ESLAB #53: The *Gaia* Universe
Unedited (Zenodo community curated by J.H.J. de Bruijne)

Towards a Complete Census of CSPNe in Gaia

Nicholas Chornay, Nicholas A. Walton

Institute of Astronomy, University of Cambridge

Abstract

Planetary Nebulae (PNe) are a brief stage through which low and intermediate mass stars pass towards the end of their evolution, between red giant and white dwarf. They are excellent probes of the chemo-kinematical structure of the Milky Way. Accurate distances to PNe are necessary to make any sort of meaningful astrophysical characterisation, from their nebula size and lifetimes to their central star luminosities. The Gaia mission provides an unprecedented opportunity to better understand these objects through measurements of the trigonometric parallaxes of their central stars. Taking full advantage of these measurements requires accurate cross matching of PN catalogues and Gaia sources. As the size of both catalogues increases, it becomes impractical and inadvisable to do this manually. We present an automated method based on the classic likelihood ratio technique that bootstraps from available data and considers both the relative positions and Gaia colours of central star candidates. Applying this method to the HASH PN Catalogue and to Gaia DR2, we find over 1000 likely central star matches, twice as many as in previous works, while also rejecting significant numbers of potential contaminants. The subsample of these matches with the most accurate parallaxes will allow us to make comparisons to distance scales in literature as well as to simultaneously trace central star and nebular evolution.

1 Introduction

Planetary Nebulae (PNe) are one of the end stages of life for low to intermediate mass stars, in which they shed their outer layers, growing brighter and hotter before cooling into white dwarfs. The rapidly expanding shell of gas is ionised by the central star, reaching a typical size of a few lightyears over the lifespan of the nebula.

PNe are important in galactic evolution for their chemical enrichment of the ISM, and their visibility over great distances makes them valuable chemical probes of not only the Milky Way but also nearby galaxies. In addition, the Planetary Nebula Luminosity Function (PNLF) provides a useful rung in the cosmic distance ladder.

Only around 3000 galactic PNe are currently known. While their numbers are small relative to the stellar population due to their relatively short lifespans (on the order of tens of thousands of years), stellar evolution models predict many more than have been discovered (cf. Frew (2016)).

The understanding of PNe is limited by their poorly constrained distances (cf. Smith (2015)). Direct distances are known to very few PNe, and fractional uncertainties are high, with the most relatively accurate distances available being to PNe located in galaxies external to our own. Distance measurement difficulties are aggravated by the rapid evolution of the central star of the PN (CSPN), which prevents the application of usual methods of distance determination such as isochrome fitting. Meaningful astrophysical characterisation of these objects, however, requires accurate distances, which has motivated many different avenues of research towards determining more accurate distance estimates.

There are two broad categories of distance measurement techniques. Primary techniques measure distances to individual PNe, either directly (through measuring the trigonometric parallax of the central star) or based on modelling: of the nebula expansion velocities (cf. Schönberner *et al.* (2018)), of the environment (e.g. extinction distances, cluster or bulge membership, or location in external galaxies), or of the CSPN.

Secondary, or statistical distance scales, rely on deriving a relationship that holds broadly for PNe and that provides a means of estimating, for example, the physical size of the PN, which can be then compared to the angular size and used to derive the distance, in a manner analogous to the distance modulus for stars whose absolute luminosity is known. These secondary methods require a calibrating set of objects whose distances are known through other means. The most recent statistical method is the Hα surface brightness to physical radius relation in Frew *et al.* (2016), which used a calibrating set of around 300 galactic and extragalactic PNe.

Secondary methods are only as good as the quality and purity of their calibrating distances. Incorrect distances or polluting objects inflate their errors well beyond the uncertainties stemming from measurement errors and intrinsic scatter. Thus, improving primary distances to a set of PNe provides a twofold benefit, as it betters not only the distances to that set but also, through improved calibration of statistical distance scales, to the population as a whole.

