UAB
Universitat Autònoma de Barcelona

UNIVERSITAT AUTÒNOMA DE BARCELONA
DEPARTAMENT DE FÍSICA

UAB
Universitat Autònoma
de Barcelona

DOCTORAL THESIS

Application and development of advanced mathematical tools for population and time series analysis in pulsar astrophysics

Author:
Carlos R. García

Advisors:
Prof. Diego F. Torres
Dr. Alessandro Patruno

Tutor:
Prof. Lluís Font Guiteras

A thesis submitted for the degree of
Doctor of Philosophy in Physics

Institute of Space Sciences (ICE-CSIC/IEEC)

"Para mi mamá y mi abuela."

Abstract

In this thesis, we introduce novel methods for analyzing pulsar populations using a variety of mathematical techniques.

These tools—particularly graph theory—have been thoroughly validated in advanced mathematics, enabling us to overcome some of the constraints (even dimensional) inherent in conventional visualization approaches.

This exploration benefits from dimensionality reduction techniques, which not only lessen computational demands but also highlight potential for describing physical characteristics. The resulting structures encode information about pulsar similarities that extend beyond standard spin parameters, revealing relationships that are not readily apparent in traditional diagrams. With a physically motivated topological perspective, we leverage the strengths of these methods and present results that span from prospective source classification and the emergence of new classes to catalog comparison, among other applications.

This new approach enables fresh interpretations of longstanding problems, laying a new foundation for visualizing the pulsar population and categorizing sources.

Building on this, we identify several sources as likely members of specific binary subclasses and investigate the potential transitional nature of others. We also introduce an algorithm capable of quickly determining the position of newly discovered pulsars within the resulting structure, which we apply to known transitional candidates and accreting systems. By evaluating their positions under reasonable assumptions about currently unknown parameters, we demonstrate that this method can effectively narrow parameter spaces, guiding future targeted observations and classification efforts.

Furthermore, we extend the use of graph theory to the boundary of machine learning, demonstrating its capability for binary separation in an unsupervised context.

We demonstrate how these topological structures offer transparent and efficient representations in high-dimensional spaces, thereby avoiding opaque, black-box models that render interpretations difficult or questionable.

Finally, we introduce and apply an innovative, flexible time-series alignment technique—adapted from applications in economics, medicine, and speech recognition—to the field of gamma-ray astrophysics. We delve into he mathematical framework behind dynamic time warping and apply it to the study of gamma-ray pulsar emission processes from the perspective of their light curve morphological similarity.

The method identifies notable similarities among light curves, unaffected by their flux levels, and frequently irrespective of their timing, physical, or spectral properties. Very different pulsars can produce extremely similar light curves.

The results presented here are promising, offering a refreshing direction for the field and new pathways for rigorous mathematical analysis, ultimately providing meaningful alternatives to traditional approaches in high-energy astrophysics.

iv

This thesis is based on the following publications:

- García, Torres, and Patruno (2022)
  Visualizing the pulsar population using graph theory
  Monthly Notices Of The Royal Astronomical Society (MNRAS), 515, 3, 3883 - 3897

- García and Torres (2023)
  Quantitative determination of minimum spanning tree structures: using the pulsar tree for analyzing the appearance of new classes of pulsars
  Monthly Notices Of The Royal Astronomical Society (MNRAS), 520, 1, 599 - 610

- García, Illiano, Torres, Papitto, Coti Zelati, de Martino, and Patruno (2024a)
  Millisecond pulsars phenomenology under the light of graph theory
  Astronomy and Astrophysics (A&A), 692, A187

- García, Torres, Zhu-Ge, and Zhang (2024b)
  Separating repeating fast radio bursts using the minimum spanning tree as an unsupervised methodology
  Astrophysical Journal (ApJ), 977, 273

- García and Torres (2025)
  Quantitative exploration of the similarity of gamma-ray pulsar light curves
  Astrophysical Journal Letters (ApJL), 982, L51

  and briefly commented here, also:

- Zhang, Torres, García, Li, and Mestre (2024)
  Analysis of the possible detection of the pulsar wind nebulae of PSR J1208-6238, J1341-6220, J1838-0537 and J1844-0346
  Astronomy and Astrophysics (A&A), 691, A332.

Acknowledgements

I want to thank Diego, who has not only been my thesis director but also someone I turn to for discussion, debate, and, above all, learning. Nothing in this thesis would have been possible without him. I also thank Alessandro for the opportunity he gave me to embark on this path, for all the advice he provided to keep me going, and most of all, for the trust he placed in me. I am grateful to Dani for all his efforts to make this journey more comfortable for me, and for all the trips we've taken to other countries, where he has been an unbeatable companion. Thanks to the entire MAP group; it has been truly inspiring to work so closely with such highly skilled people. Thanks to everyone at ICE-CSIC for creating a motivating, healthy, and supportive environment for personal growth. Thanks to my friends for being my anchor and helping me stay awake. Thank you, Mom, for holding my hand through it all. I would like to thank my family for being my home from afar, for understanding each of my absences, and for accompanying me throughout this journey. Thank you for being there, for being here, for being there, for being here, and when it was not possible, at least it seemed like it. Thank you for being my vertigo. "I've lost my mind on this game like Vincent van Gogh dedicated his life to his art, and he lost his mind in the process, that's happened to me, but foock it..it will pay, I'll be happy"-C.M.

Contents

List of Figures

List of Tables

Chapter 1

Pulsars

1.1 Introduction

The idea that neutron stars might exist preceded their observational confirmation by several decades. Baade and Zwicky made the first proposal for the existence of neutron stars in 1934 (Baade and Zwicky, 1934), an idea already suggested by Landau (1938): the existence of stars composed mainly of neutrons[1]. The former work suggested that a neutron star may be the final stage of a normal star's collapse following a supernova explosion. After that, Tolman (1939) and Oppenheimer and Volkoff (1939) independently performed the first calculations of neutron star models and analyzed the structure based on a star composed of degenerate neutron gas. This approach showed that the degeneration was such that temperature was irrelevant and that everything depended on the relationship between density and pressure. Therefore, the focus of attention from that moment on was on determining this relationship, known as the Equation of State (EoS) (Lyne and Graham-Smith, 2012).

In the following years, Harrisson, Wakano, and Wheeler 1958[2], Cameron (1959), Ambartsumyan and Saakyan (1960), and Hamada and Salpeter (1961) conducted research on EoS and neutron star models (see Harrison et al., 1965 for extensive discussions of these models). However, interest in these objects remained low, primarily among astronomers, as work on neutron stars had focused mainly on understanding how the cores of massive stars, composed solely of neutrons, could be a source of stellar energy. In addition, observing neutron stars with the optical telescopes of the time was complicated due to their small size and residual thermal energy, which meant that even less attention was paid to them (Shapiro and Teukolsky, 1983).

This changed in 1967 when the PhD student Jocelyn Bell Burnell (see Bell Burnell, 1977 for her role in the discovery), who was part of the group of astronomers led by Antony Hewish[3], detected the first pulsar. The find originated from data collected at the Mullard Radio Astronomy Observatory, where an unusual, stable radio signal in the form of pulses was observed. In an initial explanation, they proposed white dwarfs or neutron stars as the sources of these unusual stable oscillations[4] (Hewish

---

[1]In Shapiro and Teukolsky (1983), Yakovlev et al. (2013), and Haensel, Potekhin, and Yakovlev (2007), based mainly on to recollections of Rosenfeld (1974, Proc. 16th Solvay Conference on Physics, p. 174), note that Lev Landau in (Phys. Z. Sowjetunion, 1, 285, 1932) already suggests the possibility of stable structures within stars composed of neutrons (see also Landau, 1979; Brecher, 1999), as credited in Cameron, 1970), after the discovery of the neutron by Chadwick (1932).

[2]Wheeler, J. A. 1958, paper read at the Solvay Conference, Brussels. Published with B. K. Harrison and M. Wakano, in La Structure et la Révolution de l'Univers (Stoops, Brussels, 3100 pp., 1959).

[3]Antony Hewish received the Nobel Prize in 1974 for his role in the pulsars' discovery.

[4]Recorded as CP1919, the signals observed were so surprising that they persuaded astronomers to consider that an extraterrestrial civilization was responsible for them (initially called *Little Green Men* or "LGM"), see Bell Burnell (1977) and Penny (2013).

et al., 1968).

This discovery contributed to research proposing the existence of stable stars denser than white dwarfs, which would be rapidly rotating and have strong magnetic fields (Hoyle et al., 1964; Woltjer, 1964; Tsuruta and Cameron, 1966). It had even been conjectured that the energy source seen in the Crab nebula[5] came from this type of object (Wheeler, 1966), see Figure 1.2.

In the following years, both the Cambridge team and the National Radio Astronomy Observatory groups at Green Bank (West Virginia) and Arecibo (Puerto Rico), two of its facilities, detected more pulsars, which helped determine that these stable oscillations with regular patterns were due to rapidly rotating neutron stars. Gold (1968) then explained that the pulses observed were rapidly rotating neutron stars with a strong magnetic field, which would generate relativistic velocities in the plasma surrounding their magnetosphere. Pacini (1967) and Pacini (1968), in turn, explained how a simple dipolar magnetic model could be capable of carrying out this process, connecting it with the activity seen in the Crab Nebula. This leads to the simple model that a rapidly rotating neutron star would behave like a rotating beacon, or a lighthouse, generating pulsed signals with a stable pattern, which is why the name "pulsar" was coined. Subsequent observations and studies of the energy source of the Crab Nebula (Staelin and Reifenstein, 1968; Comella et al., 1969; Cocke, Disney, and Taylor, 1969) as well as the slowdown of what would be the Crab pulsar, showed that the energy lost by the latter would be the energy used to activate the Crab Nebula (Richards and Comella, 1969). Similarly, the observed Vela supernova remnant (Large, Vaughan, and Mills, 1968), together with the detection of the Vela pulsar at its center, offered one of the earliest observational links between neutron stars and supernova explosions. This interpretation was supported by the works of Radhakrishnan and Cooke (1969), Radhakrishnan et al. (1969), and Radhakrishnan and Manchester (1969), who collectively showed that the Vela pulsar was a rotating neutron star with a well-defined magnetic geometry, whose alignment with the remnant and energy loss properties were consistent with theoretical expectations. These findings were observational support for the hypothesis initially proposed by Baade and Zwicky, suggesting that neutron stars are the remnants left behind by supernova explosions.

Over the decades, advances in observational technology and instrumentation have enabled the discovery of ∼3800 pulsars (Manchester et al., 2005) detected through their radio emission by The Australian Telescope National Facility (ATNF[6]). Additionally, it was found that pulsars are also bright in X-rays, optical, and gamma rays. In particular, the Large Area Telescope (LAT) on board the *Fermi* Gamma-ray Space Telescope[7] has detected ∼300 pulsars released in the *Fermi*-LAT Third Pulsar Catalog (3PC, Smith et al., 2023).

Most known pulsars reside within our Galaxy and are concentrated along the Galactic plane.In addition, about ∼345 pulsars have been discovered in globular clusters[8] see e.g., Hessels et al. (2015), where stellar densities could favor binary evolution.

In this regard, it should be noted that while most known pulsars are isolated neutron stars, a significant fraction reside in binary systems (see Postnov and Yungelson,

---

[5]The Crab Nebula is widely accepted as the remnant of a supernova explosion observed by Chinese astronomers in 1054 AD, who referred to this event as a "guest star". Its connection with the Crab Nebula was first proposed by Duyvendak (1942), Mayall and Oort (1942), Baade (1942), and Minkowski (1942).

[6]`https://www.atnf.csiro.au/research/pulsar/psrcat/`

[7]`https://fermi.gsfc.nasa.gov/`

[8]A comprehensive and regularly updated list of pulsars in globular clusters: `https://www3.mpifr-bonn.mpg.de/staff/pfreire/GCpsr.html`
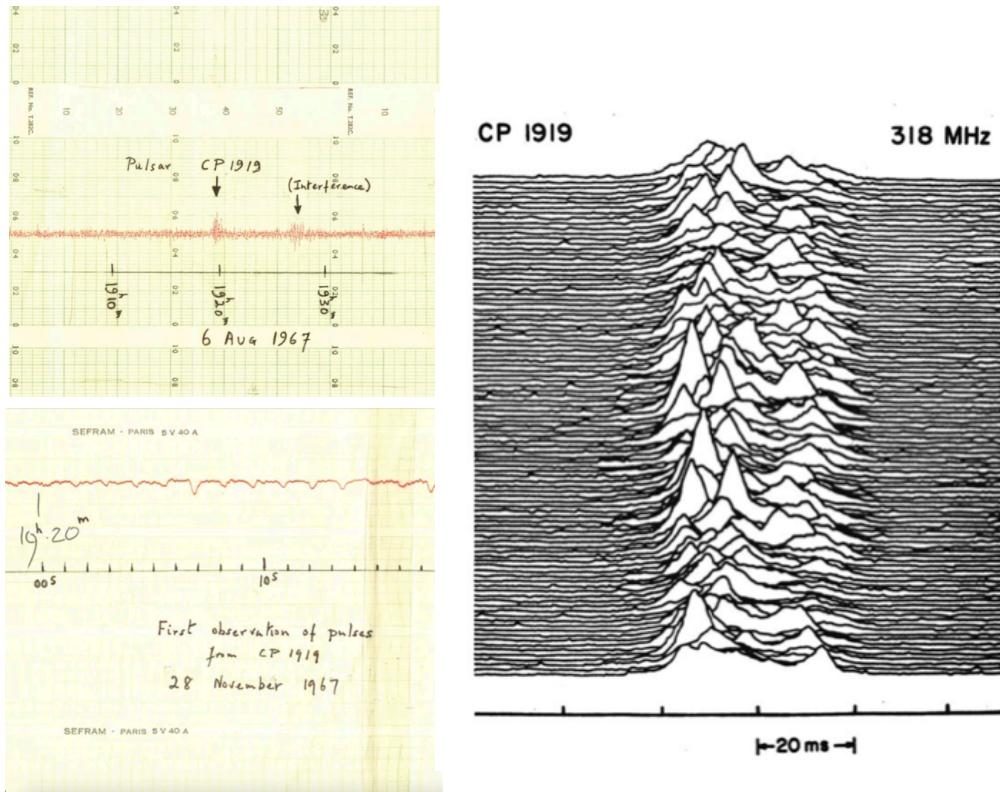
FIGURE 1.1: First column: shows the actual first detection charts taken from Jocelyn Bell (figure adopted from https://www.cam.ac.uk/stories/journeysofdiscovery-pulsars). The first panel displays the initial record of CP1919, dated August 6, 1967, in which the anomalous regular pulse was observed. The second panel presents the record dated November 28, 1967, which consolidated the previous observations and, as a result, confirmed the detection of the first pulsar: PSR B1919+21. Second column: Stacked plot on CP1919, also known as a plot illustrating successive mean pulse spectra, as referenced in Harold Dumont Craft Jr, who titled it as "Many consecutive pulses from CP1919" seen in his PhD dissertation "Radio observations of the pulse profiles and dispersion measures of twelve pulsars", Cornell University, 1970, p. 214 (figure adopted from (Turchetti, 2023), see this publication for more information).
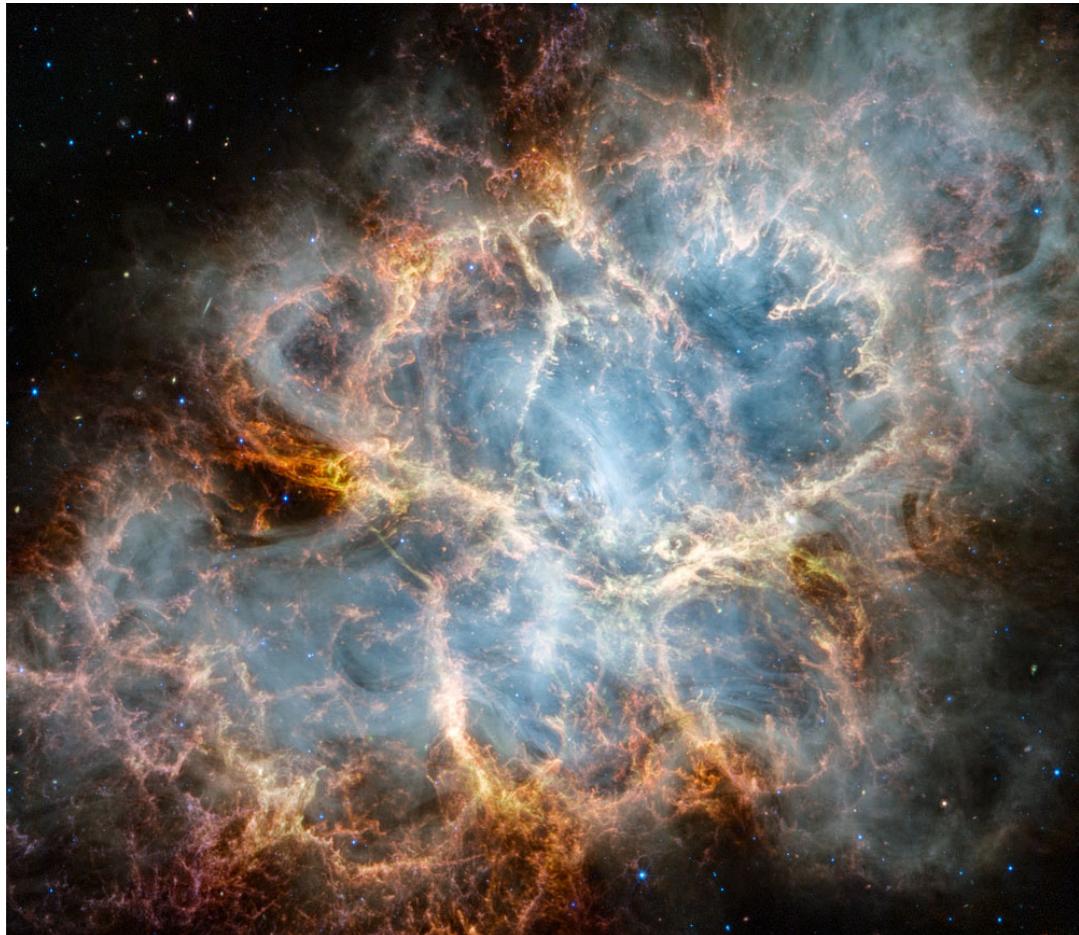
FIGURE 1.2: The Crab Nebula as seen by the James Webb Space Telescope in infrared light. At the center of this ring-shaped structure is a bright white spot: the Crab Pulsar (Credits: NASA, ESA, CSA, STScI, Jeff Hester (ASU), Allison Loll (ASU), Tea Temim (Princeton University)).

[2014](#) for an extensive review). The first radio pulsar in a binary system[9] was followed by Hulse and Taylor ([1975](#))[10] using the Arecibo radio telescope. Before 1974, neutron stars in X-ray binaries (such as Sco X-1 Giacconi et al., [1962](#)[11], Her X-1 Tananbaum et al., [1972](#), and Cen X-3 Schreier et al., [1972](#)) had already been discovered. These systems supported the idea that a neutron star could accrete matter from a companion, leading to high-energy emissions across the X-ray and gamma-ray regimes. These discoveries provided crucial observational support for binary evolution models of neutron stars, and established the foundation for the development of the recycling scenario, in which neutron stars in close binaries are spun up through accretion of material from the companion to form millisecond pulsars (MSPs) (Alpar et al., [1982](#); Bhattacharya and van den Heuvel, [1991](#)). The MSPs constitute a distinct subpopulation characterized by rapid rotation and low magnetic fields. MSPs offer unique insights into stellar evolution, the interaction between magnetic fields and plasma transferred by the donor star, and particle acceleration from compact objects, particularly in binary systems (see e.g., Manchester [2017](#); Papitto and Bhattacharyya [2022](#) for reviews).

Throughout this Chapter, we will show the formation and phenomenological processes through observed and derived properties that define the population of pulsars as a whole, and MSPs in particular. We will show the evolution of binary systems and the relevance of their classification based on their behavior. Additionally, we will briefly introduce Fast Radio Bursts, which are events of uncertain origin but are increasingly linked to compact objects.

1.1.1 Compact Object Formation

To describe what a pulsar is and its characteristics, we must revisit its origin, i.e., its progenitor. When a normal star consumes most of its nuclear fuel, it is considered to have reached the end of its stellar evolution and therefore dies as we know it. At that moment, what is known as a compact object is born: a white dwarf, neutron star, or black hole. Two fundamental characteristics differentiate compact objects from normal stars (Shapiro and Teukolsky, [1983](#)):

- Compact objects no longer sustain thermonuclear reactions in their cores, so the thermal pressure that previously counterbalanced gravity disappears. However, this collapse can be prevented by other processes, such as electron degeneracy pressure in the case of white dwarfs, or primarily neutron degeneracy pressure and nuclear interactions in the case of neutron stars. This does not happen in the case of black holes, which are the outcome of complete gravitational collapse, where no known force stops the contraction once a critical mass threshold is exceeded.

- Compact objects are significantly smaller in size compared to normal stars of comparable mass, leading to extreme compactness and resulting in much stronger surface gravitational fields.

The mass of a normal star is a significant factor in determining whether it will ultimately evolve into a white dwarf, a neutron star, or a black hole in its final stage.

---

[9]PSR B1913+16 was detected in July 1974 and confirmed in August as a result of observing a significant change in its spin period (Haensel, Potekhin, and Yakovlev, [2007](#)).

[10]Russell A. Hulse and Joseph H. Taylor were awarded with the Nobel prize in 1993 for this discovery.

[11]This detection was key to the model proposed by Shklovsky ([1967](#)), where the X-ray emission detected originated from a close binary system due to an accreting process.

White dwarfs[12] are believed to originate from stars with masses $10 \pm 2M_\odot$ (Althaus et al., 2010). These progenitors of white dwarfs are expected to form a planetary nebula through the controlled ejection of matter in their final stage. According to the Chandrasekhar limiting mass, once converted into a white dwarf, they can only reach a maximum of $1.4M_\odot$ (Chandrasekhar, 1931). Note that massive progenitors $> 20M_\odot$ may lead to black holes[13] (Fryer, 1999; Heger et al., 2003).

However, neutron stars are stellar remnants produced in Type II, Ib, or Ic supernova explosions, caused by the gravitational collapse of ordinary stars with initial masses between $8M_\odot$ and $25M_\odot$. A supernova explosion occurs when the progenitor star can no longer produce energy from nuclear fusion. As a result, the pressure gradient from radiation is not sufficient to balance gravitational attraction, causing the star to collapse. This event in the inner layers of the star releases gravitational energy, which is transferred to the outer layers, causing them to be expelled.

Neutron stars derive their name from the predominance of neutrons in their interior, whose degeneracy pressure prevents the neutron star from collapsing entirely. Neutron stars can have masses of 1-$2M_\odot$ and a radius of 10-12 km (Burgio and Vidaña, 2020).

This occurs due to the process of mutual elimination through β-decay between protons and electrons that they undergo. This process determines the equilibrium of the neutron fluid that comprises the central part of the star, which is $\sim 5\%$ protons and electrons. As a result, their density reaches nuclear values since they essentially function as a giant nucleus held together by self-gravity (Shapiro and Teukolsky, 1983).

Assuming the predicted onion structure (see Figure 1 and further details in Burgio and Vidaña (2020)) for the interior of a neutron star, as well as the parameters[15] typically accepted for describing it, with a radius of 10 km and a mass of $1.4M_\odot$, we observe how specific values are comparable to those of nuclear matter. For example, neutron stars reach, in their core, a density of 6.7×10^{14}g cm^{-3}, close to 2.7×10^{14}g cm^{-3} calculated for nuclear matter. Even under normal conditions, the mean number density of neutrons of a neutron star[16], 0.4 fm^{-3}, is close to the number density of neutrons and protons seen for nuclear matter of 0.16 fm^{-3}. The crust of neutron stars also appears to be remarkably formed. So much so that their outer crust, ~ 1 km thick and composed of iron near its surface, has a composition of heavy nuclei that become increasingly rich in neutrons as one descends toward its core. Assuming the number of neutrons N and the atomic number Z of normal matter are similar, and $N \sim 2Z$ for heavy nuclei, the neutron fluid has $N \gg Z$ (Lyne and Graham-Smith, 2012).

---

[12]White dwarfs, which emit primarily their residual thermal energy, are characterized by having higher effective temperatures than normal stars, despite having lower luminosity. This is mainly due to their small radius, since according to the black body relation, for an object with temperature T and radius R, its radiative flux varies with T^4, and its luminosity as $L \sim R^2 T^4$. White dwarfs can be observed directly through optical telescopes during their cooling period, which spans billions of years.

[13]Black holes do not produce any direct emissions, as they do not permit light or matter to escape their gravitational pull. Their observation is primarily indirect, based on the influence they exert on their surroundings, accretion disk emission detectable in X-ray and optical emission, and gravitational wave signals from mergers (see, e.g., LIGO/Virgo, which detected the first gravitational wave from the merger of stellar-mass binary black holes Abbott et al., 2016). The Event Horizon Telescope[14] produced the first image of a black hole shadow in the galaxy M87* (Event Horizon Telescope Collaboration et al., 2019).

[15]See Özel and Freire (2016) for a comprehensive summary of proposed EoSs, as well as observational constraints on neutron star masses and radii.

[16]1 femtometre (fm) = 10^{-15} m.
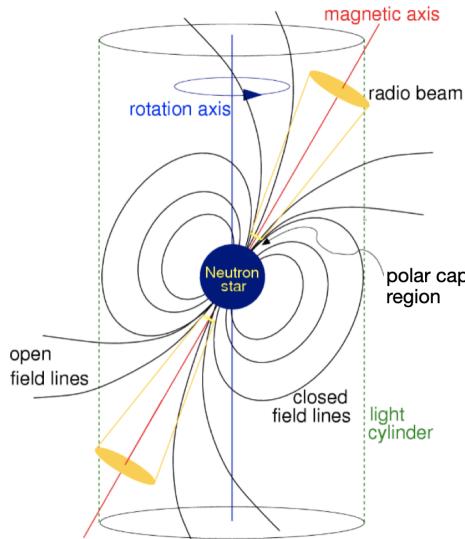
FIGURE 1.3: Scheme of the pulsar "lighthouse" model showing its
emission geometry (not drawn to scale). The rotation axis (blue) is
inclined concerning the magnetic axis (red), producing a beam (yel-
low) of radiation that sweeps across the observer's line of sight. The
description of the rest of the scheme can be found in Section 1.2.3.
(Figure adopted from Lorimer and Kramer, 2005).

Neutron stars are multi-wavelength emitters that can be observed across the entire
electromagnetic spectrum, from radio to gamma bands. More details on radio and
gamma detectability on neutron stars, which behave as pulsars, are provided in later
sections. Additionally, neutrino[17] and gravitational wave[18] observatories can also
detect these objects (Haensel, Potekhin, and Yakovlev, 2007). These fundamental
principles governing the formation of compact objects, ranging from white dwarfs
to neutron stars and black holes, offer a natural starting point for understanding how
pulsars emerge from a specific progenitor and how they differ from other possible
endpoints of stellar evolution.

1.2   Pulsar Phenomenology

As described in Section 1.1, following the discovery of the first pulsar (Hewish et
al., 1968), a large number of observations were conducted, establishing key facts
about them. Primarily, and even considering open questions based on reasonable
doubts, a basic model (Gold, 1968) has long been established, stating that pulsars
are highly magnetized neutron stars that rotate rapidly and are formed during the
supernova explosions of massive stars. The "lighthouse" model, also known as the
dipole model (Pacini, 1967; Pacini, 1968; Ostriker and Gunn, 1969), describes a ro-
tating neutron star with a misaligned magnetic axis relative to its rotation axis as
shown in Figure 1.3.

---

[17]For instance: The Sudbury Neutrino Observatory (SNO): `https://sno.phy.queensu.ca/`, and the
Ice Cube Neutrino Observatory (IceCube): `https://icecube.wisc.edu/`

[18]For instance: Laser Interferometer Gravitational-Wave Observatory (LIGO): `https://www.ligo.caltech.edu/`, European Gravitational Observatory (VIRGO detector): `https://www.virgo-gw.eu/`,
and the Large Interferometer Space Antenna (LISA): `https://lisa.nasa.gov/`.
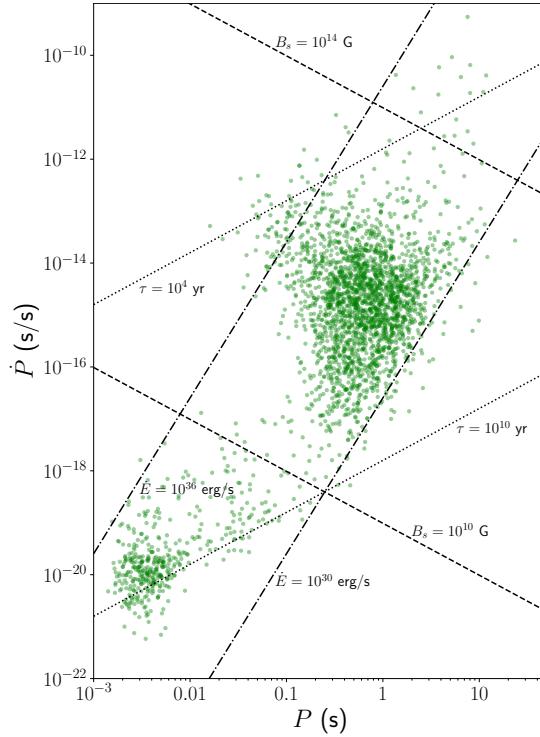
FIGURE 1.4: The $P\dot{P}$ diagram based on the the ATNF v2.6.1 restricted to $\dot{P} > 0$ resulting in 2752 pulsars. Lines of constant magnetic field at the surface (B_s) in dashed, characteristic age (τ_c) in dotted, and spin-down energy loss ($\dot{E}$) in dash-dotted are also shown.

The neutron star rotates, accelerating charged particles that are distributed along field lines to form a beam (radio beam in Figure 1.3). These accelerated particles emit electromagnetic radiation, which is typically detected in radio frequency through the observed pulses. Each pulse is produced when the radio beam, aligned with the magnetic axis, sweeps across the observer's line of sight during each rotation. The pattern seen in the pulses, defined by their repetition period, corresponds to the rotation period of the neutron star (Lorimer, 2008).

1.2.1 Spin Periods and Slowdown Rates

We will focus on the radio emission of pulsars and the properties of these that can be described through this process, adopting the formalism seen in Lorimer and Kramer (2005) and the references cited. Timing analyses in radio, through the periodicity of the arrival time of the pulses, allow us to determine the spin period (P) of pulsars and their corresponding rate of spin-down, also known as the spin period derivative ($\dot{P}$). These measurements place the population of radio-emitting pulsars on a plane called the $P\dot{P}$ diagram, which enables us to explore various characteristics that define them. Figure 1.4 shows the population of pulsars, as seen in the ATNF v2.6.1, that contains 2752 pulsars.

A remarkable differentiation comes from P, where the so-called "normal pulsars" are located from $P \sim 0.5$ s which increases at rates between $\dot{P} \sim 10^{-15}$ ss^{-1} and $\dot{P} \sim 10^{-10}$ s, while the MSPs with $P < 10$ ms and $\dot{P} \sim 10^{-20}$ ss^{-1}. Assuming a pulsar behaves as a rotating magnetic dipole, its observed P and $\dot{P}$ provide insight into key physical properties. Variations in these quantities imply differences in derived properties considering the spin evolution, such as the magnetic field at the

surface, spin-down energy loss, characteristic age, and other quantities related to the pulsar's magnetosphere (see Section 1.2.3), including the magnetic field at the light cylinder, the Goldreich–Julian charge number density, and the surface electric voltage. Note that in Figure 1.4, we restrict the sample to pulsars with $\dot{P} > 0$. This selection is essential to ensure the physical validity of the derived properties, which are based on mathematical expressions involving positive values of P and $\dot{P}$. For instance, observed negative values of $\dot{P}$ are often associated with pulsars in globular clusters, due to the acceleration in the cluster's gravity well (see Section 6.3). These values do not reliably reflect the intrinsic spin evolution of the pulsar and are therefore excluded from the analyses presented in this work. In the following sections, we examine the theoretical relationships that link these observables to the derived properties of the pulsars and explore how they influence the pulsar population, as depicted in the $P\dot{P}$ diagram. We will assume a dipole model, with $P(\text{s})$ and $\dot{P}(\text{s/s})$, where a neutron star behaves like a sphere of uniform density rotating in a vacuum, following Lorimer and Kramer (2005) and references therein. In this way, its moment of inertia $I = kMR^2$ will be calculated with the values $M = 1.4M_\odot$[19], $R = 10^6$ cm, and $k = 0.4$, resulting in the canonical value of $I = 10^{45}$ g cm^2.

1.2.2 Spin Evolution

The evolution of the pulsar's spin can be observed through the pulse period, which increases over time. Let $\dot{P} = \mathrm{d}P/\mathrm{d}t$ denote the increase in P and $E_{rot} = I\Omega^2/2$ the rotational kinetic energy, with $\Omega = 2\pi/P$ the angular frequency, we can relate the change in P to the loss of E_{rot} as:

$$\dot{E} \equiv -\frac{\mathrm{d}E_{rot}}{\mathrm{d}t} = -\frac{\mathrm{d}(I\Omega^2/2)}{\mathrm{d}t} = -I\Omega\dot{\Omega} = 4\pi^2 I\dot{P}P^{-3} . \tag{1.1}$$

The quantity $\dot{E}$ is known as the spin-down energy loss (or spin-down luminosity) and quantifies the rate at which the neutron star loses rotational energy. By substituting the canonical value of $I = 10^{45}$ g cm^2 (see Section 1.2.1) into Equation (1.1), we obtain:

$$\dot{E} \simeq 3.95 \times 10^{46} \dot{P}P^{-3} \text{ erg s}^{-1} . \tag{1.2}$$

From Equation (1.2) comes the value of $\dot{E}$ seen in the population of pulsars observed at the ATNF. If we look at Figure 1.4, the most energetic pulsars are young pulsars (see below discussion on age), unlike older ones, which are usually less active. Furthermore, not all of $\dot{E}$ is converted into radio emission; in fact, only a small part of it is obtained in that frequency band. The majority of the energy loss is emitted in the form of high-energy radiation, including X-rays and gamma rays, as well as through pulsar winds and, undetectable yet, gravitational waves (Lasky, 2015; Glampedakis and Gualtieri, 2018).

To model how a pulsar loses its rotational energy over time, the braking index n is introduced. To do this, we assume that a rotating magnetic dipole with a particular moment $|\vec{m}|$, with a separation α between its magnetic axis and its axis of rotation, emits electromagnetic radiation at a given Ω:

$$\dot{E}_{\text{dipole}} = \frac{2}{3c^3}|\vec{m}|^2\Omega^4 \sin^2\alpha . \tag{1.3}$$

---

[19]$1M_\odot$=1.989×10^{33} g.

This energy loss, seen in Equation (1.3), is assumed to be equal to the lost rotational energy described in Equation (1.1). In this way, we can evaluate the evolution of its rotational frequency over time, described by:

$$\dot{\Omega} = -\left(\frac{2|\vec{m}|^2 \sin^2 \alpha}{3Ic^3}\right)\Omega^3 .$$

(1.4)

When considering the rotational frequency $\nu = 1/P$, Equation (1.4) can be generalized as a power-law in the following form:

$$\dot{\nu} = -K\nu^n ,$$

(1.5)

where $\dot{\nu}$ is its rotational frequency over the time, and K is a constant. The n reflects the dominant mechanism by which the pulsar spins down. In other words, measuring n allows us to investigate the processes that influence the long-term spin evolution of pulsars and explore potential evolutionary connections among the different pulsar populations.

This is because n can also be measured whenever $\dot{\nu}$ can be observed, so that by differentiating Equation (1.5) and not considering K for simplifying, we have:

$$n = \nu\ddot{\nu}/\dot{\nu}^2 .$$

(1.6)

The measurement of n can be affected by timing noise and glitches in some instances. Glitches are sudden increases in the spin frequency of pulsars and are thought to offer insights into the internal structure of neutron stars.

Timing noise, on the other hand, refers to irregularities in rotational behavior that cause deviations in pulse arrival times from a steady state of spin-down. While measuring ν and $\dot{\nu}$ using standard timing models is relatively straightforward, determining $\ddot{\nu}$ is significantly more challenging due to its minimal magnitude.

This follows from Equation (1.6), which shows that accurately measuring n requires a reliable estimate of $\ddot{\nu}$. Since younger pulsars tend to spin down more rapidly, their $\ddot{\nu}$ are more pronounced, making them more favorable targets for such studies (Parthasarathy et al., 2019; Parthasarathy et al., 2020 and references therein). Archibald et al. (2016a) showed that pulsars with $n > 3$ exist, as exemplified by PSR J1640-4631, which was the first observed case where other physical mechanisms, such as magnetic quadrupoles, are essential for explaining pulsar spin-down. Also, it can be observed that there is compelling evidence for a common trend among young glitching pulsars, with several exhibiting values of $n \leq 2$ (Espinoza, Lyne, and Stappers, 2017). Furthermore, as shown in Parthasarathy et al. (2019) as well as in Parthasarathy et al. (2020), n can vary significantly across pulsars.

Estimations of $n = 5$ suggest that the pulsar behaves as a pure emitter of gravitational waves, characterized by a quadrupole moment, to explain specific observations made as the case of the high n seen in J1640-4631 (de Araujo, Coelho, and Costa, 2016).

The value of $n = 3$ corresponds to the case of pure magnetic dipole radiation in vacuum, the canonical model for pulsar spin-down proposed initially in early theoretical work (Pacini, 1968; Ostriker and Gunn, 1969), and adopted in this study. This choice is motivated by its ability to link observable quantities to physical properties of the neutron star, such as age or magnetic field strength, through simplified and interpretable expressions.

Thus, we can estimate the pulsar's age from this starting point. This can be obtain through Equation (1.5), taking $\nu = 1/P$, can be expressed as $\dot{P} = KP^{2-n}$, and integrating the latter while assuming K constant and $n \neq 1$, we have:

$$T = \frac{P}{(n-1)\dot{P}}\left[1 - \left(\frac{P_0}{P}\right)^{n-1}\right]. \tag{1.7}$$

where P_0 denotes the spin period at birth. Under the model's considerations, $n = 3$, so the dipolar magnetic radiation produces the spin-down, and $P_0 \ll P$, Equation (1.7) can be seen as:

$$\tau_c \equiv \frac{P}{2\dot{P}} \simeq 1.58 \times 10^{-8}\frac{P}{\dot{P}} \text{ yr}. \tag{1.8}$$

The quantity in Equation (1.8) is denoted as the characteristic age, and it is the value expressed in Figure 1.4. This figure shows the normal pulsars, the bulk of the population, with $\tau_c \sim 10^7$ yr, while young pulsars are those with $\tau_c \leq 10^5$ yr, and MSPs present a $\tau_c \sim 10^{10}$ yr.

Even so, it must be considered that, due to the conditions imposed on τ_c, it only serves as an estimate, especially when true ages are not accessible, for locating pulsars within a population (see Section 6.2.2 for cases where it is discarded). Thus, it can be verified that τ_c does not provide the true age of the pulsar, for example, using the Crab pulsar (see Section 1.1). According to Equation (1.8), and taking $P = 0.033$ s and $\dot{P} = 4.2 \times 10^{-13}$ ss^{-1} from the Crab pulsar, we see a $\tau_c = 1260$ yr. However, the supernova explosion of the Crab Nebula is dated to 1054 AD, so the true age of the Crab pulsar is estimated to be less than 1000 years.

Among the physical properties associated with the immediate surface of the neutron star, the magnetic field at the surface (B_s) plays a central role. Assuming a purely dipolar behavior, the magnetic field strength is related to the magnetic moment through:

$$B = \frac{|\vec{m}|}{r^3}. \tag{1.9}$$

Substituting $\Omega = 2\pi/P$ and introducing Equation (1.9) into Equation (1.4), we obtain:

$$B_s \equiv B(r = R) = \sqrt{\frac{3c^2 I P \dot{P}}{8\pi^2 R^6 \sin^2 \alpha}}. \tag{1.10}$$

Taking the canonical values of $I = 10^{45}$ g cm^2, $R = 10^6$ cm, and assuming[20] $\alpha = 90°$, by Equation (1.10) we have:

$$B_s \simeq 3.2 \times 10^{19} \sqrt{P\dot{P}} \text{ G}. \tag{1.11}$$

Equation (1.11) provides a widely used estimate of the magnetic field strength at the surface of the neutron star B_s, and its behavior is shown in Figure 1.4. We can distinguish pulsars with high $B_s \sim 10^{14}G$, usually referred to as magnetars, from normal pulsars that are around $B_s \sim 10^{12}G$, and, for instance, MSPs with $B_s \sim 10^8 G$. However, it is essential to note that this value is derived from model assumptions. Thus, the true magnetic field strength may vary from this estimation, in particular due to deviations from ideal dipole geometry, contributions from multipole fields, or

---

[20]With $\alpha = 90°$ the magnetic axis is perpendicular to the rotation axis, which equates to the maximum dipole torque.

the presence of plasma in the magnetosphere (Spitkovsky, 2006; Pétri, 2015; Gralla, Lupsasca, and Philippov, 2017).

In some cases, the magnetic field strength can be measured directly by detecting cyclotron line features in X-ray spectra (Truemper et al., 1978; Wheaton et al., 1979; Coburn et al., 2002; Tiengo et al., 2013), or even in isolated pulsars (Bignami et al., 2003; Suleimanov et al., 2010). The observed values are generally consistent in order of magnitude with B_s. Furthermore, as seen in Shapiro and Teukolsky (1983), Equation (1.11) can be interpreted as the magnetic field strength at the magnetic equator, but the field at the magnetic poles is approximately twice as strong: $B_{\text{pole}} = 2B_s$. This distinction becomes important when comparing theoretical models with observational diagnostics, particularly in studies of X-ray emission (Zavlin et al., 2000).

1.2.3   Magnetosphere

The realistic evolution of the model assumed for the rotating neutron star in a vacuum is based on the idea that it behaves like a highly magnetized, rotating conductive sphere. This is because the Lorentz forces acting on the charges inside the neutron star are stronger than the gravitational forces. Deutsch (1955) was the first to study the structure of the electromagnetic field around an isolated, rotating, magnetized neutron star in vacuum that has a magnetic axis misaligned with the rotation axis. Then, Goldreich and Julian (1969) shows that a neutron star with aligned magnetic axes and rotation generates an external electric field that extracts plasma from its surface. Although this model is an approximation with some deviation from a realistic case, it contributes to our understanding of the pulsar's behavior and facilitates its observation. Thus, the pulsar magnetosphere is the region surrounding the neutron star, filled with magnetized plasma and shaped by interactions with the intense magnetic field. The magnetosphere spans from the stellar surface to distances of several thousand kilometers and may be shaped by interactions with the interstellar medium, surrounding accretion material, or binary companions. The complexity of such an environment has led to the assumption that analytical solutions are only applicable in simple cases, providing a convincing answer. In Section 1.3, we present studies conducted on the description of the magnetosphere through models that will aid in understanding gamma emission.

To describe magnetospheric physical properties' proxies relevant to our work, we follow the treatment presented in Lorimer and Kramer (2005), see also Goldreich and Julian (1969). We start from the premise that the neutron star behaves like a perfectly conductive, rotating, and magnetized sphere. Therefore, we expect an induced electric field to be produced by the magnetic field of this rotating magnetized sphere, balanced by the electric field generated by the distribution of charges, since it is a conductive sphere. If the space outside the star were a true vacuum, these induced surface charges would generate an external electric field, which is quadrupolar and associated with a net charge of zero. This field can be described by the electric potential:

$$\Phi(r,\theta) = \frac{R^5 \Omega B_s}{6cr^3}(3\cos^2\theta - 1) \, , \tag{1.12}$$

expressed in polar coordinates (r,θ) centered on the star.

Charged particles are greatly influenced by the electric field on the surface and are expelled, making it impossible to maintain a vacuum around the neutron star. This flow of charged particles, combined with the neutron star's high magnetic field, forms the dense plasma that surrounds it. This creates a charge distribution, visible at the magnetic poles and the equator, where charges of opposite signs will be

positioned respectively. To measure the maximum value of the charge separation at the magnetic pole, we use the Goldreich-Julian charge number density (η_{GJ}):

$$\eta_{GJ} = \frac{\Omega B_s}{2\pi ce} \simeq \frac{B_s}{ceP} = 2.2 \times 10^{18} \sqrt{\frac{\dot{P}}{P}} \ \text{cm}^{-3} \ . \tag{1.13}$$

The quantity referred to in Equation (1.13) sets a baseline for magnetospheric plasma densities and serves as a reference for comparing different emission regimes.

The plasma contained in the envelope surrounding the neutron star is in a state of rigid co-rotation with it, influenced by an electromagnetic field that affects all charged particles uniformly. The plasma's dynamic behavior can only be maintained up to a certain distance, beyond which such co-rotation would require particles to move faster than the speed of light. This maximum distance can be calculated if we consider the co-rotation velocity, $v = \Omega R$, which defines the radius of an imaginary surface around the neutron star. This distance where $v = c$ is known as the radius of the light cylinder, taking again $\Omega = 2\pi/P$:

$$R_{lc} = \frac{c}{\Omega} = \frac{cP}{2\pi} = 4.77 \times 10^4 P \ \text{km} \ . \tag{1.14}$$

Now, assuming a dipolar behavior of the magnetic field at a distance r as we saw in Equation (1.9), where $B(r) \propto 1/r^3$, this can be defined using B_s, also seen as the magnetic field at the equator, and R:

$$B(r) = B_s \left(\frac{R}{r} \right)^3 \ . \tag{1.15}$$

Then, we can define the magnetic field in the light cylinder (B_{lc}) by introducing Equation (1.14) in Equation (1.15):

$$B_{lc} = B_s \left(\frac{R}{R_{lc}} \right)^3 = B_s \left(\frac{2\pi R}{cP} \right)^3 \ . \tag{1.16}$$

Assuming the canonical values seen in Equation (1.11):

$$B_{lc} \simeq 3 \times 10^8 \sqrt{\frac{\dot{P}}{P^5}} \ \text{G} \ . \tag{1.17}$$

The quantity of B_{lc} seen in Equation (1.17) gives the value taken from the ATNF to define the population of pulsars used in this work, and captures magnetospheric activity (see Chapter 4 for more information on its impact on population separation). To understand the role of the light cylinder, we refer to Figure 1.3. The light cylinder is defined by its R_{lc} and has a significant influence on the structure of the magnetic field lines. Inside the light cylinder lie the so-called closed field lines, which loop back to the surface of the neutron star. In contrast, open field lines extend beyond the light cylinder and do not return; instead, they carry relativistic plasma into the pulsar wind. These open field lines originate from a small region around the magnetic poles known as the polar cap (see Figure 1.3). In other words, the "last open field line", which touches the light cylinder tangentially, defines the edge of the polar cap over the neutron star surface. This region plays a crucial role in shaping both the radio emission beam and the structure of the particle outflow.

From that region, we aim to estimate the potential drop between the magnetic pole and the edge of the polar cap, referred to as surface electric voltage (ΔV).

First, we need to define the polar cap region, which can be described by (R, θ_p), where θ_p is the angle between the "last open field line" and the magnetic axis. More specifically, we have to compute the polar cap radius (R_p).

Considering these dipolar magnetic field lines as constant, under the boundary conditions (R, θ_p) and $(R_{lc}, \theta = \pi/2)$ due to aligned axes, we can see:

$$\frac{\sin^2 \theta}{r} = \frac{\sin^2 \theta_p}{R} = \frac{1}{R_{lc}} \ . \tag{1.18}$$

As explained above, the region of the polar cap is a circle on the neutron star surface delimited by the "last open field lines". Therefore, R_p can be seen as the projected distance from the magnetic axis to the "last open field line" under θ_p.

Thus, using Equation (1.18) and Equation (1.14) to describe θ_p , we can compute R_p as:

$$R_p = R \sin \theta_p = \sqrt{\frac{2\pi R^3}{cP}} \simeq 144 \times \sqrt{P} \ \mathrm{m} \ . \tag{1.19}$$

where $\Omega = 2\pi/P$, and $R = 10^6$ cm.

Now, we can obtain ΔV by applying Equation (1.12) at the magnetic pole $(R_{lc}, \pi/2)$ and polar cap (R_p, θ_p):

$$\Delta V = \frac{B_s \Omega^2 R^3}{2c^2} \simeq 2.1 \times 10^{20} \sqrt{\frac{\dot{P}}{P^3}} \ \mathrm{V} \ . \tag{1.20}$$

For the calculation of Equation (1.20), in addition to the canonical values set in Section 1.2.1, we recall that the CGS system[21] is assumed throughout this chapter except when the magnitude is indicated. Thus, the result in the first part of the above expression is given in Statvolts, whose equivalence to Volts is seen as 1 Statvolt $\simeq$ 300 V, giving the final result. The ΔV quantity provides an estimate of the pulsar's ability to accelerate particles, which may be relevant for emission modeling.

1.3   Alternative Emission Models

In this section, we revisit the main theoretical models developed to describe pulsar magnetospheres, with a particular focus on their ability to reproduce the gamma-ray light curves observed by instruments such as *Fermi*-LAT (Smith et al., 2023). Although pulsars emit across a broad range of wavelengths, only a limited subset of ~300 pulsars have been detected in gamma-rays (see Section 1.4 for more details about the population). This emission is highly anisotropic and originates from particles accelerated along open magnetic field lines by strong electric fields induced by the rotation of the magnetic field. As such, gamma-ray light curves offer constraints on the geometry and physical conditions within the magnetosphere. This emission is highly anisotropic and originates from particles accelerated along magnetic field lines opened by intense electric fields generated by the rotation of its magnetic field. The pulsar's radiation is explained based on assumptions regarding the location and geometry of the emission zones, as explained, e.g., by Brambilla et al. (2015), see also Philippov and Kramer (2022) for a comprehensive review.

Thus, models such as the Polar Cap (PC) (see, for example, Daugherty and Harding, 1996), Outer Gap (OG) (see, for example, Cheng, Ho, and Ruderman, 1986; Romani

---

[21]$c = 3 \times 10^{10}$ cm/s; $e = 4.8032 \times 10^{-10}$ cm$^{3/2}$ g$^{1/2}$ s^{-1}

and Yadigaroglu, 1995), and Slot Gap (SG) (Arons, 1983; Muslimov and Harding, 2004) were built upon the magnetic dipole geometry assumed in the vacuum model discussed in Section 1.2.1. However, the PC emission model, which assumes that particle acceleration occurs near the neutron star due to strong parallel electric fields, was ruled out due to the sensitivity achieved by *Fermi*-LAT, e.g., in measuring the Vela spectrum in Abdo et al. (2013a). However, none of these models can fully explain the complete phenomenology of gamma-ray light curves (see, for example, Romani and Watters, 2010; Pierbattista et al., 2012), nor are they consistent with the global properties of the magnetosphere, such as current closure. In particular, see Viganò et al. (2015a) and Viganò et al. (2015b) for a study of the approximations and assumptions behind OG models.

A more representative view of the geometry of the external field than that assumed by the magnetic dipole model in a vacuum has advanced the study of pulsar magnetospheres. This progress is facilitated by the development of, for instance, Force-Free Electrodynamics, see e.g., Spitkovsky (2006), Kalapotharakos and Contopoulos (2009), Pétri (2012), Harding and Kalapotharakos (2017), and Pétri and Mitra (2021), and Particle-In-Cell models, see Chen and Beloborodov (2014), Philippov and Spitkovsky (2014), Cerutti et al. (2015), and Cerutti, Figueiredo, and Dubus (2025). The former assumes a perfectly conducting plasma and therefore cannot account for particle acceleration due to the absence of parallel electric fields ($E_\parallel$). The latter incorporates kinetic microphysics and self-consistent particle acceleration. Each of these, and others such as MHD models e.g., Komissarov (2006) and Tchekhovskoy, Spitkovsky, and Li (2013) or resistive or dissipative magnetospheres model e.g., Kalapotharakos et al. (2012), Kalapotharakos, Harding, and Kazanas (2014), Brambilla et al. (2015), and Cao and Yang (2024) serve their own aims and have their own limitations. An Effective Synchro-curvature Model, which considers the motion of charged particles along curved magnetic field lines and calculates their radiation through synchro-curvature processes (Viganò et al., 2015c) tries to reproduce, using only three effective parameters that acts as order paraemeters, both the gamma-ray spectra across the entire population of known gamma-ray pulsars (Viganò et al., 2015c; Viganò, Torres, and Martín, 2015) and the subset with observed non-thermal X-ray pulsations (Torres, 2018; Torres et al., 2019). Recent advancements have further showcased its versatility, with successful applications in fitting gamma light curves of *Fermi*-LAT pulsars (see Íñiguez-Pascual, Torres, and Viganò, 2024; Íñiguez-Pascual, Torres, and Viganò, 2025 and Section 9.1 for more details).

1.4   Millisecond Pulsars

As mentioned at the beginning of this chapter, neutron stars can also be members of binary systems, which make up a significant fraction of the population shown in Figure 1.4. These systems offer a distinct perspective on the formation and evolution of neutron stars. Observations allow us to establish constraints that provide information about the properties of their companions. Most binary systems include white dwarfs, main sequence stars, or other neutron stars as companions, although a few systems exhibit unusual companions. For instance, PSR B1257+12 is accompanied by up to three terrestrial-mass bodies (see Table 6.2 for more details). At the same time, PSR B1620-26 is part of a triple system that includes a young white dwarf and a Jupiter-mass third body, located in the globular cluster M4 (Thorsett, Arzoumanian, and Taylor, 1993; Backer, Foster, and Sallmen, 1993; Thorsett et al., 1999; Sigurdsson et al., 2003). Noteworthy is J1719-1438, whose system produced interest due to

an uncertain companion that poses planetary mass but a remarkably high density (Bailes et al., 2011; van Haaften et al., 2012; Horvath, 2012).

Most of the binary systems observed are composed of MSPs as seen in Figure 1.4 with $P < 30$ ms, and are considered as the older neutron stars. The first MSP discovered was PSR B1937+21, as reported by Alpar et al. (1982), who proposed that these were merely the result of accretion. As we will see below, the MSPs are formed due to the spin-up that the pulsar experiences when it accretes mass from its companion, known as the recycling scenario (Bhattacharya and van den Heuvel, 1991; Tauris and Savonije, 1999; Tauris, Langer, and Kramer, 2012).

However, a subset of MSPs is found to be isolated, which likely formed in binary systems but later lost their companion through mechanisms such as ablation or disruption. Ablation, in particular, occurs when the pulsar's energetic wind evaporates the companion over time, a scenario supported by observations of so-called "black widow" and "redback" systems, denominated spider pulsars, transitioning into isolated MSPs (see eg., Fruchter, Stinebring, and Taylor, 1988; Roberts, 2013; Chen et al., 2013, and a further explanation about these systems in Section 1.4.3). Disruption, in contrast, refers to the dynamical breakup of a binary system, especially in dense stellar environments, resulting from gravitational interactions with passing stars. This is especially relevant in globular clusters, where the high number of isolated pulsars is a consequence of the increased frequency of such encounters and disruption events. It is worth noting that several alternative formation scenarios have been proposed to explain this observation, including accretion following mergers of neutron stars and main-sequence stars, as well as mergers of massive white dwarf binaries and tidal disruption events of main-sequence stars. In this regard, it is essential to note that the overall proportion of MSPs in globular clusters is significantly higher than that observed in the galactic plane (see Yin et al., 2024 and references therein). Another approach to characterizing the MSP population is through their gamma-ray emission (see Section 1.3 or, e.g., the 3PC or dedicated studies such as Torres et al., 2017). In the Third *Fermi*-LAT Pulsar Catalog (3PC), approximately 143 gamma-ray millisecond pulsars (gMSPs) have been identified.

1.4.1 Formation and Evolution of Binary Systems to Millisecond Pulsars

To explain a plausible formation pathway for MSPs, we focus on the evolution of binary systems, following the general framework described in Lorimer (2008). The process begins with the formation of a neutron star in a core-collapse supernova within a binary system. The binary can remain bound only if the mass ejected during the explosion does not exceed a critical fraction of the system's pre-supernova mass, typically, less than half. Otherwise, the sudden loss of gravitational binding leads to system disruption, leaving behind an isolated, rapidly rotating neutron star (see e.g., Blaauw 1961; Hills 1983). If the binary survives the explosion, the system may acquire a high orbital eccentricity due to the kick imparted to the neutron star at birth (Tauris et al., 2017). Over time, if the companion star is sufficiently massive, it will evolve off the main sequence and expand. Once it fills its Roche lobe, mass transfer begins, marking the initial point of the recycling scenario (Bhattacharya and van den Heuvel, 1991; Tauris and Savonije, 1999; Tauris, Langer, and Kramer, 2012; Patruno and Watts, 2021). In this process, the neutron star accretes matter and angular momentum from its companion, gradually spinning up to millisecond periods (Alpar et al., 1982; Radhakrishnan and Srinivasan, 1982). During this accretion phase, the system becomes a low-mass X-ray binary (LMXB), characterized by strong X-ray emission produced as the infalling material heats the surface

of the neutron star. Once accretion ceases, the pulsar reactivates in the radio band, now observable as a recycled MSP. Statistical analyses of pulsar spin distributions provide empirical evidence in support of this evolutionary scenario. Papitto et al. (2014) showed that different classes of accreting and rotation-powered MSPs exhibit distinct spin periods. They note that accreting MSPs and eclipsing rotation-powered MSPs seem to occupy an intermediate stage, reinforcing the notion of spin-down during the late stages of mass accretion and reflecting the diversity of evolutionary channels that can produce MSPs. The last two types of MSPs, as we will see in Section 1.4.3, can sometimes be adopted by the same transitional system as different states, referred to as transient millisecond pulsars.

1.4.2 Characterization of binary systems

In addition to the properties described in Section 1.2, which will continue to be of interest for characterizing the binary population, except τ_c (see discussion in Section 6.2.2), we will use the binary period (P_B), the projected semi-major axis (A_1), and the minimum mass of the companion (M_C).

These last parameters enable us to describe the binary nature of the MSP in question, providing a new and distinct perspective when classifying the population. These quantities will be obtained from the ATNF and analyzed in depth in Chapter 6 on their application.

It should be noted that P_B represents the orbital period observed in the system and will be expressed in days, following the ATNF. It is a quantity that can be directly measured from the periodic variations in pulse arrival times caused by the pulsar's motion around the center of mass of the binary system.

On the other hand, A_1 represents the projected semi-major axis that describes the elliptical orbit of the system as it moves. This quantity is derived directly from timing measurements and is called projected because we only see the component along our line of sight:

$$A_1 = \frac{a_1 \sin i}{c} \text{ lt-s} . \tag{1.21}$$

In Equation (1.21), a_1 denotes the semi-major axis of the pulsar's orbit and i represents the orbital inclination angle. An inclination of $i = 90°$ corresponds to an edge-on view, while $i = 0°$ describes a face-on orientation.

Assuming the third law of Kepler and the relationship of center of mass (Karttunen et al., 2007), we can explore the allowed range of M_C from the binary mass function, which relates the latter to the observables P_B, a_1, and i, alongside M and the gravitational constant G (Tauris and van den Heuvel, 2006; Bailes et al., 2011; Podsiadlowski, 2012):

$$f(M_C) = \frac{4\pi^2}{G} \frac{(a_1 \sin i)^3}{P_B^2} = \frac{(M_C \sin i)^3}{(M_C + M)^2} M_\odot . \tag{1.22}$$

For its calculation, assumptions such as $M \simeq 1.4 M_\odot$ and $i = 90°$ will give the minimum value of the companion's mass, M_C.

The steps carried out for Equation (1.22) are also applicable in cases where some observables are measured over a range of values or must be estimated through other parameters (see Section 6.10 for a specific case).

1.4.3 Spider Pulsars

Spider pulsars (e.g., Eichler and Levinson 1988; Roberts 2013; Roberts et al. 2018; Di Salvo et al. 2023), including black widows and redbacks, as well as a special state of the latter called transitional millisecond pulsars, form a special class of MSPs that reside in binary systems characterized by tight orbits and low-mass companions that are actively ablated by the pulsar wind, as we introduced in Section 1.4. They show irregular radio eclipses due to the presence of intrabinary material. For a comprehensive review of these objects and their phenomenology, see Papitto and de Martino (2022). The first black widow, B1957+20, discovered by Fruchter, Stinebring, and Taylor (1988), is an eclipsing pulsar with a $P =$ 1.6 ms and a binary period of 9.1 hr, orbiting a low-mass companion of 0.02 $M_\odot$. This discovery supported the idea that the companions of isolated MSPs were ablated by energetic particles and/or γ-rays produced by the pulsar wind. Although the recycling scenario (see Section 1.4.1) was increasingly validated by observations of fast millisecond pulsations in X-rays in transient LMXBs and X-ray bursts, it remained to be explained how the change from accretion to quiescence occurred to produce radio emission. This problem was described as the "missing link" between MSPs and LMXBs, and introduced the idea of a transitional millisecond pulsar. This is a class of redback systems in which a neutron star can alternate between accretion-powered and rotation-powered states. During the accretion stage, the system is observed to emit bright X-rays. When the mass transfer rate decreases in the final evolutionary stages, these binaries come to contain an MSP whose radio emission is enhanced by the rotating magnetic field of the neutron star. This scenario was favoured by the detection of X-ray millisecond pulsations from accreting neutrons stars, as SAX J1808.4-3658 (Chakrabarty and Morgan, 1998; Wijnands and Klis, 1998), and evidence of the existence of an accretion disk in the past of an MSP with the discovery of PSR J1023+0038, classified initially as a MSP and later found to exhibit characteristics of an accreting system (Archibald et al., 2009; Archibald et al., 2010; Patruno et al., 2014; Stappers et al., 2014). Although the latter was an early step in confirming this scenario, the first confirmed observation of a pulsar transitioning in real time between these two states was reported by Papitto et al. (2013), who detected X-ray pulsations from IGR J18245–2452 during an accretion episode followed by radio pulses a few days later, thus establishing it as the first confirmed transitional millisecond pulsar. PSR J1023+0038, now classified as a transitional millisecond pulsar, remains the best-studied system of its kind. It was the first to exhibit simultaneous optical and X-ray pulsations during the high-intensity X-ray mode, which disappear during the low-intensity mode, suggesting a shared underlying physical mechanism (Papitto et al., 2019). State transitions from XSS J12270–4859 were observed, resulting in the third confirmed transitional millisecond pulsar to date. The unusual properties seen in XSS J12270–4859 were determined to be characteristic of the sub-luminous state of tMSPs. The disappearance of optical emission lines and the significant dimming across multiple wavelengths indicated that XSS J12270–4859 underwent a transition from an accretion disc state to a radio pulsar state during a short time interval in late 2012 (Saitou et al., 2009; de Martino et al., 2010; Hill et al., 2011; de Martino et al., 2013; Bassa et al., 2014; de Martino et al., 2014).

As a summary, we must note that redbacks are eclipsing radio pulsars in tight binary systems, either with a non-degenerate main-sequence companion with a mass in the range ∼0.1–0.8 $M_\odot$. In contrast, black widows have a $\lesssim$ 0.06 $M_\odot$ semi-degenerate companion. While a transitional millisecond pulsar exhibits dramatic state changes:

transitioning from rotation-powered when they behave as redbacks, to accretion-powered and vice versa, on timescales as short as a few weeks. It is worth noting that all three confirmed transitional millisecond pulsars exhibit a peculiar state of accretion disk, characterized by an X-ray luminosity that is lower than that of outbursting AMXPs but brighter than that of both rotation-powered MSPs and quiescent AMXPs. A defining feature of this sub-luminous disk state is the presence of distinct X-ray intensity modes, typically classified as "high" (or active) and "low" (or passive), along with sporadic flaring episodes (Papitto and de Martino, 2022). One of the most efficient ways to identify candidate tMSPs has turned out to be searching for this enigmatic variability in the X-ray emission between the two intensity levels, "high" and "low" modes, in the sub-luminous disc state (see, e.g., de Martino et al., 2013; Papitto et al., 2013; Patruno et al., 2014; Archibald et al., 2015; Linares et al., 2014; Bogdanov et al., 2015; Papitto et al., 2019; Baglio et al., 2023). This method has proven successful in identifying a few candidates (see, e.g., Bogdanov and Halpern, 2015; Coti Zelati et al., 2019). In Section 6.2.3, we will show a more individual analysis of these classes of pulsars.

