Using the Host Galaxy to Help Identify Supernova Type

David Moutard and David Cinabro

Department of Physics and Astronomy,
Wayne State University

Abstract

In this study, we use existing supernova (SN) samples in an effort to explore whether host galaxy properties can be used to determine SN type. We follow earlier work of Foley and collaborators[1], using galaxy magnitude, color, and Petrosian ratio (a stand-in for galaxy morphology) to classify type Ia SN (SNIa) and core-collapse SN (CCSN). The results show that at low redshift, the g-r color is the most useful classifier, whereas at higher redshifts, the star formation rate (SFR) is the most useful. Other properties tested proved to be either highly correlated with the first two, or they provided very little separation between classes. We provide a classifier based on a linear discriminant that combines the host properties, and test it on an independent sample of SN.

1 Introduction

SNIa have been used for many years by cosmologists as standardizable candles to determine distances and radial velocities of very distant galaxies in efforts to put greater constraints on the Hubble constant or cosmological constants. However, discerning a SNIa from a CCSN can be very time consuming, involving many observations over several nights. In 2013, Ryan J. Foley and collaborators [1], hereafter F13, attempted to correlate host galaxy properties with SN type. Similarly, in this paper, we compare the properties of SNIa hosts and CCSN hosts. The galaxy properties tested in F13 were morphology, B-K color index, K Magnitude, effective offset, and pixel rank, with the latter two being indicators of a more precise location of the SN within the galaxy. These last two proved to be mostly ineffective, and morphology was found to be the most useful in separating SNIa hosts from CCSN hosts.

Morphology, color, and magnitude are all correlated with SFR, which is a sensible indicator of supernvoa type. Since CCSN occur only in high mass, short lived stars, one expects to find them in galaxies with a higher SFR. If SFR is low, there is little material left to create these large, hot stars, and so one does not expect to see many CCSN. Instead, these low SFR galaxies exhibit old, long lasting stars, with many white dwarf (WD) stars. Since it is believed that SNIa occur when a WD accretes matter from a binary companion until reaching the Chandrasekhar limit, if a galaxy is expected to have many WD stars, then one also expects more SNIa.

Morphology (that is, whether a galaxy is spiral or elliptical) is correlated with SFR because elliptical galaxies have little material left with which to create new stars, so their SFR is low. Spiral galaxies, on the other hand, still have enough material to produce the large hot stars required to create CCSN (i.e. high SFR). In terms of color and magnitude, bluer galaxies tend to have a higher SFR, and the brightest galaxies tend to be large elliptical galaxies, so their SFR is expected to be lower. These three parameter are compared in this study as well, with some variation. Since the

Sloan Digital Sky Survey (SDSS) is used in this study, we used Sloan filters (ugriz). B-K color is replaced with g-r color, and instead of K magnitude, g magnitude is used, as it is the most efficient in the SDSS filter list. One of the most significant changes is replacing morphology, which was determined by eye in F13, with Petrosian ratio. Petrosian ratio is the ratio of the radius containing 50% of the Petrosian flux to the radius containing 90% of the Petrosian flux. This has been shown in the past as a consistent, quantitative replacement for morphology [2, 3]; elliptical galaxies tend to have a lower Petrosian ratio than spiral galaxies.

2 Methods

Here, we discuss the methods of analysis. Section 3.1 provides a general overview of the data reduction and sample selection methods, then sections 3.1 and 3.2 delve into the specific methods for a low z and a high z sample, respectively.

2.1 Sample Selection and Reduction

Using the SDSS-II SN Survey, which provides SN right ascension (RA), declination (DEC), and redshift (z), we generate a list of any galaxies within 0.01° of the SN, and within 0.08 of the SN z using SDSS Data Release 13 (DR13) SQL. This list represents any potential host galaxy for the SN within a fairly wide range. This large list of potential SN hosts is then compared again to the SN location, and whichever galaxy within 0.005° of the SN is closest in redshift to the SN is chosen as the host galaxy. Using data from these galaxies provided by SDSS, plots are made to show the separation between the SNIa and the CCSN host samples. This is initially done for a low z sample, then again for a larger, higher z sample. These two samples are analyzed using different methods, which will be discussed in the following sections.

2.2 Study at Redshift < 0.1

The first sample tested is a sample containing only spectroscopically confirmed SN at $z < 0.1$. This redshift limit is chosen because at redshifts past 0.1, efficiency starts to drop off for CCSN [4]. After data reduction, a sample of 42 SNIa and 42 CCSN remains. One CCSN host is cut from the g magnitude and color measurements due to a very high uncertainty in the g magnitude, leaving 41 CCSN hosts. Since SDSS provides apparent magnitudes m, absolute magnitude M is calculated using the following equation:

$$M = m - 5log(4.28 \times 10^8 pc \times z) + 5 \tag{1}$$

which is simply the distance modulus rewritten in terms of redshift. Once all data is in terms of absolute magnitudes, histograms are made to observe the separation between SNIa hosts and CCSN hosts.

Then, linear discriminant analysis (LDA) is used to provide a three dimensional vector that best connects the two samples, where each property observed (Petrosian ratio, magnitude, and color) represents one of the dimensions. In order to combine all three properties into one value, the projection of the vectors representing each galaxy in this three dimensional space onto the LDA vector is calculated. The size of the projection can be seen as a measure of how deeply into the CCSN host sample a galaxy lies. The vector extends from the SNIa host sample into the CCSN host sample, so a small projection would imply that a galaxy is leaning towards a SNIa host, where a larger projection would imply that a galaxy is tending toward CCSN host[5].

