morphological, and kinematic properties of galaxies.

3.3.1. Optimal choice of clusters

There are several methods for finding the optimal number of clusters, present in the data e.g. technique by Sugar and James (2003), gap statistics (Tibshirani et al. (2001)) and many more . If an inherent clustering is present in the data then it is manifested by any clustering technique. We have used the Dunn index to find out the optimal number of cluster (Dunn (1974)) under the method of K-means clustering. Dunn index takes value between 0 to ∞. Initially we determined the structures of sub populations (clusters) for varying number of clusters say, k = 1, 2, 3, 4, ...etc. For each such clusters, we have computed the values of the Dunn Index. The value for which the Dunn Index comes out to be highest (here, 0.0063) is taken as the optimal choice of cluster (here, 12) present in our data set and that is the optimal choice of k taken in K-means analysis (Fig. 4). This finding is further justified by another technique i.e, GMMBC.

In this technique also, we have initially varied k over a certain range (say 1, 2, 3, ...etc.). As GMMBC is a model based clustering, the choices of k indicate the choice of the number of models being mixed among themselves and that mixture of models is taken for which the value of BIC is lowest. Here also the optimal number of cluster (i.e, the optimal number of models) appears to be 12 with a BIC value of -2.9$\times 10^5$ (Fig. 5).

3.3.2. Final selection of the reduced data set

Thus we have determined that the data set contains 12 distinct groups amongst themselves and from the previous section, the dimension of the data set can be reduced to 14 from 48 set of parameters using MSTD technique (Section 3). We have used two different indexes for choosing the best set of 14 ICs from the complete set of 48 ICs namely Average Silhouette Width (ASW) (Rousseeuw (1987)) and Within cluster Sum of Squares (WSS). WSS is used to check whether a clustering is good or poor. It is the sum of squares of the distances of the observations (x) within a cluster (C_i) from the cluster centroid (r_i). It is given as:

$$WSS = \sum_{x \in C_i} (x - r_i)^2$$

A good clustering yields a small within cluster sum of squares. In this way we choose that particular set of 14 ICs which gives the best clustering with respect to both ASW as well as WSS.

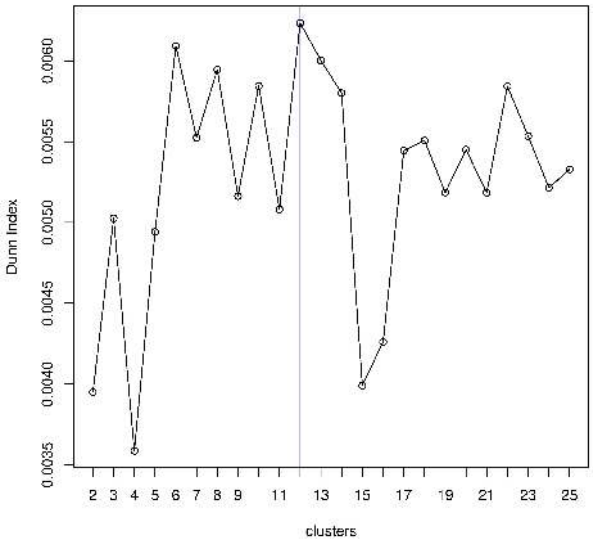

Fig. 4: Dunn Index for different k.

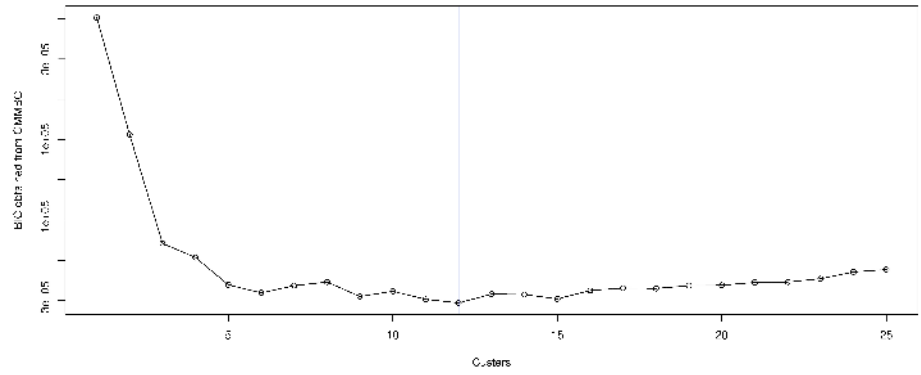

Fig. 5: BIC values obtained from GMMBC for various number of clusters.

In order to choose 14 ICs at random out of 48 under the above mentioned method we have used systematic sampling. Systematic sampling (Rao et al. (1988)) is a type of probability sampling method, in which the samples are drawn from a larger population according to a random starting point but with a fixed and periodic interval. This interval is called the sampling interval. We have used a systematic sampling scheme for the choice of 14 ICs from the set of 48 ICs. We further extend this sampling technique to all possible combinations of the initial random start so that for any random start we have a set of sample in our hand. Moreover we have taken all possible combinations of the sampling intervals also so that for any choice of the sampling interval we have a corresponding set of sample in our hand.

In doing this, some of the samples gets repeated, such as if we start with IC number 1 and goes on taking those ICs with a periodic sampling interval of 4, (with a maximum periodicity of 48, as we have 48 ICs in our hand) then we end up taking IC numbers 1,5,9,13,17,21,25,29,33,37,41,45,1,5 respectively. Here IC number 1 & 5 are repeated in the set, hence this set of ICs are not taken in our analysis. In this way those samples are rejected for our analysis where the ICs gets repeated in choosing the set of 14 different ICs from a sample of 48 ICs.

After getting all the set of ICs we calculate the ASW and WSS (mentioned earlier) for each samples and took that particular sample as our reduced data which gives the maximum ASW or the minimum WSS. ICs starting from number 39 and with a sample interval of 23 (viz., IC39, IC14, IC37, IC12, IC35, IC10, IC33, IC8, IC31, IC6, IC29, IC4, IC27, IC2) gives an ASW of 0.188904 & WSS of 5.76606 and satisfies our aforementioned criteria for the selection of the sample. We further rename the set of ICs (IC39, IC14, IC37, IC12, IC35, IC10, IC33, IC8, IC31, IC6, IC29, IC4, IC27, IC2) as (IC1 - IC14) respectively for the sake of our simplicity in addressing the ICs.

Finally, we have applied K-means cluster analysis on this data set by using 14 optimally selected Independent Components as parameters and the value of k as 12. Further, a second clustering by GMMBC is performed over the same data set for robustness analysis (Table 2).

Table 2

(The median values of the Age (Gyr), length_avg along with the standard error (in parentheses), and the mean values of the observable parameter along with the standard error (in parentheses), for each cluster is given below. The clusters obtained by K-means are denoted by 'K' and the clusters obtained by GMMBC are denoted by 'G'. The mean values for rest of the observable parameter along with the standard error (in parentheses) is given in the following Url: `https://drive.google.com/file/d/1BnELad32A6VyDahR4txNpVBWGxvrN8SE/view?usp=sharing`)

Cluster Indices	Sample_Size	Age (Median Values)	Barred_sample_Size	length_avg (Median Values)
K1	1616	2.3(0.05424)	301	8.87097(0.21045)
K2	5257	1.609(0.03186)	399	9.2719(0.17604)
K3	4103	1.609(0.03445)	143	8.05755(0.26768)
K4	126	5.875(0.3058)	15	10.68616(0.8906)
K5	425	1.0152(0.11685)	12	7.07733(0.63335)
K6	17	0.2273(1.55877)	1	10.34275(0)
K7	23	0.7187(0.80196)	1	8.10091(0)
K8	3723	1.434(0.03455)	150	7.40862(0.26954)
K9	1	13.25(0)	0	NA
K10	185	1.434(0.14679)	4	5.36336(0.40655)
K11	6365	2.1(0.03878)	570	10.13034(0.16852)
K12	4248	1.434(0.03203)	173	8.70386(0.27095)
Cluster Indices	**Sample_Size**	**Age (Median Values)**	**Barred_sample_Size**	**length_avg (Median Values)**
G1	432	1.434(0.08072)	8	7.06626(0.97356)
G2	3088	1.8(0.03348)	363	8.77866(0.20282)
G3	6803	1.2781(0.02093)	178	6.78002(0.21122)
G4	10697	1.9(0.02655)	816	9.75887(0.127)
G5	101	2.75(0.5013)	7	10.03251(1.11173)
G6	759	1.8(0.12186)	13	10.61704(1.21821)
G7	2004	1.434(0.04803)	43	8.27783(0.37798)
G8	704	1.0152(0.07775)	23	7.33386(0.60604)
G9	811	2.4(0.10432)	190	10.01833(0.28531)
G10	416	5.5(0.17722)	76	10.26973(0.46843)
G11	270	1.68(0.17767)	51	8.14877(0.5025)
G12	4	10(2.61632)	1	11.65327(0)

Cluster Indices	Sample_Size	$log(M_*)$	D_n(4000)	U	G	R	I
K1	1616	9.90439(0.00835)	1.2544(0.00299)	-19.43732(0.01757)	-20.70913(0.01635)	-21.3206(0.01633)	-21.67598(0.01666)
K2	5257	9.86208(0.00661)	1.46626(0.00287)	-18.8964(0.01304)	-20.41194(0.01188)	-21.0993(0.01197)	-21.47548(0.01205)
K3	4103	9.72752(0.00699)	1.42718(0.00278)	-18.69309(0.01346)	-20.17005(0.0125)	-20.84024(0.01263)	-21.21494(0.01267)
K4	126	10.72587(0.03902)	1.90489(0.01165)	-19.19173(0.07196)	-21.11759(0.07008)	-21.98127(0.07163)	-22.40291(0.07261)
K5	425	9.1308(0.02185)	1.18082(0.00562)	-18.61763(0.03978)	-19.72759(0.04088)	-20.14142(0.04274)	-20.40716(0.0445)
K6	17	7.83294(0.37162)	1.40902(0.04798)	-17.14303(0.61724)	-17.79457(0.73786)	-21.02199(0.33941)	-21.18044(0.33999)
K7	23	8.79826(0.06862)	1.04913(0.00631)	-18.33515(0.14112)	-19.28366(0.12926)	-19.60257(0.11795)	-19.78258(0.12601)
K8	3723	9.62969(0.00748)	1.39598(0.00307)	-18.64067(0.01387)	-20.06829(0.01294)	-20.7063(0.01308)	-21.06136(0.01333)
K9	1	8.01(0)	1.28818(0)	-9.37978(0)	-17.25025(0)	-17.85383(0)	-18.1867(0)
K10	185	9.46335(0.03027)	1.35396(0.01373)	-18.30381(0.06301)	-19.70998(0.05604)	-20.33062(0.05465)	-20.67803(0.05457)
K11	6365	10.1918(0.00699)	1.57024(0.003)	-19.15221(0.01021)	-20.76579(0.01048)	-21.5271(0.011)	-21.9312(0.01143)
K12	4248	9.72668(0.00698)	1.42778(0.00278)	-18.73717(0.01316)	-20.2036(0.01229)	-20.86652(0.01239)	-21.23347(0.01264)

Cluster Indices	Sample_Size	$log(M_*)$	D_n(4000)	U	G	R	I
G1	432	9.365(0.01722)	1.34264(0.00994)	-18.10025(0.03795)	-19.50202(0.03409)	-20.12297(0.03315)	-20.476(0.0332)
G2	3088	9.78574(0.00719)	1.29102(0.00256)	-19.22077(0.01415)	-20.52803(0.01358)	-21.13111(0.0138)	-21.47963(0.01412)
G3	6803	9.55841(0.00466)	1.35675(0.00167)	-18.65592(0.00974)	-20.04478(0.00886)	-20.65707(0.00872)	-21.00623(0.00881)
G4	10697	10.12099(0.00453)	1.57668(0.00195)	-19.06244(0.00805)	-20.67267(0.00772)	-21.42848(0.0078)	-21.83013(0.00788)
G5	101	9.33158(0.13348)	1.39201(0.01821)	-18.20581(0.18508)	-19.51398(0.21045)	-20.65595(0.17291)	-20.92904(0.18339)
G6	759	9.84625(0.01795)	1.51084(0.00884)	-18.14955(0.04125)	-19.90522(0.03106)	-20.68149(0.0311)	-21.106(0.03147)
G7	2004	9.63086(0.0095)	1.41727(0.00349)	-18.45827(0.01884)	-19.94553(0.01697)	-20.62005(0.01678)	-20.99567(0.01688)
G8	704	9.12939(0.01529)	1.20886(0.0035)	-18.48432(0.02834)	-19.64376(0.02843)	-20.08478(0.02932)	-20.35706(0.03059)
G9	811	10.26343(0.01468)	1.39753(0.0079)	-19.53831(0.02447)	-21.04197(0.02272)	-21.78353(0.02314)	-22.18099(0.02503)
G10	416	10.77909(0.01898)	1.88879(0.00606)	-19.40894(0.03545)	-21.30414(0.03478)	-22.17206(0.03502)	-22.60102(0.03539)
G11	270	9.59041(0.03426)	1.19806(0.00966)	-19.23462(0.05048)	-20.40093(0.05429)	-20.90591(0.05899)	-21.21069(0.06243)
G12	4	8.86(0.68627)	1.51563(0.11345)	-13.46851(2.36757)	-16.72492(2.0779)	-17.79778(1.81164)	-18.25704(1.74008)

Cluster Indices	Sample_Size	Z	J	H	K	SFR	U-G
K1	1616	-21.93775(0.01694)	-21.66014(0.01993)	-22.38373(0.02041)	-22.82018(0.02079)	1.0638(0.02713)	1.27181(0.00525)
K2	5257	-21.7555(0.01212)	-21.55149(0.01157)	-22.28546(0.01128)	-22.71686(0.01084)	0.29182(0.00681)	1.51554(0.0054)
K3	4103	-21.49019(0.01297)	-21.25062(0.01396)	-22.01234(0.01381)	-22.4539(0.01368)	0.2581(0.00657)	1.47697(0.00511)
K4	126	-22.73229(0.07426)	-22.70672(0.06646)	-23.40741(0.06816)	-23.76453(0.06688)	0.05794(0.0218)	1.92586(0.02941)
K5	425	-20.58465(0.04568)	-20.41109(0.04792)	-21.18188(0.04722)	-21.69543(0.04628)	0.46424(0.02539)	1.10996(0.01159)
K6	17	-17.93447(0.91007)	-21.39051(0.33567)	-22.2191(0.28551)	-22.70181(0.30081)	1.74706(0.80901)	0.65155(0.98487)
K7	23	-19.92013(0.12171)	-19.70275(0.17297)	-20.48132(0.18368)	-20.87053(0.17381)	0.41304(0.07394)	0.94851(0.02521)
K8	3723	-21.32314(0.01359)	-21.0286(0.0149)	-21.7904(0.01453)	-22.2578(0.01449)	0.26242(0.00661)	1.42762(0.00444)
K9	1	-18.31074(0)	-18.28723(0)	-18.47423(0)	-19.55423(0)	0(0)	7.87047(0)
K10	185	-20.93675(0.05492)	-20.51798(0.04727)	-21.33757(0.04803)	-21.83218(0.04165)	0.21892(0.02679)	1.40617(0.0191)
K11	6365	-22.24511(0.01161)	-21.97196(0.0123)	-22.70899(0.01216)	-23.1256(0.01199)	0.2798(0.00683)	1.61358(0.00361)
K12	4248	-21.50759(0.01276)	-21.25928(0.01321)	-22.02619(0.01295)	-22.46407(0.01271)	0.27093(0.00671)	1.46643(0.00416)