1.5 Fast Radio Bursts

Fast Radio Bursts (FRBs) were first detected in 2001 by the Parkes Telescope and reported in 2007 by Lorimer et al. (2007), named currently FRB 20010724 (also called FRB 010724 or "Lorimer Burst", see Lorimer, McLaughlin, and Bailes (2024) for a review of the discovery and highlights of the FRBs topic). They are transient radio pulses, typically lasting on millisecond timescales (see Zhang, 2023 for a comprehensive review). Although there is consensus on its astronomical origin based on all observations carried out and all events detected, this was not the case from the outset. An artificial origin occurring on Earth had to be ruled out ("perytons"; see, e.g., Burke-Spolaor et al. (2011) and Petroff et al. (2015)), while there were also speculations that it could be the pulse of other classes of isolated neutron stars observed at X-ray and gamma-ray wavelengths (McLaughlin et al., 2006).
However, the detection reported by Thornton et al. (2013) not only confirmed an extragalactic origin but also pinpointed its origin as an astronomical event.
The detectability of events originating from what is believed to be the same source establishes a distinction between two categories, at least from an observational point of view: repeaters and non-repeaters, albeit it is unclear whether all non-repeaters will eventually repeat. This is because, until Spitler et al. (2016) detected several bursts from the same source, named FRB 20121102A (also called FRB 121102, "R1", or "Spitler burst"), which demonstrated that the source survives the process that produces the bursts, all detected FRBs had been attributed to different sources. The repetition established that the FRBs do not originate in cataclysmic astrophysical events, contrary to previous hypotheses.
After that, the number of detected repeating sources increased due to observations made by the Canadian Hydrogen Intensity Mapping Experiment (CHIME, Amiri et al., 2019; CHIME/FRB Collaboration et al., 2019), the Australian Square Kilometre Array Pathfinder (ASKAP) (Kumar et al., 2019), and the Five-hundred-meter Aperture Spherical radio Telescope (FAST) in China (Luo et al., 2020; Niu et al., 2022). Furthermore, most of these sources have been detected at cosmological distances, leading to the expectation that they generate extreme, coherent radio emissions.
The repetitive behavior of FRB 20121102A prompted specific observation campaigns using the Karl G. Jansky Very Large Array (VLA) in conjunction with the Arecibo

telescope, focusing on detecting possible additional bursts and estimating the source's location using the interferometric technique (Chatterjee et al., 2017). Noteworthy are the observations via interferometry carried out by several groups, such as the ASKAP collaboration and Deep Synoptic Array (DSA) collaboration, which led to the location of both repeating and non-repeating sources in different types of host galaxies (Bannister et al., 2019; Prochaska et al., 2019; Ravi et al., 2019; Macquart et al., 2020; Marcote et al., 2020; Bhandari et al., 2022; Xu et al., 2022)

Regarding its origin, several theoretical models were proposed (e.g., Platts et al., 2019, for a summary) before the first discovery. It has been suggested that their emission could be powered by the dissipation of magnetic fields of a magnetar (Popov and Postnov, 2013; Lyubarsky, 2014; Katz, 2016; Murase, Kashiyama, and Mészáros, 2016; Kumar, Lu, and Bhattacharya, 2017; Nicholl et al., 2017; Margalit and Metzger, 2018; Lu, Kumar, and Zhang, 2020; Yang and Zhang, 2021). This has been substantiated after CHIME (CHIME/FRB Collaboration et al., 2020) and STARE-2 radio array (Bochenek et al., 2020) discovered FRB 200428 in association with a hard X-ray burst from the Galactic magnetar, SGR 1935+2154 (Li et al., 2021; Mereghetti et al., 2020; Zhong et al., 2020). In the magnetar interpretation, the leading source model for FRBs, some differences may arise from age, with young magnetars producing a higher repetition rate than older ones (Beloborodov, 2017; Metzger, Berger, and Margalit, 2017; Margalit et al., 2020; Levin, Beloborodov, and Bransgrove, 2020; Feng et al., 2022). However, this may not represent the entire FRB population; see, e.g., the analysis observed in FRB 20200120E in Pearlman et al. (2025). They argue against typical ultraluminous X-ray bursts, magnetar-like giant flares, or a SGR 1935+2154-like intermediate flare as associations for this repeater. Additionally, they suggested that it is unlikely to be an ultraluminous X-ray source or a young extragalactic pulsar embedded in a Crab-like nebula acting as the engine of FRB 20200120E. Other scenarios proposed include a giant radio pulse-emitting pulsar or a magnetar formed through a delayed channel. Zhang (2023) also offers an extensive review of FRB source models. Although initially unexpected in the context of pulsars, the behavior of FRBs immediately suggested a compact object origin. Among the most compelling models are those involving neutron stars, particularly magnetars, as progenitors. This connection, along with the discovery of repeating FRBs, has prompted the inclusion of these events in this work.

1.6   Conclusions

This section outlines the fundamental observables of pulsars, P and $\dot{P}$, and how they provide a first approach to the pulsar population. From these quantities, several derived quantities such as the spin-down energy loss, characteristic age, and surface magnetic field can be inferred, as discussed in Section 1.2.2. These parameters describe (as a proxy) the rotational energy loss and provide a first estimate of the pulsar's evolutionary stage and magnetic properties. However, they are derived under assumptions that the pulsar behaves as an isolated, rotating magnetic dipole in a vacuum. As such, they do not account for the presence of a magnetosphere and do not describe the detailed mechanisms by which radiation is produced. Some physical parameters discussed in Section 1.2.3 relate directly to the environment where pulsar emission originates. These include the Goldreich–Julian charge number density, magnetic field at the light cylinder, and the surface electric voltage. They are linked to the conditions under which particle acceleration and radiation

can occur. In particular, the structure of the magnetosphere and the associated electric and magnetic fields determine whether a pulsar can accelerate particles to the energies required for producing coherent radio waves. It can be observed that the presence of open field lines, a well-defined polar cap, and a sufficient potential drop are relevant for sustaining radio emission.

While all these derived properties are model-dependent, they serve as a proxy within the complete set of physical parameters, providing a framework for understanding how pulsars evolve, emit, and interact with one another. The $P\dot{P}$ diagram, as a visual tool, has played a central role in the pulsar population. However, it has a limitation when it comes to displaying all the information in the population's feature space. This can be seen in Figure 1.4, where the addition of constant lines from the physical property relationships would render the diagram useless. For this reason, something on which we shall further comment later, relying solely on the P–$\dot{P}$ diagram may be misleading, especially when trying to separate populations or infer connections between classes of pulsars, as it requires all available information. We will present approaches that enable the feature space to be expanded by leveraging similarity relationships, such as graphs —specifically, the Minimum Spanning Tree (see Chapter 2) —in conjunction with dimension reduction and feature space analysis techniques that rely on the Principal Component Analysis (see Chapter 3).

Particular attention is paid to MSPs, especially those in binary systems. These objects represent a key phase in the evolutionary process of neutron stars, offering insight into recycling processes, mass transfer, and spin-up mechanisms. Binary MSPs also introduce additional observables into the analysis, such as orbital period, the projected semi-major axis, and companion mass estimates, thereby enriching the feature space and enabling a more detailed classification. Additionally, it is worth noting that no methodology has been established to identify candidates among radio-emitting pulsars that are not in a sub-luminous disk state, and the Minimum Spanning Tree can contribute to this purpose.

In Chapters 4, 5, and 6, we will see specific applications of the described pulsar populations.

Furthermore, in Chapter 7, we will discuss FRBs from these approaches to offer a new form of separation based on their repetition rate.

Chapter 2

Graph theory

2.1 Introduction

Graph theory has become central to advanced mathematics and data science, offering a flexible framework for modeling relationships between objects. Originating from Euler's solution to the Königsberg bridge problem in the 18th century (see Euler, 1736; Shields, 2012), it has since evolved into a foundational tool across various fields, including computer science, engineering, biology, linguistics, and economics. Its applications are wide-ranging: in social networks, it is used to identify influential nodes (Granovetter, 1973; Rashidi, Boroujeni, Soltanaghaei, et al., 2024); in biology, to map protein-protein interactions (Ghosh, Kumar, Basu, et al., 2015); in finance, to study risk propagation through interconnected asset networks (Smerlak et al., 2015); in transportation, to optimize routes and infrastructure (Henzinger et al., 1997; Roughgarden and Tardos, 2002); in neuroscience, it supports the analysis of brain connectivity and cognitive processes (Bullmore and Sporns, 2009); in genomics, it plays a central role in genome sequencing and the reconstruction of DNA sequences (Chikomana and Hu, 2023). The list is not exhaustive. Graph theory is particularly well-suited for analyzing systems of discrete elements with pairwise interactions, as these are often represented as nodes connected by edges as seen in Figure 2.1.

In particular, the Minimum Spanning Tree (MST) is a useful structure for uncovering relationships in multidimensional data.

The MST is a graph that connects points in a multidimensional space. Each point (or node) is linked to at least one other by an edge whose length is associated with a given distance, aiming to minimize the total edge weight while forming a connected, acyclic graph.

Unlike methods that rely on Euclidean space, the MST operates without imposing assumptions about the space's global geometry. This flexibility allows it to capture complex patterns in the data that might be missed by conventional techniques. Moreover, the MST is firmly rooted in graph theory, which provides a mathematically rigorous and conceptually rich framework.

In addition, graph theory demonstrates that the MST is unique as long as the distances are distinct (Kleinberg and Tardos, 2005; Cormen et al., 2009; Goodrich and Tamassia, 2014), and its definition intuitively suggests that it is an optimization technique.

The MST was widely used in engineering problems, starting from its original application developed by Borůvka in 1926, for electricity distribution in Moravia, as seen in Nešetřil, Milková, and Nešetřilová (2001). Currently, the MST is employed for analyzing cognitive impairment (Simon et al., 2021) to risk in financial markets (Pozzi, Di Matteo, and Aste, 2013). Early applications in scientific problems include describing the interrelationship of species or genetics (see the work of Florik in the 1950s and Edwards in the 1960s, as commented in Hartigan, 1981 and Winther, 2018,
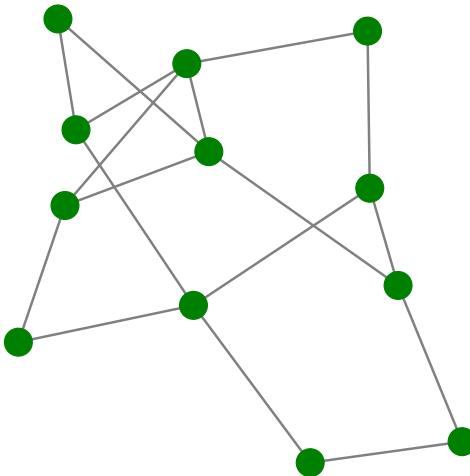
FIGURE 2.1: Representation of a graph composed of nodes (green) and edges (black). An edge indicates a direct relationship between two nodes. This relationship can be quantified, for example, by assigning a weight to the corresponding edge, which represents a distance or similarity.

respectively), disciplines in which this technique is widely exploited. This non-exhaustive reference list demonstrates growing interest in utilizing the MST across various fields.

In astronomy, the MST has been used for finding high-energy sources for photon-starved imaging techniques (Campana et al., 2013), establishing differences between cluster and field stars (Sánchez, Alfaro, and López-Martínez, 2018), detecting filaments (Pereyra et al., 2020), galaxy clustering (Barrow, Bhavsar, and Sonoda, 1985), and cosmology (Bonnaire et al., 2020; Naidoo et al., 2020; Naidoo, Massara, and Lahav, 2022; Naidoo and Lahav, 2024). This approach has also been implemented to analyze event samples in particle colliders (Lovelace Rainbolt and Schmitt, 2017) or cosmic rays (Harari, Mollerach, and Roulet, 2006). Despite this interest, as noted, the MST has barely been used for pulsars. To our knowledge, the first related publication is (Maritz, Maritz, and Meintjes, 2016), which uses 11 handpicked objects. The goal was to demonstrate that an MST could distinguish between binaries and isolated pulsars using the dispersion measure as a proxy for distance. In a more recent work, the MST has been utilized for classifying pulsars based on their profiles (Vohl, van Leeuwen, and Maan, 2024).

Graph theory, when applied through the MST, also offers various centrality estimators, such as betweenness and closeness, that serve as tools for structurally analyzing populations of objects. For example, in (Baron and Ménard, 2021), the authors present an algorithm that identifies the main trend in a dataset by constructing graphs and exploiting the characteristics of their structures to reveal global sequences, with applications in astronomy, geology, and natural image data. This chapter introduces the fundamental concepts necessary to build and work with weighted graphs, with a focus on the MST. We aim to establish it as a robust, practical framework for extracting meaningful insights from complex astronomical data. In addition, we demonstrate the utility of betweenness and closeness centrality as structural metrics for a deeper understanding of the relative importance of individual nodes within the graph.

2.2 Basic definitions for building a weighted graph

Graph theory seeks to establish relationships among objects based on their connections, ultimately representing these relationships in a graph. Let $G(V, E)$ denote a graph of a set of nodes V and a set of edges E, where each edge $e = v_n, v_m$ connects a pair of nodes. To represent the relationship between the nodes, a weight $w(e)$ (e.g., a distance value), is assigned to each edge, making G a weighted graph (see, e.g., Wilson, 2010). When the edges have no orientation imposed, $e = \{v_n, v_m\} = \{v_m, v_n\}$, then G can be considered an undirected graph. A possibility is to assume the $w(e)$ as the Euclidean distance calculated as the straight-line distance between (v_n, v_m) in a N-dimensional space. This can be calculated as follows:

$$d_{nm} = \sqrt{\sum_{j=1}^{N}(v_{jn} - v_{jm})^2} \ . \tag{2.1}$$

In Equation (2.1), d_{nm} represents the $w(e)$, where v_{jn} and v_{jm} denote the j-th variable associated with nodes v_n and v_m, respectively. A normalization process is typically required when variables have different units or scales. The total weight of G will be obtained from the sum of each specific $w(e)$ as follows:

$$w(G) = \sum_{e \in E(G)} w(e) \ . \tag{2.2}$$

The resulting value in Equation (2.2) is particularly useful in optimization problems where minimizing $w(G)$ is of central interest.

2.2.1 General properties of graphs

The following properties are general and applicable to any graph, as seen in Wilson (2010). The notation introduced in Section 2.2 for the specific case of G is retained here. A *path* in a graph is a sequence of nodes in which no node is repeated. If the path starts and ends at the same node, it is called a *cycle*. A graph is said to be *connected* if a path exists between every pair of nodes. The *degree* of a node v, denoted $g(v)$, is the number of edges incident to it, that is, the number of adjacent nodes directly connected to v. The overall structure of a graph can also be encoded in its *adjacency matrix*, A, which has dimensions $|V| \times |V|$[1]. The element A_{nm} indicates whether nodes v_n and v_m are adjacent. For unweighted graphs, $A_{nm} = 1$ if the edge $e = \{v_n, v_m\}$ exists; for weighted graphs, $A_{nm} = w(e)$. The main diagonal of A is zero when the graph contains no loops (i.e., no node is connected to itself). Finally, we introduce the notion of a *cut*, particularly relevant in the algorithms discussed in Section 2.3. A cut is a partition of V into two disjoint subsets, S and $V - S$. The associated *cut set* consists of all edges $e = \{v_n, v_m\}$ such that $v_n \in S$ and $v_m \in V - S$.

2.3 The minimum spanning tree

Since G has an edge between every pair of nodes, it is a complete, undirected, and weighted graph. Let $T(V', E')$ be a subgraph of $G(V, E)$, where $V' \subseteq V$ and $E' \subseteq E$.

---

[1]The notation $|\cdot|$ represents the cardinality (or size) of the specified set, indicating the number of elements within it.

We define T as a spanning tree of G if it connects all nodes of G, which implies $V' = V$, and contains no cycles. The following statements regarding T are equivalent (see e.g., Tarjan, 1983)

- T is a spanning tree of G.

- T is connected and contains no cycles.

- T is connected and E' contains $|V| - 1$ edges.

- T contains no cycles and E' contains $|V| - 1$ edges.

- T is minimally connected: removing an edge would disconnect it.

- T is maximally acyclic: adding an edge would form a cycle.

An MST is a spanning tree T of a graph G for which the sum of all edge weights is the smallest possible, according to Equation (2.2). Thus, the MST connects all nodes in V using locally minimal distances, subject to the global constraint of minimizing the total distance, resulting in a connected and acyclic structure. The MST is obtained via greedy algorithms, which select the best possible local choice in each iteration to pursue a globally optimal solution (see, e.g., Roughgarden, 2019 for more details).

2.3.1 The Kruskal's algorithm

We employ a specific greedy algorithm, known as Kruskal's algorithm (Kruskal, 1956), to compute the MST of $G(V, E, w)$. The algorithm constructs an MST by iteratively adding the lowest-weight edge that does not form a cycle. Its basic steps are:

- Sort the edges in E by increasing $w(e)$.

- Initialize an empty set of edges E' and let $T = (V, E')$.

- Choose the edge e with the smallest $w(e)$ and add it to E'.

- Make sure that the chosen e does not produce any cycle in the structure of T.

- Repeat until T spans all nodes in V, such that $|E'| = |V| - 1$ and the total weight $w(T)$ is minimized (see Equation 2.2).

The pseudocode for Kruskal's algorithm, following Erickson (2019), is shown below: The implementation of Kruskal's algorithm relies on a disjoint-set data structure (see, e.g., Cormen et al., 2009), which operates on a collection of disjoint subsets of V. This structure supports three key operations:

- *MakeSet*(v): Generates a new subset containing only v.

- *Find*(v): Operation returns an identifier for the subset containing v.

- *Union*(v_n, v_m): Merges the subsets containing v_n and v_m.

Figure 2.2 illustrates the application of Kruskal's algorithm to a weighted and undirected graph, yielding an MST. Let $G(7, 10)$, creating a $T(7, E')$ with $E' = 0$ the edges are included into the E' in increasing order starting from $e = \{A, B\} = 1$. Then, the edges are added to E' to connect all the nodes, avoiding cycles, forming an MST as $T(7, 6)$. Note in the last panel that $e = \{B, C\} = 6$ would create a cycle, so it is discarded and the $e = \{E, G\} = 7$ is added.
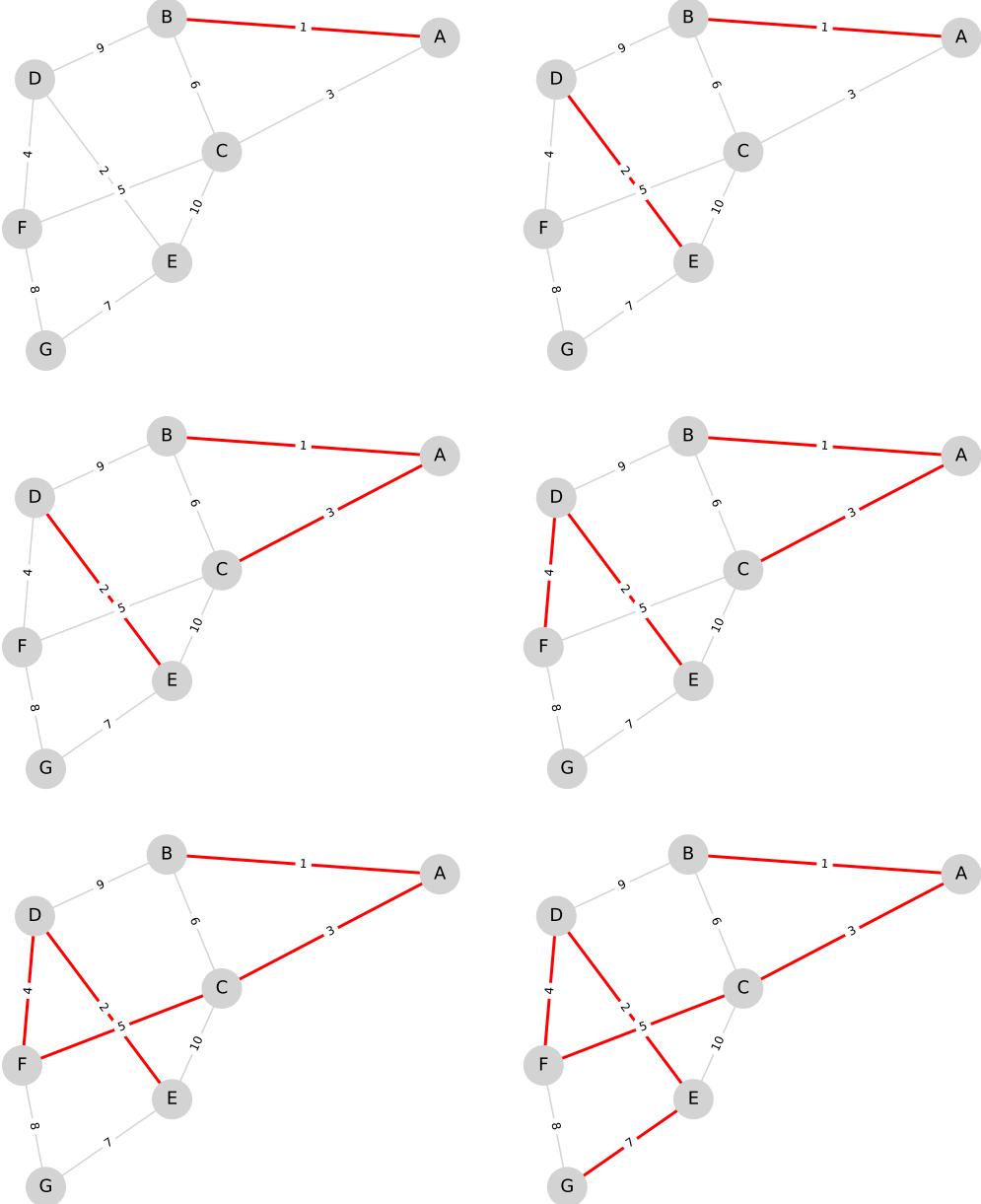
FIGURE 2.2: From left to right and top to bottom, we illustrate the application of Kruskal's algorithm. Each panel represents an iteration of the algorithm, adding an edge (marked in red) while avoiding the creation of cycles. Let $G(7, 10)$ be an undirected and weighted graph, whose edges are colored grey, except in the first panel where the $e = \{A, B\} = 1$ edge starts in red. The last panel displays the resulting MST, $T(7, 6)$, with all edges in red.

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

Require: sort E in increasing order according to $w(e)$
 1: $T \leftarrow (V, \varnothing)$
 2: **for** each $v \in V$ **do**
 3:     *MakeSet*(v)
 4: **end for**
 5: **for** each $e = \{v_n, v_m\} \in E$ (in order) **do**
 6:     **if** *Find*(v_n) $\neq$ *Find*(v_m) **then**
 7:         *Union*(v_n, v_m)
 8:         $T \leftarrow T \cup e$
 9:     **end if**
10: **end for**
11: **return** T

2.3.2 The Prim's algorithm

Prim's algorithm (Prim, 1957) is an alternative method for computing the MST of $G(V, E, w)$ and is equally valid. This algorithm grows the tree from an initial node by successively attaching the nearest node not yet in the tree, always selecting the lightest available edge that connects to the existing structure without cycles. Its basic steps are:

- Initialize the tree T with an arbitrary starting node $v_0 \in V$ and an empty set of edges E'.

- Create a priority queue storing all edges of E connecting T to nodes not yet in T.

- Choose the edge e with the smallest $w(e)$ that connects a visited node to an unvisited one and add it to E'.

- Repeat until T spans all nodes in V, such that $|E'| = |V| - 1$ and the total weight $w(T)$ is minimized (see Equation 2.2).

The pseudocode for Prim's algorithm presented below is a hybrid adaptation of the versions seen in Erickson (2019) and Cormen et al. (2009).
This implementation of Prim's algorithm is based on the following standard definitions and operations as described in Cormen et al. (2009):

- *priority(v)*: Represents the minimum weight of any edge connecting v to the growing MST.

- *parent(v)*: Stores the node through which v will be attached to the final MST.

- *PriorityQueue*: A data structure that maintains the set of candidate nodes ordered by their priority values.

- *ExtractMin*: Selects and removes the node with the lowest priority from the queue, corresponding to the next node added to the MST.

- *DecreaseKey*: Updates the priority of a node in the queue when a lighter connection is found, preserving the correct ordering.

Algorithm Prim($G = (V, E, w), v_0$)

1: **for** each $v \in V$ **do**
2:     $priority(v) \leftarrow \infty$
3:     $parent(v) \leftarrow$ None
4: **end for**
5: $priority(v_0) \leftarrow 0$
6: $Q \leftarrow$ PriorityQueue(V, keyed by $priority$)
7: **while** $Q \neq \emptyset$ **do**
8:     $v_n \leftarrow ExtractMin(Q)$
9:     **for** each neighbor v_m of v_n **do**
10:         $e \leftarrow \{v_n, v_m\}$
11:         **if** $v_m \in Q$ and $w(e) < priority(v_m)$ **then**
12:             $parent(v_m) \leftarrow v_n$
13:             $priority(v_m) \leftarrow w(e)$
14:             $DecreaseKey(v_m, w(e))$
15:         **end if**
16:     **end for**
17: **end while**
18: $E' \leftarrow \{\{v, parent(v)\} \mid v \in V \setminus \{v_0\}\}$
19: **return** $T = (V, E')$

Figure 2.3 shows the application of Prim's algorithm on a weighted and undirected graph, resulting in an MST. Let $G(7, 10)$, and $v_0 = E$ (this selection is completely random), create a $T(v_0, E')$ with $E' = 0$. We create a priority queue of the edges connected to E, $e = \{E, C\} = 10$, $e = \{E, D\} = 2$, and $e = \{E, G\} = 7$, from where the smallest one is selected, $e = \{E, D\} = 2$, and the node D is added to T. This step is repeated until all the nodes are included in V, without creating any cycle, forming an MST as $T(7, 6)$. Note in the last panel that $e = \{B, C\} = 6$ is on the priority queue, so B and C are already part of T, but would create a cycle, so it is discarded and the $e = \{E, G\} = 7$ is added.

2.3.3 MST properties

Below, we list key properties associated with the MST, see e.g., Kleinberg and Tardos (2005) and Goodrich and Tamassia (2014), where proofs can also be found. Considering that $T(V, E')$ is the MST of $G(V, E)$, the principal properties are:

- Uniqueness: The resulting MST will be unique if all $w(e)$ values in $E(G)$ are distinct. This implies that choosing any other greedy algorithm instead of Kuskal's as a solution to search for the MST (e.g., Prim's algorithm) would give the same results. This is the case, for instance, seen in Figures 2.4.

- 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 $e_c \in C$ it is determined that $T(V, E' \cup \{e\} - \{e_c\})$ is a spanning tree. Thus, if T is an MST, then $w(e_c) > w(e)$. Otherwise, the edge with the largest weight 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 $w(e)$.
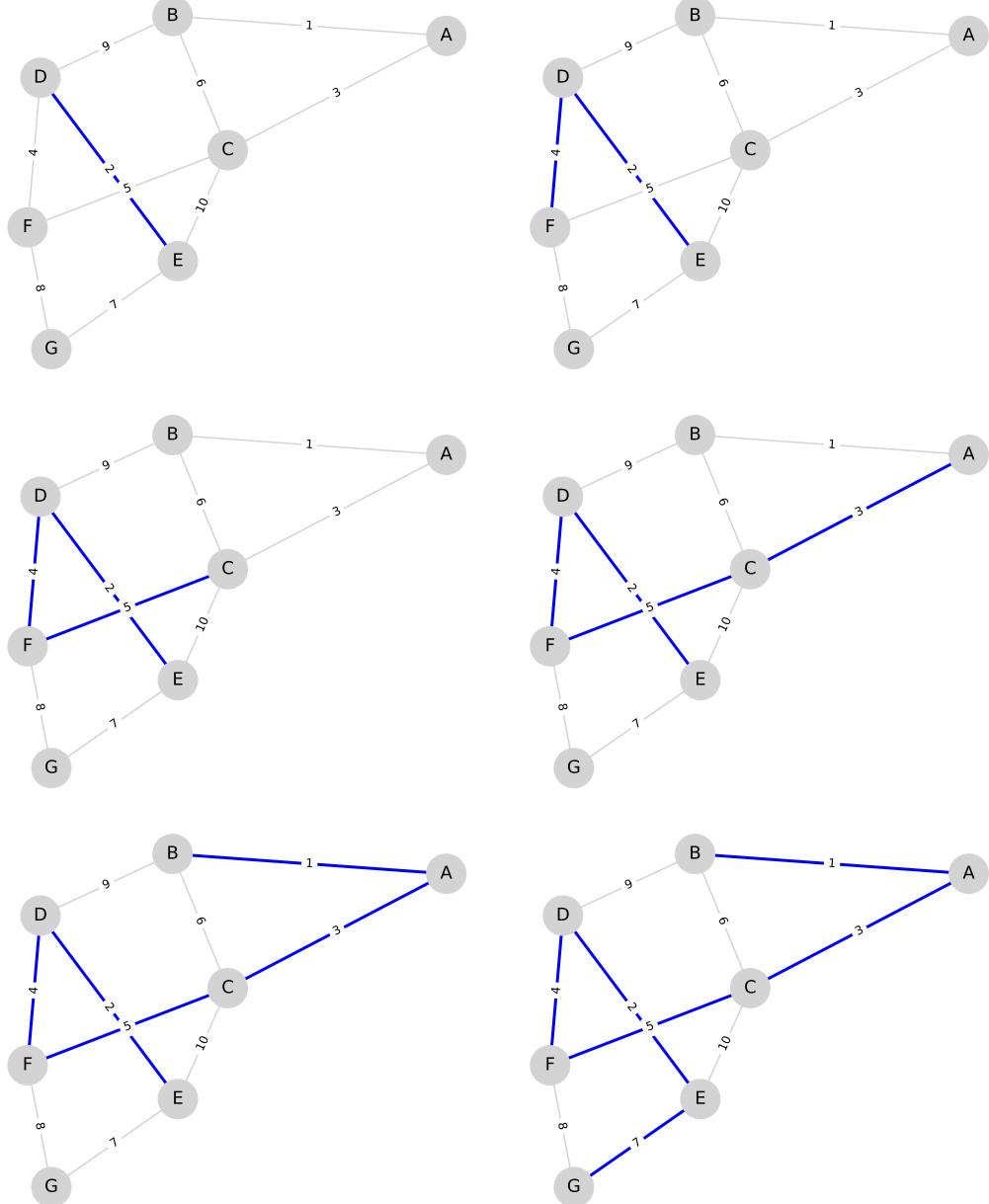
FIGURE 2.3: From left to right and top to bottom, we illustrate the
application of Prim's algorithm. Each panel represents an iteration
of the algorithm, adding an edge (marked in blue) while avoiding
the creation of cycles. Let $G(7,10)$ be an undirected and weighted
graph, whose edges are colored grey, except in the first panel where
the $e = \{E, D\} = 2$ edge starts in blue. The last panel displays the
resulting MST, $T(7,6)$, with all edges in blue.
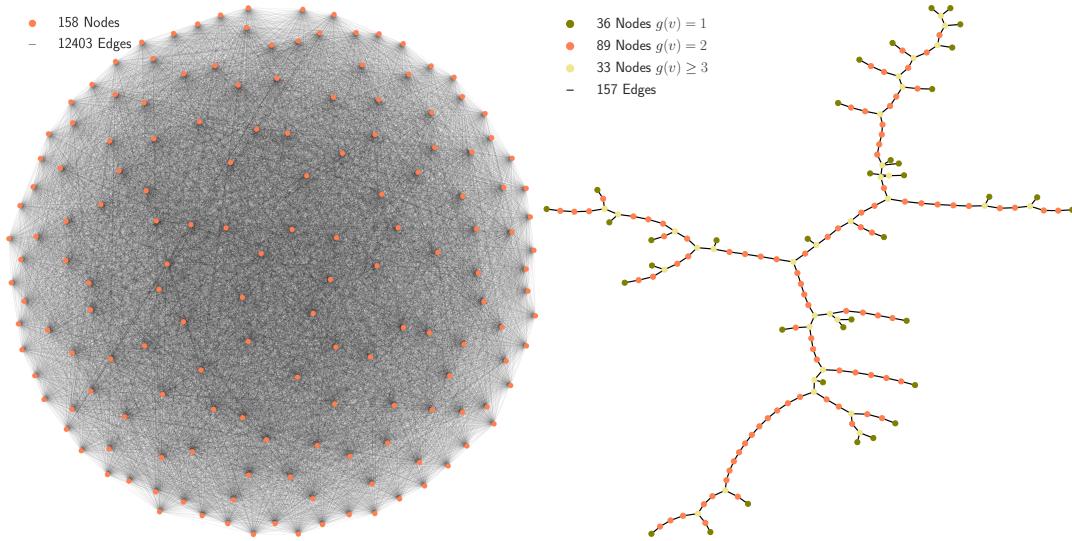
FIGURE 2.4: Left: Complete, undirected and weighted graph $G(158, 12403)$. Right: Corresponding MST, $T(158, 157)$, from the defined G. The nodes are marked according to the degree: $g(v) = 1$ –olive, $g(v) = 2$ –coral, and $g(v) \geq 3$ –yellow .

- 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 $e_D \in D$ such that $T(V, E' - \{e\} \cup \{e_D\})$ is a spanning tree. Thus, if T is an MST, then $w(e_D) > w(e)$; otherwise, the lightest edge contained in D must be included in E' for T to remain an MST.

Figure 2.4 shows an example to exemplify the structures of G and its derived MST T, together with the complexity of working with a high number of nodes $|V|$. Due to the uniqueness property described above, T is a unique MST because all weights of G are distinct. Therefore, this result is achieved by applying both the Kruskal and Prim's algorithms. A similar result is seen for the examples in Figures 2.2 and 2.3.

2.4 Centrality estimators in graphs

In graph theory, centrality measures quantify the importance of individual nodes within a graph's structure. These estimators capture how well-positioned a node is to control flow, connect different regions, or act as a structural bridge. This section focuses on two centrality measures relevant to the MST: betweenness centrality and closeness centrality. As betweenness centrality was ultimately selected for our analysis, we include a step-by-step example of its application in Appendix A to aid understanding. Closeness centrality is presented as a complementary reference to highlight the methodological contrast and clarify the rationale behind our choice; this discussion is provided in Section 2.5.

2.4.1 Betweenness centrality estimator

Betweenness centrality is an estimator that defines how central a node is, or, put otherwise, how many times a given node of a graph is in between any two others

(see Freeman, 1977, see also Moxley and Moxley, 1974). Let $G(V, E, w)$, with $v_s \in V$ to $v_j \in V$, this estimator measures centrality from the ratio between the number of times a v_j appears on the shortest path between any two other nodes (v_s, v_t), and the number of possible shortest paths that could occur between these. This can be seen via (see e.g., Brandes, 2001):

$$C_B(v_j) = \frac{2}{(N-1)(N-2)} \sum_{s \neq j \neq t \in V} \frac{\sigma_{s,t}(v_j)}{\sigma_{s,t}} . \tag{2.3}$$

Here, $\sigma_{s,t}(v_j)$ is assigned to one when any shortest path between the nodes (v_s, v_t) passes through the node v_j, or zero if it does not. Note that in unweighted graphs, shortest paths are defined by the minimal number of edges; in weighted graphs, they are determined by the paths with minimal total edge weight. However, as explained in Section 2.3, the MST is acyclic, so only one path is possible between any two nodes when we work on T. This means that for all $s \neq t$, $\sigma_{s,t} = 1$; as such, the calculation of $C_B(v_j)$ in an MST becomes independent of $w(e)$ and reflects only the topological position of the node within the tree. Let G (or T in case of the MSTs) is undirected, $\sigma_{s,t} = \sigma_{t,s}$ will be taken into account only once in Equation (2.3), which will be normalized by multiplying it by the factor $2/(N-1)(N-2)$ where $N = |V|$. This allows us to compare the results obtained between graphs of different $|V|$. Betweenness centrality helps formalize an intuitive mental concept of how central a node is in a given graph, allowing us to establish mathematical definitions based on its distribution. An explicit example in Appendix A.1 with an MST and the graph from which we obtain that MST of a few nodes will help clarify the computation.

2.4.2 Closeness centrality estimator

Closeness centrality measures the proximity of a node to all other nodes in a graph, based on the average length of the shortest paths connecting it to the rest (Sabidussi, 1966; Freeman, 1977). Using the same notation and assumptions introduced in Section 2.4.1, this estimator quantifies the proximity of v_j to every other $v_t \in V$ by computing the inverse of the shortest path distance $d_G(v_j, v_t)$. This is defined as the sum $w(e)$ along the path from v_j to v_t (with $d_G(v_j, v_t) = d_G(v_t, v_j)$ by definition). The closeness centrality of v_j is given by (see e.g., Brandes, 2001):

$$C_C(v_j) = (N-1) \sum_{j \neq t \in V} \frac{1}{d_G(v_j, v_t)} , \tag{2.4}$$

The factor $(N-1)$ in Equation (2.4) ensures that the measure is normalized and comparable across graphs of different $|V|$.

As discussed below, this estimator relies on spatial distance, even in the MST approach, making it unappealing when searching for a quantifier to highlight the different structures observed in an MST from a topological perspective, such as identifying the tree's main trunk.

An example of the calculation of this estimator can be found in Appendix A.2.

2.5 Conclusions

This chapter establishes the theoretical foundations and practical utility of graph theory, focusing on the MST to represent multidimensional data structures. The MST

provides a novel visualization of complex relationships between objects simply and clearly. Subject to the framework of graph theory, the MST allows us to explore different populations by leveraging robust mathematical tools.

Its simplicity and robustness allow both interpretability and flexibility for various scientific applications. In the broader context of this thesis, the MST framework plays a central role in the exploration and interpretation of pulsar populations and fast radio bursts.

In this work, we selected Kruskal's algorithm for computing the MSTs primarily over Prim's algorithm because of its ease of implementation and conceptual alignment with the problem setup. Since the complete set of pairwise distances between nodes was available in advance, Kruskal's edge-sorting strategy offered a direct and efficient means to construct the MST, without requiring dynamic key updates characteristic of Prim's approach. Although the latter can scale well in sparse and dense settings, depending on the data structures used, Kruskal's clarity and edge-centric design made it better suited to our study's conditions and assumptions.

Regarding the applications, as shown in Chapter 4, the MST can be used as a descriptive and alerting tool to visualize the pulsar population and their relationships in high-dimensional space, revealing patterns that are not easily captured by traditional representations, such as the $P\dot{P}$ diagram.

One can also utilize the MST to define possible subpopulations of pulsars that are distinguishable from each other by considering their physical properties through purely mathematical tools, such as centrality estimators. See cases of this approach in Chapter 5 and Chapter 7.

Specifically, we adopt betweenness centrality to quantify the relative structural importance of nodes within the MST. Rather than using closeness centrality, this choice is motivated by the desire for a measure less directly affected by the geometry of the feature space from which the edge weights are derived. Closeness centrality, computed over these weights as we see in Equation (2.4), is inherently sensitive to the distribution of distances in that space. In contrast, betweenness centrality, under the acyclic attribute of the MST, reflects how often a node serves as a bridge along shortest paths between other nodes, independent of the problem geometry, providing a more topologically robust and intuitive representation of what visually appears to be the tree's main trunk. On the other hand, it is worth noting the MST's predictive and inferential potential described in Chapter 6. It helps us characterize millisecond pulsars and promotes possible candidates for spider pulsars based on the structural position. Similarly, it enables us to suggest plausible ranges for undetermined parameters in some sources, considering their proximity in the MST to better-known neighbors. Finally, Chapter 7 presents MST as an unsupervised methodology that separates repeaters from non-repeaters of fast radio bursts, helping to identify repeater candidates and the most relevant variables for their separation. These applications underscore the value of MST as a visualization technique and framework for exploratory, descriptive, prescriptive, and even predictive analysis. Its integration into astrophysical problems provides a fresh, robust, and interpretable alternative to usual classification and clustering approaches.

Chapter 3

Principal Component Analysis

3.1 Introduction

Principal Component Analysis (PCA) is a widely used technique for reducing the dimensionality of complex datasets while preserving as much information as possible, while also retaining the importance of each original variable. Comprehensive discussions can be found in Mardia, Bibby, and Kent (1979), Jolliffe (2002), Izenman (2008), and Rencher and Christensen (2012); see also the reviews by Shlens (2014) and Jolliffe and Cadima (2016).

Initially introduced by Pearson (1901) in the context of the principal axes of ellipsoids in geometry, the method was independently developed by Hotelling (1933), known as the "Hotelling transform," for statistical analysis. It is closely related to the Kosambi-Karhunen-Loève transform (Ramaswamy, 2016; Karhunen and Selin, 1960; Lévy and Michel, 1965), and belongs to a family of classical dimensionality reduction techniques, alongside Independent Component Analysis (Herault and Ans, 1984; Ans, Herault, and Jutten, 1985; Herault, Jutten, and Ans, 1985; Comon, 1994), and Non-negative Matrix Factorization (Lee and Seung, 1999; Lee and Seung, 2000).

In astronomy, PCA has proven especially valuable for extracting the dominant sources of variation in high-dimensional datasets (Wall and Jenkins, 2003; Ivezić et al., 2014). Early applications in astronomy (see, e.g., Francis and Wills, 1999) include spectral analyses of stars Deeming (1964) and Whitney (1983), galaxies (Faber, 1973; Bujarrabal, Guibert, and Balkowski, 1981; Efstathiou and Fall, 1984), and quasars (Mittaz, Penston, and Snijders, 1990; Francis et al., 1992; Boroson and Green, 1992). It has also been applied to imaging studies of the interstellar medium (Heyer and Peter Schloerb, 1997; Brunt, Heyer, and Mac Low, 2009) and to the analysis of X-ray binaries (e.g., Malzac et al., 2006; Koljonen et al., 2012; Koljonen, 2015), blazars (Gallant, Gallo, and Parker, 2018), and symbiotic stars (Danehkar, Drake, and Luna, 2024). A notable area of use is the study of X-ray variability in Active Galactic Nuclei (e.g., Vaughan and Fabian, 2004; Miller, Turner, and Reeves, 2008; Parker et al., 2014b; Parker et al., 2014a; Gallo et al., 2015; see also a review in Danehkar, 2024).

Beyond astrophysical research, PCA is a fundamental tool in machine learning for data preprocessing, visualization, or unsupervised classification (e.g., Bishop, 2007; Raschka, 2015; Müller and Guido, 2016; Witten et al., 2017; Géron, 2019).

This Chapter introduces PCA as a complementary technique to the framework discussed in Chapter 2. While MST captures the topological relationships within the data, PCA helps identify the dominant physical variables, enabling a more comprehensive interpretation of the problem.

3.2   Understanding PCA

The core objective of PCA is to reduce the number of variables, of a dataset while preserving as much as possible its variance. Variance quantifies the extent to which data spreads around its mean; it's calculated as the average of the squared deviations from the mean and acts as a proxy for the information content in the dataset. Keep in mind that variance is not uniformly distributed across variables; some directions in the feature space exhibit more variability than others. PCA also aims to clarify the contribution of each original variable to the variance. To understand how PCA accomplishes this, it is helpful to approach it from two complementary perspectives:

- A geometric perspective explains PCA as a projection and rotation in a new coordinate system.

- A mathematical perspective shows how this projection is derived from an optimization problem involving the covariance matrix.

3.2.1   Geometric Interpretation of PCA

PCA finds new axes (principal components, PCs) for the data space such that the first axis (PC_1) points in the direction of maximum variance. Each subsequent axis (PC_2, PC_3, ...) is orthogonal to the previous ones and captures the following highest remaining variance. The maximum number of PCs that can be obtained is limited by the number of variables. Figure 3.1 helps to explain why PCA seeks to maximize variance: it ensures the projected data best preserves the structure of the original dataset. The process consists of projecting the data points onto these new axes. To preserve the structure of the original data set as much as possible, the distance from the projected point to the original point should be minimized. This quantity is referred to as the reconstruction error (hereafter error). Therefore, this new axis, denoted by the unit vector $\vec{v}$, will minimize the sum of all squared errors across all data points. On the other hand, the variance along $\vec{v}$ is equal to the average distance squared of the projection of the points to the center of the new axis. This can be interpreted using the Pythagorean theorem:

$$\|\vec{x}\|^2 = \|\vec{z}\|^2 + \|\vec{x} - \vec{z}\|^2, \tag{3.1}$$

where $\vec{x}$ is an original data point, $\vec{z} = (\vec{v}^T \vec{x})\vec{v}$ is its orthogonal projection onto $\vec{v}$, and $\vec{x} - \vec{z}$ is the error. Since $\vec{x}$ is constant, minimizing $\|\vec{x} - \vec{z}\|^2$ is equivalent to maximizing $\|\vec{z}\|^2$, which corresponds to maximizing the variance along the direction along $\vec{v}$. In other words, the new axis, which captures the maximum variance in the data, is analogous to the PC; it is the axis that least deforms the original data.

3.2.2   Mathematical Interpretation of PCA

The variance of a projection in direction $\vec{v}$, which is a unit vector, is given by:

$$\text{Var}(\vec{v}) = \frac{1}{n} \sum_{i=1}^{n} \|\vec{z}_i\|^2 = \frac{1}{n} \sum_{i=1}^{n} (\vec{v}^T \vec{x}_i)^2 = \frac{1}{n} \|\tilde{X}\vec{v}\|^2 = \frac{1}{n} \vec{v}^T \tilde{X}^T \tilde{X} \vec{v} = \vec{v}^T C \vec{v}. \tag{3.2}$$

Formally, $\tilde{X} \in \mathbb{R}^{n \times N}$ is the mean-centered data matrix, where n represents the number of data points, rows in $\tilde{X}$, and N represents the variables of the data, columns in $\tilde{X}$. Note $\|\vec{z}_i\|^2 = (\vec{v}^T \vec{x}_i)^2$ represents the variance contribution of data point $\vec{x}_i$ in $\vec{v}$. In addition, the covariance matrix (C) of the dataset is defined as:
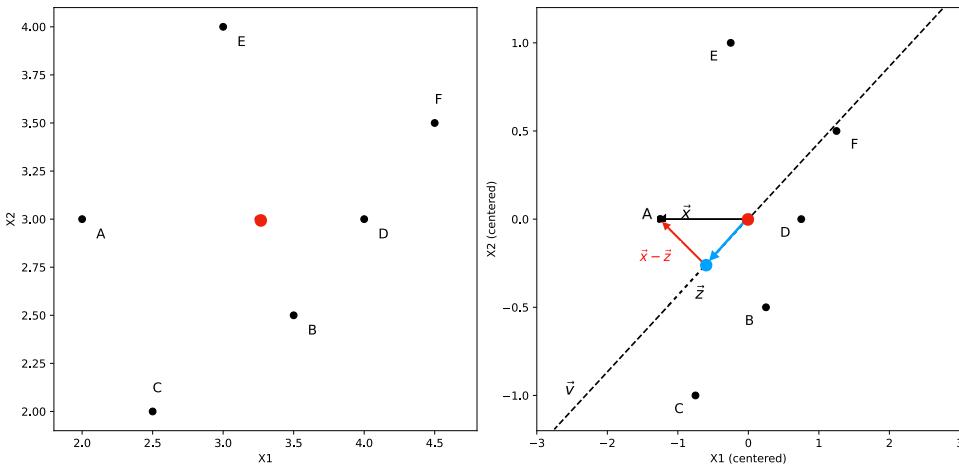
FIGURE 3.1: Left panel. Data points in a feature space of two variables, x1 and x2. The red dot represents the mean of the data. Right panel. Data points centered (mean subtracted) on the red dot (origin), illustrating the PCA geometry considering Equation (3.1). The data point (A) represented by $\vec{x}$ is projected onto the direction of maximum variance ($\vec{v}$), resulting in the projection $\vec{z}$. The squared distance from $\vec{z}$ to the origin is the variance contribution of that point in $\vec{v}$; the distance to the axis is the error ($\vec{x} - \vec{z}$).

$$C = \frac{1}{n-1}\tilde{X}^T\tilde{X}\,. \tag{3.3}$$

As a note, while PCA is traditionally applied to mean-centered data, it can also be performed using median-centered variables. This approach is often more robust when the data distribution is non-Gaussian or has skewed tails, as the presence of outliers can affect the mean as a proper measure of central tendency. The chosen type of normalization will manage this situation; for instance, data standardization involves subtracting the mean and then dividing by the standard deviation of the data. However, scalarization consists in subtracting the median and dividing by the range between the first and third quartiles (see de Amorim, Cavalcanti, and Cruz, 2022 or the following Chapters, 4, 6, and 7, for an idea about this discussion in real cases). The directions of the PCs are found by solving the following constrained quadratic optimization problem, considering the maximization of Equation (3.2):

$$\max_{\vec{v}} \quad \vec{v}^T C \vec{v} \quad \text{subject to} \quad ||\vec{v}|| = 1\,. \tag{3.4}$$

Introducing a Lagrange multiplier λ in Equation (3.4), we define:

$$\mathcal{L}(\vec{v}, \lambda) = \vec{v}^T C \vec{v} - \lambda(\vec{v}^T \vec{v} - 1)\,. \tag{3.5}$$

Taking the derivative with respect to Equation (3.5) and setting it to zero leads to the eigenvalue decomposition (EVD, see Section 3.3):

$$\nabla_{\vec{v}}\mathcal{L} = 2C\vec{v} - 2\lambda = 0 \implies C\vec{v} = \lambda\vec{v}\,. \tag{3.6}$$

Equation (3.6) shows that the PCs are the eigenvectors ($\vec{v}$) of C defined by Equation (3.3). At the same time, the amount of variance captured by each $\vec{v}$ is given by

the corresponding eigenvalue (λ).

The transformation of the dataset seen through $\tilde{X}$ into the PC plane is:

$$PC_m = \vec{v}_m \tilde{X}^T \,, \tag{3.7}$$

where each resulting PC is a linear combination of the original variables. The result of this projection, the PC scores, represents the coordinates of the original data points in the new basis defined by the PCs. Moreover, as shown in Equation (3.7), the PCs are uncorrelated because they are obtained by projecting the dataset onto the $\vec{v}_s$ of C, which are orthogonal by construction. In addition to measuring the variance that each PC retains, we define the explained variance:

$$\text{Explained Variance}(PC_m) = \frac{\lambda_m}{\sum_{j=1}^{N} \lambda_j} \,. \tag{3.8}$$

Through the Equation (3.8), we can establish an order of importance of the PCs by quantifying the dimensionality reduction.

It is essential to note that PCA fundamentally assumes linearity in the data, as outlined in Equation (3.7). Specifically, it identifies directions that are linear combinations of the original variables and maximizes variance along those directions. This assumption works well when the variables are linearly correlated, but may be limiting when relationships are nonlinear. For instance, in the case of pulsars (see Chapter 4), variables derived from the dipole model, such as the characteristic age, magnetic field at the surface, and spin-down luminosity, are nonlinear functions of the spin period and its spin period derivative. However, logarithmic transformations make these relationships linear. On the other hand, PCA may not fully capture the data's variance when non-linearity is present, even if statistical dependencies exist. Nevertheless, at that point, see, e.g., Chapters 6 and 7, PCA can be employed primarily as an exploratory method to uncover patterns, reduce dimensionality, and assess the relative contributions of variables.

3.2.3 Visualizing Principal Components

Once the PCs have been computed, several standard plots are used to visualize and interpret their usefulness and relevance. These visualizations are crucial for understanding how many PCs to retain and how the original variables contribute to each PC. The most common visual tools include:

Explained Variance Plot (Scree Plot).   This plot displays the variance each PC explains according to Equation (3.8). The λ_s (or the squared singular values σ_s^2, see the following section for further details) are plotted in descending order. It helps identify how many PCs account for most of the variance.

Cumulative Explained Variance.   The cumulative sum of explained variance seen in Equation (3.8) is plotted as a function of the number of PCs. This plot illustrates the number of PCs required to retain a specified percentage of the total variance in the data.

Variable Contribution Heatmap.   A heatmap of the PCA application corresponds to the coefficients (or "weights") of the original variables, also known as loadings,

which indicate the extent to which each original variable contributes to each PC. Each row represents a variable, and each column represents a PC. This visualization highlights the variables that contribute most to each PC.

Plane of Principal Components (PC plane). A PC plane is a scatter plot that displays the PC scores in the space defined by the selected PCs.

As examples of these visual tools, see Figures 4.2, 4.3, 6.3, and 7.2, generated on real data of pulsars, millisecond pulsars, and fast radio bursts, respectively.

3.3 Methods for Principal Components Computation

The optimization approach seen in Equation (3.6) leads to an eigenvalue problem whose solution yields the PCs and their associated variances, denoted by $\vec{v}_s$ and λ_s, respectively. However, different mathematical strategies can be used to apply the PCA. The most commonly used methods are the EVD and the Singular Value Decomposition (SVD), which are mathematically equivalent in their outcomes but differ in their computational approaches. In Appendix B, we show the application of both methods for the same example.

3.3.1 Eigenvalue Decomposition

This method seen in Equation (3.6) is applied directly to C considering Equation (3.3), giving rise to Equation (3.7). This approach is conceptually straightforward and well-suited for datasets with a relatively small N. However, for high-dimensional data where N is large or comparable to n, working directly with C can make this method computationally intensive. Alternatives, such as SVD, which do not strictly deal with C as shown below, are preferred for computational efficiency, numerical precision, and stability.

3.3.2 Singular Value Decomposition

The SVD method computes the PCs directly from $X \in \mathbb{R}^{n \times N}$, factorizing it as:

$$X = U\Sigma V^T . \tag{3.9}$$

Equation (3.9) shows $U \in \mathbb{R}^{n \times n}$ that contains the left singular vectors, $\Sigma \in \mathbb{R}^{n \times N}$ which is a diagonal matrix with non-negative singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_N \geq 0$, and $V \in \mathbb{R}^{N \times N}$ that contains the right singular vectors. The columns of V correspond to the $\vec{v}_s$ of C. To see how this connects with the PCA, consider Equation (3.3), where introducing Equation (3.9), and since $U^T U = I$ due to U is orthogonal, we get:

$$X^T X = V\Sigma^T \Sigma V^T \quad \Rightarrow \quad C = \frac{1}{n-1} X^T X = V \left(\frac{\Sigma^T \Sigma}{n-1} \right) V^T , \tag{3.10}$$

which is the eigendecomposition of C. Thus, the λ_s of C, considering $\left(\frac{\Sigma^T \Sigma}{n-1} \right)$ in Equation (3.10), can be obtained as:

$$\lambda_m = \frac{\sigma_m^2}{n-1} . \tag{3.11}$$

Finally, the left singular vectors U can be computed multiplying both sides of Equation (3.9) by V, taking $V^T V = I$ by orthogonality, obtaining:

$$XV = U\Sigma \, . \tag{3.12}$$

Equation (3.12) corresponds to projecting the original data onto the $\vec{v}_s$, where each row of $U\Sigma$ corresponds to the coordinates of a data point in the PC plane.

3.4 Conclusions

PCA offers a rigorous and intuitive framework for reducing dimensionality and analyzing the internal structure of multivariate data. The following chapters apply PCA to datasets that include physical and observational parameters from various astrophysical sources, such as pulsars, millisecond pulsars, and fast radio bursts.
In Chapter 4, we apply PCA to the pulsar population to unravel whether the intrinsic properties, spin period, and spin period derivative, are sufficient to determine how they relate. Additionally, we observe a significant reduction in the number of necessary variables (now, just two PCs) to collect the whole variance without losing information, even when the population is defined by up to eight variables, enabling manageable visualization. This feature space is extended to define the population of millisecond pulsars in Chapter 6. We can observe the contributions of binary parameters versus intrinsic properties through PCA, which helps in analyzing clustering methods. In Chapter 7, we utilize the PCs derived from the PCA application over the population of fast radio bursts, described by both observed and derived properties, as an alternative to these. Using population separation methods, we investigate whether PCs can reveal hidden patterns that distinguish classes of fast radio bursts.

Chapter 4

Visualizing the pulsar population using graph theory

4.1 Introduction

Ever since the discovery of the first pulsar (Hewish et al., 1969), the $P\dot{P}$ diagram, has been used to summarize our knowledge and guide our research on the pulsar population. Classes of pulsars and possible links among them are referred to in this diagram, as it is what we know about their possible evolutionary tracks along the pulsar's lifetime (see, e.g., Enoto, Kisaka, and Shibata, 2019 for a review). However, we have looked at the same diagram for over five decades, so a fresh appraisal may be healthy. Is the $P\dot{P}$ diagram the most practical or complete way to visualize the pulsars we know? Does it introduce any unwarranted bias on what we consider to be similar pulsars?

In this chapter, building on the work presented in García, Torres, and Patruno (2022), we apply PCA (see Chapter 3) to show that even when all the variables of the rotating dipole model used to describe a pulsar depend on P and $\dot{P}$, the population's variance is not fully captured by these two physical properties alone. Therefore, any classification or visualization of similarity based solely on P and $\dot{P}$ is potentially misleading. To address this multidimensional problem, which arises from the need to utilize additional properties to complete our vision, we introduce the MST (see Chapter 4), also referred to as the Pulsar Tree, and discuss its potential applications. It is computed based on a properly normalized distance that measures the closeness between pulsars.

The Pulsar Tree hosts information about pulsar similarities that go beyond P and $\dot{P}$ and are thus naturally challenging to read from the $P\dot{P}$ diagram. In addition, we shall introduce an online tool that encompasses all our results and allows users to focus on user-defined problems.

4.2 The pulsar variance

4.2.1 Variables definition

We consider pulsars listed in the Australia Telescope National Facility catalog, version 1.67 (Manchester et al., 2005) (ATNF v1.67, as of March 2022), which includes 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 considered. The number of pulsars listed in ATNF v1.67 is 3282, of which 2509 have a known spin period and spin 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 chapter will be applied to this set

without distinction. For characterizing the pulsar population and ultimately defining a distance from one pulsar to another to establish a relationship, we consider the following physical set of pulsar variables (see Section 1.2.2)[1]:

- Spin period:
  P [s],

- Spin period derivative:
  $\dot{P}$ [s s^{-1}],

- Surface magnetic flux density (equator):
  $B_s = (3c^3I)^{1/2}/(8\pi^2R^6\sin^2\alpha)^{1/2}\sqrt{P\dot{P}} \simeq 3.2 \times 10^{19}P^{1/2}\dot{P}^{1/2}$G,

- Magnetic field at the light cylinder:
  $B_{lc} = B_s(\Omega R)^3/c^3 \simeq 3 \times 10^8 P^{-5/2}\dot{P}^{1/2}$G,

- Spin-down energy loss rate:
  $\dot{E}_{sd} = 4\pi^2I\dot{P}P^{-3} \simeq 3.95 \times 10^{46}P^{-3}\dot{P}$ erg s^{-1},

- Characteristic age:
  $\tau_c = P/2\dot{P} \simeq 15.8 \times 10^{-9}P\dot{P}^{-1}$yr,

- Surface electric voltage:
  $\Delta\Phi = (B_s4\pi^2R^3)/(2cP^2) \simeq 2.1 \times 10^{20}P^{-3/2}\dot{P}^{1/2}$V,

- Goldreich-Julian charge number density:
  $\eta_{GJ} = (\Omega B_s)/(2\pi ce) \simeq 2.21 \times 10^{18}P^{-1/2}\dot{P}^{1/2}$cm^{-3}

As a summary of the above, 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 and e) have their usual meanings. 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 fundamental magnitudes that characterize the energetics and magnetospheres of pulsars. In our set, others suggested that we incorporate the idea that dissimilar pulsars (e.g., those with millisecond and regular periods) can have similar magnetospheres. The magnetic field at the light cylinder partly describes this, which may be similar for both. The voltage gives the potential drop between the magnetic pole and the edge of the polar cap. It is thought to represent the variety introduced by the electromagnetic configuration, for which another parameter of interest is the Goldreich-Julian charge number density, $\propto B_s/P$. Note that these magnitudes, being all functions of P and $\dot{P}$, can have a relationship between themselves, as just noted. Ideally, the mass and radius of the neutron stars, would also be considered variables of interest in our study. However, this information is only available for a tiny 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 timescale. However, these are not known for most pulsars in our sample. They all depend on the unknown (except for a few) pulsars' actual age (for which the characteristic

---

[1]In García, Torres, and Patruno (2022), the Equations for the τ_c, $\Delta\Phi$, and η_{GJ} contain incomplete constant factors. However, their dependencies on P and $\dot{P}$ were correctly stated, and all numerical computations used the proper expressions.
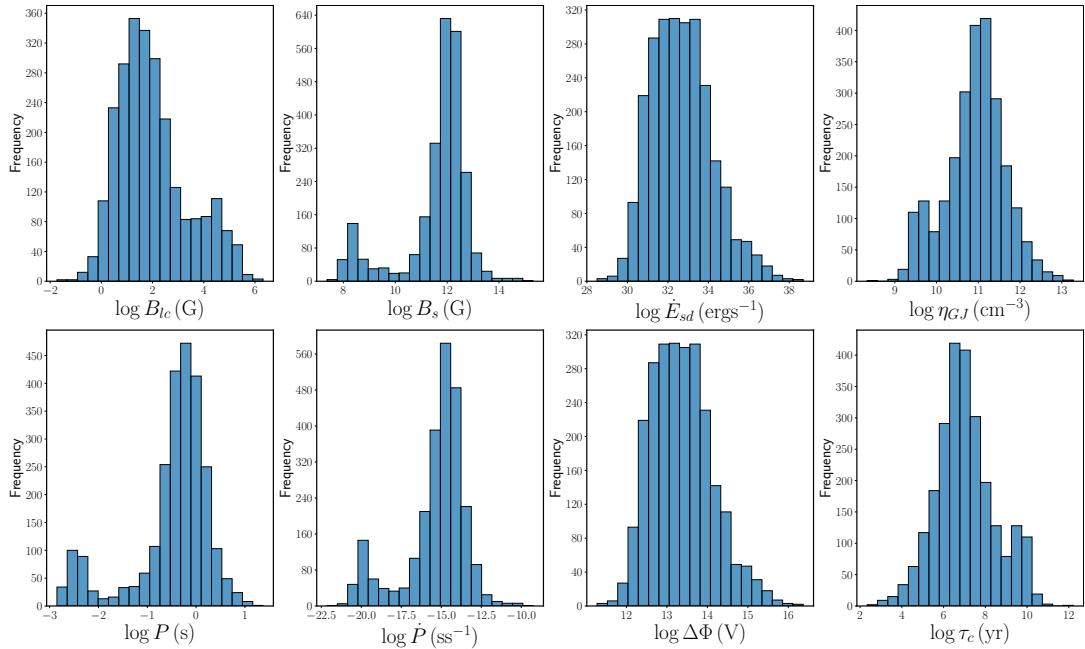
FIGURE 4.1: Distribution of the logarithm of the eight variables used
for the complete set of 2509 pulsars.

age τ_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 $\dddot{P}$, would dramatically reduce the sample size. Finally, other measurable quantities exist 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) would also significantly reduce the sample, as it would be affected by extrinsic conditions (absorption, distance), and/or would incorporate parameters that are difficult to compare for the population as a whole.

