Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **515,** 3883–3897 (2022) https://doi.org/10.1093/mnras/stac1997

Visualizing the pulsar population using graph theory

C. R. García,[1,2]★ Diego F. Torres[1,2,3]★ and Alessandro Patruno[1,2]★

[1] *Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans s/n, E-08193 Barcelona, Spain*
[2] *Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[3] *Institució Catalana de Recerca i Estudis Avançats (ICREA), E-08010 Barcelona, Spain*

Accepted 2022 July 8. Received 2022 July 7; in original form 2022 June 8

ABSTRACT

The $P\dot{P}$ diagram is a cornerstone of pulsar research. It is used in multiple ways for classifying the population, understanding evolutionary tracks, identifying issues in our theoretical reach, and more. However, we have been looking at the same plot for more than five decades. A fresh appraisal may be healthy. Is the $P\dot{P}$-diagram the most useful or complete way to visualize the pulsars we know? Here we pose a fresh look at the information we have on the pulsar population. First, we use principal component analysis over magnitudes depending on the intrinsic pulsar's timing properties (proxies to relevant physical pulsar features), to analyse whether the information contained by the pulsar's period and period derivative is enough to describe the variety of the pulsar population. Even when the variables of interest depend on P and $\dot{P}$, we show that $P\dot{P}$ are not principal components. Thus, any distance ranking or visualization based only on P and $\dot{P}$ is potentially misleading. Next, we define and compute a properly normalized distance to measure pulsar nearness, calculate the minimum spanning tree of the population, and discuss possible applications. The pulsar tree hosts information about pulsar similarities that go beyond P and $\dot{P}$, and are thus naturally difficult to read from the $P\dot{P}$-diagram. We use this work to introduce the pulsar tree website containing visualization tools and data to allow users to gather information in terms of MST and distance ranking.

Key words: methods: data analysis – stars: neutron – pulsars: general.

1 INTRODUCTION

Ever since the discovery of the first pulsar (Hewish et al. 1969), the plot containing the period derivative versus the period of every pulsar found, or simply, the $P\dot{P}$-diagram, has been used as a way of summarizing our knowledge and guiding our research on the pulsar population. Classes of pulsars and possible links among them are referred to in this plot, as it is what we know about their possible evolutionary tracks along the pulsar's lifetime (see e.g. Enoto, Kisaka & Shibata (2019) for a review). However, we have been looking at the same plot for more than five decades. A fresh appraisal may be healthy. Is the $P\dot{P}$-diagram the most useful or complete way to visualize the pulsars we know? Does it introduce any unwarranted bias on what we consider to be similar pulsars? Here, we use principal component analysis (PCA; e.g. Pearson 1901; Shlens 2014) to show that even when the variables considered to describe a pulsar would all depend on P and $\dot{P}$, the variance of the population is not contained in the variance of the latter quantities. Thus, we note that any nearness ranking or visualization based only on P and $\dot{P}$ is potentially misleading. Next, we define and compute a properly normalized distance to measure nearness from one pulsar to another, calculate the minimum spanning tree (MST; e.g. Gower & Ross 1969) of the pulsar population, and discuss possible applications. The pulsar tree hosts information about pulsar similarities that go beyond P and $\dot{P}$, and are thus naturally difficult to read from the $P\dot{P}$-diagram. We also introduce here an online tool encompassing all our results, as well as allowing a user to focus on user-defined problems.

The MST is a graph that connects points in a multidimensional space. Each point (or node) is linked to at least another by an edge, whose length is associated with a given distance. The edges of an MST are chosen so that the the sum of their lengths is minimal and all nodes are linked, implying no cycles are present (no paths starting and returning to the same node). Graph theory shows that as long as distances are distinct, the MST is unique. Its very definition intuitively implies that the MST is an optimization technique. In fact, it was widely used in engineering problems, starting from their original application developed by Boruvka in 1926; for the distribution of electricity in Moravia, see Nešetřil, Milková & Nešetrilová (2001). Currently, MSTs are used from analysing cognitive impairment (Simon et al. 2021) to risk in financial markets (Pozzi, Di Matteo & Aste 2013). Early usage in scientific problems include describing the interrelationship of species or genetics (see the work of Florik in the 1950's and Edwards in 1960's as commented in Hartigan (1981) and (Winther 2018), respectively), disciplines in which it is a widespread technique. In astronomy, it has been used for finding high-energy sources (Campana et al. 2013), establishing differences between cluster and field stars (Sánchez, Alfaro & López-Martínez 2018), detecting filaments (Pereyra et al. 2020), galaxy clustering (Barrow, Bhavsar & Sonoda 1985), and cosmology (Naidoo et al. 2020). It has also been used in the analysis of event samples in particle colliders (Lovelace Rainbolt & Schmitt 2017), or cosmic rays

★ E-mail: crodriguez@ice.csic.es (CRG); dtorres@ice.csic.es (DFT); patruno@ice.csic.es (AP)

(Harari, Mollerach & Roulet 2006). This non-exhaustive reference list is just an example of growing interest in MST use across different fields. Despite this interest, it has been barely been used in relation to pulsars. To our knowledge there is only one related publication (Maritz, Maritz & Meintjes 2016) using 11 handpicked objects. The aim was to show that an MST could distinguish binaries from isolated pulsars using the dispersion measure as distance. In this paper, we explore using the MST to provide a novel classification, alerting, and visualization tool for pulsars.

2 THE PULSAR VARIANCE

2.1 Variable definition

We consider pulsars listed in the ATNF Catalogue (Manchester et al. 2005), including radio pulsars, X-ray and/or gamma-ray pulsars, and magnetars for which coherent pulsations have been detected. Accretion-powered pulsars such as, e.g. SAX J1808.4−3658, are not included in this table, however. The current number of pulsars listed in the catalogue is 3282, of which 2509 have a known period and period derivative (larger than 0). From the latter, 2242 are isolated pulsars and 267 are pulsars residing in binary systems. All the methods considered in this work will be applied to this set as a whole, without making any distinction among the pulsars in it. For characterizing the pulsar population, and ultimately defining a ‘distance’ from one pulsar to another, we shall consider the following physical set of pulsar variables (see, e.g. Lorimer & Kramer 2012 for reference):

(i) Spin period:
P [s],
(ii) Spin period derivative:
$\dot{P}$ [s s^{-1}],
(iii) Surface magnetic flux density (equator):
$B_{\rm s} = (3c^3 I)^{1/2}/(8\pi^2 R^6 \sin^2\alpha)^{1/2}\sqrt{P\dot{P}} \simeq 3.2\times 10^{19} P^{1/2}\dot{P}^{1/2}{\rm G}$,
(iv) Magnetic field at the light cylinder:
$B_{\rm lc} = B_{\rm s}(\Omega R)^3/c^3 \simeq 3\times 10^{8} P^{-5/2}\dot{P}^{1/2}{\rm G}$,
(v) Spin-down energy loss rate:
$\dot{E}_{\rm sd} = 4\pi^2 I\dot{P}P^{-3} \simeq 3.95\times 10^{46} P^{-3}\dot{P}{\rm ergs}^{-1}$,
(vi) Characteristic age:
$\tau_{\rm c} = P/2\dot{P} \simeq 15.8\times P\dot{P}^{-1}{\rm Myr}$,
(vii) Surface electric voltage:
$\Delta\Phi = (B_{\rm s}4\pi^2 R^3)/(2cP^2) \simeq 6.3\times 10^{5} P^{-3/2}\dot{P}^{1/2}{\rm V}$,
(viii) Goldreich–Julian charge density:
$\eta_{\rm GJ} = (\Omega B_{\rm s})/(2\pi c e) \simeq 7\times 10^{10} P^{-1/2}\dot{P}^{1/2}{\rm cm}^{-3}$.

Here the moment of inertia I was assumed as 10^{45} g cm^{-2}, the radius of the star R was assumed as 10 km, and the inclination α between the magnetic and rotation axes as 90°. The remaining constants (c, e) have the usual meaning.

The measurable quantities P and $\dot{P}$ are the leading magnitudes in this set of variables, from which all others are calculated using the rotating dipole model, as is usual for pulsar estimations.

The surface magnetic field and spin-down power are basic magnitudes critical to characterize the pulsars’ energetics and their magnetospheres. In our set, they are complemented with others to incorporate the fact that dissimilar pulsars (e.g. millisecond and normal) can have similar magnetospheres. This is partly described by the magnetic field at the light cylinder, which is similar in both cases. The voltage gives the potential drop between the magnetic pole and the edge of the polar cap. It is thought here to represent the variety introduced by the electromagnetic configuration, for which another parameter of interest is the Goldreich–Julian charge density, $\propto B_{\rm s}/P$. Note that these magnitudes, being all functions of P and $\dot{P}$, can have a relationship between themselves, as just noted.

The mass and the radius of the neutron stars would ideally also be considered as part of the variables of interest in our study. However, this information is only available for a very small percentage of the sample.

Other variables of interest are those related to the birth properties of pulsars, such as the initial spin-down power, the initial magnetic field, or the spin-down time-scale. However, these are not known for most pulsars in our sample. They all depend on the unknown (except for a few) pulsars’ real age (for which the characteristic age $\tau_{\rm c}$ is only a proxy). Similarly, the braking index is measured for just a handful of pulsars (and, in addition, it is known that it may vary significantly). Estimates from it using $\ddot{P}$ would significantly reduce the sample. Finally, other measurable quantities exist that are unrelated to intrinsic properties (transverse velocities, DM, distances) and/or are known for a limited number of objects. Using luminosities and other properties at different frequencies (e.g. fluxes, pulse shapes, peak separation, etc.) would also significantly cut the sample, would be affected by extrinsic conditions (absorption, distance), and/or would incorporate parameters that are difficult to compare for the population as a whole. These are left for analysis in future work, where particular sub-samples will be looked at in more detail taking into account different variables, focusing here only on intrinsic pulsar features to introduce the technique.

2.2 Treating variables

The values of the magnitudes considered may differ by several orders of magnitude for different pulsars. The distributions of the logarithm of these variables are shown in Fig. 1.

The distributions are not normal (as is also the case for the set of original variables, without logarithm). Two clear populations of millisecond and normal pulsars appear separately in all plots, except the two related to the spin-down power and voltage. As is obvious from their centralization (mean and median) and dispersion (standard deviation and interquartile range, IQR), the log-variables are orders of magnitude closer together. Given that the distributions are not normal, we use the *robust scaler* for normalizing the log variables,

$$x_i^{\dagger} = \frac{x_i - Q_2}{IQR}, \tag{1}$$

where the †-symbol represents that the quantity x_i has been normalized; Q_1, Q_2, and Q_3 represent the first quartile, median, and third quartile of the distribution, respectively; and IQR is the interquartile range, $(Q_3 - Q_1)$. It can be seen that the distributions of the variables after being normalized have a median equal to zero and an IQR equal to 1. Note that if we take the logarithm of the variables and then apply equation (1) to normalize it, relations between variables are more clearly uncovered. For instance, $\dot{E}_{\rm sd}$ and $\Delta\Phi$ lead to $\log \dot{E}_{\rm sd}^{\dagger} = \log \Delta\Phi^{\dagger}$, that is also visible in the corresponding distributions of Fig. 1. Considering both of them at once in defining the nearness of two given pulsars relates to the fact that the physical meaning represented by the two original magnitudes is different.

