AST425 Project Proposal: Relative Velocities Within Galaxies in Relation to Galactic Morphological Evolution Using the MaNGA Survey

Juan Pablo Alfonzo
juanpablo.alfonzo@mail.utoronto.ca

October 22 2020

Abstract

This study will focus on studying morphological changes that galaxies undergo as they age. It will explore the the transition period between star forming galaxies (SFG) to quiescent galaxies (QG). This transition is a relatively fast one in comparison to a galaxy's life span, and hence not much is known about the drastic change in morphology that occurs when a galaxy's star forming rate (SFR) plummets. Galaxies in this transitional phase with rapidly decreasing SFR are known as green valley galaxies (GVG), and relatively few of them have been observed given how short the transition phase is. The hope of this study is to gain deeper insights into this transitional phase of galaxy evolution by studying the relative velocities of stars and gasses within galaxies. The goal is to see if there is a trend between these parameters in SFGs, QGs, and GVGs, and whether we can notice any trends amongst these parameters across galaxies of the same type. The study will use principal component analysis (PCA) to look for these trends in the data. The hope is that with this we can get a better idea of how the morphology of the galaxy is changing during this rapid transition phase. Additionally, the study aims to find a relation between galaxy type and its internal relative velocities.

1 Motivation of Proposed Project

The study of galaxy formation and their evolution through time is an area where much research has been done, as it is critical to getting a proper picture of our universe, yet one where there is still much which we do not understand. The main way we can study how galaxies form and how they change in time is by studying their stellar populations. Not only how the stars happen to be spread throughout the galaxy, but also where we find star forming regions. The most important metric being the star forming rate (SFR), or how many stars the galaxy is making in a given amount of time. In fact one of the key ways we can tell the age of galaxy is by observing its SFR regions that may be scattered throughout the galaxy. By studying these star forming regions physical properties, such as their relative velocity compared to the galaxy or their flux, we can deduce information about the galaxy as a whole. Younger galaxies tend to have bigger and more active star forming regions compared to that of older galaxies, hence they have a higher SFR. These younger galaxies are often called star forming galaxies and late type galaxies (Spiral Galaxies), while the older galaxies are often called quiescent galaxies and early type galaxies (Elliptical Galaxies)[3]. The early/late naming distinction arises from the idea of early type stars (young and hot stars of spectral class O and B), and late type stars (old stars of cooler spectral types). The irony in this is that it is now evident that late type galaxies actually contain mostly early type stars, while early type galaxies actually contain mostly late type stars. To avoid this confusion in naming I will simply refer to late style galaxies as star forming galaxies (SFG), and to early type galaxies as quiescent galaxies (QG) from here on.

A question that is of specific interest to understanding how galaxies evolve in time is understanding the transition period between an SFG to a QG. We have seen that although this transition last a very short time, a few Gyr, it involves profound morphological and structural transformations of those galaxies, as well as dramatic ageing of their average stellar populations.[3] This rapid transition of a galaxy going from an SFG to a QG is known as quenching or rapid halting of the star-formation activity. Additionally, we see that this process happens in a relatively fast burst as current SFGs are observed to steadily decrease their SFR over a period time of time while showing only slightly older stellar populations on average. This means the aging of galaxies seems to be a smooth process until some threshold is reached in which rapid change begins to occur within the galaxy. Galaxies which are in this rapid transition phase between SFG and QG are often called green valley galaxies (GVG). There is a much lower observed amount of GVGs in comparison to SFGs and QGs. This is believed to be a consequence of how fast the transformation to SFG to QG occurs, and hence GVGs do not exist for very long in comparison to other galaxies[3]. Figure 1 below shows their scarcity, as well as showing why they are called green valley galaxies.

2 Proposed Research

The research I intend to do will focus on this rapid transition period of galaxies when they go from SFGs to QGs. The main aim of this research will be to examine properties of GVGs and compare them to those of SFGs and QGs galaxies. This will be done using information from the MaNGA (Mapping Nearby Galaxies at APO (Apache Point Observatory)) survey which gives spectrographic data on galaxies. Unlike previous Sloan Digital Sky Survey (SDSS) surveys which obtained spectra only at the centers of target galaxies, MaNGA enables spectral measurements across the face of each of 10,000 nearby galaxies thanks to