Finally, the significance of the separations between each property is found by first quantifying the separation between the samples as the following:

$$\frac{\bar{x}_{CC} - \bar{x}_{Ia}}{\sqrt{\sigma^2_{x,CC} + \sigma^2_{x,Ia}}} \tag{2}$$

where $\bar{x}_{type}$ is the average value of a given property x for either type of SN, and $\sigma_{x,type}$ is the standard deviation of the sample for either type. Then, a sample of 82,483 random galaxies is pulled from the same 2.5° by 120° patch of sky that the SDSS-II SN Survey covered. These galaxies (hereafter the field sample) are chosen to have roughly the same redshift distribution as the host galaxy sample, since many of the properties studied tend to vary with redshift. Two random subsamples of the same sizes as the SNIa host sample and the CCSN sample are pulled from the field sample and the separation is measured in the same way for these random galaxies. This is repeated 1000 times, which provides a roughly Gaussian distribution of separations. The separation from equation (2) is compared to this distribution of separations in order to find the statistical significance of the separation for SN host galaxies. The value of the statistical significance is reported in the results section.

The resulting populations for each of these properties are represented in Figure 1, as well as the combination in Figure 2. As one can see, the strongest separation is apparent in the Petrosian ratio and the g-r color. The significance of Petrosian ratio is roughly 2.7σ, magnitude shows a significance of only about 0.5σ, and color displays a significance of roughly 3.3σ. The combination, unsurprisingly shows the highest statistical significance, at about 4.3σ with a very strong visual separation.

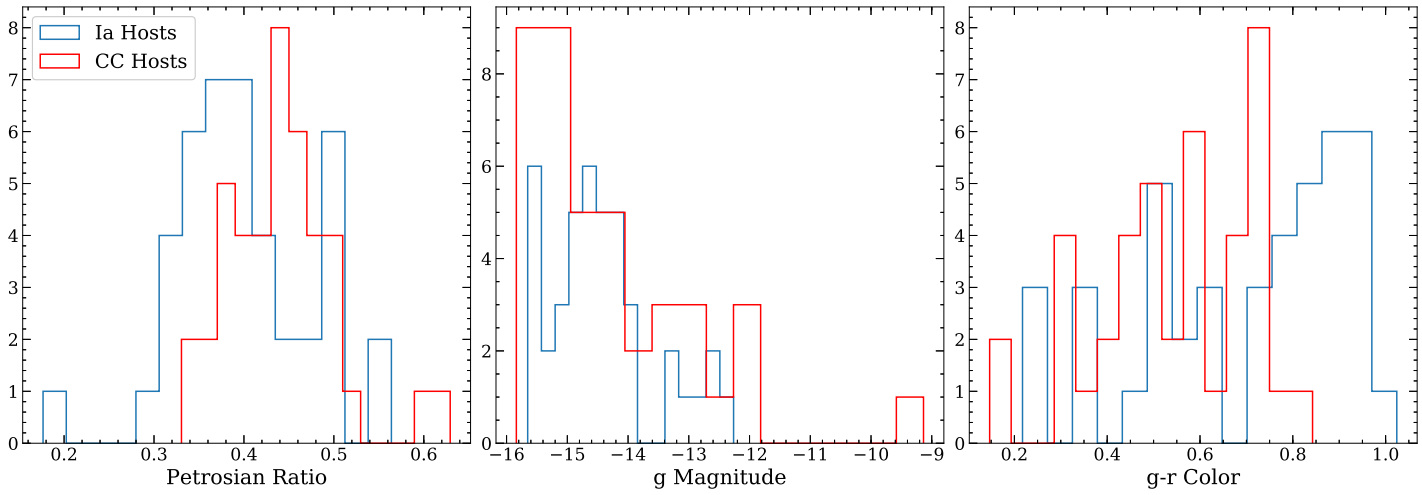


Figure 1: These plots represent the low redshift host galaxy samples. As one can see, the Petrosian ratio and the color are separated visually, and magnitude shows almost no separation.

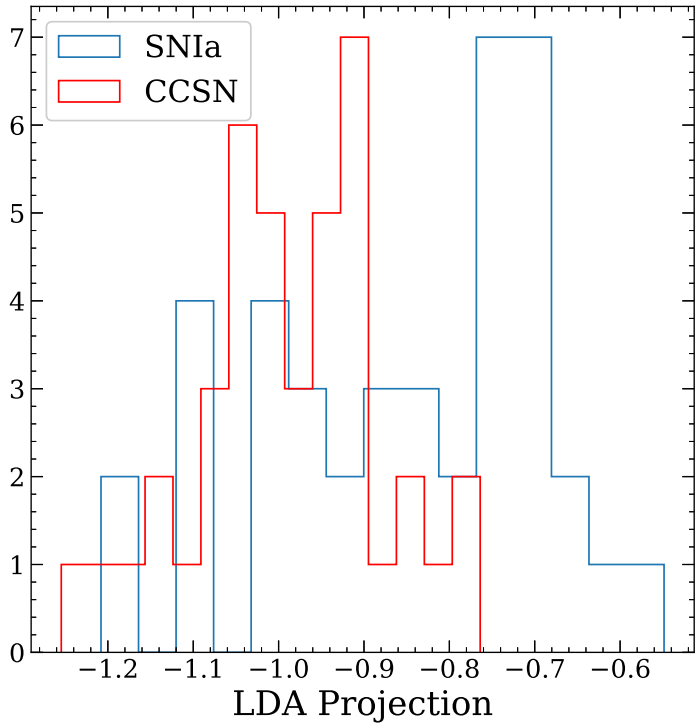


Figure 2: This lower redshift LDA sample shows a strong separation in populations.