Cluster Indices	Sample_Size	Z	J	H	K	SFR	U-G
G1	432	-20.73857(0.03356)	-20.44912(0.03313)	-21.35022(0.03301)	-21.87317(0.03276)	0.16505(0.01129)	1.40178(0.01126)
G2	3088	-21.73313(0.01438)	-21.43728(0.01589)	-22.16386(0.01593)	-22.60139(0.0161)	0.72847(0.01466)	1.30725(0.00388)
G3	6803	-21.26019(0.00889)	-20.92901(0.00985)	-21.72344(0.00988)	-22.20691(0.00984)	0.26987(0.00437)	1.38886(0.00289)
G4	10697	-22.1379(0.00796)	-21.92457(0.00789)	-22.63046(0.00805)	-23.02487(0.00805)	0.24855(0.00469)	1.61023(0.0024)
G5	101	-20.59199(0.24866)	-21.21329(0.11427)	-21.9583(0.11029)	-22.4513(0.1088)	0.71089(0.17431)	1.30818(0.22496)
G6	759	-21.42326(0.03195)	-21.46964(0.03088)	-22.2503(0.02953)	-22.71963(0.02979)	0.10198(0.00863)	1.75566(0.02819)
G7	2004	-21.27896(0.01708)	-21.01205(0.01676)	-21.81426(0.01646)	-22.27631(0.01605)	0.19596(0.00756)	1.48726(0.00633)
G8	704	-20.54253(0.03133)	-20.35409(0.03294)	-21.17807(0.03232)	-21.69468(0.03311)	0.34261(0.01514)	1.15943(0.00788)
G9	811	-22.49597(0.02391)	-22.26266(0.02501)	-23.00113(0.02572)	-23.44847(0.02589)	0.79063(0.03991)	1.50365(0.01025)
G10	416	-22.93372(0.03575)	-22.77189(0.02965)	-23.46636(0.03046)	-23.82615(0.03035)	0.05288(0.00979)	1.8952(0.01016)
G11	270	-21.42426(0.06471)	-21.05023(0.06567)	-21.80613(0.06565)	-22.23182(0.06423)	0.90222(0.06998)	1.16632(0.01662)
G12	4	-18.67572(1.60019)	-20.17417(1.14757)	-20.74217(1.22211)	-21.17692(1.101)	0(0)	3.25642(1.54529)

Cluster Indices	Sample_Size	G-R	R-I	I-Z	U-Z	J-H	H-K
K1	1616	0.61146(0.00341)	0.35539(0.00166)	0.26177(0.00171)	2.50043(0.01098)	0.72359(0.00568)	0.43645(0.00606)
K2	5257	0.68736(0.00201)	0.37617(0.00095)	0.28002(0.00103)	2.8591(0.00793)	0.73397(0.00451)	0.4314(0.00466)
K3	4103	0.67019(0.00235)	0.37471(0.0018)	0.27525(0.00161)	2.79711(0.00861)	0.76172(0.00562)	0.44155(0.00577)
K4	126	0.86367(0.00633)	0.42165(0.00452)	0.32938(0.00546)	3.54056(0.03394)	0.70069(0.01356)	0.35712(0.01082)
K5	425	0.41384(0.00715)	0.26574(0.004)	0.17749(0.00355)	1.96702(0.02333)	0.77079(0.01829)	0.51355(0.01996)
K6	17	3.22742(0.67742)	0.15845(0.18426)	-3.24598(0.93807)	0.79144(0.79693)	0.82859(0.09596)	0.48271(0.08432)
K7	23	0.3189(0.02483)	0.18001(0.01508)	0.13755(0.01397)	1.58497(0.05698)	0.77857(0.1127)	0.38922(0.08759)
K8	3723	0.63801(0.00252)	0.35506(0.00128)	0.26178(0.00131)	2.68247(0.00883)	0.7618(0.00632)	0.4674(0.00648)
K9	1	0.60358(0)	0.33287(0)	0.12404(0)	8.93096(0)	0.187(0)	1.08(0)
K10	185	0.62065(0.01164)	0.34741(0.00594)	0.25872(0.00618)	2.63294(0.03915)	0.81959(0.03145)	0.49461(0.03302)
K11	6365	0.76131(0.00188)	0.4041(0.0019)	0.31391(0.00195)	3.0929(0.00683)	0.73703(0.00303)	0.41661(0.00317)
K12	4248	0.66292(0.00235)	0.36695(0.00159)	0.27412(0.00165)	2.77042(0.00818)	0.76692(0.00562)	0.43787(0.00577)

Cluster Indices	Sample_Size	G-R	R-I	I-Z	U-Z	J-H	H-K
G1	432	0.62095(0.00689)	0.35303(0.00391)	0.26257(0.00424)	2.63833(0.02381)	0.90109(0.02149)	0.52295(0.022)
G2	3088	0.60308(0.00247)	0.34852(0.00126)	0.2535(0.00122)	2.51236(0.00821)	0.72658(0.00502)	0.43753(0.00517)
G3	6803	0.6123(0.00169)	0.34915(0.00091)	0.25396(0.00093)	2.60427(0.00591)	0.79443(0.00488)	0.48347(0.00501)
G4	10697	0.75581(0.00125)	0.40164(0.00062)	0.30777(0.00063)	3.07546(0.00454)	0.7059(0.00239)	0.3944(0.00256)
G5	101	1.14197(0.14978)	0.27309(0.11667)	-0.33705(0.23185)	2.38619(0.21734)	0.74501(0.03715)	0.493(0.03616)
G6	759	0.77627(0.00509)	0.42452(0.00292)	0.31725(0.00322)	3.2737(0.03104)	0.78066(0.01374)	0.46933(0.01393)
G7	2004	0.67452(0.00332)	0.37561(0.00178)	0.28329(0.00189)	2.82068(0.01205)	0.80222(0.00945)	0.46205(0.00947)
G8	704	0.44103(0.00492)	0.27228(0.00314)	0.18547(0.00296)	2.05821(0.01635)	0.82398(0.01488)	0.51661(0.01582)
G9	811	0.74156(0.00487)	0.39747(0.00772)	0.31498(0.00776)	2.95766(0.01799)	0.73847(0.00631)	0.44734(0.00622)
G10	416	0.86792(0.00318)	0.42896(0.00217)	0.3327(0.0023)	3.52479(0.01341)	0.69446(0.00656)	0.3598(0.00661)
G11	270	0.50498(0.01069)	0.30478(0.00538)	0.21357(0.0048)	2.18964(0.03459)	0.7559(0.01862)	0.42569(0.01913)
G12	4	1.07286(0.30623)	0.45926(0.08625)	0.41868(0.18051)	5.20721(1.29849)	0.568(0.12765)	0.43475(0.23419)

4. Results

In the present work we have compiled a data set upto a red shift of $z < 0.06$, consisting of unbarred, weak barred and strong barred spiral galaxies and including starburst, AGN and LINER. At first we have checked the data set for Gaussianity and found it to be non-Gaussian. We performed Independent Component Analysis for dimentionality reduction and used MSTD to find out the optimal number of Independent Components. Then we clustered the data set with respect to these Independent Components by K-means cluster analysis followed by finding the number of optimum groups. The optimum number of ICs is 14 and the number of co-herrent groups is 12 (Table 2). From Table 2, it is clear that group 9 (K9) is an outlier and groups 6 and 7 (K6 and K7) contain a very small number of galaxies. Hence we have not considered K9. Subsequently we performed Gaussian Mixture Model Based Clustering (GMMBC) for checking the robustness of the groups. We have found similar number of groups with one group containing only 4 members like K9. Remaining 11 groups found by both the methods are more or less compatible with respect to membership as well as average values of the parameters (K1 → G2, K2 → G6, K3 → G7, K4 → G10, K5 → G8, K6 → G9, K7 → G11, K8 → G1, K9 → G12, K10 → G3, K11 → G4, K12 → G5). Therefore we have discussed the physical properties of the groups found with respect to K-means cluster analysis.

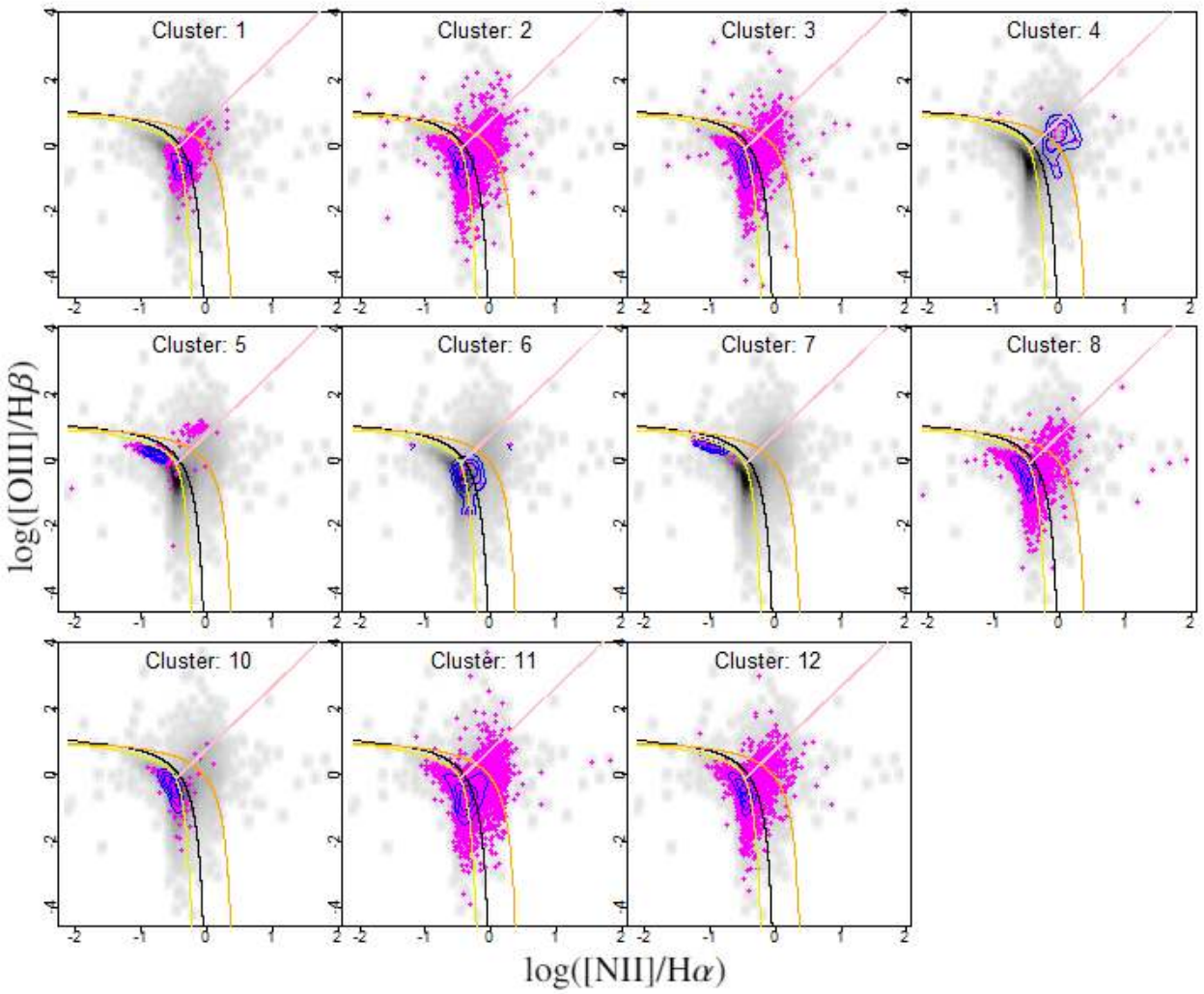

Fig. 6(a): log([OIII]/Hβ) line ratios are plotted against log([NII]/Hα) values for the whole sample (in grey) along with groups K1 - K12 (in magenta dots and blue iso contours) separately, except K9. Curves are from Stasińska et al. (2006); Kauffmann et al. (2003); and Kewley et al. (2001) (left to right respectively). The bottom left zone is for star-forming regions, the upper zone is for AGNs and Seyfert galaxies, and the bottom right zone is for LINERs.

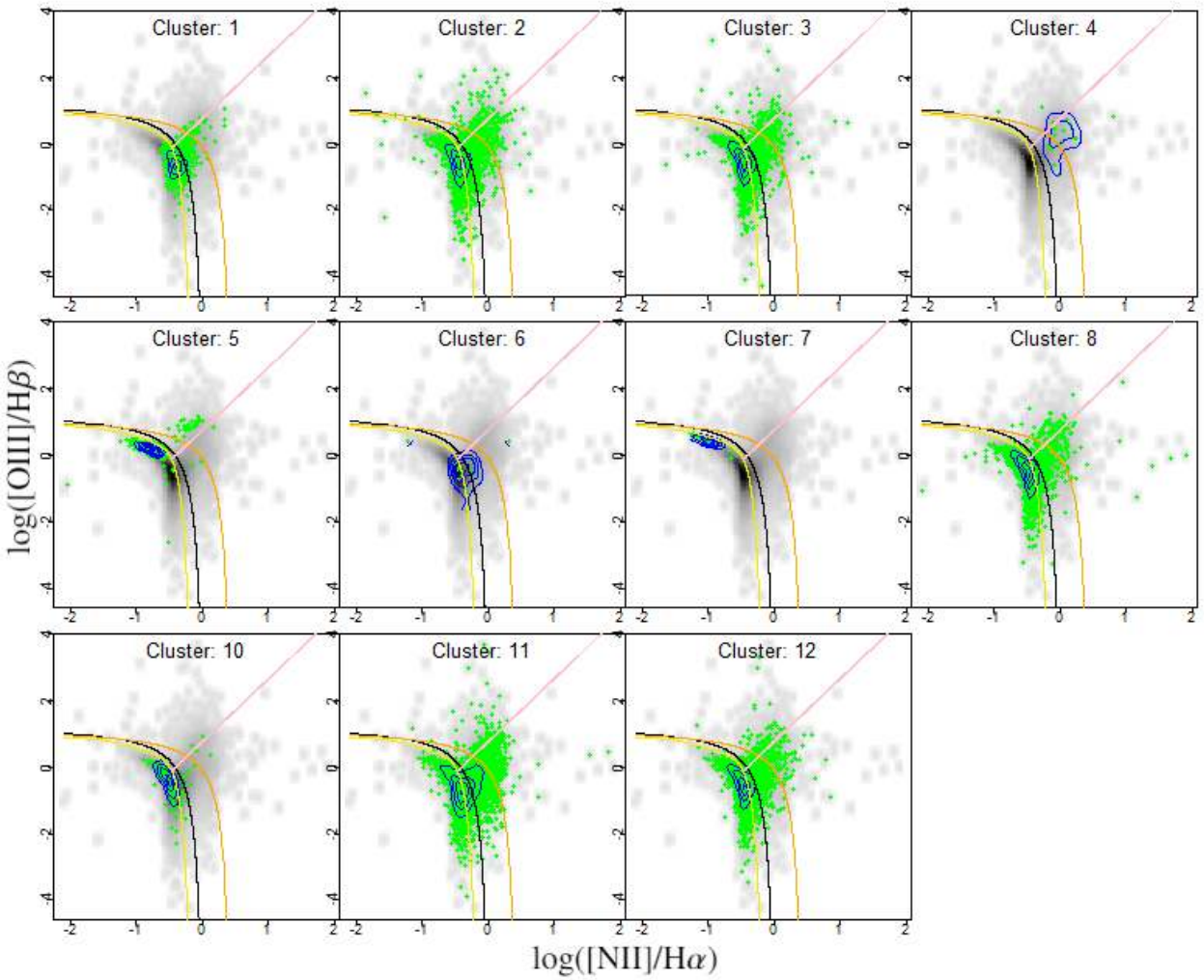

Fig. 6(b): Same as Fig. 6(a) but only for unbarred spiral galaxies (green dots).