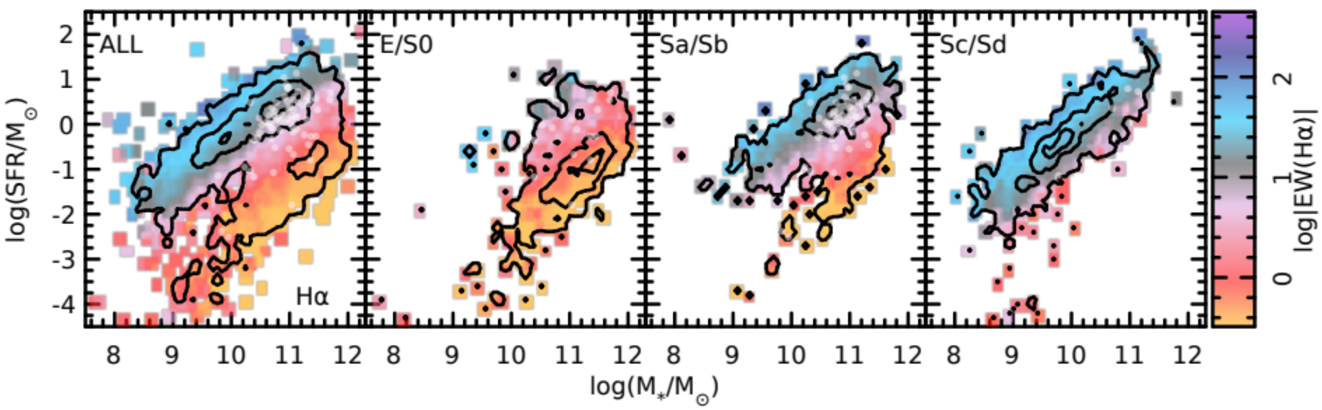

Figure 1: A graph that compares stellar mass of galaxies to their SFR as well as the effective width of their Hα lines. The red regions indicate QG type galaxies, and the blue regions indicate SFG type galaxies. As expected for late type spiral galaxies (Sa,Sb,Sc,Sd) we see mostly blue, with more blue dominance with the later the galaxy type, and red dominance in early type galaxies (E/S0). The so called GVGs lie in between these two large red and blue "mountains", hence they are in a "green valley". This can best been seen in the "ALL" part of the figure, but still very faintly showing how few there are in comparison to SFGs and QGs.[3]

17 simultaneous "integral field units" (IFUs), each composed of tightly-packed arrays of optical fibers[1]. This allows for tracking of various important parameters within galaxies such as: two-dimensional maps of stellar velocity and velocity dispersion, mean stellar age and star formation history, stellar metallicity, element abundance ratio, stellar mass surface density, ionized gas velocity, ionized gas metallicity, star formation rate and dust extinction. The main parameters that I wish to compare amongst these galaxies are their stellar velocities and the velocities of gases that give us information about the star forming going on within the galaxy. Of course the most important parameter will be the SFR of the galaxies as this number will tell us if the galaxy we are looking at should be classified as a SFG, QG or GVG. In figure 2 below I have shown two sample maps provided by MaNGA survey as an illustration of the type of data that is available.

The way I will be accessing the raw data from the MaNGA survey will be through an interface called Marvin SDSS. Marvin is a complete ecosystem designed for overcoming the challenge of searching, accessing, and visualizing the MaNGA data. It consists of three components: a Web App, a Python package of Tools, and an API (application programming interface). Marvin combines these components to provide a seamless experience when using MaNGA Data[2]. The plots in figure 2 were plotted using Marvin via the python packages