2.3 Study at Redshift < 0.4

After analysis at low z is complete, we pull the entirety of the SDSS-II SN Survey sample. This includes both photometrically and spectroscopically confirmed SN out to $z = 0.4$. After data reduction, we are left with 97 CCSN hosts and 981 SNIa hosts. This extreme difference is to be expected, since SDSS is a brightness-limited survey and SNIa are on the order of 100 times brighter than CCSN. These samples could then be broken into subsamples based on whether they are photometrically or spectroscopically confirmed. The breakdown is shown in the table below. Since galaxy properties like color and magnitude change drastically with redshift, a program called

Table 1: High z SN Distribution by Type

SN Type	# SN	Confirmation Type	# SN
SNIa	981	Photometric	607
		Spectroscopic	374
CCSN	97	Photometric	35
		Spectroscopic	62

ZPEG, part of the PGASE family of galaxy evolution simulation code, is used to provide rest-frame galaxy information for this higher z sample. ZPEG works by fitting input information about a galaxy, such as redshift and ugriz filter magnitudes, with a set of galaxy templates.These galaxy templates are then evolved to a target redshift and compared with observed properties. The best matching template gives us the properties used in our analysis. This helps account for effects such as reddening. ZPEG has many potentially interesting outputs, such as SFR, WD mass, and galaxy age, as well as adjusted magnitudes. Many of these outputs prove to be not very interesting, and

some of the seemingly useful outputs often correlate strongly with another property being tested. One such example is WD Mass and SFR shown in Figure 3.

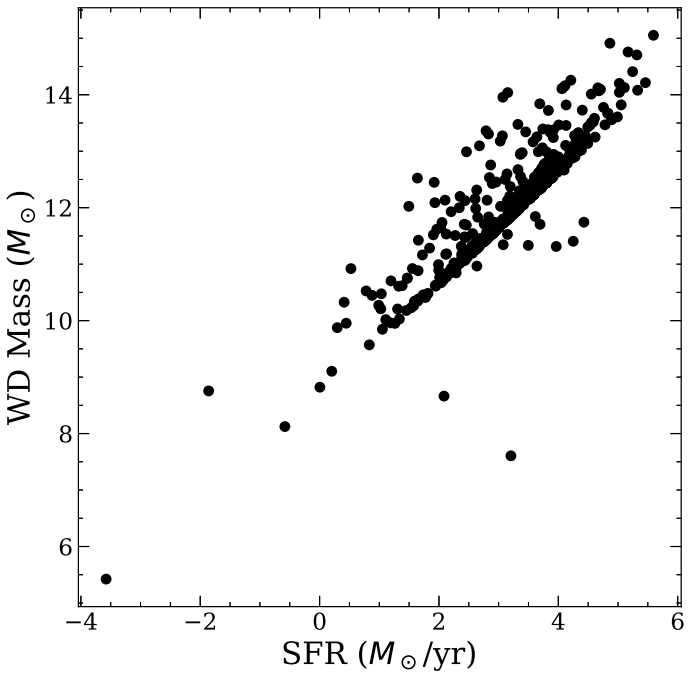


Figure 3: Here we observe the strongly linear correlation between WD Mass of a galaxy and its SFR. Because of this strong correlation, only SFR is used for our analysis.

Petrosian ratio proves to show almost no separation at this higher redshift (only roughly 0.47σ, shown in Figure 4), so it is replaced by SFR. At higher redshift, galaxies are fainter and show significantly less detail, often showing just the galactic core in the image. A brightness ratio fails in this case, since not enough of the galaxy is visible to the SDSS telescope. Figure 5 illustrates this well. Perhaps better telescopes could yield more detail, but for this study, the telescope used is not sufficient at these higher redshifts. The properties we explore with the high z sample are therefore SFR, color, and magnitude. LDA is again used to find the best vector connecting the spectroscopically confirmed SN host galaxies, and the same vector is tested on the photometrically confirmed sample, in an effort to show the effectiveness and generality of this LDA method of combination. A sample of 100,010 field galaxies is pulled with the same redshift distribution as the host samples, similar to the method discussed in section 2.2. Again, this sample is taken from the same patch of sky covered by the SDSS-II SN Survey. The significance is calculated in exactly the same way as discussed in section 2.2, and is reported in Table 2.

Figure 6 represents the populations studied at a higher redshift, for both the photometrically and spectroscopically confirmed subsamples, as well as the sample in it's entirety. SFR, as one would expect shows the highest separation, with an exception in the photometrically confirmed sample, which shows little separation at all. Table 2 shows the statistical significance for the separation all of three high z samples. Interestingly, whiel statistical significance is on average lower for the photometric subsample, the LDA combination does yield a slightly higher statistical significance. It is also unexpected that for the high z sample, the LDA combination has a lower statistical significance than some of the individual properties.