4.1. Properties of the ICs

In the analysis of ICA, we randomly selected set of ICs for which the variation in the multivariate set up is maximum with respect to Dunn index (Fig. 4). For the optimum set there are 14 ICs. ICs are actually linear combinations of several parameters (here 48 observable parameters) with various co-efficients. Few ICs are denoted by some features specific to a particular physical property of galaxies without any prior selection. At the same time each IC is not limited by the dominant feature. It also includes other with some lower weights and thus takes into account the complex interplay between observable parameters. Table 3 shows that among 14 ICs, 7 represent about five kind of properties:

1) Metallicty (IC1,IC2,IC5),
2) Balmer absorption feature and low level ionisation (IC3),
3) Color (IC4),
4) Velocity dispersion (IC14),
5) Metallicity and high level ionisation (IC6, IC11).

The remaining ICs are not dominated by any particular physical characteristic and the correlations with these ICs show negligibly small values. Thus even though we have got 14 ICs following IC1-IC14 (Subsections 3.2.1 & 3.3.2)

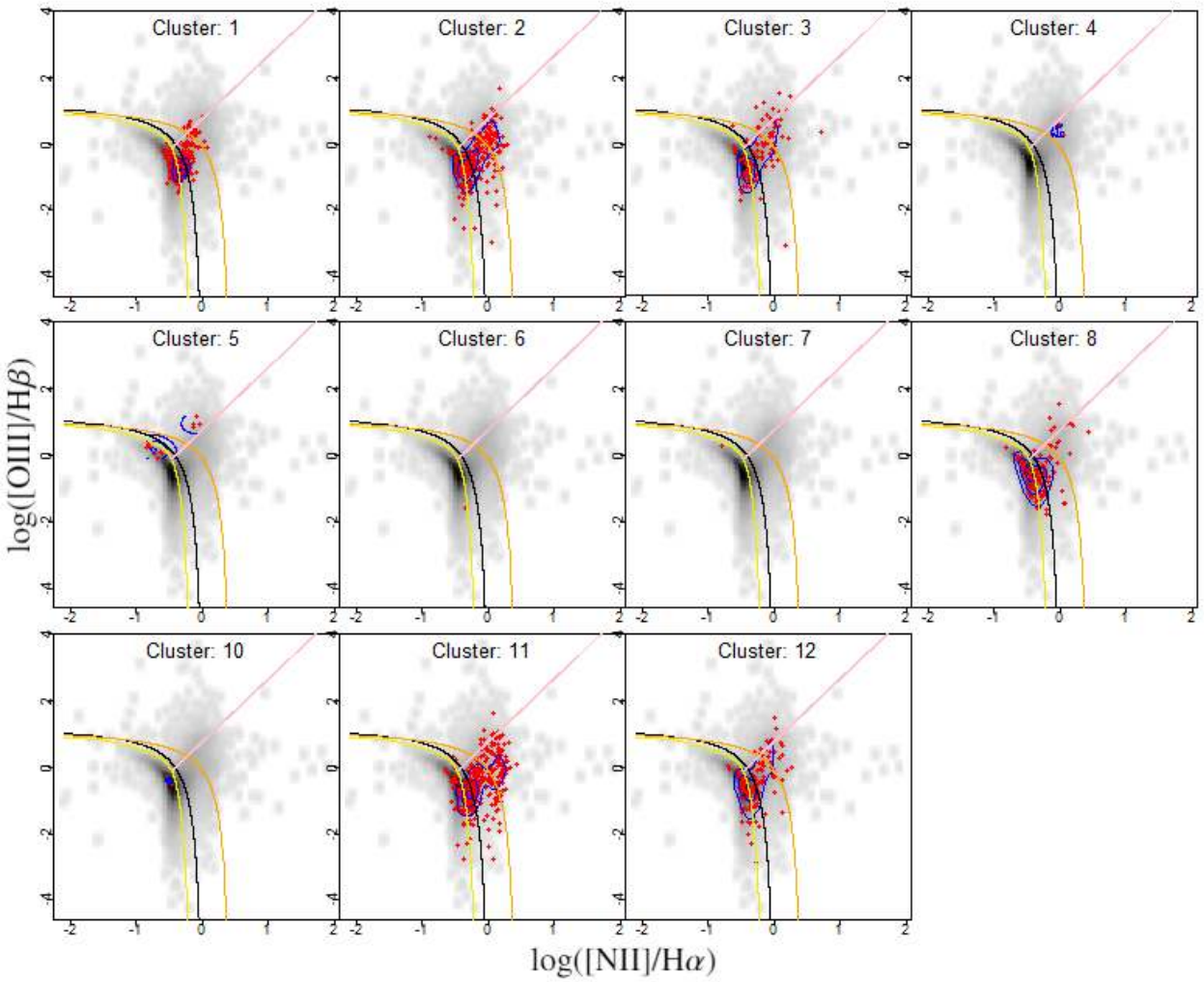

Fig. 6(c): Same as Fig. 6(a) but only for barred spiral galaxies (red dots).

among those, 6, ICs have negligibly small effects on the total variation and 8 of them are the most significant ones.

Table 3

(Observed parameters with highest correlation coefficients with significant ICs .)

IC	**Influential observed parameters**
$IC1$	Lick_Fe 5406 (0.54)
$IC2$	Lick_Ca 4455 (-0.30)
$IC3$	Lick_H_β (0.62), EW (Nii 6584) (0.65), EW (H_α) (0.57), EW (H_β) (0.65), EW (H_γ) (0.65), EW (H_δ) (0.65)
$IC4$	i-z (0.30)
$IC5$	Lick_Nad (0.76), Lick_Ca 4227 (0.70), Lick_g 4300 (0.76), Lick_Fe 4383 (0.74), Lick_Fe 4531 (0.71), Lick_C 4668 (0.79), Lick_Mgb (0.77), Lick_Fe 5406 (0.72), $D_n(4000)$ (0.74), g-r (0.76), r-i (0.74), u-z (0.76)

$IC6$	Lick_Fe 5709 (-0.51)
$IC11$	EW (Oiii 5007) (-0.64)
$IC14$	σ-forbidden (0.38)

4.2. Properties of the galaxies in the groups.

There are 11 effective groups (K1 - K8, K10 - K12), as a result of Cluster analysis. Among these K2, K4, K11 (viz. Table 2) consist of oldest galaxies with respect to average ages and $D_n(4000)$ values (though average age of K1 is higher than K2 but the average $D_n(4000)$ values are just the opposite and since it is an observed parameter hence is more reliable one than the estimated value. Also K1 group of galaxies are similar to the groups of galaxies in the medium age range with respect to other physical properties e.g. colour or metallicity etc.). K1, K3 ,K8, K10, K12 consist of galaxies of medium age, K6, K7 are the youngest groups of galaxies and K5 consists of unbarred galaxies. In each group we have classified 3 subgroups as strong barred (where the size of the bars are at least 30% of their host galaxy size), weak barred (where the size of the bars is smaller than 30% of the size of the host galaxy) and unbarred galaxies. It is clear from Table 2 that galaxies in groups K2, K4 and K11 have galaxies of median ages, and $D_n(4000)$ values which increase as, K2 < K11 < K4 and the median bar lengths of these groups also increase as K2 < K11 < K4. The metallicities are higher in these galaxies. On the contrary for galaxies in the groups K1, K3, K8, K10, K12 the median ages and $D_n(4000)$ values vary more or less in the medium range ($\sim$ median age, 1.4 Gyr - 1.6 Gyr) with their bar lengths smaller compared to the oldest groups of galaxies. Finally K6, K7 are the youngest groups of unbarred galaxies with minimum sample size and K5 is the youngest group (Median age $\sim$ 1.0 Gyr) containing barred and unbarred galaxies.

4.3. Emission line diagnostics and CM diagrams

Emission line ratios have been recommended by various authors (Baldwin et al. (1981); Veilleux and Osterbrock (1987); Kauffmann et al. (2003)) for qualitative classification of galaxies. According to the above scheme, scatter diagram of two emission line ratios ($\log(\mathrm{NII}/\mathrm{H}_\alpha)$ and $\log(\mathrm{OIII}/\mathrm{H}_\beta)$) are classified by equations of curves separating the different classes of starburst galaxies, AGN and LINERs (Fig. 6). It is clear from Fig. 6 that K2, K4 and K11 contain all types of galaxies and the starburst galaxies have a wide range of ionization ($\log(\mathrm{OIII}/\mathrm{H}_\beta) \sim$ -4 to 0.5). K1, K3, K8, K10 and K12 groups are primarily dominated by starburst with few AGN and LINERs, wheras K6, K7, K5 are populated by starburst galaxies.

These observations are more or less consistent with the groups. As K2, K4 and K11 are the oldest groups of galaxies and they contain both unbarred and barred galaxies. Star formation still occurs in some unbarred galaxies accompanied by AGN and LINERs which are oldest in ages. In galaxies of medium ages in the groups K1, K3, K8, K10 and K12, they are dominated by star forming galaxies rather than AGN or LINERs where star formation is being affected by the presence of bars (Athanassoula (1983); Buta and Combes (1996)). This is also reflected in Fig. 7 where for the oldest groups (K2, K4, K11) the contours peak to redder color, medium aged groups (K1, K3, K8, K10, K12) concentrate around bluer zone and in particular in Fig. 8 where for the strong barred galaxies the contours comparatively peak at redder zone contrary to weak barred galaxies which peak at bluer zone in the color magnitude (CM) diagrams.

The color magnitude diagrams in ($g - r$, R) and color-color diagrams in ($g - r$, $u - r$) (Figs. 9(a) and 9(b)) for all galaxies (strong barred, weak barred and unbarred) show similar features as in Fig. 8. One interesting feature in color-color diagram shows that in K2 and K11 the weak barred galaxies occupy the same redder region as the strong barred galaxies. Since these galaxies fall in the oldest groups of galaxies, it indicates that these weak barred galaxies were previously strong barred galaxies but the bars are getting dissolved, supporting a recurrent phenomenon of bar formation. In the medium aged groups of galaxies the effect is not very pronounced due to non-availability of data on weak barred galaxies.

4.4. Star formation efficiency

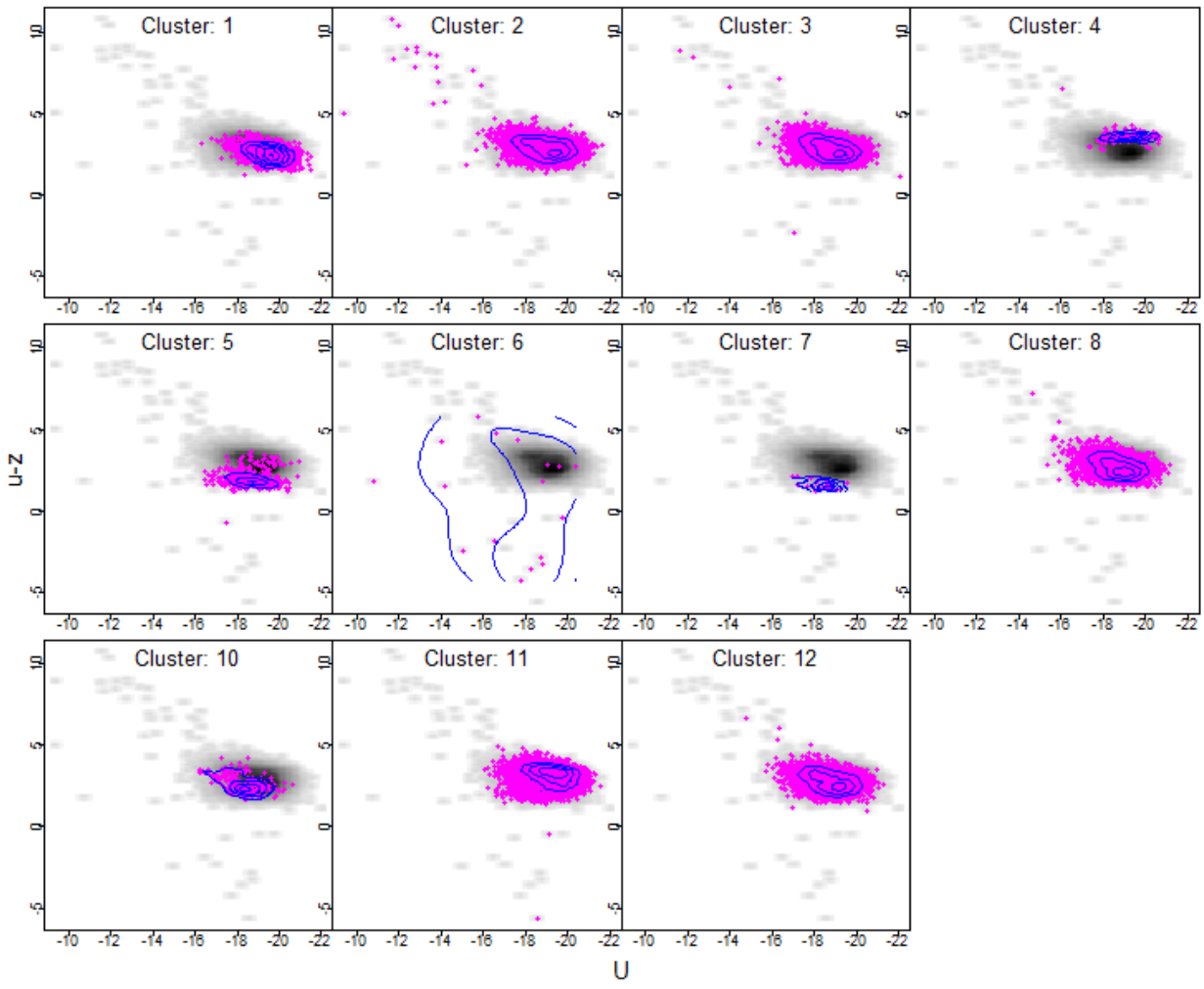

Fig. 7(a): Color–magnitude ($u-z$ vs. U) diagrams of the whole sample and for each of the groups K1 - K12 except K9.