3 PRINCIPAL COMPONENT ANALYSIS

PCA is especially suitable to isolate the main factors introducing variance in a population when the variables known for it are not independent (see Appendix A). Since six of the variables taken to describe the intrinsic properties of the pulsar population are in fact computed from P and $\dot{P}$, one can intuitively conclude that two

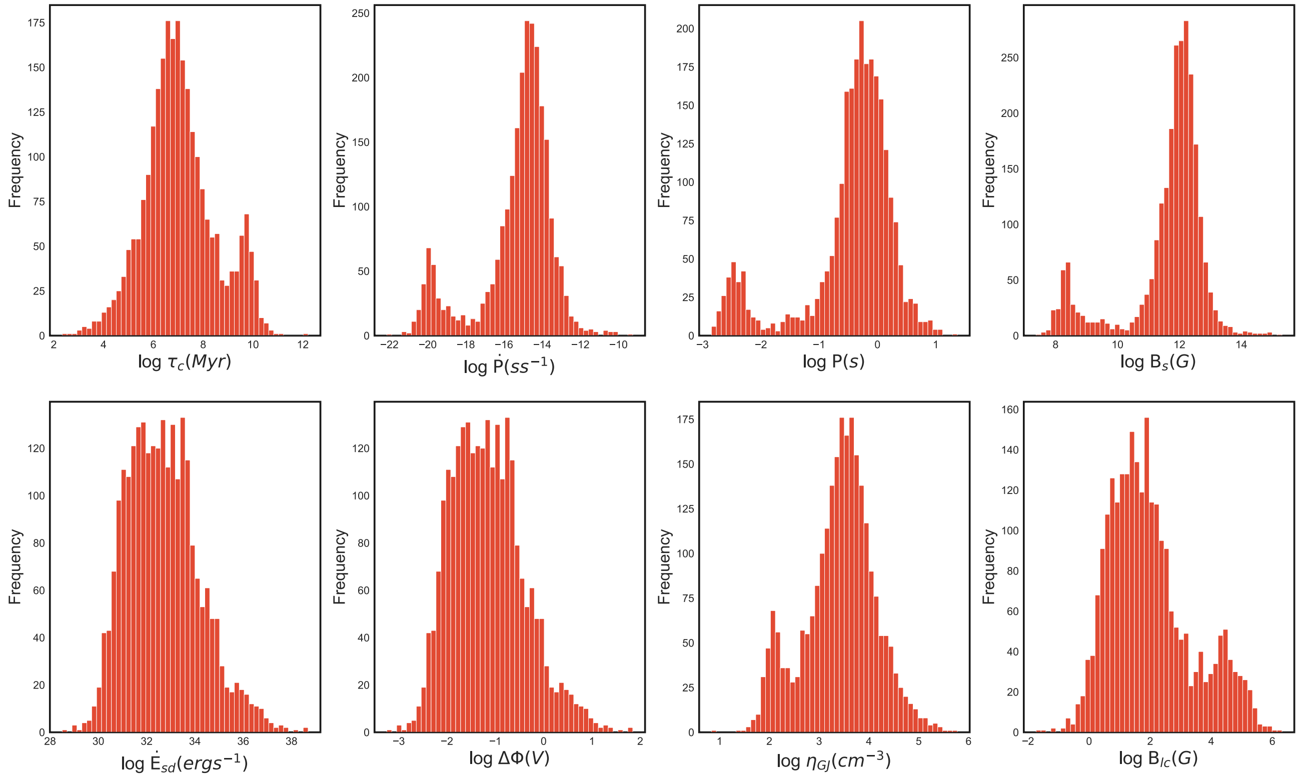

Figure 1. Distribution of the logarithm of the eight variables used for the complete set of 2509 pulsars.

principal components (PCs) are needed to describe our population variance. However, the latter *is not* contained in the variance of P and $\dot{P}$. Said otherwise, P and $\dot{P}$ are not the two PCs needed. Thinking only in terms of P and $\dot{P}$ to compare pulsars may thus be misleading, except in the extreme case where these values are simply the same. Fig. 2 shows the result of the PCA analysis. The two PCs are

$$PC_1 = 0.21B^{\dagger}_{\mathrm{lc_l}} - 0.29\eta^{\dagger}_{\mathrm{GJ_l}} + 0.05\Delta\Phi^{\dagger}_{\mathrm{l}} + 0.05\dot{E}^{\dagger}_{\mathrm{sd_l}} - 0.46\dot{P}^{\dagger}_{\mathrm{l}} - 0.59B^{\dagger}_{\mathrm{s_l}} - 0.47P^{\dagger}_{\mathrm{l}} + 0.29\tau^{\dagger}_{\mathrm{c_l}}$$
$$\text{(71.6 per cent of the explained variance),} \quad (2)$$

$$PC_2 = 0.43B^{\dagger}_{\mathrm{lc_l}} + 0.32\eta^{\dagger}_{\mathrm{GJ_l}} + 0.47\Delta\Phi^{\dagger}_{\mathrm{l}} + 0.47\dot{E}^{\dagger}_{\mathrm{sd_l}} + 0.19\dot{P}^{\dagger}_{\mathrm{l}} + 0.05B^{\dagger}_{\mathrm{s_l}} - 0.36P^{\dagger}_{\mathrm{l}} - 0.32\tau^{\dagger}_{\mathrm{c_l}}$$
$$\text{(28.4 per cent of the explained variance),} \quad (3)$$

where the †-quantities refer to the normalized ones as in equation (1), and the sub-index l stands to note that the normalization is applied to the logarithm of the variable.

After some algebra (see Appendix A) these latter equations can be reformulated as

$$PC_1 = -8.471 - 1.178\log P - 0.832\log\dot{P}, \quad (4)$$

$$PC_2 = 14.182 - 2.931\log P + 1.105\log\dot{P}. \quad (5)$$

Note that no dag marking is herein needed, as we have absorbed the corresponding IQR and median of each variable into the coefficients and that the units of P and $\dot{P}$ are as in equation (3).

The left-hand panel of Fig. 3 shows the pulsar population in the $\log P - \log\dot{P}$ together with lines representing equal values of the PCs. The right-hand panel of Fig. 3 shows the same pulsars but directly in the plane PC_1, PC_2. Nearness in one plane has not the same meaning as it has in the other. To exemplify this, we plot a circle in the P, $\dot{P}$ diagram and see how the circle transforms to the PC_1, PC_2 plane via the equations above. This is shown in the second row of Fig. 3; the difference in relative distances can be up to a factor of 3 or more. This change of shape advances the idea that any sort of nearness ranking will be affected if considering the PCs, instead of P and $\dot{P}$ (further comments about this can be found in the Appendix).

4 MST OF THE PULSAR POPULATION

Appendix B introduces all concepts needed to understand and compute an MST, and we take this for granted in what follows. We define an Euclidean distance using the eight normalized (via equation 1) logarithm of the variables introduced above. Equivalently, after our analysis of the previous section, we can use just two variables according to the PCA analysis, that contains the whole population variance. Both choices end up producing the same MST (and thus the results from the analysis that follows from it are the same) but the latter is less demanding given the reduced dimensionality of the problem. With this Euclidean distance, we first obtain a complete, undirected, and weighted graph $G(V, E) = G(2509, 314\,6286)$, with $|V|$ nodes and $|E|$ edges, where each edge is defined by a specific weight value. From that, we obtain the pulsar MST, which is shown in Fig. 4.

4.1 Branch analysis and pulsar classification in the MST

As shown in Fig. 5, mixing nodes from different branches of the MST produces a scattered distribution of variables. This is a generic

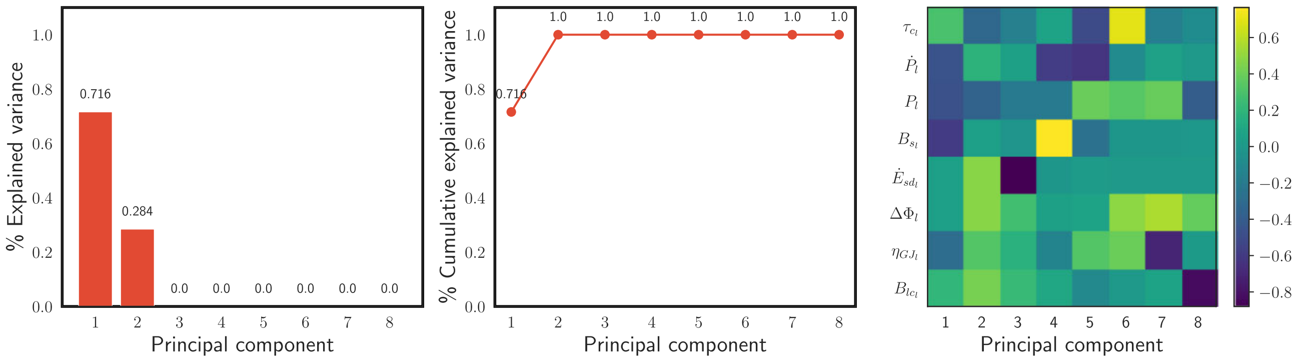

Figure 2. PCA results using the logarithm of the set of variables for the whole population of 2509 pulsars. The left-hand panel shows the explained variance contributed by each of the PCs according to the eigenvalues of the covariance matrix. It represents the amount of information contained in each PC. The central panel shows the cumulative variance explained by the new set of variables that has been defined through the PCA analysis. The right-hand panel shows the 'weight' that each variable has with respect to each PC. This value is the coefficient that is held in each eigenvector. Negative values imply that the variable and the PC in question are negatively correlated. Conversely, a positive value shows a positive correlation between the PC and the variable.