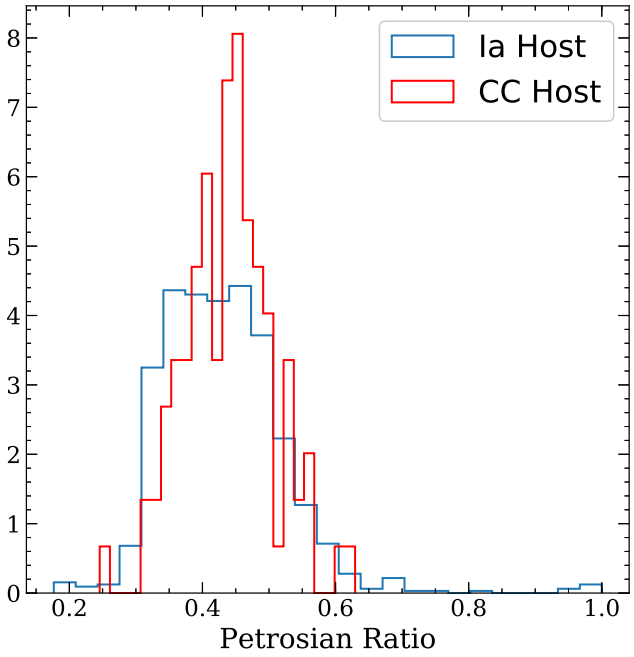
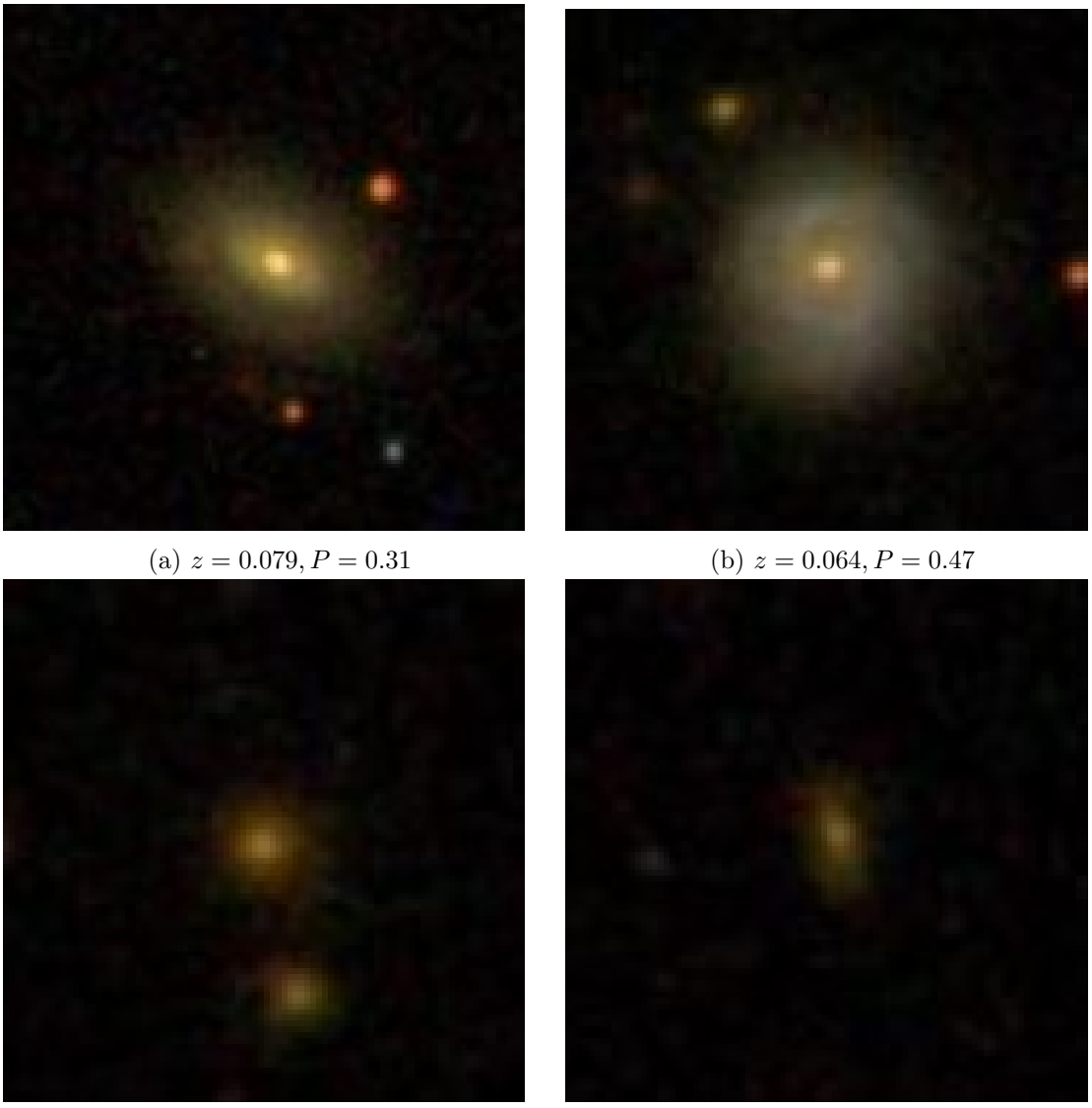


Figure 4: One can see that in the higher redshift sample, the readily apparent separation seen in the low z sample is no longer present.