Our aim is to assess the effect of bar on various properties of galaxies. Therefore we use the specific star formation rate parameter (Star Formation Efficiency, hereafter SFE) denoted by $\log(SFR/M_*)$ (viz. Table 1) as a good measure for star formation activity in galaxies (Brinchmann et al. (2004)). Also $D_n(4000)$ (Kauffmann et al. (2003); Balogh et al. (1999)) has been used for the second estimate of age of stellar populations. Fig. 10 shows the SFE and $D_n(4000)$ vs $log(M_*)$ for various groups K1 - K12. For the oldest groups i.e, in K2, K4, K11, data for K4 are not available.

For K2, low mass weak barred galaxies have lower SFE compared to strong barred and unbarred galaxies ($M_* \leq 10^{10.5} M_\odot$). This might indicate that these galaxies were initially strong barred and their bars are getting dissolved and their SFE got attenuated in the presence of strong bar initially. Thus they have low SFE (Vera et al. (2016)). High mass weak barred galaxies ($M_* > 10^{10.5} M_\odot$) have their bars growing from unbarred galaxies, which is quenching their star formation activity and they have again low SFE. Generally SFE is always lower in strong barred galaxies than unbarred galaxies. Various authors (Masters et al. (2012); Ho et al. (1997); Sheth et al. (2005); Ellison et al. (2011); Lee et al. (2012)) have suggested that the presence of bars funnel the gas into the central region of the galaxy. Subsequently the material turned into molecular gas which when becomes gravitationally unstable undergo fragmentation and trigger star formation activity. Therefore presence of bar could accelerate gas consumption ceasing formation of new stars in the outer region of the discs as they become redder. Thus in the star burst phase large amount of gas is transported towards the galactic central region for triggering star formation activity and in the post star burst phase the gas is consumed by circum nuclear star burst which shows low star formation rate (SFR) (Jogee et al. (2005); Sheth et al. (2005)).

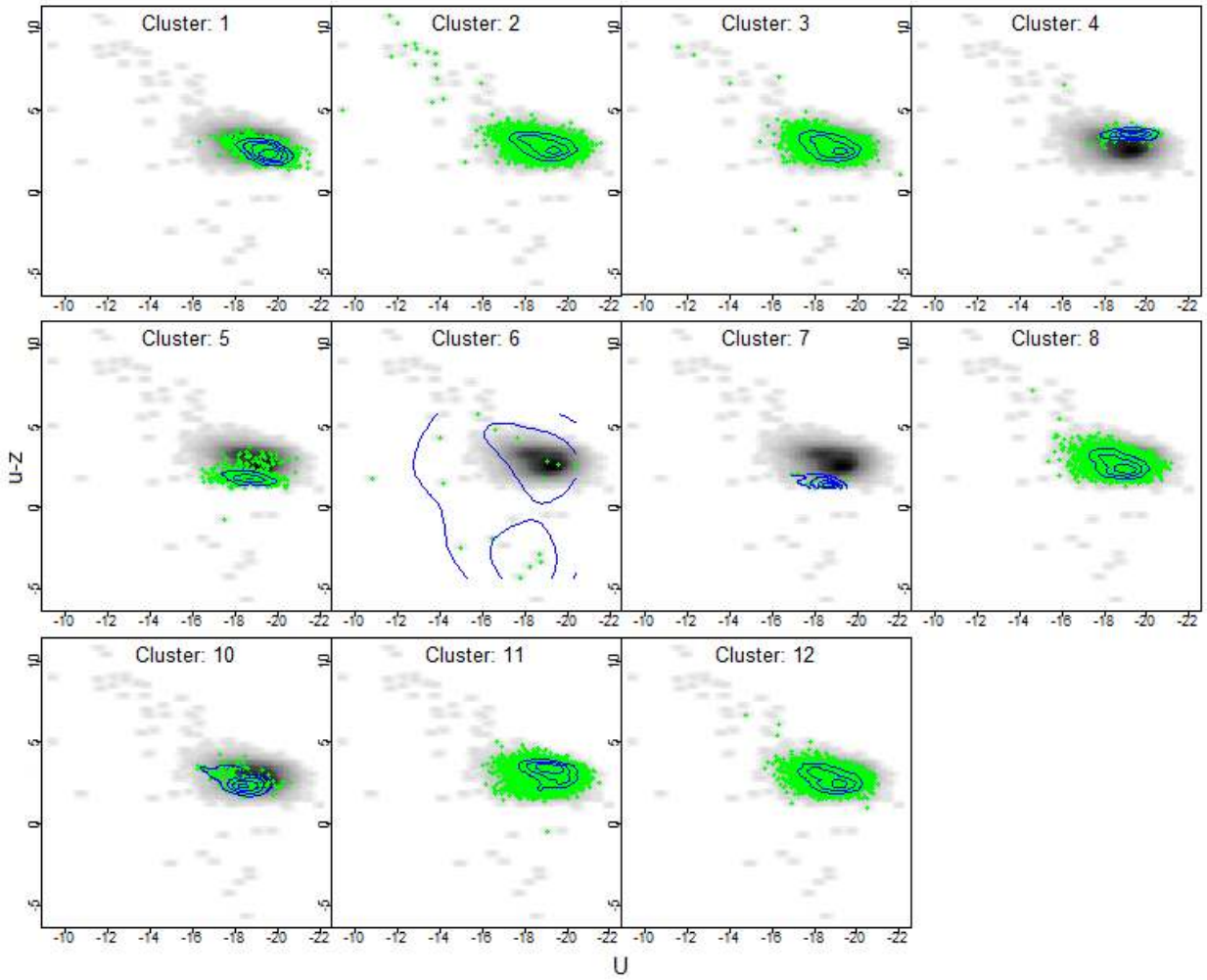

Fig. 7(b): Color–magnitude ($u-z$ vs. U) diagrams of Unbarred galaxies and for each of the groups K1 - K12 except K9.

In K4, SFE values are not available for barred galaxies. In K11, for high mass range ($10^{10} M_{\odot}$ > M > $10^{9} M_{\odot}$) strong barred and weak barred galaxies have similar SFE, which may be as low as $\sim 10^{-11}\ yr^{-1}$). Few unbarred galaxies have very low SFE ($\sim 10^{-11}\ yr^{-1}$). Since they are of oldest ages their SFEs are low. Few weak barred galaxies in the high mass range have low SFE. This is indicative of the fact that these galaxies had strong bar initially but is getting dissolved. So, the SFE is low in these weak-barred galaxies. Some of this fact is consistent with the recurrent bar formation scenario in galaxies (Kormendy and Kennicutt Jr (2004); Bournaud and Combes (2002); Berentzen et al. (2004); Gadotti and de Souza (2006); Katz et al. (2018); Hilmi et al. (2020); de Lorenzo-Cáceres et al. (2019))

In the medium aged groups of galaxies K1, K3, K12, they have similar phenomenon of recurrent bar formation. e.g. in K1, K3 and K12 some features have been observed as in case of oldest groups.

4.5. Mass-metallicity relation

In Fig. 11 the mass-metallicity relation is shown for all the groups (K1 - K12 except K9). In the oldest groups of galaxies (K2, K4, K11), barred galaxy mettalicity are always larger than unbarred ones. Metallicity increases with galaxy mass but the increase is almost constant after z > 0.02. For the same metallicity barred galaxy masses are higher

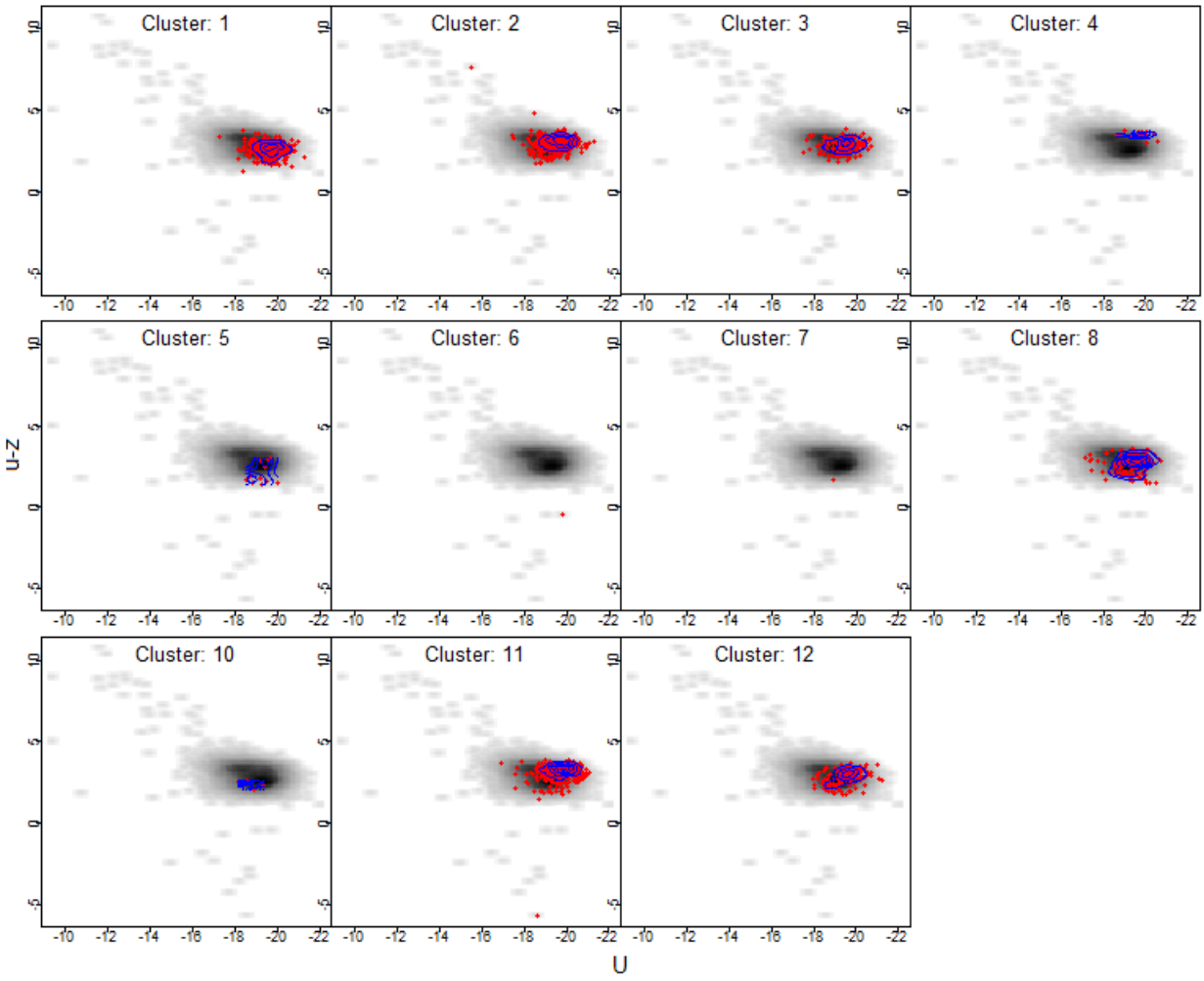

Fig. 7(c): Color–magnitude ($u - z$ vs. U) diagrams of Barred galaxies and for each of the groups K1 - K12 except K9.

than unbarred ones. In these groups the SFE decreases with increasing mass. This might be due to the fact that (Ellison et al. (2007); Ellison et al. (2011)) metal enhancement without an accompanying increase in star formation activity may be due to a short lived phase of bar-triggered star formation in the past. Also the fall in the metallicity is rapid for strong barred galaxies rather than weak or unbarred galaxies. Weak barred galaxy curve crossed the strong bar curve in the high mass zone as well as in low mass range. This might be due to the fact that in the low mass range they can grow from strong barred ones as their bars are getting dissolved and from unbarred ones, in the high mass zone, when SFEs are low (viz. Fig. 10). For the intermediate aged groups (K1, K3, K8, K10, K12) in most cases the data are not available from sdss but the more-or-less trend is similar as that of oldest groups. Moreover the crossing of weak bar curve with strong and unbarred ones indicates recurrent phenomenon of bar formation.

5. Conclusion

The present work deals with a large data set of unbarred, strong barred and weak barred galaxies taken from sdss DR15 and cross-matched with zooniverse, for collecting bar properties of the barred galaxies. We have considered several significant observable parameters (e.g. Lick indices, Metallicity, SFR, Color Magnitudes etc.) (Table 1.) for performing the statistical analyses and the estimated parameters (e.g. age, SFE etc) along with observable ones are used for physical interpretation of the homogeneous groups. We have studied the influence of bars on the various properties

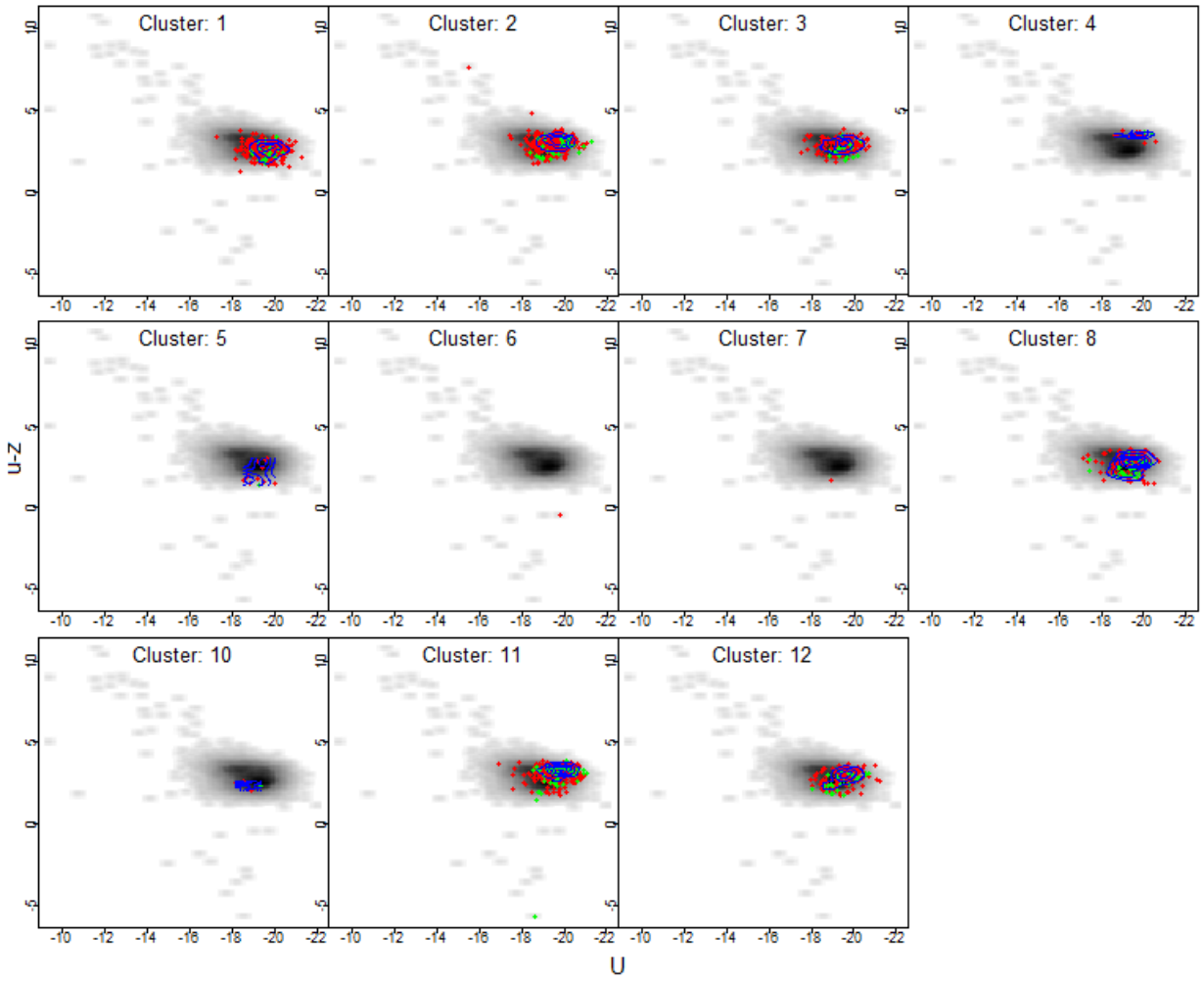

Fig. 8(a): Color–magnitude ($u - z$ vs. U) diagrams of the Barred galaxies and for each of the groups K1 - K12 except K9, where the red dots indicate the strong barred galaxies and the green dots indicate the weak barred galaxies.