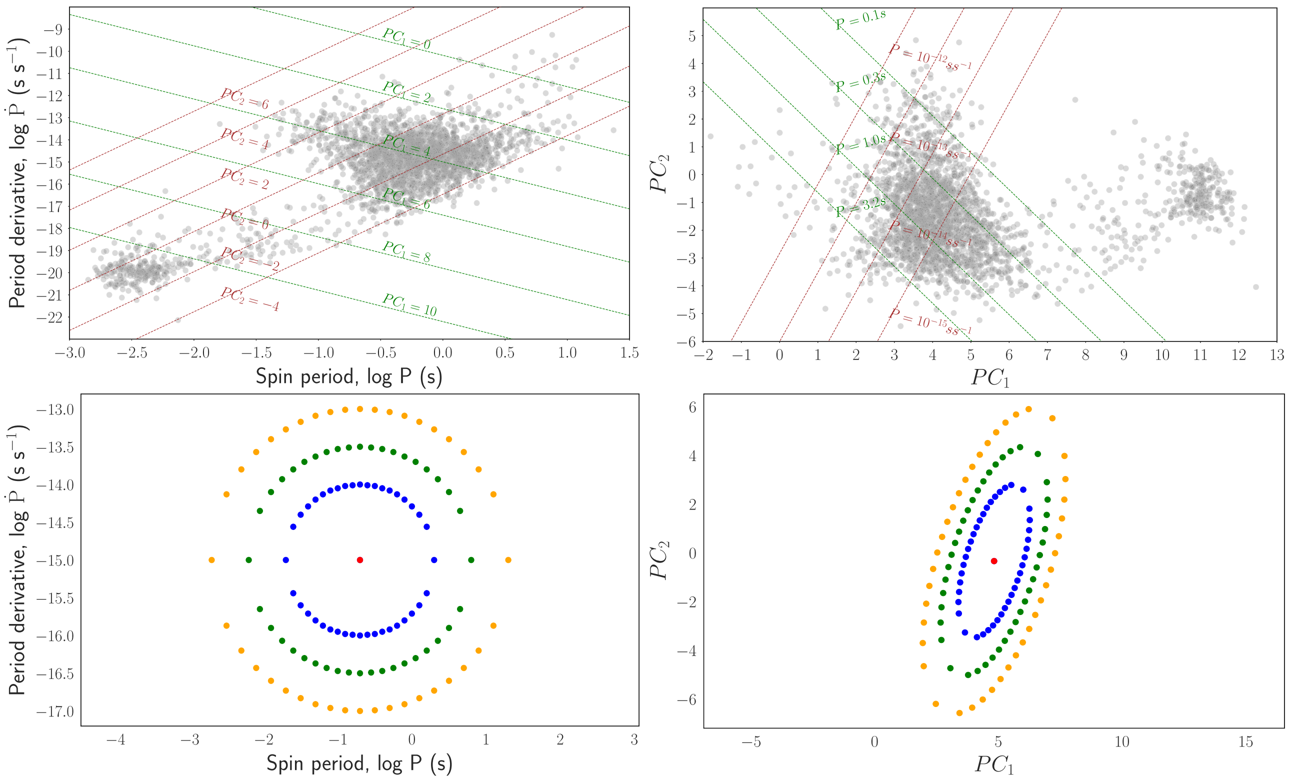

Figure 3. Top left-hand panel: representation of the total pulsar population as a function of the P and $\dot{P}$ variables. Constant values of PC_1 and PC_2 are shown in green and brown lines, respectively, according to equations (4) and (5). Top right-hand panel: representation of the total pulsar population as a function of PC_1 and PC_2 whose values are obtained according to the eight variables taken into account as shown in equations (2) and (3). Constant values of P and $\dot{P}$ are shown in green and brown lines, respectively, taking into account equations (4) and (5). Bottom left-hand panel: synthetic pulsars positioned circularly at different radii from a given centre. Bottom right-hand panel: transformation of the circle through equations (4) and (5).

behaviour that happens for any mixing of the branches in the MST, and for any mixing of the nodes even within a single branch (see the panels in the three last rows of Fig. 5). If we read the MST in a disordered manner, nothing is learned from it. Instead, Fig. 6 shows that if we choose one of the branches at a time and run along with the nodes in it in an orderly manner, a smooth behaviour of the variables naturally appears. Mixing branches of the pulsar tree is equivalent to grouping pulsars by their nearness in the $P\dot{P}$-diagram. This is visually shown in Fig. 7. The pulsar tree hosts information that is difficult to read from the $P\dot{P}$-diagram.

4.1.1 The MST as a descriptive tool

The ordering introduced by each of the branches is indicative that there is an internal physical grouping in the MST. This is shown in Fig. 8, where we also show the variation of the principal

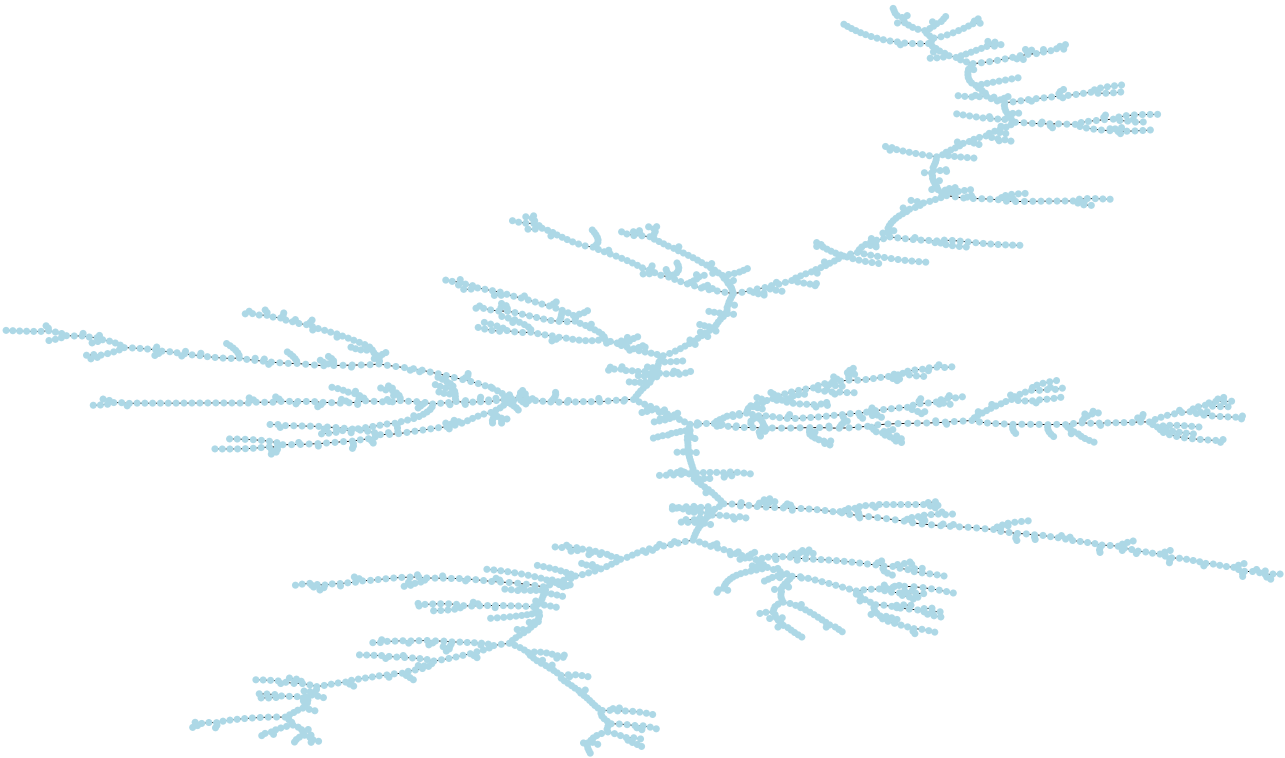

Figure 4. A different look at the pulsar population. MST-graph (2509, 2508) based on the complete, undirected, and weighted graph $G(2509, 3146286)$ for the 2509 pulsars and their full combination of weights computed from their Euclidean distance among eight normalized variables (or the equivalent two PCs). Each node in the MST represent a pulsar. Branches group pulsars with particular characteristics.

components PC_1 and PC_2. These variations serve to visualize the physical properties of different pulsar classes, and understanding them may lead to physical connections among pulsars, or links in their evolution. To emphasize this, we shall observe how some known groups of pulsars locate in the MST.

The main tree trunk travels from young and energetic pulsars in the bottom[1] to millisecond pulsars in the top (see panels – referred from the left to right and top to bottom – 1, 2, and 5 of Fig. 8). However, other than representing the overall separation of the binary pulsars from the rest of the sample, the variables used in this MST do not allow us to dig deeper into sub-population of binaries: they are heavily affected by the evolutionary processes that occur for MSPs during the accreting (recycling) phase. As an example, consider black-widows (BWs), redbacks (RBs), and transitional pulsars (tMSPs; see Papitto & Bhattacharyya 2022 for reviews). Black Widows and redbacks are binary MSPs where the companion is a semi-degenerate star or a main-sequence star, respectively. The tMSPs (of which only two tMSPs (of the three known) have measured values of P and $\dot{P}$) are instead binary MSPs that switch between an accreting/high X-ray emission state and a radio pulsar state. No clear grouping appears for these binaries sub-samples (see Fig. 9), as the intrinsic pulsar in them are similar. In future work, we shall supplement intrinsic variables with other representing the companion, the orbital parameters, and the environment of binary pulsars to try to address these issues.

The branches departing from the main trunk cover deviations that are better represented by the variability in one or a few of the magnitudes considered. The extremes of each branch are thus extreme pulsars of the population in a particular way.

For instance, the longest rightwards branch (moving along it towards the rightmost node) groups pulsars with increasing age, period, and period derivative. Pulsars in this branch are not particularly energetic nor do they have large magnetic fields at the light cylinder. Instead, they have an increasing surface field, reaching up to extreme values. The second-longest rightwards branch has similar behaviour, but is formed by less magnetized objects at their surface, and also less energetic, slower, and older. Not surprisingly, then, magnetars are located in both of these branches, as is also the only XDIN, J1856−3754, quoted in the ATNF catalogue. We show this in more detail in Fig. 9. Interestingly, the XDIN and the magnetars do not share the same branch.

Some of the low-field magnetars, since they are more energetic, less magnetized objects that have nevertheless shown magnetar-flaring behaviour, appear quite separate from the rest. In fact, the low-field magnetars depicted in the MST having $P < 1$ are J1846−0258 (Gavriil et al. 2008) and J1119−6127 (Archibald et al. 2016), and they appear in Fig. 9 in the lower leftmost branch, one almost on top of the other in this scale. This is a very different location from where J2301+5852, J1647−4552, J1822−1604, and J0418+5732, the other low-field magnetars, are. They are all located at the end of the branch going rightwards, above the magnetars.

The different branches at the bottom of the MST contain all energetic pulsars. They share the same values of light cylinder magnetic fields of the millisecond pulsars, at the other end of the MST, despite they are very different in almost every other aspect. The small branches in which the energetic pulsars divide at the bottom of

[1] In what follows, we use ‘bottom’, ‘top’, ‘right’, or ‘left’ to facilitate the referral to a particular position in the MST shown. However, we emphasize that (in lack of axis) what matters is not the representation but the edges among nodes – or, in technical terms, the adjacency matrix.