Trigonometric parallaxes can offer the most accurate, assumption-free primary distances. Before the Gaia mission (Gaia Collaboration *et al.* (2016)), the set of parallaxes was limited to less than 15 PNe with parallaxes from USNO (Harris *et al.* (2007)) and the HST (Benedict *et al.* (2009)). A small number of parallaxes were found in the first Gaia data release by Stanghellini *et al.* (2017), but now with Gaia DR2 (Gaia Collaboration *et al.* (2018)) we expect a huge increase. Some work has been done previously by Kimeswenger & Barría (2018) on Gaia DR2 parallaxes, but with an incomplete input catalogue, and a manual search process.

In the remainder of this paper, we present an automated method for finding CSPNe in Gaia DR2, with the aim of gathering more parallax measurements that can be used to better constrain the distances to galactic PNe.

2 Methods

We use as our input catalogue the HASH PN catalogue[1] from Parker *et al.* (2016). This is the most complete catalogue of galactic PNe, containing over 2500 spectroscopically confirmed PNe (cf. Frew & Parker (2010)) as well as possible and likely PNe, and a range of other objects that are prone to being confused with PNe. Positional accuracy has been an issue in previous PN catalogues (cf. comments in Parker *et al.* (2006)); HASH PN represents an attempt to enforce consistent positional measurements, drawing primarily on the centroiding of the nebula itself in narrowband Hα imagery.

In Gaia DR2 we search for sources near the centres of these nebulae. It is important to note that HASH PN is a catalogue of nebulae rather than of central stars, and furthermore that because of the varied morphologies and the possibility of the true extent of the nebula not being visible, central stars may not always be the most central. We need not only rely on position of candidate stars, however, as we can also consider their colours: the source of the UV radiation that ionises the nebula will be a hot, blue star. The observed star may not appear blue, though, due to reddening in the line of sight to the PN or internal to the nebula itself, or due to the visible light being dominated by a main sequence binary companion.

Not all CSPNe are visible and/or pass Gaia's detection threshold. The problem is thus to weigh the evidence for a Gaia source being the true central star against the possibility of it being a spurious contaminating object. Previous methods for central star matching have been largely manual, relying on a subjective assessment of position as well as assumptions about the colour of the star being searched for. The manual approach is labour intensive and prone to inaccuracy and inconsistency, and, as the sizes and update frequencies of data sets increase, becomes increasingly impractical. Moreover it fails to capture quantitative uncertainty in matching.

2.1 Catalogue Matching

This more general problem of disentangling chance from genuine coincidences is by no means a new one. It arose in the past in the context of matching catalogues from surveys at different wavelengths, for example finding optical counterparts to radio sources. A popular solution to this problem is the likelihood ratio method (Sutherland & Saunders (1992)), and it still finds use today, with a similar quantity appearing for example in the figure of merit used to assess match quality between Gaia and external catalogues (Marrese *et al.* (2017)).

Given a candidate counterpart for a source in the input catalogue, the idea of the method is to compare two competing hypotheses: the candidate being a genuine match versus merely coincidental. In the simplest version, if the positions in the catalogues have an angular separation r, the likelihood ratio is the ratio of the probability of finding true counterparts separated by r to the probability of finding a chance object with that separation. That is, it is a ratio of two densities. In the case of Gaussian positional uncertainties and a constant density of background objects ρ, the distribution of separations between true counterparts follows a Rayleigh distribution, and the density of spurious objects simply increases linearly with r. This gives a likelihood ratio

$$L = \frac{\mathrm{Rayleigh}(r; \sigma)}{2\pi r \rho} \tag{1}$$

The notion of a likelihood ratio can be extended to densities in spaces other than separation - for example, magnitude or colour. It can also take into account the possibility of a non-detection. For example, given a candidate with colour c, and an assumption that the colour and separation are independent, the likelihood ratio becomes

$$L = Q \frac{P(c|\mathrm{genuine})}{P(c|\mathrm{chance})} \frac{P(r|\mathrm{genuine})}{P(r|\mathrm{chance})} \tag{2}$$

where Q is the identification rate - the prior probability of finding a match.