of spiral galaxies. The following conclusions have been drawn from the above study:

1) The entire data set is classified into 12 homogeneous groups among which group 9 is an outlier. Instead of parameters we have classified the data set with respect to 14, ICs which are linear combinations of various parameters with different weights. This is suitable for a non-Gaussian data set in a multivariate set up. We have found 12 groups by two independent methods, K-means cluster analysis (CA) and Gaussian Mixture Model Based Clustering (GMMBC) with respect to 14, ICs which establishes the robustness of the classification.
2) We have 14 Independent Components among which 8 components are found to be significant and they represent various influential physical galaxy properties like metallicity, ionisation, color, absorption features and velocity dispersion etc. Remaining 6 components do not carry much variation in the galaxy properties. With respect to these 14 components the data set has been classified into 12 homogeneous groups by K-means clustering and the robustness has been established by another widely used method GMMBC.
3) Among these 12 groups, four groups (K2, K4, K11) fall in the oldest age ($\sim$ 2.6 Gyr - 6.75 Gyr) category and the groups (K1, K3, K8, K12) fall in the medium range ($\sim$ 1.68 Gyr - 1.8 Gyr). One group is an outlier (K9) and the remaining two groups (K6 - K7) are the youngest smallest groups of unbarred galaxies.
4) Oldest groups have longest bar lengths, highest metallicities and fall in the redder zone of color-magnitude diagram.

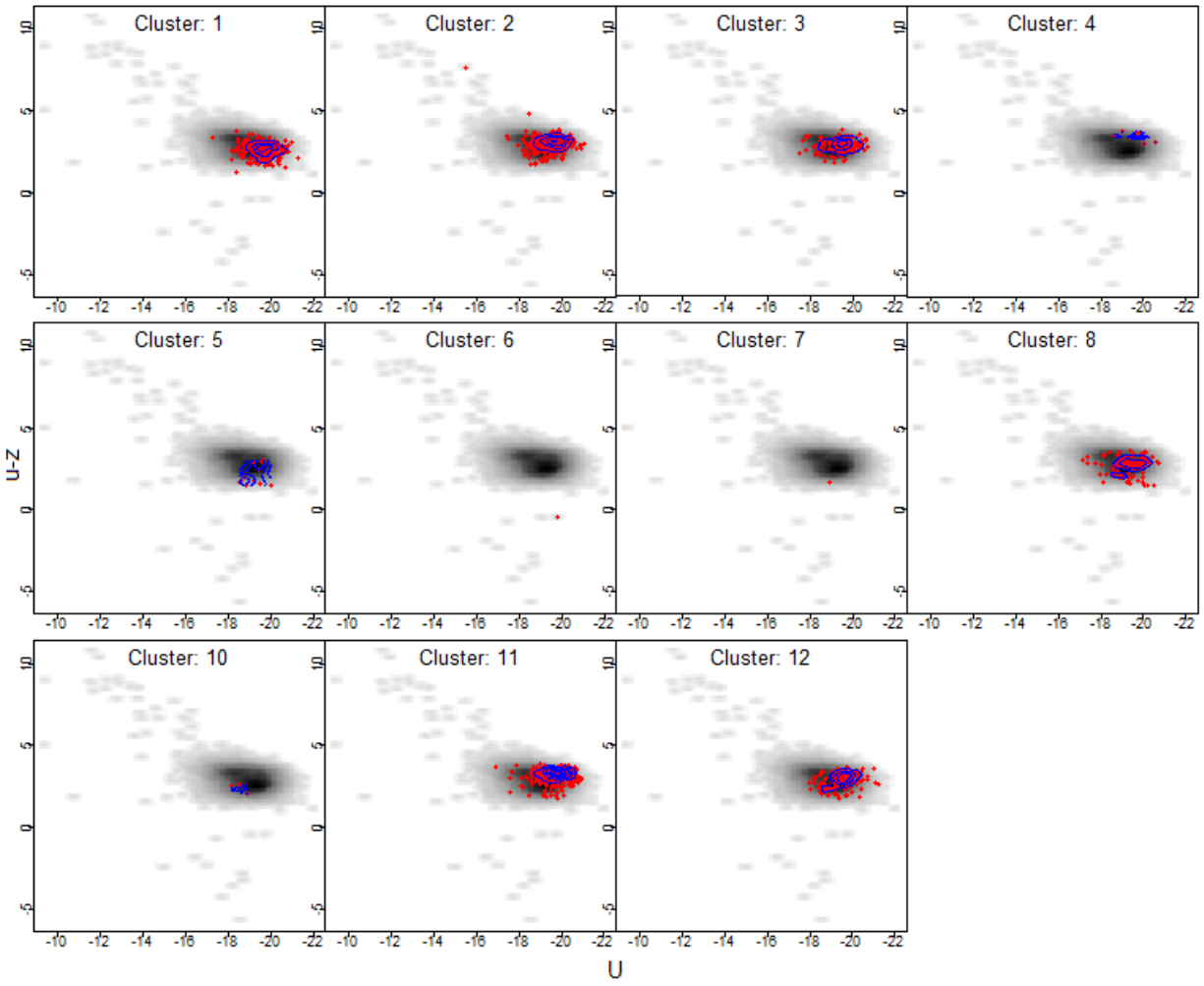

Fig. 8(b): Color–magnitude ($u - z$ vs. U) diagrams of the strong barred galaxies and for each of the groups K1 - K12 except K9.

5) Galaxies of medium age range have shorter bar lengths and the groups are predominated by star burst galaxies, showing a kind of bluer zone.
6) In particular, weak barred galaxies show indication of recurrent bar formation scenario when the color-magnitude, color-color diagrams are studied thoroughly. This is consistent with the theoretical works suggested by various authors (Kormendy and Kennicutt Jr (2004); Gadotti and de Souza (2006)).
7) It has been found that presence of bars may affect the SFE, metallicity, color magnitude and nature of galaxies. When the barred galaxies are oldest, they are redder, their bar lengths are longer and SFE are lower. Few weak barred galaxies, which are precursors of strong barred galaxies, have lower SFE and few weak barred galaxies which are precursors of unbarred galaxies of lower masses have higher SFE. This is the very new feature reflected in the present study and concludes that bar formation is not always one way phenomenon but may get dissolved in course of time in oldest and medium age ranged galaxies.

Acknowledgement

The authors are grateful to the sdss Data Release 15 (sdss DR15) [††] for giving the opportunity to retrieve the data and

[††] https://skyserver.sdss.org/dr15/en/tools/search/sql.aspx

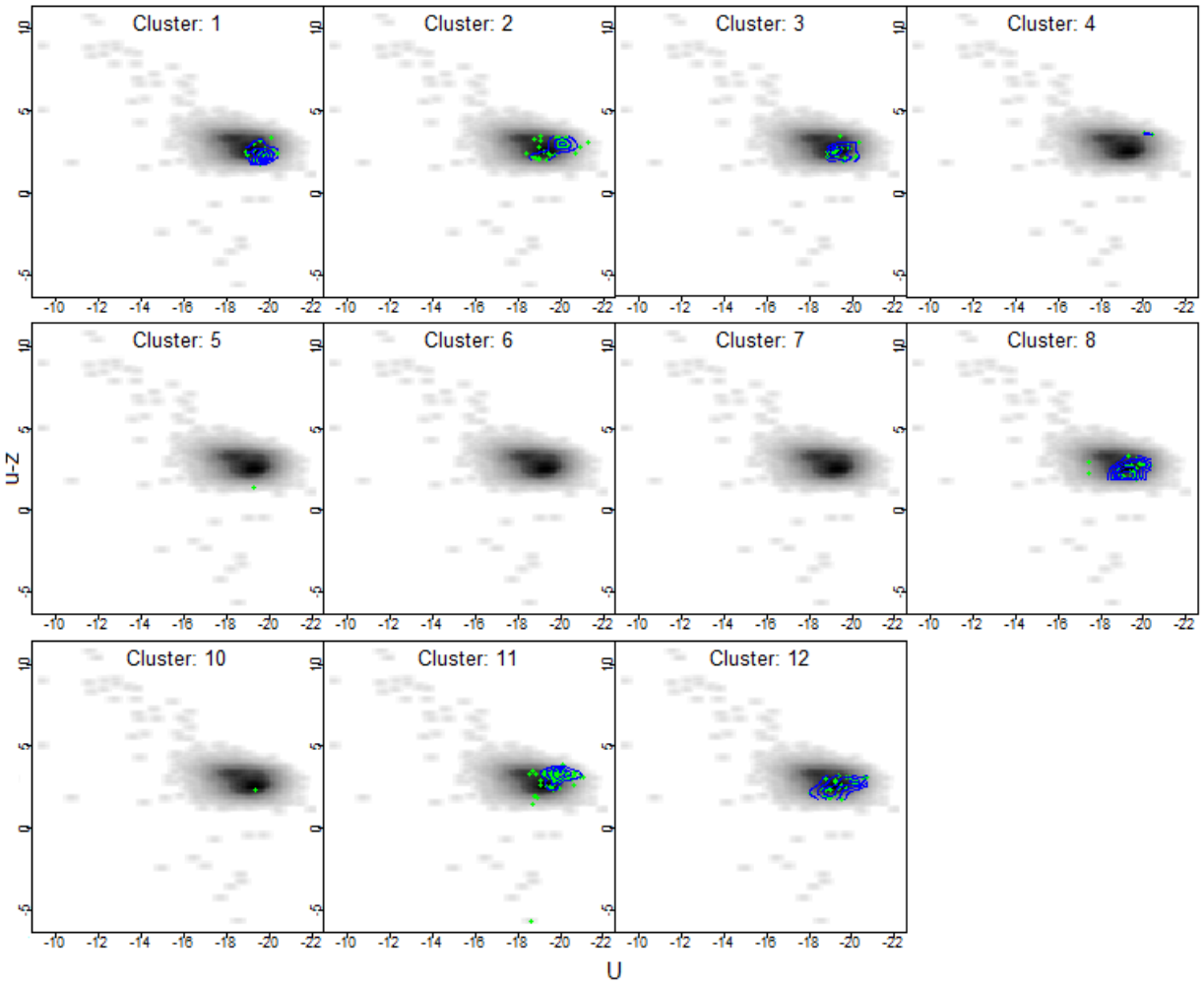

Fig. 8(c): Color–magnitude ($u - z$ vs. U) diagrams of the weak barred galaxies and for each of the groups K1 - K12 except K9.

to work with it. One of the authors Prasenjit Banerjee acknowledges to a research fellowship provided by The University Grants Commission , with UGC Id. NOV2017-422665 & UGC Reference No. 1141/(CSIR-UGC Net June 2017).

References

Vera, M., Alonso, S., and Coldwell, G., "Effect of bars on the galaxy properties," *Astronomy & Astrophysics*, Vol. 595, 2016, p. A63.

Athanassoula, E., "Formation of rings and lenses," *Symposium-International Astronomical Union*, Vol. 100, Cambridge University Press, 1983, pp. 243–252.

Buta, R. J., and Combes, F., *Galactic rings*, Overseas Publishers Association, 1996.

Combes, F., and Elmegreen, B., "Bars in Early Type and Late Type Galaxies," *Astronomy and Astrophysics*, Vol. 271, 1993, p. 391.

Martin, P., "Quantitative morphology of bars in spiral galaxies," *The Astronomical Journal*, Vol. 109, 1995, p. 2428.

Ellison, S. L., Nair, P., Patton, D. R., Scudder, J. M., Mendel, J. T., and Simard, L., "The impact of gas inflows on star formation rates and metallicities in barred galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 416, No. 3, 2011, pp. 2182–2192.

Zhou, Z.-M., Cao, C., and Wu, H., "Star formation properties in barred galaxies. III. Statistical study of bar-driven secular evolution using a sample of nearby barred spirals," *The Astronomical Journal*, Vol. 149, No. 1, 2014, p. 1.