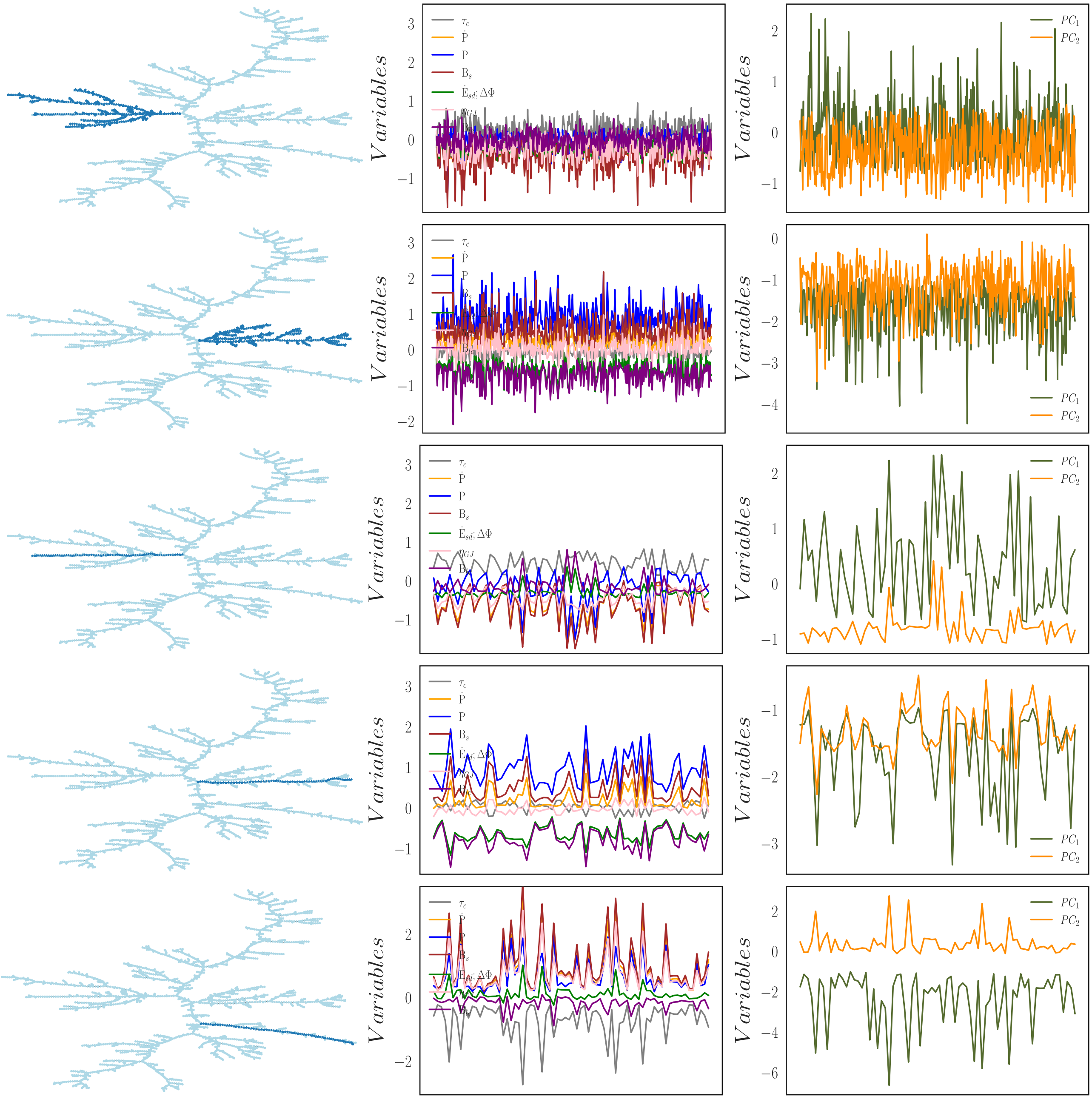

Figure 5. Top two rows: In the left-hand panel, nodes from different branches are highlighted in dark blue in the MST. The middle and right-hand panels show the behaviour of the eight normalized variables and two PCs when following an arbitrary mixing of the nodes in the highlighted part of the MST. Third, fourth ,and fifth rows: the same, but now arbitrarily mixing the nodes of only one branch. Note that for the purpose of plotting both PC_1 and PC_2 in the same scale in the right-hand panels of Fig. 5, and also Fig. 6, we have subtracted the corresponding mean – from the whole set – from each set of values.

the plot also separate them into those having a significantly higher value of different variables; like $\dot{E}$, $B_{\rm lc}$, $B_{\rm s}$. Fig. 9 consistently shows how the pulsars associated with TeV confirmed or candidate PWNe (H. E. S. S. Collaboration et al. 2018) are essentially all located in these branches. Only three TeV PWNe, B1742−30(1)/J1745−3040, J1858+020/J1857+0143, and CTA 1/J0007+7303 are somewhat outliers. The former two are TeV PWN candidates, the oldest and less energetic PWNe in the population (see table 4 of H. E. S. S. Collaboration et al. 2018). The MST by itself cannot judge the reality of the association proposed, but emphasizes how distinct these two are with respect to the rest. The case of CTA 1 has been studied in detail as a possible PWN in the reverberation phase, and/or having a higher magnetization (Martín, Torres & Pedaletti 2016). Its peculiarity has been already noted from a physical standpoint, although it is less of an outlier in comparison to the rest of the PWNe population (both in the MST and in comparing PWNe models).

Another interesting example of the MST view on pulsars is to note where the *Fermi*-LAT-detected gamma-ray pulsars that are part of the ATNF fall on it (see Abdo et al. 2013; Fermi-LAT Collaboration 2021). Fig. 9 shows two panels to this effect, where the ranges they have for the light cylinder magnetic field and spin-down power are

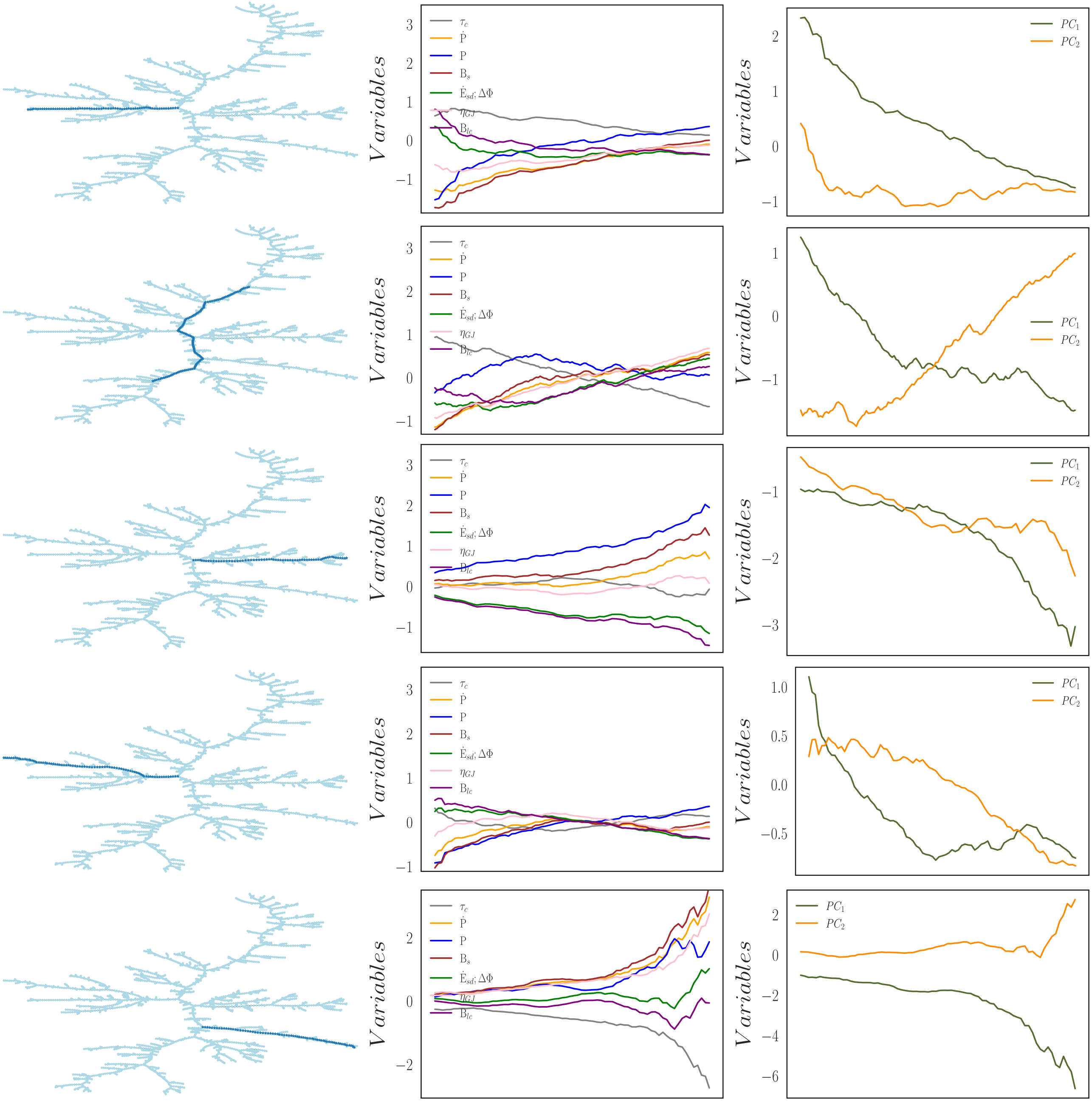

Figure 6. Similar to the panels in Fig. 5, but here following the node sequence as appearing along the marked branch of the MST. The node sequence is here read from the left to right (except in those branches that appear mostly vertical, where they are read from the top to bottom).

noted. The detected gamma-ray pulsars cluster at specific locations of the tree, most of it being empty of gamma-ray emission. This corresponds to the fact that gamma-ray pulsars have high values of light cylinder magnetic field, $B_{\rm lc} > 100$ G, with most having actually $B_{\rm lc} > 10^3$ G, and relatively high spin-down power. Some other magnitudes are less decisive, e.g. there are gamma-ray pulsars across the full range of $\eta_{\rm GJ}$-values. Along the branches where the *Fermi*-LAT pulsars lie there might be a close sequence of detected and non-detected pulsars, despite the MST clearly showing the similarity of their intrinsic properties. This is a representation that extrinsic features such as distance, environment, or geometry play a role in *Fermi*-LAT detectability at an individual level.

The *Fermi*-LAT pulsar isolated in the central part of the MST is J2208+4056, the only one depicted with $B_{\rm lc} < 100$ G. This pulsar has been noted by Smith et al. (2019) as having a spin-down ($\sim 8 \times 10^{32}$ erg s^{-1}) about three times lower than the previously observed gamma-ray emission death-line. The outlier *Fermi*-LAT pulsar in the left-hand part of the MST is J1231−5113 and has an even lower spin-down power ($\sim 5 \times 10^{32}$ erg s^{-1}). In comparison, the other magnitudes $B_{\rm lc}$, τ, $\eta_{\rm GJ}$, are similar to the rest of the gamma-ray population.