4.2.2 Treating variables

The magnitudes' values may differ by several orders of magnitude for different pulsars, so we consider the logarithm of them to mitigate this as much as possible. The distributions of the logarithm of these variables are shown in Figure 4.1.

These distributions are not normal (as is the case for the set of original variables, without logarithms). Two clear populations of millisecond and normal pulsars appear separately in all variables except those related to the $\dot{E}$ and $\Delta\Phi$. As is evident 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*, e.g., see de Amorim, Cavalcanti, and Cruz (2022), for scaling the log variables,

$$x_i^\dagger = \frac{x_i - Q_2}{IQR} \ . \tag{4.1}$$

The †-symbol represents that the quantity x_i has been scaled, Q_1, Q_2 and Q_3 represent the 1st quartile, median, and 3rd quartile of the distribution, respectively, with
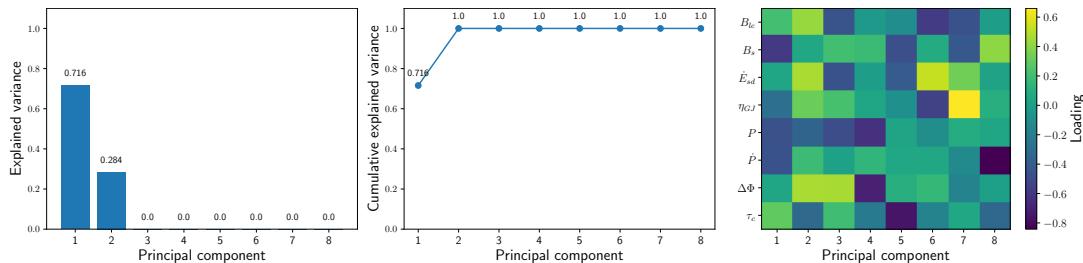
FIGURE 4.2: PCA results for the logarithm of the set of variables for the population of 2509 pulsars, see Section 3.2.3 for further explanations about PCA definitions. The left panel, referred to as the scree plot, displays each PC's explained variance based on the eigenvalues of the covariance matrix. Note that the covariance matrix calculates the relationships between pairs of variables, showing how changes in one variable are associated with changes in another. It represents the amount of information contained in each PC. The central panel shows the cumulative explained variance by the new variables defined through the PCA analysis. The right panel shows the "weight", also called loading, each variable has concerning each PC, indicating its contribution to the variance captured by that PC. This value is the coefficient held in each eigenvector of the covariance matrix. Negative values imply that the variable and the PC are negatively correlated. Conversely, a positive value shows a positive correlation between the PC and the variable.

the IQR as $(Q_3 - Q_1)$. After being scaled, the variables' distributions have a median equal to zero and an IQR equal to one.

Note that if we take the logarithm of the variables and then apply Equation (4.1) to scale them, the relations between variables are more clearly revealed. For instance, $\dot{E}_{sd}$ and $\Delta\Phi$, lead to $\log \dot{E}_{sd}^{\dagger} = \log \Delta\Phi^{\dagger}$, that is also visible in the corresponding distributions of Figure 4.1. Considering both at once in defining the nearness of two given pulsars is adopted to represent that the physical meaning of the two original magnitudes is different.

4.3 Principal components analysis

PCA is especially suitable for identifying the main factors that introduce variance in a population, particularly when the variables involved are linear (see Section 3.2.2 for further explanations).

In this case, since six of the variables that describe the intrinsic properties of the pulsar population (see Section 4.2.1) are derived from P and $\dot{P}$, one might intuitively conclude that two PCs would suffice to describe the population's variance. However, Figure 4.2 illustrates that the PCs identified to reproduce the whole variance, PC_1 and PC_2, are not simply P and $\dot{P}$ themselves, but rather linear combinations of all the variables.

In other words, the variance of the population is not entirely contained within the individual variances of P and $\dot{P}$ alone.

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 the same.

The first two PCs, containing 71.6% and 28.4% of explained variance (see left panel of Figure 4.2), respectively, can be seen as (see Equation (3.7)):

$$PC_1 = \ 0.21 B^{\dagger}_{lc_l} - 0.29 \eta^{\dagger}_{GJ_l} + 0.05 \Delta \Phi^{\dagger}_l + 0.05 \dot{E}^{\dagger}_{sd_l} - 0.46 \dot{P}^{\dagger}_l$$
$$- 0.59 B^{\dagger}_{s_l} - 0.47 P^{\dagger}_l + 0.29 \tau^{\dagger}_{c_l}$$
$$PC_2 = \ 0.43 B^{\dagger}_{lc_l} + 0.32 \eta^{\dagger}_{GJ_l} + 0.47 \Delta \Phi^{\dagger}_l + 0.47 \dot{E}^{\dagger}_{sd_l} + 0.19 \dot{P}^{\dagger}_l$$
$$+ 0.05 B^{\dagger}_{s_l} - 0.36 P^{\dagger}_l - 0.32 \tau^{\dagger}_{c_l} \ . \tag{4.2}$$

In Equation (4.2) the variables are scaled following Equation (4.1), the sub-index l shows the logarithm of them. We note that all physical properties have a non-zero loading associated with them, as shown in the right panel of Figure (4.2).
After some computations (see Appendix C.1 for more details), Equation (4.2) can be reformulated as:

$$PC_1 = -8.471 - 1.178 \log P - 0.832 \log \dot{P}$$
$$PC_2 = 14.182 - 2.931 \log P + 1.105 \log \dot{P} \ . \tag{4.3}$$

Equation (4.3)[2] corroborates the direct non-equivalence between P and $\dot{P}$, and PC_1 and PC_2, as discussed above. These equations also indicate that several variables are equally important (with similar loadings) in defining the PCs.
The left panel of Figure 4.3 shows the pulsar population in the $P\dot{P}$ diagram together with lines representing equal values of the PCs, i.e., constant PC scores. The right panel of Figure 4.3 shows the same pulsars but directly in the $PC_1 PC_2$ plane with the PC scores derived from Equation (4.2). Note that nearness in one plane does not have the same meaning as in another. 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 Equation (4.3). This is illustrated in the second row of Figure 4.3, where the difference in relative distances can be as much as a factor of 3 or more. This change of shape advances the idea that any nearness ranking will be affected if considering the PCs instead of P and $\dot{P}$ (further comments about this can be found in Appendix C.2).

4.4 Minimum Spanning Tree of the pulsar population

Chapter 2 presents all the concepts needed to understand and compute an MST, and we take this for granted in what follows. We define a Euclidean distance (see Equation (2.1)) using the eight-scaled (via Equation (4.1)) logarithm of the variables introduced in Section 4.2.1. Equivalently, after analyzing the data in Section 4.3, we can use just PC_1 and PC_2, which contain the population variance in the ATNF v1.67. Both choices produce the same MST (and thus, the results from the study that follows are the same), but the latter is less demanding due to the reduced dimensionality of the problem.
With the Euclidean distance defined over the population, we first obtain a complete, undirected, and weighted graph $G(V, E) = G(2509, 3146286)$. From that, we calculate the MST, $T(2509, 2508)$, which we call the Pulsar Tree, which is shown in Figure 4.4.

---

[2]Note that no dag marking is needed, as the corresponding IQR and median of each variable is absorbed into the coefficients, and that the units of P and $\dot{P}$ are as in Equation (4.2).
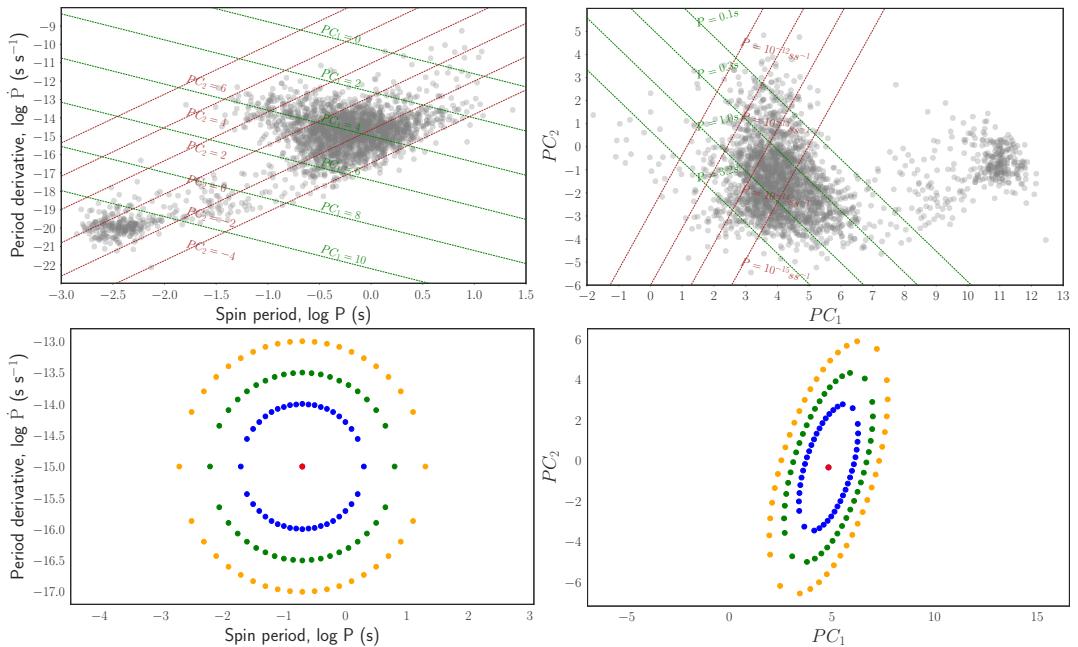
FIGURE 4.3: Top-Left panel: $P\dot{P}$ diagram of the pulsar population. Constant values of PC_1 and PC_2 are shown in green and brown lines, respectively, according to Equation (4.3). Top-right panel: PC_1PC_2 plane of the pulsar population containing the PC scores according to Equation (4.2).
Constant values of P and $\dot{P}$ in green and brown lines, respectively, according to Equation (4.3). Bottom-left panel: Synthetic pulsars positioned circularly at different radii from a given center in the $P\dot{P}$ diagram. Bottom-right panel: Transformation of the circle into PC_1PC_2 plane through the Equation (4.3).
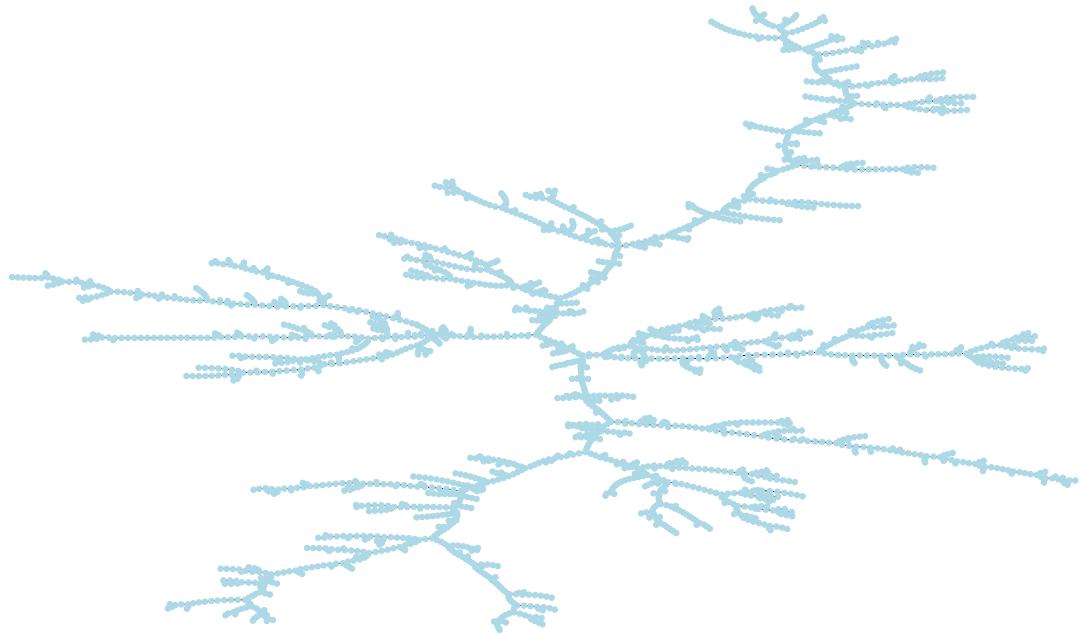
FIGURE 4.4: A different look at the population. MST, $T(2509, 2508)$, based on the complete, undirected, and weighted graph $G(2509, 3146286)$ for the 2509 pulsars (each node represents a pulsar) from the ATNF v1.67 and their full combination of weights computed from their Euclidean distance among eight scaled variables (or the equivalent 2 PCs). Each node in the MST represents a pulsar.

4.4.1 Branch analysis and pulsar classification in the MST

As shown in Figure 4.5, mixing nodes from different branches of the MST produces a scattered distribution of variables. This generic behavior occurs when mixing branches in the MST and any blending of nodes, even within a path (see the panels in the last three rows of Figure 4.5). If we read the MST in a disordered manner, nothing is learned from it. Instead, Figure 4.6 shows that if we choose one of the paths at a time and run along with the nodes in it in an orderly manner, a smooth behavior of the variables naturally appears. Mixing branches or paths of the Pulsar Tree is equivalent to grouping pulsars by their nearness in the $P\dot{P}$ diagram, as shown in Figure 4.7. The Pulsar Tree hosts information that is challenging to read from the $P\dot{P}$ diagram.

4.4.2 The MST as a descriptive tool

The ordering introduced by each branch and path indicates an internal physical grouping in the MST, as shown in the variations of the variables seen in Figure 4.8. These variations illustrate the physical properties of different pulsar classes, and understanding them may reveal physical connections among pulsars or insights into their evolution. To emphasize this, we shall observe how some known groups of pulsars are located in the MST.

The main tree trunk travels from young and energetic pulsars at the bottom[3] to millisecond pulsars (MSPs) in the top (see panels -referred from left to right and top

---

[3]In what follows, we use "bottom", "top", "right", or "left" to refer to a particular position in the MST shown. However, we emphasize that (in the absence of axes) what matters is not the representation but the edges among nodes, see Appendix C.3 for a further explanation.
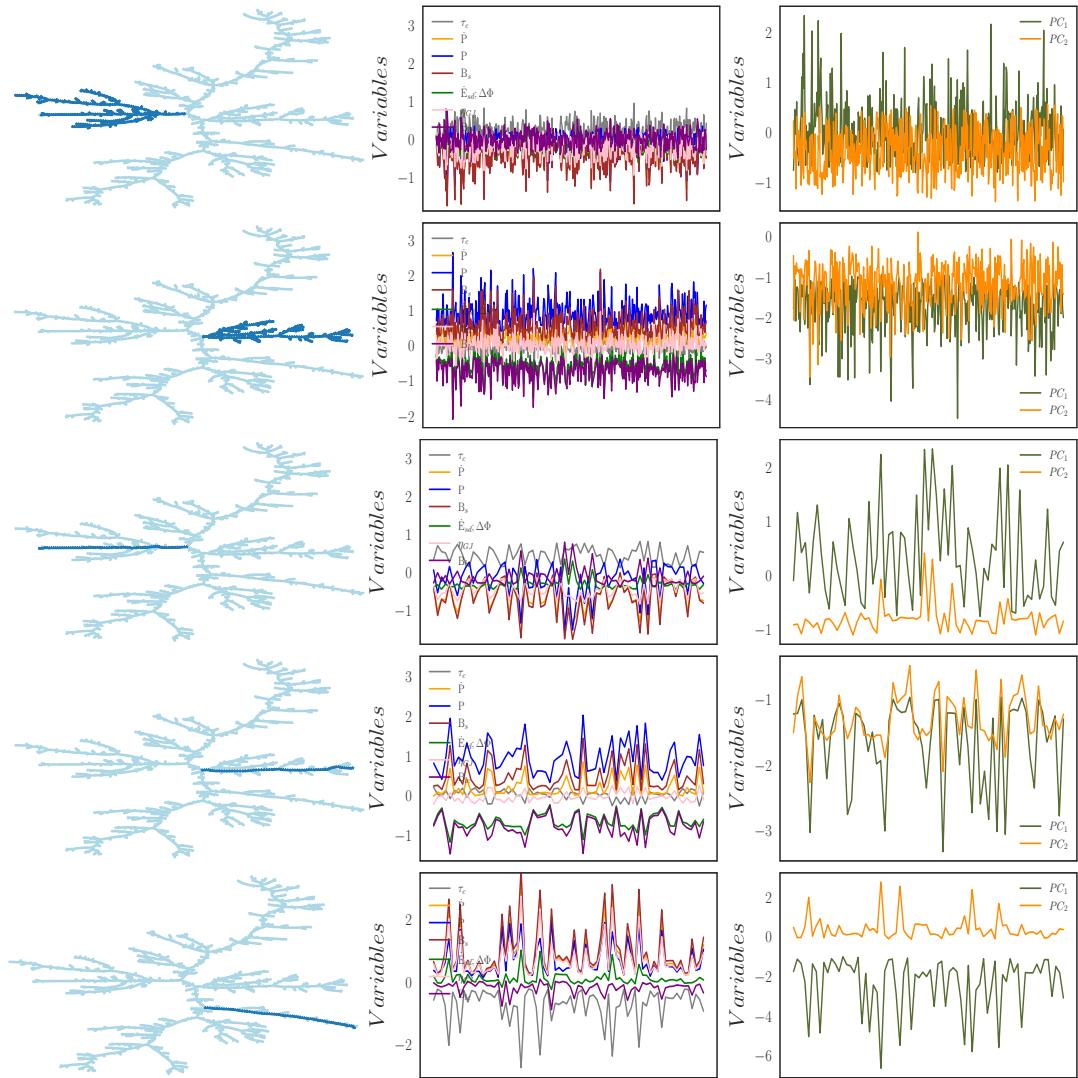
FIGURE 4.5: Top two rows: On the left, nodes from different branches are highlighted in dark blue in the MST. The middle and right panels display the behavior of the eight scaled variables and the two PCs when following an arbitrary mixing of the nodes in the highlighted part of the MST, as shown in the corresponding panel in the left column. The third, fourth, and fifth rows are taken from a unique branch, respectively; however, the nodes of the selected path are now arbitrarily mixed. For visualization purposes, we subtracted the corresponding mean for each PC in the right column for PC_1 and PC_2.
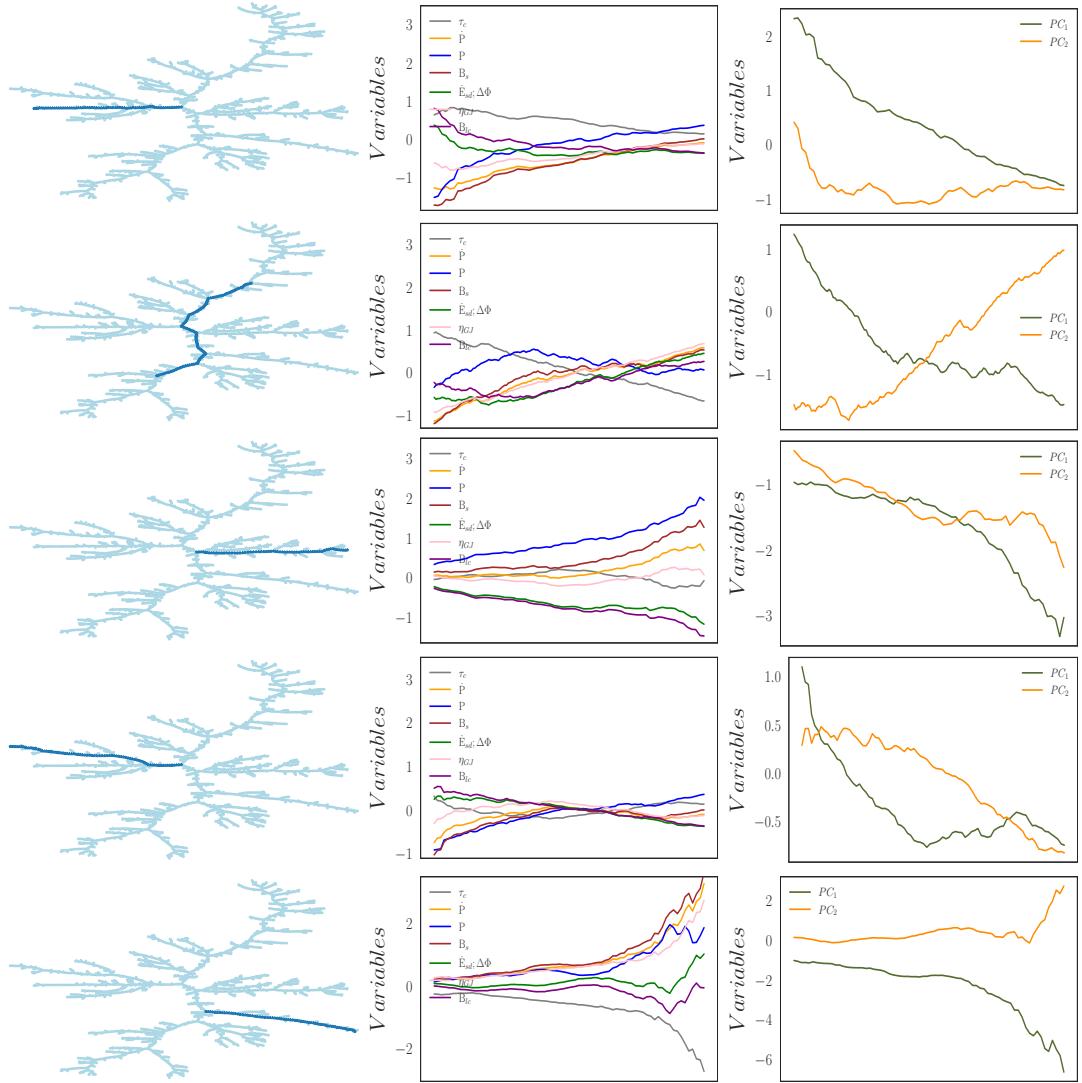
FIGURE 4.6: Similar to the panels in Figure 4.5, but following the path nodes as appearing along the marked branch of the MST in its corresponding panel in the left column. The path is read from left to right (except in those branches that appear mostly vertical, where they are read from top to bottom). The first, third, and fifth panels in the left column mark the same nodes as the third, fourth, and fifth panels, respectively, in the left column in Figure 4.5.
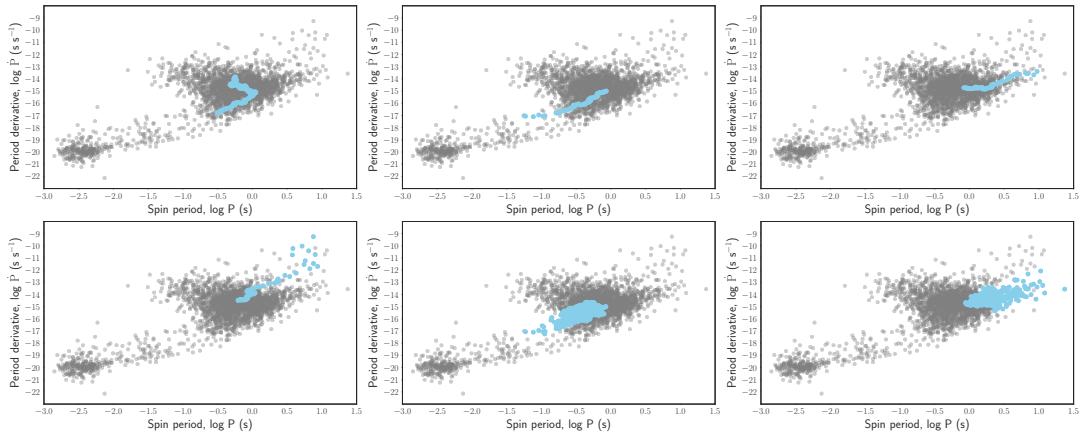
FIGURE 4.7: Positions of the branches and paths analyzed in Figures
4.5 and 4.6 according to the $P\dot{P}$ diagram. From left to right, and top
to bottom: second row of Figure 4.5; and, third, fourth, fifth, first, and
second rows, respectively, following Figure 4.6.

to bottom- first, second, and fifth of Figure 4.8). 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 delve deeper into the sub-population of
binaries, as they are heavily influenced by the evolutionary processes that occur for
MSPs during the accreting (recycling) phase. As an example, consider BWs, RBs,
and tMSPs, see Papitto and Bhattacharyya (2022) for reviews, and also Swihart et al.
(2022), Koljonen and Linares (2023), Freire et al. (2017), Lynch et al. (2012), Douglas
et al. (2022), and Strader et al. (2019) for identifying the sources. As their physical
properties are similar, no clear grouping appears for these binary sub-samples (see
Figure 4.9). In Chapter 6, we explore supplementing intrinsic variables with others
that represent the companion, orbital parameters, and the environment of binary
pulsars to address these issues.

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

For instance, the most extended rightward branch (moving along it towards the
rightmost node) groups pulsars with increasing τ_c, P, and $\dot{P}$. Pulsars in this branch
are not particularly energetic, nor have significant B_{lc}; instead, they have an increased B_s, reaching extreme values. The second-longest rightward branch has similar behavior but is formed by less magnetized objects at their B_s and is also less
energetic, slower, and older. Not surprisingly, magnetars and the only XDIN, J1856-
3754, quoted in the ATNF v1.67, are located in both branches (see Olausen and Kaspi,
2014; Coti Zelati et al., 2018). Figure 4.9 shows this in more detail. 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 behavior, appear pretty separate from the rest. The low-field magnetars depicted in the MST having $P < 1$ s
are J1846-0258 (Gavriil et al., 2008) and J1119-6127 (Archibald et al., 2016b), and they
appear in Figure 4.9 in the bottom leftmost branch. This location differs from that of
J2301+5852, J1647-4552, J1822-1604, and J0418+5732, the other low-field magnetars,
which are located on the second-longest rightward branch above the magnetars.

The different branches at the bottom of the MST contain all energetic pulsars. Despite being very different in almost every aspect, they share the same B_{lc} as MSPs at
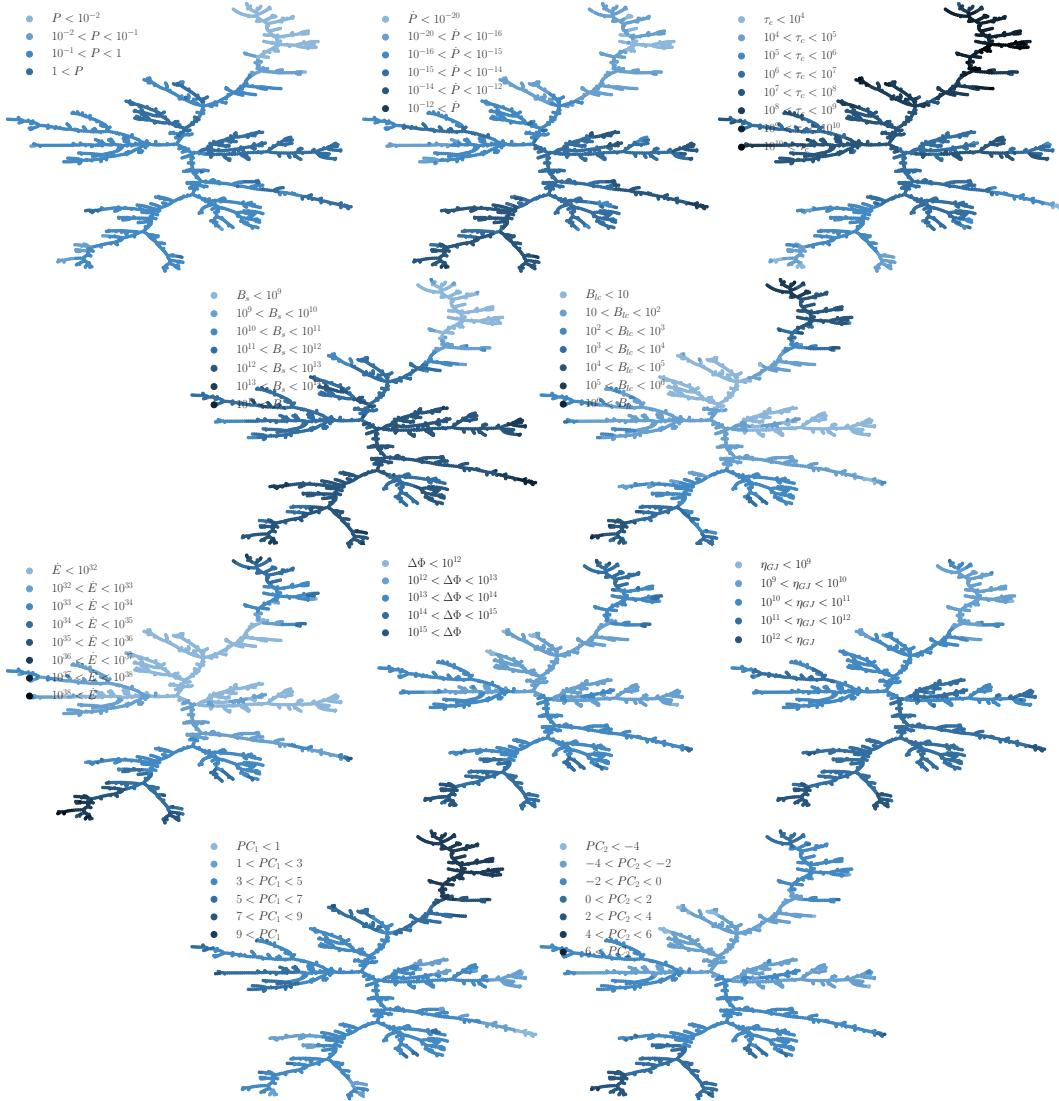
FIGURE 4.8: Representation of the variables considered directly onto the MST. The first row shows P, $\dot{P}$, and τ; the second row shows the magnetic field at the surface and at the light cylinder; the third row shows the spin-down power, the surface electric voltage, and the Goldreich-Julian current, and finally, the last row shows directly the PC_1 and PC_2.
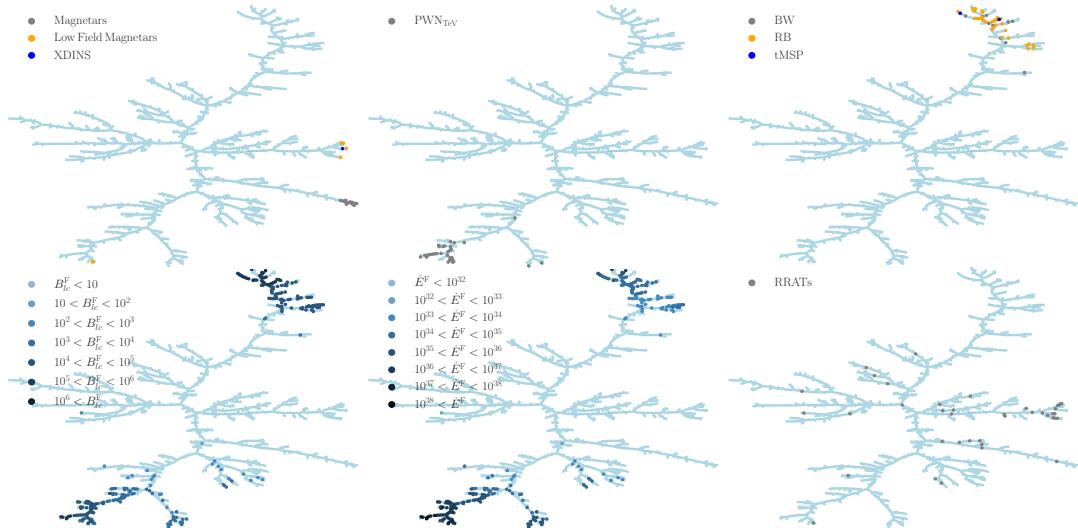
FIGURE 4.9: Representation of different pulsar classes on the MST. The first row shows magnetars, low field magnetars, X-ray Dim Isolated Neutron Stars (XDINS), pulsars associated with TeV pulsar wind nebulae (PWN in TeV), redbacks (RB), black widows (BW), and transitional millisecond pulsars (tMSPs). The first two panels of the second row show the *Fermi*-LAT pulsars (thus the superscript F), further characterized by their magnetic field at the light cylinder and spin-down power in radio, as depicted in Figure 4.8. The third panel in the second row shows the Rotation Radio Transients (RRATs).

the top of the MST. The small branches in which the energetic pulsars divide at the bottom of the plot also separate them into those having a significantly higher value of different variables, like $\dot{E}$, B_{lc}, B_s. Figure 4.9 consistently shows how the pulsars associated with TeV confirmed or candidate pulsar wind nebulae (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 alone cannot judge the reality of the proposed association, but it emphasizes how distinct these two are from the rest (see Section 4.4.3 for more details about the MST in this context). The case of CTA 1 has been studied in detail as a possible PWN in the reverberation phase and/or due to having a higher magnetization Martín, Torres, and Pedaletti (2016); see also Torres et al. (2014). Its peculiarity has already been noted from a physical standpoint. However, it is less of an outlier than 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 v1.67 fall on it (see Abdo et al., 2013b; Fermi-LAT Collaboration, 2021). Figure 4.9 shows two panels to this effect, where B_{lc} and $\dot{E}$ are noted. The detected gamma-ray pulsars cluster at specific locations, with the most empty of gamma-ray emission, corresponds to the fact that gamma-ray pulsars have $B_{lc} > 100$ G, with most having $B_{lc} > 10^3$ G, and relatively high $\dot{E}$. Some other magnitudes are less decisive, e.g., there are gamma-ray pulsars across the full range of η_{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 suggests

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 $B_{lc} < 100$ G. This pulsar has been noted by Smith et al. (2019) as having a $\dot{E}$ ($\sim 8 \times 10^{32}$ erg s^{-1}) about three times lower than the previously observed gamma-ray emission death-line. The outlier *Fermi*-LAT pulsar in the left part of the MST is J1231-5113 and has an even lower $\dot{E}$ ($\sim 5 \times 10^{32}$ erg s^{-1}), while its other magnitudes B_{lc}, τ, η_{GJ} are similar to the rest of the gamma-ray population.

We also investigate the position in the MST of the 40 RRATs (taken from Abhishek et al., 2022; Cui and McLaughlin, 2022) with known P and $\dot{P}$ appearing in the ATNF v1.67. RRATS are pulsars showing extreme radio variability, as most are discovered through their single, isolated pulses, and only one of the RRATs considered is located in the main trunk. This pulsar is located near the degree 3 node J1828-1336, which separates the main trunk into the two central branches.

The MST can also be used to analyze any other pulsar population, for instance, those with known glitches or the intermittent and nulling pulsar group.

The online tool, the Pulsar Tree web (see Section 4.5), provided with this work, promotes this kind of analysis.

4.4.3 The MST view on PWNe

A dedicated application of the MST methodology to the study of PWNs is presented in the work by Zhang et al. (2024). This analysis, focused on the potential detection of PWNs around PSR J1208-6238, J1341-6220, J1838-0537, and J1844-0346, provides a thorough implementation of the method, along with insightful results that validate its significance beyond population studies. Considering the Pulsar Tree, this study identifies potential TeV PWNe based on their locations, utilizing a one-zone PWN leptonic model to predict their possible energy spectrum characteristics. As illustrated in Figure 1 in Zhang et al. (2024), most pulsars with detected TeV PWNe (H. E. S. S. Collaboration et al., 2018) are positioned in the lower-left branch of the Pulsar Tree. The noteworthy quantity of TeV-detected PWNe in a specific part of the Pulsar Tree is not a result of random selection, as obtaining a non-trivial number of a particular type of pulsar is not straightforward. The MST relates equally energetic and relatively young sources that are likely to exhibit TeV PWN behavior. However, there is no correlation with the environment, or the progenitor is included in the distances on which the Pulsar Tree is computed.

4.4.4 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 based on the theoretical tracks presented in Figure 10 of Viganò et al. (2013) and individually studied where they would fall in the MST should they be part of our sample. This is shown in Figure 4.10, where the arrows show increasing age at a fixed initial magnetic field (B_P^0), and the rounded cap point in the origin of the arrows shows possible birthplaces. In agreement with what was already discussed in Figure 4.6, we find that tracks are not randomly found in the MST. There are two birth zones: at the bottom part, where we find all energetic pulsars, and at the rightmost branch, where we see the magnetars. Then, for low B_P^0, pulsars die on the main trunk of
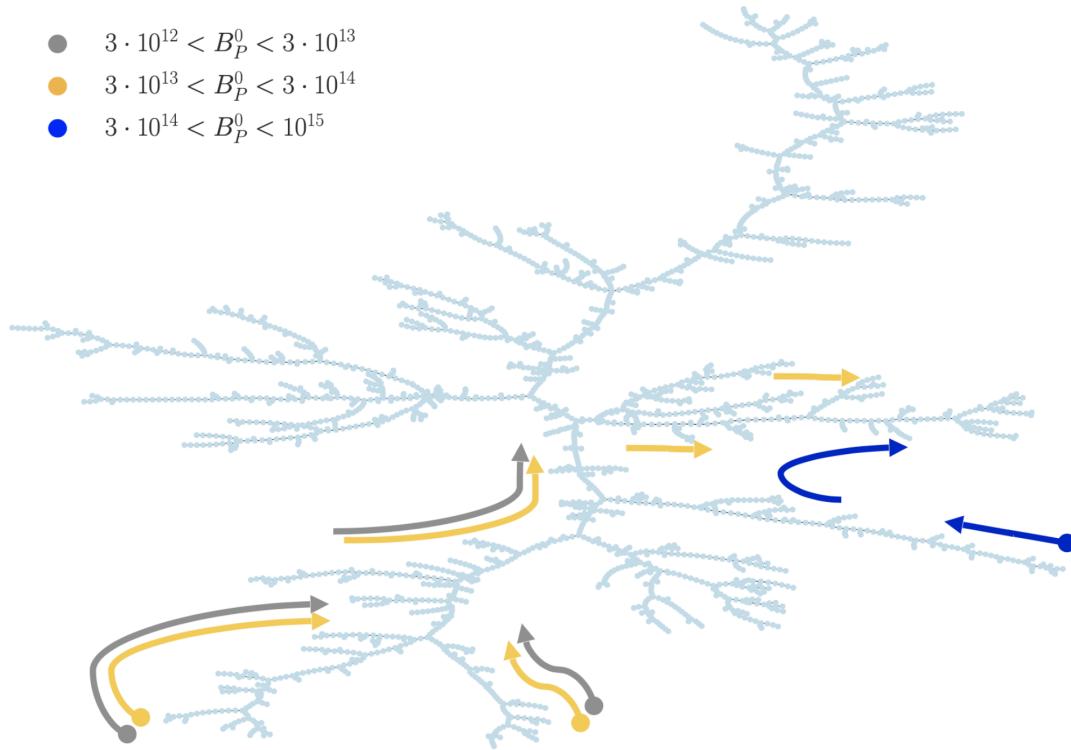
FIGURE 4.10: Representation of evolutionary tracks described in Vi-
ganò et al. (2013) into the MST, according to their initial magnetic field
(B_P^0).

the MST. The pulsars populate the middle branches when the B_P^0 increases. At the
same time, this is high enough to shift the birthplace from the energetic pulsar zone
to the classical magnetar range; the evolution is mainly confined to the two right-
most branches in Figure 4.10. In these branches, we find pulsars evolving in two
directions (from the main trunk to the extreme and vice versa) depending on their
origin.

4.4.5   The MST as an alerting tool

In addition to providing a descriptive perspective, the MST may serve as an alerting
tool for pulsars of interest (see Chapter 6). These examples consider the MST location
and the ranking of distances. 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
Appendix C.2). We use The Pulsar Tree web (see Section 4.5) to note the following:

- Based on the location of the energetic low-field magnetars J1846-0258, J1119-
  6127 at the bottom part of the MST, and due to its nearness ranking, other
  pulsars with essentially the same characteristics are noted, particularly 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). PSR J1513-5908 (in the
  composite SNR 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.

- Panel 5 of Figure 4.8 shows that few locations of the MST showing $B_{lc} \sim 100$ 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 (already detected) in the MST appear as promising potential targets. A few neighboring pulsars to the outlier J1231-5113, at the end of this branch, also show a relatively large B_{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, located at the end of the nearby branch, are of interest. Again, note that both have $B_{lc} > 10^3$ G, and $\dot{E} > 10^{33}$ erg s^{-1}.

- The detection of the radio pulsar PSR J2208+4056 by the *Fermi*-LAT despite its low $\dot{E}$ has been ascribed by Smith et al. (2019) to a possible case of favorable geometry. If this is the case, it may remain isolated in the MST, which appears close to the main trunk. Its closest neighbors (J0532-6639, J0502+4654, and J1848-0123) call for attention to test this.

- Others pulsars-of-interest regarding their possible detection in gamma-rays maybe 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, 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 of which were detected by *Fermi*-LAT). In the case of the pulsar J1550-5418, among its ten 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, with 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 compared 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.

- The pulsars J1842-0905 and J1457-5902, and J1413-6141 and J1907+0631 are the closest neighbours to the pulsar wind nebulae J1745-3040/PWN B1742-30(1) and J0007+7303/PWN CTA 1, respectively. The latter is an outlier of the PWN population (see above). Thus, their neighbors are interested in testing whether this pulsar parameter space region is prone to producing observable nebulae.

- Finally, nodes with higher degrees, particularly those connected to the main trunk or located at the ends of significant branches, are suggested for individual examination. While their positions in the MST are partly determined by the spanning nature of the tree, some of these nodes (as identified in Chapter 5) also serve as key structural points, potentially revealing subclass distinctions.

We remind that the MST is constructed solely from intrinsic pulsar properties and does not incorporate spatial position or distance. As a result, it may cluster together sources with similar emission characteristics regardless of their actual location. For instance, one of the pulsars highlighted (J0532–6639) is located in the Large Magellanic Cloud. This suggests that, despite its extragalactic origin, its emission properties could be intrinsically similar to those of some Galactic pulsars included in the analysis.

4.5   The Pulsar Tree web

The Pulsar Tree web[4] 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 regarding 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 neighboring 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; obtain tables of the distance ranking for any pulsar, and more. The continuation of the research led to the development of new functionalities for the Pulsar Tree web, specifically related to the binary pulsar population (see Chapter 6 for details).

4.6   Discussion

We have examined the pulsar population in a manner different from the usual $P\dot{P}$ diagram. Instead of considering just P and $\dot{P}$, we used a set of 8 variables as proxies for the 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 equivalent to PC_1 and PC_2. Distance ranking (or visualizations) based only on P and $\dot{P}$ may hide interesting connections and mislead our intuition. We subsequently computed the Pulsar Tree using an adequately scaled Euclidean distance, discussing its properties and how the different classes of pulsars find their ordered place within it.

Based on a more comprehensive set of variables, this more accurate source definition yields a multidimensional problem that is challenging to solve using traditional techniques. We have seen how objects such as the MST can be robust solutions within the graph theory framework. Additionally, we have demonstrated that, under suitable conditions, the MST provides a meaningful visualization of the pulsar population, reflecting the underlying physical relationships.

The MST approach offers applications beyond what we have described above. For instance, advanced analysis of the MST, including clustering, centralization, betweenness, and closeness, can illuminate physical connections and link pulsars to one another. In Chapter 5, we show a methodology for separating the MST using the betweenness centrality estimator (discussed in Section 2.4.1). The goal is to cluster the pulsar population from a graph theory context and provide a technique to predict the emergence of new classes of pulsars. 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, where the focus is on binary pulsars, see Chapter 6. Orbital parameters (orbital period, the projected semi-major axis, mass of the companion, etc.) are considered part of the distance definition. Chapter 7 demonstrates the scalability of the MST as a separation technique and its predictability when dealing with various events, such as fast radio bursts. Although the MST technique is not as widely used in astrophysics as in other fields, it can offer new perspectives in classification, clustering, source identification, and cross-correlation of source properties, including those of pulsars.

---

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

Chapter 5

Quantitative determination of minimum spanning tree structures: Using the Pulsar Tree for analyzing the appearance of new classes of pulsars

5.1 Introduction

In this chapter, based on the work García and Torres (2023), we introduce a quantitative methodology to define relevant parts of the Pulsar Tree, primarily the trunk and what we will refer to as significant branches. This method of clustering is generalizable to any spanning tree (and, of course, to any minimum spanning tree), as we will see in Chapter 6, where its application is demonstrated on an MST based on the MSPs within binary systems. In particular, we aim to introduce a qualifier to identify which parts of the MST exhibit statistically significant differences, allowing us to observe them as independent clusters. This enables the analysis of these clusters based on their underlying physical properties, which are used to calculate the MST, and helps to identify potential subpopulations. Furthermore, once we establish the trunk and significant branches, we can determine whether the resulting groups have evolved in previous iterations of the pulsar population. To do this, we start with the ATNF v1.67 pulsar population and the physical properties we used in Chapter 4 for its characterization (see Section 4.2.1 for a more detailed explanation of the variables used to calculate the Pulsar Tree and the catalog content). Note that this version contains 2509 pulsars with $\dot{P} > 0$, 2242 of which are isolated, and 267 belong to binary systems. To delimit these incarnations, we use the date label from the ATNF catalog as a filter, which refers to the date of pulsar discovery, and consider the historical evolution of the pulsar population. As a derivative of the latter study, we analyze how to determine, through defined evolutionary metrics, whether a new class of pulsars emerges in new data, future surveys, or new incarnations of pulsar catalogs.

5.2 Clustering algorithm: Significant Branches

We introduce a clustering algorithm based on significant branches using the Pulsar Tree, the MST of the pulsar population of the ATNF v1.67, as seen in Figure 4.4. Its purpose is to separate the population on which the MST is built into distinct groups or clusters by dividing the graph into separate structures.
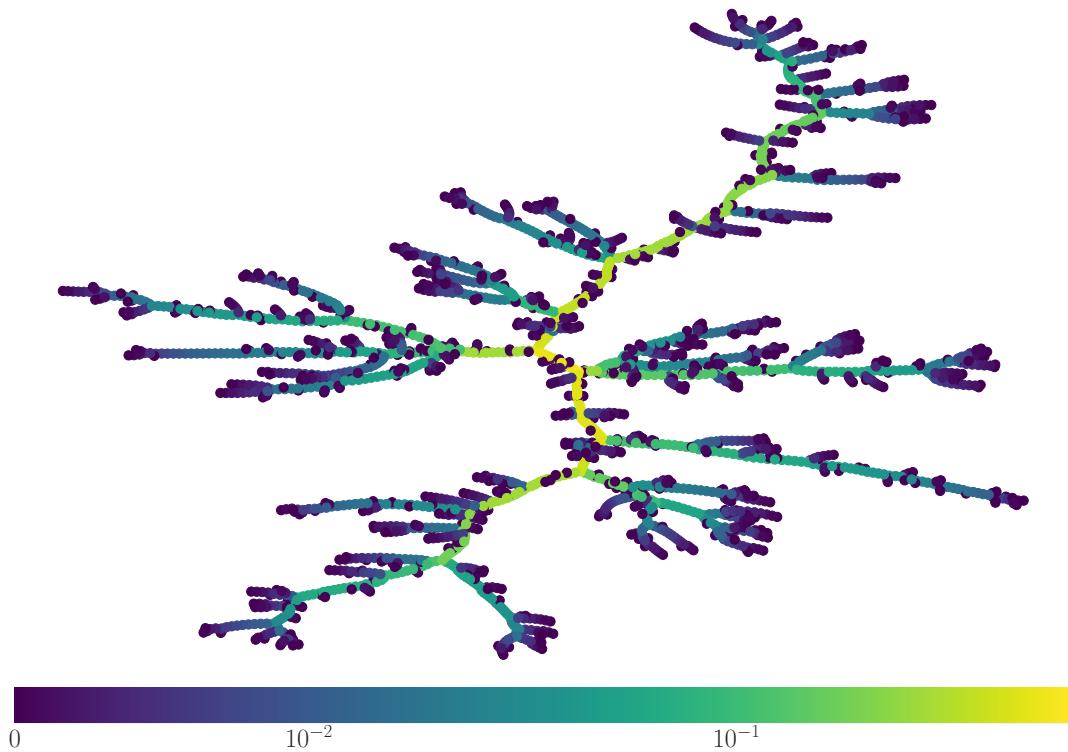
FIGURE 5.1: A color-coded representation of MST(2509, 2508) according to Equation (2.3). The color intensity varies from zero (with the strongest color) for nodes located at the termination of the graph, which are nodes of degree 1, to nodes in the central part of the MST, where C_B reaches its highest values, around 0.6.

This approach includes a topological stage, where the most central part of the MST structure, denoted as the trunk due to its resemblance to a real tree, is identified using betweenness centrality (see Section 2.4.1) as an indicator of node centrality under certain topological conditions. Once the trunk is established, we can locate the branches as the components that extend from it.

We then incorporate a statistical stage that considers the physical meaning underlying these branches. This approach enables us to select the branches that demonstrate a statistically significant difference among them, thereby positioning the clusters as potential distinct classes of pulsars.

5.2.1 Definition of the main trunk and branches

Figure 5.1 shows the Pulsar Tree after the application of Equation (2.3). As explained in Section 2.4.1, the value of the betweenness centrality estimator (C_B) enables us to establish a hierarchical structure in the MST by identifying nodes of higher centrality In other words, the higher the C_B, the greater the centrality in the MST, and therefore, making the node eventually part of the trunk.

Accordingly, Figure 5.2 shows the distribution of the C_B computed for all nodes in the Pulsar Tree. The distribution is strongly right-skewed, with a few outliers exhibiting extreme values. As shown in Figure 5.1, and consistent with the definition of betweenness centrality, these outliers correspond to the most central nodes of the graph. We interpret that from these will come the trunk of the Pulsar Tree.
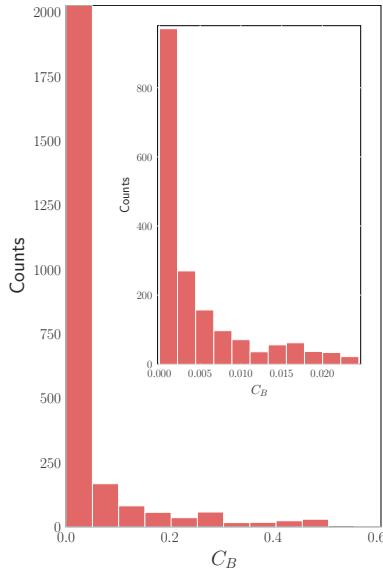
FIGURE 5.2: Distribution of the C_B shown in Figure 5.1. A zoom-in of
the lower 70% of the data is displayed in the top right.

We shall use an appropriate technique for asymmetric distributions to delimit the
nodes that will form the trunk (see, e.g., Tukey, 1977):

$$[Q_1 - k \times IQR]; [Q_3 + k \times IQR] . \tag{5.1}$$

The use of quartiles (Q_1, Q_3) to identify central values and outliers are validated for
both symmetric and asymmetric distributions: they do not assume anything about
the mean or the standard deviation, and their use is compatible with distributions of
positive rightward skew, as ours. Furthermore, any asymmetric distribution is more
robustly defined by the median as a measure of central tendency and the interquar-
tile range (IQR= $Q_3 - Q_1$) as a measure of its dispersion, as both are less sensitive
to extreme values. For our case, the right-hand side in Equation (5.1), together with
the usual value of 3 for *k*, sets the condition for a node to be considered belonging to
the trunk (i.e., an outlier of the distribution of C_B). We designated these as potential
trunk nodes, or pTN. This condition is necessary, but not sufficient, for a node to be
considered part of the trunk. In addition to being formed by pTNs (something that
relates to a bare-eye identification as the trunk in Figure 5.1), we need to provide a
criterion to establish where the trunk starts and ends. To that aim, we shall request
a topological condition: The trunk must be a path, i.e., a sequence of consecutive
nodes containing no duplicates (see Section 2.2.1), and we require that it starts and
ends at nodes, called contour nodes (CN), whose degrees are greater than 2. In this
way, we ensure that nodes give rise to substructures known as contour branches
(CBs) at the terminations of the trunk. These can be related to the mental image of
branches opening from the trunk in a real tree. To prevent CBs formed just by a few
nodes (noise), we define a significant threshold (α_s) so that the number of nodes in
these branches must exceed this lower limit. Under the validation of a given α_s for
the consideration of CBs, all the possible CNs are identified, and all the potential
trunks[1] (pTs), as the paths that pair these CNs, can be computed. Concluding, for
each pT, we have a set of branches starting from every node of degree larger than 2
(including the CBs) and having α_s as the minimum number of nodes.

---

[1]The number of potential trunks: #pTs = CNs × (CNs − 1)/2

5.2.2   Significant branches

Once the branches are obtained, we adopt a general conceptual definition: A significant branch is a group of nodes departing from the main trunk containing at least a α_s of the graph and can be statistically distinguished from other branches. To differentiate the branches, i.e., to select the significant branches, we will use the Kolmogorov-Smirnov (KS) statistics to define the significance threshold and immediately measure whether one branch distinguishes itself statistically from the others. The KS test compares two distributions under study by measuring the maximum distance between their empirical cumulative distribution functions, as described in Yadolah (2008), Wolfe (2012), and Lehmann (2012). The KS test does not assume any form of distribution beforehand, making it a non-parametric test that can be used for any distribution. The aim here will be to determine if we can reject the null hypothesis (H_0): the distribution of the properties of the nodes of two given branches is consistent with them being drawn from the same parent distribution. We will seek to reject this null hypothesis at a 95% confidence level (CL) or better. When this happens, we shall establish that whatever distance is used to compute the weights between nodes and form the MST separates branches whose nodes are drawn from statistically distinct parent populations.

Therefore, each set obtained will consist of a trunk and its significant branches. Depending on the conditions, the significant attribute of the branches can be adaptable to each case study. This will allow us to focus our work on a given clustering and discard those that do not accord with the proposed objectives (see, e.g., Section 6.3).

5.2.3   Significant Branches in the Pulsar Tree

To compute the significant branches derived after applying the clustering algorithm to the Pulsar Tree, we consider the distribution of C_B shown in Figure 5.1 and the outlier condition defined in Equation (5.1). From this, we identify that approximately 10% of the nodes in the Pulsar Tree qualify as pTNs.

To ensure statistical robustness in comparing the branches obtained, we impose an $\alpha_s \sim 5\%$ of the population. This value ensures that each branch contains enough values for the KS test to operate reliably. Additionally, we prioritize a set of significant branches with comparable sizes to improve the interpretability of the KS test results and minimize the impact of sample size heterogeneity. The seven significant branches of the Pulsar Tree are shown in Figure 5.3.

The significant branches identified resemble those one would point by hand if looking at the Pulsar Tree. When commenting on the MST as a descriptive tool, some branches were already used as examples in Section 4.4.5. The top and bottom branches of Figure 5.3 roughly correspond to binary pulsars and to the more energetic isolated pulsars of the sample (the youngest and those with the highest magnetic field at the light cylinder pulsars are located towards the end of this branch), respectively. Rightwards departing branches are characterized by increasing values of the surface magnetic field, ending with magnetars at their extremes. The outgoing leftward branch contains the oldest isolated pulsars. Figure 5.4 shows examples of the distribution of variables for some of the significant branches of the pulsar tree, as can be extracted from the Pulsar Tree web[2] introduced in Section 4.5.

The significant branches, by definition, separate different physical properties. To emphasize this point, Figure 5.5 compares the PC_1 and PC_2 (see Section 4.3) corresponding to each significant branch, showing differences in their distribution.

---

[2] http://www.pulsartree.ice.csic.es
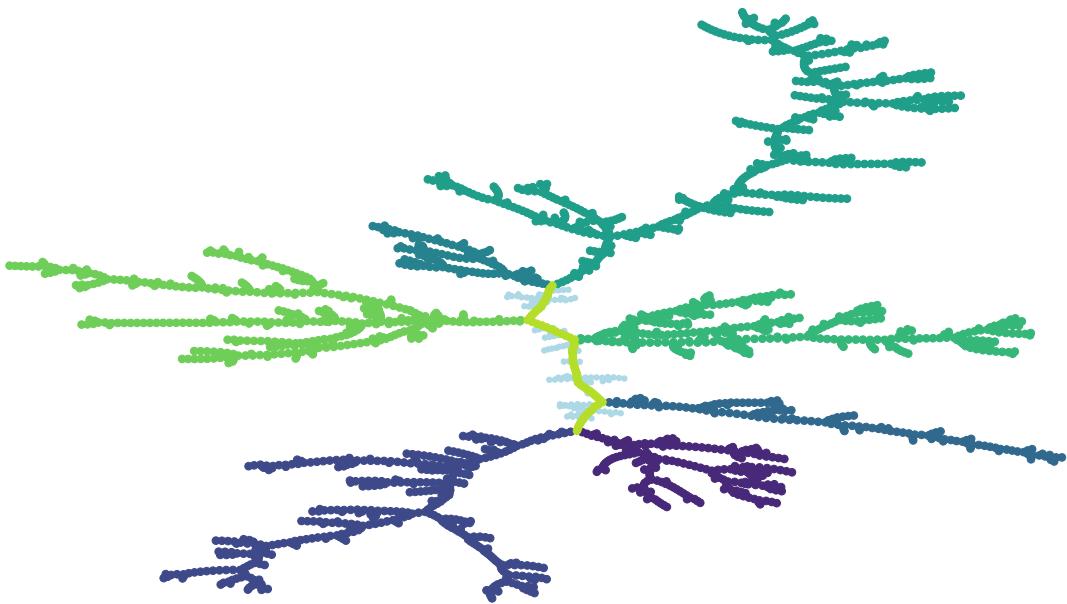
FIGURE 5.3: MST according to the significant branches of the pulsar population of the ATNF v1.67 after applying the clustering algorithm described in Section 5.2.1. These have at least 5% of the total population size; following the branches clockwise (from 12:00), they have the following pulsars: 585, 364, 157, 224, 444, 427, and 119, respectively. The MST nodes depicted in the main part are 64 and represent the trunk. Note that the remaining light blue (125) nodes belong to non-significant branches and are considered the noise of the trunk.
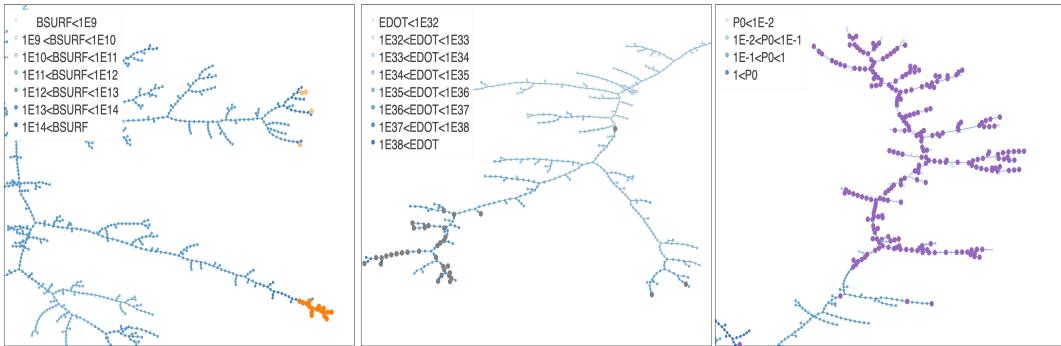

FIGURE 5.4: Three examples of the distribution of variables for some of the significant branches of the Pulsar Tree can be extracted from the online tool at `http://www.pulsartree.ice.csic.es`. We suggest looking at these examples directly to access additional functionalities, zoom in, and see individual values for each pulsar. The left panel shows the distribution of the surface magnetic field in the middle branches. It increases towards the outskirts of the MST and ends in low-field magnetars (light orange) and classical magnetars (orange). The central panel shows the spin-down power distribution of the bottom branch, with those having a pulsar wind nebula detected in TeV (H. E. S. S. Collaboration et al., 2018) (grey) noted. The right panel shows the period distribution in the upper branches of the MST, where binaries (purple) and long-period pulsars are located.
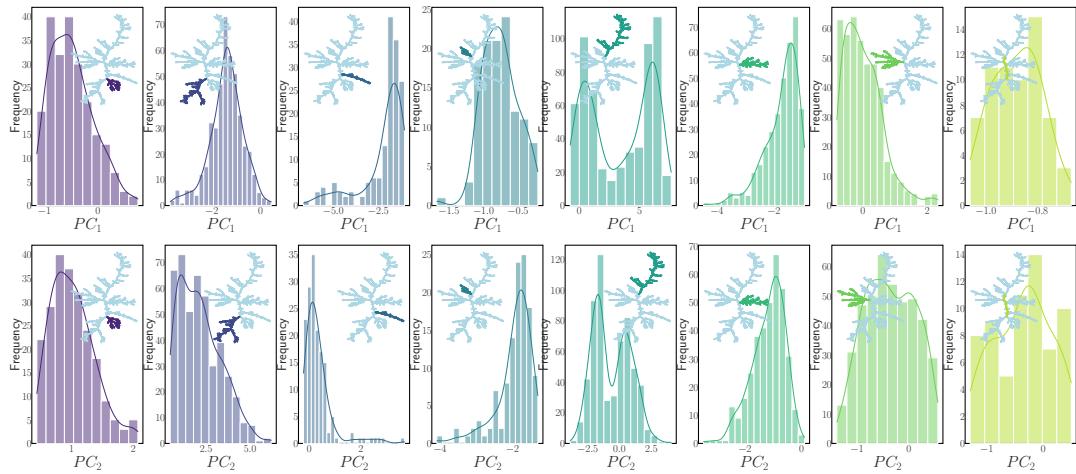
FIGURE 5.5: The distributions of the PC_1 and PC_2 according to the significant branches seen in Figure 5.3, as noted in the inset. All panels are plotted in the corresponding scale for PC_1 and PC_2. The last panel in each row shows the distributions for the trunk.

We also note the case of the fifth panel from the left in Figure 5.5, showing the distribution of the significant branch containing the binary pulsars. This significant branch is large and encompasses both pulsars in binaries and isolated ones. As such, it exhibits a double-peak structure in the distribution, which is not present in any other significant branch of the MST. Looking at the latter, the long-period, isolated pulsars are located at the bottom part of the significant branch. This significant branch can also be separately examined by focusing on the two smaller branches (73 and 45 pulsars), which are notably located in its lower part. None of these branches exceeds α_s and, according to the definition, cannot yet be classified as individually significant branches. However, analyzing them through the KS test, they reject H_0 concerning the rest of the significant branches as well as each other, over the distributions of all the variables used in this work. We expect that a higher number of pulsars will probably join these branches into one, generating an additional significant branch, or that the number of nodes of each branch will increase to make them individually significant.

Figure 5.6 shows the distributions of each underlying physical magnitude considered according to the significant branches of the Pulsar Tree seen in Figure 5.9.