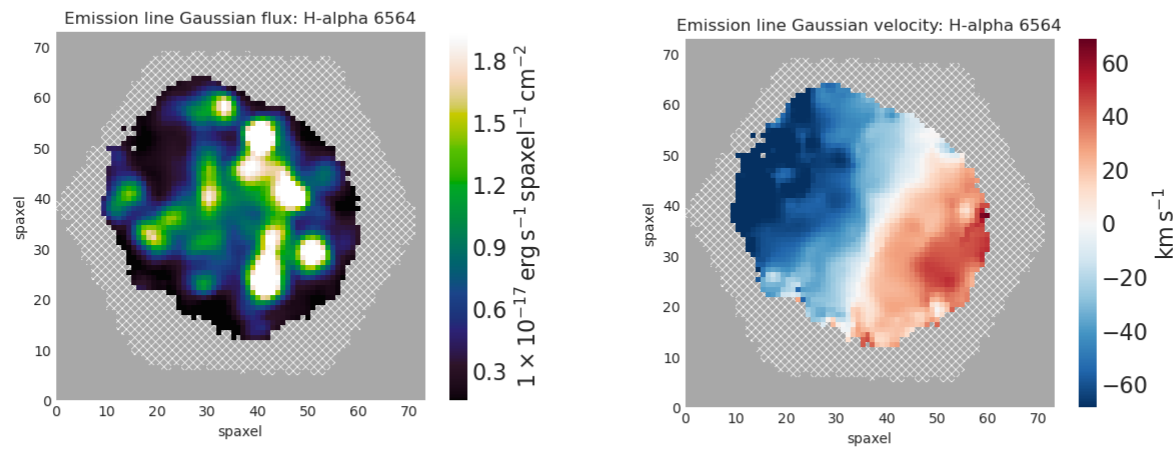

Figure 2: A sample of two parameters within galaxies that are contained in the MaNGA survey. The figure on the left shows us the Hα (which is a specific deep-red visible spectral line in the Balmer series with a wavelength of 656.28 nm in air) flux using emission lines from the galaxy. It tells us both the strength of the flux, and its location within the galaxy on an xy plane that can be converted into a RA/DEC coordinates. The figure on the right shows the relative velocity of this Hα gas with respect to the center of the galaxy. Additionally the colour tells us if the gas is spinning into the plane (blue) or out of the plane (red), and the magnitude of the velocity. Both of these plots were generated using python based code that relied on the Marvin SDSS library[2].

it provides. The package additionally allows to extract the data for every pixel seen in the graph as a numpy array. Using this I will look for trends in the relative velocities of both stars and gasses which are indicative of star formation. The types of gasses present in a given galaxy will help indicate which type of galaxy we are dealing with (SFG, QG, or GVG). Additionally time will spend at the start of the project to replicate the right most rectangle 1. Based on the results from this we will have a sample of galaxies to work with from each region. Additionally, this exercise will allow for the project to have a better scope in terms of a better understanding of the galaxies being used, and also allow for a clear control to be established on the data being used. Additionally, further literature from the review cited in this proposal will be consulted to see how others have gone about making a distinction between these three types of galaxies. Having classified a galaxy in question into one of the groups, or possible seeing which ones seem like they could fall into multiple groups or none at all, I will begin to study the relative velocities within the galaxies. Since the data MaNGA provides is very rich, meaning there are many dimensions of data given each pixel is 3 dimensional and each galaxy image contains thousands of pixels. Since I will be looking at large sample size of galaxies to look for trends I will use PCA (Principle

Component Analysis) algorithms to help make trends in this vast amount of data visible. I will first explain here briefly how PCA techniques work, and then explain how I plan to use them in this project.

Briefly, PCA works by taking multiple variables and plotting them against each other. This in essence turns each variable at play into a dimension. What PCA then does is create a line of best fit using least squares, and ensures this line also passes through the origin of the graph. This line is called PC1 (Principle Component 1). From here it makes n (where n is the number of dimensions or variables) lines that are orthogonal to PC1, and also pass through the origin. These lines are known as the PC1-PCn lines. We can then graph this on what is known as a PCA graph, in which the axes of the graph are the PC lines and the projections of the data points indicate where to graph them on the PCA graph. We can consider this as a change of basis transformation in which the PC lines become the new basis vectors of the space. From here we can measure the variance of each line with respect to the data as follows:

$$\sigma_n = \frac{d_1^2 + d_2^2 + ... + d_k^2}{l - 1} \tag{1}$$

Where σ_n is the variance of a given PC line, $(d_1, .., d_k)$ are the distances of the projected points on the PC line to the origin, and l is the number of total data points

What this formula allows us to do is see is which PC is responsible for the most variance in the data. In other words which PC has the biggest impact on the data. This is extremely useful because it allows us to do dimension reduction on our data. This means if we have a data set with 4 dimensions, but we find 95% of the variance lies in two of the PC lines. Then we can simply use these two PC lines as axis and reduce our 4 dimensional data to 2 dimensions, making it much easier to work with and find trends within.

I will now describe how I plan to implement PCA into my research project. The main reason it will be beneficial to use PCA is this project is that I will be dealing with data that has a very large number of dimensions. Each pixel has three dimensions and there are thousands of pixel per galaxy image. However, it is highly likely that only a few of these dimensions will be largely responsible for any variation in the data and hence PCA is an ideal tool here to make sense of this much data. An idea that would be worth exploring in this project is graphing the relative velocities of the various gasses along with the relative velocities of the stars in the galaxy. We could apply a PCA to this data and see which points tend to group together. This would then tell us if there is any correlation between stellar velocities and any given gas. We could then compare something like this across our three galaxy types and see which trends emerge from the data. We can also analyze the trend in these parameters amongst galaxies of the same type (eg. across all SFGs that are analyzed), and see if there is any trend in that PCA plot.