We also investigate the position in the MST of the 40 RRATs with known P and $\dot{P}$ appearing in the ATNF (Abhishek et al. 2022; Cui & McLaughlin 2022). RRATs are pulsars showing extreme radio

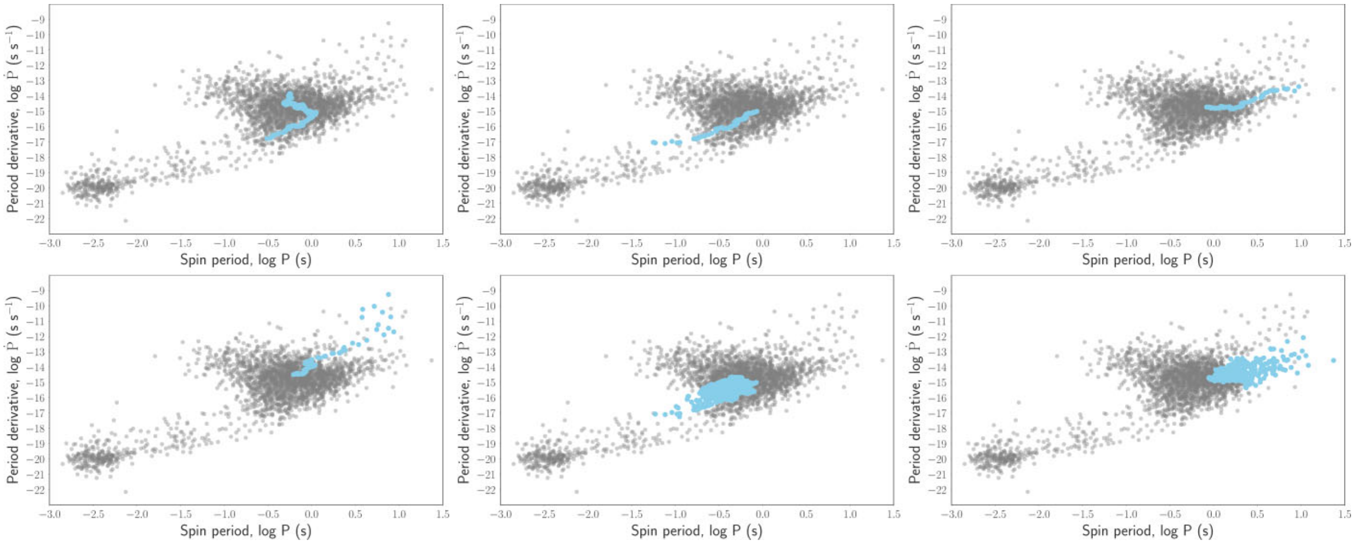

Figure 7. Positions of the branches analysed in Figs 5 and 6. We show the main trunk of the tree (top left-hand panel) followed by the branch depicted in row 3 of Fig. 5 (or row 1 of Fig. 6), row 4 of Fig. 5 (or row 3 of Fig. 6), and row 5 of Fig. 5 (or row 5 of Fig. 6). the last two panels show the corresponding pulsars to the branches marked in rows 1 and 2 of Fig. 5.

variability, as most of them are discovered through their single, isolated pulses. Only one of the 40 RRATs considered is located in the main trunk. This pulsar is near the degree 3 node J1828−1336 that separates the main trunk in the two central branches.

The MST can also be used to analyse any other pulsar population, for instance, where do the pulsar's known glitches, or the intermittent and nulling pulsars, group. The online tool provided with this work promotes this kind of analysis.

4.1.2 *The MST view of evolutionary tracks*

While pulsars evolve, they change their timing parameters and move across the $P\dot{P}$-diagram. Evolutionary models are then constructed following the fully coupled evolution of temperature and magnetic field in neutron stars (e.g. see Viganò et al. 2013). To simulate evolutionary tracks, we have created synthetic pulsars over the theoretical tracks of fig. 10 of Viganò et al. (2013) and individually studied where would they fall in the MST should they be part of our sample. This is shown in Fig. 10 where the arrows show increasing age at a fixed initial magnetic field and the rounded cap point in the origin of the arrows shows possible birthplaces. We find, in agreement with what was already discussed in Fig. 6, tracks are not randomly found in the MST. There are two birth zones in it, at the bottom part where we find all energetic pulsars, and at the rightmost branch, where we find the magnetars. Then, for low initial magnetic field values, pulsars go on to die on the main trunk of the MST. When the field increases, the pulsars populate the middle branches. When the initial field is high enough to shift the birthplace from the energetic pulsar zone to the classical magnetar range, the evolution is mostly confined to the two rightmost branches in Fig. 10. In these branches then, we find pulsars evolving in the two directions (from the main trunk to the extreme and vice versa) depending on their place of origin.

4.1.3 *The MST as an alerting tool*

In addition of providing a descriptive perspective, the MST might be used as an alerting tool for pulsars of interest. These are some examples taking into account the MST location and the distance ranking. Implied connections are not always obvious using $(P, \dot{P})$ only (see the discussion on the PCA and the distance ranking above and in the appendices). We use the web application provided with this work to note the following:

(i) Based on the location of the energetic low-field magnetars J1846−0258, J1119−6127 at the bottom part of the MST, and due as well to its nearness ranking, other pulsars with essentially the same characteristics are noted, in particular, J1208−6238. It has been suggested as a possible low-field magnetar in the literature (Clark et al. 2016) and is second (first) in the distance ranking of J1846−0258 (J1119−6127) after J1119−6127 (followed by J1846−0258). PSRs J1513−5908 (in the composite S/N MSH 15−52), J1640−4631, and J1930+1852 follow in the ranking of J1846−0258; and J1640+4631, J1614−5048, and J1513−5908 do so in the ranking of J1119−6127.

(ii) Panel 5 of Fig. 8 shows that few locations of the MST showing $B_{\rm lc} \sim 100$ G or beyond and no detected gamma-ray pulsars yet. These regions become of special interest for future searches. In particular, those near J1231−5113 (which is already detected) in the MST appear to be promising potential targets. A few neighbouring pulsars to the outlier J1231−5113, at the end of this branch, also show a relatively large $B_{\rm lc}$ with a similar range of spin-down power, and other variables, in comparison to *Fermi*-LAT pulsars. Likewise, PSR J1915+1616 and J2129+1210B at the end of the nearby branch are of interest. Again, note that both have $B_{\rm lc} > 10^3$ G, and $\dot{E} > 10^{33}$ erg s^{-1}.

(iii) The detection of the radio pulsar PSR J2208+4056 by the *Fermi*-LAT in spite of its low spin-down power has been ascribed by Smith et al. (2019) to a possible case of favorable geometry. If this is the case, it may remain indeed isolated in the MST, where appears close to the main trunk. Its closest neighbours (J0532−6639, J0502+4654, and J1848−0123) call for attention in order to test this.

(iv) Others pulsars of interest regarding their possible detection in gamma-rays may be J1818-1607 and J1550−5418. These lie in the magnetar branch, where no other detected *Fermi*-LAT pulsar is located (Li et al. 2017). The latter authors established a *Fermi*-LAT integrated upper limit for J1550−5418 and attempted folding,

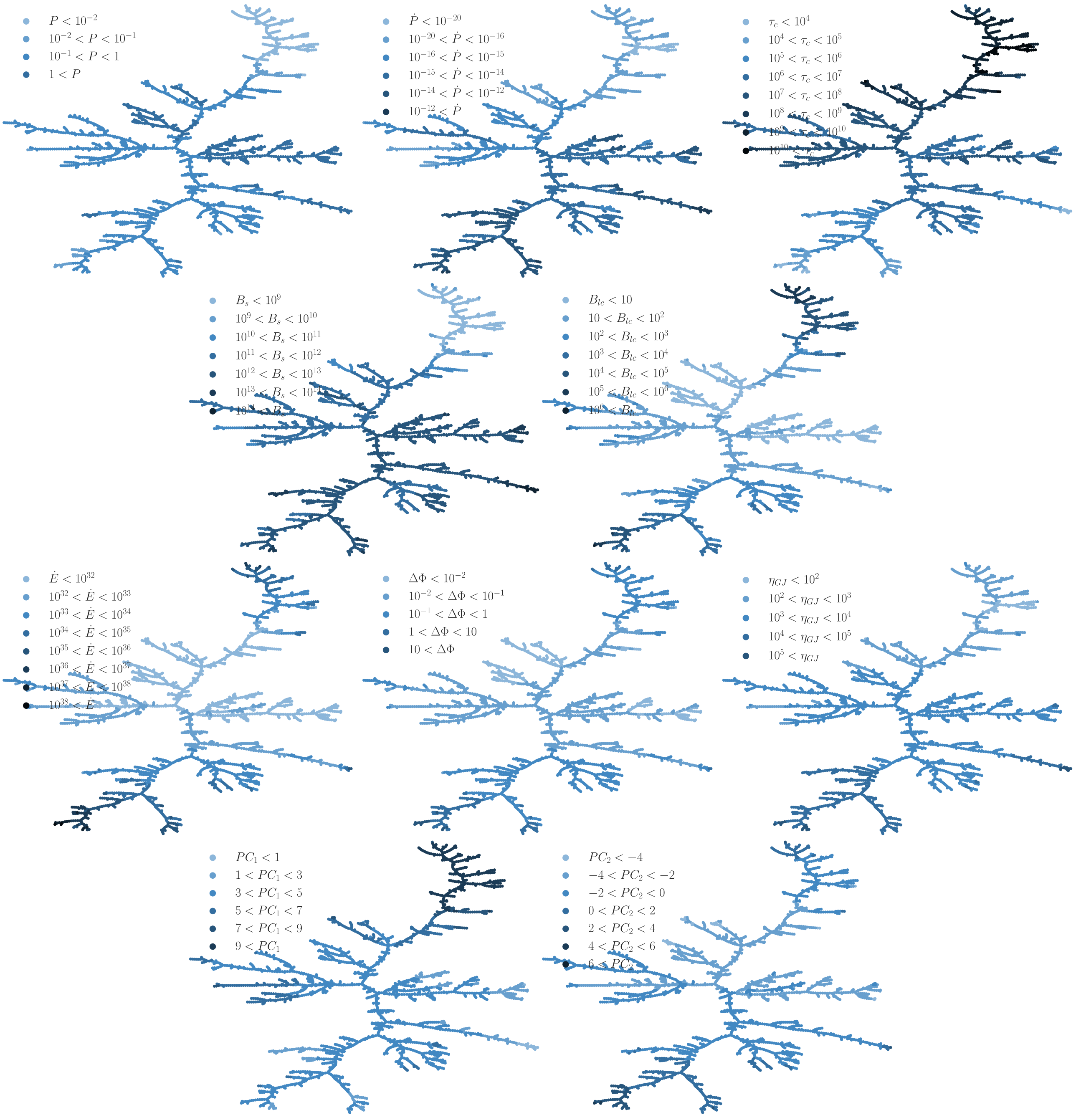

Figure 8. Representation of the values of the different pulsar magnitudes considered directly on to the MST. The first row shows P, $\dot{P}$ and τ; the second row shows the magnetic field at the surface and at the light cylinder; the third row shows the spin-down power, the voltage and the Goldreich–Julian current, and, finally, the last row shows directly the principal components PC_1 and PC_2 values.

finding no signal. However, these pulsars have similar properties to other pulsars already detected by *Fermi*-LAT (ordered by distance, J1208−6238 occupies the third position, followed by J1119−6217, both detected by *Fermi*-LAT). In the case of the pulsar J1550−5418, among its 10 closest pulsars according to the calculation of the Euclidean distance, the pulsars J1734−3333, J1746−2850, and J1726−3530 are in the sixth, eighth, and ninth positions, respectively. They are located in the lower part of the MST, where there is a high density of *Fermi*-LAT pulsars, although they have not yet been detected in gamma-rays. The distance ranking of these two magnetars is uncommon to others: other magnetars in the branch have neither a *Fermi*-LAT pulsar nor other pulsars located in high-density areas of *Fermi*-LAT detections in the first positions of their distance ranking.

(v) The pulsars J1842−0905 and J1457−5902, and J1413−6141 and J1907+0631 are the closest neighbours to the PWNe J1745−3040/PWN B1742−30(1) and J0007+7303/PWN CTA 1, respectively. The latter are somewhat outliers of the PWNe population (see above) and thus their neighbours are of interest to test whether this region of the pulsar parameter space is prone to producing observable nebulae.