5.2.4 Monte Carlo validation of significant branches

To evaluate the statistical robustness of the observed branches, we conducted a series of Monte Carlo simulations to test whether a similar level of separation could arise purely by chance. Specifically, we generated synthetic groupings by randomly partitioning the pulsar population into seven groups, ensuring that their sizes match those of the original significant branches. These random groupings are entirely independent of the MST structure and serve as a null model for comparison. We repeated this process 10^5 times, and for each simulation, we applied the KS test to assess the separability of the random "branches" based on the physical properties under consideration. The results show that achieving comparable separation by chance is highly unlikely. In none of the 10^5 simulations were all seven random branches simultaneously distinguishable using the KS test. In fact, in 55% of the simulations,
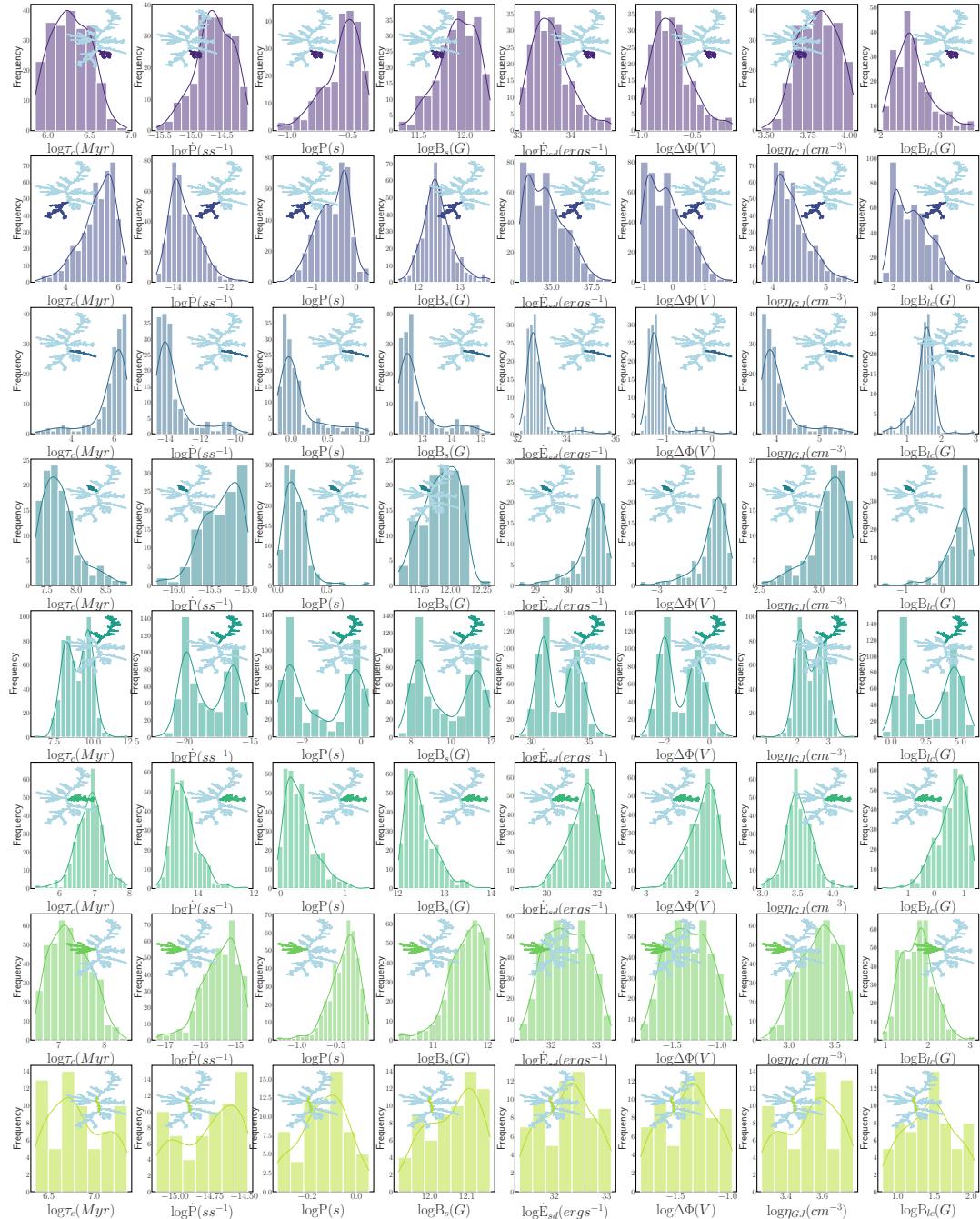
FIGURE 5.6: The distributions of the logarithm of the eight intrinsic variables considered to compute the significant branches (1 for each row) seen in Figure 5.3, as noted in the inset. The variables shown are spin period, spin period derivative, characteristic age, spin-down energy loss rate, surface magnetic flux density, the magnetic field at the light cylinder, surface electric voltage, and Goldreich-Julian current density. All panels are plotted on the corresponding scale for each magnitude. The last row shows the distributions for the trunk.

64

Chapter 5. Quantitative determination of minimum spanning tree structures: Using the Pulsar Tree for analyzing the appearance of new classes of pulsars

even a separation between just two of the seven random groups could not be statistically supported; the null hypothesis could not be rejected in those cases.

5.3 Evolution of the MST with the pulsar population

We now consider how the clustering algorithm applies to the evolution of the pulsar population throughout history. According to ATNF v1.67, we shall consider the pulsars known up to 1978, 1988, 1998, 2008, 2018, and 2022, resulting in sets containing 147, 439, 662, 1660, 2267, and 2509 pulsars. The clustering algorithm is applied individually to these subpopulations. Figure 5.7 shows the MST computed based on each subpopulation, considering the physical variables described for this work, with the C_B values noted as in Figure 5.1.

Equation (5.1) establishes the following pTNs for each subpopulation: 0%, 7%, 9%, 9%, 13%, and 10%. Due to the size of the subpopulations, we reduce the minimum branch size to consider it significant. That is, we set $\alpha_s \sim 3\%$ to obtain significant branches in all subpopulations. Thus, the number of significant branches extracted from the population of pulsars throughout history, that is, for each subpopulation, is: 0, 5, 5, 4, 9, and 7. It should be noted that the number of significant branches observed in each subpopulation is comparable despite introducing up to 20 times more pulsars than those known until 1978. The significant branches for each subpopulation are shown in Figure 5.8. Interestingly, due to the small sample size, the MST of the population of pulsars known up to 1978 does not reveal any outliers in its distribution of C_B. Correspondingly, we observe that the MST for this subpopulation has fewer substructures than those with more pulsars.

As stated before, these MSTs are still frames of the knowledge gathered from the pulsar population known up to each of the years quoted, where even whole classes of pulsars were undiscovered.

5.4 Pulsar classes appearance: *togetherness* and *growth* of the Pulsar Tree's significant branches

Figure 5.9 illustrates how the nodes of each significant branch observed in the pulsar tree today were distributed in previous versions of the MSTs according to the pulsars known at the time. Over time, all the significant branches of the Pulsar Tree converge, as most of the nodes in these branches were also previously connected. To quantify this evolutionary process, we define a measure called *togetherness*, which shows the percentage of nodes belonging to a given branch (or the trunk) that were together at earlier times, considering the significant branches produced in the MSTs of the other subpopulations. In cases where the given branch cuts into smaller pieces (e.g., its nodes populate two significant branches of the previous incarnation of the catalog), we shall also consider the group containing the most substantial number of nodes, as this is necessarily more representative of the final significant branch. This involves starting from a significant branch in the pulsar population of the ATNF v1.67, where we shall count the number of nodes that existed and were located together in a branch of the previous version (2018). Starting from the latter set, we checked how many existed and were already together in the catalog before it (1998), and so on.

We also introduce a measure called *growth*, which, unlike *togetherness*, focuses solely on the pulsars within the significant branch under analysis. Specifically, *growth*
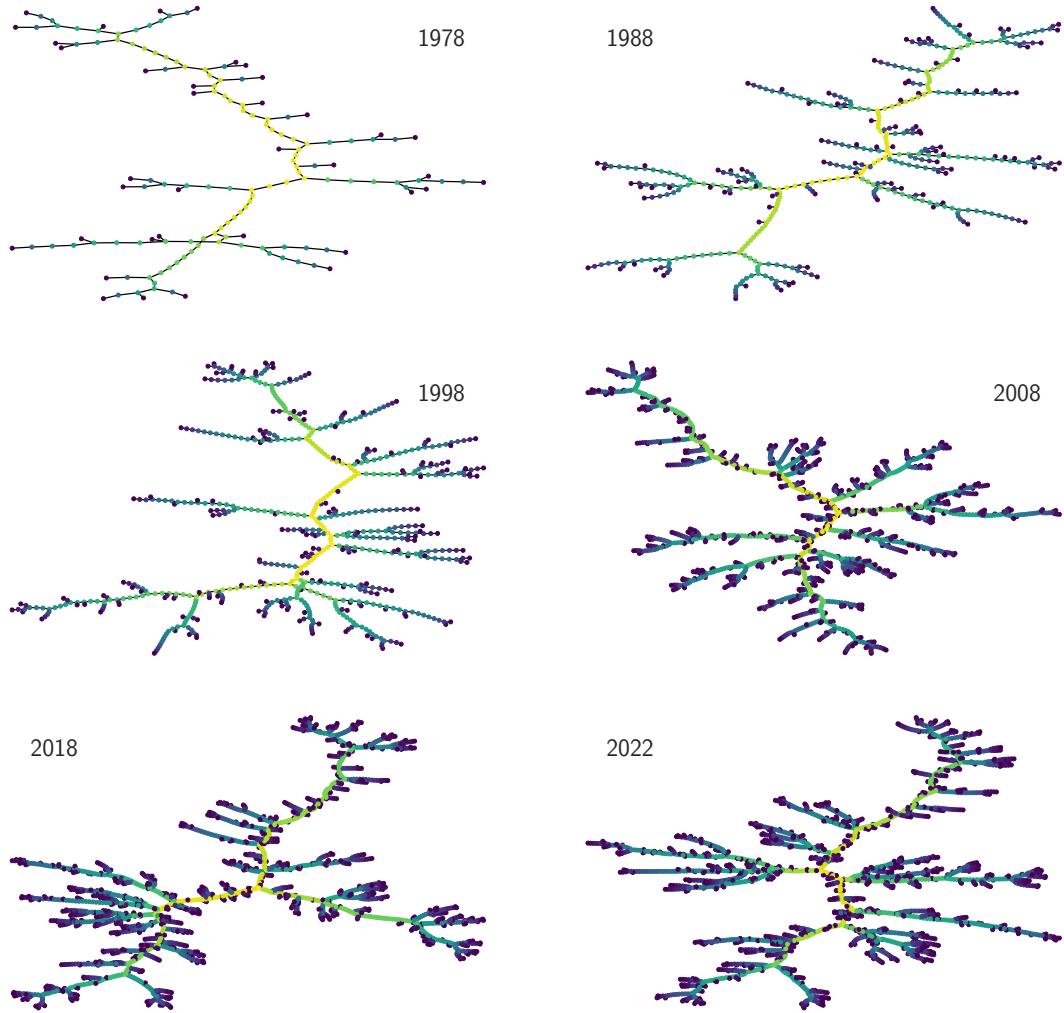
FIGURE 5.7: Representation of the C_B values found after the application of Equation (2.3) in the MSTs of the pulsar population throughout history, with nodes representing all pulsars known up to the different years according to the legend (up to December 31st of the prior year, and up to the date of version 1.67 of the ATNF catalog in March 2022). Similar to Figure 5.1, and on the same scale, the color map represents the highest C_B values in yellow colors and increases the darkness of the paint to dark purple for those values that reach zero.
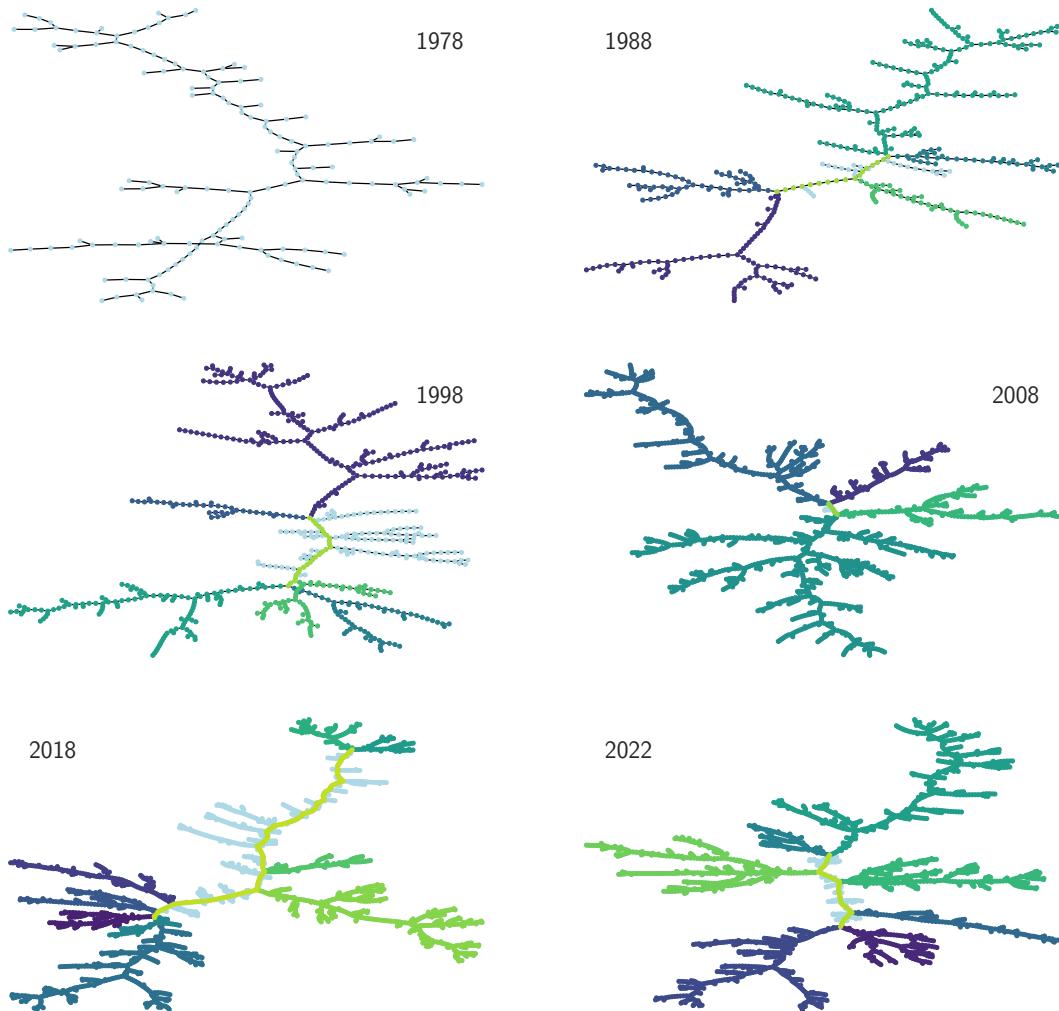
FIGURE 5.8: Significant branches and trunks for each MST corresponding to each subpopulation filtered by the time indicated in the legend. No part of the MST in 1978 is indicated because the application of Equation (5.1) does not find nodes that can represent the trunk (there are no outliers in the distribution of C_B). The different branches for each epoch are assigned a random color (MSTs are not color-coded for intercomparison). The trunk always shows the most intense green color. From 1988 to 2022, the number of significant branches was 5, 5, 4, 9, and 7, respectively.
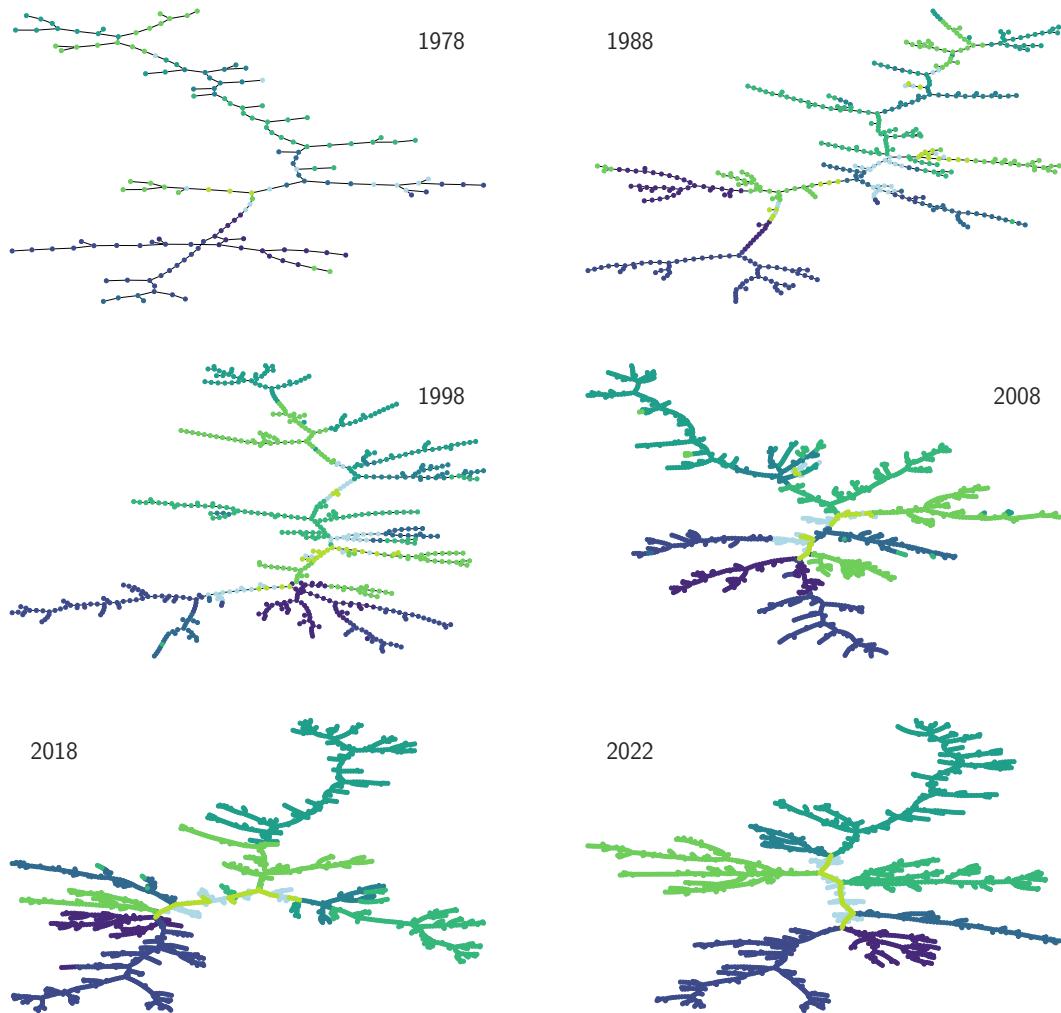
FIGURE 5.9: Projection of the significant branches seen in the Pulsar Tree (the last panel corresponds to that shown in Figure 5.3) into the MSTs corresponding to each epoch, following a chronological order as indicated within each panel.

tracks whether these pulsars remain grouped over time across all subsequent MSTs, as additional pulsars are incrementally added until the branch reaches its final size. Figure 5.10 shows that all branches in the Pulsar Tree not only join groups of pulsars that were earlier separated (*togetherness* in the first row) but also have a core group of pulsars that stay together along all subpopulations (*growth* in the second row).

5.5 Conclusions

A new class of pulsars is conceptually a new set with distinct properties. In this chapter, we introduce a methodology based on the graph theory-motivated pulsar tree and the underlying physical properties of the pulsar population, which enables the study of this phenomenon from a clustering perspective. In general terms (since it may find applications for other problems beyond pulsars), we have introduced a methodology to utilize the MST to qualify the nodes that form it. The logic is as follows:

1. Produce the MST based on the Euclidean distance of the variables containing the full variance of the population as determined from PCA.

2. Compute betweenness centrality for all nodes, produce its distribution, and determine which nodes are outliers.

3. Set the significance threshold, defined as the minimum percentage of the total population of pulsars for which the branches are statistically distinct (measured by a KS test).

4. If having former incarnations of the sample, with fewer nodes (in our cases, those provided by the historical discovery of pulsars), compute *togetherness* and *growth* to see convergence stability and establishment of significant branches.

The former methodology enables us to quantitatively define the trunk and significant branches of any tree, for which the nodes have statistically different distributions of the selected variables. In the Pulsar Tree, using the intrinsic variables, the significant branches can be directly associated with "classes" or, at the very least, with connections among the respective nodes, which quantitatively distinguish them from others in different tree locations. Studying the population evolution throughout history allows us to see how our nodes that were initially non-connected, become connected, while others remain linked all the time. In the future, we may see new branches appear, with no or minimal projection onto the ATNF v1.67 pulsar catalog, indicating a new class. We may also see, as described, non-significant sub-branches develop into full-fledged significant ones. The MST will provide a quantitative perspective on whether we see something "new" or just "more of the same" when new pulsars are discovered and whether currently unknown pulsars connect dislocated groups.

This methodology has been applied to the population of MSPs in binary systems (see Chapter 6) to identify clusters that may suggest the existence of new pulsar classes or provide further insight into the behavior of known systems, such as spider pulsars. This is characterized not only by physical properties but also by binary parameters.
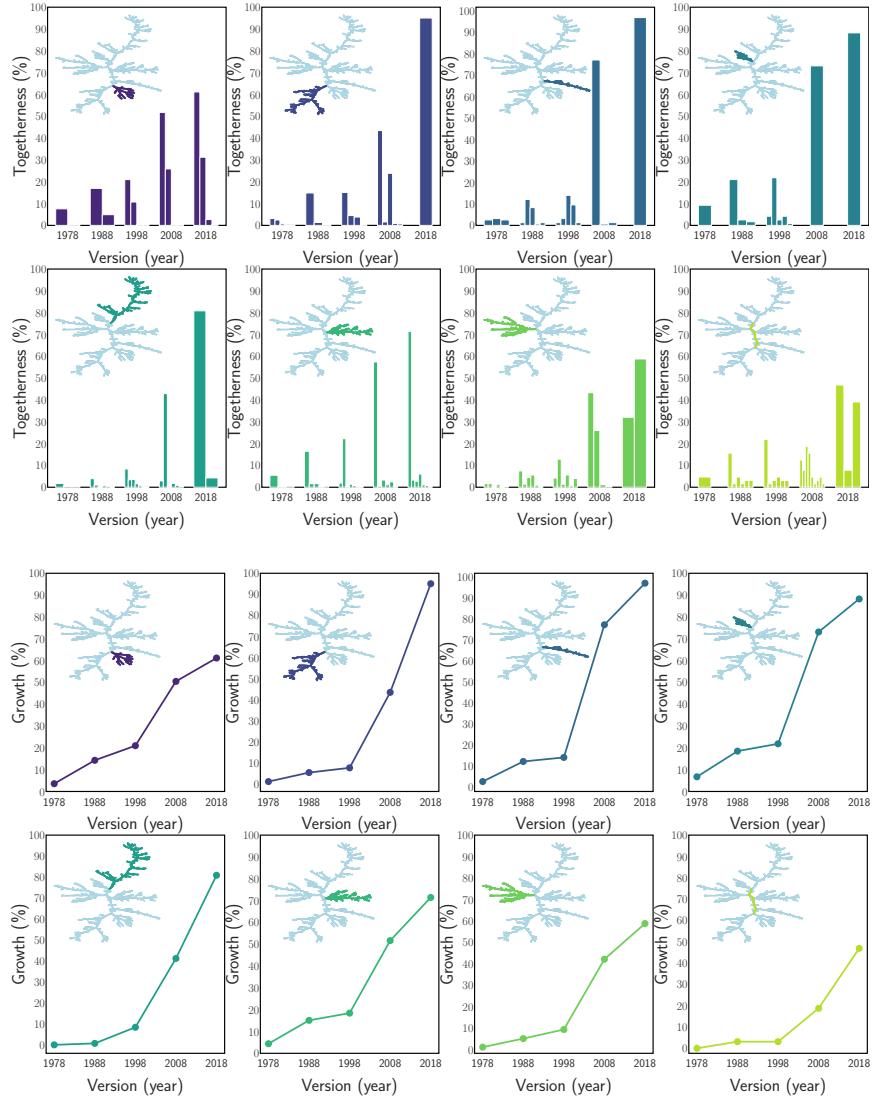
FIGURE 5.10: First row: The *togetherness* is shown, representing the percentage of nodes related to each significant branch of the Pulsar Tree, according to the marking in the inner MST depicted in each panel, which was already together in an MST at earlier times in 2022. When the given significant branch of the Pulsar Tree appeared in different branches in the earlier versions of the Pulsar Tree (see Figure 5.9), different bars are shown; whereas, when the nodes in a significant branch were located all together in a single group, only a bar appears. Second row: The *growth* is shown, representing the percentage of nodes related to each significant branch of the Pulsar Tree, according to the marking in the inner MST depicted in each panel, that remain together throughout history after appearing. To obtain these percentages, we have taken into account the best representation (the largest bar for each branch in 2018, as seen in each panel of the first row) and analyzed how this group has grown from the previous time, repeating the process until 1978.

Chapter 6

Millisecond pulsars phenomenology under the light of graph theory

6.1 Introduction

This chapter is based on García et al. (2024a) and continues the study that integrates graph theory (see Chapter 2) and the application of PCA (see Chapter 3) to the field of pulsar astrophysics.

Here, we specifically consider the millisecond pulsars (MSPs) population, i.e., weakly magnetized and rotating neutron stars with spinning periods shorter than 10 ms. They are usually hosted in tight binary systems in a low mass ($< 1\,\mathrm{M}_\odot$) companion star. MSPs offer unique insights into stellar evolution, the interaction between magnetic fields and plasma transferred by the donor star, and particle acceleration from compact objects, particularly in binary systems (see Section 1.4 for more details and Manchester 2017; Papitto and Bhattacharyya 2022 for comprehensive reviews).

By treating each MSP as a node, we shall compute the MST (see Section 2.3) of the MSP population and use it to describe the population as a whole and to identify individual pulsars that may warrant further investigation due to their unique attributes or positions within the graph. We shall specifically examine locations in the MST in which spider pulsars reside (e.g., Eichler and Levinson 1988; Roberts 2013; Roberts et al. 2018; Di Salvo et al. 2023), i.e., eclipsing radio pulsars in tight binary systems either with a non-degenerate main-sequence companion with a mass in the range ∼0.1–0.8 M$_\odot$ (redbacks, RBs) or with a $\lesssim 0.06\,\mathrm{M}_\odot$ semi-degenerate companion (black widows, BWs). We also consider transitional millisecond pulsars (tMSPs, Papitto and de Martino 2022) that exhibit dramatic state changes, i.e., go from rotation-powered to accretion-powered and vice versa, on timescales as short as a few weeks (see Section 1.4.3 for more details). Driven by the method introduced in Chapter 5, which allows for dividing an MST into distinct parts based on specific variables, we explore how to classify MSPs into different groups. Finally, we shall implement an algorithm to search for the position of pulsars within the graph, even when some properties have not yet been measured. This will allow us to predict ranges for specific variables and their characteristic phenomenology.

6.2 The sample variance and the MST

6.2.1 Sample selection

The sample used is taken from the Australia Telescope National Facility catalog, v2.1.1 (ATNF v2.1.1, see Manchester et al., 2005) imposing that pulsars have a spin period in the millisecond range ($P < 10^{-2}$ s) and a positive spin period derivative ($\dot{P} > 0$ s s^{-1}). The latter allows us to calculate the intrinsic variables derived from P and $\dot{P}$ as described in Section 6.2.2. A total of 218 pulsars result from these cuts. Of these, 43 are confirmed as BWs, 9 of them in globular clusters (see, eg., Swihart et al. 2022; Koljonen and Linares 2023; Freire et al. 2017; Lynch et al. 2012; Douglas et al. 2022), 16 as RBs, 2 of them in globular clusters, (see, e.g., Strader et al. 2019; Koljonen and Linares 2023), and two as tMSPs, J1023+0038 (Archibald et al., 2009) and J1227-4853 (Bassa et al., 2014). Note that tMSP J1824-2452I (Eckert et al., 2013) is excluded from the sample as it does not have a $\dot{P}$ measurement, although it will be analyzed in Section 6.4.1. These pulsars are listed in Table 6.2. According to the third pulsar catalog by the Fermi Large Area Telescope (*Fermi*-LAT, see Smith et al. 2023), 103 pulsars out of the 218 are found to be gamma-ray emitters.

6.2.2 Variables and PCA

Figure 6.1 shows the logarithmic distribution of the spin period (P) and spin period derivative, ($\dot{P}$), surface magnetic flux density (B_s), the magnetic field at the light cylinder (B_{lc}), spin-down energy loss rate ($\dot{E}_{sd}$), surface electric voltage ($\Delta\Phi$), and Goldreich-Julian charge number density (η_{GJ}), binary period (P_B), projected semi-major axis of the orbit (A_1) and the median mass of the companion star for each system (M_C). We do not consider here the characteristic age ($\tau_c = P/2\dot{P}$) because, in binary systems, additional torques imparted on the pulsar during accretion phases can render this estimate unreliable (see, e.g., Kiziltan and Thorsett 2010; Tauris, Langer, and Kramer 2012; Jiang et al. 2013). The distributions are not normal (which is also the case for the set of distributions of the original variables, without logarithm), and due to this fact, we use the robust scaler to scale them (see, Equation (4.1)). Figure 6.2 shows how the P_B, A_1, and M_C distribution distinguishes BWs from other classes. However, when none of these variables are used in the analysis, the distinction is no longer evident, and spider pulsars are seen as a more homogeneous group.

Figure 6.3 shows the results of the PCA analysis applied to these variables (see Chapter 3 for more details). In the left panel of Figure 6.3, we observe how most of the variance is distributed in the first two PCs. Likewise, the central panel shows that the whole variance is represented in a 4-dimensional space, i.e., we need the first 4 PCs to cover 100% of the variance of the sample. This relatively low number of PCs results from the fact that within the dipolar model used as a proxy (see Section 4.2.1), all physical variables depend on P and $\dot{P}$, and also that A_1 of a binary system is related to the P_B and the M_C (in this case through the total mass of the system, i.e., the sum of the masses of the pulsar and its companion) via Kepler's third law. The PCA, as shown in the right panel of Figure 6.3, does not highlight the dominant influence of key variables for the separation between pulsar spiders, such as the M_C, but provides a more complex classification where its dominance is balanced with other variables. Technically, the four PCs needed to describe all the variance have similarly significant loadings in several variables (see the right panel of Figure 6.3).
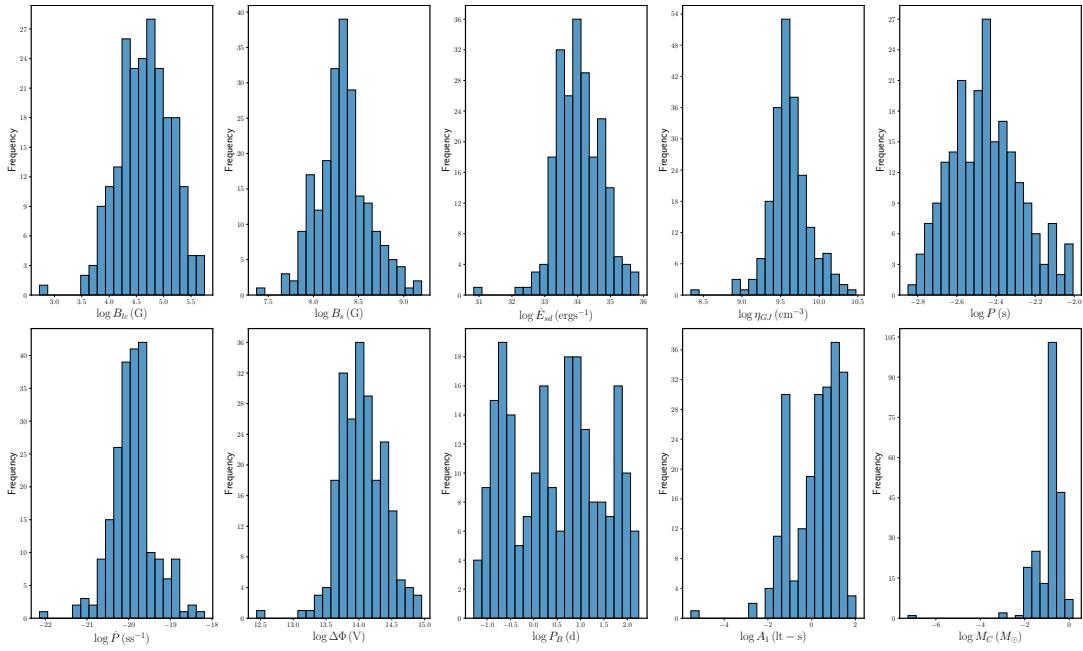
FIGURE 6.1: Distribution of the logarithm of the 10 variables considered for the sample of 218 pulsars.

6.2.3   MST

In Chapter 2, all the necessary concepts to calculate an MST are explained. We define a Euclidean distance (see Equation (2.1)) over the variables or the PCs described in Section 6.2.2. Note, as explained in Section 4.2, that using the first four PCs ($\sim 100\%$ of the explained variance) produces the same MST as using all variables (and thus the results from the analysis that follows from the graph are the same) but is less demanding given the reduced dimensionality of the problem. With this Euclidean distance, we first compute a complete, undirected, and weighted graph $G(V, E) = G(218, 23653)$, with $|V| = 218$ nodes and $|E| = 23653$ edges, with a specific weight value defining each edge. From that, we obtain the MST of this sample, $T(V, E') = (218, 217)$, with $|V| = 218$ nodes and $|E'| = 217$ edges, shown in Figure 6.4, where each node represents a pulsar.

The MST can separate BWs from RBs, with tMSPs appearing close in the same structure, as shown in Figure 6.4. These tree´s regions are zoomed in Figure 6.5, where several pulsars that will be discussed next are highlighted[1]

Black widows pulsars in the graph

We will now focus on some specific pulsars following Figure 6.5. Two of these are J1300+1240 and J1630+3550, which have not been explicitly classified as BWs yet but have been discussed by Yan et al. (2013) and Sobey et al. (2022). A possible formation scenario for J1300+1240, suggested by Yan et al. (2013), is in low-mass, narrow-orbit bound binaries. Here, the initial point of departure is a low-mass binary pulsar with a very narrow orbit, such as the known BW J1959+2048. Due to their comparable velocity and rotation period to those of J1959+2048, J1300+1240 may also evolve in this scenario. J1630+3550 is an MSP in a binary system of 7.6h orbiting a companion with a minimum mass of 0.0098 M$_\odot$. These parameters can indicate a BW as is

---

[1]Details on the Pulsar Tree web are given in Section 4.5, http://www.pulsartree.ice.csic.es.

FIGURE 6.2: Cross-dependence of the 10 variables considered. BWs (in red) and RBs (in green) confirmed are separately noted. For visualization purposes, no distinction is made between MSPs residing in globular clusters, and the tMSPs are marked in green due to their behavior as RB. The pulsars not yet assigned to any of these classes are shown in blue. The main diagonal shows the distribution for each variable. The panels with the shaded background show the pairs formed with P_B, M_C, and A_1, which best differentiate the BWs from the rest.
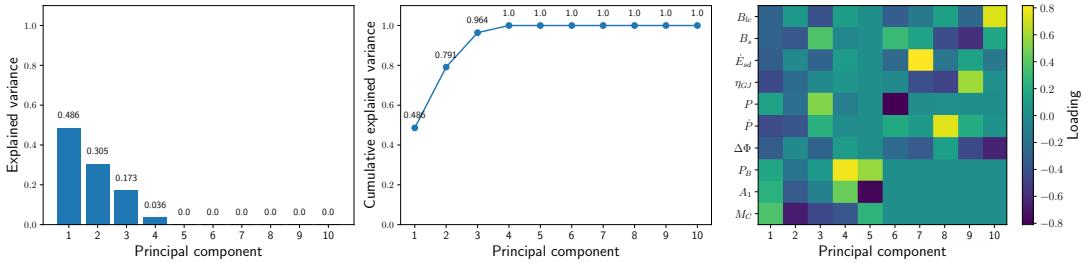
FIGURE 6.3: PCA results for the logarithm of the set of variables for the population of 218 MSPs, see Section 3.2.3 for further explanations about PCA definitions. The left panel, called the scree plot, shows each PC's explained variance according to the covariance matrix's eigenvalues. Note that the covariance matrix calculates the relationships between pairs of variables, showing how changes in one variable are associated with changes in another. It represents the amount of information contained in each PC. The central panel shows the cumulative explained variance by the new variables defined through the PCA analysis. The right panel shows the "weight", also called loading, each variable has concerning each PC, indicating its contribution to the variance captured by that PC. This value is the coefficient held in each eigenvector of the covariance matrix. Negative values imply that the variable and the PC are negatively correlated. Conversely, a positive value shows a positive correlation between the PC and the variable.


FIGURE 6.4: The MST of the binary pulsar population defined as $T(218, 217)$ based on the complete, undirected, and weighted graph $G(218, 23653)$ computed from the Euclidean distance among 10 scaled variables (or the equivalent 4 PCs that described their whole variance). Each node in the MST represents a pulsar. The MST notes separately confirmed BWs (red), RBs (green), and tMSPs J1023+0038 and J1227-4853 (yellow), respectively. Also, the BWs and RBs in globular clusters (light red and light green, respectively) are highlighted. The RBs J1622-0315 (green) and the RB candidate J1302-3258 (light teal) are also noted in the rightmost branch of the MST. The unclassified ones appear in blue. See Table 6.2 for more details.
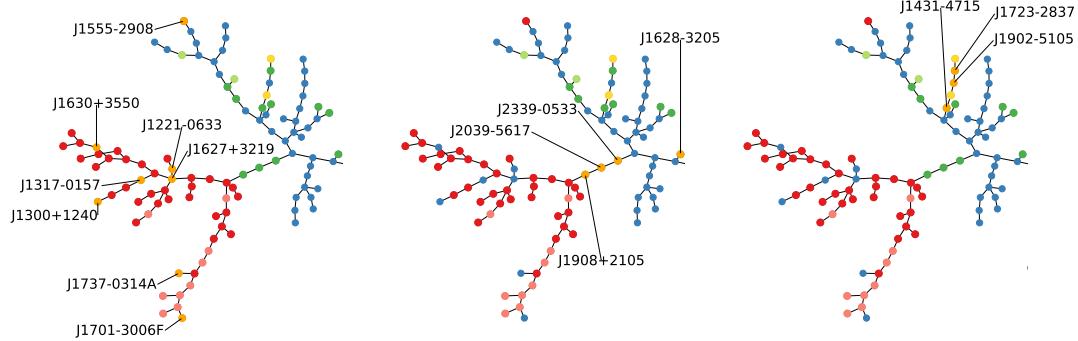
FIGURE 6.5: Zoom of the leftmost part of the MST shown in Figure 6.4, to observe more clearly the location of the pulsars (shown in orange) discussed in Section 6.2.3 (left panel), in Section 6.2.3 (central panel), and Section 6.2.3 (right panel). The same color code of Figure 6.4 is shown, where the confirmed BWs appear in red, RBs in green, tMSPs in yellow, and the BWs and RBs in globular clusters in light red and light green, keeping the unclassified pulsars in blue.

stated in Lewis et al. (2023). From the MST graph, we can see that they are located in the BW region. We can further consolidate their potential categorization as BWs by observing the ranking of the nearest pulsars based on the Euclidean distance derived from the defined variables, which can be obtained using the Pulsar Tree web[2]. Note that the ranking provides a distance-ordered list of one pulsar relative to all others. In contrast, the MST provides a global manner of optimizing the total pathway, considering all distances, to connect all pulsars. Only the nearest neighbor to any given pulsar will have an immediately recognized position within the MST; as such, the closest neighbor will always be one of the pulsars connected to the one of interest in the MST. For a deeper discussion and examples regarding the differences between the Ranking and the MST, see Appendix C.2 for a more detailed explanation. J1300+1240 has nine pulsars among its 10 nearest neighbors, all of which have already been classified as BWs. The remaining one is J1630+3550, whose 10 nearest neighbors are also BWs.

Among the interesting sources highlighted in Figure 6.5 are J1317-0157 and J1221-0633, classified as BW systems by Swiggum et al. (2023), having tight orbits, low-mass companions, and exhibiting eclipses. In both cases, their nearest neighbor ranking contains nine confirmed BWs except for J1627+3219.

J1627+3219's position in the MST, underscored by its ranking, where eight of its ten nearest neighbors are confirmed BW pulsars, supports its categorization in this group. This assertion is further bolstered by Braglia et al. (2020), who identified a candidate optical periodicity, necessitating verification through subsequent observations. Recent research by Corcoran, Ransom, and Lynch (2023) also underscores the potential BW nature of this pulsar. Still, the lack of detected radio pulsations highlights the need for further studies to understand its nature fully.

We also pinpoint J1737-0314A, which is likely a BW system according to the timing solution provided by Pan et al. (2021). Looking at its ranking, its categorization as a BW is favored since it contains nine confirmed BWs among its 10 nearest neighbors, the remaining being J1701-3006F. In addition, the latter, J1701-3006F, considering its mass limit and orbital features, is also favored as a BW, as outlined in Freire et al. (2005), although eclipses have not been observed. However, this absence could be

---

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

due to a low orbital inclination (Vleeschower et al., 2024). On the other hand, its nearest neighbor ranking shows six BWs, with the closest being J17001-3006F; see also King et al. (2005). Both J17001-3006E and J1701-3006F appear connected in the MST in Figure 6.5, filling the end of a branch that contains many of these cases.

Notably, J1555-2908 is the only confirmed BW away from the rest in the MST. In the work by Kennedy et al. (2022), extensive analysis was performed using optical spectroscopy and photometry. This study, combined with gamma-ray pulsation timing information alongside a companion mass of 0.06 $M_\odot$, concluded that J1555-2908 lies at the observed upper boundary of what is typically classified as a BW system. Note that its nearest neighbor, J1835-3259B, is not a BW.

Of interest to all sources discussed in this section, we remark that no other non-BW pulsar, except J1908+2105 discussed in the next section, has more than five confirmed BWs in its ranking.

Redbacks pulsars in the graph

Figure 6.5 shows that RB pulsars do not appear clustered like BWs but are mainly close to each other, with a small fraction near the border of BWs. The caveat here is the smaller number of RBs known so far.

The pulsar J1908+2105 is an interesting case within the RB class. This one is right on the frontier with the part of the MST mostly populated by BWs and could be classified as such, mainly due to its minimum companion mass of 0.06 $M_\odot$ and its short P_B of 0.14 days (Strader et al., 2019). However, the pronounced radio eclipses of this system align it more closely with RBs (also see, e.g., Koljonen and Linares 2023; Deneva et al. 2021). Note that ranking the 10 nearest neighbor pulsars to J1908+2105 shows that 8 out of 10 are BWs. The fifth and ninth pulsars in the ranking, J2039-5617 and J0337+1715, respectively, are RB and uncategorized. This situation, where the ranking of a given pulsar thought to pertain to one class (RB) is dominated by pulsars belonging to the other, only arises in the case mentioned above. Thus, the MST location and the ranking favor, or at the very least do not rule out, a BW classification for J1908+2105.

On the other hand, the gamma-ray source 3FGLJ2039-5617 (PSR J2039-5617) is almost certainly associated with an optical binary listed as an RB candidate by Strader et al. (2019). It validates its predicted nature of RB Corongiu et al. (2021), where they found clear evidence of eclipses of the radio signal for about half of the orbit, which they associate with the presence of intra-binary gas. It appears as a confirmed RB in Koljonen and Linares (2023). Furthermore, J2039-5617 shares similarities with the confirmed RB J2339-0533, exhibiting a peak in gamma-ray emission close to a minimum in its optical emission, which contrasts with the expected phase of the intrabinary shock (IBS) as seen in An et al. (2020). This similarity between J2039-5618 and J2339-0533 and their position in the MST highlight intriguing aspects of their behavior, alerting us to further analysis.

The distinct positions of J1622-0315, J1302-3258, and J1628-3205, seen in Figure 6.4, located far from the rest of the RBs in the MST, also capture attention. J1622-0315 is one of the lightest known RB systems, a companion mass of 0.15 $M_\odot$, with a relatively hot companion (Yap, Kong, and Li, 2023; Sen et al., 2024). This system is notable for its P_B of 0.16 days, marginally smaller than other RBs that typically have one of 0.2 days or more.

TABLE 6.1: Ranking based on the Euclidean distances calculated over the 10 scaled variables considered for the nodes in the region of the tMSPs.

J1023+0038	J1723-2837	J1902-5105	J1227-4853	J1431-4715
J1723-2837	J1902-5105	J1723-2837	J1902-5105	J2205+6012
J1227-4853	J1023+0038	J1227-4853	J1431-4715	J1227-4853
J1902-5105	J1227-4853	J1431-4715	J1723-2837	J1342+2822B

On the other hand, J1302–3258 is classified as an RB candidate and reported to have a minimum companion mass of 0.15 $M_\odot$ (Strader et al., 2019). However, it is distinguished by the absence of an identified optical companion and the lack of published evidence of extensive radio eclipses. Koljonen and Linares (2023) note that, unlike most known RBs and candidates, J1302-3258 does not have a *Gaia* counterpart, a trend more typical of BWs, which generally possess more fabulous companion stars with fainter optical magnitudes. J1628-3205 is a RB in a 0.21 day-long orbit around and has a 0.17–0.24 $M_\odot$ companion star (Ray et al., 2012; Li, Halpern, and Thorstensen, 2014; Strader et al., 2019). Its optical counterpart shows two peaks per orbital cycle. Such a shape is reminiscent of the ellipsoidal deformation of a star that nearly fills its Roche lobe in a high-inclination binary. It suggests that heating by the pulsar wind is not a major effect. In this context, J1628-3205 is an intermediate case among RBs, falling between systems where the companion star's emission is dominated by pulsar irradiation (e.g., J1023+0038) and those that are not (e.g., J1723-2837, J1622-0315, J1431-4715)(see Li, Halpern, and Thorstensen 2014; Yap, Kong, and Li 2023; de Martino et al. 2024 and references therein).

Transitional pulsars in the graph

The pulsars J1723-2837, J1902-5105, and J1431-4715 appear in the same part of the MST in Figure 6.5 as the known tMSPs J1227-4853 and J1023+0038. In addition, these three are the only ones in the sample containing at least one known tMSP in their top three neighbors; see Table 6.1, which also lists the closest neighbors of the two confirmed tMSPs.

J1723-2837 is a nearby (d∼1 kpc based on *Gaia* parallax measure) 1.86 ms RB in a 15 h-period binary system (Crawford et al., 2013). With an X-ray luminosity of 10^{32}erg s^{-1}, it ranks among the brightest RBs in that band (see, e.g. Lee et al., 2018). Based on the similarity with the X-ray output of tMSPs in the rotation-powered pulsar state, Linares (2014) suggested that it is one of the most promising candidates to observe a transition to an accretion state. The observed emission, in both soft (Bogdanov et al., 2014; Hui, 2014) and hard X-rays (Kong et al., 2017), is modulated at the P_B, indicative of an origin in the IBS between the pulsar wind and mass outflow from the companion star. On the other hand, its optical counterpart shows no sign of significant irradiation with two peaks per orbital cycle (Li, Halpern, and Thorstensen, 2014).

J1902–5105, a 1.74 ms radio MSP in a 2-day-period binary system, was discovered within the Parkes telescope surveys targeting unidentified *Fermi-LAT* sources (Kerr et al., 2012). It is a relatively bright MSP located at a distance of ∼1.2 kpc (Camilo et al., 2015). It was suggested that the companion is a $0.2 - 0.3 M_\odot$ white dwarf with a helium core (Camilo et al., 2015).

Finally, J1431-4715, discovered in the High Time Resolution Universe (HTRU) survey with the Parkes radio telescope, is an RB MSP with a P of 2.01 ms in a 10.8 hr

orbit with a companion mass of 0.20 $M_\odot$ (Bates et al., 2015; Miraval Zanon et al., 2018; de Martino et al., 2024).

Results reported in García et al. (2024a) indicate a lack of moding[3], suggesting a lack of conclusive evidence regarding the transitional nature of J1723-2837, J1902-5105, and J1431-4715. Still, their position in the MST and neighbor ranking confirms them as subjects of interest as potential transitional systems, though they are currently in the radio pulsar state. The relative proximity of J1723-2837, particularly J1902-5105, renders them optimal candidates for potentially unveiling a future transition from a rotation-powered state to an accretion-powered one.

6.3 MST clustering

In this Section, we apply the clustering algorithm introduced in Chapter 5 to separate an MST into distinct parts according to a quantitative prescription. The methodology and specific details of this case are presented in the Appendix D. Figure 6.6 shows separated branches, and Figure 6.7 shows the distributions for each magnitude considering them. The MST visually represents the binary millisecond pulsar population in this way without making any prior assumptions about the nodes' nature.

Two of the branches indicated by the method in Figure 6.6 (depicted in dark green and pink) contain all BWs except J1555-2908, which was already deemed as lying at the edge of the BWs' population as discussed in Section 6.2.3. In addition, one of these branches (pink) contains the most candidate or confirmed BWs seen in globular clusters. The BWs in globular clusters, together with the rest of the pulsars seen in the pink branch, show a $\dot{P}$ and $\dot{E}$ in general higher than the others; see Figure 6.7. Note that these values cannot be reliably measured in a globular cluster due to the acceleration in the cluster's gravity well. Similarly, tMSPs and most RBs fall in another branch (depicted in gray according to Figure 6.6). As we see in Figure 6.7, pulsars in the orange branch exhibit a P_B, M_C, and A_1 close to those in the gray branch. However, in contrast, they show a lower $\dot{P}$ and thus a smaller B_s. The rest of the population, particularly those pulsars in the dark blue and brown branches, which are significantly less different to each other, are markedly different from the other groups, as shown in Table D.1 (see Appendix D for details on the similarity criteria). Figure 6.7 shows that although these nodes are less obviously separated in the usual representations, such as the typical $P, \dot{P}$ diagram, they show larger P_B, M_C, P, and smaller $\dot{E}$ than the BW pulsars.

We can discuss differences among branches based on the distributions of the variables seen in the main diagonal of Figure 6.7. We observe that M_C exhibits sharp distributions around similar values, except for the (pink) branch, which primarily contains BWs in globular clusters, for which M_C shows smaller values. The dark green branch, plagued by BWs, displays a broader distribution, always within small values. This is reflected in A_1 and P_B, where the behavior is similar even with somewhat broader distributions, mainly for the branches that do not contain BWs. The widest distributions are observed in P for all branches except the one depicted in orange, which has a high density of cases around low values. The dark blue branch contains the pulsars with more extended periods. The $\dot{P}$ shows different trends for

---

[3]Although J1723-2837, J1902-5105, and J1431-4715 are currently detected as radio MSP and are therefore not expected to exhibit the typical bimodality in the X-ray light curve, we analyzed *XMM-Newton* archival observations of these sources (ObsID 0653830101 for J1723-2837, ObsID 0841920101 for J1902-5105, and ObsID 0860430101 for J1431-4715) to check for possible mode switching. This analysis is shown in Appendix E.
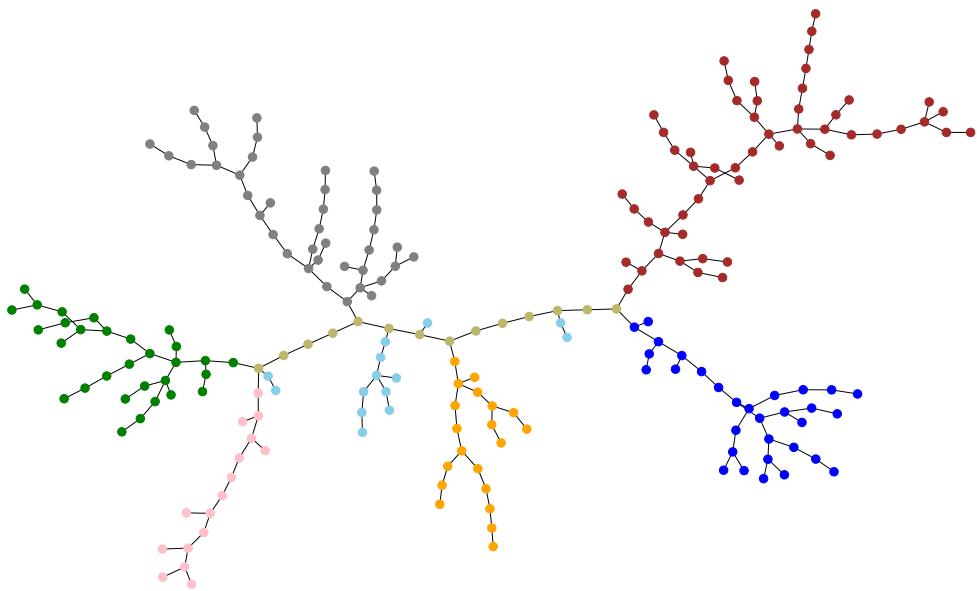
FIGURE 6.6: The MST defined as $T(218, 217)$ of the MSPs population is separated into significant branches according to the algorithm described in the text (see Section 6.3), is shown. The branches group a comparable number of pulsars: gray (39), orange (20), dark blue (31), brown (54), dark green (30), and pink (16). The trunk (in dark khaki) has only 14 nodes, with 14 others (in sky blue) not considered to pertain to any significant branch or trunk, being the noise of the former structures.
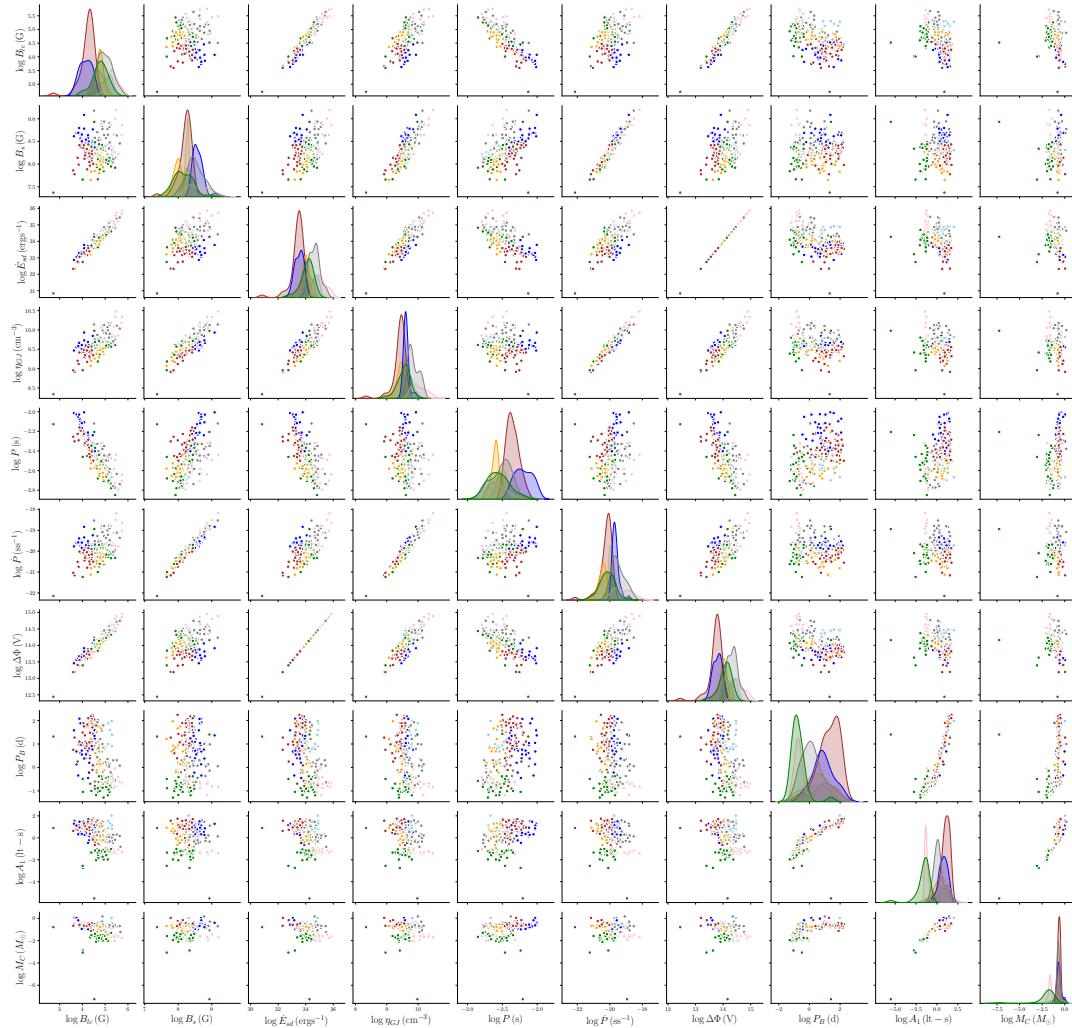
FIGURE 6.7: Cross-dependence and distributions of the 10 variables considered separated according to the branches in Figure 6.6.

some branches. In that order, the orange, brown, and dark green branches exhibit sharp behaviors that enable us to distinguish the pulsars they contain. Most of them are not classified as BWs or RBs. This is reflected in $\dot{E}$, η_{GJ}, B_{lc}, and B_S; branches containing BWs and RBs show a wider distribution, shifted towards larger values for the dark green, gray, and pink branches, considering that the latter two contain different types of spider pulsars. In addition, Figure 6.7 can be compared with Figure 6.2. The former is more informative, as the clustering technique separates the nodes in groups that are significantly different from one another (see Table D.1 in Appendix D for details on the similarity criteria) beyond the known RBs and RBs.

6.3.1   Tracking the variables along the MST

In closing this Section, we note that the MST technique orders the pulsars according to one or several of the variables considered (this was discussed at length in Section 4.4.1). Figure 6.8 below serves as an example (data to construct similar figures for other paths are provided in the Pulsar Tree web that accompanies this work) of how the variables are ordered along a given path in the graph theory context. Figure 6.8 shows the values of P, $\dot{P}$, B_s, $\dot{E}$, M_C, A_1, and P_B along a path in the MST.

6.4   Localizing a new node

We introduce an incremental algorithm for updating the MST when new nodes are considered. This can infer where a new node with given properties will fall in the MST. When we add a new node and compute its distance from all other nodes in the graph, we preserve the known distances among the previously existing nodes. In this process, we keep the scaling of the distance used in calculating the original MST, denoted by T, to which the new node will be added. This involves identifying potential cycles that would be formed with the existing T, adding new edges, and discarding them so that the resulting graph remains an MST. While avoiding a complete recalculation of the MST, this approach provides the new node's location relative to the former ones, resulting in an updated T, which we shall refer to as T'. This process follows the principles outlined in traditional graph theory, e.g., the correctness of Prim's algorithm as per the MST theorem and the MST lemma as the cycle property (see Chapter 2 for technical details).

6.4.1   IGR J18245-2452—PSR J1824-2452I

Since tMSP IGR J18245-2452 (or PSR J1824-2452I, see, e.g., Papitto et al. 2013) lacks a measured value of the $\dot{P}$, it is initially excluded from the sample used in this work (see Section 6.2.1). Considering its measured values for P, P_B, M_C, and A_1, we shall assume a value for the $\dot{P}$ and add this pulsar to T, in Figure 6.4. Based on the estimated $B_s \sim (0.7 - 35) \times 10^8$ G by Ferrigno et al. (2014), and its known P, we can compute[4] a range of $\dot{P} \sim (1.2 \times 10^{-21} - 3 \times 10^{-18})$ss^{-1}. We take 10 equispaced values in this range (and their concurrent values of all physical variables derived taking $\dot{P}$ into account, and we construct for each one a T'.

When we assume the lowest values of the given range, $\dot{P} \lesssim 10^{-20}$ss^{-1}, we find that the added pulsar is located along one of the rightmost branches (brown branch, see Figure 6.6) of the MST, see the left panel in Figure 6.9. As $\dot{P}$ increases to the range $(10^{-20}, 3 \times 10^{-19})$ ss^{-1}, its MST location would climb until it reaches the end of the

---

[4]For consistency, we assume $B_s = 3.2 \times 10^{19} \sqrt{P\dot{P}}$ G
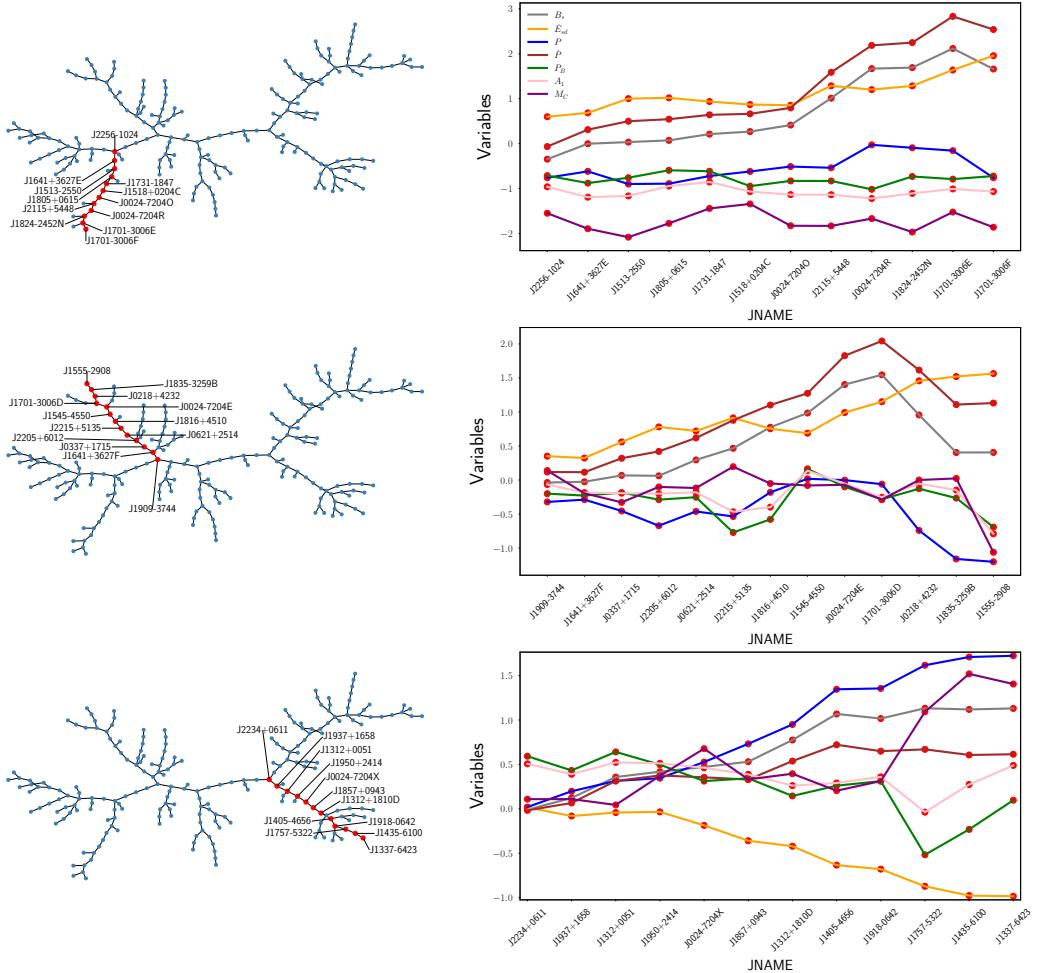
FIGURE 6.8: Examples of the values of P, $\dot{P}$, B_s, $\dot{E}$, M_C, A_1, and P_B (the legend shown in the first row applies to all rows) along a path in the MST. Left column: Selected paths and nodes in the MST are highlighted in red. Right column: The y-axis shows the scaled variables. The x-axis displays the pulsars along the selected path (nodes in red) by their identifier, according to the ATNF v2.1.1. The order shown responds to the path reading in the MST, which always goes from the central part (trunk) to the ends of the MST.
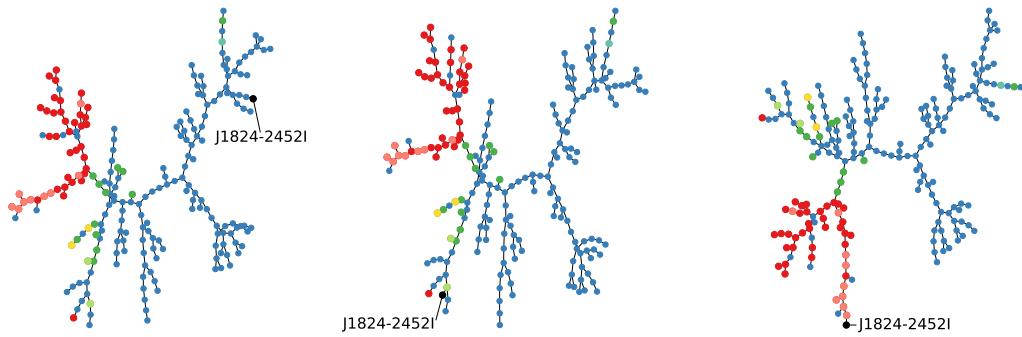
FIGURE 6.9: The MSTs containing different realizations for $\dot{P}$, $\dot{P} \sim$ 1.22×10^{-20}ss^{-1} (left panel); $\dot{P} \sim 3.34 \times 10^{-19}$ss^{-1} (central panel); $\dot{P} \sim 3 \times 10^{-18}$ss^{-1} (right panel), of the tMSP J1824-2452I (in black). Each MST is defined as a $T'(219, 218)$ and shown with the spider type of pulsars as seen in Figure 6.4, where the confirmed BWs appear in red, RBs in green, tMSPs in yellow, the BWs and RBs in globular clusters in light red and light green, the RB candidate in light teal, and keeping the unclassified ones in blue.

branch where the rest of the RBs, and therefore the known tMSPs, appear. It will remain at that end until it $\dot{P} \lesssim 6 \times 10^{-19}$ss^{-1} is considered, see the central panel in Figure 6.9. When this value of $\dot{P}$ is exceeded and up to the largest ones explored $\dot{P} \sim 3 \times 10^{-18}$ss^{-1}, the pulsar would be located in the BW branch as we can see in the right panel of Figure 6.9. Thus, as IGR J18245-2452 is an RB transitional pulsar, it would be reasonable to expect that it falls near the majority of the nodes of its class, implying we can limit searches for its $\dot{P}$ from $\dot{P} \sim (1.2 \times 10^{-21} - 3 \times 10^{-18})$ ss^{-1} to just around $\dot{P} \sim (1 \times 10^{-20}, 6 \times 10^{-19})$ ss^{-1} or $B_s \sim (2 \times 10^8, 1.55 \times 10^9)$G.