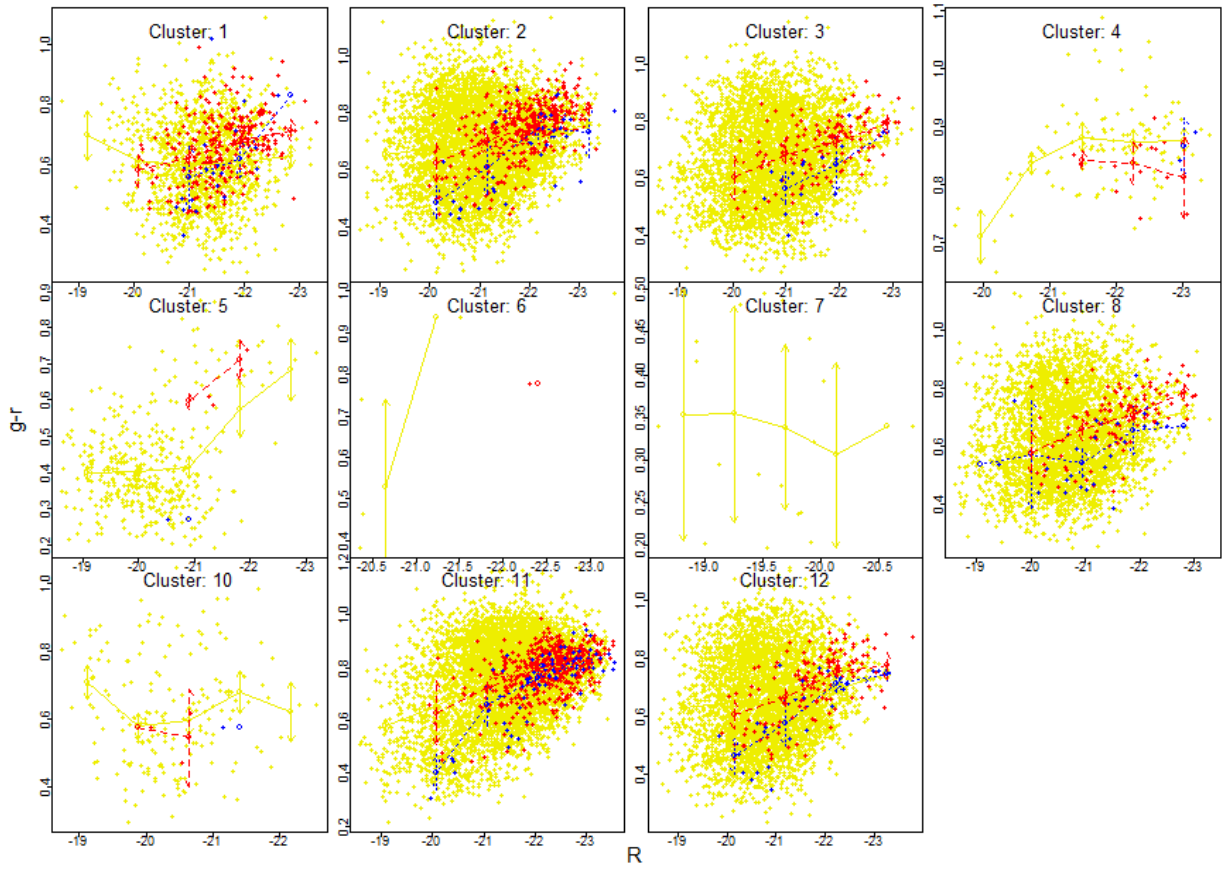

Fig. 9(a): Colour distributions $(g - r)$ are plotted against R values for the groups K1 - K12 except K9 where the yellow dots indicate unbarred galaxies, red dots indicate strong barred galaxies and the blue dots indicate weak barred galaxies.

Debattista, V. P., and Sellwood, J., "Dynamical friction and the distribution of dark matter in barred galaxies," *The Astrophysical Journal Letters*, Vol. 493, No. 1, 1998, p. L5.

Weinberg, M. D., "Evolution of barred galaxies by dynamical friction," *Monthly Notices of the Royal Astronomical Society*, Vol. 213, No. 3, 1985, pp. 451–471.

Debattista, V. P., and Sellwood, J., "Constraints from dynamical friction on the dark matter content of barred galaxies," *The Astrophysical Journal*, Vol. 543, No. 2, 2000, p. 704.

Athanassoula, E., "What determines the strength and the slowdown rate of bars?" *Monthly Notices of the Royal Astronomical Society*, Vol. 341, No. 4, 2003, pp. 1179–1198.

Erwin, P., "What determines the sizes of bars in spiral galaxies?" *Monthly Notices of the Royal Astronomical Society*, Vol. 489, No. 3, 2019, pp. 3553–3564.

Kim, M., Choi, Y.-Y., and Kim, S. S., "Effect of bars on evolution of SDSS spiral galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 494, No. 4, 2020a, pp. 5839–5850.

Garma-Oehmichen, L., Cano-Díaz, M., Hernández-Toledo, H., Aquino-Ortíz, E., Valenzuela, O., Aguerri, J., Sánchez, S., and Merrifield, M., "SDSS-IV MaNGA: bar pattern speed estimates with the Tremaine–Weinberg method and their error sources," *Monthly Notices of the Royal Astronomical Society*, Vol. 491, No. 3, 2020, pp. 3655–3671.

Roberts Jr, W. W., Huntley, J. M., and van Albada, G. D., "Gas dynamics in barred spirals-Gaseous density waves and galactic shocks," *The Astrophysical Journal*, Vol. 233, 1979, pp. 67–84.

Norman, C. A., Sellwood, J., and Hasan, H., "Bar dissolution and bulge formation: An example of secular dynamical evolution in galaxies," *The Astrophysical Journal*, Vol. 462, 1996, p. 114.

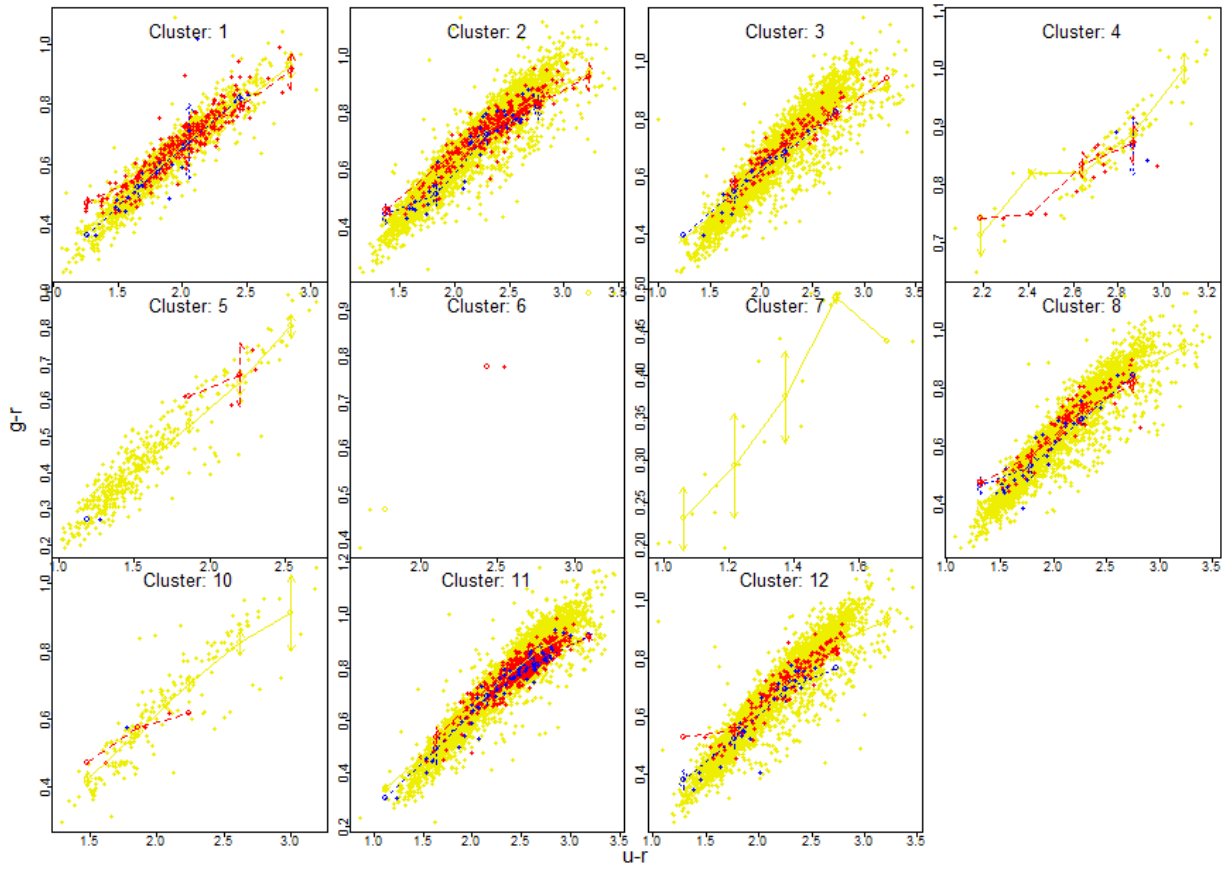

Fig. 9(b): Colour distributions $(g - r)$ are plotted against $(u - r)$ values for the groups K1 - K12 except K9 where the yellow dots indicate unbarred galaxies, red dots indicate strong barred galaxies and the blue dots indicate weak barred galaxies.

Sellwood, J., and Moore, E., "On the formation of disk galaxies and massive central objects," *The Astrophysical Journal*, Vol. 510, No. 1, 1999, p. 125.

Athanassoula, E., Lambert, J., and Dehnen, W., "Can bars be destroyed by a central mass concentration?—I. Simulations," *Monthly Notices of the Royal Astronomical Society*, Vol. 363, No. 2, 2005, pp. 496–508.

Spinoso, D., Bonoli, S., Dotti, M., Mayer, L., Madau, P., and Bellovary, J., "Bar-driven evolution and quenching of spiral galaxies in cosmological simulations," *Monthly Notices of the Royal Astronomical Society*, Vol. 465, No. 3, 2016, pp. 3729–3740.

Barbuy, B., Chiappini, C., and Gerhard, O., "Chemodynamical history of the Galactic Bulge," *Annual Review of Astronomy and Astrophysics*, Vol. 56, 2018, pp. 223–276.

Guo, M., Du, M., Ho, L. C., Debattista, V. P., and Zhao, D., "A New Channel of Bulge Formation via the Destruction of Short Bars," *The Astrophysical Journal*, Vol. 888, No. 2, 2020, p. 65.

Rosas-Guevara, Y., Bonoli, S., Dotti, M., Zana, T., Nelson, D., Pillepich, A., Ho, L. C., Izquierdo-Villalba, D., Hernquist, L., and Pakmor, R., "The buildup of strongly barred galaxies in the TNG100 simulation," *Monthly Notices of the Royal Astronomical Society*, Vol. 491, No. 2, 2020, pp. 2547–2564.

Bournaud, F., and Combes, F., "Gas accretion on spiral galaxies: Bar formation and renewal," *Astronomy & Astrophysics*, Vol. 392, No. 1, 2002, pp. 83–102.

Berentzen, I., Athanassoula, E., Heller, C., and Fricke, K., "The regeneration of stellar bars by tidal interactions: numerical simulations of fly-by encounters," *Monthly Notices of the Royal Astronomical Society*, Vol. 347, No. 1, 2004, pp. 220–236.

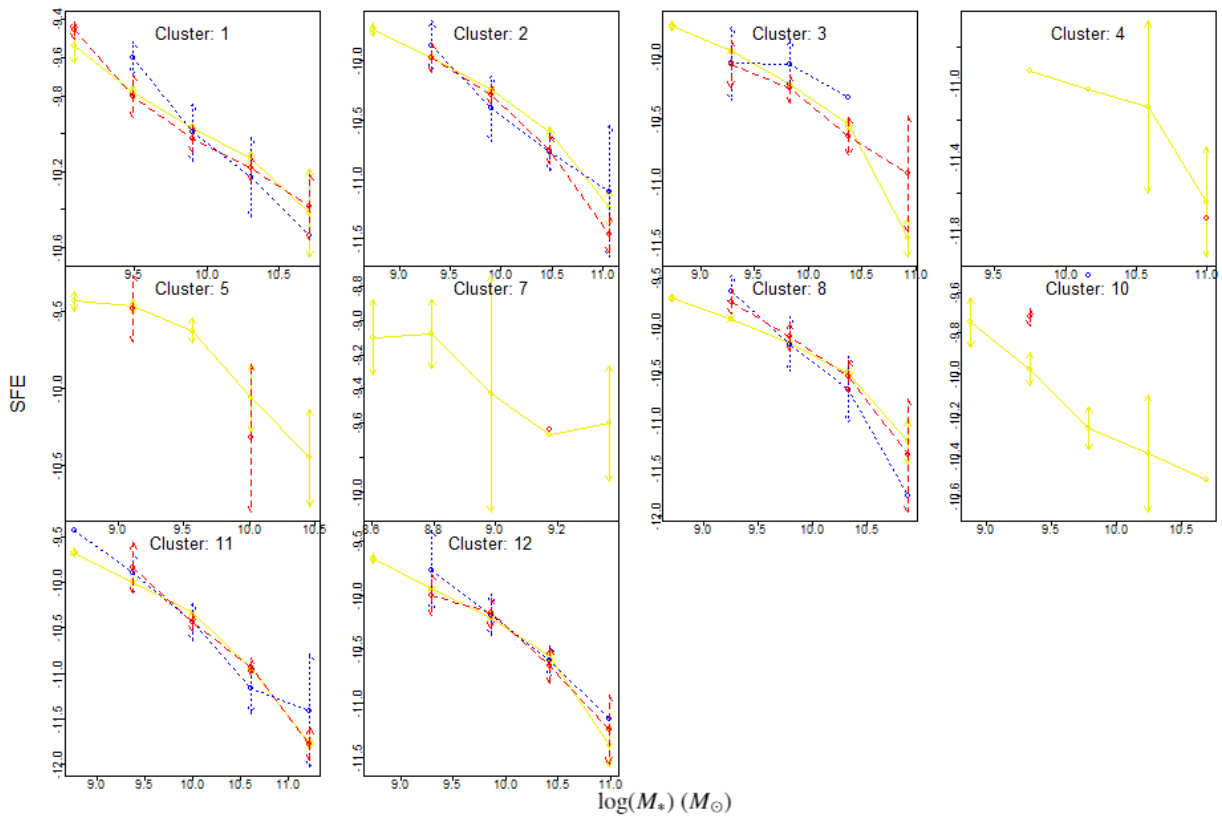

Fig. 10(a): SFE ($yr^{-}1$) values are plotted against log(M_*) ($M_\odot$) for the groups K1 - K12 except K9 where the yellow line is for unbarred galaxies, red line is for strong barred galaxies and the blue line is for weak barred galaxies.

Gadotti, D. A., and de Souza, R. E., "On the lengths, colors, and ages of 18 face-on bars," *The Astrophysical Journal Supplement Series*, Vol. 163, No. 2, 2006, p. 270.

Pettitt, A. R., and Wadsley, J. W., "Bars and spirals in tidal interactions with an ensemble of galaxy mass models," *Monthly notices of the royal astronomical society*, Vol. 474, No. 4, 2018, pp. 5645–5671.

Katz, D., Antoja, T., Romero-Gómez, M., Drimmel, R., Reylé, C., Seabroke, G., Soubiran, C., Babusiaux, C., Di Matteo, P., Figueras, F., et al., "Gaia data release 2-mapping the milky way disc kinematics," *Astronomy & Astrophysics*, Vol. 616, 2018, p. A11.

Hilmi, T., Minchev, I., Buck, T., Martig, M., Quillen, A., Monari, G., Famaey, B., de Jong, R., Laporte, C., Read, J., et al., "Fluctuations in galactic bar parameters due to bar–spiral interaction," *Monthly Notices of the Royal Astronomical Society*, Vol. 497, No. 1, 2020, pp. 933–955.

Kormendy, J., and Kennicutt Jr, R. C., "Secular evolution and the formation of pseudobulges in disk galaxies," *Annu. Rev. Astron. Astrophys.*, Vol. 42, 2004, pp. 603–683.

Debattista, V. P., Carollo, C. M., Mayer, L., and Moore, B., "The kinematic signature of face-on peanut-shaped bulges," *The Astrophysical Journal*, Vol. 628, No. 2, 2005, p. 678.

Debattista, V. P., Mayer, L., Carollo, C. M., Moore, B., Wadsley, J., and Quinn, T., "The secular evolution of disk structural parameters," *The Astrophysical Journal*, Vol. 645, No. 1, 2006, p. 209.

Martinez-Valpuesta, I., Shlosman, I., and Heller, C., "Evolution of stellar bars in live axisymmetric halos: recurrent buckling and secular growth," *The Astrophysical Journal*, Vol. 637, No. 1, 2006, p. 214.