(vi) Finally, the higher degree nodes –particularly those connected with the main trunk, the nodes that are extremes of significant

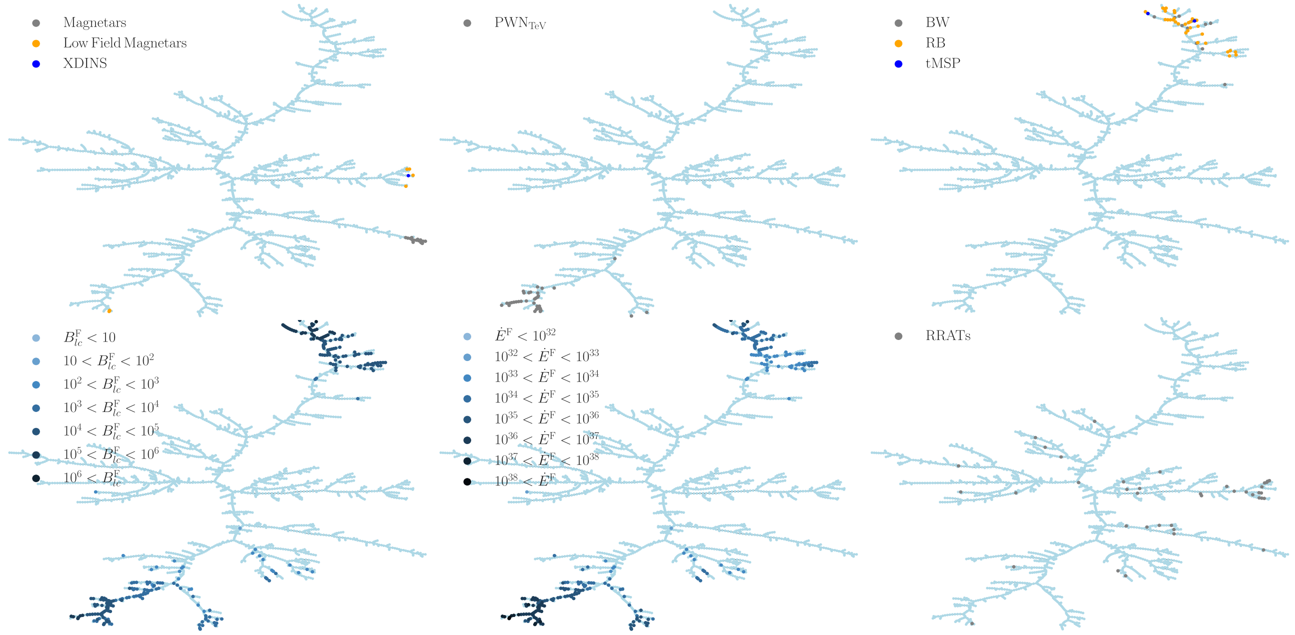

Figure 9. Representation of the members of different pulsar classes on to the (PC_1, PC_2) MST. The first row shows magnetars; redbacks, black widows and transitional pulsars, and pulsars associated with TeV pulsar wind nebulae (PWNe). The first two panels in the second row shows the *Fermi*-LAT pulsars (thus the superscript F), further characterized against their spin-down power and light cylinder radius, as depicted in Fig. 8. The last panel shows the rotating radio transients (RRATs).

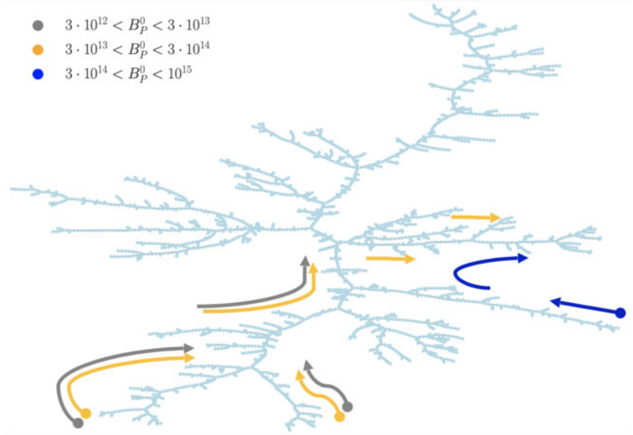
$3 \cdot 10^{12} < B_P^0 < 3 \cdot 10^{13}$
$3 \cdot 10^{13} < B_P^0 < 3 \cdot 10^{14}$
$3 \cdot 10^{14} < B_P^0 < 10^{15}$

Figure 10. Representation of evolutionary tracks described in Viganò et al. (2013) into the MST, according to their initial magnetic field.

branches, and possibly other topologically selected nodes, are suggested for individual study, as they may have undiscovered relevance.

Note that we are not considering the physical distances to the pulsars in our MST, which is only based on intrinsic pulsar properties. Thus, it may well be the case (as it so happens) that one of the noticed pulsars above is in the LMC (J0532−6639): In such occasions, the MST may just be pointing that intrinsically this pulsar may emit similarly to others detected, despite it maybe too far to be seen with current instrumentation.

6 DISCUSSION

We have looked at the pulsar population from a different perspective from the usual $P\dot{P}$ diagram. Instead of considering just P and $\dot{P}$ we used a set of eight variables as proxies for intrinsic physical properties of all pulsars. While these variables all depend on P and $\dot{P}$, the variance of the population is not fully contained by the variance of the latter quantities. P and $\dot{P}$ are not the PCs. Distance ranking (or visualizations) based only on P and $\dot{P}$ may hide interesting connections and mislead our intuition. We subsequently computed the pulsar MST, using a properly normalized Euclidean distance, discussed its properties and how the different classes of pulsars find an ordered place into it.

The MST approach offers applications beyond what we have described above. For instance, advanced analysis of the MST regarding clustering, centralization, betweenness, and closeness can illuminate physical connections and link pulsars among themselves. Similarly, a quantitative comparison of MSTs constructed for synthetic populations of pulsars offers a way to qualify the goodness of the population synthesis model that created them. The method used here can also be generalized to consider other variables in the distance definition and the PCA, and then in the MST done with them, allowing different problems to be treated. For instance, focus can be put on binary pulsars, and for them, environmental and orbital parameters (DM, orbital period, orbital size, nature of the companion, etc.) can be thought of as part of the distance definition. Another possibility in this regard would be to consider emission properties in particular regimes, distance definitions containing spectral parameters in a given energy range, or light-curve properties. The forthcoming third *Fermi*-LAT pulsar catalogue may be especially appropriate for working in this direction. The MST technique usage in astrophysics is not as widespread yet as it is in other fields, albeit it has the potential to offer new perspectives in classification, clustering, source identification, and cross-correlation of source properties, including pulsars.

ACKNOWLEDGEMENTS

We have made an extensive use of the ATNF Pulsar catalogue v1.67: as updated on 2022 March 7. This work has been supported by the grants PID2021-124581OB-I00 and PGC2018-095512-B-I00. CR is funded by the PhD FPI fellowship PRE2019-090828, and acknowledges the graduate program of the Universitat Autònoma of Barcelona. DFT acknowledges as well USTC and the Chinese Academy of Sciences Presidential Fellowship Initiative 2021VMA0001. AP acknowledges partial support from a Ramon y Cajal fellowship RYC-2017-21810. This work was also supported by the Spanish program Unidad de Excelencia María de Maeztu CEX2020-001058-M.

5 DATA AVAILABILITY AND THE PULSAR TREE WEB

The pulsar tree web[2] accompanying this paper contains visualization tools and data to produce all plots and go beyond what has been presented in this paper. It allows the readers to gather information in terms of MST localization, $P\dot{P}$ comparison, and distance ranking. Among the functionalities included already, it can locate a given pulsar of the sample in the MST, $P\dot{P}$, PC_1PC_2-diagrams and on any other diagram using the variables adopted in this paper; identify all properties of the given pulsar and all neighbouring nodes both in the MST; zoom around a given portion of the MST (or on any of the other diagrams); obtain tables of the properties of the nodes in the region of interest; and obtain tables of the distance ranking for any pulsar, and more. As our research continues, we expect to upgrade the pulsar tree web with new functionalities.

DATA AVAILABILITY

The data underlying this paper are available in 'The pulsar tree' web at the Institute of Space Sciences (ICE, CSIC) http://pulsartree.ice.csic.es.

REFERENCES

Abdo A. A. et al., 2013, ApJS, 208, 17

Abhishek, Malusare N., Tanushree N., Hegde G., Konar S., 2022, preprint (arXiv:2201.00295)

Archibald R. F., Kaspi V. M., Tendulkar S. P., Scholz P., 2016, ApJ, 829, L21

Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, MNRAS, 216, 17

Buitinck L. et al., 2013, ECML PKDD Workshop: Languages for Data Mining and Machine Learning

Campana R., Bernieri E., Massaro E., Tinebra F., Tosti G., 2013, Ap&SS, 347, 169

Clark C. J. et al., 2016, ApJ, 832, L15

Cui B., McLaughlin M., 2022, RRATalog (on-line catalog), http://astro.phys.wvu.edu/rratalog/

Enoto T., Kisaka S., Shibata S., 2019, Rep. Prog. Phys., 82, 106901

Erickson J., 2019, Lecture Notes, Algorithms. University of Illinois at Urbana-Champaign, Urbana-Champaign, IL

Fermi-LAT Collaboration, 2021, Public List of LAT Detected Gamma-Ray Pulsars (on-line catalog), https://confluence.slac.stanford.edu/display/GLAMCOG/Public+List+of+LAT-Detected+Gamma-Ray+Pulsars

Gavriil F. P., Gonzalez M. E., Gotthelf E. V., Kaspi V. M., Livingstone M. A., Woods P. M., 2008, Science, 319, 1802

Gower J. C., Ross G. J. S., 1969, J. R. Stat. Soc. C, 18, 54

H. E. S. S. Collaboration et al., 2018, A&A, 612, A2

Hagberg A. A., Schult D. A., Swart P. J., 2008, in Varoquaux G., Vaught T., Millman J., eds, Proceedings of the 7th Python in Science Conference, Pasadena, CA, p. 11

Harari D., Mollerach S., Roulet E., 2006, Astropart. Phys., 25, 412

Harris C. R. et al., 2020, Nature, 585, 357

Hartigan J. A., 1981, J. Am. Stat. Assoc., 76, 388

Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1969, Nature, 224, 472

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Kleinberg J., Tardos E., 2005, Algorithm Design. Addison Wesley, The United States of America

Kruskal J. B., 1956, Proc. Am. Math. Soc., 7, 48

Li J., Rea N., Torres D. F., de Oña-Wilhelmi E., 2017, ApJ, 835, 30

Lorimer D. R., Kramer M., 2012, Handbook of Pulsar Astronomy. Cambridge Univ. Press, Cambridge

Lovelace Rainbolt J., Schmitt M., 2017, J. Instrum., 12, P02009

McKinney W., et al., 2010, Proceedings of the 9th Python in Science Conference, p. 51

Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, AJ, 129, 1993

Maritz J., Maritz E., Meintjes P., 2016, Peterson S., Yacoob S., eds, The Proceedings of SAIP2016, the 61st Annual Conference of the South African Institute of Physics, p. 243

Martín J., Torres D. F., Pedaletti G., 2016, MNRAS, 459, 3868

Naidoo K., Whiteway L., Massara E., Gualdi D., Lahav O., Viel M., Gil-Marín H., Font-Ribera A., 2020, MNRAS, 491, 1709