6.4.2   tMSP candidate: 3FGL J1544.6-1125

3FGL J1544.6-1125 has shown variability with high and low modes exclusively seen in tMSPs, see Bogdanov and Halpern (2015). In Britt et al. (2017), the P_B of this pulsar is estimated to be $P_B = 0.2415361(36)$ days $= 20868.72(31)$ s. The semi-amplitude of the radial velocity of the companion star $K_2 = 39.3 \pm 1.5$ km s^{-1} yields a companion mass $M_2 \lesssim 0.7\,M_\odot$. To estimate the range of A_1, we start defining the center of mass location, relying on the fundamental relationship $M_1 a_1 = M_2 a_2$ (see, e.g., Karttunen et al. 2007), where a_1 and a_2 represent the distances of the objects from the center of mass, and M_1 and M_2 are the masses of the neutron star and the companion star, respectively. Additionally, the radial velocity of the companion is defined as (see, e.g., Tauris and van den Heuvel 2023) $K_2 = \omega_B \cdot a_2 \cdot \sin(i)$, where $\omega_B = 2\pi/P_B$. By using the above relationships, we derive $A_1 = a_1 \sin(i) = \frac{K_2 \cdot P_B}{2\pi} \cdot \frac{M_2}{M_1}$. We make several assumptions, where for M_1 we have $M_{1min} \sim 1.4M_\odot$ and $M_{1max} \sim 2M_\odot$, and for M_2 we have $M_{2min} \sim 0.05M_\odot$ and $M_{2max} \sim 0.7M_\odot$. Consequently, we calculate the range $A_1 \sim (0.0088, 0.259)$ lt-s. We consider values from M_{2min} up to M_{2max} for the range of M_C. Also, as the pulsar has an unknown value of P, we impose no constraint on it and shall range it within $\sim (1, 10) \times 10^{-3}$ s. Similarly, we shall also inspect the range between the minimum and maximum measurement of $\dot{P}$ seen in the sample according to $(7 \times 10^{-23}, 7 \times 10^{-19})$ ss^{-1}. We consider 10 values spanning each range for $(P, \dot{P}, A_1, M_C)$, resulting in 10^4 distinct combinations for which we apply the incremental algorithm.
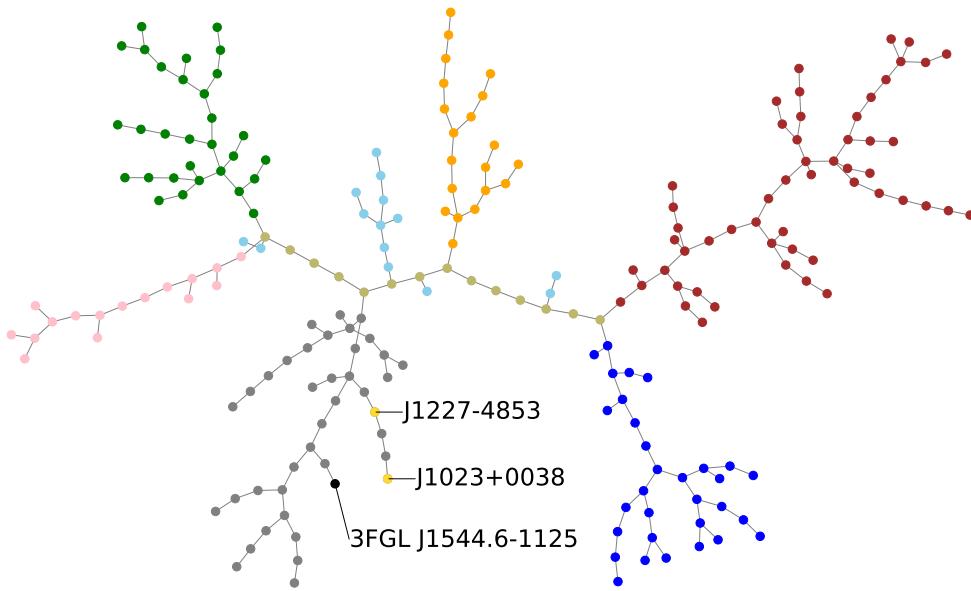
FIGURE 6.10: The MST defined as $T'(219, 218)$ containing the 3FGL J1544.6-1125 (in black) is shown with the known tMSPs (in yellow) labeled J1023+0038 and J1227-4853, as seen in Figure 6.4 and the color code of the significant branches, trunk, and the remaining pulsars as seen in Figure 6.6.

We observe that the structural stability[5] of T' is extremely high compared to T, since in more than 99% of the cases, they only differ on one edge. As seen in the previous analysis, this allows us to position 3FGL J1544.6-1125 considering known parts of T as the regions with the marked classes of pulsars as in Figure 6.4 or the branches seen in Figure 6.6. In 67% of the cases, the position of 3FGL J1544.6-1125 falls in a branch (gray), where the known tMSPs are located, as we show in Figure 6.10. Despite the uncertainties of the variables, the MST promotes the tMSP classification of 3FGL J1544.6-1125.

6.4.3 SAX J1808.4-3658

The location of SAX J1808.4-3658, an accreting millisecond X-ray pulsar (AMXP, see, e.g., Patruno and Watts 2021), within the MST is also of considerable interest. It has exhibited approximately ten month-long outbursts, with a recurrence of about 2–3 years, making it the AMXP with the most numerous outbursts, which is suitable for in-depth investigation of its long-term timing properties, as discussed by Illiano et al. (2023). SAX J1808.4-3658 is likely a gamma-ray source (de Oña Wilhelmi et al., 2016), and there is an ambiguity about its activity status as a rotation-powered millisecond pulsar during quiescent periods despite the lack of detected radio pulsations. Focusing on the long-term first derivative of the spin frequency, as detailed in Figure 2 and Section 3.2 of Illiano et al. (2023), reveals a $\dot{\nu} = -1.152(56) \times 10^{-15}\,\text{Hz}\,\text{s}^{-1}$, in alignment with findings from previous works (Patruno et al., 2012; Sanna et al., 2017; Bult et al., 2020). The variables taken are $P = 0.00249391976403(31)$ s, $A_1 =$

---

[5]The percentage of edges preserved between the original and updated MST after adding a new node. Although minor changes (even a single edge) can impact global properties, a high level of edge retention, such as the 99%, serves as a useful indicator of consistency.
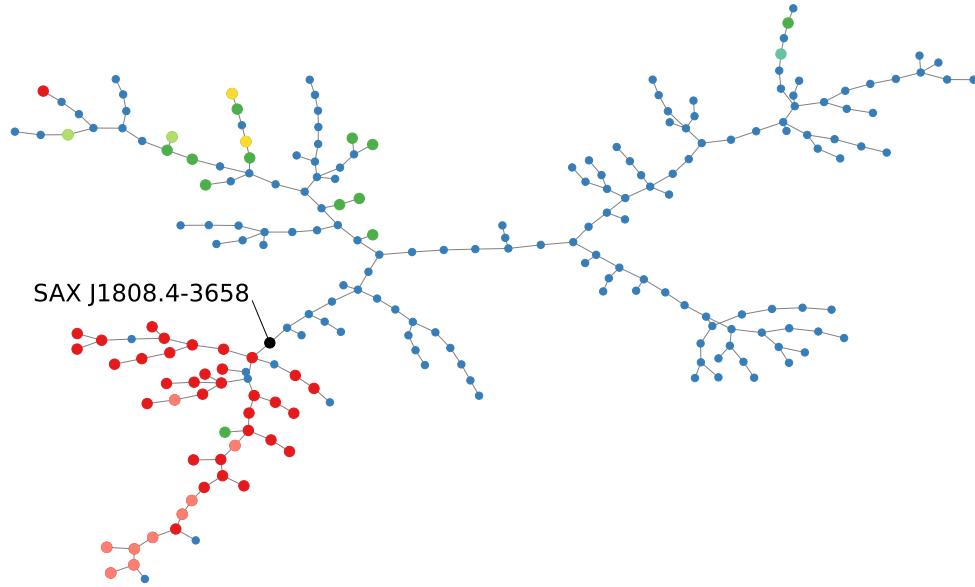
FIGURE 6.11: The MST defined as $T'(219, 218)$, containing the AMXP SAX J1808.4-3658 (in black) is shown. The confirmed BWs appear in red, RBs in green, tMSPs in yellow, the BWs and RBs in globular clusters in light red and light green, the RB candidate in light teal, and keeping the unclassified ones in blue, as seen in Figure 6.4.

0.0628033(57) lt-s, and the $P_B = 0.083902314(96)$ d (as listed in Table 1 of Illiano et al. 2023). Consequently, we estimate $\dot{P} = -\dot{\nu}/\nu^2 = -P^2\dot{\nu} \sim 7.2 \times 10^{-21}\text{s s}^{-1}$. Past timing analysis indicates that the neutron star orbits a semi-degenerate companion of $M_C \sim 0.05$ M$_\odot$ (see Bildsten and Chakrabarty 2001). Considering the values computed for the other variables, given the known spin period and spin period derivative, we apply the incremental algorithm to obtain T'. Using the BWs and RBs seen in Figure 6.4 as a reference, we show T' in Figure 6.11 together with the AMXP SAX J1808.4-3658, which falls at the gate of the high-density zone of BWs.

6.5   Concluding remarks

Graph theory provides an elegant way to inspect a population. A combination of the MST and the distance ranking underlying helps distinguish components and identify candidates for membership in a particular group. In this work, we have focused on the MSP population and observed that its MST clearly categorizes BWs and RBs into distinct categories. This separation is evident when plotting the M_C for each system. When the values of M_C are not clear-cut between these populations, it is increasingly complex to make a safe classification without further studies. Using the location in the MST, the individual distance ranking for a given pulsar, and the fact that in only very few cases, the ranking of a given pulsar thought to pertain to one class is dominated by pulsars belonging to another, suggests that

- We promote the BW classification of J1300+1240, J1630+3550, J1317-0157, J1221-0633, J1627+3219, J1701-3006F, and J1737-0314A.

- The MST location and the ranking favor, or at the very least do not rule out, a BW classification for J1908+2105.

- The pulsars J1723-2837, J1902-5105, and J1431-4715 are the only ones in the sample containing at least one known tMSP is in their top 3 neighbors. We promote further studies of these pulsars to observe them transitioning into an accretion-powered state, analogous to the known tMSPs.

- The MST location of IGR J18245-2452/PSR J1824-2452I as an RB transitional pulsar implies that the yet unknown $\dot{P}$ is in the range $\dot{P} \sim (1 \times 10^{-20}, 6 \times 10^{-19})\,\text{s}\,\text{s}^{-1}$. Searches limited to this range could empower the techniques and lead to detection.

- Spanning over the uncertainty in the variables of the putative transitional pulsar behind 3FGL J1544.6-1125, we see that it is located near the others in most cases.

- The AMXP SAX J1808.4-3658 is close to the group BWs and can be classified as such from its location in the MST. Indeed, di Salvo et al. (2008) proposed that the system is a hidden BW and that the source is ejecting matter in X-ray quiescence. Note that this fact alone would not necessarily imply its possible BW nature.

- If we use graph theory to cluster the population of MSPs via distinguishing branches (unsupervised methodology), we recover a clear distinction between BWs, RBs, and a significant number of other MSPs showing different features, larger P_B, M_C, P, and smaller $\dot{E}$ releases.

- This process also separates BWs in the field and these in globular clusters, placing them in different branches within the BW region.

TABLE 6.2: The truncated values according to the ATNF v2.1.1 (visit https://www.atnf.csiro.au/research/pulsar/psrcat/ for updated versions) of the 10 variables considered for the confirmed BWs, RBs, and tMSPs (candidates for BWs or RBs are placed below the corresponding line). The tMSPs are included in the RBs section and are denoted by an asterisk. Pulsars belonging to globular clusters are marked with a star. Note that the pulsar J1300+1240 (also known as PSR B1257+12, see Wolszczan and Frail 1992; Wolszczan 1994; Wolszczan et al. 2000) is a planetary system, and its orbital parameters can be considered as lower limits depending on the mass of its planets (see, e.g., Konacki and Wolszczan 2003).

JNAME	P_B(d)	$\dot{E}_{sd}$(erg s^{-1})	P(s)	B_s(G)	$\dot{P}$(s s^{-1})	M_C(M$_\odot$)	A_1(lt $-$ s)	$\Delta\Phi$(V)	η_{GJ}(cm^{-3})	B_{lc}(G)
					Black widows					
J0023+0923	0.13	1.58×10^{34}	0.0030	1.88×10^{8}	1.14×10^{-20}	0.018	0.034	1.33×10^{14}	4.28×10^{9}	62401.91
J0024-7204O*	0.13	6.48×10^{34}	0.0026	2.86×10^{8}	3.03×10^{-20}	0.024	0.045	2.69×10^{14}	7.50×10^{9}	145482.85
J0024-7204P*	0.14	5.41×10^{35}	0.0036	1.57×10^{9}	6.63×10^{-19}	0.019	0.038	7.77×10^{14}	2.98×10^{10}	305080.04
J0024-7204R*	0.06	1.38×10^{35}	0.0034	7.27×10^{8}	1.48×10^{-19}	0.029	0.033	3.93×10^{14}	1.44×10^{10}	161687.07
J0251+2606	0.20	1.82×10^{34}	0.0025	1.40×10^{8}	7.57×10^{-21}	0.027	0.065	1.42×10^{14}	3.82×10^{9}	80162.82
J0312-0921	0.09	1.53×10^{34}	0.0037	2.73×10^{8}	1.97×10^{-20}	0.010	0.015	1.30×10^{14}	5.10×10^{9}	50446.32
J0610-2100	0.28	8.45×10^{33}	0.0038	2.20×10^{8}	1.23×10^{-20}	0.024	0.073	9.71×10^{13}	3.95×10^{9}	35950.28
J0636+5128	0.06	5.76×10^{33}	0.0028	1.00×10^{8}	3.44×10^{-21}	0.007	0.008	8.02×10^{13}	2.42×10^{9}	39963.19
J0952-0607	0.26	6.66×10^{34}	0.0014	8.31×10^{7}	4.77×10^{-21}	0.022	0.062	2.72×10^{14}	4.06×10^{9}	275776.46
J1124-3653	0.22	1.69×10^{34}	0.0024	1.21×10^{8}	6.01×10^{-21}	0.031	0.079	1.37×10^{14}	3.49×10^{9}	81614.11
J1301+0833	0.27	6.64×10^{34}	0.0018	1.41×10^{8}	1.05×10^{-20}	0.027	0.078	2.72×10^{14}	5.29×10^{9}	211158.03
J1311-3430	0.06	4.93×10^{34}	0.0025	2.34×10^{8}	2.09×10^{-20}	0.009	0.010	2.34×10^{14}	6.33×10^{9}	130950.18
J1446-4701	0.27	3.66×10^{34}	0.0021	1.48×10^{8}	9.80×10^{-21}	0.021	0.064	2.02×10^{14}	4.67×10^{9}	131662.94
J1513-2550	0.17	8.96×10^{34}	0.0021	2.16×10^{8}	2.16×10^{-20}	0.018	0.040	3.16×10^{14}	7.06×10^{9}	213347.04
J1518+0204C*	0.08	6.71×10^{34}	0.0024	2.57×10^{8}	2.60×10^{-20}	0.043	0.057	2.73×10^{14}	7.17×10^{9}	157520.28

Table 6.2 (continued): Properties of the MSPs population.

JNAME	P_B(d)	$\dot{E}_{sd}$(erg s^{-1})	P(s)	B_s(G)	$\dot{P}$(s s^{-1})	M_C(M$_\odot$)	A_1(lt−s)	$\Delta\Phi$(V)	η_{GJ}(cm^{-3})	B_{lc}(G)
J1544+4937	0.12	1.09×10^{34}	0.0021	7.86×10^{7}	2.79×10^{-21}	0.019	0.032	1.10×10^{14}	2.51×10^{9}	73224.85
J1555-2908	0.23	3.07×10^{35}	0.0017	2.85×10^{8}	4.45×10^{-20}	0.059	0.151	5.86×10^{14}	1.10×10^{10}	468636.58
J1641+3627E*	0.11	4.47×10^{34}	0.0024	2.10×10^{8}	1.74×10^{-20}	0.023	0.037	2.23×10^{14}	5.86×10^{9}	128450.77
J1641+8049	0.09	4.27×10^{34}	0.0020	1.36×10^{8}	8.94×10^{-21}	0.046	0.064	2.18×10^{14}	4.65×10^{9}	154502.68
J1653-0158	0.05	1.24×10^{34}	0.0019	6.95×10^{7}	2.40×10^{-21}	0.011	0.010	1.17×10^{14}	2.44×10^{9}	85609.20
J1701-3006E*	0.15	3.62×10^{35}	0.0032	1.01×10^{9}	3.10×10^{-19}	0.035	0.070	6.36×10^{14}	2.16×10^{10}	281028.45
J1719-1438	0.09	1.63×10^{33}	0.0057	2.18×10^{8}	8.04×10^{-21}	0.001	0.001	4.27×10^{13}	2.60×10^{9}	10546.33
J1731-1847	0.31	7.78×10^{34}	0.0023	2.46×10^{8}	2.54×10^{-20}	0.038	0.120	2.94×10^{14}	7.28×10^{9}	179657.07
J1745+1017	0.73	5.77×10^{33}	0.0026	8.60×10^{7}	2.72×10^{-21}	0.015	0.088	8.03×10^{13}	2.24×10^{9}	43265.82
J1805+0615	0.33	9.31×10^{34}	0.0021	2.22×10^{8}	2.27×10^{-20}	0.026	0.087	3.22×10^{14}	7.23×10^{9}	216420.69
J1810+1744	0.15	3.97×10^{34}	0.0016	8.84×10^{7}	4.60×10^{-21}	0.049	0.095	2.10×10^{14}	3.68×10^{9}	181229.94
J1824-2452M*	0.24	4.42×10^{34}	0.0047	7.75×10^{8}	1.22×10^{-19}	0.012	0.032	2.22×10^{14}	1.12×10^{10}	66406.37
J1824-2452N*	0.19	1.66×10^{35}	0.0033	7.39×10^{8}	1.59×10^{-19}	0.021	0.049	4.31×10^{14}	1.52×10^{10}	183891.70
J1833-3840	0.90	1.07×10^{35}	0.0018	1.84×10^{8}	1.77×10^{-20}	0.009	0.061	3.46×10^{14}	6.82×10^{9}	265593.33
J1836-2354A*	0.20	2.42×10^{33}	0.0033	8.92×10^{7}	2.31×10^{-21}	0.019	0.046	5.20×10^{13}	1.84×10^{9}	22164.68
J1928+1245	0.13	2.40×10^{34}	0.0030	2.27×10^{8}	1.67×10^{-20}	0.010	0.018	1.63×10^{14}	5.21×10^{9}	77477.09
J1959+2048	0.38	1.60×10^{35}	0.0016	1.66×10^{8}	1.68×10^{-20}	0.024	0.089	4.23×10^{14}	7.16×10^{9}	375949.48
J2017-1614	0.09	7.80×10^{33}	0.0023	7.61×10^{7}	2.45×10^{-21}	0.030	0.043	9.33×10^{13}	2.27×10^{9}	57631.66
J2047+1053	0.12	1.04×10^{34}	0.0042	3.02×10^{8}	2.08×10^{-20}	0.040	0.069	1.07×10^{14}	4.87×10^{9}	35979.39
J2051-0827	0.09	5.48×10^{33}	0.0045	2.42×10^{8}	1.27×10^{-20}	0.030	0.045	7.82×10^{13}	3.72×10^{9}	24801.53
J2052+1219	0.11	3.38×10^{34}	0.0019	1.16×10^{8}	6.70×10^{-21}	0.038	0.061	1.94×10^{14}	4.06×10^{9}	139873.60
J2055+3829	0.12	4.32×10^{33}	0.0020	4.62×10^{7}	9.99×10^{-22}	0.025	0.045	6.95×10^{13}	1.53×10^{9}	47537.98
J2115+5448	0.13	1.67×10^{35}	0.0026	4.46×10^{8}	7.49×10^{-20}	0.024	0.044	4.32×10^{14}	1.18×10^{10}	237535.29
J2214+3000	0.41	1.91×10^{34}	0.0031	2.16×10^{8}	1.47×10^{-20}	0.015	0.059	1.46×10^{14}	4.81×10^{9}	67001.37
J2234+0944	0.41	1.66×10^{34}	0.0036	2.73×10^{8}	2.01×10^{-20}	0.017	0.068	1.36×10^{14}	5.21×10^{9}	53684.04
J2241-5236	0.14	2.60×10^{34}	0.0021	1.24×10^{8}	6.89×10^{-21}	0.013	0.025	1.70×10^{14}	3.93×10^{9}	111420.40
J2256-1024	0.21	3.71×10^{34}	0.0022	1.63×10^{8}	1.13×10^{-20}	0.034	0.082	2.03×10^{14}	4.92×10^{9}	126750.83
J2322-2650	0.32	5.54×10^{32}	0.0034	4.54×10^{7}	5.83×10^{-22}	0.0008	0.002	2.48×10^{13}	9.08×10^{8}	10266.96
J1221-0633	0.38	2.87×10^{34}	0.0019	1.02×10^{8}	5.26×10^{-21}	0.015	0.05	1.79×10^{14}	3.65×10^{9}	132280.06
J1300+1240†	25.2	1.87×10^{34}	0.0062	8.53×10^{8}	1.14×10^{-19}	5.6×10^{-8}	3×10^{-6}	1.44×10^{14}	9.49×10^{9}	33265.14
J1317-0157	0.08	8.82×10^{33}	0.0029	1.27×10^{8}	5.49×10^{-21}	0.020	0.02	9.92×10^{13}	3.04×10^{9}	48767.61
J1627+3219	0.16	2.07×10^{34}	0.0021	1.10×10^{8}	5.47×10^{-21}	0.025	0.05	1.52×10^{14}	3.50×10^{9}	99738.68
J1630+3550	0.31	2.44×10^{34}	0.0032	2.62×10^{8}	2.08×10^{-20}	0.011	0.03	1.65×10^{14}	5.62×10^{9}	73150.90
J1701-3006F*	0.20	7.25×10^{35}	0.0022	7.22×10^{8}	2.22×10^{-19}	0.024	0.05	9.00×10^{14}	2.17×10^{10}	560490.08
J1737-0314A*	0.22	4.86×10^{35}	0.0019	4.40×10^{8}	9.55×10^{-20}	0.018	0.04	7.36×10^{14}	1.53×10^{10}	531633.70
Redbacks and transitional millisecond pulsars										
J1023+0038*	0.19	5.68×10^{34}	0.0016	1.09×10^{8}	6.92×10^{-21}	0.15	0.34	2.52×10^{14}	4.48×10^{9}	213328.66
J1048+2339	0.25	1.16×10^{34}	0.0046	3.79×10^{8}	3.00×10^{-20}	0.35	0.83	1.14×10^{14}	5.62×10^{9}	35000.56
J1227-4853*	0.28	9.12×10^{34}	0.0016	1.38×10^{8}	1.10×10^{-20}	0.24	0.66	3.19×10^{14}	5.67×10^{9}	270534.19
J1431-4715	0.44	6.83×10^{34}	0.0020	1.70×10^{8}	1.41×10^{-20}	0.14	0.55	2.76×10^{14}	5.86×10^{9}	196263.68
J1622-0315	0.16	7.93×10^{33}	0.0038	2.12×10^{8}	1.14×10^{-20}	0.11	0.21	9.41×10^{13}	3.81×10^{9}	34977.90
J1628-3205	0.21	1.42×10^{34}	0.0032	1.98×10^{8}	1.19×10^{-20}	0.18	0.41	1.26×10^{14}	4.27×10^{9}	56116.06
J1717+4308A*	0.20	7.65×10^{34}	0.0031	4.44×10^{8}	6.11×10^{-20}	0.18	0.39	2.92×10^{14}	9.73×10^{9}	132169.23
J1723-2837	0.61	4.65×10^{34}	0.0018	1.19×10^{8}	7.54×10^{-21}	0.27	1.22	2.28×10^{14}	4.46×10^{9}	175618.45
J1740-5340A*	1.35	1.36×10^{35}	0.0036	7.92×10^{8}	1.68×10^{-19}	0.21	1.65	3.90×10^{14}	1.50×10^{10}	152737.57
J1803-6707	0.38	7.49×10^{34}	0.0021	2.00×10^{8}	1.84×10^{-20}	0.33	1.06	2.89×10^{14}	6.51×10^{9}	193679.03
J1816+4510	0.36	5.22×10^{34}	0.0031	3.75×10^{8}	4.31×10^{-20}	0.18	0.59	2.41×10^{14}	8.13×10^{9}	108100.42
J1908+2105	0.14	3.23×10^{34}	0.0025	1.90×10^{8}	1.38×10^{-20}	0.06	0.11	1.90×10^{14}	5.14×10^{9}	105961.15
J1957+2516	0.23	1.74×10^{34}	0.0039	3.33×10^{8}	2.74×10^{-20}	0.11	0.28	1.39×10^{14}	5.82×10^{9}	50306.20
J2039-5617	0.22	3.00×10^{34}	0.0026	1.96×10^{8}	1.41×10^{-20}	0.19	0.47	1.83×10^{14}	5.11×10^{9}	98678.12
J2215+5135	0.17	7.41×10^{34}	0.0026	2.98×10^{8}	3.33×10^{-20}	0.24	0.46	2.87×10^{14}	7.91×10^{9}	157526.43
J2339-0533	0.19	2.32×10^{34}	0.0028	2.04×10^{8}	1.41×10^{-20}	0.30	0.61	1.60×10^{14}	4.89×10^{9}	79742.09
J1902-5105*	2.01	6.86×10^{34}	0.0017	1.28×10^{8}	9.19×10^{-21}	0.18	1.90	2.76×10^{14}	5.08×10^{9}	227053.37
J1302-3258	0.78	4.81×10^{33}	0.0037	1.58×10^{8}	6.54×10^{-21}	0.17	0.92	7.33×10^{13}	2.91×10^{9}	27786.94

Chapter 7

Separating repeating fast radio bursts using the minimum spanning tree as an unsupervised methodology

7.1 Introduction

Fast Radio Bursts (FRBs) are transient radio pulses, typically lasting on the order of milliseconds and originating from cosmological distances (see Section 1.5). They have attracted considerable interest and numerous physical interpretations; see, e.g., Katz (2018), Popov, Postnov, and Pshirkov (2018), Cordes and Chatterjee (2019), Petroff, Hessels, and Lorimer (2019), Platts et al. (2019), Zhang (2020), Xiao, Wang, and Dai (2021), Petroff, Hessels, and Lorimer (2022), Xiao, Wang, and Dai (2022), and Zhang (2023). Their observed properties, such as their dispersion (DM) and rotation (RM) measures, provide clues into the physical characteristics of the intervening medium, including density and magnetic field strength. Their energy or brightness temperature also offers valuable hints regarding the underlying mechanisms driving their emission.

After the discovery of the first repeater in 2016, FRB 20121102A (Spitler et al., 2016; Scholz et al., 2016), FRBs are broadly categorized into two groups based on their repetition properties: repeaters and (apparently) non-repeaters. However, synthesizing the physical causes behind these differences and escaping from often-not-fully-understood observational biases remains a significant challenge. See Chapter 1 for a detailed discussion of FRBs.

In this chapter based on the work (García et al., 2024b), we aim to explore a novel classification tool for FRBs, separating them into repeaters and non-repeaters (for other approaches, see, e.g., Palaniswamy, Li, and Zhang, 2018; Caleb et al., 2019; Ai, Gao, and Zhang, 2021; Luo, Zhu-Ge, and Zhang, 2023; Zhu-Ge, Luo, and Zhang, 2023; Sun et al., 2025). We utilize graph theory (see Chapter 2) as a framework for this methodology, and PCA (see Chapter 3) as an alternative to describe the properties of FRBs. In particular, we introduce the MST as an unsupervised learning approach with a supervised evaluation. The MST, where each node will represent an FRB, provides a manageable and compact structure calculated in an N-dimensional space. Aided by the MST's fundamental properties, we will explore its capacity to establish the variables with the best separation power and then indicate a group of likely repeating FRBs that have yet to be classified. Using an unsupervised learning approach is particularly appealing when one does not control which properties within a data set drive a classification, as is the case here. In addition, even the labels

themselves are subject to possible change in time, as non-repeaters could become repeaters. Also, such methods do not necessarily require sample splitting for proper application. This is valuable since the sample is insignificant and contains at most 750 FRBs, so we are not dealing with big data. Randomly cutting the sample in a training set (say, 80%/20% in size) would imply that 150 FRBs are sufficient to test the pattern behind a repeater/non-repeater classification, which we consider a risk. Finally, the MST, as an unsupervised method, provides transparency in identifying the variables that best distinguish repeaters from non-repeaters. This becomes less clear in supervised methods when techniques such as feature importance (see, e.g., Breiman, 2001; Altmann et al., 2010; Casalicchio, Molnar, and Bischl, 2018; Fisher, Rudin, and Dominici, 2018) or Shapley Additive exPlanations (SHAP, see Lundberg and Lee, 2017; Lundberg, Erion, and Lee, 2018 for more details) are applied for that purpose.

7.2   Sample, variables, PCA, and MST

We use the sample from the Canadian Hydrogen Intensity Mapping Experiment Fast Radio Burst (CHIME) catalog (CHIME/FRB Collaboration et al., 2021; The Chime/FRB Collaboration et al., 2023). We consider their 750 FRBs, comprising 265 repeaters and 485 non-repeaters. The repeaters group corresponds to events from 70 localizations. Each FRB is considered individually as a distinct event, labeled as a repeater (as in CHIME). Following the CHIME catalog, sub-bursts can be attributed to repeaters and non-repeaters. They are also considered individual bursts, each with the corresponding label. We exclude the six FRBs for which neither flux nor fluence has been measured.

To characterize an FRB, we shall consider the logarithm of the following variables (this is done just for convenience, as their values may differ by several orders of magnitude for different FRBs): (CHIME/FRB Collaboration et al., 2021; The Chime/FRB Collaboration et al., 2023)

- peak frequency ν_c (MHz),

- flux S_ν (Jy),

- fluence F_ν (Jy ms),

- boxcar width Δt_{BC} (ms),

and also derived parameters based on the DM-z relation (as presented in Luo, Zhu-Ge, and Zhang (2023) and Zhu-Ge, Luo, and Zhang (2023) and references therein) such as:

- redshift z,

- rest-frame frequency width $\Delta \nu$ (MHz),

- rest-frame width Δt_r (ms),

- burst energy E (erg),

- luminosity L (erg/s), and

- brightness temperature T_B (K).
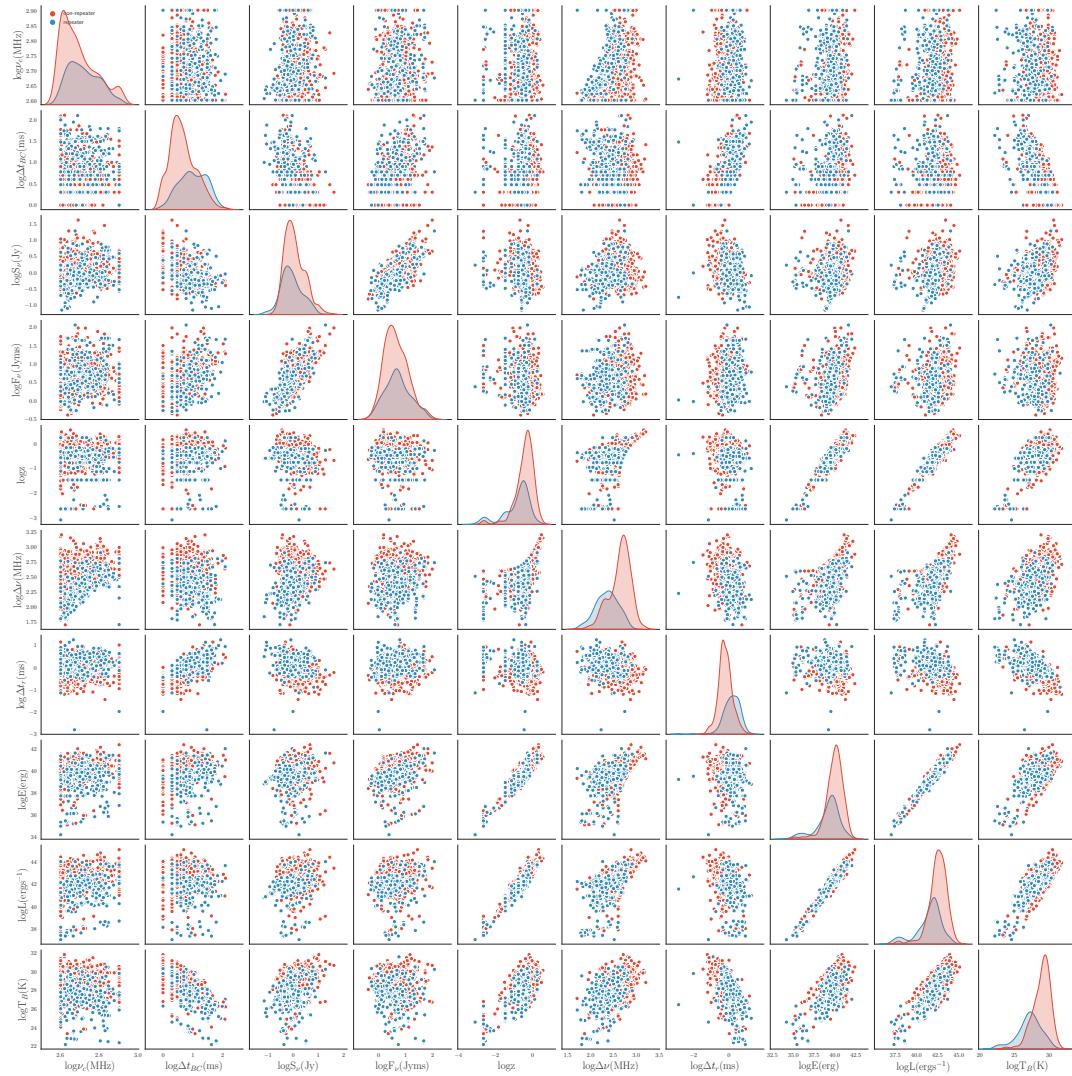
FIGURE 7.1: Cross-correlation of the logarithm of the 10 magnitudes considered. Repeaters (in blue) and non-repeaters (in red) are separately noted. The main diagonal shows the distribution for each variable.

As their distribution is not normal (nor is it that of the original variables without logarithm), we use the robust scaler to scale them (see Equation (4.1)).

As we can observe from the cross-correlations shown in Figure 7.1, the physical properties exhibit different classification powers. This can be corroborated by applying a simple separation method: for each of the magnitudes considered, we note the median of each distribution and count the number of repeaters positioned on each side of this central location. The rest-frame frequency width provides the best separation, yielding 217 (82%) grouped FRB repeaters (all with a width lower than 397.24 MHz). This is far off from the random expectation: After producing 10^5 simulations in which repeater labels are randomly assigned, the average separation achieved is 52%, that is, a separation consistent with an entirely random process in which labels are separated in equal numbers on both sides of the median. This high separation power of the rest-frame frequency width will serve as a threshold to validate the performance of other algorithms or classifiers. We shall look below for classifiers that will perform better than 82% in separating repeaters from non-repeaters.
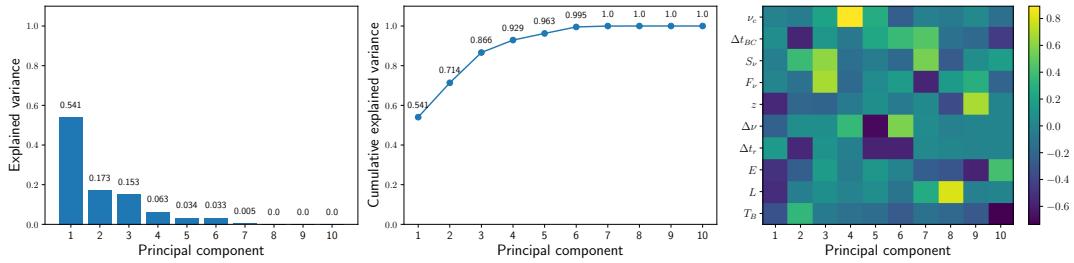
FIGURE 7.2: PCA results for the logarithm of the set of variables defined for the population of 750 FRBs, see Section 3.2.3 for further explanations about PCA definitions.

Figure 7.2 presents the outcome of conducting PCA over the logarithm of all the variables defined in this section, see Chapter 3 for more details. While the first six PCs contain the majority of variance, retaining over 99% Full variance coverage needs nine PCs, so the dimensionality reduction is not extreme if it is to be entirely retained. Thus, the number of PCs needed to cover the total variance, combined with the relatively flat distribution of the loadings of the first PCs, renders its use a priori unappealing. In other words, if almost all PCs are needed (9 out of 10), dimensionality is not significantly reduced. If the loadings are flat, no strong directional variance is identified, thus no dominant underlying structure appears. This justifies saying that PCA does not extract meaningful simplification or patterns.

Considering the MST we shall introduce the Euclidean distance (see Equation (2.1)), using all or part of the defined physical variables mentioned above, or the PCs (employing the PCs that account for 100% of the explained variance yields an MST identical to the one derived using all magnitudes). We initially generate a complete, undirected, weighted graph $G(V, E)$, characterized by a set V of nodes (each of them is an FRB) and set E of edges to define the relationships between the FRBs. The latter are assigned to a specific weight w equal to the Euclidean distance between the FRBs, using all or part of the variables or PCs, as described. We will start from the graph $G(750, 280875)$ for each combination of variables and derive an MST, $T(750, 749)$.

7.3 Algorithm and results

The total number of combinations (without repetition) from a set of 10 variables is $\sum_1^{10} \binom{10}{k} = 2^{10} - 1 = 1023$, reflecting the number of MSTs we compute. For each of these MSTs, we identify the node with the highest centrality of the graph through the betweenness centrality estimator (see Section 2.4.1). This estimator determines the frequency at which a particular node appears on the shortest path between any two other nodes, thereby measuring the node's importance in terms of connectivity within the MST. Eliminating this node would separate T into several connected parts (or branches). The branch with the highest number of repeaters will be named the repeater branch. All other branches will be considered as non-repeater branches. A high density of repeaters in the repeater branch would qualify the variables under which the underlying Euclidean distance is built as a well-separated filter via the MST algorithm. To judge the performance of each MST seen as a classifier, we use the metrics precision, recall, F_1, and F_2 score plus the area under the curve (AUC or ROC-AUC), see Appendix F for a more detailed explanation. Precision measures the ratio of correctly identified repeaters regarding the repeater branch's total FRBs

(both repeaters and non-repeaters). At the same time, recall assesses the proportion of actual repeaters that are correctly identified. The F_1 score balances precision and recall equally, whereas the F_2 score places more on recall. AUC evaluates the classifier's overall discrimination ability.In identifying high-density locations of repeaters, metrics such as recall and precision focus on correctly identifying actual repeaters and minimizing non-repeaters that are misclassified as repeaters. A high F_1 score indicates a good balance between recall and precision. On the other hand, the F_2 score, as seen, places more emphasis on recall than precision, effectively tolerating a certain level of misclassification of non-repeaters as repeaters. Lastly, a high AUC suggests that MST predictions rank actual repeater instances higher than non-repeaters, which is crucial for identifying many repeaters among the whole sample. To determine the best combination of variables, we look for a combination that offers a balanced trade-off between precision and recall (as reflected in the F_1 Score), also considering the F_2 Score (which weighs recall higher than precision) and the ROC-AUC value for overall performance. This is, in practice, implemented by an overall rank (see, e.g., Li, Wang, and Xiao, 2017), which averages all ranks across these evaluating metrics.

Armed with the MST/betweenness centrality methodology and the rankings, we find 25 combinations of variables and three combinations of PCs for which we obtain a separating, i.e., recall better than 82%. Results are shown in Table 7.1. Recall values are notably high across the board, which is excellent if the primary concern is to minimize missing repeaters. In some combinations, precision is lower than recall, which is expected when recall is prioritized. The F_1 scores are still reasonably good, indicating a balanced performance between precision and recall. The F_2 scores are also high, reinforcing the model orientation towards recall, which is desirable in contexts where the consequences of missing repeaters outweigh those of non-repeaters seen as repeaters. The ROC-AUC values are also far from what could be considered a random guess, showing a reasonable degree of separability. Based on the overall rank, a combination of peak frequency, rest-frame frequency width, and brightness temperature (#1) achieves the best balance. Although this is not the best combination for recall, for which it shows 0.8528, it is not far from the best 0.8604 (a difference of just two repeaters) obtained with the combination (#7). Also, it shows a precision of 0.5825, the second highest, which, together with the recall above, makes it the combination with the best F_1 score, F_2 score, and ROC value.

Table 7.1 shows the results obtained using the PCs. Recalling that we need 9 PCs to define the entire variance, we consider $\sum_1^9 \binom{9}{k} = 2^9 - 1 = 511$. Only three combinations of PCs present results exceeding the recall threshold, although they are not particularly informative or ranked better than those obtained directly using physical variables.

We consistently observe high recall rates, exceeding 0.82 across all combinations, which highlights their effectiveness in accurately identifying repeaters. The precision values, ranging from 0.51 to 0.55, and corresponding F_1 scores between 0.63 and 0.67, suggest a reasonable balance between the accuracy of robust predictions and the method's ability to identify actual repeaters. Furthermore, the F_2 scores, centered around 0.75, emphasize the evaluation's focus on recall over precision, aligning with our classification goal to reduce misclassified repeaters. The approximately 0.65 to 0.68 ROC-AUC values demonstrate the MST based on PCs' robust capability to distinguish between repeaters and non-repeaters, indicating significant discriminative power. Considering these points through the overall rank, the combination (PC_1, PC_2, PC_3) emerges as the most effective. This represents 87% of the total variance and has the best separation power of all those using principal components.

However, describing more variance with more PCs does not yield better separation power, implying that something other than the variables used to describe the sample here likely plays a role in the separation.

7.3.1   Significance testing via Monte Carlo label permutations

To evaluate whether the observed separation power could arise by chance, we test the randomness of the process in two ways, employing a Monte Carlo permutation test considering the labels of the FRBs. First, we question whether, if there is no link between repeating/non-repeating labels and the variables, we could still find a group of 25 combinations out of 1023 possible ones that could lead to a separating power of over 82%. We do so by randomly relabelling the FRBs in repeaters and non-repeaters, maintaining the sample proportions, and building 10^5 fake samples for each MST using the same variables as the original.

In any of the simulations performed, no combination of variables out of the 1023 possible exceeded the threshold of 82% separation. Second, for each combination selected in Table 7.1, we compare the actual result (i.e., its real separation power) with the average recall of the fake samples using the Z-score. This statistical measure quantifies how many standard deviations a data point (in this case, the recall) is from the mean of the distribution (see the Appendix F for more details). Table 7.1 shows that Z-scores obtained are at 9σ or higher for recall values, along all combinations, compared to distributions from fake sets. Additionally, considering the PCs, the high Z-scores of 10 and 12 across the combinations underscore the statistical significance of these results, affirming that the observed recall rates are not due to chance. These randomness tests are a strong indicator of the effectiveness and robustness of this classification approach. In particular, the randomness of the best-ranked combination of variables (#1), viewed in the Z-score of the first row in Table 7.1, is one of the lowest, which proves a robust classification.

Combination #1 successfully identified a significant number of repeaters (226/265), ensuring that relatively few repeaters are missed (39/265). Some non-repeaters are located in the repeater branch of the MST (162/485). If there is any physical link between the variables and the repeating property, some of them should be candidates to appear as repeaters in the next catalogs, depending on duty cycles. We will return to this below. Figure 7.3 shows the MST with the arrangement of repeaters and non-repeaters of combination #1; it separates the two types of FRBs.

7.3.2   Results considering selection effects

As explained in CHIME/FRB Collaboration et al. (2021), assessing selection effects is a challenge to be addressed in the study of FRBs. We have approached this question under the considerations outlined in Section 3.3 of The Chime/FRB Collaboration et al. (2023). The following cuts have been imposed in the sample:

1. Events measured by the *bonsai*[1] real-time detection pipeline S/N < 12 are excluded due to being more likely misclassified as noise.

2. Events with DM $< 1.5 \max(\mathrm{DM_{NE2001}}, \mathrm{DM_{YMW16}})$ are excluded to minimize the possibility of having a wrong identification of rotating radio transients or radio pulsars as FRBs considering that a $\sim 50\%$ errors may be common in models used to estimate the Galactic DM.

---

[1]FRB search code implemented to address computational challenges and generates SNR estimations, see CHIME/FRB Collaboration et al. (2018) and reference therein for more details.

TABLE 7.1: Summary of performance metrics of the combination of variables selected as those exceeding the threshold of 0.82 recall is seen in the rest-frame frequency width, sorted by overall rank. The combinations seen in the 'variables' column are named as follows: (1): $\log \nu_c$, (2): $\log \Delta t_{BC}$, (3): $\log S_\nu$, (4): $\log F_\nu$, (5): $\log z$, (6): $\log \Delta \nu$, (7): $\log \Delta t_r$, (8): $\log E$, (9): $\log L$, and (10): $\log T_B$.

#	Variables	Precision	Recall	F_1 Score	F_2 Score	ROC-AUC	Z Score
1	[1, 6, 10]	0.5825	0.8528	0.6922	0.7804	0.7773	14
2	[1, 5, 6, 7]	0.5622	0.8528	0.6777	0.7729	0.7033	13
3	[1, 2, 5, 6, 10]	0.5589	0.8415	0.6717	0.7642	0.7285	13
4	[1, 3, 4, 6, 7, 8, 9, 10]	0.5784	0.8491	0.6881	0.7764	0.6834	13
5	[1, 5, 6, 8, 10]	0.5878	0.8340	0.6895	0.7695	0.6961	14
6	[6]	0.5722	0.8226	0.6749	0.7564	0.7522	19
7	[1, 2, 3, 6, 7, 8]	0.5352	0.8604	0.6599	0.7672	0.6941	12
8	[1, 3, 5, 6, 9]	0.5510	0.8566	0.6706	0.7711	0.6577	13
9	[1, 2, 3, 4, 6, 7, 9, 10]	0.5457	0.8340	0.6597	0.7543	0.7007	12
10	[2, 5, 10]	0.5598	0.8302	0.6687	0.7571	0.6999	13
11	[1, 3, 5, 6]	0.5509	0.8377	0.6647	0.7587	0.6645	12
12	[1, 4, 5, 6, 9]	0.5688	0.8264	0.6738	0.7578	0.6583	13
13	[5, 6, 9, 10]	0.5198	0.8415	0.6427	0.7488	0.6883	11
14	[8, 10]	0.5298	0.8377	0.6491	0.7505	0.6859	11
15	[2, 6, 8]	0.5366	0.8302	0.6519	0.7483	0.7006	12
16	[1, 2, 4, 6, 7, 9, 10]	0.5239	0.8264	0.6413	0.7409	0.7230	11
17	[1, 2, 4, 5, 6, 10]	0.5201	0.8302	0.6395	0.7417	0.7118	11
18	[1, 4, 6, 9, 10]	0.5116	0.8302	0.6331	0.7383	0.7013	10
19	[1, 2, 3, 6, 9, 10]	0.5117	0.8226	0.6310	0.7335	0.7180	10
20	[1, 2, 4, 5, 6, 9, 10]	0.5034	0.8302	0.6268	0.7348	0.7002	10
21	[1, 2, 3, 4, 6, 7, 8, 9]	0.4966	0.8377	0.6236	0.7366	0.6590	10
22	[1, 2, 4, 5, 6]	0.4944	0.8340	0.6208	0.7332	0.6866	10
23	[1, 6, 8, 9, 10]	0.4846	0.8302	0.6120	0.7266	0.7026	9
24	[1, 5, 6, 9, 10]	0.4944	0.8302	0.6197	0.7309	0.6987	10
25	[1, 4, 5, 6, 9, 10]	0.4943	0.8226	0.6176	0.7262	0.6851	10
1	[PC_1, PC_2, PC_3]	0.5504	0.8453	0.6667	0.7635	0.6858	12
2	[PC_1, PC_7]	0.5178	0.8226	0.6356	0.7360	0.6716	11
3	[$PC_1, PC_3, PC_6, PC_7, PC_8$]	0.5092	0.8377	0.6334	0.7420	0.6523	11
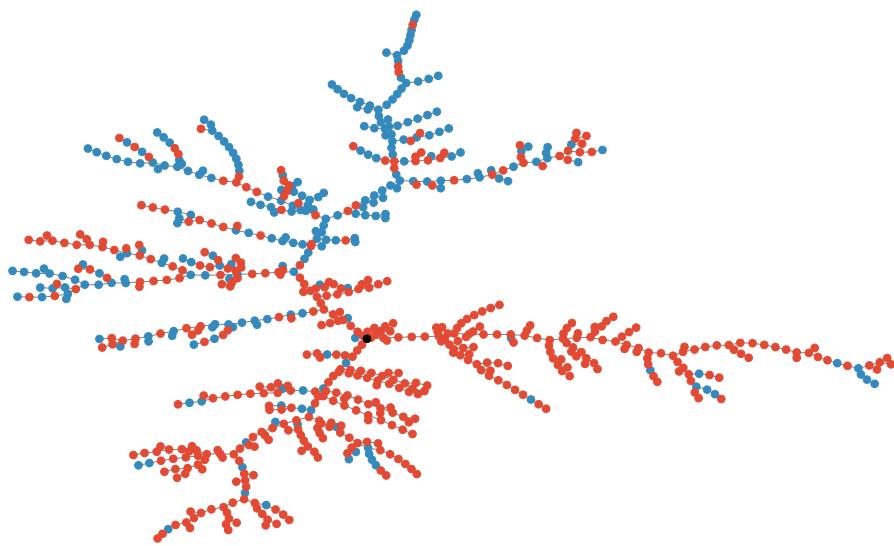
FIGURE 7.3: $T(750, 749)$ computed from the Euclidean distance based on the combination of peak frequency, rest-frame frequency width, and brightness temperature. Repeaters are shown in blue, and non-repeaters in red. The most central node in terms of betweenness centrality appears in black.

TABLE 7.2: Properties of the currently labeled non-repeater FRBs seen as repeaters for all combinations shown in Table 7.1. The "Sub" column shows the sub-burst number to which each FRB corresponds. Candidate FRBs reported in Luo, Zhu-Ge, and Zhang (2023) and Zhu-Ge, Luo, and Zhang (2023) are highlighted in black. Repeaters not classified as such in any of the selected combinations (outliers) are shown below the line.

Name	Sub	RA (°)	Dec (°)	ν_c (MHz)	Δt_{BC} (ms)	S_ν (Jy)	F_ν (Jy ms)	z	$\Delta\nu$ (MHz)	Δt_r (ms)	E (erg)	L (erg/s)	T_B (K)
FRB20180907E	0	167.88	47.09	400.20	11.80	0.73	6.90	0.3118	178.54	3.16	7.11×10^{39}	9.87×10^{41}	7.74×10^{27}
FRB20180920B	0	191.09	63.52	421.10	10.81	0.35	1.70	0.4007	116.53	1.66	3.10×10^{39}	8.93×10^{41}	6.71×10^{27}
FRB20180928A	0	312.95	30.85	400.20	2.95	1.34	2.50	0.0022	92.11	0.27	1.20×10^{35}	6.44×10^{37}	1.06×10^{25}
FRB20181017B	0	237.76	78.50	593.20	12.78	1.06	6.50	0.2067	247.97	1.91	4.24×10^{39}	8.35×10^{41}	1.86×10^{27}
FRB20181022E	0	221.18	27.13	443.70	2.95	0.69	2.08	0.2073	193.28	0.33	1.02×10^{39}	4.09×10^{41}	4.09×10^{28}
FRB20181125A	2	147.94	33.93	434.50	14.75	0.39	3.20	0.1710	156.33	1.09	1.04×10^{39}	1.48×10^{41}	6.49×10^{26}
FRB20181125A	1	147.94	33.93	436.60	14.75	0.39	3.20	0.1710	177.76	1.23	1.04×10^{39}	1.49×10^{41}	6.43×10^{26}
FRB20181125A	0	147.94	33.93	426.50	14.75	0.39	3.20	0.1710	141.34	1.35	1.02×10^{39}	1.45×10^{41}	6.73×10^{26}
FRB20181214A	0	70.00	43.07	435.00	2.95	0.156	0.41	0.2308	116.19	0.43	2.47×10^{38}	1.15×10^{41}	1.20×10^{28}
FRB20181220A	0	346.11	48.43	400.20	2.95	1.33	3.00	0.0022	196.64	0.43	1.44×10^{35}	6.39×10^{37}	1.05×10^{25}
FRB20181226E	0	303.56	73.64	400.20	2.95	0.48	1.35	0.1779	186.23	0.99	4.37×10^{38}	1.83×10^{41}	2.55×10^{28}
FRB20181229B	0	238.37	19.78	445.50	20.64	0.42	4.90	0.3197	154.80	2.55	5.92×10^{39}	6.70×10^{41}	1.24×10^{27}
FRB20190112A	0	257.98	61.20	697.70	9.83	1.40	16.20	0.3476	317.48	1.22	3.64×10^{40}	4.24×10^{42}	8.80×10^{27}
FRB20190128C	0	69.80	78.94	491.60	15.73	0.71	5.90	0.1772	238.50	5.23	2.32×10^{39}	3.29×10^{41}	8.73×10^{26}
FRB20190206B	0	49.76	79.50	506.40	19.66	0.95	9.60	0.2190	350.34	5.82	6.03×10^{39}	7.27×10^{41}	1.09×10^{27}
FRB20190206A	0	244.85	9.36	534.50	5.90	1.40	9.10	0.0618	213.84	0.76	4.53×10^{38}	7.40×10^{40}	1.20×10^{27}
FRB20190218B	0	268.70	17.93	588.00	17.69	0.57	5.90	0.4416	334.17	1.42	1.84×10^{40}	2.56×10^{42}	2.56×10^{27}
FRB20190223A	0	64.72	87.65	444.80	3.93	0.47	1.58	0.2865	149.61	0.59	1.52×10^{39}	5.81×10^{41}	3.05×10^{28}
FRB20190308C	1	188.36	44.39	453.40	21.63	0.47	4.80	0.4542	218.85	0.28	1.22×10^{40}	1.74×10^{42}	2.51×10^{27}
FRB20190308C	0	188.36	44.39	449.00	21.63	0.47	4.80	0.4542	211.58	0.38	1.21×10^{40}	1.72×10^{42}	2.56×10^{27}
FRB20190323D	0	56.88	46.93	400.20	12.78	0.37	2.49	0.5930	204.86	3.41	9.66×10^{39}	2.29×10^{42}	1.26×10^{28}
FRB20190329A	0	65.54	73.63	432.30	11.80	0.52	2.24	0.0022	73.87	1.04	1.16×10^{35}	2.70×10^{37}	2.20×10^{23}
FRB20190410A	0	263.47	-2.37	515.70	6.88	1.59	5.80	0.0728	182.92	0.94	3.89×10^{38}	1.14×10^{41}	1.51×10^{27}
FRB20190412B	0	285.65	19.25	400.20	42.27	0.68	12.80	0.0146	228.59	6.70	2.60×10^{37}	1.40×10^{39}	1.11×10^{24}
FRB20190423B	1	298.58	26.19	537.60	9.83	0.87	7.00	0.0031	159.79	8.54	1.06×10^{35}	6.48×10^{38}	6.48×10^{23}
FRB20190423B	0	298.58	26.19	524.60	9.83	0.87	7.00	0.0031	148.96	8.47	1.04×10^{35}	6.81×10^{38}	6.81×10^{23}
FRB20190429B	0	329.93	3.96	422.40	16.71	0.74	5.00	0.1944	50.64	5.34	3.62×10^{39}	1.32×10^{41}	1.32×10^{27}
FRB20190430A	0	77.70	87.01	433.80	19.66	0.75	7.70	0.2278	214.13	2.75	4.50×10^{39}	5.38×10^{41}	1.27×10^{27}
FRB20190527A	1	12.45	7.99	484.70	57.02	0.47	10.10	0.5367	205.46	1.74	3.87×10^{40}	2.77×10^{42}	4.46×10^{26}
FRB20190527A	0	12.45	7.99	449.10	57.02	0.47	10.10	0.5367	172.11	1.61	3.59×10^{40}	2.56×10^{42}	5.20×10^{26}
FRB20190601C	1	88.52	28.47	517.00	5.90	1.32	5.80	0.1753	223.54	0.58	2.35×10^{39}	6.28×10^{41}	1.02×10^{28}
FRB20190601C	0	88.52	28.47	502.20	5.90	1.32	5.80	0.1753	201.91	0.43	2.28×10^{39}	6.10×10^{41}	1.08×10^{28}
FRB20190617B	0	56.43	1.16	459.30	13.76	0.99	9.20	0.1655	217.37	6.50	2.94×10^{39}	3.69×10^{41}	1.58×10^{27}
FRB20180910A	0	354.83	89.01	417.60	0.98	6.50	5.60	0.126	649.22	0.126	2.51×10^{40}	4.73×10^{43}	3.82×10^{31}
FRB20190210C	0	313.90	89.19	448.50	1.97	2.37	3.60	0.5798	631.93	0.181	1.50×10^{40}	1.56×10^{43}	2.58×10^{30}
FRB20200726D	0	294.75	59.40	684.00	4.92	0.76	3.48	1.3812	541.72	0.407	1.23×10^{41}	6.37×10^{43}	3.17×10^{29}

3. Events detected in the telescope's sidelobes (see Lin et al., 2023b; Lin et al., 2023a) are excluded since understanding the primary beam shape at large zenith angles is limited. This makes it challenging to characterize events.

After considering such cuts, we are left with 459 FRBs. This is a significant reduction from the earlier sample (by about 40%). In this sample, we find 135 FRBs classified as repeaters. Again, for each variable considered, we note the median of each distribution and count the number of repeaters positioned on each side of this central location. The rest-frame frequency width, which can separate 109 of the 135 repeaters (80%), provides the best classification for this reduced sample. Thus, we test whether our methodology provides a separating power beyond this threshold. After applying it similarly to what has been done for the entire sample, we find 15 different combinations of variables for which the separating performance exceeds that provided by the rest-frame frequency width.

We note two aspects of interest: selection effects may force us to work with reduced samples whose properties are likely not well understood yet, due to low sample sizes and changes in total variance. As a result, the PCA analysis and the combination of variables with high separating power differ from those corresponding to the entire sample. However, it can be recognized that peak frequency, rest-frame frequency width, and brightness temperature (variables in #1, see Table 7.1) are among the variables that appear the most in all combinations with ample separating power, whereas, on the other hand, the burst energy is the least implied, offering little classification power.

7.4 Conclusions

To our knowledge, this is the first application of the graph theory MST technique to separate the repetition properties of FRBs. To some extent, this work relates to that of Bhatporia et al. (2023), which uses topological data analysis to cluster the population of FRBs; however, performance metrics are not provided in their case. The performance metrics seen in Table 7.1 indicate that the MST classifiers do a good job of separating repeaters among the FRBs, excelling especially in recall and thus minimizing false negatives, i.e., repeaters seen as non-repeaters. Furthermore, the clarity in its construction can contribute to the scientific understanding of FRBs by highlighting the importance of certain variables over others in determining what best describes a source's repetitive character. Due to these motivations, this approach demonstrates the potential for practical application. Methodologically, we do not define clusters over the MST; instead, we compute the betweenness centrality for each node and identify the one with the highest value (the black node in Figure 7.3). This node acts as a bridge within the MST; we partition the graph into separate branches by removing it. We then use the labels to count the number of repeaters on each branch, which allows us to study the distribution of repeaters and non-repeaters. Since we do not cluster, we do not encounter issues such as chaining or unbalanced clustering, as seen, for example, in single linkage algorithms (see Everitt et al., 2011 for a discussion). Importantly, we do not use the labels during the construction of the MST or the identification of the most central node; our method is unsupervised and requires no training. As a result, the MST works with the entire sample at once, without dividing it into smaller pieces and thereby risking the loss of relevant information.

Table 7.1 shows that while rest-frame frequency width and brightness temperature demonstrated strong discriminatory power per se via median separation of the sample, the frequent occurrence of peak frequency in the selected combinations highlights its potential contribution to the classification process. This information can be valuable for optimizing the current method and informing feature selection strategies in other classification methods, such as machine learning.

The performance of this method is on par with or exceeds that of others, for instance, in Chen et al. (2022), Luo, Zhu-Ge, and Zhang (2023), and Zhu-Ge, Luo, and Zhang (2023), where they utilize samples from CHIME/FRB Collaboration et al. (2021) under a machine learning approach. In more detail, our methodology is applied under the same conditions as in Luo, Zhu-Ge, and Zhang (2023) and Zhu-Ge, Luo, and Zhang (2023), where the sample comprises 594 FRBs, consisting of 500 non-repeaters and 94 repeaters. Under these assumptions, the results are a 97% recall and an F_2 score of 77%, comparable to that obtained in the quoted papers. The recall is higher than those seen in Chen et al. (2022), where the sample is the same, but they identify one repeater less (501 non-repeaters and 93 repeaters). In this quoted work, the variable set adds more properties to deal with a total of 13 (10 observational and three model-dependent parameters, as they define), several of which are set.

Regarding the forecasting power of this methodology, in Table 7.2, we show 33 currently labeled as non-repeaters FRBs that appear in the repeater branch, identified using the 25 selected combinations of variables that show a separating power exceeding a recall threshold of 82%. In the context of the MST strategy, these are considered strong candidates for repeating bursts. Moreover, a few have been identified as well using machine learning methodologies in Luo, Zhu-Ge, and Zhang (2023) and Zhu-Ge, Luo, and Zhang (2023) Also, Table 7.2 shows the FRBs denoted as outliers, known repeaters, but are consistently classified in the non-repeater branch of any of the 25 MSTs considered. Both candidates and outliers provide a handle on the technique's power and caveats, which should be monitored with future samples of FRBs.

Lastly, reducing false positives can be essential to gain robustness, and attempts could be made to improve precision without significantly compromising recall. To refine the classification, a methodology that involves dividing the graph into significant branches (as introduced in Chapter 5) could be helpful.