Aguerri, J. A. L., and González-García, A. C., "On the origin of dwarf elliptical galaxies: the fundamental plane," *Astronomy & Astrophysics*, Vol. 494, No. 3, 2009, pp. 891–904.

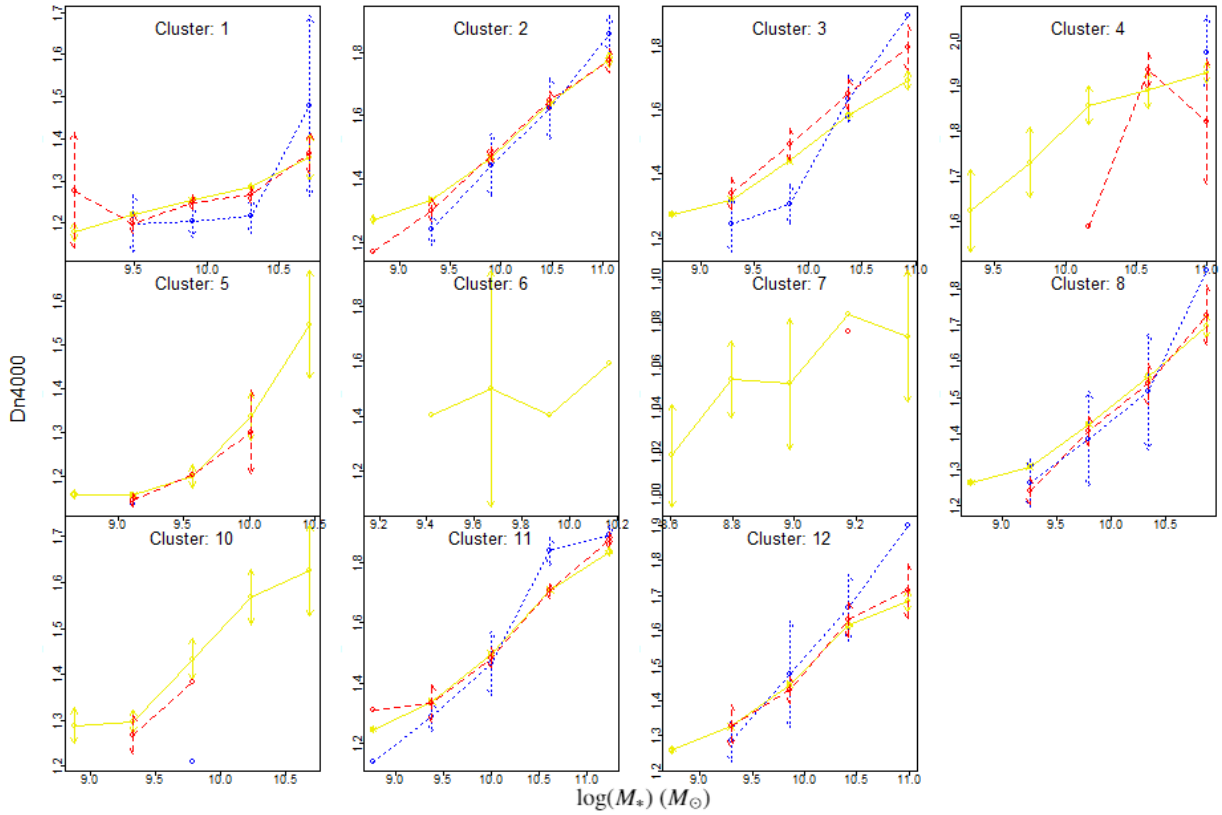

Fig. 10(b): D_n4000 values are plotted against $\log(M_*)(M_\odot)$ for the groups K1 - K12 except K9 where the yellow line is for unbarred galaxies, red line is for strong barred galaxies and the blue line is for weak barred galaxies.

de Lorenzo-Cáceres, A., Sánchez-Blázquez, P., Méndez-Abreu, J., Gadotti, D. A., Falcón-Barroso, J., Martínez-Valpuesta, I., Coelho, P., Fragkoudi, F., Husemann, B., Leaman, R., et al., "Clocking the assembly of double-barred galaxies with the MUSE TIMER project," *Monthly Notices of the Royal Astronomical Society*, Vol. 484, No. 4, 2019, pp. 5296–5314.

Fragkoudi, F., Grand, R. J., Pakmor, R., Blázquez-Calero, G., Gargiulo, I., Gomez, F., Marinacci, F., Monachesi, A., Ness, M., Perez, I., et al., "Chemodynamics of barred galaxies in cosmological simulations: On the Milky Way's quiescent merger history and in-situ bulge," *Monthly Notices of the Royal Astronomical Society*, Vol. 494, No. 4, 2020, pp. 5936–5960.

Barazza, F. D., Jogee, S., and Marinova, I., "Bars in Disk-dominated and Bulge-dominated Galaxies at z˜ 0: New Insights from˜ 3600 SDSS Galaxies," *The Astrophysical Journal*, Vol. 675, No. 2, 2008, p. 1194.

Masters, K. L., Nichol, R., Bamford, S., Mosleh, M., Lintott, C. J., Andreescu, D., Edmondson, E. M., Keel, W. C., Murray, P., Raddick, M. J., et al., "Galaxy Zoo: dust in spiral galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 404, No. 2, 2010, pp. 792–810.

Lintott, C., Schawinski, K., Bamford, S., Slosar, A., Land, K., Thomas, D., Edmondson, E., Masters, K., Nichol, R. C., Raddick, M. J., et al., "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 410, No. 1, 2011, pp. 166–178.

Oh, S., Oh, K., and Sukyoung, K. Y., "Bar Effects on Central Star Formation and Active Galactic Nucleus Activity," *The Astrophysical Journal Supplement Series*, Vol. 198, No. 1, 2011, p. 4.

Alonso, M., Coldwell, G., and Lambas, D. G., "Effect of bars in AGN host galaxies and black hole activity," *Astronomy & Astrophysics*, Vol. 549, 2013, p. A141.

Alonso, S., Coldwell, G., and Lambas, D. G., "AGN spiral galaxies in groups: effects of bars," *Astronomy & Astrophysics*, Vol. 572, 2014, p. A86.

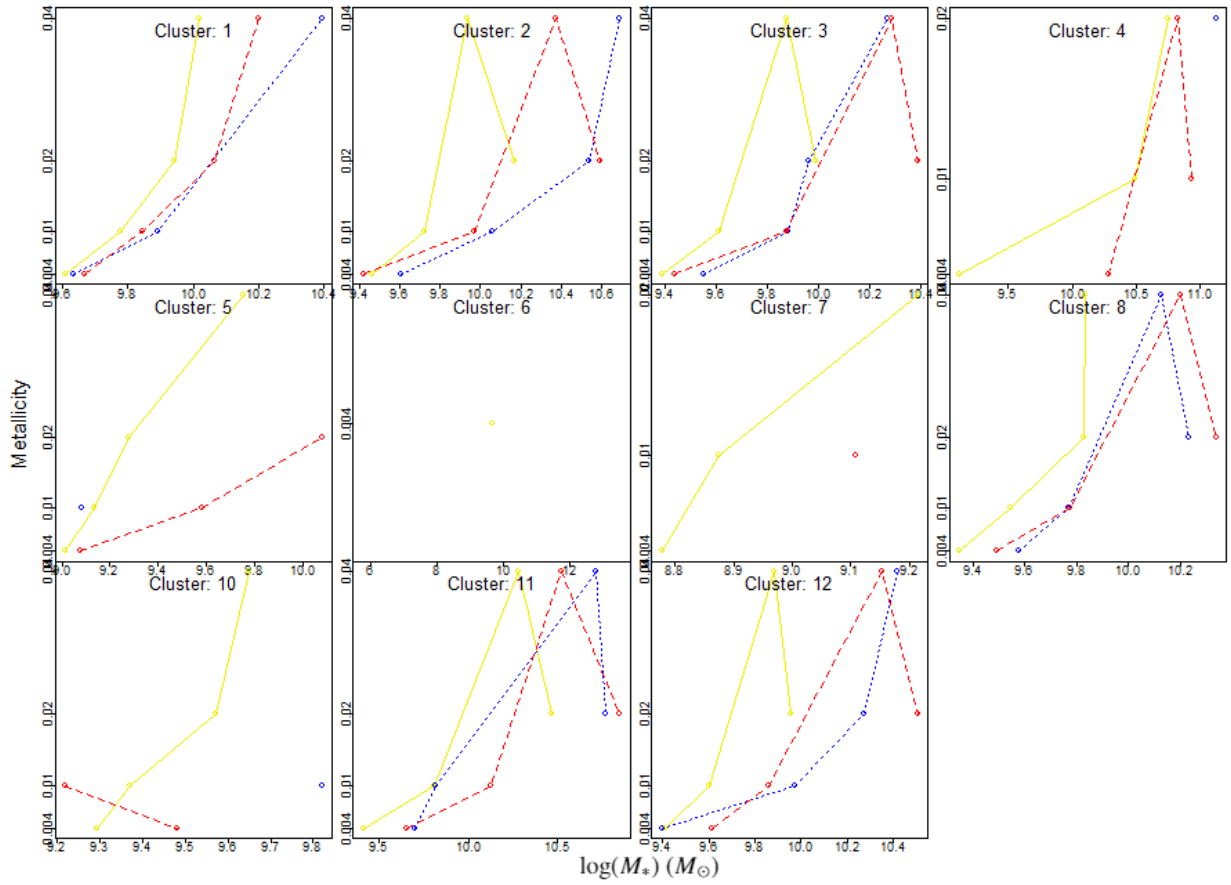

Fig. 11: Metallicity are plotted against $\log(M_*)(M_\odot)$ values for the groups K1 - K12 except K9 where the yellow line is for unbarred galaxies, red line is for strong barred galaxies and the blue line is for weak barred galaxies.

Kim, M., Choi, Y.-Y., and Kim, S. S., "Direct effects of the environment on AGN triggering in SDSS spiral galaxies: merger-AGN connection," *Monthly Notices of the Royal Astronomical Society*, Vol. 491, No. 3, 2020b, pp. 4045–4056.

Cuomo, V., Aguerri, J. A. L., Corsini, E. M., Debattista, V. P., Méndez-Abreu, J., and Pizzella, A., "Bar pattern speeds in CALIFA galaxies-II. The case of weakly barred galaxies," *Astronomy & Astrophysics*, Vol. 632, 2019, p. A51.

Hawarden, T., Mountain, C., Leggett, S., and Puxley, P., "Enhanced star formation–The importance of bars in spiral galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 221, No. 1, 1986, pp. 41P–45P.

Devereux, N., "The spatial distribution of 10 micron luminosity in spiral galaxies," *The Astrophysical Journal*, Vol. 323, 1987, pp. 91–107.

Hummel, E., Van der Hulst, J., Kennicutt, R., and Keel, W., "Environmental impact on the nuclear radio activity in spiral galaxies," *Astronomy and Astrophysics*, Vol. 236, 1990, pp. 333–345.

Pompea, S. M., and Rieke, G. H., "A test of the association of infrared activity with bars," *The Astrophysical Journal*, Vol. 356, 1990, pp. 416–429.

Martinet, L., and Friedli, D., "Bar strength and star formation activity in late-type barred galaxies," *arXiv preprint astro-ph/9701091*, 1997.

Chapelon, S., Contini, T., and Davoust, E., "Starbursts in barred spiral galaxies. V. Morphological analysis of bars," *arXiv preprint astro-ph/9902174*, 1999.

Donohoe-Keyes, C., Martig, M., James, P., and Kraljic, K., "Redistribution of stars and gas in the star formation deserts of barred galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 489, No. 4, 2019, pp. 4992–5003.

Wang, J., Athanassoula, E., Yu, S.-Y., Wolf, C., Shao, L., Gao, H., and Randriamampandry, T., "Suppressed or enhanced central star formation rates in late-type barred galaxies," *The Astrophysical Journal*, Vol. 893, No. 1, 2020, p. 19.

Kim, E., Hwang, H. S., Chung, H., Lee, G.-H., Park, C., Sodi, B. C., and Kim, S. S., "Star Formation Activity of Barred Spiral Galaxies," *The Astrophysical Journal*, Vol. 845, No. 2, 2017, p. 93.

Newnham, L., Hess, K. M., Masters, K. L., Kruk, S., Penny, S. J., Lingard, T., and Smethurst, R., "The H i morphology and stellar properties of strongly barred galaxies: support for bar quenching in massive spirals," *Monthly Notices of the Royal Astronomical Society*, Vol. 492, No. 4, 2020, pp. 4697–4715.

Vila-Costas, M., and Edmunds, M., "The relation between abundance gradients and the physical properties of spiral galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 259, No. 1, 1992, pp. 121–145.

Martin, P., and Roy, J.-R., "The influence of bars on the chemical composition of spiral galaxies," *The Astrophysical Journal*, Vol. 424, 1994, pp. 599–614.

Sánchez-Blázquez, P., Rosales-Ortega, F., Méndez-Abreu, J., Pérez, I., Sánchez, S., Zibetti, S., Aguerri, J., Bland-Hawthorn, J., Catalán-Torrecilla, C., Fernandes, R. C., et al., "Stellar population gradients in galaxy discs from the CALIFA survey-The influence of bars," *Astronomy & Astrophysics*, Vol. 570, 2014, p. A6.

Carles, C., Martel, H., Ellison, S. L., and Kawata, D., "The mass dependence of star formation histories in barred spiral galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 463, No. 1, 2016, pp. 1074–1087.

Kruk, S. J., Lintott, C. J., Bamford, S. P., Masters, K. L., Simmons, B. D., Häußler, B., Cardamone, C. N., Hart, R. E., Kelvin, L., Schawinski, K., et al., "Galaxy Zoo: secular evolution of barred galaxies from structural decomposition of multiband images," *Monthly Notices of the Royal Astronomical Society*, Vol. 473, No. 4, 2018, pp. 4731–4753.

Seo, W.-Y., Kim, W.-T., Kwak, S., Hsieh, P.-Y., Han, C., and Hopkins, P. F., "Effects of Gas on Formation and Evolution of Stellar Bars and Nuclear Rings in Disk Galaxies," *The Astrophysical Journal*, Vol. 872, No. 1, 2019, p. 5.

Cavanagh, M., and Bekki, K., "Bars formed in galaxy merging and their classification with deep learning," *Astronomy & Astrophysics*, Vol. 641, 2020, p. A77.

Chattopadhyay, T., Fraix-Burnet, D., and Mondal, S., "Unsupervised Classification of Galaxies. I. Independent Component Analysis Feature Selection," *Publications of the Astronomical Society of the Pacific*, Vol. 131, No. 1004, 2019, p. 108010.

Fraix-Burnet, D., Thuillard, M., and Chattopadhyay, A. K., "Multivariate approaches to classification in extragalactic astronomy," *Frontiers in Astronomy and Space Sciences*, Vol. 2, 2015, p. 3.