Table 2: The statistical significances for each of the high z subsamples, broken down by property. Unsurprisingly, the highest statistical significance occurred when observing the sample in its entirety

Sample	Confirmation Type	# σ
Spectroscopic	SFR	2.9
	g Magnitude	3.4
	g-r Color	1.2
	LDA Combination	2.4
Photometric	SFR	1.6
	g Magnitude	1.5
	g-r Color	0.1
	LDA Combination	2.8
Entire Sample	SFR	4.1
	g Magnitude	4.6
	g-r Color	1.0
	LDA Combination	2.9

Early attempts are also made at determining correlations between SN Host galaxies and the random field galaxy sample. The concept is that if CC host galaxies had a property with a mean value of a, and that same property in Ia hosts shows a mean value of b, then the mean for that property in a large sample of random field galaxies should be somewhere in between a and b. Unfortunately, these results prove to be inconclusive. This is possibly because, although the redshift range of the field sample is chosen to match that of the SN sample, other properties are not made to share a distribution. SDSS possibly returns the brightest galaxies first for the field sample, which might have different properties than those that hosted SN. In Section 4, we discuss this further.



(a) $z = 0.079, P = 0.31$

(b) $z = 0.064, P = 0.47$

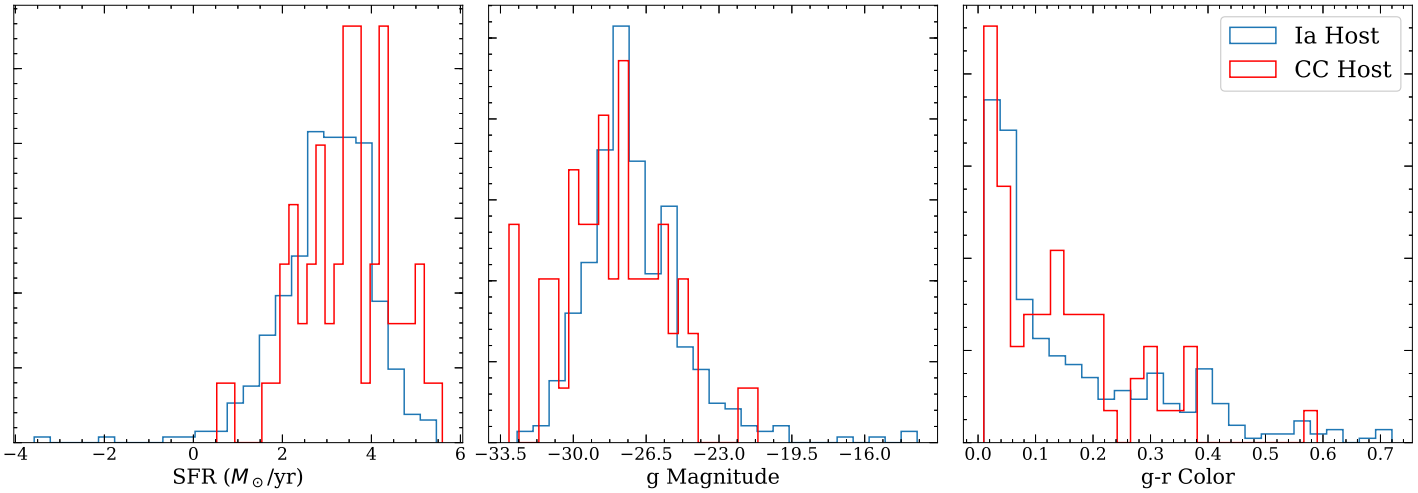
(c) $z = 0.299, P = 0.35$

(d) $z = 0.318, P = 0.46$

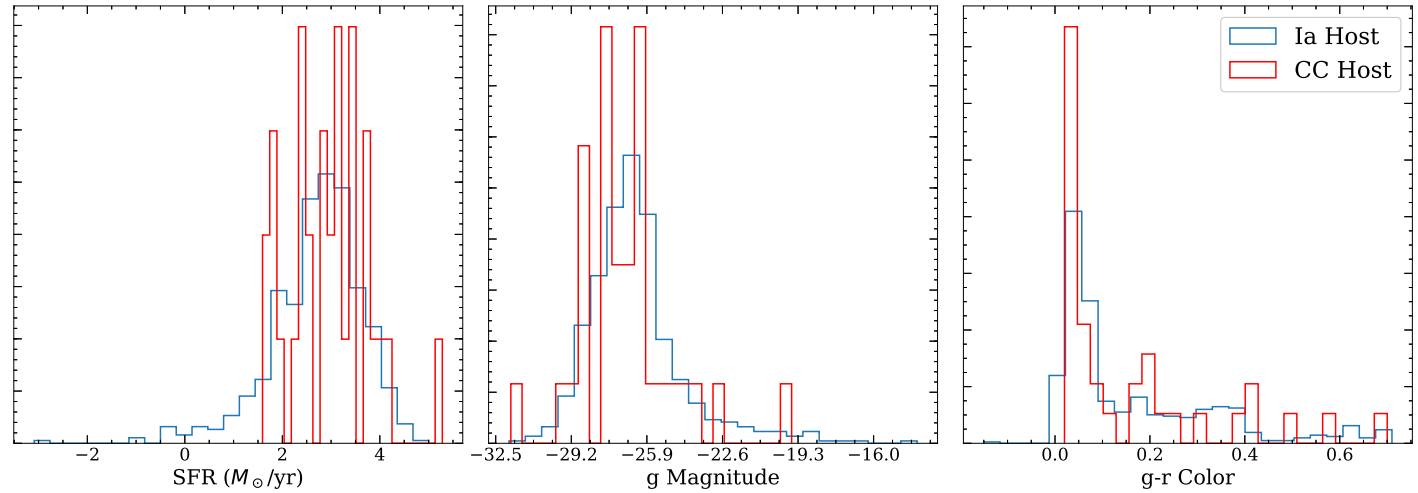
Figure 5: Here, P represents the Petrosian ratio. One can see that at higher redshifts, galaxies lose detail, even when compared to low z galaxies with a similar Petrosian ratio. This image also captures reddening, which is the primary reason for using ZPEG. These galaxies were found using the SDSS Explore feature.