If given the opportunity, we would like to consider the repetition rate of each of the repeating FRBs, which would be a noticeable enhancement. In particular, we would like to see whether the MST grouping can distinguish a soft transition across the sample from the most commonly repeating FRBs to the non-repeating ones. However, this exercise is hindered by the current availability of data, as the rate should consider the optimal integration time for a particular field of view, which is not provided in the catalog information CHIME/FRB Collaboration et al. (2021) and The Chime/FRB Collaboration et al. (2023). We encourage providing such data in future catalog versions, as it would benefit this and similar analyses, comparing it with any modeling of the repeating mechanism.

Chapter 8

Dynamic Time Warping

8.1 Introduction

The development of robust methods for time series alignment has been a long-standing area of study. The goal has been to develop techniques that are sufficiently stable to be adapted effectively. Early efforts focused on the application of Euclidean distance Faloutsos, Ranganathan, and Manolopoulos (1994), Korn, Jagadish, and Faloutsos (1997), Das et al. (1998), Debregeas and Hebrail (1998), Keogh et al. (2001b), Yi and Faloutsos (2000), Keogh et al. (2001a), Chan, Fu, and Yu (2003), and Agrawal et al. (1995). However, works such as Berndt and Clifford (1994), Kadous (1999), Keogh and Pazzani (2000), Aach and Church (2001), Bar-Joseph et al. (2002), Chu et al. (n.d.), and Vlachos, Kollios, and Gunopulos (2002) have observed crucial weaknesses, including its rigidity in the alignment.

These caveats led to the search for methods that would allow flexibility when producing alignments, which would solve, for example, the issue of similar time series that were out of phase or shortened relative to each other. Thus, based on the concept of dynamic programming, as seen in Bellman and Kalaba (1959) and Bellman (1961) and widely used in speech recognition (Vintsyuk, 1968; Itakura, 1975; Sakoe and Chiba, 1978; Myers, Rabiner, and Rosenberg, 1980; Myers and Rabiner, 1981; Godin and Lockwood, 1989; Rabiner and Juang, 1993), Berndt and Clifford (1994) introduced the idea of Dynamic Time Warping (DTW), see Figure 8.1.

The goal of DTW is to account for differences in phase or time by finding a potentially nonlinear alignment between the elements of the two time series. Intuitively, this can be achieved by either skipping certain elements of a time series or using some of them multiple times. Methodologically, DTW identifies the optimal alignment between the two time series under specific constraints. Based on this alignment, the time series can be warped in a nonlinear manner to match each other (see Keogh and Ratanamahatana, 2005; Müller, 2015 for more details).

The DTW approach is broadly utilized across various fields: it is used for human activity classification (Ten Holt, Reinders, and Hendriks, 2007; Sempena, Maulidevi, and Aryan, 2011), music and motion analysis (Muller, 2007), signature or fingerprint recognition (Munich and Perona, 1999; Kovacs-Vajna, 2000), shape matching (Bartolini, Ciaccia, and Patella, 2005), bioinformatics (Aach and Church, 2001), chemical engineering (Dai and Zhao, 2011), or medicine (Kiani majd et al., 2017; Gogolou et al., 2019). Beyond these applications, it has also been commonly used to analyze economic trends in stock markets and financial time series, particularly for predicting economic behavior (Raihan, 2017; Franses and Wiemann, 2020; Arya, Deepa, and Gandhi, 2021). DTW has also recently started to be used in astrophysics: Examples include the similarity between light curves of Gamma-ray bursts, which offers insight into their central engine activities (Zhang, Zhang, and Castro-Tirado, 2016),
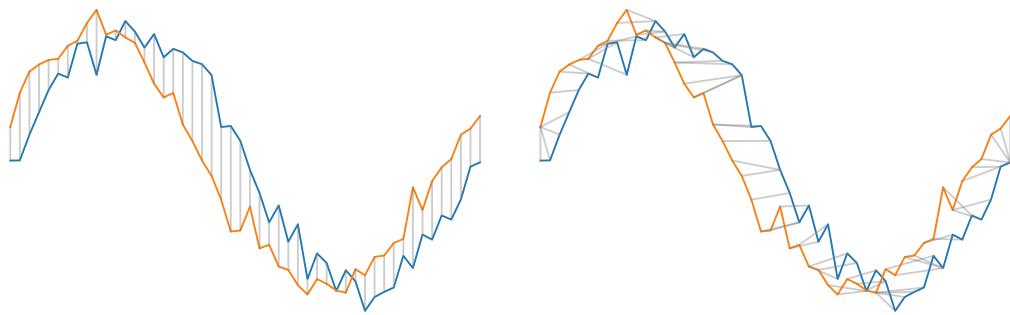
FIGURE 8.1: Left panel: A Euclidean alignment is shown where the elements of one time series (blue) are aligned (gray lines) point by point, vertically, with the elements of the other time series (orange). Right panel: shows DTW alignment on the same time series as the left panel, where points are not aligned vertically, allowing elements of one series to be aligned with different elements of the other series more than once.

the reconstruction of gravitational wave signals from core-collapse supernovae (Suvorova, Powell, and Melatos, 2019), and the assessment of solar wind time series (Samara et al., 2022). Recently, Vohl, van Leeuwen, and Maan (2024) has employed an unsupervised method involving DTW to categorize radio pulsars by profile shape similarity using graph theory.

In this chapter, we introduce the fundamental aspects of DTW, as well as the methodological steps required for its application. Additionally, we will examine the advantages of its implementation as a similarity measure compared to more conventional methods, such as Euclidean distance.

8.2 Computing Dynamic Time Warping

Consider two time series X and Y of length n and m,

$$
\begin{aligned}
X &= [x_1, x_2, \ldots, x_i, \ldots, x_n] \\
Y &= [y_1, y_2, \ldots, y_j, \ldots, y_m] .
\end{aligned}
\tag{8.1}
$$

The indices (i, j) run along the elements of each time series, respectively ($i = 1, \ldots, n$, $j = 1, \ldots, m$). To be considered a warping path, i.e., a set of connected pairs of points (x_i, y_j) covering all the elements of both series in Equation (8.1), we shall require three conditions that ensure efficient, meaningful, and temporally consistent alignments (see, e.g., Berndt and Clifford, 1994):

1. Boundary condition. First, each of the time series's initial and final elements faces each other so that any warping path has constrained endpoints, that is, $i_1 = 1$, $j_1 = 1$, and $i_k = n$, $j_k = m$.

2. Monotonicity condition. We shall also require that no elements of the series that break the temporal order be connected, i.e., no backward steps. Said mathematically, $i_{k-1} \leq i_k$ and $j_{k-1} \leq j_k$.

3. Continuity condition. Finally, we shall also request that the elements of the series being connected do so without gaps or simply that all elements in the series are involved in the path, $i_k - i_{k-1} \leq 1$ and $j_k - j_{k-1} \leq 1$.

Taking these conditions into account, the DTW is computed as (see, e.g., Li, 2014):

$$\text{DTW}(X, Y) = \min_{W(X,Y) \in \mathcal{A}} \left(\sum_{(i,j) \in W} d(x_i, y_j)^2 \right)^{\frac{1}{2}}, \qquad (8.2)$$

where W represents an explored warping path that pertains to the set $\mathcal{A}$ of all paths fulfilling the above-referred conditions (hereafter warping path conditions). The DTW value from Equation (8.2) is the cost of the optimal warping path (W_o), that is, the minimum value obtained after adding the distance d between all pairs (i, j) of W_o.

8.2.1 Dynamic programming approach

From the perspective of DTW calculation, dynamic programming plays a crucial role in the efficient computation of DTW. Following Kruskal and Liberman (1983), Rabiner and Juang (1993), Keogh and Ratanamahatana (2005), and Müller (2015) as a reference, we compute the DTW on the time series seen in Equation (8.1). We construct a cost matrix (E) of size $n \times m$ where each cell (i^{th}, j^{th}) represents an alignment, and contains the distance between x_i and y_j seen as $d(x_i, y_j) = (x_i - y_j)^2$. In this case, W can be viewed as a series of E cells that determine the alignment between X and Y. In such a way, we can define:

$$W = w_1, w_2, \ldots, w_k, \ldots, w_\text{K}, \qquad (8.3)$$

where $w_k = (i, j)_k$ represents the kth of W and contains the values seen in (i^{th}, j^{th}) of E, i.e., $d(x_i, y_j)$ defined above. Equation (8.3) satisfies $\max(m, n) \leq \text{K} < m + n - 1$. As discussed, there exists an $\mathcal{A}$ composed of an exponential number of W satisfying the warping path conditions, so analogously to Equation (8.2):

$$\text{DTW}(X, Y) = \min_{W \in \mathcal{A}} \left(\sqrt{\sum_{w \in W} w} \right) = \left(\sqrt{\sum_{k=1}^{K} w_k} \right)_{W_\text{o}}. \qquad (8.4)$$

In Equation (8.4), the value of DTW is given by W_o whom is defined by Equation (8.3). However, the application of Equation (8.4) on every W is not feasible due to the enormous size of $\mathcal{A}$. For this purpose, a recurrence relation is implemented, which is at the core of dynamic programming, to find W_o. The idea behind this process is to solve the problem by combining the solutions from the simplified subproblems, which were derived from the original problem. In other words, an expression that defines a sequence based on conditions that gives a term as a function of the previous terms. From this, we work out what is known as the accumulative cost matrix (D) computed by:

$$D(i, j) = E(i, j) + \min \left(D(i-1, j-1), D(i-1, j), D(i, j-1) \right). \qquad (8.5)$$

As derives from Equation (8.5), D is a matrix ($n \times m$) where $D(1, 1) = E(1, 1)$ by definition. Therefore, W, according to the warping path conditions, is constrained in D as follows:
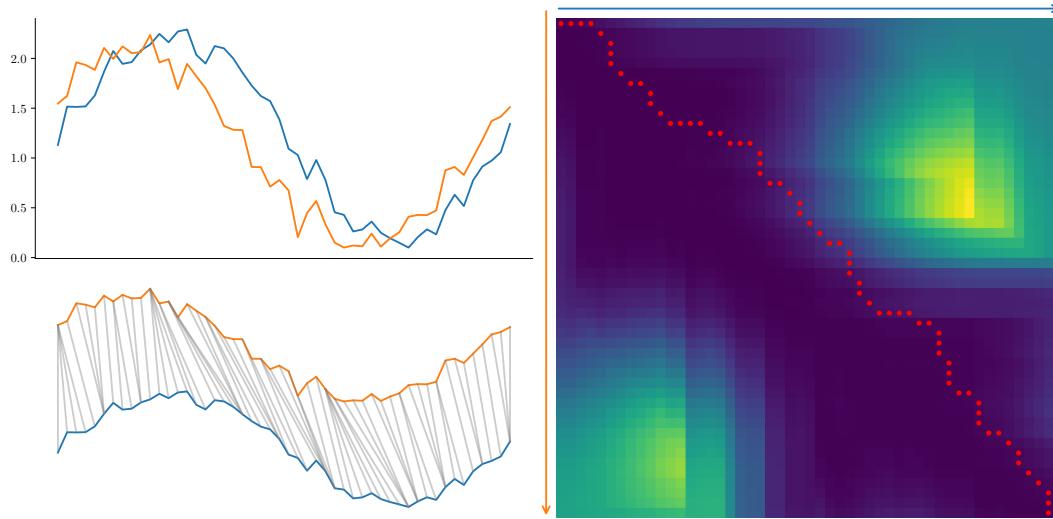
FIGURE 8.2: First column: two time series, X in orange and Y in blue, are shown unaligned (top panel), and aligned from W_o by DTW, where each gray line is a $w_k \in W_o$ (bottom panel). Second column: Shown is D calculated over the time series in the first column, with orange and blue arrows indicating the temporal order, respectively. The path of W_o is shown in red. The values tending towards yellow represent the cells containing higher cumulative costs, while the darker cells represent those with lower costs.

1. The W has to start and end on opposite diagonals of D: $w_1 = (1,1)$ and $w_k = (m,n)$.

2. The W has to traverse D evenly spaced in terms of time with no backward jump: let $w_k = (i,j)$, then $w_{k-1} = (i',j')$ with $i - i' \geq 0$ and $j - j' \geq 0$.

3. The W has restricted advancement in D to adjacent cells: let $w_k = (i,j)$, then $w_{k-1} = (i',j')$ with $i - i' \leq 1$ and $j - j' \leq 1$.

Thus, the value of DTW seen in Equation (8.4) is given by the W_o found in D whose cost is determined by E (see Equation(8.3)), see Figure 8.2 to get an idea of this computation.

An example of the DTW application, along with its computation step by step, can be seen in Appendix G.1.

8.2.2   Euclidean distance in the Dynamic Timer Warping context

The Euclidean distance (ED) can be viewed as a restricted case of DTW. In such a way, let W according to Equation (8.3), w_k takes $k = i = j =$, defining what we denote as Euclidean path (W_E). Therefore, W_E represents the main diagonal of D, so it can only be computed when $n = m$ in Equation (8.1). Furthermore, W_E, since it satisfies the warping path conditions, is always explored in Equation (8.4). This leads to the conclusion that for any X and Y with $n = m$, we have $\text{DTW}(X,Y) \leq \text{ED}(X,Y)$. Figure 8.3 highlights how W_o deviates from W_E, allowing flexibility in aligning time series looking for the minimum cost. An example of W_E calculated step by step in comparison with W_o can be seen in Appendix G.2.
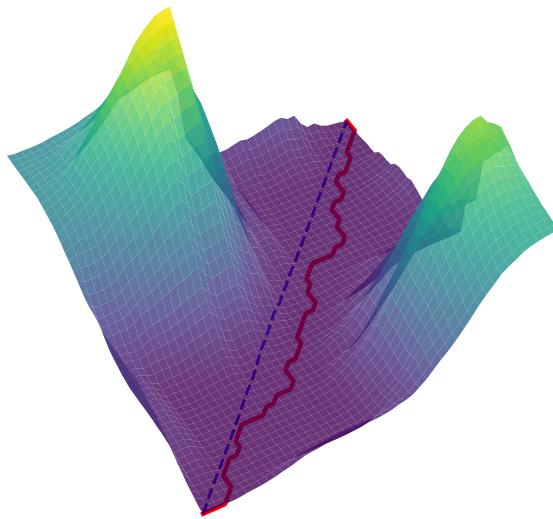
FIGURE 8.3: The D depicted in Figure 8.2 is represented as a three-dimensional surface, where the vertical z-axis indicates the cumulative cost of the alignment between the elements of X and Y. The W_o is shown in red and the W_E in blue.

8.2.3   Global constraints

In addition to the warping path conditions, other global constraints limit the number of W seen in A. This not only helps to speed up the DTW calculation but also provides more global control over the path of W by avoiding what is known as 'pathological' warpings, where few elements of X are aligned with many elements of Y. These types of constraints determine how far W can be moved away from the diagonal of D by determining what is known as a warping window or global constraint region. The two best-known global constraints are the Sakoe-Chiba band (Sakoe and Chiba, 1978) and the Itakura parallelogram (Itakura, 1975). Figure 8.4 shows how the scope of W is restricted to the highlighted area. While the Sakoe-Chiba band traverses the diagonal of D with a width fixed by the global constraint area, the Itakura parallelogram restricts the slope of W (see Geler et al., 2019 for a discussion about a comparison of their performance).

8.3   Properties of DTW

DTW can be defined as a similarity distance; however, it is not strictly a metric (Vidal Ruiz, Casacuberta Nolla, and Rulot Segovia, 1985 for a discussion). Let X e Y as defined in Equation (8.1), we can state the following properties:

- Self-similarity: $\text{DTW}(X, X) = 0$.

- Symmetry: $\text{DTW}(X, Y) = \text{DTW}(Y, X)$.

- Positivity: If $X \neq Y$, then $\text{DTW}(X, Y) \geq 0$.
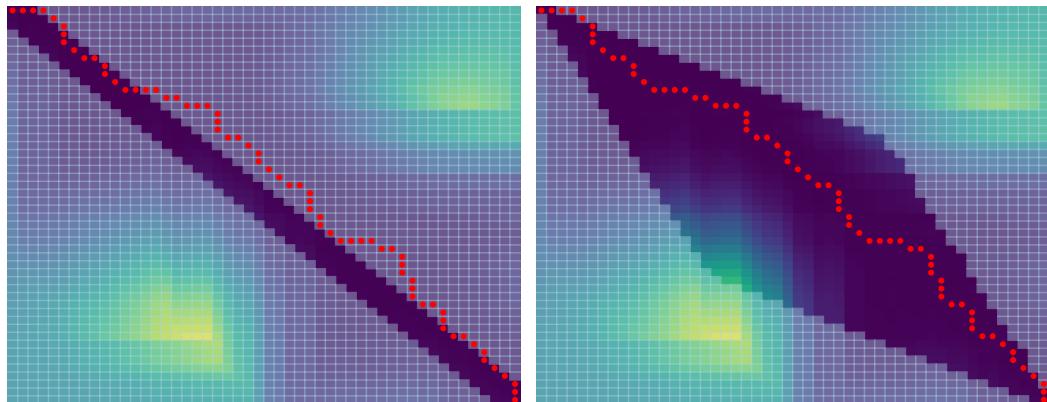
FIGURE 8.4: The Sakue-Chiba band and the Itakura parallelogram constraints on the D are shown in the right and left panels, respectively, highlighted as the only possible region for computing DTW. The W_o is marked in red when no constraint is applied, as seen in Figure 8.2 where the D shown was computed.

However, to be considered a metric, DTW should also meet the triangular inequality: $DTW(X, Y) \leq DTW(X, Z) + DTW(Y, Z)$ with Z as a time series with non restriction length. This property is not always satisfied by DTW and has served as a starting point, among others, for the search of new and improved forms of its implementation; see, e.g., Keogh and Ratanamahatana (2005), Athitsos et al. (2008), Ding et al. (2008), and Sart et al. (2010). Furthermore, it must be noted that due to the dynamic alignment performed by DTW, the following property may not hold: if $DTW(X, Y) = 0$, then $X = Y$. However, considering this situation, we enter the realm of the concept of similarity since X and Y may differ due to $n \neq m$. For instance, a DTW ~ 0 would allow us to establish a strong similarity relation, which is highly useful in fields where time series represent physical phenomena observed with temporal differences but that somehow respond to similar phenomenological processes.

8.4   Conclusions

The advent of DTW as a distance measure between time series stems from the weaknesses imposed by the usual distance measures, such as ED. Methods such as ED limit the cases where they can be applied, as they require fundamental conditions, such as $n = m$ in Equation (8.1).

Moreover, they show less efficiency when, even under the same length condition, the aim is to find the difference between morphologies due to their static alignments, point by point. This causes time series that are similar in terms of morphological characteristics, such as the number of peaks, the value of higher intensity, or the width of these, but out of phase, to show high values.

However, a dynamic alignment such as the one carried out by DTW allows us to solve these problems.

In Chapter 9, we use these premises to conduct a morphological comparison of the gamma-ray light curves of a sample of approximately 300 pulsars collected by the *Fermi*-LAT Third Pulsar Catalog (3PC, Smith et al., 2023) The high variability of these detections, as well as the different observation times that produce light curves of varying sizes, makes DTW an optimal approach for establishing a similarity distance

between them. In this way, we have been able to establish a ranking of similarity between pulsars through their light curves, leading to a clustering of them and to suggestions about their emission processes and the factors influencing them.

Chapter 9

Quantitative exploration of the similarity of gamma-ray pulsar light curves

9.1   Introduction

Comparing light curves of two or more gamma-ray pulsars has been chiefly an artisanal business. One would usually compare how similar two light curves are to the eye of the beholder, mentally disregarding the different flux levels. However, this is not always possible, as the light curve count rates can vary from a few to more than a thousand for the same phase bin. To improve this, consider global light curve indicators like the number of peaks, their separation, width, and height. This is not as obvious as it seems for several reasons. First, a solid definition of a light curve peak is lacking. Any characterization of light curves would inherently depend on this prior definition. Secondly, light curves with two peaks of different heights can have equal relative heights, or light curves with a distinct peak separation can be identical in other respects. It is clear, then, that the former indicators do not exhaust the need for a similarity assessment and do not provide a quantitative measurement for it.

Recent theoretical studies of light curve morphology, as seen in Íñiguez-Pascual, Torres, and Viganò (2024), Cerutti, Figueiredo, and Dubus (2024), and Íñiguez-Pascual, Torres, and Viganò (2025), emphasize the importance of such similarity measurements. Results from theoretical models suggest that light curve properties are primarily determined by geometry and do not correlate with other physical parameters of the pulsars or their magnetospheres. Íñiguez-Pascual, Torres, and Viganò (2024) has, in particular, demonstrated that pulsars with different timing and spectral parameters have similar radiation skymaps. This implies a likely similarity among their magnetospheres, with the geometry widely dominating over the spectra in shaping the light curves. If so, considering a sample of light curves generated for the same pulsar but for different geometries and observers should relate to exploring the population of all observed pulsars. Thus, it is essential to advance our understanding of pulsar magnetospheres to know to what extent pulsars with different properties have similar light curves.

This point is further supported by Íñiguez-Pascual, Torres, and Viganò (2025), who, considering the same population and the same synchro-curvature model above, presented a simultaneous fit of spectra and light curves. The results obtained show that light curve morphology is largely insensitive to timing and spectral parameters, reinforcing the dominant role of geometry in shaping pulsar emission.

Section 1.3 briefly reviewed the development of gamma-ray emission models to give some theoretical context to the physical origin of the light curves analyzed in this chapter.

Dynamic time warping (DTW) is a well-established method for comparing time series and identifying similar behaviors, even when they exhibit variations in speed or other temporal differences. The fundamental concepts and the DTW implementation can be found in Chapter 8. In this chapter, we present the DTW methodology proposed in García and Torres (2025) for the first time to quantitatively assess the degree of similarity between two gamma-ray light curves and apply it to the whole set of light curves in the *Fermi*-LAT Third Pulsar Catalog (3PC, Smith et al., 2023).

9.2 *Fermi*-LAT 3PC Light curves

The set of time series considered for this work comprises 294 light curves associated with the pulsars reported in the 3PC. In this sample, there are 143 millisecond pulsars (MSPs) with periods shorter than 10 ms ($P < 10$ ms). We consider the pulsars' weighted counts above $E \geq 100$ MeV, subtracting the background level for each pulsar as reported in the 3PC. When the resulting value is lower than zero, we set it to zero (no pulsar contribution above the background). As we are interested in comparing the morphology of the light curves, they should have the same scale; thus, we consider the *Min-Max scaler* after background subtraction, i.e., all light curves are transformed to lie between 0 and 1 via.

$$X^\dagger = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}. \tag{9.1}$$

Gamma-ray light curves have different sizes, i.e., are discretized in different bins (N). The 3PC contains 48 light curves described with 25, 116 with 50, 98 with 100, 27 with 200, 4 with 400, and 1 with 800 bins. These bins correspond to the rotational phase of a pulsar ($\phi \in (0, 1)$).

9.3 Results

Considering the definitions in Section 8.2, we will apply DTW to the 3PC light curves. The arrangement of its weighted counts in representation of its flux levels can be seen according to Equation (8.1). Then, Equation (9.1) is applied after subtracting the corresponding background. The result of this light curve processing enables us to effectively apply Equation (8.2) and obtain a DTW value that shows the similarity between the given light curves. However, as ϕ=0 is arbitrary, we shall consider all possible phase shifts, getting a DTW for each using Equation (8.2). The minimum value of the DTW measures how similar two light curves are, given the most favorable phase shift between them. Finally, in addition to comparing light curves using DTW, we shall also directly compute the Euclidean distance (ED) when the lengths in the two-time series are the same or have been correspondingly rebinned to the same. The Appendix G.1 provides a worked-out example that will help generate intuition regarding these applications.

9.3.1 Comparing light curves with the same number of bins via ED

Figure 9.1 shows the three most similar pairs of light curves quantified by ED. Figure 9.1 also shows how these light curves look as read from the 3PC before processing, i.e., before background subtraction, before the Equation (9.1) is used to compare the shape disregarding different flux levels, and before rotations are performed to produce the best alignment leading to the minimum value of ED. The morphological similarity of these pairs of light curves is striking. In particular, MSPs and young pulsars share detailed light curve morphology. The pulsar spin period (P) and the spin period derivative ($\dot{P}$), and the quantities derived from them using the rotating dipole model (see Section 1.2.1), such as the surface magnetic at the surface (B_s), the magnetic field at the light cylinder (B_{lc}), and the spin-down energy loss rate ($\dot{E}_{sd}$), can be very different from one pulsar to another, even when their light curves are very similar. Differences up to a factor of 70 in $\dot{E}_{sd}$, of 4 in B_s, and of 30 in B_{lc} can be found even between young pulsar pairs sharing light curve morphology. The differences in parameters, when the pair mixes MSPs and young pulsars, can be much larger, e.g., reaching up to a factor of more than 3×10^4 in B_s. The gamma-ray emission properties, as reported in the 3PC, e.g., phase-averaged integral energy flux in the 0.1 to 300 GeV energy band, G_{100} ($F_{E>100\text{MeV}}$), and the energy of the maximum peak of the SED (E_p) also span an extensive range, representing very different gamma-ray spectra. Light curve similarity is also immune to such differences.

9.3.2 Comparing light curves of the same size via DTW

We also employ the DTW methodology to determine whether it offers any advantages over ED. Results for similar, equal-size light curves under DTW are shown in Figure 9.2. We recall that the cumulative cost matrix in DTW (D, see details in Equation (8.5)) includes the ED comparison. Thus, all paths considered via ED are also considered when using DTW. In some cases, the DTW ranking of the three most similar pairs for each subset of equal-N light curves is the same as those in ED. This is not the case; see Section 8.2.2 for more details.

The similarity of the pulsar light curves showcased by DTW goes beyond simple visual inspection (which is what ED matches more closely). It does not focus on, e.g., having similar peaks separated by the same phase interval, which our eye would do directly if we could rotate all light curves to consider all possible shifts. Instead, it focuses on the global morphology. Consider, for example, the case of the group of $N = 200$ bins. The DTW ranking selects the most similar pair, J0614-3329 and J2032+4127, which is not even among the first three pairs ranked by ED. The pulsar pairs selected by ED are visually appealing to the eye because the ϕ separation is the same, despite the height of each peak being different. Under DTW, however, these do not rank as highly, as this method prioritizes overall morphological similarity rather than strict phase alignment. One example is the top pair ranked by DTW, which has a slightly different phase separation between peaks (which for DTW is not a problem but would increase ED significantly) and an almost exact match of peak heights. When comparing these light curves, DTW emphasizes peak values more, whereas ED focuses more on phasing. As before, they involve pulsars of all kinds with significantly different physical and spectral parameters.

9.3.3 Caveats of ED and rebinning the whole sample

ED has a restricted application to light curves of equal size. This restriction would exclude the best-described light curves from most similarity comparisons unless
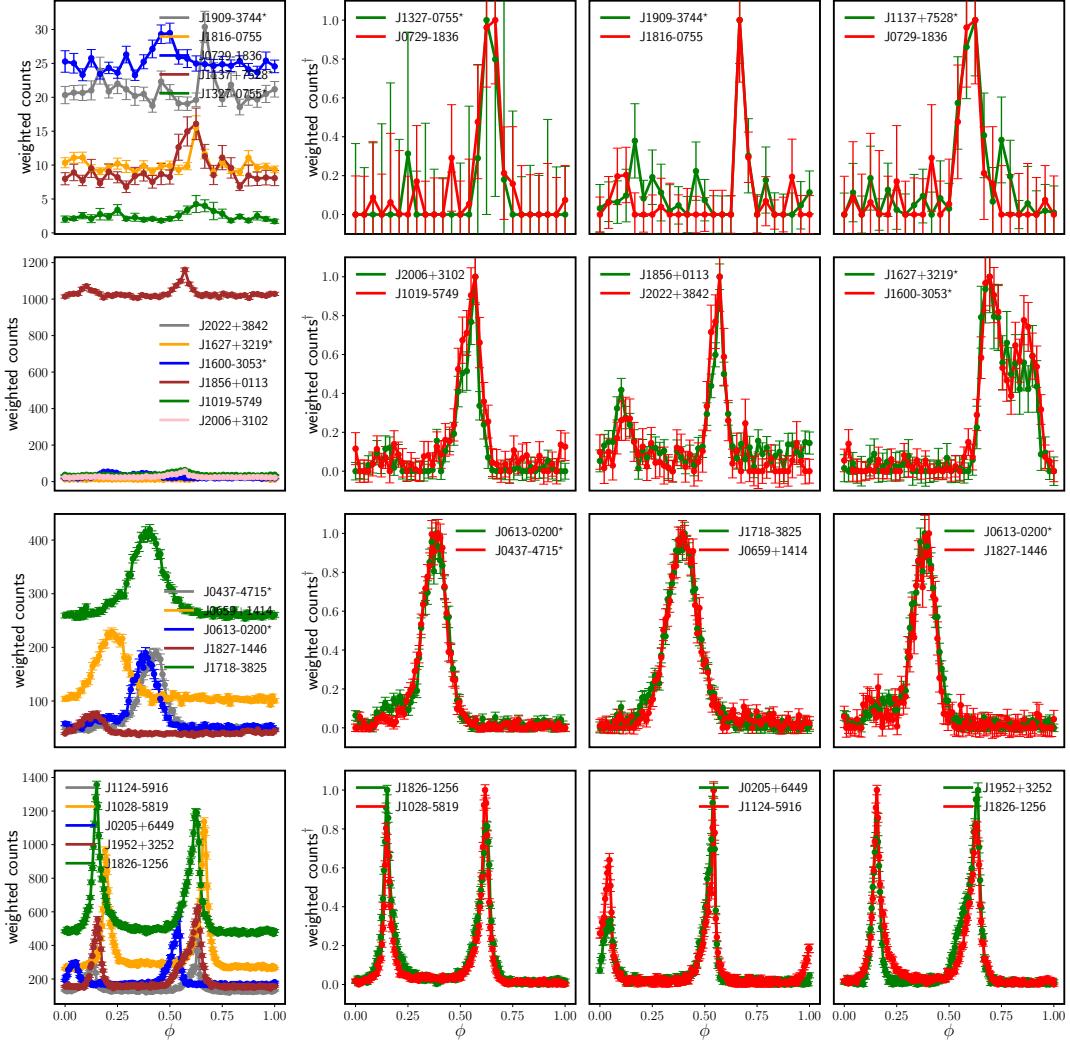
FIGURE 9.1: Each row shows the three most similar pairs of light curves according to ED for light curves having 25, 50, 100, and 200 bins. The plot in the first column of each row shows the 3PC data directly before processing. MSPs are denoted with a star on the pulsar name.
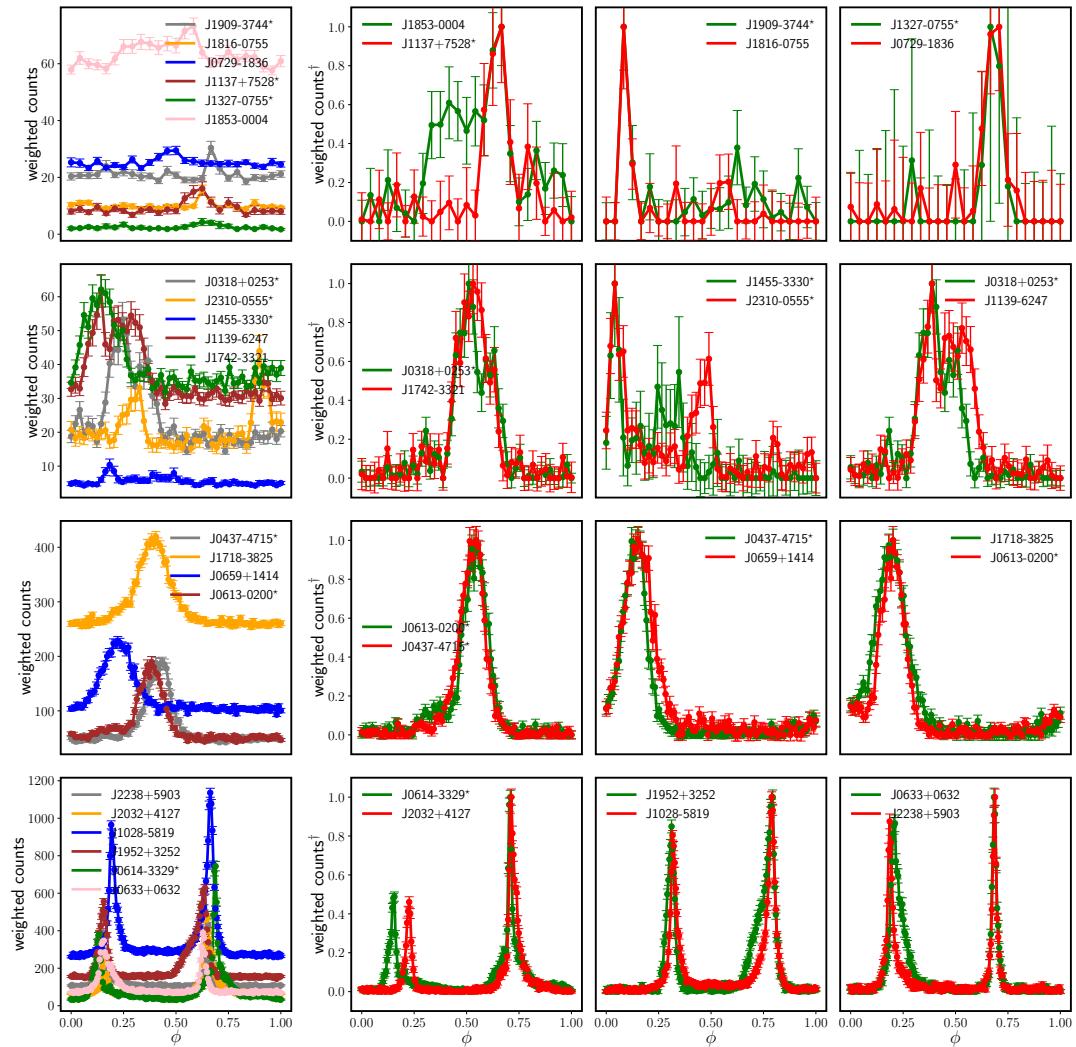
FIGURE 9.2: The three most similar pairs of light curves according to DTW for light curves having 25, 50, 100, and 200 bins. The plot in the first column of each row shows the 3PC data directly before processing. MSPs are denoted with a star on the pulsar name.
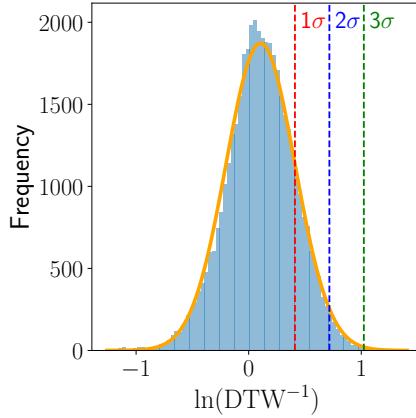
FIGURE 9.3: Distribution of the natural logarithm of DTW^{-1} values, $\ln(DTW^{-1})$. The orange line shows the normal distribution identified as the best fit for the data. The dashed vertical lines denoted with 1σ (red), 2σ (blue), and 3σ (green) represent the intervals of the distribution according to the empirical rule.

they were rebinned to a lower resolution, resulting in a loss of details. Rebinning, thus, (unless rebinning to the worst-resolved light curve, with $N = 25$) will produce the loss of significant portions of the sample. DTW remains stable despite rebinning. For instance, over the sample of light curves rebinned to 100 bins (formed by 130 light curves, as those with $N < 100$ bins are ignored), DTW rankings of the five most similar light curves from any given one retain three pulsars ranked also when no rebinning is applied in 85% of the cases. If we look at the two best matches, at least one is reproduced by both situations (with and without rebinning) in 88% of the cases. In addition to reducing the sample, we have demonstrated that ED is capable of producing large measures (signaling dissimilarity) even in cases where the overall structure remains the same. This will occur mainly when two or more peaks are not precisely aligned, there is a slight difference in phase separation, or the peak widths differ slightly. When rebinning, as the resolution at which the comparison will be made will, at most, be that of the lesser-sampled light curve, these issues will appear more often. DTW solves both problems simultaneously.

9.3.4 Using DTW to compare all light curves

Since the DTW methodology does not require reapplication over time series of equal length, we now consider all light curves without rebinning. We obtain 294 ordered rankings, one for each pulsar; we have $(294 \times 293)/2 = 43071$ DTW values. The log-normal distribution is identified as the best fit for both DTW and DTW^{-1} values (we take the inverse values to facilitate looking at the most similar light curve pairs in the far end tail of the distribution). Thus, considering the distribution of $\ln(DTW^{-1})$ values, as shown in Figure 9.3, allows us to define confidence intervals in the usual way, as a Gaussian is the best fit for this distribution. We use the mean μ and the standard deviation (σ) of the distribution of $\ln(DTW^{-1})$, equal to 0.105 and 0.306, respectively, to obtain a value of 3σ deviation equal to 2.78. That is, a value of $DTW^{-1} = 2.78$ represents two pairs of light curves whose similarity measured by DTW is 3σ beyond the mean. Beyond this value, 70 pairs of light curve comparisons are found, involving 63 pulsars (13 MSPs).

We can use the DTW distribution to partition the light curve sample into clusters. If we request that clusters contain at least four pulsars, we find six clusters, and members are related by at least one similarity at the 3σ level. We can depict these connections in a graph. Figure 9.4 shows these six clusters, where 41 pulsars are displayed as the graph nodes. The properties of these pulsars are shown in Table 9.1. The different similarity levels according to the distribution of DTW values are shown by the color-coded edges in the corresponding complete cluster graphs (where all nodes are connected). Figure 9.4 also shows the light curves of the members of each cluster. The complete graphs show that the vast majority of the light curves within a given cluster also belong in the tail of the distribution, with similarities beyond > 1 and $> 2\sigma$.

9.4 Conclusions

This chapter provides a quantitative assessment of gamma-ray light curve similarity that extends beyond estimating global properties, such as the number of peaks, peak width, or peak separation. Studying via DTW the similarity of all light curves (more than 43000 pairs, each with their corresponding rotations arising by considering that phase zero is arbitrary), a well-behaved distribution is found, from which similarity can be assessed. Referring to that distribution of DTW values, we could define confidence intervals and, for the first time, determine whether a pair of pulsars has light curve similarity above or below a given threshold. This methodology lacks the caveats affecting the point-by-point comparison provided by the Euclidean distance between two time series. The latter can only be applied to light curves described with equal resolution. Rebinning not only loses details of the morphology in cases that are better sampled but also requires cutting the sample unless light curves with the worst resolution drive the morphological comparison. Even if we rebin and cut the sample accordingly, light curves that, for instance, have two very similar peaks with similar heights and widths but with slightly different peak separations would never be considered identical by the Euclidean distance estimator. Instead, DTW would mark those as such, since it examines the global light curves. Using this methodology, we could cluster the light curve sample quantitatively.

The variety of the physical and spectral properties of the light curves that are morphologically similar is large. Having the whole distribution of the DTW, we have explored whether the similarity of any two light curves, as measured with the DTW technique, is correlated with nearness in any other quantity of interest. In particular, we tested for correlations between the DTW values for any two pulsars of the sample and the Euclidean distance under each one of the pulsar properties considered above. For instance, we considered the $(DTW)^{-1}$ measure between two given pulsars versus the correctly normalized difference of their periods or their magnetic field at the light cylinder. No correlation arises for any of the physical or spectral parameters. Similar light curves do not imply that pulsars are also identical in terms of their timing properties and can mix MSPs with young ones. Even for pulsars of the same kind, similar morphologies do not necessarily imply identical period derivatives, magnetic fields, or spin-down energies; they can all vary significantly. This suggests that the exact mechanism in all pulsars radiates gamma-ray emission and that geometry dominates the morphology of the light curve.
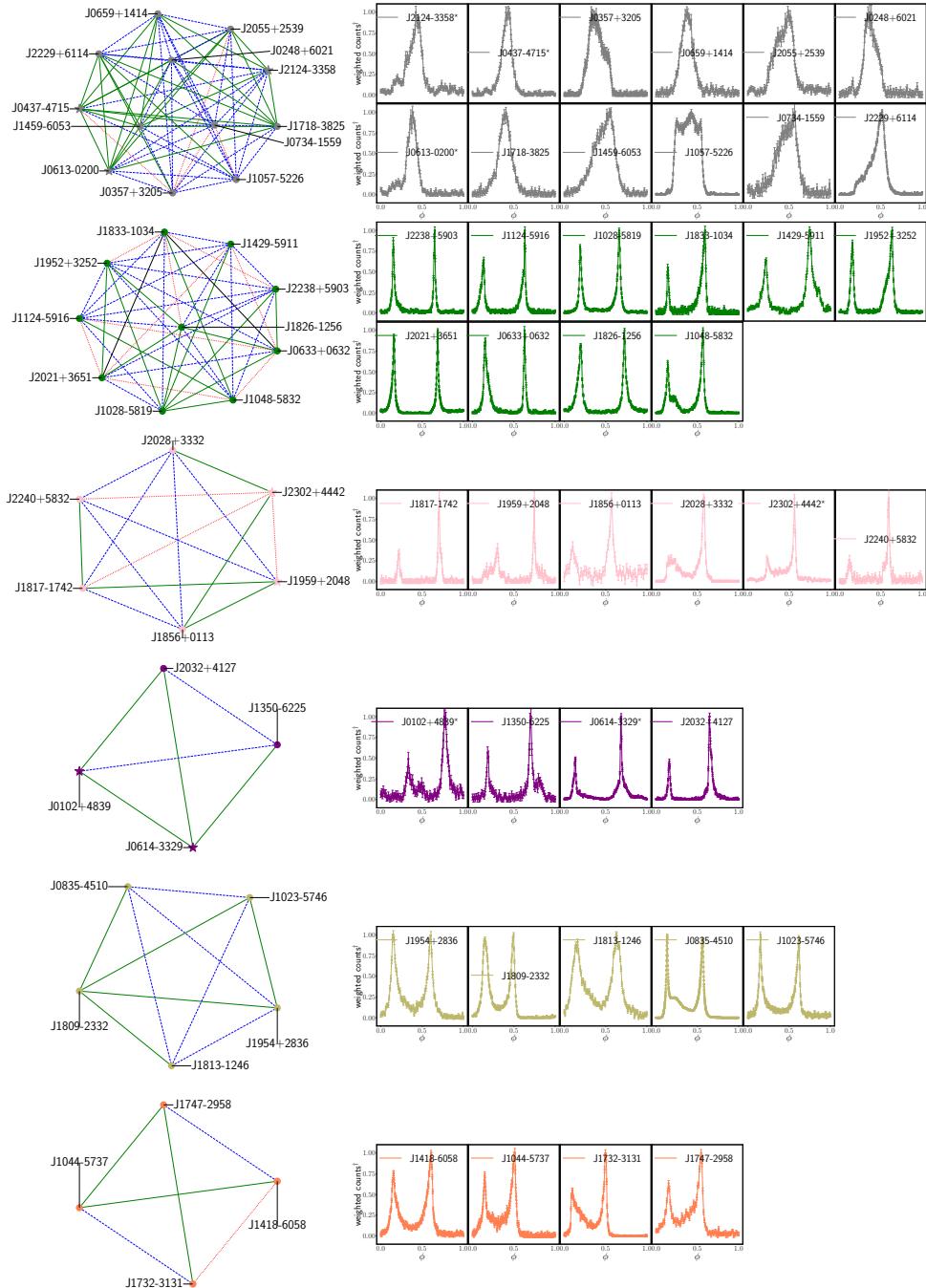
FIGURE 9.4: Clusters with four or more pulsars and all members related to another by at least one similarity at a 3σ significance level. Each node represents a pulsar, and each colored edge connecting the nodes represents the similarity significance of the corresponding light curves (black for $< 1\sigma$, red for cases between 1 and 2σ, blue for cases between 2 and 3σ, and green for $> 3\sigma$). The light curves associated with each cluster are shown next. MSPs are denoted with a star-shaped node on the pulsar name.

TABLE 9.1: The columns show the pulsar timing, derived properties using the rotating dipole model, and gamma-ray properties of the clusters, including color coding, shown in Figure 9.4. MSPs are denoted with a star on the pulsar name, and their corresponding $\dot{P}$ has an additional factor 10^{-6} multiplying the column label. J1856+0113 has TS<25, so no spectral fit results are provided. See the 3PC for more details.

PSRJ	P [s]	$\dot{P}$ [s/s] $[\times 10^{-14}]$	$\dot{E}_{sd}$ [erg/s] $[\times 10^{35}]$	B_s [G] $[\times 10^{10}]$	B_{lc} [G] $[\times 10^{4}]$	$F_{E>100MeV}$ [erg/cm^2/s] $[\times 10^{-11}]$	E_p [GeV]
J0248+6021	0.217	5.52	2.13	350.35	0.32	2.94	0.774
J0357+3205	0.444	1.31	0.06	244.02	0.03	6.01	0.869
J0437−4715*	0.005	5.72	0.12	0.06	2.80	1.76	0.700
J0613−0200*	0.003	0.96	0.13	0.02	5.57	3.82	1.156
J0659+1414	0.384	5.50	0.38	465.38	0.08	2.65	0.156
J0734−1559	0.155	1.25	1.32	140.95	0.35	4.57	0.623
J1057−5226	0.197	0.58	0.30	108.54	0.13	29.59	1.213
J1459−6053	0.103	2.53	9.08	163.32	1.37	12.27	0.372
J1718−3825	0.074	1.32	12.49	100.35	2.22	10.37	0.564
J2055+2539	0.319	0.41	0.05	115.85	0.03	5.32	1.215
J2124−3358*	0.004	2.06	0.07	0.03	2.48	3.88	1.949
J2229+6114	0.051	7.53	215.58	199.49	13.33	24.01	0.740
J0633+0632	0.297	7.96	1.19	492.24	0.17	9.56	1.427
J1028−5819	0.091	1.42	7.34	115.28	1.39	24.51	0.888
J1048−5832	0.124	9.55	19.92	347.86	1.69	18.52	0.968
J1124−5916	0.136	75.15	119.15	1021.18	3.78	6.11	0.362
J1429−5911	0.116	2.39	6.06	168.30	1.00	11.28	0.540
J1826−1256	0.110	12.12	35.72	369.86	2.54	41.35	0.870
J1833−1034	0.062	20.20	336.24	357.81	13.89	8.91	0.305
J1952+3252	0.040	0.58	37.23	48.56	7.24	14.77	0.993
J2021+3651	0.104	9.48	33.53	317.40	2.62	49.21	1.064
J2238+5903	0.163	9.69	8.87	401.71	0.86	6.64	0.673
J1817−1742	0.150	2.06	2.42	177.62	0.49	2.85	1.123
J1856+0113	0.267	20.59	4.25	750.91	0.36	-	-
J1959+2048*	0.002	1.68	1.60	0.02	36.73	1.57	1.038
J2028+3332	0.177	0.49	0.35	93.74	0.16	5.71	1.511
J2240+5832	0.140	1.53	2.20	147.82	0.50	0.98	1.138
J2302+4442*	0.005	1.33	0.04	0.03	1.75	3.90	2.355
J0102+4839*	0.003	1.17	0.17	0.02	6.68	1.50	1.700
J0614−3329*	0.003	1.78	0.22	0.02	7.06	11.46	2.835
J1350−6225	0.138	0.89	1.33	112.10	0.39	3.60	1.951
J2032+4127	0.143	1.16	1.56	130.56	0.41	14.24	2.606
J0835−4510	0.089	12.23	67.64	334.51	4.31	929.32	1.267
J1023−5746	0.111	37.99	108.21	658.53	4.37	14.60	0.621
J1809−2332	0.147	3.44	4.29	227.33	0.66	42.48	1.285
J1813−1246	0.048	1.76	62.39	92.96	7.70	24.78	0.483
J1954+2836	0.093	2.12	10.48	141.71	1.64	10.70	1.172
J1044−5737	0.139	5.46	8.02	278.71	0.95	11.40	0.650
J1418−6058	0.111	17.10	49.92	440.00	3.00	29.70	1.001
J1732−3131	0.197	2.80	1.46	237.56	0.29	17.93	2.089
J1747−2958	0.099	6.11	24.98	248.61	2.37	15.91	0.526

Chapter 10

General Conclusions and Future Work

In this Chapter, I summarize the main points of the thesis and discuss some avenues for future exploration.

10.1   General Conclusions

This thesis has concentrated on advanced mathematical techniques, with a special focus on graph theory, PCA, and DTW, to analyze pulsar and FRB populations.

The population of pulsars has been extensively described in Chapter 1, from their origin to the physical properties that best characterize their evolution and emission, highlighting the caveats they present. In this Chapter, we have also contextualized FRBs, as well as their possible origin and their more than likely connection to pulsars, specifically magnetars.

In Chapters 2 and 3, we have presented the mathematical concepts that comprise the frameworks of this thesis, along with Chapter 8, where we do the same for DTW.

The innovative classification, grouping, and prediction methodologies developed are supported in these more conceptual chapters.

The MST was brought in Chapter 4 as a tool to visualize the pulsar population collected in the ATNF, demonstrating that the traditional $P\dot{P}$ diagram does not fully capture its variance. Using PCA, we showed that physical properties derived must be considered to define as much information as possible. We have shown that the pulsar MST (or simply, the pulsar tree) beyond being a mathematical approach, encodes physical meaning within its structure. This work led to the development of the Pulsar Tree App, an interactive platform for exploring pulsar populations from a graph theory perspective.

Building on this approach, Chapter 5 shows the application of the MST to automatically identify pulsar subpopulations, utilizing betweenness centrality estimators and non-parametric statistical tests, as the KS test, to define significant branches. This clustering method provided information about the historical evolution of pulsars. The latter is quantified through the implementation of *growth* and *togetherness*, two measures designed to capture the stability of significant branches over time. This algorithm also represents a new member of the family of clustering methods, which has the advantage of relying on the underlying topology of the population, making it a rigorous approach.

Furthering the MST application, Chapter 6 focused on MSPs in binaries seen in the ATNF, incorporating binary parameters alongside physical properties. Based on their MST proximity to confirmed systems, we identified candidates for BWs, RBs, and tMSPs. This approach allowed us to estimate missing parameters and the

possible nature of specific pulsars. Using the clustering method introduced in Chapter 5, we found that the clusters organize according to notably distinct physical and binary properties. Some align with known classes, such as spider pulsars, while others suggest new avenues for exploration. These advancements were integrated into the Pulsar Tree App, expanding its capabilities for MSPs.

In a separate application, Chapter 7 employed MST as an unsupervised classification method to distinguish between repeating and non-repeating FRBs observed by CHIME. We developed a graph-based methodology that outperforms traditional machine learning classifiers by computing multiple MSTs based on observational and derived properties. The process identified new repeater candidates and highlighted the most discriminative features of FRB separation. This suggests that the MST could serve as a valuable preprocessing tool for complex, computationally intensive machine learning methods that rely on robust, well-validated inputs.

Shifting our focus to time-series analysis, we introduced the DTW method for gamma-ray pulsar light curves from the *Fermi*-LAT 3PC in Chapter 9. The mathematical context and advantages of this approach are discussed in Chapter 8. This work represents the first quantitative clustering of gamma-ray pulsars based on light curve morphology. It defines similarity thresholds and demonstrates that pulsars with nearly identical light curves can exhibit distinct timing properties. These findings suggest that geometry dominates gamma-ray emission rather than intrinsic properties.

The results presented throughout this thesis reveal physical insights that support the establishment of these techniques as valid approaches for studying high-energy astrophysics.

10.2   Future Work

The reach of the methodologies developed in this thesis for graph theory and DTW applications in high-energy astrophysics is far from over. While the MST has provided insights into pulsar populations and FRBs, further analysis could reveal evolutionary trends, transitional sources, new classes, and hidden connections. One example is how specific parameters evolve along branch paths in the MST, which could provide insight into pulsar evolution and classification. Another example is an ongoing project we are currently pursuing, which utilizes DTW to quantify the similarity between confirmed and candidate tMSPs through their light curves. This approach could be extended to different time series, such as Gamma Ray Bursts (GRBs) or magnetar flares, which exhibit various light curve structures, often with complex variability. There are also opportunities to combine graph theory and DTW. The MST could be used to cluster pulsars based on the DTW distances of their gamma-ray light curves and explore the underlying behaviour of the clusters. A natural extension of this thesis is the application of Graph Neural Networks, which enable the direct encoding and learning of relational information between astrophysical sources. Having developed algorithms based on graph structures, the context is well-prepared to extend these concepts into dynamic learning frameworks.

Appendix A

Graph centrality

A.1 Betweenness centrality estimator: Practical example

Here, we provide a simple example of the betweenness centrality estimator. To calculate the coefficient C_B by applying Equation (2.3) on a graph, we recall to take into account the following considerations:

- Select all pairs of nodes (v_s, v_t). One can exclude the adjacent ones, as there cannot be another node between them.

- Select all shortest paths within the graph joining (v_s, v_t), eliminating duplicities (i.e., $\sigma_{s,t} = \sigma_{t,s}$). If the graph is unweighted, shortest paths are defined by the minimum number of edges that must be traversed. If weighted, shortest paths minimize the total edge weight. If the shortest path can be reached by different paths due to cycles, $\sigma_{s,t}$ takes a value equal to the number of such paths.

- Each node, v_j, appearing on these paths will be assigned the numerical value 1, i.e., $\sigma_{s,t}(v_j)$=1.

- Make the total sum of the values according to the previous steps. We will normalize by multiplying the result obtained by $2/(N-1)(N-2)$.

We apply these steps to two different examples, as shown in Figure A.1, to illustrate how the arrangement of nodes in the graph affects C_B. Note that for simplicity, both graphs are considered unweighted when calculating C_B. The left panel shows a graph $G(5,5)$, which is not complete but connected, whose node b has the highest C_B. To calculate this value, we consider that b appears in the shortest paths $\sigma_{a,c}(b)$, $\sigma_{a,d}(b)$, $\sigma_{a,e}(b)$ and $\sigma_{c,e}(b)$, and we assign the value 1.Of those pairs, on the other hand, $\sigma_{a,d}$ and $\sigma_{c,e}$ have a path that reaches the same value by two different paths; therefore, a value of 2 will be assigned to each of them. Concurrently, $\sigma_{a,c}$ and $\sigma_{a,e}$ will equal one because there is only one shortest path to link the corresponding pairs. So the sum total for node b is $1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + 1$=3.5 and if we normalize ($N = 5$) the final value is $C_B(b)$=0.58. The second example (right panel of Figure A.1) represents the MST, $T(5,4)$, of the graph G. An MST does not contain loops, so the shortest distance between a pair of nodes can only be reached through a single path. This path will, therefore, always be the shortest one so that $\sigma_{s,t}$ in Equation (2.3) will always be 1. Then, $\sigma_{s,t}(v_j)$ will, at most, be equal to 1 for each node v_j. Focusing again on node b, it appears in the following paths $\sigma_{a,c}(b)$, $\sigma_{a,d}(b)$ and $\sigma_{a,e}(b)$, which will all contribute 1 to the sum. The final sum, then 3, normalized as defined above, is 0.5.
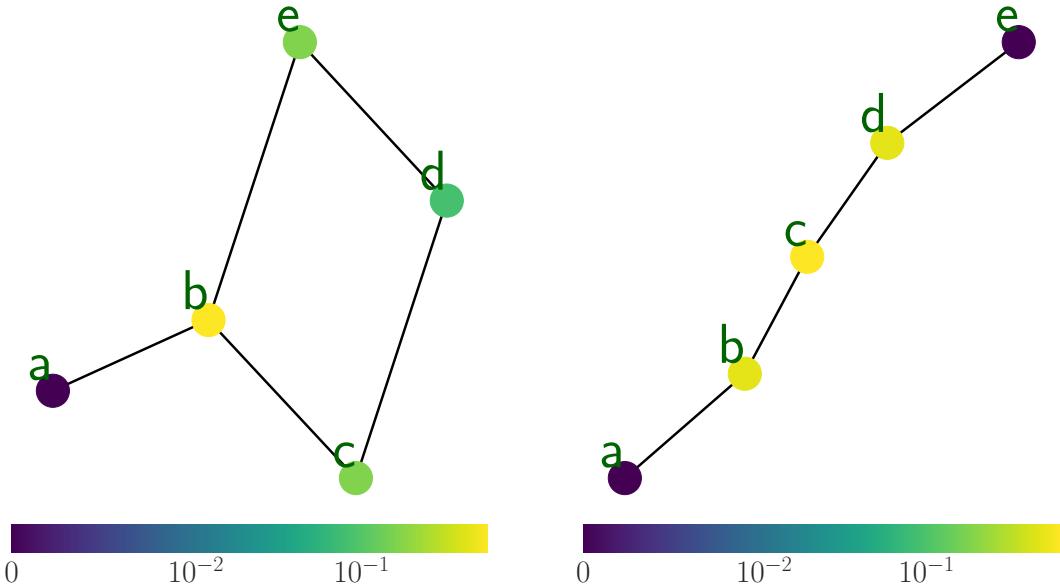
FIGURE A.1: Left: A graph $G(5,5)$ with a cycle composed of nodes (b,c,d,e). The node with the largest C_B is the node b, whose value would be 0.58, while the node a would take the value of 0. Right: A graph $T(5,4)$ is the MST from G. The node with the largest C_B is the node c, whose value is 0.66, while the nodes a and e would take the value of 0. Similar to Figure 5.1, and on the same scale, the color map represents the highest C_B values in yellow colors with increasing darkness for values toward zero.

A.2   Closeness centrality estimator: Practical example

Here, we provide a simple example of the closeness centrality estimator. To calculate the coefficient C_C by applying Equation (2.4) on a graph, we recall to take into account the following considerations:

- Select the node v_j for which you want to compute the closeness centrality.

- Compute the shortest path length $d(v_j, v_t)$ from node v_j to every other node v_t in the graph, with $v_j \neq v_t$. If the graph is unweighted, shortest paths are defined by the minimum number of edges that must be traversed. If weighted, shortest paths minimize the total edge weight. If the shortest path can be reached by different paths due to cycles, $\sigma_{s,t}$ takes a value equal to the number of such paths.

- Sum all shortest path distances from v_j to the other nodes: $\sum_{j \neq t} d(v_j, v_t)$.

- Take the reciprocal of the step above. We will normalize by multiplying the result obtained by $N-1$.

Following Appendix A.1, we apply these steps to two different examples, as shown in Figure A.2, to illustrate how the arrangement of nodes in the graph affects C_C. Note that for simplicity, both graphs are considered unweighted when calculating C_C. The left panel shows a graph $G(5,5)$, which is not complete but connected, whose node b has the highest C_C. The distances from node b to the rest are considered the number of edges in the shortest path: $d(b,a) = 1$, $d(b,c) = 1$, $d(b,d) = 2$, and $d(b,e) = 1$.
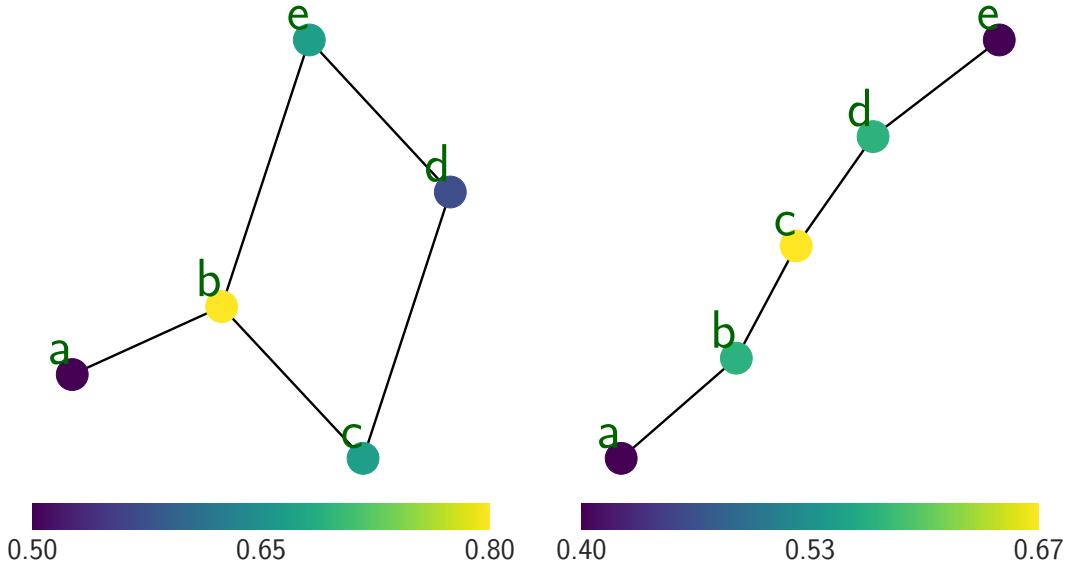
FIGURE A.2: Left: A graph $G(5,5)$ with a cycle composed of nodes (b,c,d,e). The node with the largest C_C is the node b, whose value would be 0.8, while the node a would take the value of 0.5. Right: A graph $T(5,4)$ is the MST from G. The node with the largest C_C is the node c, whose value is 0.66, while the nodes a and e would take the value of 0.4. The color map represents the C_C values.

The total distance is: $\sum_{b \neq t} d(b,t) = 1 + 1 + 2 + 1 = 5$.

The C_C value is then given by Equation (2.4): $C_C(b) = \frac{4}{5} = 0.8$. The second example (right panel of Figure A.2) represents the MST, $T(5,4)$, of the graph G. An MST does not contain loops, so the shortest distance between a pair of nodes can only be reached through a single path.

Focusing again on node b: $d(b,a) = 1$, $d(b,c) = 1$, $d(b,d) = 2$, and $d(b,e) = 3$.

The total distance is: $\sum_{b \neq t} d(b,t) = 1 + 1 + 2 + 3 = 7$.

The C_C value is then given by Equation (2.4): $C_C(b) = \frac{4}{7} = 0.57$.

Appendix B

PCA using EVD and SVD

We illustrate PCA step by step using a small dataset with $n = 4$ data points and $N = 2$ variables. Let us consider the dataset:

$$X = \begin{bmatrix} 3 & 2 \\ 4 & 2 \\ 1 & 1 \\ 0 & 4 \end{bmatrix}, \tag{B.1}$$

where each row corresponds to a data point with two features.

B.1   Computing PCA using EVD

Step 1: Center the data

PCA requires the data to be mean-centered. We compute the mean of each column:

$$\mu = \frac{1}{4} \sum_{i=1}^{4} X_i = \begin{bmatrix} 2 \\ 2.25 \end{bmatrix}. \tag{B.2}$$

The centered data matrix $\tilde{X}$ is obtained by subtracting μ from Equation (B.1):

$$\tilde{X} = X - \mu^T = \begin{bmatrix} 1 & -0.25 \\ 2 & -0.25 \\ -1 & -1.25 \\ -2 & 1.75 \end{bmatrix}. \tag{B.3}$$

Step 2: Compute the Covariance matrix

The covariance matrix C is computed as described in Equation (3.3) using Equation (B.3):

$$\tilde{X}^T \tilde{X} = \begin{bmatrix} 10 & -3 \\ -3 & 4.75 \end{bmatrix}. \tag{B.4}$$

Then:

$$C = \frac{1}{3} \begin{bmatrix} 10 & -3 \\ -3 & 4.75 \end{bmatrix} = \begin{bmatrix} 3.3 & -1 \\ -1 & 1.58 \end{bmatrix}. \tag{B.5}$$

Step 3: Apply Eigenvalue Decomposition

We compute the eigenvalues λ and eigenvectors $\vec{v}$ of C, see Equation (3.6), through the characteristic polynomial:

$$\det(C - \lambda I) = 0 \,. \tag{B.6}$$

Substituting Equation (B.5) in Equation (B.6), and the identity matrix I, we obtain:

$$\det\left(\begin{bmatrix} 3.33 - \lambda & -1 \\ -1 & 1.58 - \lambda \end{bmatrix}\right) = (3.33 - \lambda)(1.58 - \lambda) - 1 = 0 \,. \tag{B.7}$$

Expanding Equation (B.7):

$$\lambda^2 - (3.33 + 1.58)\lambda + (3.33)(1.58) - 1 = 0 \Rightarrow \lambda^2 - 4.91\lambda + 4.26 = 0 \,. \tag{B.8}$$

Solving Equation (B.8), we obtain:

$$\lambda_1 \approx 3.78, \quad \lambda_2 \approx 1.12 \,, \tag{B.9}$$

where λ_1 and λ_2 are the eigenvalues. To find the corresponding $\vec{v}$, we have to solve:

$$(C - \lambda_i I)\vec{v}_i = 0 \,. \tag{B.10}$$

Considering Equation (B.10), for each λ seen in Equation (B.9), we obtain:

$$\vec{v}_1 \approx \begin{bmatrix} -0.91 \\ 0.41 \end{bmatrix} \,, \quad \vec{v}_2 \approx \begin{bmatrix} -0.41 \\ -0.91 \end{bmatrix} \,. \tag{B.11}$$

Equation (B.11) shows the final results after being normalized by $||\vec{v}_i||$.