Whitmore, B. C., and Forbes, D. A., "An Objective Classification System for Spiral Galaxies and," *The World of Galaxies: Proceedings of the Conference "Le Monde des Galaxies" Held 12–14 April 1988 at the Institut d'Astrophysique de Paris in Honor of Gérard and Antoinette de Vaucouleurs on the Occasion of His 70th Birthday*, Springer Science & Business Media, 2012, p. 95.

Cabanac, R. A., de Lapparent, V., and Hickson, P., "Classification and redshift estimation by principal component analysis," *Astronomy & Astrophysics*, Vol. 389, No. 3, 2002, pp. 1090–1116.

Chattopadhyay, T., and Chattopadhyay, A. K., "Objective classification of spiral galaxies having extended rotation curves beyond the optical radius," *The Astronomical Journal*, Vol. 131, No. 5, 2006, p. 2452.

Peth, M. A., Lotz, J. M., Freeman, P. E., McPartland, C., Mortazavi, S. A., Snyder, G. F., Barro, G., Grogin, N. A., Guo, Y., Hemmati, S., et al., "Beyond spheroids and discs: classifications of CANDELS galaxy structure at 1.4< z< 2 via principal component analysis," *Monthly Notices of the Royal Astronomical Society*, Vol. 458, No. 1, 2016, pp. 963–987.

Chang, W.-C., "On using principal components before separating a mixture of two multivariate normal distributions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 32, No. 3, 1983, pp. 267–275.

Ellis, S., Driver, S., Allen, P., Liske, J., Bland-Hawthorn, J., and De Propris, R., "The Millennium Galaxy Catalogue: on the natural subdivision of galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 363, No. 4, 2005, pp. 1257–1271.

Chattopadhyay, T., and Chattopadhyay, A. K., "Globular clusters of the Local Group–statistical classification," *Astronomy & Astrophysics*, Vol. 472, No. 1, 2007, pp. 131–140.

Chattopadhyay, T., Misra, R., Chattopadhyay, A. K., and Naskar, M., "Statistical evidence for three classes of gamma-ray bursts," *The Astrophysical Journal*, Vol. 667, No. 2, 2007, p. 1017.

Mondal, S., Chattopadhyay, A. K., and Chattopadhyay, T., "Globular clusters in the Milky Way and dwarf galaxies: A distribution-free statistical comparison," *The Astrophysical Journal*, Vol. 683, No. 1, 2008, p. 172.

Chattopadhyay, A. K., Chattopadhyay, T., Davoust, E., Mondal, S., and Sharina, M., "Study of ngc 5128 globular clusters under multivariate statistical paradigm," *The Astrophysical Journal*, Vol. 705, No. 2, 2009, p. 1533.

Babu, G. J., Chattopadhyay, T., Chattopadhyay, A. K., and Mondal, S., "Horizontal branch morphology of globular clusters: A multivariate statistical analysis," *The Astrophysical Journal*, Vol. 700, No. 2, 2009, p. 1768.

Almeida, J. S., Aguerri, J. A. L., Munoz-Tunón, C., and De Vicente, A., "Automatic unsupervised classification of all sloan digital sky survey data release 7 galaxy spectra," *The Astrophysical Journal*, Vol. 714, No. 1, 2010, p. 487.

Fraix-Burnet, D., Dugué, M., Chattopadhyay, T., Chattopadhyay, A. K., and Davoust, E., "Structures in the fundamental plane of early-type galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 407, No. 4, 2010, pp. 2207–2222.

Fraix-Burnet, D., Chattopadhyay, T., Chattopadhyay, A. K., Davoust, E., and Thuillard, M., "A six-parameter space to describe galaxy diversification," *Astronomy & Astrophysics*, Vol. 545, 2012, p. A80.

De, T., Fraix Burnet, D., and Chattopadhyay, A. K., "Clustering large number of extragalactic spectra of galaxies and quasars through canopies," *Communications in Statistics-Theory and Methods*, Vol. 45, No. 9, 2016, pp. 2638–2653.

Modak, S., Chattopadhyay, T., and Chattopadhyay, A. K., "Two phase formation of massive elliptical galaxies: study through cross-correlation including spatial effect," *Astrophysics and Space Science*, Vol. 362, No. 11, 2017, p. 206.

Modak, S., Chattopadhyay, T., and Chattopadhyay, A. K., "Unsupervised classification of eclipsing binary light curves through k-medoids clustering," *Journal of Applied Statistics*, Vol. 47, No. 2, 2020, pp. 376–392.

Pires, S., Juin, J., Yvon, D., Moudden, Y., Anthoine, S., and Pierpaoli, E., "Sunyaev-Zel'dovich cluster reconstruction in multiband bolometer camera surveys," *Astronomy & Astrophysics*, Vol. 455, No. 2, 2006, pp. 741–755.

Pike, S., Ebisawa, K., Ikeda, S., Morii, M., Mizumoto, M., and Kusunoki, E., "Application of data science techniques to disentangle X-ray spectral variation of super-massive black holes," *arXiv preprint arXiv:1701.05386*, 2017.

Martins-Filho, W., Griffith, C., Pearson, K., Waldmann, I., Alvarez-Candal, A., and Zellem, R. T., "Independent Component Analysis applied to Ground-based observations," *AAS*, Vol. 231, 2018, pp. 148–26.

Sheldon, K., and Richards, G., "Emission Line Correlations as Diagnostics of Quasar Winds," *AAS*, Vol. 231, 2018, pp. 250–23.

Richardson, C. T., Allen, J. T., Baldwin, J. A., Hewett, P. C., Ferland, G. J., Crider, A., and Meskhidze, H., "Interpreting the ionization sequence in star-forming galaxy emission-line spectra," *Monthly Notices of the Royal Astronomical Society*, Vol. 458, No. 1, 2016, pp. 988–1012.

Sarro, L., Ordieres-Meré, J., Bello-García, A., González-Marcos, A., and Solano, E., "Estimates of the atmospheric parameters of M-type stars: a machine-learning perspective," *Monthly Notices of the Royal Astronomical Society*, Vol. 476, No. 1, 2018, pp. 1120–1139.

Mu, B., "Unsupervised spectral classification of astronomical x-ray sources based on independent component analysis," *Thesis. Rochester Institute of Technology*, 2007.

Das, S., Chattopadhayay, T., and Davoust, E., "Multivariate analysis of the globular clusters in M87," *Publications of the Astronomical Society of Australia*, Vol. 32, 2015.

Baldwin, J. A., Phillips, M. M., and Terlevich, R., "Classification parameters for the emission-line spectra of extragalactic objects." *Publications of the Astronomical Society of the Pacific*, Vol. 93, No. 551, 1981, p. 5.

Veilleux, S., and Osterbrock, D. E., "Spectral classification of emission-line galaxies," *The Astrophysical Journal Supplement Series*, Vol. 63, 1987, pp. 295–310.

Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., White, S. D., Ridgway, S. E., Brinkmann, J., Fukugita, M., Hall, P. B., et al., "The host galaxies of active galactic nuclei," *Monthly Notices of the Royal Astronomical Society*, Vol. 346, No. 4, 2003, pp. 1055–1077.

Kewley, L. J., Groves, B., Kauffmann, G., and Heckman, T., "The host galaxies and classification of active galactic nuclei," *Monthly Notices of the Royal Astronomical Society*, Vol. 372, No. 3, 2006, pp. 961–976.

Kewley, L. J., Dopita, M. A., Leitherer, C., Davé, R., Yuan, T., Allen, M., Groves, B., and Sutherland, R., "Theoretical evolution of optical strong lines across cosmic time," *The Astrophysical Journal*, Vol. 774, No. 2, 2013, p. 100.

Shapiro, S. S., and Wilk, M. B., "An analysis of variance test for normality (complete samples)," *Biometrika*, Vol. 52, No. 3/4, 1965, pp. 591–611.

Stephens, M. A., "EDF statistics for goodness of fit and some comparisons," *Journal of the American statistical Association*, Vol. 69, No. 347, 1974, pp. 730–737.

Lilliefors, H. W., "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American statistical Association*, Vol. 62, No. 318, 1967, pp. 399–402.

Villasenor Alva, J. A., and Estrada, E. G., "A generalization of Shapiro–Wilk's test for multivariate normality," *Communications in Statistics—Theory and Methods*, Vol. 38, No. 11, 2009, pp. 1870–1883.

Brosche, P., "The manifold of galaxies. Galaxies with known dynamical parameters," *Astronomy and Astrophysics*, Vol. 23, 1973, pp. 259–268.

Murtagh, F., and Heck, A., *Multivariate data analysis*, Vol. 131, Springer Science & Business Media, 2012.

Hyvärinen, A., "Independent component analysis in the presence of gaussian noise by maximizing joint likelihood," *Neurocomputing*, Vol. 22, No. 1-3, 1998, pp. 49–67.

Hyvärinen, A., "The fixed-point algorithm and maximum likelihood estimation for independent component analysis," *Neural Processing Letters*, Vol. 10, No. 1, 1999a, pp. 1–5.

Hyvarinen, A., "Gaussian moments for noisy independent component analysis," *IEEE signal processing letters*, Vol. 6, No. 6, 1999b, pp. 145–147.

Pfister, N., Weichwald, S., Bühlmann, P., and Schölkopf, B., "Robustifying independent component analysis by adjusting for group-wise stationary noise," *Journal of Machine Learning Research*, Vol. 20, No. 147, 2019, pp. 1–50.

Chattopadhyay, A. K., Mondal, S., and Chattopadhyay, T., "Independent Component Analysis for the objective classification of globular clusters of the galaxy NGC 5128," *Computational Statistics & Data Analysis*, Vol. 57, No. 1, 2013a, pp. 17–32.

Comon, P., "Independent component analysis, a new concept?" *Signal processing*, Vol. 36, No. 3, 1994, pp. 287–314.

Chattopadhyay, A. K., Chattyopadhyay, T., De, T., and Mondal, S., "Independent component analysis for dimension reduction classification: Hough transform and CASH algorithm," *Astrostatistical Challenges for the New Astronomy*, Springer, 2013b, pp. 185–202.

Hastie, T., and Tibshirani, R., "Independent components analysis through product density estimation," *Advances in neural information processing systems*, 2003, pp. 665–672.

Hyvärinen, A., and Oja, E., "Independent component analysis: algorithms and applications," *Neural networks*, Vol. 13, No. 4-5, 2000, pp. 411–430.

Albazzaz, H., and Wang, X. Z., "Statistical process control charts for batch operations based on independent component analysis," *Industrial & engineering chemistry research*, Vol. 43, No. 21, 2004, pp. 6731–6741.

Eloyan, A., and Ghosh, S. K., "A semiparametric approach to source separation using independent component analysis," *Computational statistics & data analysis*, Vol. 58, 2013, pp. 383–396.

Chattopadhyay, T., Sharina, M., and Karmakar, P., "Statistical analysis of dwarf galaxies and their globular clusters in the Local Volume," *The Astrophysical Journal*, Vol. 724, No. 1, 2010, p. 678.

Kairov, U., Cantini, L., Greco, A., Molkenov, A., Czerwinska, U., Barillot, E., and Zinovyev, A., "Determining the optimal number of independent components for reproducible transcriptomic data analysis," *BMC genomics*, Vol. 18, No. 1, 2017, p. 712.

Agarwal, S., Lim, J., Zelnik-Manor, L., Perona, P., Kriegman, D., and Belongie, S., "Beyond pairwise clustering," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2, IEEE, 2005, pp. 838–845.

Himberg, J., Hyvärinen, A., and Esposito, F., "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage*, Vol. 22, No. 3, 2004, pp. 1214–1222.

Feldman, D., *Algorithms for Finding the Optimal K-line-means*, Tel Aviv University, 2003.

Sugar, C. A., and James, G. M., "Finding the number of clusters in a dataset: An information-theoretic approach," *Journal of the American Statistical Association*, Vol. 98, No. 463, 2003, pp. 750–763.

Tibshirani, R., Walther, G., and Hastie, T., "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 2, 2001, pp. 411–423.

Dunn, J. C., "Well-separated clusters and optimal fuzzy partitions," *Journal of cybernetics*, Vol. 4, No. 1, 1974, pp. 95–104.

Rousseeuw, P. J., "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, Vol. 20, 1987, pp. 53–65.

Rao, C. R., RadhakrishnaRao, C., Kleffe, J., and Kleffe, J., *Estimation of variance components and applications*, Vol. 3, North Holland, 1988.

Stasińska, G., Fernandes, R. C., Mateus, A., Sodré, L., and Asari, N. V., "Semi-empirical analysis of Sloan Digital Sky Survey galaxies–III. How to distinguish AGN hosts," *Monthly Notices of the Royal Astronomical Society*, Vol. 371, No. 2, 2006, pp. 972–982.

Kewley, L. J., Dopita, M., Sutherland, R., Heisler, C., and Trevena, J., "Theoretical modeling of starburst galaxies," *The Astrophysical Journal*, Vol. 556, No. 1, 2001, p. 121.

Brinchmann, J., Charlot, S., White, S., Tremonti, C., Kauffmann, G., Heckman, T., and Brinkmann, J., "The physical properties of star-forming galaxies in the low-redshift Universe," *Monthly Notices of the Royal Astronomical Society*, Vol. 351, No. 4, 2004, pp. 1151–1179.

Balogh, M. L., Morris, S. L., Yee, H., Carlberg, R., and Ellingson, E., "Differential galaxy evolution in cluster and field galaxies at z 0.3," *The Astrophysical Journal*, Vol. 527, No. 1, 1999, p. 54.

Masters, K. L., Nichol, R. C., Haynes, M. P., Keel, W. C., Lintott, C., Simmons, B., Skibba, R., Bamford, S., Giovanelli, R., and Schawinski, K., "Galaxy Zoo and ALFALFA: atomic gas and the regulation of star formation in barred disc galaxies," *Monthly Notices of the Royal Astronomical Society*, Vol. 424, No. 3, 2012, pp. 2180–2192.

Ho, L. C., Filippenko, A. V., and Sargent, W. L., "A Search for "Dwarf"Seyfert Nuclei. V. Demographics of Nuclear Activity in Nearby Galaxies," *The Astrophysical Journal*, Vol. 487, No. 2, 1997, p. 568.

Sheth, K., Vogel, S. N., Regan, M. W., Thornley, M. D., and Teuben, P. J., "Secular evolution via Bar-driven gas inflow: Results from BIMA SONG," *The Astrophysical Journal*, Vol. 632, No. 1, 2005, p. 217.

Lee, G.-H., Woo, J.-H., Lee, M. G., Hwang, H. S., Lee, J. C., Sohn, J., and Lee, J. H., "Do bars trigger activity in galactic nuclei?" *The Astrophysical Journal*, Vol. 750, No. 2, 2012, p. 141.

Jogee, S., Scoville, N., and Kenney, J. D., "The central region of barred galaxies: Molecular environment, starbursts, and secular evolution," *The Astrophysical Journal*, Vol. 630, No. 2, 2005, p. 837.

Ellison, S. L., Patton, D. R., Simard, L., and McConnachie, A. W., "Clues to the origin of the mass-metallicity relation: Dependence on star formation rate and galaxy size," *The Astrophysical Journal Letters*, Vol. 672, No. 2, 2007, p. L107.