Nešetřil J., Milková E., Nešetrilová H., 2001, Discrete Math., 233, 3

Papitto A., Bhattacharyya S., 2022, Millisecond Pulsars. Springer, Netherlands

Pearson K., 1901, London Edinburgh Dublin Phil. Mag. J. Sci., 2, 559

Pereyra L. A., Sgró M. A., Merchán M. E., Stasyszyn F. A., Paz D. J., 2020, MNRAS, 499, 4876

Pitkin M., 2018, J. Open Source Softw., 3, 538

Pozzi F., Di Matteo T., Aste T., 2013, Sci. Rep., 3, 1665

Prim R. C., 1957, Bell Syst. Tech. J., 36, 1389

Roughgarden T., 2019, Algorithms Illuminated (Part 3): Greedy Algorithms and Dynamic Programming. Soundlikeyourself Publishing LLC, New York, The United States of America

Sánchez N., Alfaro E. J., López-Martínez F., 2018, MNRAS, 475, 4122

Shlens J., 2014, preprint (arXiv:1404.1100)

Simon O. B., Buard I., Rojas D. C., Holden S. K., Kluger B. M., Ghosh D., 2021, Sci. Rep., 11, 19704

Smith D. A. et al., 2019, ApJ, 871, 78

Tarjan R. E., 1983, Data Structures and Network Algorithms. Society for Industrial and Applied Mathematics,Philadelphia, The United States of America

The Bokeh team, 2022, Bokeh Documentation

The Graphviz team, 2022, Graphviz Documentation

Van Rossum G., Drake F. L., Jr, 1995, Python Reference Manual. Centrum voor Wiskunde en Informatica, Amsterdam

Viganò D., Rea N., Pons J. A., Perna R., Aguilera D. N., Miralles J. A., 2013, MNRAS, 434, 123

Virtanen P. et al., 2020, Nat. Methods, 17, 261

Wilson R. J., 2010, Introduction to Graph Theory. Pearson, The United Kingdom

Winther R. G., 2018, Phylogenetic Inference, Selection Theory, and History of Science Selected Papers of A. W. F. Edwards with Commentaries. Cambridge Univ. Press, Cambridge

APPENDIX A: PRINCIPAL COMPONENT ANALYSIS

The principal component analysis (PCA) is a technique used to reduce the number of variables of a problem without having a

[2] http://www.pulsartree.ice.csic.es.

significant loss of information; see, e.g. Shlens (2014) for a review. PCA is based on the search for existing correlations between the original variables involved, so as to find a new system to represent the data. The axes of this new system will be vectors that are linearly independent of each other, i.e. geometrically orthogonal, and oriented in the direction where they are able to encompass the largest possible variance of the processed data. This linear transformation is based on the eigenvector decomposition of the covariance matrix C of the variables involved:

$$C = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{X})(X_i - \hat{X})^T, \tag{A1}$$

where X is the data set in a matrix form, taking into account that $X \in R^{n \times N}$ and $C \in R^{N \times N}$, n the size of the data set, N is the number of variables describing each member of the data set, and $\hat{X}$ is the mean value of the variable. Furthermore, C will by definition be a symmetric and positive semi-definite matrix that contains on its main diagonal the variance of the variables used. The eigenvector decomposition can then be done as

$$C\boldsymbol{v} = \lambda\boldsymbol{v}, \tag{A2}$$

where the eigenvectors $\boldsymbol{v}$ found from the matrix C together with their eigenvalues λ allow one to perform the above transformation from the original set of variables to the set of principal components (hereafter PCs). In this way, the eigenvectors $\boldsymbol{v}$ are the new axes on which the new dimensional space will be built, which have the intrinsic property of maximizing the variance of the treated data. Each $\boldsymbol{v}$ will have an associated λ according to equation (A2). The eigenvalues serve to measure in which direction the data are most dispersed. Thus, it is useful to use the ratio of each λ to the sum total of all eigenvalues to define what is known as explained variance. Said otherwise, the explained variance ratio is the percentage of variance that is associated with each of the PCs. The larger λ is, the greater the explained variance covered by the resulting $\boldsymbol{v}$ will be. The cumulative explained variance is the sum of each explained variance. The new space may at most be N-dimensional, when all PCs have a non-null contribution to the variance of the population. This final set of PCs will be obtained by relating the original set of variables to the $\boldsymbol{v}$ that have been defined. This linear transformation follows

$$PC_m = \boldsymbol{v}_m X^T, \tag{A3}$$

where X^T is the transpose matrix denoted in equation (A1). In this way, as $\boldsymbol{v}$ has dimension ($1 \times N$) and X^T has dimension ($N \times n$), the result is a row vector of size ($1 \times n$) that will contain the PC_m values for each member of the space.

A1 Algebra with PC_1 and PC_2

The variables used depend on powers of P and $\dot{P}$, and therefore once the logarithm is taken, the expressions defining each variable are linear. For example, $\log B_s = K + \log P + \log \dot{P}$, where K is a constant (i.e. the equation of a plane, $Ax + By + Cz + D = 0$). The eigenvalues associated with the PCs beyond PC_1 and PC_2 will be strictly null and therefore all the explained variance is accumulated in the first two PCs.

Equations (2) and (3) are expressed as a function of the normalization of the logarithm of the variables, $(V_l)^\dagger$, where the sub-index l indicates the logarithm of the generic variable V. Here, given that any of the variables we are considering can be written as $V(P, \dot{P}) = K P^a \dot{P}^b$, $V_l = \log V = K_l + a P_l + b \dot{P}_l$, the normalization is

$$(V_l)^\dagger = \frac{V_l - Q_{2,V_l}}{IQR_{V_l}} = \frac{K_l + aP_l + b\dot{P}_l - Q_{2,V_l}}{IQR_{V_l}}. \tag{A4}$$

Once equation (A4) is applied to all the variables, a summation will be obtained, which leads to equations (4) and (5).

We now consider Fig. 3, where we show how a circle in the $P\dot{P}$-diagram would look in the PC plane. The simulation of points located along a circle in the P, $\dot{P}$-diagram, defined by its logarithmic coordinates (P_l, $\dot{P}_l$), can be done by considering a fixed value $r^2 = \Delta P_l^2 + \Delta \dot{P}_l^2$, where Δ is used to represent that the circle can be displaced from the origin. Introducing equations (4) and (5) into it, we obtain

$$r^2 = 0.702\Delta PC_1^2 + 0.148\Delta PC_2^2 - 0.363\Delta PC_1 \Delta PC_2, \tag{A5}$$

which corresponds to an ellipse (i.e. $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$) with an angle of rotation $\cot(2\theta) = (A - C)/B$). Starting counter-clockwise from the positive axis of PC_1, this leads to $\theta = 73\overset{\circ}{.}38$.

APPENDIX B: BASICS OF GRAPH THEORY

B1 Basic definitions for building a weighted graph

The aim of graph theory is to establish relationships among objects based on the connections they have, ultimately representing them in a graph. $G(V, E)$ will denote a graph of a set V, of nodes v; and a set E, of edges e joining these nodes. The latter are assigned with a weight w (e.g. a distance value) representing the relationship between the nodes (see e.g. Wilson 2010). These edges will not have any privileged addresses, and will therefore be treated as undirected connections. To assign the value of w for a given edge, the simplest possibility is to assume the weight to be the Euclidean distance between the nodes in an N-dimensional space of interest (this is usually called an Euclidean graph),

$$d_{nm} = \sqrt{\sum_{j=1}^{N}\sum_{n=1}^{V}\sum_{m>n}^{V}(v_{jn} - v_{jm})^2}, \tag{B1}$$

where d_{nm} represents the distance between variables that define each node individually.[3] The total weight of a graph G will be obtained from the sum of each specific weight on each edge,

$$w(G) = \sum_{e \in E(G)} w(e). \tag{B2}$$

Note that $e = \{v_n, v_m\} = \{v_m, v_n\}$. Moreover, G has an edge between every pair of nodes, which makes it a complete (equivalent to fully connected) graph as well as undirected and weighted due to the edge characteristics described above. A path is a sequence of nodes through G in which no node is encountered more than once. Then G is a connected graph where any pair of nodes can communicate through a path. If a path is closed, it is called a cycle. In addition, the partition of V into two sets (S, $V - S$) is called a cut. The set containing the edges $e = \{v_n, v_m\}$ whose $v_n \in$ S and $v_m \in$ V-S is defined as the cut set. The nodes of each graph can be classified according to the number of edges that connect to it, what is also known as the node degree $g(v)$. In turn, this value is the same as the number of adjacent nodes connected via incident edges to the node

[3] As usual, a normalization process is adopted when there is the need to compare magnitudes with different units.

in question. The adjacency matrix is the backend of the graph. This matrix has dimensions ($|V| \times |V|$) where each *nm*-th element of this matrix indicates whether the nodes $e = \{v_n, v_m\}$ are adjacent. The value used for signalling the existence of adjacent nodes can either be 1 or $w(e)$ corresponding to the edge defined by both nodes. Note that the main diagonal of this matrix will be null as the graph *G* does not contain loops (i.e. it does not contain cycles with just one node). The adjacency matrix can be used to compare two graphs quantitatively.

B2 The minimum spanning tree (MST)

$T(V, E^{'})$ is a subgraph of $G(V, E)$, i.e. one whose nodes and edges come from the sets of nodes and edges of G, when $V \subseteq V$ – in our case the same – and $E^{'} \subseteq E$. Furthermore, we will say that *T* is a spanning tree of *G* when it does not contain cycles and connects all nodes of *G*. The following statements regarding *T* are thus equivalent (see e.g. Tarjan 1983):

(i) *T* is a spanning tree of *G*.
(ii) *T* contains no cycles and is connected.
(iii) *T* is connected and $E^{'}$ contains $|V| - 1$ edges.
(iv) *T* has no cycles and $E^{'}$ contains $|V| - 1$ edges.
(v) *T* is minimally connected: there is only one path to connect any two vertices of *T*, so by removing any edge *T* it would be disconnected.
(vi) *T* is maximally acyclic: adding an edge to *T* would form a cycle.

The MST is a sub-graph of *G* such that the sum of all weights, $\sum_{(e) \in T} w(e)$, is the smallest possible. In this way, the MST connects all the nodes v taking into account the minimum distance between them at a local level, but under the condition that there is a global minimization for connecting the whole population of nodes. The solution of this kind of problem is obtain via greedy algorithms, which search the best possible solution in each iteration (Roughgarden 2019). Using $G(V, E)$, such an algorithm joins $T \cup e_i$ (where e_i as the edge with the smallest possible weight contained in *E*), at every step, unless the addition of this causes a cycle (see e.g. Wilson 2010). Here we use a special case of the greedy algorithm, called Kruskal's algorithm (see e.g. (Kruskal 1956)) for getting an MST out of the complete *G*.To implement this algorithm, the following steps must be carried out:

(i) Order the edges by increasing weight.
(ii) Choose the edge with the smallest $w(e)$ and add it to $E^{'}$.
(iii) Make sure that the chosen *e* does not produce any cycle in the structure of *T*.
(iv) The process finishes when all nodes *V* are connected, thus resulting in a graph $T(V, E^{'}, w^{'})$, where $|E^{'}| = |V| - 1$ and $w^{'}(T)$ is the weight of *T* according to equation (B2)

The description of this algorithm would follow the pseudocode shown (see e.g. Erickson 2019):

Algorithm 1 Kruskal $G(V, E, w)$

```
Require: sort $E$ in increasing order according to $w(e)$
T ← $(V, \varnothing)$
for each $v \in V$ do
    $MakeSet(v)$
end for
for i ← 1 to $|E|$ do
    $e_i = \{v_n, v_m\} \leftarrow e_i$ the lowest edge in $E$
    if $Find(v_n) \neq Find(v_m)$ then
        $Union(v_n, v_m)$
        $T \leftarrow T \cup e_i$
    end if
end for
return $T$
```

This description is supported by the implementation of the algorithm based on a union data structure, which operates on disjoint subsets of *V*, having the ability to support *MakeSet*, *Find*, and *Union* operations. Note that *MakeSet* generates a subset for each node v. On the other hand, the *Find* operation returns an identifier for the subset to which v belongs. Finally, *Union* decreases the number of subsets by merging those containing v_n and v_m.