After LDA combination, the separations for the high z sample (both photometrically confirmed, spectroscopically confirmed, and combined) are shown in Figure 7. These do not show the same strong separation seen in the low z sample, but there is a slight difference between the samples, where the CCSN host sample appears to be shifted to the right.

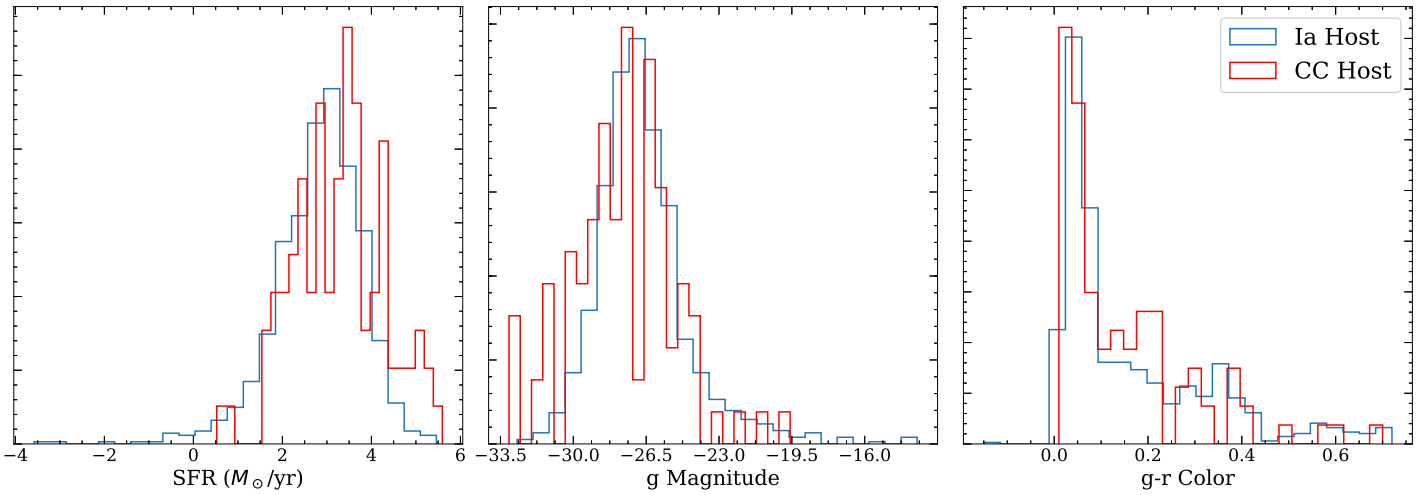
The program used to provide a linear discriminant also provides second and third order vectors that represent the second and third best connections between populations. These are tested to see if higher order LDA vectors can be useful, but the results showed no separation, and are therefore ignored for the rest of the study. These second and third order LDA histograms can be seen in Figure 8.