Step 4: Project the data

To transform the original data into the new basis defined by the PCs, we project $\tilde{X}$ onto $\vec{v}$:

$$Z = \tilde{X} \cdot V \tag{B.12}$$

where V is the matrix of $\vec{v}$ seen as $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$. Applying Equation (B.12) and considering Equation (B.11), we obtain:

$$Z = \begin{bmatrix} -1.013937 & -0.185558 \\ -1.924570 & -0.598774 \\ 0.394113 & 1.551507 \\ 2.544394 & -0.767175 \end{bmatrix} \,. \tag{B.13}$$

where each row in Equation (B.13) gives the new coordinates, PC scores, of X seen in Equation (B.1).

In other words, each column in Equation (B.13) is a PC, which can be calculated as we saw in Equation (3.7):

$$PC_1 = -0.91 \cdot \tilde{X}_{:,1} + 0.41 \cdot \tilde{X}_{:,2} \tag{B.14}$$
$$PC_2 = -0.41 \cdot \tilde{X}_{:,1} - 0.91 \cdot \tilde{X}_{:,2} \,. \tag{B.15}$$

In Equation (B.15) we consider Equation (B.11) and Equation (B.1), where $\tilde{X}_{:,1}$ and $\tilde{X}_{:,2}$ represent each column of $\tilde{X}$, respectively.

B.2 Computing PCA using SVD

Given the X seen in Equation (B.1), the SVD factorizes it as according to Equation (3.9), where:

- $U \in \mathbb{R}^{4 \times 2}$ contains the left singular vectors,

- $\Sigma \in \mathbb{R}^{2 \times 2}$ is a diagonal matrix of singular values $\sigma_i \geq 0$,

- $V \in \mathbb{R}^{2 \times 2}$ contains the right singular vectors.

To calculate U, Σ, and V, according Equation (3.10), we have to use Equation (B.4).

Step 1: Apply Eigenvalue Decomposition

Let's solve the characteristic polynomial:

$$\det(\tilde{X}^T \tilde{X} - \lambda I) = 0 . \tag{B.16}$$

From Equation (B.16):

$$\det\left(\begin{bmatrix} 10 - \lambda & -3 \\ -3 & 4.75 - \lambda \end{bmatrix}\right) = (10 - \lambda)(4.75 - \lambda) - 9 = 0 . \tag{B.17}$$

Expanding Equation (B.17):

$$\lambda^2 - (10 + 4.75)\lambda + (10)(4.75) - 9 = 0 \Rightarrow \lambda^2 - 14.75\lambda + 38.5 = 0 . \tag{B.18}$$

Solving Equation (B.18), we obtain:

$$\lambda_1 \approx 11.36, \quad \lambda_2 \approx 3.39 , \tag{B.19}$$

where λ_1 and λ_2 are the eigenvalues of $\tilde{X}^T \tilde{X}$.
To find the corresponding $\vec{v}$, we have to solve:

$$(\tilde{X}^T \tilde{X} - \lambda_i I)\vec{v}_i = 0 . \tag{B.20}$$

Considering Equation (B.20), for each λ seen in Equation (B.19), we obtain:

$$\vec{v}_1 \approx \begin{bmatrix} -0.91 \\ 0.41 \end{bmatrix} , \quad \vec{v}_2 \approx \begin{bmatrix} -0.41 \\ -0.91 \end{bmatrix} . \tag{B.21}$$

Equation (B.21) shows the final results after being normalized by $||\vec{v}_i||$.

Step 2: Singular values (Σ)

We compute singular values considering Equation (B.19):

$$\sigma_1 = \sqrt{11.36} \approx 3.37, \quad \sigma_2 = \sqrt{3.39} \approx 1.84 . \tag{B.22}$$

Then, considering Equation (B.22):

$$\Sigma = \begin{bmatrix} 3.37 & 0 \\ 0 & 1.84 \end{bmatrix} \tag{B.23}$$

Note applying Equation (3.11), we recover Equation (B.9) based on C.

Step 3: Right singular vectors (V)

Using Equation (B.21), we get:

$$V = \begin{bmatrix} -0.911 & -0.412 \\ 0.412 & -0.911 \end{bmatrix} \tag{B.24}$$

Step 4: Left singular vectors (U)

Using Equation (3.12), we have:

$$U = \tilde{X}V\Sigma^{-1} . \tag{B.25}$$

Then, computing Σ^{-1} considering Equation (B.23):

$$\Sigma^{-1} \approx \begin{bmatrix} 0.2967 & 0 \\ 0 & 0.5435 \end{bmatrix} , \tag{B.26}$$

and considering Equations (B.3) and (B.24):

$$\tilde{X}V \approx \begin{bmatrix} -1.014 & -0.185 \\ -1.925 & -0.598 \\ 0.394 & 1.551 \\ 2.544 & -0.770 \end{bmatrix} , \tag{B.27}$$

following Equation (B.25), we obtain:

$$U = \tilde{X}V\Sigma^{-1} \approx \begin{bmatrix} -0.300 & -0.100 \\ -0.570 & -0.325 \\ 0.116 & 0.842 \\ 0.754 & -0.416 \end{bmatrix} . \tag{B.28}$$

Step 5: Project the data

The PC scores can be obtained via Equation (B.27) or equivalently, using Equation (B.28) and Equation (B.23):

$$Z = \tilde{X}V = U\Sigma \approx \begin{bmatrix} -1.014 & -0.185 \\ -1.925 & -0.598 \\ 0.394 & 1.551 \\ 2.544 & -0.770 \end{bmatrix} . \tag{B.29}$$

Equation (B.29) and Equation (B.13) give the same results, and are equivalent ways of calculating PCs.

Appendix C

Details on algebra with principal components, distance, and rendering of the MST

C.1 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, $\log B_s = K + logP + \log \dot{P}$, where K is a constant, geometrically speaking, 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.

Equation (4.2) is 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 $V(P, \dot{P}) = KP^a \dot{P}^b$, $V_l = \log V = K_l + aP_l + b\dot{P}_l$, the normalization via Equation (4.1) is

$$(V_l)^\dagger = \frac{V_l - Q_{2,V_l}}{IQR_{V_l}} = \frac{K_l + aP_l + b\dot{P}_l - Q_{2,V_l}}{IQR_{V_l}}. \tag{C.1}$$

Once Equation (C.1) is applied to all the variables, a summation will be obtained, which leads to Equation (4.3).

In Figure 4.3, we show how a circle in the $P\dot{P}$ diagram would appear in the plane defined by the PCs. The simulated points located along a circle in the $P\dot{P}$ diagram, defined by its logarithmic coordinates $(P_l, \dot{P}_l)$, can be described by 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. Considering Equation (4.3), we get:

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

Equation (C.2) describes 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 $\theta = 73.38°$.
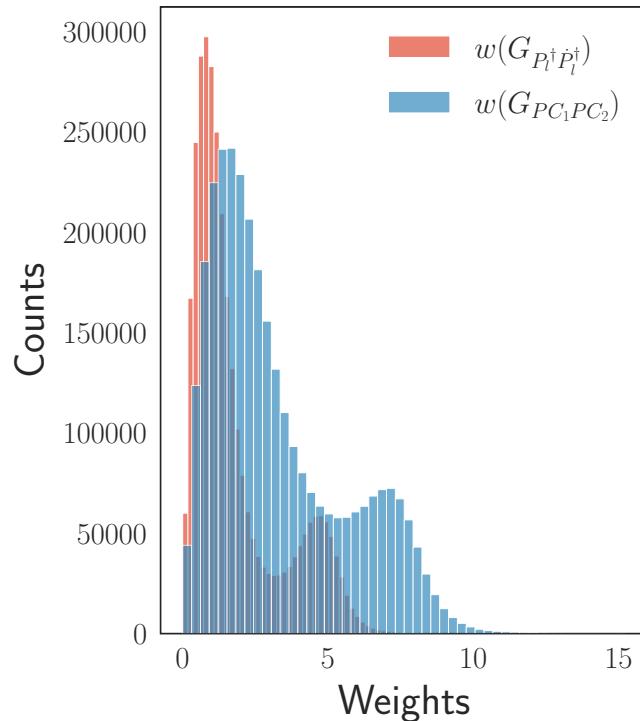
FIGURE C.1: Distributions of the weights associated with the complete, undirected, and weighted graph for two different distance definitions. We show in red the distribution of distances based on the logarithms of the variables P and $\dot{P}$, once they have been normalized, and in blue, the distribution of distances based on PC_1 and PC_2.

C.2 Nearness in the distance context and the MST

The nearness between two pulsars depends on the definition of distance and primarily on the variables considered to compute it; thus, the more comprehensive this distance is, the more accurate it becomes. One way to see this is by comparing the distributions of the weights associated with a complete, undirected and weighted graph considering only P, $\dot{P}$ with those obtained from a similar graph but considering the set of 8 variables described in Section 4.2.1 (equivalently, the first two PCs). This is illustrated in Figure C.1, where the distributions differ, which relates to the discussion derived from the bottom panels of Figure 4.3. In addition, from the set of weights associated with the graph, it is possible to know the nearest neighbors of any randomly chosen pulsar. Considering the ranking, a distance-ordered list of one pulsar relative to all others, it can be seen that neighboring pulsars differ according to each distance definition. For instance, for every pulsar, we identify the nearest three neighbors using a distance based on P and $\dot{P}$, and compare these with the ones obtained using the complete set of variables of interest. We find that 45% of the population incorporates a new pulsar. Even when the same pulsars are concerned, the ranking positions may change in many cases: 40% of the population has at most one pulsar in the same position.

Due to MST properties, we know that the first neighbor in the distance ranking from a given source is indeed one of the nodes attached to the MST. However, the second neighbor and beyond in the distance ranking do not necessarily reflect vis-à-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 minimizes the distance to a given node and the total distance required to visit the entire population. This introduces a global perspective in selecting how particular sources are connected to the rest, ultimately serving as a basis for clustering analysis techniques.

C.3 Representing the MST: how rendering is done

Given that an MST lacks a fixed set of axes, i.e., it lives in a non-Euclidean space, its rendering (angles among branches, orientation, branch direction) does not hold any particular meaning; only the connections among nodes do. A different representation could be chosen, while conserving the fundamental characteristic —i.e., preserving the exact distance definition —which will keep each node linked to the same neighbors and in the same order. Despite the appearance difference, the nearness and connections described in one MST would be the same as in the other. An obvious example is to take one of the branches that go rightwards in the Pulsar Tree (see Figure 4.4) and force it to go leftwards. This would produce the same MST (the same adjacency matrix, in technical terms, see Chapter 2.2), and everything we have said looking at the MST with the same branch pointing rightwards would also apply to the new rendering. It is not the appearance of the overall graph that is important, but rather the properties of the graph.

For the MST representations in this work, we utilize the NEATO program, which employs the Graphviz Python library (The graphviz team, 2022) to minimize a pseudo-energy and select the orientation of the different components.

Appendix D

Details on significant branches and the MST for the MSP population

Here, we present a concise and focused explanation of how the clustering algorithm introduced in Chapter 4 is applied to the population of MSPs. We apply the betweenness centrality estimator (see Section 2.4.1) to identify the most central nodes of the graph T. Values identified as outliers in betweenness centrality will be referred to as potential trunk nodes, or pTNs. The trunk is defined as a path, i.e., a sequence of non-repeated nodes. At the end of the trunk, the pTNs that delimit are called contour nodes (CN). These CNs must be nodes of degree greater than 2, acting as articulation points for the graph. Removing these nodes partitions the graph into connected components. The connected components originating in a CN are called contour branches (CB). This leaves a clustered structure with a minimum of two groupings at each extreme of the trunk. To avoid limiting the trunk in nodes of degree 2 or larger but from which the departing structures are of non-representative size (i.e., we aim to prevent noise near the trunk), we define a significant threshold (α_s) so that the number of nodes in the CBs is requested to exceed this lower limit. We set the threshold at 5%, implying that branches will contain at least 11 pulsars of the total population. Once all the possible CNs are identified, we calculate all the possible trunks as the paths that pair these CNs. The number of trunks obtained: $\#Trunks = \text{CNs} \times (\text{CNs} - 1)/2$. For each possible trunk, we have a set of branches constituted by the connected components starting from every node of a degree larger than 2 (including the CBs) and having α_s as a lower limit to the number of nodes each one must contain.

This results in six sets of branches and trunks. We choose the set that provides more excellent uniformity in the size of the resulting groups, which contributes to the robustness of a comparison analysis, such as a non-parametric test, such as the Kolmogorov-Smirnov (KS) statistics (see, e.g., Yadolah, 2008; Wolfe, 2012; Lehmann, 2012). In this case, the null hypothesis (H_0) is that the distribution of the variables of the nodes of two given branches is consistent with them coming from the same parent distribution. If we reject this null hypothesis, say at the 95% confidence level (CL) or better, we may conclude that two branches may be formed by different pulsars or at different evolutionary stages. Table D.1 shows the KS test results for these branches, where the H_0 column indicates which branches differ the most under the above assumptions. It is observed that most branches are distinguished by most individual variables.

TABLE D.1: Comparison of the branches in Figure 6.6. We denote with a number 1 when two branches, according to the KS-test for a 95% CL, reject H_0 for the analyzed variable. Instead, we denote the opposite case with a number of 0, where we can not reject H_0. The last column shows the accumulated sum for each confrontation between branches, counting for which variables these branches reject H_0. For visualization reasons, we denote with '*' those colors in which dark is suppressed, alluding to the dark green and dark blue branches.

Branches	B_{lc}	B_s	$\dot{E}$	η_{GJ}	P	$\dot{P}$	$\Delta\Phi$	P_B	A_1	M_C	H_0
Green* - Gray	0	1	1	1	0	1	1	1	1	1	8
Green* - Pink	1	1	1	1	0	1	1	0	0	0	6
Green* - Blue*	1	1	1	0	1	1	1	1	1	1	9
Green* - Orange	0	1	1	1	1	1	1	1	1	1	9
Green* - Brown	1	1	1	1	1	0	1	1	1	1	9
Gray - Pink	1	0	1	1	0	0	1	1	1	1	7
Gray - Blue*	1	1	1	1	1	1	1	1	1	0	9
Gray - Orange	1	1	1	1	1	1	1	0	0	0	7
Gray - Brown	1	1	1	1	1	1	1	1	1	1	10
Pink - Blue*	1	0	1	1	1	1	1	1	1	1	9
Pink - Orange	1	1	1	1	0	1	1	1	1	1	9
Pink - Brown	1	1	1	1	1	1	1	1	1	1	10
Blue* - Orange	1	1	1	1	1	1	1	0	0	0	7
Blue* - Brown	0	1	0	1	1	1	0	1	0	0	5
Orange - Brown	1	1	1	0	1	1	1	1	1	1	9

Appendix E

No moding in identified tMSP candidates in the radio state

The EPIC-pn (Strüder et al., 2001) operated with a time resolution of 73.4 ms (full frame mode) for 1723-2837 and J1902-5105, and 47.4 ms (large window mode) for J1431-4715. The two EPIC-MOS (Turner et al., 2001) were employed with a time resolution of 2.6 s (full frame mode) in the first two cases and 0.9 s (large window mode) for J1431-4715. We processed and analyzed the data using the Science Analysis Software (SAS; v.21.0.0). All three observations exhibited high background activity in the 10–12 keV light curve, so we excluded the contaminated data intervals. For the EPIC-pn and each MOS, we extracted source photons within a circular region centered on the source position with a 40° radius, and background photons from an 80″-wide, source-free circular region. Using the `epiclccorr` task, 0.3-10 keV background-subtracted light curves were extracted from the three EPIC instruments over the simultaneous coverage time interval, binned with a time resolution of 100 s. As previously noted, we recall that the X-ray flux of J1723-2837 (top panel of Figure E.1) is variable with the orbital phase as discussed in Bogdanov et al., 2014. We searched the *XMM-Newton* light curves for the typical bimodality between 'high' and 'low' intensity modes (see Figure E.1; see de Martino et al., 2024, see also (Koljonen and Linares, 2023) for a detailed analysis of J1431-4715). We experimented with different bin sizes but did not observe the expected bimodal feature. Further analysis of the count rate distributions confirmed the absence of this bimodality. A deep search for pulsation has identified PSR J1431-4715 as a gamma-ray pulsar Smith et
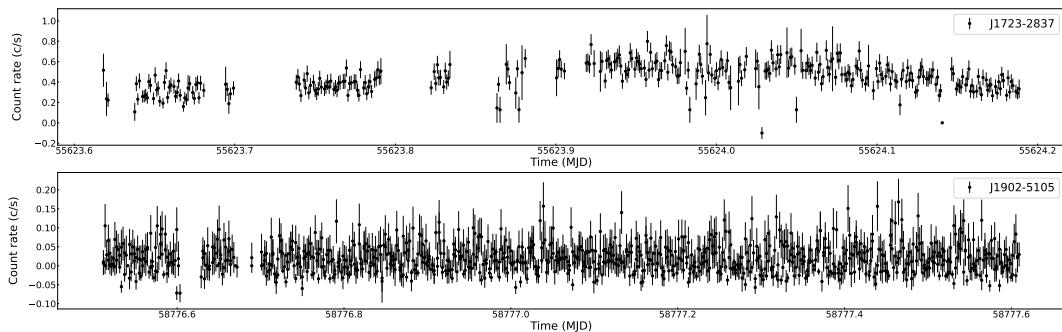
FIGURE E.1: *XMM-Newton* 0.3-10 keV background-subtracted light curves over the time interval in which the three EPIC instruments collected data simultaneously for J1723-2837 (top panel) and J1902-5105 (bottom panel). The light curves are binned with a time resolution of 100 s. For a detailed analysis of J1431-4715 see de Martino et al., 2024; Koljonen and Linares, 2023.

al., 2023; its gamma-ray emission shows variable signature as those found for prototypical tMSPs Torres et al., 2017. Note that the prototype for tMSPs, J1023+0038, is located at ∼1.37 kpc (Deller et al., 2012), while the estimated distances for J1723-2837 and J1902–5105 are ∼0.75 kpc (Bogdanov et al., 2014) and ∼1.2 kpc (Camilo et al., 2015), respectively. Assuming comparable luminosities, if a similar bimodal behavior were present, we would expect to observe it in these closer sources. Moreover, distinct high and low modes are evident in the faint candidate tMSP 3FGL J1544.6-1125 (Bogdanov and Halpern, 2015). We have re-analyzed the *XMM-Newton* archival observation ObsID 072080101 as a check for the latter source. The bimodal pattern was detected in the EPIC/MOS1 0.3-10 keV background-subtracted light curve, binned with a time resolution of 100 s and exhibiting an average count rate comparable to that of J1723-2837. Thus, should it be present in J1723-2837, it should have become similarly apparent.

Appendix F

Evaluating metrics

We summarize the key statistical metrics used to evaluate the separation between repeater and non-repeater FRBs seen in Chapter 7. The evaluating metrics provided are precision, recall, F-scores, the Receiver Operating Characteristic (ROC) analysis, and the Z-score used for assessing statistical significance relative to randomized baselines, see e.g., Han, Kamber, and Pei, 2012; Bishop, 2007; Géron, 2019. These tools allow for interpretable assessment of how well the selected properties through the MST approach distinguish between the determined FRB categories in a manner consistent with machine learning methods.

Precision quantifies the ratio of repeaters that were correctly identified compared to the total in the branch, while *recall* (also known as *sensitivity*) measures the proportion of actual repeaters that were correctly classified. The *F measure*, defined as

$$F_\beta = (1 + \beta^2) \frac{Precision\ Recall}{(\beta^2\ Precision) + Recall} \tag{F.1}$$

evaluates the predictive performance by combining precision and recall through a weighted harmonic mean. The importance of precision ($\beta < 1$) versus recall ($\beta > 1$) is modulated by the β parameter. When $beta = 1$, precision and recall are equally weighted. The F_1 score is obtained when $\beta = 1$, resulting in equal emphasis on precision and recall. The F_2 score, alternatively, is derived when $\beta = 2$, giving more importance to recall over precision. The *Receiver Operating Characteristic (ROC)* curve is a graphical representation of a binary classification model's performance across different threshold settings. It plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold values. Here, Specificity, also known as the true negative rate, measures the proportion of actual negative instances (non-repeater FRBs) correctly identified as negative. The ROC curve helps assess the trade-off between sensitivity and specificity, providing insights into the classifier's ability to discriminate between positive and negative instances. The Area Under the Curve (AUC) value quantifies the overall performance of a binary classification model represented by the ROC curve. It measures the area under the ROC curve, with values ranging from 0 to 1. AUC values closer to 1 indicate the classifier's better discrimination ability, while values closer to 0.5 suggest poor discrimination (similar to random guessing).

The labels assigned to repeaters and non-repeaters according to the repeater branch are as follows:

- **True Positive (TP)**: Repeaters correctly classified as repeaters

- **False Positive (FP)**: Non-repeaters incorrectly classified as repeaters.

- **True Negative (TN)**: Non-repeaters correctly classified as non-repeaters.

- **False Negative (FN)**: Repeaters incorrectly classified as non-repeaters.

The scores are defined as

$$Precision = \frac{TP}{TP + FP} \tag{F.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{F.3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{F.4}$$

$$F_1\ Score = \frac{2\ Precision\ Recall}{Precision + Recall} \tag{F.5}$$

$$F_2\ Score = \frac{5\ Precision\ Recall}{4\ Precision + Recall} \tag{F.6}$$

The *Z-score*, also known as the standard score or z-value, is a statistical measure that quantifies how many standard deviations a data point is from the mean of a dataset. It is defined as

$$Z = \frac{x - \mu}{\sigma}. \tag{F.7}$$

Here, x is the value of the data point (the classifier's recall), μ and σ are the mean and the standard deviation of the dataset (distribution from the fake cases), respectively, see Section 7.3.1 for a proper application. A Z-score of 0 indicates that the data point is strictly at the mean of the dataset, while positive and negative Z-scores suggest that the data point is above and below the mean, respectively. A Z-score higher than 2σ or 3σ in absolute value is significant at the 5% or even 1% level.

Appendix G

Working with DTW

G.1 Optimal warping path computation

Consider, as an example, the following two-time series, see Figure G.1, one with five and another with seven instances

$$
\begin{aligned}
ts_1 &= \quad\ 3,1,2,2,1 \\
ts_2 &= \ 2,0,0,3,3,1,0.
\end{aligned}
\tag{G.1}
$$

We calculate the Euclidean distance matrix (E), see eg., Dokmanic et al., 2015), as

$$
E = E(d(i,j)^2) =
\begin{bmatrix}
1 & 9 & 9 & 0 & 0 & 4 & 9 \\
1 & 1 & 1 & 4 & 4 & 0 & 1 \\
0 & 4 & 4 & 1 & 1 & 1 & 4 \\
0 & 4 & 4 & 1 & 1 & 1 & 4 \\
1 & 1 & 1 & 4 & 4 & 0 & 1
\end{bmatrix}.
\tag{G.2}
$$

Here, each value is the square of the difference between the elements of the series located in (i,j), that is, $d = (ts_1(i) - ts_2(j))^2$, where i is an index running on ts_1, from 1 to 5, and j runs on ts_2, from 1 to 7.
In the next step, we compute the cumulative distance matrix (D) through the following recurrence relation seen in the application of Equation (8.5):

$$
D = D(i,j) =
\begin{bmatrix}
1 & 10 & 19 & 19 & 19 & 23 & 32 \\
2 & 2 & 3 & 7 & 11 & 11 & 12 \\
2 & 6 & 6 & 4 & 5 & 6 & 10 \\
2 & 6 & 10 & 5 & 5 & 6 & 10 \\
3 & 3 & 4 & 8 & 9 & 5 & 6
\end{bmatrix}
\tag{G.3}
$$

For instance, the value seen in row 2 and column 2 in Equation (G.3) is obtained as $D(2,2) = E_d(2,2) + \min\big(D(1,2), D(2,1), D(1,1)\big) = 1 + \min(2,10,1) = 2$.
The aim will be to find the optimal path (the one with the minimal cost), referred to as W_o, that relates the two series. This can be found by tracing backward (backtracking) in D, taking the previous cells' lowest cumulative values from the initial cell. The optimization will be subject to the following conditions. First, it needs to connect the beginning and the end of each time series, i.e., it will be a path, W, that starts at cell (5, 7) and ends at cell (1, 1). Additionally, we must ensure that the path W respects the sequential order of the time series, so it should move through the cells either to the left or upward, but never to the right or downward. For instance, from cell (5, 7), the next lowest value is 5, which occupies the cell (5, 6). Finally, to avoid skipping any time series element, each step in W is constrained to reach only neighboring cells. For instance, the next cell to get from (5, 6) can only be the cells
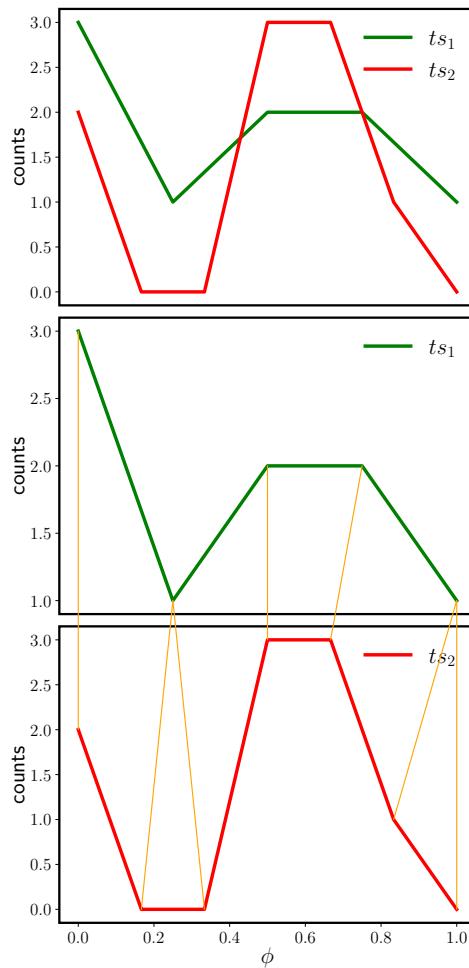
FIGURE G.1: Top panel: Time series ts_1 (green) and ts_2 (yellow) used in the DTW example. According to Equation (G.1), the y-axis represents an arbitrary number of counts. The x-axis represents the phases for each value, from 0 to 1, defined by the number of values in each time series. Bottom panel: Application of DTW. The orange lines represent W_o, the connection between the pairs of the time series (i.e., the chosen (i, j) pairs from Equation (G.2)).

(4, 5), (4, 6), or (5, 5), whose values, respectively, are 5, 6, and 9, thus W goes through (4, 5) which is the cell with the minimum cost. When more than one possible cell with the same value exists, the diagonal cell is prioritized as the most natural one-to-one alignment. In this case, the colored values seen in Equation (G.3) represent W_o, which connects the elements (1, 1), (2, 2), (2, 3), (3, 4), (4, 5), (5, 6) and, (5, 7). Its final cost, i.e., the DTW value, is the square root of the sum of each of the above-selected elements, as per Equation (G.2) and yields 2.45. Visually, W_o is shown in the bottom panel of Figure G.1, where each orange line represents the connection of each of the elements extracted from Equation (G.3).

G.2 The Euclidean path

Consider the following time series, having the same length

$$
\begin{aligned}
ts_3 &= 3,1,2,2,1 \\
ts_4 &= 2,0,0,3,3
\end{aligned}
\tag{G.4}
$$

Calculating E_d and D, as shown before we get

$$
E_d =
\begin{bmatrix}
1 & 9 & 9 & 0 & 0 \\
1 & 1 & 1 & 4 & 4 \\
0 & 4 & 4 & 1 & 1 \\
0 & 4 & 4 & 1 & 1 \\
1 & 1 & 1 & 4 & 4
\end{bmatrix},
\tag{G.5}
$$

and

$$
D =
\begin{bmatrix}
1 & 10 & 19 & 19 & 19 \\
2 & 2 & 3 & 7 & 11 \\
2 & 6 & 6 & 4 & 5 \\
2 & 6 & 10 & 5 & 5 \\
3 & 3 & 4 & 8 & 9
\end{bmatrix}.
\tag{G.6}
$$

The main diagonal in Equation (G.6) (noted in blue) represents a W that has been explored in computing Equation (8.2), arising by the connection of all (i, j) with $i = j$; a point-by-point comparison of the time series. According to Equation (G.5), its cost would be 3.31. However, when applying Equation (8.2), we see that another W, that would involve a different set of pairs (i, j), and has a lower cost, being W_o in Equation (G.6) (noted in orange). According to Equation (G.5), the cost of the blue path would be 3.00. In Figure G.2, the alignment through the set (i, j) following W and W_o, corresponding to both methods, is shown with the same color code used in Equation (G.6). Therefore, if the time series have the same length, the direct alignment between them (also referred to as the Euclidean alignment or just the Euclidean distance (ED) of the time series) will be computed and compared in the search for the minimal value. This may or may not be the W_o; in this example, it is not.
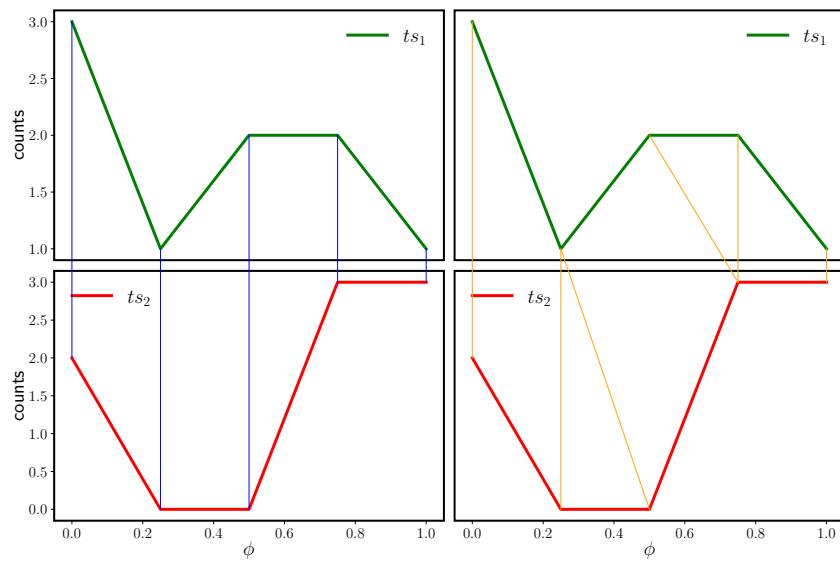
FIGURE G.2: Time series ts_3 (green) and ts_4 (yellow) seen in Equation (G.4) connected via the Euclidean alignment (left) or DTW (right).

Appendix H

Implementation and computational cost

All the codes of this thesis are built on *Python* v 3.11.2 (Van Rossum and Drake Jr, 1995), in which we implement the *Psrqpy* package (Pitkin, 2018) to deal with the population of pulsars from the ATNF catalogue. 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 Pulsar Tree web is done using *Bokeh* (The Bokeh team, 2022). All computations, except those shown in Section H.3, were performed on a laptop equipped with a 2.0 GHz Quad-Core Intel Core i5 processor, 16 GB of 3733 MHz LPDDR4X memory, and Intel Iris Plus Graphics (1536 MB).

H.1 Computing MSTs

To create the graphs, we use the *NetworkX* (Hagberg, Schult, and 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, which we have used to calculate the MST.

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, and 2509 nodes from a complete, undirected, and weighted graph with 2509 nodes and 3146286 edges can be addressed from two approaches:

- A brute-force method that builds the complete graph explicitly, computing all pairwise Euclidean distances and storing them as edge weights in a fully connected graph using *NetworkX*. This graph is then processed using *NetworkX*'s `minimum_spanning_tree` function to extract the MST. This method yields all the necessary edges: 3146286 edges for the given dataset of 2509 nodes. In practice, this implementation took ∼300 seconds per execution.

- This method computes the MST directly from the pairwise distance matrix using `scipy.sparse.csgraph.minimum_spanning_tree`. It avoids creating the full graph with all edges and instead uses a more efficient internal representation. The MST for the given dataset is computed in about 1.2 seconds.

H.2   Computing Significant Branches

The clustering algorithm designed to identify significant branches in a given MST begins by computing betweenness centrality using *NetworkX*'s `betweenness_centrality` function. This allows us to detect the most central nodes, which are considered candidates for forming the main trunk of the tree. From these nodes, the graph is then partitioned around these paths to identify the branches. In the second stage, the algorithm evaluates the statistical significance of each branch using the two-sample KS test, `scipy.stats.ks_2samp`, applied to physical properties across branch pairs. For the dataset of 2509 pulsars, computing the significant branches and storing the potential ones takes approximately 35 seconds.

H.3   Computing DTW

The work in Chapter 9 was conducted utilizing the *dtaidistance* 2.3.10 package (wannesm et al., 2022) for the DTW application. The number of DTW values for each light curve comparison, $DTW(lc_1, lc_2)$, depends on the N of each light curve. Due to the applied rotation being N_1, and N_2, the size corresponding to lc_1 and lc_2, respectively, to this comparison is $N_1 \times N_2$. Considering the 43071 possible pairs, the N of each light curve, the number of computations is higher than 3×10^8. The implementation has been carried out on the Hidra High-Performance Computing (HPC) Cluster, which is hosted at the Institute of Space Sciences (ICE, CSIC). For example, in a comparison involving light curves of 25 bins and 50 bins, the time usage required for a single (bin-to-bin) DTW computation is ~ 0.0014 s, while the total time considering the shift is 1.407 s. In this case, the memory usage is negligible, 0.78 MB. For a comparison, taking light curves of 100 bins and 200 bins, the time usage for a single DTW is ~ 0.0190 s, with a total of 371.186 s if looking at the complete rotation set. Let $lc1, lc2$=400 bins, the number of DTW computations performed for a complete rotation set is $\sim$160000. The total time taken for all calculations, assuming the complete sample was no less than 72 hours, was mainly driven by the lc of 400 bins. The memory usage is around 3.75 MB. The computations were distributed using multiprocessing across 16 cores for a more efficient process.

Bibliography

Aach, John and George Church (July 2001). "Aligning Gene Expression Time Series With Time Warping Algorithms". In: *Bioinformatics (Oxford, England)* 17, pp. 495–508. DOI: 10.1093/bioinformatics/17.6.495.

Abbott, B. P. et al. (Feb. 2016). "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Phys. Rev. Lett.* 116.6, 061102, p. 061102. DOI: 10.1103/PhysRevLett.116.061102. arXiv: 1602.03837 [gr-qc].

Abdo, A. A. et al. (Oct. 2013a). "The Second Fermi Large Area Telescope Catalog of Gamma-Ray Pulsars". In: *ApJS* 208.2, 17, p. 17. DOI: 10.1088/0067-0049/208/2/17. arXiv: 1305.4385 [astro-ph.HE].

— (Oct. 2013b). "The Second Fermi Large Area Telescope Catalog of Gamma-Ray Pulsars". In: *ApJS* 208.2, 17, p. 17. DOI: 10.1088/0067-0049/208/2/17. arXiv: 1305.4385 [astro-ph.HE].

Abhishek et al. (Jan. 2022). "Radio Pulsar Sub-Populations (II) : The Mysterious RRATs". In: *arXiv e-prints*, arXiv:2201.00295, arXiv:2201.00295. arXiv: 2201.00295 [astro-ph.HE].

Agrawal, Rakesh et al. (1995). "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases". In: *Proceedings of the 21th International Conference on Very Large Data Bases*. VLDB '95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 490–501. ISBN: 1558603794.

Ai, Shunke, He Gao, and Bing Zhang (Jan. 2021). "On the True Fractions of Repeating and Nonrepeating Fast Radio Burst Sources". In: *ApJ* 906.1, L5, p. L5. DOI: 10.3847/2041-8213/abcec9. arXiv: 2007.02400 [astro-ph.HE].

Alpar, M. A. et al. (Dec. 1982). "A new class of radio pulsars". In: *Nature* 300.5894, pp. 728–730. DOI: 10.1038/300728a0.

Althaus, Leandro G. et al. (Oct. 2010). "Evolutionary and pulsational properties of white dwarf stars". In: *A&ARv* 18.4, pp. 471–566. DOI: 10.1007/s00159-010-0033-1. arXiv: 1007.2659 [astro-ph.SR].

Altmann, André et al. (Apr. 2010). "Permutation importance: a corrected feature importance measure". In: *Bioinformatics* 26.10, pp. 1340–1347. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq134. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/10/1340/48851160/bioinformatics_26_10_1340.pdf. URL: https://doi.org/10.1093/bioinformatics/btq134.

Ambartsumyan, V. A. and G. S. Saakyan (Oct. 1960). "The Degenerate Superdense Gas of Elementary Particles". In: *Soviet Ast.* 4, p. 187.

Amiri, M. et al. (2019). "A second source of repeating fast radio bursts". In: *Nature* 566.7743, pp. 235–238. DOI: 10.1038/s41586-018-0864-x. URL: https://doi.org/10.1038/s41586-018-0864-x.

An, Hongjun et al. (2020). "Orbital Modulation of Gamma Rays from PSR J2339-0533". In: *The Astrophysical Journal* 897.1, p. 52. DOI: 10.3847/1538-4357/ab93ba. URL: https://dx.doi.org/10.3847/1538-4357/ab93ba.

Ans, Bernard, Jeanny Herault, and Christian Jutten (Jan. 1985). "Architectures neuromimétiques adaptatives: Détection de primitives". In: *Proceedings of Cognitiva 85* 2, pp. 593–597.

Archibald, Anne M. et al. (June 2009). "A Radio Pulsar/X-ray Binary Link". In: *Science* 324.5933, p. 1411. DOI: 10.1126/science.1172740. arXiv: 0905.3397 [astro-ph.HE].

Archibald, Anne M. et al. (Oct. 2010). "X-ray Variability and Evidence for Pulsations from the Unique Radio Pulsar/X-ray Binary Transition Object FIRST J102347.6+003841". In: *ApJ* 722.1, pp. 88–95. DOI: 10.1088/0004-637X/722/1/88. arXiv: 1008.1068 [astro-ph.HE].

Archibald, Anne M. et al. (July 2015). "Accretion-powered Pulsations in an Apparently Quiescent Neutron Star Binary". In: *ApJ* 807.1, 62, p. 62. DOI: 10.1088/0004-637X/807/1/62. arXiv: 1412.1306 [astro-ph.HE].

Archibald, R. F. et al. (Mar. 2016a). "A High Braking Index for a Pulsar". In: *ApJ* 819.1, L16, p. L16. DOI: 10.3847/2041-8205/819/1/L16. arXiv: 1603.00305 [astro-ph.HE].

Archibald, R. F. et al. (Sept. 2016b). "A Magnetar-like Outburst from a High-B Radio Pulsar". In: *ApJ* 829.1, L21, p. L21. DOI: 10.3847/2041-8205/829/1/L21. arXiv: 1608.01007 [astro-ph.HE].

Arons, J. (Mar. 1983). "Pair creation above pulsar polar caps : geometrical structure and energetics of slot gaps." In: *ApJ* 266, pp. 215–241. DOI: 10.1086/160771.

Arya, Meenakshi S., R. Deepa, and Jayati Gandhi (2021). "Dynamic Time Warping-Based Technique for Predictive Analysis in Stock Market". In: *Proceedings of 6th International Conference on Recent Trends in Computing*. Ed. by Rajendra Prasad Mahapatra et al. Singapore: Springer Singapore, pp. 23–36. ISBN: 978-981-33-4501-0.

Athitsos, Vassilis et al. (2008). "Approximate embedding-based subsequence matching of time series". In: *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. SIGMOD '08. Vancouver, Canada: Association for Computing Machinery, 365–378. ISBN: 9781605581026. DOI: 10.1145/1376616.1376656. URL: https://doi.org/10.1145/1376616.1376656.

Baade, W. (Sept. 1942). "The Crab Nebula." In: *ApJ* 96, p. 188. DOI: 10.1086/144446.

Baade, W. and F. Zwicky (1934). "Remarks on Super-Novae and Cosmic Rays". In: *Phys. Rev.* 46 (1), pp. 76–77. DOI: 10.1103/PhysRev.46.76.2. URL: https://link.aps.org/doi/10.1103/PhysRev.46.76.2.

Backer, D. C., R. S. Foster, and S. Sallmen (1993). "A second companion of the millisecond pulsar 1620 –26". In: *Nature* 365.6449, pp. 817–819. DOI: 10.1038/365817a0. URL: https://doi.org/10.1038/365817a0.

Baglio, M. C. et al. (Sept. 2023). "Matter ejections behind the highs and lows of the transitional millisecond pulsar PSR J1023+0038". In: *A&A* 677, A30, A30. DOI: 10.1051/0004-6361/202346418. arXiv: 2305.14509 [astro-ph.HE].

Bailes, M. et al. (Sept. 2011). "Transformation of a Star into a Planet in a Millisecond Pulsar Binary". In: *Science* 333.6050, p. 1717. DOI: 10.1126/science.1208890. arXiv: 1108.5201 [astro-ph.SR].

Bannister, K. W. et al. (Aug. 2019). "A single fast radio burst localized to a massive galaxy at cosmological distance". In: *Science* 365.6453, pp. 565–570. DOI: 10.1126/science.aaw5903. arXiv: 1906.11476 [astro-ph.HE].

Bar-Joseph, Ziv et al. (2002). "A new approach to analyzing gene expression time series data". In: *Proceedings of the Sixth Annual International Conference on Computational Biology*. RECOMB '02. Washington, DC, USA: Association for Computing Machinery, 39–48. ISBN: 1581134983. DOI: 10.1145/565196.565202. URL: https://doi.org/10.1145/565196.565202.

Baron, Dalya and Brice Ménard (Aug. 2021). "Extracting the Main Trend in a Data Set: The Sequencer Algorithm". In: *ApJ* 916.2, 91, p. 91. DOI: 10.3847/1538-4357/abfc4d. arXiv: 2006.13948 [cs.LG].

Barrow, J. D., S. P. Bhavsar, and D. H. Sonoda (Sept. 1985). "Minimal spanning trees, filaments and galaxy clustering". In: *MNRAS* 216, pp. 17–35. DOI: 10.1093/mnras/216.1.17.

Bartolini, I., P. Ciaccia, and M. Patella (2005). "WARP: accurate retrieval of shapes using phase of Fourier descriptors and time warping distance". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.1, pp. 142–147. DOI: 10.1109/TPAMI.2005.21.

Bassa, C. G. et al. (June 2014). "A state change in the low-mass X-ray binary XSS J12270-4859". In: *MNRAS* 441.2, pp. 1825–1830. DOI: 10.1093/mnras/stu708. arXiv: 1402.0765 [astro-ph.HE].

Bates, S. D. et al. (Feb. 2015). "The High Time Resolution Universe survey - XI. Discovery of five recycled pulsars and the optical detectability of survey white dwarf companions". In: *MNRAS* 446.4, pp. 4019–4028. DOI: 10.1093/mnras/stu2350. arXiv: 1411.1288 [astro-ph.SR].

Bell Burnell, S. Jocelyn (1977). "PETIT FOUR". In: *Annals of the New York Academy of Sciences* 302.1, pp. 685–689. DOI: https://doi.org/10.1111/j.1749-6632.1977.tb37085.x. eprint: https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1977.tb37085.x. URL: https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1977.tb37085.x.

Bellman, R. and R. Kalaba (1959). "On adaptive control processes". In: *IRE Transactions on Automatic Control* 4.2, pp. 1–9. DOI: 10.1109/TAC.1959.1104847.

Bellman, Richard E. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton: Princeton University Press. ISBN: 9781400874668. DOI: doi:10.1515/9781400874668. URL: https://doi.org/10.1515/9781400874668.

Beloborodov, Andrei M. (July 2017). "A Flaring Magnetar in FRB 121102?" In: *ApJ* 843.2, L26, p. L26. DOI: 10.3847/2041-8213/aa78f3. arXiv: 1702.08644 [astro-ph.HE].

Berndt, Donald J. and James Clifford (1994). "Using Dynamic Time Warping to Find Patterns in Time Series". In: *KDD Workshop*. URL: https://api.semanticscholar.org/CorpusID:929893.

Bhandari, Shivani et al. (Feb. 2022). "Characterizing the Fast Radio Burst Host Galaxy Population and its Connection to Transients in the Local and Extragalactic Universe". In: *AJ* 163.2, 69, p. 69. DOI: 10.3847/1538-3881/ac3aec. arXiv: 2108.01282 [astro-ph.HE].

Bhatporia, Shruti et al. (Nov. 2023). "A Topological Data Analysis of the CHIME/FRB Catalogues". In: *arXiv e-prints*, arXiv:2311.03456, arXiv:2311.03456. DOI: 10.48550/arXiv.2311.03456. arXiv: 2311.03456 [astro-ph.HE].

Bhattacharya, D. and E. P. J. van den Heuvel (Jan. 1991). "Formation and evolution of binary and millisecond radio pulsars". In: *Phys. Rep.* 203.1-2, pp. 1–124. DOI: 10.1016/0370-1573(91)90064-S.

Bignami, G et al. (July 2003). "The magnetic field of an isolated neutron star from X-ray cyclotron absorption lines". In: *Nature* 423, pp. 725–7. DOI: 10.1038/nature01703.

Bildsten, Lars and Deepto Chakrabarty (2001). "A Brown Dwarf Companion for the Accreting Millisecond Pulsar SAX J1808.4–3658". In: *The Astrophysical Journal* 557.1, p. 292. DOI: 10.1086/321633. URL: https://dx.doi.org/10.1086/321633.

Bishop, Christopher M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 1st ed. Springer. ISBN: 0387310738. URL: http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738%3FSubscriptionId%3D13CT5CVB80YFWJEPWS02%26tag%3Dws%

26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%
3D0387310738.

Blaauw, A. (May 1961). "On the origin of the O- and B-type stars with high veloc-
ities (the "run-away" stars), and some related problems". In: *Bull. Astron. Inst.
Netherlands* 15, p. 265.

Bochenek, C. D. et al. (Nov. 2020). "A fast radio burst associated with a Galactic
magnetar". In: *Nature* 587.7832, pp. 59–62. DOI: 10.1038/s41586-020-2872-x.
arXiv: 2005.10828 [astro-ph.HE].

Bogdanov, Slavko and Jules P. Halpern (2015). "IDENTIFICATION OF THE HIGH-
ENERGY GAMMA-RAY SOURCE 3FGL J1544.6–1125 AS A TRANSITIONAL
MILLISECOND PULSAR BINARY IN AN ACCRETING STATE". In: *The Astro-
physical Journal Letters* 803.2, p. L27. DOI: 10.1088/2041-8205/803/2/L27. URL:
https://dx.doi.org/10.1088/2041-8205/803/2/L27.

Bogdanov, Slavko et al. (Jan. 2014). "X-Ray and γ-Ray Studies of the Millisecond Pul-
sar and Possible X-Ray Binary/Radio Pulsar Transition Object PSR J1723-2837".
In: *ApJ* 781.1, 6, p. 6. DOI: 10.1088/0004-637X/781/1/6. arXiv: 1311.2024
[astro-ph.HE].

Bogdanov, Slavko et al. (June 2015). "Coordinated X-Ray, Ultraviolet, Optical, and
Radio Observations of the PSR J1023+0038 System in a Low-mass X-Ray Binary
State". In: *ApJ* 806.2, 148, p. 148. DOI: 10.1088/0004-637X/806/2/148. arXiv:
1412.5145 [astro-ph.HE].

Bonnaire, Tony et al. (May 2020). "T-ReX: a graph-based filament detection method".
In: *A&A* 637, A18, A18. DOI: 10.1051/0004-6361/201936859. arXiv: 1912.00732
[astro-ph.CO].

Boroson, Todd A. and Richard F. Green (May 1992). "The Emission-Line Properties
of Low-Redshift Quasi-stellar Objects". In: *ApJS* 80, p. 109. DOI: 10.1086/191661.

Braglia, C et al. (Aug. 2020). "A multiwavelength search for black widow and red-
back counterparts of candidate γ-ray millisecond pulsars". In: *Monthly Notices
of the Royal Astronomical Society* 497.4, pp. 5364–5382. ISSN: 0035-8711. DOI: 10.
1093/mnras/staa2339. eprint: https://academic.oup.com/mnras/article-
pdf/497/4/5364/33692150/staa2339.pdf. URL: https://doi.org/10.1093/
mnras/staa2339.

Brambilla, Gabriele et al. (May 2015). "Testing Dissipative Magnetosphere Model
Light Curves and Spectra with Fermi Pulsars". In: *ApJ* 804.2, 84, p. 84. DOI: 10.
1088/0004-637X/804/2/84. arXiv: 1503.00744 [astro-ph.HE].

Brandes, Ulrik (2001). "A faster algorithm for betweenness centrality". In: *The Jour-
nal of Mathematical Sociology* 25.2, pp. 163–177. DOI: 10.1080/0022250X.2001.
9990249. eprint: https://doi.org/10.1080/0022250X.2001.9990249. URL:
https://doi.org/10.1080/0022250X.2001.9990249.

Brecher, K. (Dec. 1999). "Origin of Neutron Stars". In: *American Astronomical Society
Meeting Abstracts*. Vol. 195. American Astronomical Society Meeting Abstracts,
130.05, p. 130.05.

Breiman, Leo (Jan. 2001). "Random Forests." In: *Machine Learning* 45, pp. 5–32. DOI:
10.1023/A:1010933404324.

Britt, C. T. et al. (2017). "Orbital Dynamics of Candidate Transitional Millisecond
Pulsar 3FGL J1544.6-1125: An Unusually Face-on System". In: *The Astrophysical
Journal* 849.1, p. 21. DOI: 10.3847/1538-4357/aa8e41. URL: https://dx.doi.
org/10.3847/1538-4357/aa8e41.

Brunt, C. M., M. H. Heyer, and M. M. Mac Low (Sept. 2009). "Turbulent driving
scales in molecular clouds". In: *A&A* 504.3, pp. 883–890. DOI: 10.1051/0004-
6361/200911797. arXiv: 0910.0398 [astro-ph.GA].

Buitinck, Lars et al. (2013). "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

Bujarrabal, V., J. Guibert, and C. Balkowski (Dec. 1981). "Multidimensional statistical analysis of normal galaxies". In: *A&A* 104.1, pp. 1–9.

Bullmore, Edward and Olaf Sporns (Mar. 2009). "Complex brain networks: Graph theoretical analysis of structural and functional systems". In: *Nature reviews. Neuroscience* 10, pp. 186–98. DOI: 10.1038/nrn2575.

Bult, Peter et al. (July 2020). "Timing the Pulsations of the Accreting Millisecond Pulsar SAX J1808.4-3658 during Its 2019 Outburst". In: *ApJ* 898.1, 38, p. 38. DOI: 10.3847/1538-4357/ab9827. arXiv: 1910.03062 [astro-ph.HE].

Burgio, G. Fiorella and Isaac Vidaña (Aug. 2020). "The Equation of State of Nuclear Matter: From Finite Nuclei to Neutron Stars". In: *Universe* 6.8, 119, p. 119. DOI: 10.3390/universe6080119. arXiv: 2007.04427 [nucl-th].

Burke-Spolaor, S. et al. (Jan. 2011). "Radio Bursts with Extragalactic Spectral Characteristics Show Terrestrial Origins". In: *ApJ* 727.1, 18, p. 18. DOI: 10.1088/0004-637X/727/1/18. arXiv: 1009.5392 [astro-ph.CO].

Caleb, M. et al. (Apr. 2019). "Are all fast radio bursts repeating sources?" In: *MNRAS* 484.4, pp. 5500–5508. DOI: 10.1093/mnras/stz386. arXiv: 1902.00272 [astro-ph.HE].

Cameron, A. G. (Nov. 1959). "Neutron Star Models." In: *ApJ* 130, p. 884. DOI: 10.1086/146780.

Cameron, A. G. W. (Jan. 1970). "Neutron Stars". In: *ARA&A* 8, p. 179. DOI: 10.1146/annurev.aa.08.090170.001143.

Camilo, F. et al. (Sept. 2015). "Parkes Radio Searches of Fermi Gamma-Ray Sources and Millisecond Pulsar Discoveries". In: *ApJ* 810.2, 85, p. 85. DOI: 10.1088/0004-637X/810/2/85. arXiv: 1507.04451 [astro-ph.HE].

Campana, R. et al. (Sept. 2013). "Minimal spanning tree algorithm for γ-ray source detection in sparse photon images: cluster parameters and selection strategies". In: *Ap&SS* 347.1, pp. 169–182. DOI: 10.1007/s10509-013-1488-0. arXiv: 1305.2025 [astro-ph.IM].

Cao, Gang and Xiongbang Yang (Feb. 2024). "The Energy-dependent Gamma-Ray Light Curves and Spectra of the Vela Pulsar in the Dissipative Magnetospheres". In: *ApJ* 962.2, 184, p. 184. DOI: 10.3847/1538-4357/ad1f67. arXiv: 2403.06179 [astro-ph.HE].

Casalicchio, Giuseppe, Christoph Molnar, and Bernd Bischl (Apr. 2018). "Visualizing the Feature Importance for Black Box Models". In: *arXiv e-prints*, arXiv:1804.06620, arXiv:1804.06620. DOI: 10.48550/arXiv.1804.06620. arXiv: 1804.06620 [stat.ML].

Cerutti, Benoît, Enzo Figueiredo, and Guillaume Dubus (Dec. 2024). "Synthetic pulsar lightcurves from global kinetic simulations and comparison with the Fermi-LAT catalog". In: *arXiv e-prints*, arXiv:2412.02307, arXiv:2412.02307. DOI: 10.48550/arXiv.2412.02307. arXiv: 2412.02307 [astro-ph.HE].

— (Mar. 2025). "Synthetic pulsar light curves from global kinetic simulations and comparison with the Fermi-LAT catalog". In: *A&A* 695, A93, A93. DOI: 10.1051/0004-6361/202451948. arXiv: 2412.02307 [astro-ph.HE].

Cerutti, Benoît et al. (Mar. 2015). "Particle acceleration in axisymmetric pulsar current sheets". In: *MNRAS* 448.1, pp. 606–619. DOI: 10.1093/mnras/stv042. arXiv: 1410.3757 [astro-ph.HE].

Chadwick, J. (1932). "Possible Existence of a Neutron". In: *Nature* 129.3252, pp. 312–312. DOI: 10.1038/129312a0. URL: https://doi.org/10.1038/129312a0.

Chakrabarty, Deepto and Edward H. Morgan (July 1998). "The two-hour orbit of a binary millisecond X-ray pulsar". In: *Nature* 394.6691, pp. 346–348. DOI: 10.1038/28561. arXiv: astro-ph/9804248 [astro-ph].

Chan, Franky, Ada Fu, and Clement Yu (June 2003). "Haar Wavelets for Efficient Similarity Search of Time-Series: With and Without Time Warping". In: *Knowledge and Data Engineering, IEEE Transactions on* 15, pp. 686–705. DOI: 10.1109/TKDE.2003.1198399.

Chandrasekhar, S. (July 1931). "The Maximum Mass of Ideal White Dwarfs". In: *ApJ* 74, p. 81. DOI: 10.1086/143324.

Chatterjee, S. et al. (Jan. 2017). "A direct localization of a fast radio burst and its host". In: *Nature* 541.7635, pp. 58–61. DOI: 10.1038/nature20797. arXiv: 1701.01098 [astro-ph.HE].

Chen, Alexander Y. and Andrei M. Beloborodov (Nov. 2014). "Electrodynamics of Axisymmetric Pulsar Magnetosphere with Electron-Positron Discharge: A Numerical Experiment". In: *ApJ* 795.1, L22, p. L22. DOI: 10.1088/2041-8205/795/1/L22. arXiv: 1406.7834 [astro-ph.HE].

Chen, Bo Han et al. (Jan. 2022). "Uncloaking hidden repeating fast radio bursts with unsupervised machine learning". In: *MNRAS* 509.1, pp. 1227–1236. DOI: 10.1093/mnras/stab2994. arXiv: 2110.09440 [astro-ph.HE].

Chen, Hai-Liang et al. (Sept. 2013). "Formation of Black Widows and Redbacks—Two Distinct Populations of Eclipsing Binary Millisecond Pulsars". In: *ApJ* 775.1, 27, p. 27. DOI: 10.1088/0004-637X/775/1/27. arXiv: 1308.4107 [astro-ph.SR].

Cheng, K. S., C. Ho, and M. Ruderman (Jan. 1986). "Energetic Radiation from Rapidly Spinning Pulsars. I. Outer Magnetosphere Gaps". In: *ApJ* 300, p. 500. DOI: 10.1086/163829.

Chikomana, S. and X. Hu (2023). "Genome Sequencing Using Graph Theory Approach". In: *Open Journal of Discrete Mathematics* 13, pp. 39–48. DOI: 10.4236/ojdm.2023.132004.

CHIME/FRB Collaboration et al. (Aug. 2018). "The CHIME Fast Radio Burst Project: System Overview". In: *ApJ* 863.1, 48, p. 48. DOI: 10.3847/1538-4357/aad188. arXiv: 1803.11235 [astro-ph.IM].

CHIME/FRB Collaboration et al. (Nov. 2019). "CHIME/FRB Discovery of Eight New Repeating Fast Radio Burst Sources". In: *ApJ* 885.1, L24, p. L24. DOI: 10.3847/2041-8213/ab4a80. arXiv: 1908.03507 [astro-ph.HE].

CHIME/FRB Collaboration et al. (Nov. 2020). "A bright millisecond-duration radio burst from a Galactic magnetar". In: *Nature* 587.7832, pp. 54–58. DOI: 10.1038/s41586-020-2863-y. arXiv: 2005.10324 [astro-ph.HE].

CHIME/FRB Collaboration et al. (Dec. 2021). "The First CHIME/FRB Fast Radio Burst Catalog". In: *ApJS* 257.2, 59, p. 59. DOI: 10.3847/1538-4365/ac33ab. arXiv: 2106.04352 [astro-ph.HE].

Chu, Selina et al. (n.d.). "Iterative Deepening Dynamic Time Warping for Time Series". In: *Proceedings of the 2002 SIAM International Conference on Data Mining (SDM)*, pp. 195–212. DOI: 10.1137/1.9781611972726.12. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611972726.12. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611972726.12.

Clark, C. J. et al. (Nov. 2016). "The Braking Index of a Radio-quiet Gamma-Ray Pulsar". In: *ApJ* 832.1, L15, p. L15. DOI: 10.3847/2041-8205/832/1/L15. arXiv: 1611.01292 [astro-ph.HE].

Coburn, W. et al. (Nov. 2002). "Magnetic Fields of Accreting X-Ray Pulsars with the Rossi X-Ray Timing Explorer". In: *ApJ* 580.1, pp. 394–412. DOI: 10.1086/343033. arXiv: astro-ph/0207325 [astro-ph].

Cocke, W. J., M. J. Disney, and D. J. Taylor (1969). "Discovery of Optical Signals from Pulsar NP 0532". In: *Nature* 221.5180, pp. 525–527. DOI: 10.1038/221525a0. URL: https://doi.org/10.1038/221525a0.

Comella, J. M. et al. (1969). "Crab Nebula Pulsar NP 0532". In: *Nature* 221.5179, pp. 453–454. DOI: 10.1038/221453a0. URL: https://doi.org/10.1038/221453a0.

Comon, Pierre (1994). "Independent component analysis, A new concept?" In: *Signal Processing* 36.3. Higher Order Statistics, pp. 287–314. ISSN: 0165-1684. DOI: https://doi.org/10.1016/0165-1684(94)90029-9. URL: https://www.sciencedirect.com/science/article/pii/0165168494900299.

Corcoran, Kyle A., Scott M. Ransom, and Ryan S. Lynch (Mar. 2023). "No MSP Counterparts Detected in GBT Searches of Spider Candidates 4FGL J0935.3+0901, 4FGL J1627.7+3219, and 4FGL J2212.4+0708". In: *Research Notes of the American Astronomical Society* 7.3, 41, p. 41. DOI: 10.3847/2515-5172/acc17d. arXiv: 2303.02200 [astro-ph.HE].

Cordes, James M. and Shami Chatterjee (Aug. 2019). "Fast Radio Bursts: An Extragalactic Enigma". In: *ARA&A* 57, pp. 417–465. DOI: 10.1146/annurev-astro-091918-104501. arXiv: 1906.05878 [astro-ph.HE].

Cormen, Thomas H. et al. (2009). *Introduction to Algorithms*. 3rd. MIT Press. ISBN: 9780262033848.

Corongiu, A. et al. (Mar. 2021). "Radio pulsations from the γ-ray millisecond pulsar PSR J2039-5617". In: *MNRAS* 502.1, pp. 935–952. DOI: 10.1093/mnras/staa3463. arXiv: 2007.14889 [astro-ph.HE].

Coti Zelati, Francesco et al. (Feb. 2018). "Systematic study of magnetar outbursts". In: *MNRAS* 474.1, pp. 961–1017. DOI: 10.1093/mnras/stx2679. arXiv: 1710.04671 [astro-ph.HE].

Coti Zelati, Francesco et al. (Feb. 2019). "Prolonged sub-luminous state of the new transitional pulsar candidate CXOU J110926.4-650224". In: *A&A* 622, A211, A211. DOI: 10.1051/0004-6361/201834835. arXiv: 1903.04526 [astro-ph.HE].

Crawford, Fronefield et al. (Oct. 2013). "PSR J1723-2837: An Eclipsing Binary Radio Millisecond Pulsar". In: *ApJ* 776.1, 20, p. 20. DOI: 10.1088/0004-637X/776/1/20. arXiv: 1308.4956 [astro-ph.SR].

Cui, B. and M. McLaughlin (2022). *RRATalog*. http://astro.phys.wvu.edu/rratalog/. [Online; accessed April 1, 2022].

Dai, Yiyang and Jinsong Zhao (Mar. 2011). "Fault Diagnosis of Batch Chemical Processes Using a Dynamic Time Warping (DTW)-Based Artificial Immune System". In: *Industrial & Engineering Chemistry Research* 50. DOI: 10.1021/ie101465b.

Danehkar, A. (Oct. 2024). "Relativistic reflection modeling in AGN and related variability from PCA: a brief review". In: *Frontiers in Astronomy and Space Sciences* 11, 1479301, p. 1479301. DOI: 10.3389/fspas.2024.1479301. arXiv: 2410.01852 [astro-ph.HE].

Danehkar, A., J. J. Drake, and G. J. M. Luna (Sept. 2024). "X-Ray Variability in the Symbiotic Binary RT Cru: Principal Component Analysis". In: *ApJ* 972.1, 109, p. 109. DOI: 10.3847/1538-4357/ad5cf6. arXiv: 2406.17161 [astro-ph.HE].