3 Expected Outcomes

The expected outcome of this project is to gain further insight into how galaxies are evolving as they leave their star forming phase and enter into the quiescent phase. More specifically, to try and better understand the physics at play that occurs as galaxies enter their GVG phase and as they come out of it. Questions that this project could provide further insight to are as follows. Why is this transition period between SFGs to QGs so short relative to the life span of the galaxies? Why does the transition take such a drastic turn when coming into the green valley? What physically is happening within the galaxy as the SFR starts to rapidly go down? How does a slowing SFR contribute to the morphological change we see in galaxies as they go from SFGs (usually spirals) to QGs (usually elliptical)? These and possibly other questions which may arise during the course of the project are what I hope to answer, or at least get a partial insight into. While it is well beyond the scope of this project to fully answer any of the aforementioned questions, results of this project can help us gain insight into them. This is because the PCA method that is to be used on the data during this project lends it self very well to finding the type of trends that can give us insights into these types of questions. This is due to the fact that PCA is excellent for finding trends amongst many different variables or parameters. I will explain this in more detail later on in this section of the proposal. This project then aims to both understand the morphological evolution of galaxies through studying its relative velocities, as well as trying to find velocity trends that can help us to categorize galaxies.

Now, I will discuss the scientific significance of the outcomes I expect from this project. Firstly, as I stated in the first section of this proposal galaxy evolution is an area of study where there is still much left to uncover. Understanding how galaxies transition from star forming to non-star forming is a huge piece of the evolutionary puzzle[3]. Studying the properties of GVGs can help us pin down the exact physical phenomena that occur during these rapid transition, which can yield tremendous insight into how galaxies evolve overall. This could possibly help us better understand why late type galaxies even form in the shapes they do to begin with. Furthermore, if we can better understand the internal kinematics of galaxies as they rapidly undergo a change we can perhaps get a better idea into the nature of their creation as this too must have had a similar rapid transition point[3]. With a clearer picture of kinematics within galaxies, and how this kinematic system undergoes rapid change, we can gain a better compression of the dark matter problem within galaxies. In essence then, this project potentially will not only help provide us with a better understanding of the green valley transition phase, but also give us unforeseen results that can help us answer other questions about galaxy formation and evolution. This is the advantage of a study that will focus on galaxy morphology as function of internal galaxy kinematics. It can allow us to potentially have both results with major impact on the study of galaxy morphology, as well as give us new insights into the internal kinematics of galaxies.

4 Timeline

1. **October 2020:** Finalize Marvin tutorials from documentation to have a complete grasps of the tools that can be used for statistical manipulation, and to see exactly what raw data is available. Replicate the right most rectangle of figure 1 above to have a control of which galaxies are going to be called SFG, QG, and GVG. Begin to familiarize myself with PCA algorithms that run using python code, and begin playing around and testing how it works with small data sets.

2. **November 2020:** Begin experimenting with MaNGA data in PCA algorithms. Plug in different variables and see how strong of a correlation can be observed from the PCA plots. Hone in one specific trend of interest amongst galaxies of the same type, and how trend that is clear when viewing the SFG to GVG to QG transition.

3. **December 2020:** Continue to hammer down on trends found in the data during November. Being asking critical questions about the trends, i.e why is trend occurring, what could possibly be the physical explanation etc.

4. **January 2021:** Have mid-term progress presentation ready by first week of January to present in mid-January. Include plots generated using PCA algorithms above. Talk about what this trends might physical mean, and how I will continue to search/verify this by the end of the project.

5. **February 2021:** Continue to explore the physical significance of the trends observed by the mid-term point of the project. Read literature to see if these trends have been observed before, or if they are new. If they have been found before see what areas are still open for exploration in them.

6. **March 2021:** Finalize the final report and presentation. Get all major conclusions of the project in order, and find good visual ways to show these conclusions with data.

References

[1] Mapping Nearby Galaxies at APO (MaNGA) https://www.sdss.org/surveys/manga/.

[2] Cherinka et al. Marvin: A Toolkit for Streamlined Access and Visualization of the SDSS-IV MaNGA Data Set. *The Astronomical Journal*, 158(2), 2019.

[3] Sebastian F. Sanchez. Spatially-Resolved Spectroscopic Properties of Low-Redshift Star-Forming Galaxies. *Annual Review of Astronomy and Astrophysics*, 58, 2020.