2.2 Likelihood Ratio Method for CSPNe

We treat the problem of finding CSPNe as a catalogue matching problem and apply the likelihood ratio method. The colour c is the BP-RP colour for a source in Gaia DR2 (Evans *et al.* (2018)), and the separation r is the angular separation between that source and the centre of the nebula from HASH PN. Q is the probability of the CSPN being detected by Gaia. The challenges are the unknown colour distribution of true central stars and the lack of positional uncertainties for PNe.

However, we can infer these distributions from the data itself. The idea is to consider separately the colours and the separations, and iteratively use the objects that are high confidence matches from one representation as "training" objects for the other representation. In particular if we can find some Gaia sources that are highly likely matches due to their colours, we can use the separation between these objects and the PN positions as an estimate of the positional uncertainties. The ideas behind this approach have appeared before in the context of semi-supervised learning, with a technique known as co-training (Blum & Mitchell (1998)).

Most central stars will be the nearest to the PN centre, so an initial estimate of the colour distribution of true central stars can be derived from colours of the nearest neighbouring Gaia source to each PN. These can be compared to the colours of other nearby Gaia sources (we consider all sources within one arcminute) to derive a density ratio, as shown in Fig. 1.

Switching representations, we then take the set of PNe that have nearby Gaia sources whose colours lie in a region of colour space where the colour density ratio is high (as expected, these are the ones with very blue colours), and consider the angular separations between those sources and the centres of their PNe (Fig. 2). The long tailed distribution of these separations does not fit well with a single Rayleigh distribution, so a Rayleigh mixture model is used instead. The mixture can also be re-weighted to better reflect the positional uncertainties for different sizes of nebula.

The density of background sources, also required for the separation density ratio, is estimated locally for each PN, taking the count of Gaia sources found within one arcminute. It is the same for all candidates for a given PN, so it does not affect the ranking, only the confidence.

Finally, we use the estimated positional uncertainties to

[1] https://hashpn.space

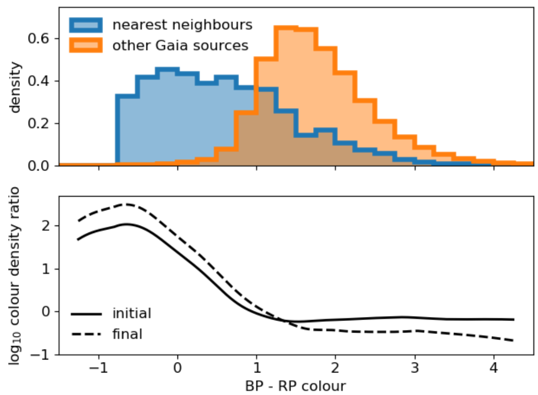

Figure 1: The densities in BP-RP colour space of Gaia sources near PN centres, and the corresponding colour density ratio. The initial colour density ratio (plotted on a log scale) is derived from the comparison between the colours of nearest and non-nearest neighbours, while the final one considers estimated positional uncertainties, as described in the text. The effect of the update is to push the density ratio away from 1, in particular penalising redder sources.

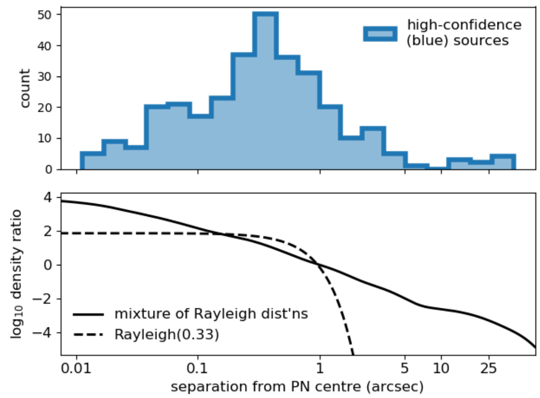

Figure 2: Histogram of separations of blue Gaia sources from PN centres, and the derived distance density ratio. The distance density ratio depends on the radius of the PN and the density of background objects, which here are taken to be 100 arcseconds and 0.02 objects per arcsecond^2 respectively. For comparison, the distance density ratio for a single Rayleigh distribution is also shown, with a parameter equal to the median separation of the blue sources. It fails to capture both the peak and the long tail of the distribution, motivating the use of a mixture model.

refine the estimate of the colour density ratio. Instead of splitting sources by nearest versus non nearest neighbour, we consider and compare the colours of objects whose status is as CSPNe or non-CSPNe is relatively certain from their position alone. This removes many of the contaminants from the initial estimate and effectively increases the contrast in the colour density ratio.