(a) Spectroscopic subsample

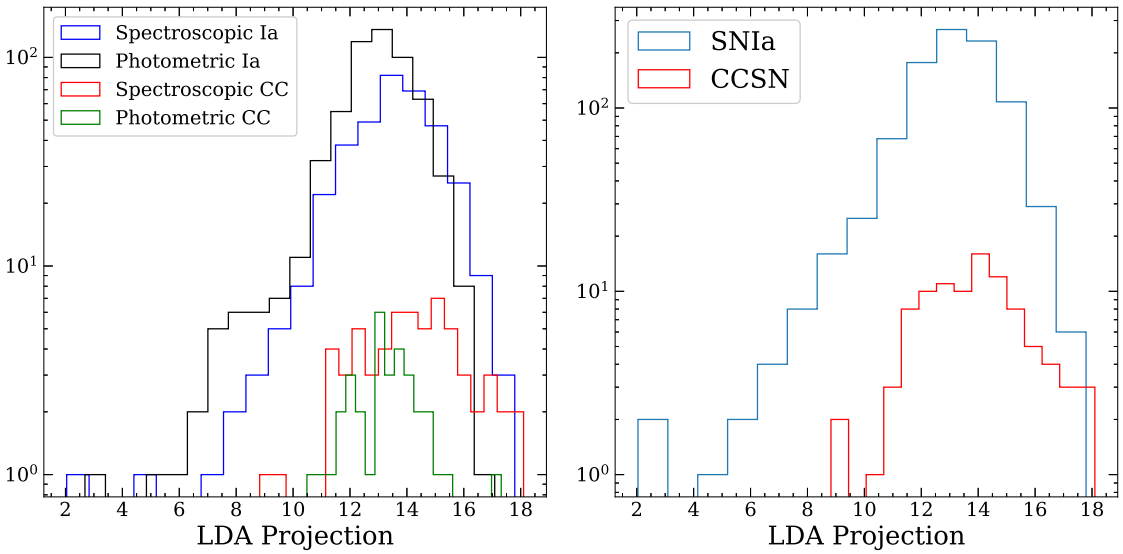


(b) Photometric subsample



(c) Complete high z sample

Figure 6: These plots represent the higher redshift host galaxy samples. SFR shows the strongest separation in (3a) and (3c), while any other parameters seem largely unhelpful. These plots have been normalized to 1 account for the large difference in sample sizes.



(a) High z, separated populations

(b) High z, populations combined

Figure 7: The y-axis is changed to a log scale in order to account for the difference in sample sizes. Normalization to 1 is not chosen in this case, since parsing the information is much more difficult, especially with 7(a)

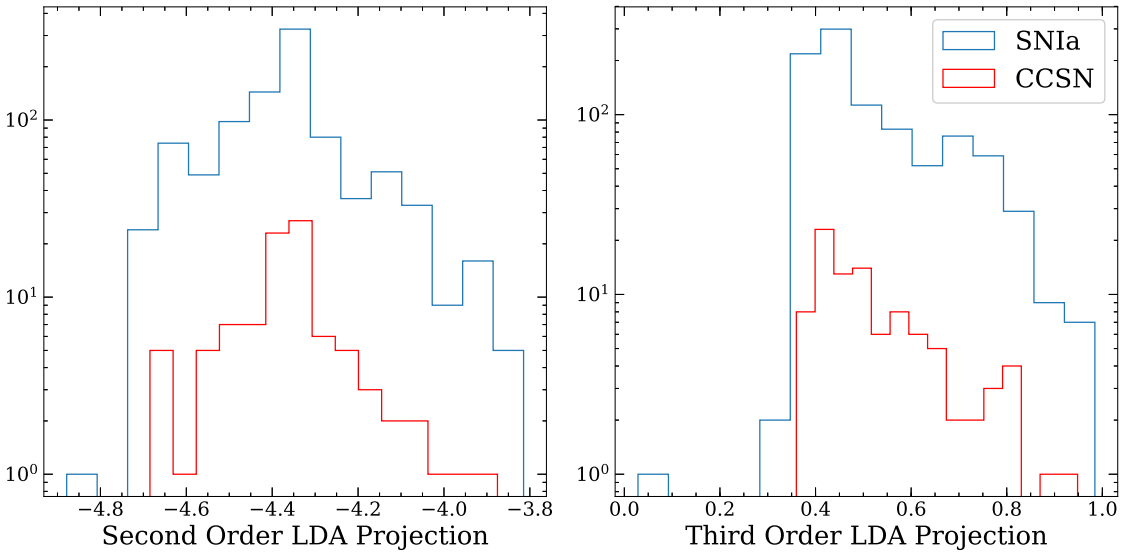


Figure 8: These plots represent the projections onto the second and third order LDA vectors for the combined high z sample

3 Results

For the high z sample, the linear discriminant is trained using the spectroscopic sample, and the provided vector is used on the photometric sample. We believe that the similarity in statistical significance between the two subsamples is evidence that observing projections onto a linear discriminant vector is at least somewhat effective as a means of combining galaxy properties.

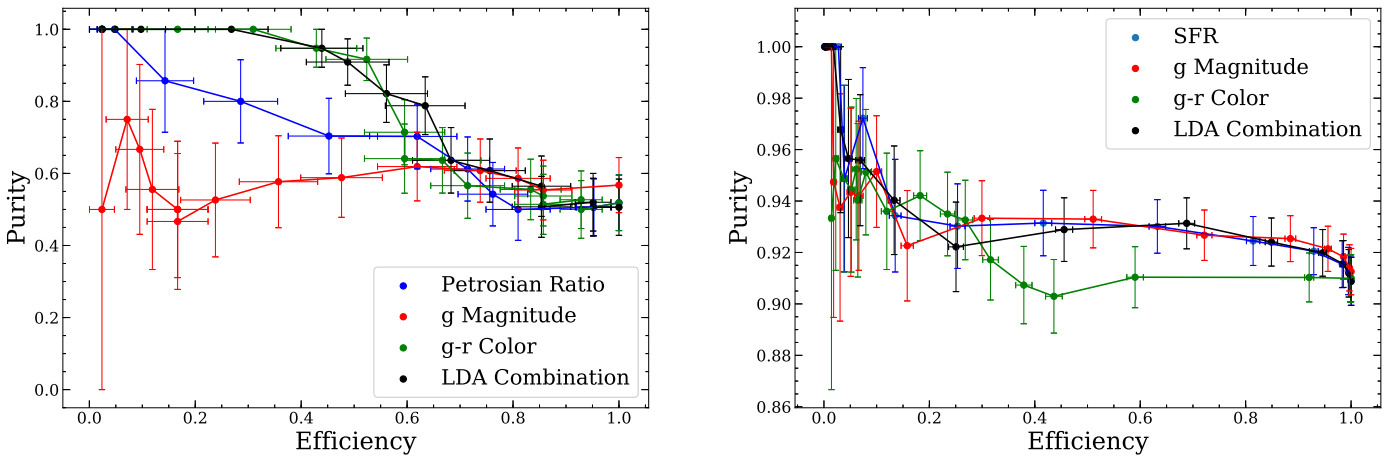
While F13 argued that brighter galaxies are more likely to host Ia since they are more likely to be large ellipticals, our samples show the opposite to be true. This could be caused by the difference in filters used to measure magnitude. F13 utilized the absolute K Magnitude, while we use the g magnitude. If a galaxy has many hot, blue stars that could result in CCSN, it will register as brighter in the g band. Once all histograms are plotted, efficiency is plotted against the purity for the Ia host samples of each parameter. Efficiency is calculated as:

$$\frac{N_{Ia,Sub}}{N_{Ia}} \tag{3}$$

where $N_{Ia,Sub}$ is the number of SNIa within a particular cut and N_{Ia} is the total number of SNIa in a sample. The purity is calculated using:

$$\frac{N_{Ia,Sub}}{N_{Ia,Sub} + N_{CC,Sub}} \tag{4}$$

where $N_{CC,Sub}$ is the number of CCSN that are within the same cut made for the SNIa sample. The direction of the cuts is made such that the first cut would only contain SNIa, making the first point 100% pure. These plots prove to be quite similar for the high z spectroscopic and photometric samples, so the efficiency versus purity is shown for the complete high z sample and the complete low z sample. The resulting plots can be seen in 9.



(a) Low z sample

(b) High z sample

Figure 9: In the low z sample, Using these properties can yield a moderately pure sample to an efficeiency of about 0.5 before dropping off. The high z sample, however, drops off drastically, and to have any sort of pure sample, efficiency must be very low.

4 Conclusion and Future Work

As one can see, the separation between SNIa host populations and CCSN host populations, while statistically significant, is not especially large. At low z, we can see that the use of galaxy properties can be a somewhat effective way to generate a moderately pure, moderately efficient sample of SNIa. However, at even slightly higher z, the separation becomes quite small, until no meaningful separation between populations remains. There is a possibility that with further research, host galaxy properties may be used in conjunction with other methods to confirm SN type, but it seems unlikely that this will ever be an efficient method to determine with any strong probability whether a SN is type Ia.

At the time of writing this report, work is continuing on this project in an effort to improve the results. Currently, we are reducing a larger data set (roughly 4000 SN) from the Unified Supernova Catalog (USC) [6] in order to test our methods on a larger data set. Also, ZPEG default parameters are used for this study, so future work may involve adjusting the galaxy evolution parameters and observing how the results differ. Other color indices may be used, such as g-i, to closer replicate those used by F13. As mentioned in Section 2, we attempt to measure field galaxy correlation as well. Since this did not return useful results, we will attempt to try again, this time matching both redshift and magnitude distributions.

This study lacks robustness in some areas. One such area is the localization of SN. In a recent paper [7], it was concluded that local colors are more representative of SN type than host galaxy colors as a whole; that is, the color in the immediate area of the SN is often more useful than using the color of the entire galaxy. While this is not entirely surprising, it is also not especially practical when considering the size and limitations of a survey such as SDSS. SDSS contains data for thousands of galaxies, many of which are very small in angular size, so one pixel might represent a few light years of actual space, making localization difficult if not impossible. Also, as mentioned in Section 1, F13 attempted to measure local SN "neighborhood" properties, the results of which were largely inconclusive. This is likely due to the previously mentioned reasons.

It seems that as Petrosian ratio becomes less and less effective at separating populations, this method starts to fail. At low z, when Petrosian ratio effectively describes different galaxy morphology, one can develop a moderately efficient and quite pure sample of SNIa. However, as z increases, and the Petrosian ratio becomes less and less indicative, the only pure sample one can create is also severely inefficient.

References

[1] Ryan J. Foley and Kaisey Mandel. Classifying Supernovae Using Only Galaxy Data. *The Astrophysical Journal*, 778(2):167, 2013.

[2] Iskra Strateva et al. Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data. *The Astronomical Journal*, 122(4):1861, 2001.

[3] Kazuhiro Shimasaku et al. Statistical Properties of Bright Galaxies in the Sloan Digital Sky Survey Photometric System. *The Astronomical Journal*, 122(3):1238, 2001.

[4] Taylor, M. and others. The Core Collapse Supernova Rate from the SDSS-II Supernova Survey. *The Astrophysical Journal*, 792:135, sep 2014.

[5] Wikipedia. Linear Discriminant Analysis. `https://en.wikipedia.org/w/index.php?title=Linear_discriminant_analysis&oldid=782803282`, 2017.

[6] D. Lennarz et al. A Unified Supernova Catalogue. *A&A*, 538:A120, 2012.

[7] M. Roman et al. The Dependence of Type Ia Supernova Luminosities on Their Local Environment. *ArXiv e-prints*, 1706.07697, June 2017.