To exemplify the structures of *G* and *T* as well as the complexity of working with a high number of nodes, Fig. B1 shows a worked example of a complete, undirected and weighted graph *G* with 158 nodes and 12 403 edges (i.e. 158 × 157/2), using as weights the values of the Euclidean distances between nodes according to equation (B1). It can be seen that the number of edges is a quickly growing function of the number of nodes, and so are the computational time requirements. For example, the computational time to calculate the MST with eight variables is about 288 s. The right-hand panel of Fig. B1 shows *T*, the MST of *G*, where we found only 157 edges in addition to the 158 nodes.

B3 MST properties

Below we list properties associated with the MST (proofs can be found in graph theory books; e.g. Tarjan 1983; Kleinberg & Tardos 2005; Erickson 2019). Considering that $T(V, E^{'})$ is the MST of *G*(*V*, *E*), the main properties are as follows:

(i) Uniqueness: If all $w(e)$ of $E(G)$ are distinct values, the resulting MST will be unique. This implies that the choice of any other greedy algorithm instead of Kuskal's as a solution to search for the MST (e.g. Prim's algorithm Prim 1957) would give the same results.

(ii) Cycle property: From the construction of *T* itself, any edge $e \in E$ such that $E^{'} \cup \{e\}$ creates a cycle *C* in *T*. On the other hand, if for some $e_{\rm c} \in C$ it is determined that $T(V, E^{'} \cup \{e\} - \{e_{\rm c}\})$ is a spanning tree. Thus, if *T* is an MST, then $w(e_{\rm c}) > w(e)$; otherwise, the edge with the largest weight contained in *C* will not be contained in $E^{'}$ for *T* to remain an MST. Therefore, every node $v \in T$ has among its incident edges the *e* with the smallest value $w(e)$.

(iii) Cut property: From the construction of *T* itself, any edge $e \in E^{'}$ such that $T(V, E^{'} - \{e\})$ will cause a cut in *T* separating it into two connected components. For the resulting cut set, denoted as *D*, it is observed that for any $e_{\rm D} \in D$ such that $T(V, E^{'} - \{e\} \cup \{e_{\rm D}\})$ is a spanning tree. Thus, if *T* is an MST, then $w(e_{\rm D}) > w(e)$, otherwise, the lightest edge contained in *D* must be contained in $E^{'}$ for *T* to remain an MST.

B4 Nearness in the context of different distances

Nearness between two pulsars depends on the definition of distance, and of the variables considered to compute it, thus the more complete this distance is, the better. One way to see this is by comparing the distributions of the weights associated to a complete, undirected and weighted graph using only P, $\dot{P}$ with those obtained with the full set of eight magnitudes considered in the text (equivalently, their two PCs). This is shown in Fig. B2, and relates to the discussion of the bottom panels of Fig. 3. The distributions are different. In addition, just from the full set of weights $w(G)$, it is possible to know which would be the nearest neighbours of any randomly chosen pulsar. It can be seen that neighbouring pulsars differ according to

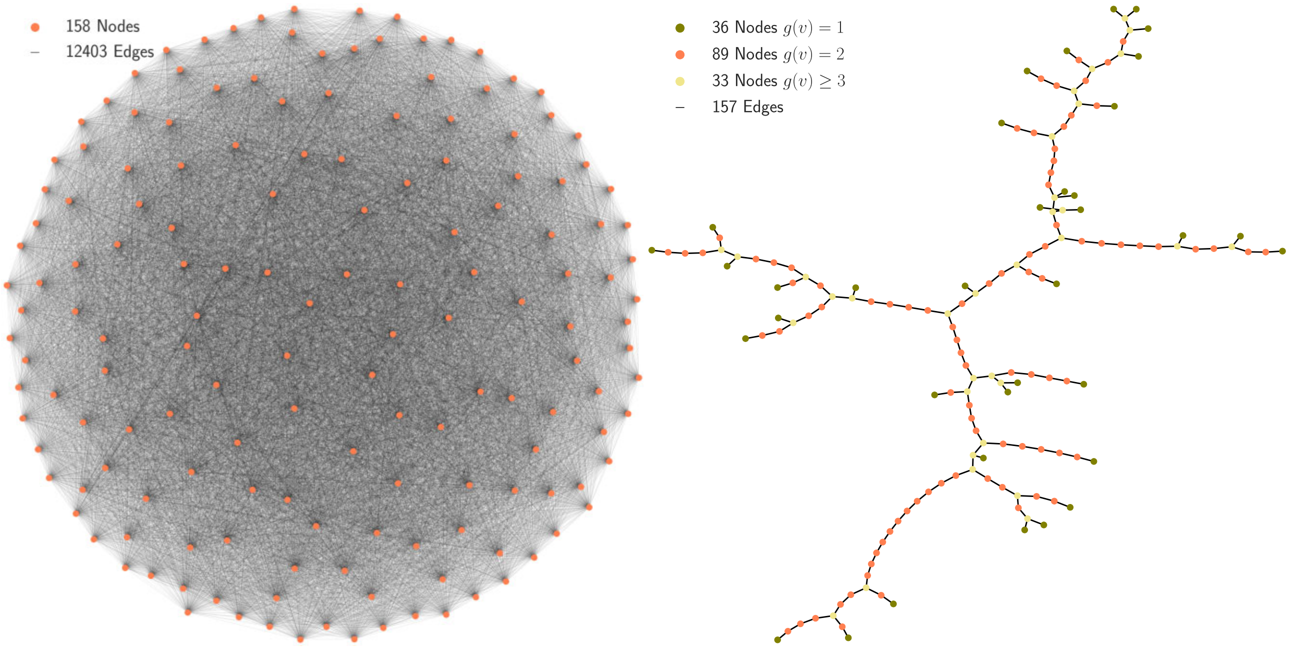

Figure B1. Left-hand panel: complete, undirected, and weighted graph *G*(158, 12403). Right-hand panel: corresponding MST *T*(158, 157), from the graph *G*(158, 12403) shown in the left-hand panel. The nodes are noted according to the degree: $g(v) = 1$ – olive, $g(v) = 2$ – coral, and $g(v) \geq 3$ – yellow.

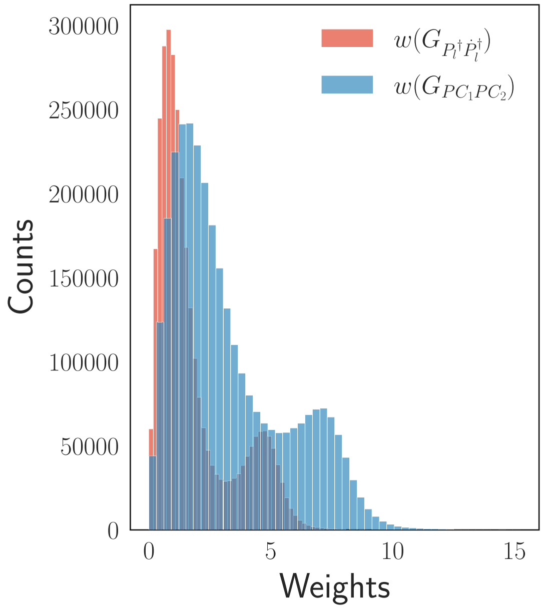

Figure B2. Distributions of the weights associated with the complete graph for two different distance definitions. We show in red distribution of distances based on the logarithms of the variables *P* and $\dot{P}$ only – once they have been normalized; and in blue, the distribution of distances based on PC_1 and PC_2.

each distance definition. Considering the nearest three neighbours for every pulsar using a distance based on P, $\dot{P}$ and comparing with the ones obtained using the full set of variables of interest, we find that 45 per cent of the population incorporates a new pulsar even in the first three places of the ranking. The ranking positions, even when the same pulsars are concerned, may change in many cases: 40 per cent of the population has at most one pulsar in the same position.

B5 Visual nearness in the MST and ranking

Due to MST properties, we know that the first neighbour in the distance ranking from a given source is indeed one of the attached nodes in the MST. However, the second neighbour and beyond in the distance ranking do not necessarily reflect vis-a-vis the visual appearance of the MST. Thus, visual nearness from a given source in the MST is not a one-to-one overall translation of the distance ranking from that source. The MST does not minimize the distance to a given node only, but the total distance needed to visit the whole population. This introduces a global perspective in the selection of how particular sources are connected to the rest, which is what ultimately serves for clustering analysis techniques.

B6 Representing the MST: how rendering is done

Given that an MST lacks a fixed set of axes, the rendering of the MST (angles among branches, orientation, branch direction) does not hold any particular meaning, only the connections among nodes do. A different representation could be chosen conserving the same properties (e.g. each node being linked to the same neighbours, being of the same order, preserving the same sequence of all variables along all branches, etc.). The nearness and connections described in one MST would be the same as in the other, despite the overall appearance differing. An obvious example is to take one of the branches that go rightwards in our MST and force it to go leftwards. This would produce exactly the same MST (the same adjacency matrix, in technical terms), and everything we have said looking to the MST with the same branch pointing rightwards would apply for the new rendering as well. It is not the appearance of the overall graph what is important, but the graph properties. In our representation, we are using the NEATO program that uses the Graphviz python library (The graphviz team 2022), which searches the minimization of a pseudo-energy to select the orientation of the different components. We have verified using earlier versions of the ATNF catalogue how the global appearance of the MST changes whenever a significant number of nodes are added (usually, the rendering does not change when adding a few nodes). Even when the visual appearance of the tree may differ, all similarity properties (e.g. the relative localization of members of sub-populations) and physical connections described are maintained.

B7 Code implementation

Our code is built on PYTHON v 3.10.4 (Van Rossum & Drake 1995), in which we implement the Psrqpy package (Pitkin 2018) to deal with the population of pulsars from the ATNF catalogue. To create the graphs, we use the NetworkX (Hagberg, Schult & Swart 2008) and the Graphviz (The graphviz team 2022) libraries. On the other hand, the SCIPY library (Virtanen et al. 2020) contains Kruskal's algorithm according to which we have been able to calculate the MST. For the application of the PCA, we used the Scikit-learn library (Buitinck et al. 2013). The previous libraries and packages contain as requirements other well-known libraries such as NUMPY (Harris et al. 2020), PANDAS (McKinney et al. 2010), and MATPLOTLIB (Hunter 2007) through which, in addition, we have been able to develop those parts of the code that were necessary to obtain the results seen. The app is done using BOKEH (The Bokeh team 2022).

This paper has been typeset from a TEX/LATEX file prepared by the author.