The likelihood ratio by itself only considers a single candidate match at a time, and its values are unbounded above. A somewhat more useful confidence measure is the reliability, also introduced by Sutherland & Saunders (1992):

$$R_i = \frac{L_i}{\sum_j L_j + (1 - Q)} \tag{3}$$

Reliability can be thought of as the probability of an individual candidate being a genuine match. It takes into account all candidate matches as well as the the possibility of a non-detection. We adopt reliability as our score measure, and use thresholds on the reliability to divide our results into likely, possible, and non-matches.

3 Results

We run the matching algorithm over the full HASH PN Catalogue (as of 10 October 2018), including true, likely, and possible PNe, but we focus our results on the subset of nearly 2500 true PNe (as characterised by Parker *et al.* (2006)). The resulting reliabilities follow a bimodal distribution, an encouraging sign that the algorithm commits in most cases to either a single match, or to there being no good matches at all. The average sum of reliabilities per PN is 0.53, agreeing well with our choice of $Q = 0.5$.

As a sanity check and to better assign meaning to the reliabilities, we compare the results from our algorithm to those from best effort visual inspection from a variety of imaging surveys on a subset of 100 PNe. The test set contains 41 good matches, 44 non-matches, and 15 potential matches. All of the non-matches have reliability less than 0.2, while all but 3 of the good matches have reliabilities greater than 0.8. Thus we choose 0.8 as our threshold for likely matches, and 0.2 as our threshold for possible matches.

Applying these thresholds to the full set of candidate matches, we find over 1000 likely matches and over 1000 non-matches, with the remaining best candidates having uncertain match status. The highest success rates are for more extended PNe away from the galactic centre and away from the disc, where the PNe tend to be nearer, the density of background objects lower, and the stars less reddened by dust.

We can compare our matching results to those from Kimeswenger & Barría (2018), a previous work on central stars of PNe in Gaia DR2. This work relied on visual searches, and used a smaller catalogue of 728 PNe with radio distances from Stanghellini & Haywood (2010). Their paper reported 382 matches, all of which are in the HASH PN input catalogue that we used, and 361 of which are in the confirmed PNe subsample on which we focused. We find good agreement with about 90% of the matched sources, with the remaining sources being split evenly between cases where our method found no match and cases where it found a better candidate (Fig. 3). Comparisons with the imaging data suggest that that the consideration of both colour and position in our method, combined with an input catalogue with more accurate positions, generally performs better. It is worthwhile to note that these differences do not substantially change the results of their distance comparisons, because of restrictions on colour and the outlier cuts that they used in the regressions.

4 Conclusions

We have created a catalogue of over 1000 likely Gaia DR2 detections of central stars of Planetary Nebulae, doubling the counts in previous work. We have done so with an automated procedure based on classical likelihood ratio methods that take minutes to run, which will allow us to easily and quickly

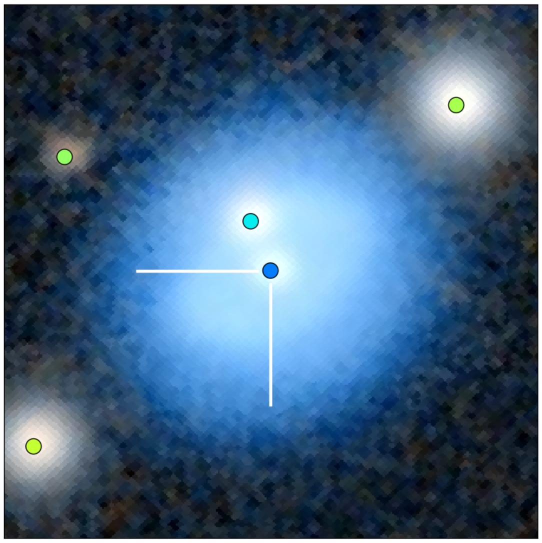

Figure 3: PNG 226.4-03.7, an example of a disagreement with the matching of Kimeswenger & Barría (2018). The white lines have length 5 arcseconds and indicate the position of the centre of the PN. Coloured dots indicate Gaia sources, with bluer dots corresponding to bluer BP-RP colours. Our method selects the blue Gaia source closest to the centre, while the match from the previous work is the next nearest, which is unlikely to be the true CSPNe from the geometry. The background image is taken from Pan-STARRS.

update the sample with further Gaia data releases and new discoveries. Chornay & Walton (2019) provides the complete catalogue of matched sources and describes the use of these to improve the calibration of statistical distance scales.

Looking forward to future Gaia data releases, we expect that Gaia EDR3 will have more detections, more colours that can be used to improve matching, and lower parallax uncertainties. In Gaia DR3, we look forward to the low resolution BP-RP spectra, which promise to be helpful in the search for new PNe.

Acknowledgments

This research has made use of data from the European Space Agency (ESA) mission *Gaia*[2], processed by the *Gaia* Data Processing and Analysis Consortium (DPAC)[3]. Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This research has also made use of the HASH PN database[4], of Astropy[5], a community-developed core Python package for Astronomy (Astropy Collaboration *et al.*, 2013; Price-Whelan *et al.*, 2018), and of "Aladin sky atlas" developed at CDS, Strasbourg Observatory, France.

This research was supported through the Cancer Research UK grant A24042.

[2] https://www.cosmos.esa.int/gaia

[3] https://www.cosmos.esa.int/web/gaia/dpac/consortium

[4] http://hashpn.space

[5] http://www.astropy.org

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., *et al.* 2013, A&A, 558, A33.

Benedict, G. F., McArthur, B. E., Napiwotzki, R., Harrison, T. E., Harris, H. C., *et al.* 2009, AJ, 138, 1969.

Blum, A. & Mitchell, T. 1998, In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pp. 92–100 (New York, NY, USA: ACM). ISBN 1-58113-057-0.

Chornay, N. & Walton, N. A. 2019. In prep.

Evans, D. W., Riello, M., De Angeli, F., Carrasco, J. M., Montegriffo, P., *et al.* 2018, A&A, 616, A4.

Frew, D. & Parker, Q. 2010, Publications of the Astronomical Society of Australia, 27, 129.

Frew, D. J. 2016, Proceedings of the International Astronomical Union, 12, 11âĂŞ19.

Frew, D. J., Parker, Q. A., & Bojičić, I. S. 2016, MNRAS, 455, 1459.

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., *et al.* 2018, A&A, 616, A1.

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., *et al.* 2016, A&A, 595, A1.

Harris, H. C., Dahn, C. C., Canzian, B., Guetter, H. H., Leggett, S. K., *et al.* 2007, AJ, 133, 631.

Kimeswenger, S. & Barría, D. 2018, A&A, 616, L2.

Marrese, P. M., Marinoni, S., Fabrizio, M., & Giuffrida, G. 2017, A&A, 607, A105.

Parker, Q. A., Acker, A., Frew, D. J., Hartley, M., Peyaud, A. E. J., *et al.* 2006, Monthly Notices of the Royal Astronomical Society, 373, 79. ISSN 0035-8711.

Parker, Q. A., Bojičić, I. S., & Frew, D. J. 2016, In *Journal of Physics Conference Series*, *Journal of Physics Conference Series*, vol. 728, p. 032008.

Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., *et al.* 2018, AJ, 156, 123.

Schönberner, D., Balick, B., & Jacob, R. 2018, A&A, 609, A126.

Smith, H. 2015, MNRAS, 449, 2980.

Stanghellini, L., Bucciarelli, B., Lattanzi, M. G., & Morbidelli, R. 2017, New Astronomy, 57, 6.

Stanghellini, L. & Haywood, M. 2010, ApJ, 714, 1096.

Sutherland, W. & Saunders, W. 1992, MNRAS, 259, 413.