Das, Gautam et al. (1998). "Rule discovery from time series". In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*. KDD'98. New York, NY: AAAI Press, 16–22.

Daugherty, Joseph K. and Alice K. Harding (Feb. 1996). "Gamma-Ray Pulsars: Emission from Extended Polar CAP Cascades". In: *ApJ* 458, p. 278. DOI: 10.1086/176811. arXiv: astro-ph/9508155 [astro-ph].

de Amorim, Lucas B. V., George D. C. Cavalcanti, and Rafael M. O. Cruz (Dec. 2022). "The choice of scaling technique matters for classification performance". In: *arXiv e-prints*, arXiv:2212.12343, arXiv:2212.12343. DOI: 10.48550/arXiv.2212.12343. arXiv: 2212.12343 [cs.LG].

de Araujo, José C. N., Jaziel G. Coelho, and Cesar A. Costa (July 2016). "Gravitational wave emission by the high braking index pulsar PSR J1640-4631". In: *J. Cosmology Astropart. Phys.* 2016.7, 023, p. 023. DOI: 10.1088/1475-7516/2016/07/023. arXiv: 1603.05975 [astro-ph.HE].

de Martino, D. et al. (June 2010). "The intriguing nature of the high-energy gamma ray source XSS J12270-4859". In: *A&A* 515, A25, A25. DOI: 10.1051/0004-6361/200913802. arXiv: 1002.3740 [astro-ph.HE].

de Martino, D. et al. (Feb. 2013). "X-ray follow-ups of XSS J12270-4859: a low-mass X-ray binary with gamma-ray Fermi-LAT association". In: *A&A* 550, A89, A89. DOI: 10.1051/0004-6361/201220393. arXiv: 1212.1615 [astro-ph.HE].

de Martino, D. et al. (Nov. 2014). "Unveiling the redback nature of the low-mass X-ray binary XSS J1227.0-4859 through optical observations". In: *MNRAS* 444.4, pp. 3004–3014. DOI: 10.1093/mnras/stu1640. arXiv: 1408.6138 [astro-ph.HE].

de Martino, D. et al. (Sept. 2024). "X-ray and optical observations of the millisecond pulsar binary PSRJ1431-4715". In: *arXiv e-prints*, arXiv:2409.02075, arXiv:2409.02075. DOI: 10.48550/arXiv.2409.02075. arXiv: 2409.02075 [astro-ph.HE].

de Oña Wilhelmi, E. et al. (Mar. 2016). "SAX J1808.4-3658, an accreting millisecond pulsar shining in gamma rays?" In: *MNRAS* 456.3, pp. 2647–2653. DOI: 10.1093/mnras/stv2695. arXiv: 1511.05032 [astro-ph.HE].

Debregeas, Anne and Georges Hebrail (1998). "Interactive interpretation of Kohonen maps applied to curves". In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*. KDD'98. New York, NY: AAAI Press, 179–183.

Deeming, Terence J. (May 1964). "Stellar Spectral Classification: I. Application of Component Analysis". In: *Monthly Notices of the Royal Astronomical Society* 127.6, pp. 493–516. ISSN: 0035-8711. DOI: 10.1093/mnras/127.6.493. eprint: https://academic.oup.com/mnras/article-pdf/127/6/493/8072401/mnras127-0493.pdf. URL: https://doi.org/10.1093/mnras/127.6.493.

Deller, A. T. et al. (Sept. 2012). "A Parallax Distance and Mass Estimate for the Transitional Millisecond Pulsar System J1023+0038". In: *ApJ* 756.2, L25, p. L25. DOI: 10.1088/2041-8205/756/2/L25. arXiv: 1207.5670 [astro-ph.SR].

Deneva, J. S. et al. (Mar. 2021). "Timing of Eight Binary Millisecond Pulsars Found with Arecibo in Fermi-LAT Unidentified Sources". In: *ApJ* 909.1, 6, p. 6. DOI: 10.3847/1538-4357/abd7a1. arXiv: 2012.15185 [astro-ph.HE].

Deutsch, Arnim J. (Jan. 1955). "The electromagnetic field of an idealized star in rigid rotation in vacuo". In: *Annales d'Astrophysique* 18, p. 1.

di Salvo, T. et al. (Oct. 2008). "Orbital evolution of an accreting millisecond pulsar: witnessing the banquet of a hidden black widow?" In: *MNRAS* 389.4, pp. 1851–1857. DOI: 10.1111/j.1365-2966.2008.13709.x. arXiv: 0708.0498 [astro-ph].

Di Salvo, Tiziana et al. (Nov. 2023). "Low Magnetic-Field Neutron Stars in X-ray Binaries". In: *arXiv e-prints*, arXiv:2311.12516, arXiv:2311.12516. DOI: 10.48550/arXiv.2311.12516. arXiv: 2311.12516 [astro-ph.HE].

Ding, Hui et al. (Aug. 2008). "Querying and mining of time series data: experimental comparison of representations and distance measures". In: *Proc. VLDB Endow.* 1.2, 1542–1552. ISSN: 2150-8097. DOI: 10.14778/1454159.1454226. URL: https://doi.org/10.14778/1454159.1454226.

Dokmanic, Ivan et al. (Nov. 2015). "Euclidean Distance Matrices: Essential theory, algorithms, and applications". In: *IEEE Signal Processing Magazine* 32.6, pp. 12–30. DOI: 10.1109/MSP.2015.2398954. arXiv: 1502.07541 [cs.OH].

Douglas, Andrew et al. (Mar. 2022). "Two New Black Widow Millisecond Pulsars in M28". In: *ApJ* 927.1, 126, p. 126. DOI: 10.3847/1538-4357/ac4744. arXiv: 2201.11238 [astro-ph.HE].

Duyvendak, J. J. L. (Apr. 1942). "Further Data Bearing on the Identification of the Crab Nebula with the Supernova of 1054 A.D. Part I. The Ancient Oriental Chronicles". In: *PASP* 54.318, pp. 91–94. DOI: 10.1086/125409.

Eckert, D. et al. (Mar. 2013). "IGR J18245-2452: a new hard X-ray transient discovered by INTEGRAL". In: *The Astronomer's Telegram* 4925, p. 1.

Efstathiou, George and S. Michael Fall (Feb. 1984). "Multivariate analysis of elliptical galaxies". In: *Monthly Notices of the Royal Astronomical Society* 206.3, pp. 453–464. ISSN: 0035-8711. DOI: 10.1093/mnras/206.3.453. eprint: https://academic.oup.com/mnras/article-pdf/206/3/453/2831195/mnras206-0453.pdf. URL: https://doi.org/10.1093/mnras/206.3.453.

Eichler, David and Amir Levinson (Dec. 1988). "On Black Widow Evolutionary Scenarios for Binary Neutron Stars". In: *ApJ* 335, p. L67. DOI: 10.1086/185341.

Enoto, Teruaki, Shota Kisaka, and Shinpei Shibata (Oct. 2019). "Observational diversity of magnetized neutron stars". In: *Reports on Progress in Physics* 82.10, 106901, p. 106901. DOI: 10.1088/1361-6633/ab3def.

Erickson, J. (2019). *Algorithms*. Jeff Erickson, Lecture Notes University of Illinois at Urbana-Champaign.

Espinoza, C. M., A. G. Lyne, and B. W. Stappers (Apr. 2017). "New long-term braking index measurements for glitching pulsars using a glitch-template method". In: *MNRAS* 466.1, pp. 147–162. DOI: 10.1093/mnras/stw3081. arXiv: 1611.08314 [astro-ph.HE].

Euler, Leonhard (1736). "Solutio problematis ad geometriam situs pertinentis". In: *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 8, pp. 128–140.

Event Horizon Telescope Collaboration et al. (Apr. 2019). "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole". In: *ApJ* 875.1, L1, p. L1. DOI: 10.3847/2041-8213/ab0ec7. arXiv: 1906.11238 [astro-ph.GA].

Everitt, Brian et al. (2011). *Cluster analysis*. English. 5th. Wiley. ISBN: 978-0-470-74991-3.

Faber, S. M. (Feb. 1973). "Variations in Spectral-Energy Distributions and Absorption-Line Strengths among Elliptical Galaxies". In: *ApJ* 179, pp. 731–754. DOI: 10.1086/151912.

Faloutsos, Christos, M. Ranganathan, and Yannis Manolopoulos (1994). "Fast subsequence matching in time-series databases". In: *Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data*. SIGMOD '94. Minneapolis, Minnesota, USA: Association for Computing Machinery, 419–429. ISBN: 0897916395. DOI: 10.1145/191839.191925. URL: https://doi.org/10.1145/191839.191925.

Feng, Yi et al. (Mar. 2022). "Frequency-dependent polarization of repeating fast radio bursts—implications for their origin". In: *Science* 375.6586, pp. 1266–1270. DOI: 10.1126/science.abl7759. arXiv: 2202.09601 [astro-ph.HE].

Fermi-LAT Collaboration (2021). *Public List of LAT Detected Gamma-Ray Pulsars*. https://confluence.slac.stanford.edu/display/GLAMCOG/Public+List+of+LAT-Detected+Gamma-Ray+Pulsars. [Online; accessed April 1, 2022].

Ferrigno, C. et al. (July 2014). "Hiccup accretion in the swinging pulsar IGR J18245-2452". In: *A&A* 567, A77, A77. DOI: 10.1051/0004-6361/201322904. arXiv: 1310.7784 [astro-ph.HE].

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici (Jan. 2018). "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously". In: *arXiv e-prints*, arXiv:1801.01489, arXiv:1801.01489. DOI: 10.48550/arXiv.1801.01489. arXiv: 1801.01489 [stat.ME].

Francis, Paul J. and Beverly J. Wills (Jan. 1999). "Introduction to Principal Components Analysis". In: *Quasars and Cosmology*. Ed. by Gary Ferland and Jack Baldwin. Vol. 162. Astronomical Society of the Pacific Conference Series, p. 363. DOI: 10.48550/arXiv.astro-ph/9905079. arXiv: astro-ph/9905079 [astro-ph].

Francis, Paul J. et al. (Oct. 1992). "An Objective Classification Scheme for QSO Spectra". In: *ApJ* 398, p. 476. DOI: 10.1086/171870.

Franses, Philip and Thomas Wiemann (June 2020). "Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping". In: *Computational Economics* 56. DOI: 10.1007/s10614-020-09986-0.

Freeman, Linton C. (1977). "A Set of Measures of Centrality Based on Betweenness". In: *Sociometry* 40.1, pp. 35–41. ISSN: 00380431. URL: http://www.jstor.org/stable/3033543 (visited on 09/14/2022).

Freire, P. C. C. et al. (Oct. 2017). "Long-term observations of the pulsars in 47 Tucanae - II. Proper motions, accelerations and jerks". In: *MNRAS* 471.1, pp. 857–876. DOI: 10.1093/mnras/stx1533. arXiv: 1706.04908 [astro-ph.HE].

Freire, Paulo C. C. et al. (2005). "The Millisecond Pulsars in NGC 6760". In: *The Astrophysical Journal* 621.2, p. 959. DOI: 10.1086/427748. URL: https://dx.doi.org/10.1086/427748.

Fruchter, A. S., D. R. Stinebring, and J. H. Taylor (1988). "A millisecond pulsar in an eclipsing binary". In: *Nature* 333.6170, pp. 237–239. DOI: 10.1038/333237a0. URL: https://doi.org/10.1038/333237a0.

Fryer, Chris L. (1999). "Mass Limits For Black Hole Formation". In: *The Astrophysical Journal* 522.1, p. 413. DOI: 10.1086/307647. URL: https://dx.doi.org/10.1086/307647.

Gallant, D, L C Gallo, and M L Parker (July 2018). "X-ray spectral variability of blazars using principal component analysis". In: *Monthly Notices of the Royal Astronomical Society* 480.2, pp. 1999–2010. ISSN: 0035-8711. DOI: 10.1093/mnras/sty1987. eprint: https://academic.oup.com/mnras/article-pdf/480/2/1999/25440546/sty1987.pdf. URL: https://doi.org/10.1093/mnras/sty1987.

Gallo, L. C. et al. (Jan. 2015). "Suzaku observations of Mrk 335: confronting partial covering and relativistic reflection". In: *MNRAS* 446.1, pp. 633–650. DOI: 10.1093/mnras/stu2108. arXiv: 1410.2330 [astro-ph.HE].

García, C. R. and Diego F. Torres (Mar. 2023). "Quantitative determination of minimum spanning tree structures: using the pulsar tree for analysing the appearance of new classes of pulsars". In: *MNRAS* 520.1, pp. 599–610. DOI: 10.1093/mnras/stad183. arXiv: 2301.05408 [astro-ph.HE].

— (Apr. 2025). "Quantitative Exploration of the Similarity of Gamma-Ray Pulsar Light Curves". In: *ApJ* 982.2, L51, p. L51. DOI: 10.3847/2041-8213/adbf0f. arXiv: 2503.02750 [astro-ph.HE].

García, C. R., Diego F. Torres, and Alessandro Patruno (Sept. 2022). "Visualizing the pulsar population using graph theory". In: *MNRAS* 515.3, pp. 3883–3897. DOI: 10.1093/mnras/stac1997. arXiv: 2207.06311 [astro-ph.HE].

García, C. R. et al. (Dec. 2024a). "Millisecond pulsars phenomenology under the light of graph theory". In: *A&A* 692, A187, A187. DOI: 10.1051/0004-6361/202450758. arXiv: 2410.13650 [astro-ph.HE].

García, C. R. et al. (Dec. 2024b). "Separating Repeating Fast Radio Bursts Using the Minimum Spanning Tree as an Unsupervised Methodology". In: *ApJ* 977.2, 273, p. 273. DOI: 10.3847/1538-4357/ad9020. arXiv: 2411.02216 [astro-ph.HE].

Gavriil, F. P. et al. (Mar. 2008). "Magnetar-Like Emission from the Young Pulsar in Kes 75". In: *Science* 319.5871, p. 1802. DOI: 10.1126/science.1153465. arXiv: 0802.1704 [astro-ph].

Geler, Zoltan et al. (July 2019). "Dynamic Time Warping: Itakura vs Sakoe-Chiba". In: pp. 1–6. DOI: 10.1109/INISTA.2019.8778300.

Géron, Aurélien (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd. O'Reilly Media.

Ghosh, S., G. Kumar, A. Basu, et al. (2015). "Graph Theoretic Network Analysis Reveals Protein Pathways Underlying Cell Death Following Neurotropic Viral Infection". In: *Scientific Reports* 5, p. 14438. DOI: 10.1038/srep14438. URL: https://doi.org/10.1038/srep14438.

Giacconi, Riccardo et al. (Dec. 1962). "Evidence for x Rays From Sources Outside the Solar System". In: *Phys. Rev. Lett.* 9.11, pp. 439–443. DOI: 10.1103/PhysRevLett.9.439.

Glampedakis, Kostas and Leonardo Gualtieri (Jan. 2018). "Gravitational Waves from Single Neutron Stars: An Advanced Detector Era Survey". In: *Astrophysics and Space Science Library*. Ed. by Luciano Rezzolla et al. Vol. 457. Astrophysics and Space Science Library, p. 673. DOI: 10.1007/978-3-319-97616-7_12. arXiv: 1709.07049 [astro-ph.HE].

Godin, C. and P. Lockwood (1989). "DTW schemes for continuous speech recognition: a unified view". In: *Computer Speech & Language* 3.2, pp. 169–198. ISSN: 0885-2308. DOI: https://doi.org/10.1016/0885-2308(89)90028-4. URL: https://www.sciencedirect.com/science/article/pii/0885230889900284.

Gogolou, Anna et al. (2019). "Comparing Similarity Perception in Time Series Visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1, pp. 523–533. DOI: 10.1109/TVCG.2018.2865077.

Gold, T. (1968). "Rotating Neutron Stars as the Origin of the Pulsating Radio Sources". In: *Nature* 218.5143, pp. 731–732. DOI: 10.1038/218731a0. URL: https://doi.org/10.1038/218731a0.

Goldreich, Peter and William H. Julian (Aug. 1969). "Pulsar Electrodynamics". In: *ApJ* 157, p. 869. DOI: 10.1086/150119.

Goodrich, Michael T. and Roberto Tamassia (2014). *Algorithm Design and Applications*. 1st. Wiley Publishing. ISBN: 1118335910.

Gralla, Samuel E., Alexandru Lupsasca, and Alexander Philippov (Dec. 2017). "Inclined Pulsar Magnetospheres in General Relativity: Polar Caps for the Dipole, Quadrudipole, and Beyond". In: *ApJ* 851.2, 137, p. 137. DOI: 10.3847/1538-4357/aa978d. arXiv: 1704.05062 [astro-ph.HE].

Granovetter, Mark (1973). "The Strength of Weak Ties". In: *American Journal of Sociology* 78.6, pp. 1360–1380. DOI: 10.1086/225469. URL: https://doi.org/10.1086/225469.

H. E. S. S. Collaboration et al. (Apr. 2018). "The population of TeV pulsar wind nebulae in the H.E.S.S. Galactic Plane Survey". In: *A&A* 612, A2, A2. DOI: 10.1051/0004-6361/201629377. arXiv: 1702.08280 [astro-ph.HE].

Haensel, P., A. Y. Potekhin, and D. G. Yakovlev (2007). *Neutron Stars 1 : Equation of State and Structure*. Vol. 326.

Hagberg, Aric A., Daniel A. Schult, and Pieter J. Swart (2008). "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th*

Python in Science Conference. Ed. by Gaël Varoquaux, Travis Vaught, and Jarrod Millman. Pasadena, CA USA, pp. 11 –15. URL: http://conference.scipy.org/proceedings/SciPy2008/paper_2/.

Hamada, T. and E. E. Salpeter (Nov. 1961). "Models for Zero-Temperature Stars." In: *ApJ* 134, p. 683. DOI: 10.1086/147195.

Han, J., M. Kamber, and J. Pei (2012). *Data Mining: Concepts and Techniques*. 3rd. Morgan Kaufmann Publishers.

Harari, Diego, Silvia Mollerach, and Esteban Roulet (July 2006). "Detecting filaments in the ultra-high energy cosmic ray distribution". In: *Astroparticle Physics* 25.6, pp. 412–418. DOI: 10.1016/j.astropartphys.2006.04.003. arXiv: astro-ph/0602153 [astro-ph].

Harding, Alice K. and Constantinos Kalapotharakos (May 2017). "Multiwavelength Polarization of Rotation-powered Pulsars". In: *ApJ* 840.2, 73, p. 73. DOI: 10.3847/1538-4357/aa6ead. arXiv: 1704.06183 [astro-ph.HE].

Harris, Charles R. et al. (2020). "Array programming with NumPy". In: *Nature* 585, 357–362. DOI: 10.1038/s41586-020-2649-2.

Harrison, B. K. et al. (1965). *Gravitation Theory and Gravitational Collapse*.

Hartigan, J. A. (1981). "Consistency of Single Linkage for High-Density Clusters". In: *Journal of the American Statistical Association* 76.374, pp. 388–394. DOI: 10.1080/01621459.1981.10477658. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1981.10477658. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477658.

Heger, A. et al. (2003). "How Massive Single Stars End Their Life". In: *The Astrophysical Journal* 591.1, p. 288. DOI: 10.1086/375341. URL: https://dx.doi.org/10.1086/375341.

Henzinger, Monika R et al. (1997). "Faster Shortest-Path Algorithms for Planar Graphs". In: *Journal of Computer and System Sciences* 55.1, pp. 3–23. ISSN: 0022-0000. DOI: https://doi.org/10.1006/jcss.1997.1493. URL: https://www.sciencedirect.com/science/article/pii/S0022000097914938.

Herault, Jeanny and Bernard Ans (July 1984). "Réseaux de neurones a synapses modifiables: Décodage de messages sensoriels composites par un apprentissage non supervisé et permanent". In: *Comptes rendus des séances de l'Académie des sciences. Série III, Sciences de la vie* 299, pp. 525–528.

Herault, Jeanny, Christian Jutten, and Bernard Ans (Jan. 1985). "Detection de grandeurs primitives dans un message composite par une architeture de calcul neuromimetique en apprentissage non supervise". In: *10° Colloque sur le traitement du signal et des images, 1985 ; p. 1017-1022*.

Hessels, J. et al. (Apr. 2015). "Pulsars in Globular Clusters with the SKA". In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 47, p. 47. DOI: 10.22323/1.215.0047. arXiv: 1501.00086 [astro-ph.HE].

Hewish, A. et al. (1968). "Observation of a Rapidly Pulsating Radio Source". In: *Nature* 217.5130, pp. 709–713. DOI: 10.1038/217709a0. URL: https://doi.org/10.1038/217709a0.

Hewish, A. et al. (Nov. 1969). "Observation of a Rapidly Pulsating Radio Source (Reprinted from Nature, February 24, 1968)". In: *Nature* 224.5218, p. 472. DOI: 10.1038/224472b0.

Heyer, Mark H. and F. Peter Schloerb (1997). "Application of Principal Component Analysis to Large-Scale Spectral Line Imaging Studies of the Interstellar Medium". In: *The Astrophysical Journal* 475.1, p. 173. DOI: 10.1086/303514. URL: https://dx.doi.org/10.1086/303514.

Hill, A. B. et al. (July 2011). "The bright unidentifiedγ-ray source 1FGL J1227.9-4852: can it be associated with a low-mass X-ray binary?" In: *MNRAS* 415.1, pp. 235–243. DOI: `10.1111/j.1365-2966.2011.18692.x`. arXiv: `1103.2637 [astro-ph.HE]`.

Hills, J. G. (Apr. 1983). "The effects of sudden mass loss and a random kick velocity produced in a supernova explosion on the dynamics of a binary star of arbitrary orbital eccentricity. Applications to X-ray binaries and to the binarypulsars." In: *ApJ* 267, pp. 322–333. DOI: `10.1086/160871`.

Horvath, J. E. (July 2012). "The nature of the companion of PSR J1719-1438: a white dwarf or an exotic object?" In: *Research in Astronomy and Astrophysics* 12.7, pp. 813–816. DOI: `10.1088/1674-4527/12/7/009`. arXiv: `1205.1410 [astro-ph.HE]`.

Hotelling, Harold (1933). "Analysis of a complex of statistical variables into principal components." In: *Journal of Educational Psychology* 24, pp. 498–520. URL: `https://api.semanticscholar.org/CorpusID:144828484`.

Hoyle, F. et al. (Apr. 1964). "On Relativistic Astrophysics." In: *ApJ* 139, p. 909. DOI: `10.1086/147825`.

Hui, Chung-Yue (June 2014). "Spider Invasion Across the Galaxy". In: *Journal of Astronomy and Space Sciences* 31, pp. 101–120. DOI: `10.5140/JASS.2014.31.2.101`.

Hulse, R. A. and J. H. Taylor (Jan. 1975). "Discovery of a pulsar in a binary system." In: *ApJ* 195, pp. L51–L53. DOI: `10.1086/181708`.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: `10.1109/MCSE.2007.55`.

Íñiguez-Pascual, Daniel, Diego F. Torres, and Daniele Viganò (May 2024). "Synchro-curvature description of γ-ray light curves and spectra of pulsars: global properties". In: *MNRAS* 530.2, pp. 1550–1562. DOI: `10.1093/mnras/stae933`. arXiv: `2404.01926 [astro-ph.HE]`.

— (Apr. 2025). "Synchro-curvature description of γ-ray light curves and spectra of pulsars: concurrent fitting". In: *arXiv e-prints*, arXiv:2504.01892, arXiv:2504.01892. DOI: `10.48550/arXiv.2504.01892`. arXiv: `2504.01892 [astro-ph.HE]`.

Illiano, Giulia et al. (Jan. 2023). "Timing Analysis of the 2022 Outburst of the Accreting Millisecond X-Ray Pulsar SAX J1808.4-3658: Hints of an Orbital Shrinking". In: *ApJ* 942.2, L40, p. L40. DOI: `10.3847/2041-8213/acad81`. arXiv: `2212.09778 [astro-ph.HE]`.

Itakura, F. (1975). "Minimum prediction residual principle applied to speech recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23.1, pp. 67–72. DOI: `10.1109/TASSP.1975.1162641`.

Ivezić, Željko et al. (2014). *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*. DOI: `10.1515/9781400848911`.

Izenman, Alan Julian (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. 1st ed. Springer Publishing Company, Incorporated. ISBN: 0387781889.

Jiang, Long et al. (Jan. 2013). "Characteristic Age and True Age of Pulsars". In: *International Journal of Modern Physics Conference Series*. Vol. 23. International Journal of Modern Physics Conference Series, pp. 95–98. DOI: `10.1142/S2010194513011124`.

Jolliffe, Ian (2002). *Principal component analysis*. New York: Springer Verlag.

Jolliffe, Ian T. and Jorge Cadima (2016). "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065, p. 20150202. DOI: `10.1098/rsta.2015.0202`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.`

1098/rsta.2015.0202. URL: https://royalsocietypublishing.org/doi/abs/
10.1098/rsta.2015.0202.

Kadous, Mohammed Waleed (1999). "Learning Comprehensible Descriptions of Mul-
tivariate Time Series". In: *Proceedings of the Sixteenth International Conference on
Machine Learning*. ICML '99. San Francisco, CA, USA: Morgan Kaufmann Pub-
lishers Inc., 454–463. ISBN: 1558606122.

Kalapotharakos, C. and I. Contopoulos (Mar. 2009). "Three-dimensional numerical
simulations of the pulsar magnetosphere: preliminary results". In: *A&A* 496.2,
pp. 495–502. DOI: 10.1051/0004-6361:200810281. arXiv: 0811.2863 [astro-ph].

Kalapotharakos, Constantinos, Alice K. Harding, and Demosthenes Kazanas (Oct.
2014). "Gamma-Ray Emission in Dissipative Pulsar Magnetospheres: From The-
ory to Fermi Observations". In: *ApJ* 793.2, 97, p. 97. DOI: 10.1088/0004-637X/
793/2/97. arXiv: 1310.3545 [astro-ph.HE].

Kalapotharakos, Constantinos et al. (July 2012). "Gamma-Ray Light Curves from
Pulsar Magnetospheres with Finite Conductivity". In: *ApJ* 754.1, L1, p. L1. DOI:
10.1088/2041-8205/754/1/L1. arXiv: 1205.5769 [astro-ph.HE].

Karhunen, K. and Ivan Selin (1960). *On Linear Methods in Probability Theory*. Santa
Monica, CA: RAND Corporation.

Karttunen, Hannu et al. (2007). *Fundamental Astronomy*. 5th ed. Springer.

Katz, J. I. (Aug. 2016). "How Soft Gamma Repeaters Might Make Fast Radio Bursts".
In: *ApJ* 826.2, 226, p. 226. DOI: 10.3847/0004-637X/826/2/226. arXiv: 1512.04503
[astro-ph.HE].

— (Nov. 2018). "Fast radio bursts". In: *Progress in Particle and Nuclear Physics* 103,
pp. 1–18. DOI: 10.1016/j.ppnp.2018.07.001. arXiv: 1804.09092 [astro-ph.HE].

Kennedy, M R et al. (Feb. 2022). "Measuring the mass of the black widow PSR J1555-
2908". In: *Monthly Notices of the Royal Astronomical Society* 512.2, 3001–3014. ISSN:
1365-2966. DOI: 10.1093/mnras/stac379. URL: http://dx.doi.org/10.1093/
mnras/stac379.

Keogh, Eamonn and Chotirat Ratanamahatana (Jan. 2005). "Exact indexing of dy-
namic time warping". In: *Knowledge and Information Systems* 7, pp. 358–386. DOI:
10.1007/s10115-004-0154-9.

Keogh, Eamonn et al. (June 2001a). "Locally Adaptive Dimensionality Reduction for
Indexing Large Time Series Databases." In: vol. 30, pp. 151–162. DOI: 10.1145/
376284.375680.

Keogh, Eamonn J. and Michael J. Pazzani (2000). "Scaling up dynamic time warping
for datamining applications". In: *Proceedings of the Sixth ACM SIGKDD Interna-
tional Conference on Knowledge Discovery and Data Mining*. KDD '00. Boston, Mas-
sachusetts, USA: Association for Computing Machinery, 285–289. ISBN: 1581132336.
DOI: 10.1145/347090.347153. URL: https://doi.org/10.1145/347090.347153.

Keogh, Eamonn J. et al. (2001b). "Dimensionality Reduction for Fast Similarity Search
in Large Time Series Databases". In: *Knowledge and Information Systems* 3, pp. 263–
286. URL: https://api.semanticscholar.org/CorpusID:15462747.

Kerr, M. et al. (Mar. 2012). "Five New Millisecond Pulsars from a Radio Survey of
14 Unidentified Fermi-LAT Gamma-Ray Sources". In: *ApJ* 748.1, L2, p. L2. DOI:
10.1088/2041-8205/748/1/L2. arXiv: 1201.5160 [astro-ph.HE].

Kiani majd, Adell et al. (July 2017). "Comparison of different methods of measuring
similarity in physiologic time series". In: *IFAC-PapersOnLine* 50, pp. 11005–11010.
DOI: 10.1016/j.ifacol.2017.08.2479.

King, A. R. et al. (Apr. 2005). "The population of black widow pulsars". In: *MNRAS*
358.4, pp. 1501–1504. DOI: 10.1111/j.1365-2966.2005.08877.x. arXiv: astro-
ph/0502511 [astro-ph].

Kiziltan, Bülent and Stephen E. Thorsett (Apr. 2010). "MILLISECOND PULSAR AGES: IMPLICATIONS OF BINARY EVOLUTION AND A MAXIMUM SPIN LIMIT". In: *The Astrophysical Journal* 715.1, 335–341. ISSN: 1538-4357. DOI: 10.1088/0004-637x/715/1/335. URL: http://dx.doi.org/10.1088/0004-637X/715/1/335.

Kleinberg, J. and E. Tardos (2005). *Algorithm design*. Addison Wesley.

Koljonen, K. I. I. (Jan. 2015). "Unsupervised spectral decomposition of X-ray binaries with application to GX 339-4". In: *Monthly Notices of the Royal Astronomical Society* 447.4, pp. 2981–2991. ISSN: 0035-8711. DOI: 10.1093/mnras/stu2663. eprint: https://academic.oup.com/mnras/article-pdf/447/4/2981/5699737/stu2663.pdf. URL: https://doi.org/10.1093/mnras/stu2663.

Koljonen, K. I. I. et al. (Dec. 2012). "2006 May–July major radio flare episodes in Cygnus X-3: spectrotiming analysis of the X-ray data". In: *Monthly Notices of the Royal Astronomical Society* 429.2, pp. 1173–1188. ISSN: 0035-8711. DOI: 10.1093/mnras/sts404. eprint: https://academic.oup.com/mnras/article-pdf/429/2/1173/18454238/sts404.pdf. URL: https://doi.org/10.1093/mnras/sts404.

Koljonen, Karri I. I. and Manuel Linares (Nov. 2023). "A Gaia view of the optical and X-ray luminosities of compact binary millisecond pulsars". In: *MNRAS* 525.3, pp. 3963–3985. DOI: 10.1093/mnras/stad2485. arXiv: 2308.07377 [astro-ph.HE].

Komissarov, S. S. (Mar. 2006). "Simulations of the axisymmetric magnetospheres of neutron stars". In: *MNRAS* 367.1, pp. 19–31. DOI: 10.1111/j.1365-2966.2005.09932.x. arXiv: astro-ph/0510310 [astro-ph].

Konacki, Maciej and Alex Wolszczan (July 2003). "Masses and Orbital Inclinations of Planets in the PSR B1257+12 System". In: *ApJ* 591.2, pp. L147–L150. DOI: 10.1086/377093. arXiv: astro-ph/0305536 [astro-ph].

Kong, A. K. H. et al. (Apr. 2017). "A NuSTAR Observation of the Gamma-Ray Emitting Millisecond Pulsar PSR J1723-2837". In: *ApJ* 839.2, 130, p. 130. DOI: 10.3847/1538-4357/aa6aa2. arXiv: 1703.10164 [astro-ph.HE].

Korn, Flip, H. V. Jagadish, and Christos Faloutsos (1997). "Efficiently supporting ad hoc queries in large datasets of time sequences". In: *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*. SIGMOD '97. Tucson, Arizona, USA: Association for Computing Machinery, 289–300. ISBN: 0897919114. DOI: 10.1145/253260.253332. URL: https://doi.org/10.1145/253260.253332.

Kovacs-Vajna, Z.M. (2000). "A fingerprint verification system based on triangular matching and dynamic time warping". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11, pp. 1266–1276. DOI: 10.1109/34.888711.

Kruskal, J. B. (Feb. 1956). "On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem". In: *Proc. Amer. Math. Soc.* 7, pp. 48–50. DOI: 10.1090/S0002-9939-1956-0078686-7.

Kruskal, JB and Mark Liberman (Jan. 1983). "The symmetric time-warping problem: From continuous to discrete". In: *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*.

Kumar, Pawan, Wenbin Lu, and Mukul Bhattacharya (July 2017). "Fast radio burst source properties and curvature radiation model". In: *MNRAS* 468.3, pp. 2726–2739. DOI: 10.1093/mnras/stx665. arXiv: 1703.06139 [astro-ph.HE].

Kumar, Pravir et al. (Dec. 2019). "Faint Repetitions from a Bright Fast Radio Burst Source". In: *ApJ* 887.2, L30, p. L30. DOI: 10.3847/2041-8213/ab5b08. arXiv: 1908.10026 [astro-ph.HE].

Landau, L. (1938). "Origin of Stellar Energy". In: *Nature* 141.3564, pp. 333–334. DOI: 10.1038/141333b0. URL: https://doi.org/10.1038/141333b0.

Landau, Lev Davidovich (1979). "66. On the Theory of Stars". In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Ed. by Kenneth R. Lang and Owen

Gingerich. Cambridge, MA and London, England: Harvard University Press, pp. 456–459. ISBN: 9780674366688. DOI: `doi:10.4159/harvard.9780674366688.` `c72`. URL: `https://doi.org/10.4159/harvard.9780674366688.c72`.

Large, M. I., A. E. Vaughan, and B. Y. Mills (1968). "A Pulsar Supernova Association?" In: *Nature* 220.5165, pp. 340–341. DOI: `10.1038/220340a0`. URL: `https://doi.org/10.1038/220340a0`.

Lasky, Paul D. (2015). "Gravitational Waves from Neutron Stars: A Review". In: *Publications of the Astronomical Society of Australia* 32, e034. DOI: `10.1017/pasa.2015.35`.

Lee, Daniel and H. Seung (Nov. 1999). "Learning the Parts of Objects by Non-Negative Matrix Factorization". In: *Nature* 401, pp. 788–91. DOI: `10.1038/44565`.

Lee, Daniel D. and H. Sebastian Seung (2000). "Algorithms for non-negative matrix factorization". In: *Proceedings of the 14th International Conference on Neural Information Processing Systems*. NIPS'00. Denver, CO: MIT Press, 535–541.

Lee, Jongsu et al. (Sept. 2018). "X-Ray Census of Millisecond Pulsars in the Galactic Field". In: *ApJ* 864.1, 23, p. 23. DOI: `10.3847/1538-4357/aad284`. arXiv: `1807.06505 [astro-ph.HE]`.

Lehmann, Erich L (2012). "The power of rank tests". In: *Selected Works of EL Lehmann*. Springer, pp. 373–390.

Levin, Yuri, Andrei M. Beloborodov, and Ashley Bransgrove (June 2020). "Precessing Flaring Magnetar as a Source of Repeating FRB 180916.J0158+65". In: *ApJ* 895.2, L30, p. L30. DOI: `10.3847/2041-8213/ab8c4c`. arXiv: `2002.04595 [astro-ph.HE]`.

Lewis, Evan F. et al. (Oct. 2023). "Discovery and Timing of Millisecond Pulsars with the Arecibo 327 MHz Drift-scan Survey". In: *ApJ* 956.2, 132, p. 132. DOI: `10.3847/1538-4357/acf99d`. arXiv: `2306.10156 [astro-ph.HE]`.

Li, C. K. et al. (Apr. 2021). "HXMT identification of a non-thermal X-ray burst from SGR J1935+2154 and with FRB 200428". In: *Nature Astronomy* 5, pp. 378–384. DOI: `10.1038/s41550-021-01302-6`. arXiv: `2005.11071 [astro-ph.HE]`.

Li, Hailin (2014). "On-line and dynamic time warping for time series data mining". In: *International Journal of Machine Learning and Cybernetics* 6, pp. 145 –153. URL: `https://api.semanticscholar.org/CorpusID:255979595`.

Li, Jian et al. (Jan. 2017). "Gamma-ray Upper Limits on Magnetars with Six Years of Fermi-LAT Observations". In: *ApJ* 835.1, 30, p. 30. DOI: `10.3847/1538-4357/835/1/30`. arXiv: `1607.03778 [astro-ph.HE]`.

Li, Miao, Jules P. Halpern, and John R. Thorstensen (Nov. 2014). "Optical Counterparts of Two Fermi Millisecond Pulsars: PSR J1301+0833 and PSR J1628-3205". In: *ApJ* 795.2, 115, p. 115. DOI: `10.1088/0004-637X/795/2/115`. arXiv: `1409.3877 [astro-ph.HE]`.

Li, Xue, Xinlei Wang, and Guanghua Xiao (Aug. 2017). "A comparative study of rank aggregation methods for partial and top ranked lists in genomic applications". In: *Briefings in Bioinformatics* 20.1, pp. 178–189. ISSN: 1477-4054. DOI: `10.1093/bib/bbx101`. eprint: `https://academic.oup.com/bib/article-pdf/20/1/178/27689776/bbx101.pdf`. URL: `https://doi.org/10.1093/bib/bbx101`.

Lin, Hsiu-Hsien et al. (July 2023a). "Constraints on the Intergalactic and Local Dispersion Measure of Fast Radio Bursts with the CHIME/FRB far side-lobe events". In: *arXiv e-prints*, arXiv:2307.05262, arXiv:2307.05262. DOI: `10.48550/arXiv.2307.05262`. arXiv: `2307.05262 [astro-ph.HE]`.

Lin, Hsiu-Hsien et al. (July 2023b). "Do All Fast Radio Bursts Repeat? Constraints from CHIME/FRB Far Side-Lobe FRBs". In: *arXiv e-prints*, arXiv:2307.05261, arXiv:2307.05261. DOI: `10.48550/arXiv.2307.05261`. arXiv: `2307.05261 [astro-ph.HE]`.

Linares, M. (Nov. 2014). "X-Ray States of Redback Millisecond Pulsars". In: *ApJ* 795.1, 72, p. 72. DOI: 10.1088/0004-637X/795/1/72. arXiv: 1406.2384 [astro-ph.HE].

Linares, Manuel et al. (Feb. 2014). "The neutron star transient and millisecond pulsar in M28: from sub-luminous accretion to rotation-powered quiescence". In: *MNRAS* 438.1, pp. 251–261. DOI: 10.1093/mnras/stt2167. arXiv: 1310.7937 [astro-ph.HE].

Lorimer, D. R. and M. Kramer (2005). *Handbook of Pulsar Astronomy*. Cambridge University Press.

Lorimer, D. R. et al. (Nov. 2007). "A Bright Millisecond Radio Burst of Extragalactic Origin". In: *Science* 318.5851, p. 777. DOI: 10.1126/science.1147532. arXiv: 0709.4301 [astro-ph].

Lorimer, Duncan R. (Dec. 2008). "Binary and Millisecond Pulsars". In: *Living Reviews in Relativity* 11.1, 8, p. 8. DOI: 10.12942/lrr-2008-8. arXiv: 0811.0762 [astro-ph].

Lorimer, Duncan R., Maura A. McLaughlin, and Matthew Bailes (June 2024). "The discovery and significance of fast radio bursts". In: *Ap&SS* 369.6, 59, p. 59. DOI: 10.1007/s10509-024-04322-6. arXiv: 2405.19106 [astro-ph.HE].

Lovelace Rainbolt, J. and M. Schmitt (Feb. 2017). "The use of minimal spanning trees in particle physics". In: *Journal of Instrumentation* 12.2, P02009. DOI: 10.1088/1748-0221/12/02/P02009. arXiv: 1608.04772 [stat.AP].

Lu, Wenbin, Pawan Kumar, and Bing Zhang (Oct. 2020). "A unified picture of Galactic and cosmological fast radio bursts". In: *MNRAS* 498.1, pp. 1397–1405. DOI: 10.1093/mnras/staa2450. arXiv: 2005.06736 [astro-ph.HE].

Lundberg, Scott and Su-In Lee (May 2017). "A Unified Approach to Interpreting Model Predictions". In: *arXiv e-prints*, arXiv:1705.07874, arXiv:1705.07874. DOI: 10.48550/arXiv.1705.07874. arXiv: 1705.07874 [cs.AI].

Lundberg, Scott M., Gabriel G. Erion, and Su-In Lee (Feb. 2018). "Consistent Individualized Feature Attribution for Tree Ensembles". In: *arXiv e-prints*, arXiv:1802.03888, arXiv:1802.03888. DOI: 10.48550/arXiv.1802.03888. arXiv: 1802.03888 [cs.LG].

Luo, Jia-Wei, Jia-Ming Zhu-Ge, and Bing Zhang (Jan. 2023). "Machine learning classification of CHIME fast radio bursts - I. Supervised methods". In: *MNRAS* 518.2, pp. 1629–1641. DOI: 10.1093/mnras/stac3206. arXiv: 2210.02463 [astro-ph.HE].

Luo, R. et al. (Oct. 2020). "Diverse polarization angle swings from a repeating fast radio burst source". In: *Nature* 586.7831, pp. 693–696. DOI: 10.1038/s41586-020-2827-2. arXiv: 2011.00171 [astro-ph.HE].

Lynch, Ryan S. et al. (2012). "THE TIMING OF NINE GLOBULAR CLUSTER PULSARS". In: *The Astrophysical Journal* 745.2, p. 109. DOI: 10.1088/0004-637X/745/2/109. URL: https://dx.doi.org/10.1088/0004-637X/745/2/109.

Lyne, Andrew and Francis Graham-Smith (2012). *Pulsar Astronomy*.

Lyubarsky, Yu. (July 2014). "A model for fast extragalactic radio bursts." In: *MNRAS* 442, pp. L9–L13. DOI: 10.1093/mnrasl/slu046. arXiv: 1401.6674 [astro-ph.HE].

Lévy, Paul and Loève Michel (1965). *Processus stochastiques et mouvement brownien . suivi d'une note de M. Loève*. fre. 2e édition revue et augmentée. Paris: Gauthier-Villars & Cie, imprimeur-éditeur.

Macquart, J. P. et al. (May 2020). "A census of baryons in the Universe from localized fast radio bursts". In: *Nature* 581.7809, pp. 391–395. DOI: 10.1038/s41586-020-2300-2. arXiv: 2005.13161 [astro-ph.CO].

Malzac, J. et al. (Mar. 2006). "Bimodal spectral variability of <ASTROBJ>Cygnus X-1</ASTROBJ> in an intermediate state". In: *A&A* 448.3, pp. 1125–1137. DOI: 10.1051/0004-6361:20053614. arXiv: astro-ph/0511725 [astro-ph].

Manchester, R. N. (Sept. 2017). "Millisecond Pulsars, their Evolution and Applications". In: *Journal of Astrophysics and Astronomy* 38.3, 42, p. 42. DOI: 10.1007/s12036-017-9469-2. arXiv: 1709.09434 [astro-ph.HE].

Manchester, R. N. et al. (Apr. 2005). "The Australia Telescope National Facility Pulsar Catalogue". In: *AJ* 129.4, pp. 1993–2006. DOI: 10.1086/428488. arXiv: astro-ph/0412641 [astro-ph].

Marcote, B. et al. (Jan. 2020). "A repeating fast radio burst source localized to a nearby spiral galaxy". In: *Nature* 577.7789, pp. 190–194. DOI: 10.1038/s41586-019-1866-z. arXiv: 2001.02222 [astro-ph.HE].

Mardia, K V, J M Bibby, and J T Kent (1979). *Multivariate analysis*. Academic Press, xv, 521 p. : ISBN: 0124712509 0124712525 0124712525. URL: http://www.loc.gov/catdir/toc/els031/79040922.html.

Margalit, Ben and Brian D. Metzger (Nov. 2018). "A Concordance Picture of FRB 121102 as a Flaring Magnetar Embedded in a Magnetized Ion-Electron Wind Nebula". In: *ApJ* 868.1, L4, p. L4. DOI: 10.3847/2041-8213/aaedad. arXiv: 1808.09969 [astro-ph.HE].

Margalit, Ben et al. (2020). "Implications of a Fast Radio Burst from a Galactic Magnetar". In: *The Astrophysical Journal Letters* 899.2, p. L27. DOI: 10.3847/2041-8213/abac57. URL: https://dx.doi.org/10.3847/2041-8213/abac57.

Maritz, J., E. Maritz, and P. Meintjes (2016). "Graph theory and pulsar astronomy tie the knot: the use of labeled graph kernels in exploring the pulsar $\dot{P}$-P diagram". In: *The Proceedings of SAIP2016, the 61st Annual Conference of the South African Institute of Physics, edited by Steve Peterson and Sahal Yacoob (UCT/2016), pp. 243 - 248. ISBN: 978-0-620-77094-1.*

Martín, Jonatan, Diego F. Torres, and Giovanna Pedaletti (July 2016). "Molecular environment, reverberation, and radiation from the pulsar wind nebula in CTA 1". In: *MNRAS* 459.4, pp. 3868–3879. DOI: 10.1093/mnras/stw684. arXiv: 1603.09328 [astro-ph.HE].

Mayall, N. U. and J. H. Oort (Apr. 1942). "Further Data Bearing on the Identification of the Crab Nebula with the Supernova of 1054 A.D. Part II. The Astronomical Aspects". In: *PASP* 54.318, pp. 95–104. DOI: 10.1086/125410.

McKinney, Wes et al. (2010). "Data structures for statistical computing in python". In: *Proceedings of the 9th Python in Science Conference*. Vol. 445. Austin, TX, pp. 51–56.

McLaughlin, M. A. et al. (Feb. 2006). "Transient radio bursts from rotating neutron stars". In: *Nature* 439.7078, pp. 817–820. DOI: 10.1038/nature04440. arXiv: astro-ph/0511587 [astro-ph].

Mereghetti, S. et al. (Aug. 2020). "INTEGRAL Discovery of a Burst with Associated Radio Emission from the Magnetar SGR 1935+2154". In: *ApJ* 898.2, L29, p. L29. DOI: 10.3847/2041-8213/aba2cf. arXiv: 2005.06335 [astro-ph.HE].

Metzger, Brian D., Edo Berger, and Ben Margalit (May 2017). "Millisecond Magnetar Birth Connects FRB 121102 to Superluminous Supernovae and Long-duration Gamma-Ray Bursts". In: *ApJ* 841.1, 14, p. 14. DOI: 10.3847/1538-4357/aa633d. arXiv: 1701.02370 [astro-ph.HE].

Miller, L., T. J. Turner, and J. N. Reeves (May 2008). "An absorption origin for the X-ray spectral variability of MCG-6-30-15". In: *A&A* 483.2, pp. 437–452. DOI: 10.1051/0004-6361:200809590. arXiv: 0803.2680 [astro-ph].

Minkowski, R. (Sept. 1942). "The Crab Nebula." In: *ApJ* 96, p. 199. DOI: 10.1086/144447.

Miraval Zanon, A. et al. (Jan. 2018). "Study of the eclipse region of the redback millisecond pulsar J1431-4715". In: *Journal of Physics Conference Series*. Vol. 956. Journal of Physics Conference Series. IOP, 012004, p. 012004. DOI: 10.1088/1742-6596/956/1/012004.

Mittaz, J. P. D., M. V. Penston, and M. A. J. Snijders (June 1990). "Ultraviolet variability of NGC 4151: a study using principal component analysis". In: *Monthly Notices of the Royal Astronomical Society* 242.3, pp. 370–378. ISSN: 0035-8711. DOI: 10.1093/mnras/242.3.370. eprint: https://academic.oup.com/mnras/article-pdf/242/3/370/3863623/mnras242-0370.pdf. URL: https://doi.org/10.1093/mnras/242.3.370.

Moxley, Robert L. and Nancy F. Moxley (1974). "Determining Point-Centrality in Uncontrived Social Networks". In: *Sociometry* 37.1, pp. 122–130. ISSN: 00380431. URL: http://www.jstor.org/stable/2786472 (visited on 09/14/2022).

Müller, Andreas C. and Sarah Guido (2016). *Introduction to Machine Learning with Python*. Sebastopol, CA: O'Reilly.

Muller, Meinard (2007). *Dynamic Time Warping*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 69–84. ISBN: 978-3-540-74048-3. DOI: 10.1007/978-3-540-74048-3_4. URL: https://doi.org/10.1007/978-3-540-74048-3_4.

Müller, Meinard (2015). *Fundamentals of Music Processing - Audio, Analysis, Algorithms, Applications*. Springer. ISBN: 978-3-319-21944-8. DOI: 10.1007/978-3-319-21945-5. URL: https://doi.org/10.1007/978-3-319-21945-5.

Munich, M.E. and P. Perona (1999). "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 1, 108–115 vol.1. DOI: 10.1109/ICCV.1999.791205.

Murase, Kohta, Kazumi Kashiyama, and Peter Mészáros (Sept. 2016). "A burst in a wind bubble and the impact on baryonic ejecta: high-energy gamma-ray flashes and afterglows from fast radio bursts and pulsar-driven supernova remnants". In: *MNRAS* 461.2, pp. 1498–1511. DOI: 10.1093/mnras/stw1328. arXiv: 1603.08875 [astro-ph.HE].

Muslimov, Alex G. and Alice K. Harding (May 2004). "High-Altitude Particle Acceleration and Radiation in Pulsar Slot Gaps". In: *ApJ* 606.2, pp. 1143–1153. DOI: 10.1086/383079. arXiv: astro-ph/0402462 [astro-ph].

Myers, C., L. Rabiner, and A. Rosenberg (1980). "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28.6, pp. 623–635. DOI: 10.1109/TASSP.1980.1163491.

Myers, C. S. and L. R. Rabiner (1981). "A comparative study of several dynamic time-warping algorithms for connected-word recognition". In: *The Bell System Technical Journal* 60.7, pp. 1389–1409. DOI: 10.1002/j.1538-7305.1981.tb00272.x.

Naidoo, Krishna and Ofer Lahav (Oct. 2024). "Methods for robustly measuring the minimum spanning tree and other field level statistics from galaxy surveys". In: *arXiv e-prints*, arXiv:2410.06202, arXiv:2410.06202. DOI: 10.48550/arXiv.2410.06202. arXiv: 2410.06202 [astro-ph.CO].

Naidoo, Krishna, Elena Massara, and Ofer Lahav (July 2022). "Cosmology and neutrino mass with the minimum spanning tree". In: *MNRAS* 513.3, pp. 3596–3609. DOI: 10.1093/mnras/stac1138. arXiv: 2111.12088 [astro-ph.CO].

Naidoo, Krishna et al. (Jan. 2020). "Beyond two-point statistics: using the minimum spanning tree as a tool for cosmology". In: *MNRAS* 491.2, pp. 1709–1726. DOI: 10.1093/mnras/stz3075. arXiv: 1907.00989 [astro-ph.CO].

Nešetřil, Jaroslav, Eva Milková, and Helena Nešetřilová (2001). "Otakar Borůvka on minimum spanning tree problem Translation of both the 1926 papers, comments, history". In: *Discrete Mathematics* 233.1. Czech and Slovak 2, pp. 3–36. ISSN: 0012-365X. DOI: https://doi.org/10.1016/S0012-365X(00)00224-7. URL: https://www.sciencedirect.com/science/article/pii/S0012365X00002247.

Nicholl, M. et al. (July 2017). "Empirical Constraints on the Origin of Fast Radio Bursts: Volumetric Rates and Host Galaxy Demographics as a Test of Millisecond Magnetar Connection". In: *ApJ* 843.2, 84, p. 84. DOI: 10.3847/1538-4357/aa794d. arXiv: 1704.00022 [astro-ph.HE].

Niu, C. H. et al. (2022). "A repeating fast radio burst associated with a persistent radio source". In: *Nature* 606.7916, pp. 873–877. DOI: 10.1038/s41586-022-04755-5. URL: https://doi.org/10.1038/s41586-022-04755-5.

Olausen, S. A. and V. M. Kaspi (May 2014). "The McGill Magnetar Catalog". In: *ApJS* 212.1, 6, p. 6. DOI: 10.1088/0067-0049/212/1/6. arXiv: 1309.4167 [astro-ph.HE].

Oppenheimer, J. R. and G. M. Volkoff (1939). "On Massive Neutron Cores". In: *Phys. Rev.* 55 (4), pp. 374–381. DOI: 10.1103/PhysRev.55.374. URL: https://link.aps.org/doi/10.1103/PhysRev.55.374.

Ostriker, J. P. and J. E. Gunn (Sept. 1969). "On the Nature of Pulsars. I. Theory". In: *ApJ* 157, p. 1395. DOI: 10.1086/150160.

Özel, Feryal and Paulo Freire (Sept. 2016). "Masses, Radii, and the Equation of State of Neutron Stars". In: *ARA&A* 54, pp. 401–440. DOI: 10.1146/annurev-astro-081915-023322. arXiv: 1603.02698 [astro-ph.HE].

Pacini, F. (1967). "Energy Emission from a Neutron Star". In: *Nature* 216.5115, pp. 567–568. DOI: 10.1038/216567a0. URL: https://doi.org/10.1038/216567a0.

— (1968). "Rotating Neutron Stars, Pulsars and Supernova Remnants". In: *Nature* 219.5150, pp. 145–146. DOI: 10.1038/219145a0. URL: https://doi.org/10.1038/219145a0.

Palaniswamy, Divya, Ye Li, and Bing Zhang (Feb. 2018). "Are There Multiple Populations of Fast Radio Bursts?" In: *ApJ* 854.1, L12, p. L12. DOI: 10.3847/2041-8213/aaaa63. arXiv: 1703.09232 [astro-ph.HE].

Pan, Zhichen et al. (July 2021). "FAST Globular Cluster Pulsar Survey: Twenty-four Pulsars Discovered in 15 Globular Clusters". In: *ApJ* 915.2, L28, p. L28. DOI: 10.3847/2041-8213/ac0bbd. arXiv: 2106.08559 [astro-ph.HE].

Papitto, A. and S. Bhattacharyya (2022). *Millisecond pulsars*. Springer ASSL.

Papitto, A. et al. (2013). "Swings between rotation and accretion power in a binary millisecond pulsar". In: *nature* 501.7468, p. 517. ISSN: 1476-4687. URL: https://doi.org/10.1038/nature12470.

Papitto, A. et al. (June 2014). "Spin frequency distributions of binary millisecond pulsars". In: *A&A* 566, A64, A64. DOI: 10.1051/0004-6361/201321724. arXiv: 1403.6775 [astro-ph.HE].

Papitto, A. et al. (Sept. 2019). "Pulsating in Unison at Optical and X-Ray Energies: Simultaneous High Time Resolution Observations of the Transitional Millisecond Pulsar PSR J1023+0038". In: *ApJ* 882.2, 104, p. 104. DOI: 10.3847/1538-4357/ab2fdf. arXiv: 1904.10433 [astro-ph.HE].

Papitto, Alessandro and Domitilla de Martino (Jan. 2022). "Transitional Millisecond Pulsars". In: *Astrophysics and Space Science Library*. Ed. by Sudip Bhattacharyya, Alessandro Papitto, and Dipankar Bhattacharya. Vol. 465. Astrophysics and Space Science Library, pp. 157–200. DOI: 10.1007/978-3-030-85198-9_6. arXiv: 2010.09060 [astro-ph.HE].

Parker, M. L. et al. (June 2014a). "PCA of PCA: principal component analysis of partial covering absorption in NGC 1365". In: *MNRAS* 441.2, pp. 1817–1824. DOI: 10.1093/mnras/stu712. arXiv: 1404.2611 [astro-ph.HE].

Parker, M. L. et al. (Jan. 2014b). "Principal component analysis of MCG-06-30-15 with XMM-Newton". In: *MNRAS* 437.1, pp. 721–729. DOI: 10.1093/mnras/stt1925. arXiv: 1310.1945 [astro-ph.HE].

Parthasarathy, A. et al. (Nov. 2019). "Timing of young radio pulsars - I. Timing noise, periodic modulation, and proper motion". In: *MNRAS* 489.3, pp. 3810–3826. DOI: 10.1093/mnras/stz2383. arXiv: 1908.11709 [astro-ph.HE].

Parthasarathy, A. et al. (May 2020). "Timing of young radio pulsars - II. Braking indices and their interpretation". In: *MNRAS* 494.2, pp. 2012–2026. DOI: 10.1093/mnras/staa882. arXiv: 2003.13303 [astro-ph.HE].

Patruno, A. et al. (Jan. 2014). "A New Accretion Disk around the Missing Link Binary System PSR J1023+0038". In: *ApJ* 781.1, L3, p. L3. DOI: 10.1088/2041-8205/781/1/L3. arXiv: 1310.7549 [astro-ph.HE].

Patruno, Alessandro and Anna L. Watts (Jan. 2021). "Accreting Millisecond X-ray Pulsars". In: *Timing Neutron Stars: Pulsations, Oscillations and Explosions*. Ed. by Tomaso M. Belloni, Mariano Méndez, and Chengmin Zhang. Vol. 461. Astrophysics and Space Science Library, pp. 143–208. DOI: 10.1007/978-3-662-62110-3_4. arXiv: 1206.2727 [astro-ph.HE].

Patruno, Alessandro et al. (Feb. 2012). "Accelerated Orbital Expansion and Secular Spin-down of the Accreting Millisecond Pulsar SAX J1808.4-3658". In: *ApJ* 746.2, L27, p. L27. DOI: 10.1088/2041-8205/746/2/L27. arXiv: 1111.6967 [astro-ph.HE].

Pearlman, Aaron B. et al. (Jan. 2025). "Multiwavelength constraints on the origin of a nearby repeating fast radio burst source in a globular cluster". In: *Nature Astronomy* 9, pp. 111–127. DOI: 10.1038/s41550-024-02386-6. arXiv: 2308.10930 [astro-ph.HE].

Pearson, K. (1901). "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11, pp. 559–572. DOI: 10.1080/14786440109462720. eprint: https://doi.org/10.1080/14786440109462720.

Penny, Alan John (Sept. 2013). "The SETI episode in the 1967 discovery of pulsars". In: *European Physical Journal H* 38.4, pp. 535–547. DOI: 10.1140/epjh/e2012-30052-6. arXiv: 1302.0641 [physics.hist-ph].

Pereyra, Luis A. et al. (Dec. 2020). "Detection and analysis of cluster-cluster filaments". In: *MNRAS* 499.4, pp. 4876–4886. DOI: 10.1093/mnras/staa3112. arXiv: 1911.06768 [astro-ph.CO].

Pétri, J. (July 2012). "The pulsar force-free magnetosphere linked to its striped wind: time-dependent pseudo-spectral simulations". In: *MNRAS* 424.1, pp. 605–619. DOI: 10.1111/j.1365-2966.2012.21238.x. arXiv: 1205.0889 [astro-ph.HE].

— (June 2015). "Multipolar electromagnetic fields around neutron stars: exact vacuum solutions and related properties". In: *MNRAS* 450.1, pp. 714–742. DOI: 10.1093/mnras/stv598. arXiv: 1503.05307 [astro-ph.HE].

Pétri, J. and D. Mitra (Oct. 2021). "Young radio-loud gamma-ray pulsar light curve fitting". In: *A&A* 654, A106, A106. DOI: 10.1051/0004-6361/202141272. arXiv: 2107.06032 [astro-ph.HE].

Petroff, E., J. W. T. Hessels, and D. R. Lorimer (May 2019). "Fast radio bursts". In: *A&ARv* 27.1, 4, p. 4. DOI: 10.1007/s00159-019-0116-6. arXiv: 1904.07947 [astro-ph.HE].

Petroff, E., J. W. T. Hessels, and D. R. Lorimer (Dec. 2022). "Fast radio bursts at the dawn of the 2020s". In: *A&ARv* 30.1, 2, p. 2. DOI: 10.1007/s00159-022-00139-w. arXiv: 2107.10113 [astro-ph.HE].

Petroff, E. et al. (Aug. 2015). "Identifying the source of perytons at the Parkes radio telescope". In: *MNRAS* 451.4, pp. 3933–3940. DOI: 10.1093/mnras/stv1242. arXiv: 1504.02165 [astro-ph.IM].

Philippov, A. and M. Kramer (Aug. 2022). "Pulsar Magnetospheres and Their Radiation". In: *ARA&A* 60, pp. 495–558. DOI: 10.1146/annurev-astro-052920-112338.

Philippov, Alexander A. and Anatoly Spitkovsky (Apr. 2014). "Ab Initio Pulsar Magnetosphere: Three-dimensional Particle-in-cell Simulations of Axisymmetric Pulsars". In: *ApJ* 785.2, L33, p. L33. DOI: 10.1088/2041-8205/785/2/L33. arXiv: 1312.4970 [astro-ph.HE].

Pierbattista, M. et al. (Sept. 2012). "Constraining γ-ray pulsar gap models with a simulated pulsar population". In: *A&A* 545, A42, A42. DOI: 10.1051/0004-6361/201219135. arXiv: 1206.5634 [astro-ph.HE].

Pitkin, M. (Feb. 2018). "psrqpy: a python interface for querying the ATNF pulsar catalogue". In: *Journal of Open Source Software* 3.22, p. 538. DOI: 10.21105/joss.00538. URL: https://doi.org/10.21105/joss.00538.

Platts, E. et al. (Aug. 2019). "A living theory catalogue for fast radio bursts". In: *Phys. Rep.* 821, pp. 1–27. DOI: 10.1016/j.physrep.2019.06.003. arXiv: 1810.05836 [astro-ph.HE].

Podsiadlowski, Philipp (Jan. 2012). "The evolution of binary systems". In: *Accretion Processes In Astrophysics: XXI Canary Islands Winter School Of Astrophysics*, pp. 45–88. DOI: 10.1017/CBO9781139343268.003.

Popov, S. B. and K. A. Postnov (July 2013). "Millisecond extragalactic radio bursts as magnetar flares". In: *arXiv e-prints*, arXiv:1307.4924, arXiv:1307.4924. DOI: 10.48550/arXiv.1307.4924. arXiv: 1307.4924 [astro-ph.HE].

Popov, S. B., K. A. Postnov, and M. S. Pshirkov (Oct. 2018). "Fast radio bursts". In: *Physics Uspekhi* 61.10, p. 965. DOI: 10.3367/UFNe.2018.03.038313. arXiv: 1806.03628 [astro-ph.HE].

Postnov, Konstantin A. and Lev R. Yungelson (2014). "The Evolution of Compact Binary Star Systems". In: *Living Reviews in Relativity* 17.1, p. 3. DOI: 10.12942/lrr-2014-3. URL: https://doi.org/10.12942/lrr-2014-3.

Pozzi, F., T. Di Matteo, and T. Aste (Apr. 2013). "Spread of risk across financial markets: better to invest in the peripheries". In: *Scientific Reports* 3, 1665, p. 1665. DOI: 10.1038/srep01665.

Prim, R. C. (Nov. 1957). "Shortest Connection Networks And Some Generalizations". In: *Bell System Technical Journal* 36.6, pp. 1389–1401. DOI: 10.1002/j.1538-7305.1957.tb01515.x.

Prochaska, J. Xavier et al. (Oct. 2019). "The low density and magnetization of a massive galaxy halo exposed by a fast radio burst". In: *Science* 366.6462, pp. 231–234. DOI: 10.1126/science.aay0073. arXiv: 1909.11681 [astro-ph.GA].

Rabiner, L. and B. Juang (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs, New Jersey: Prentice-Hall Signal Processing Series.

Radhakrishnan, V. and D. J. Cooke (Jan. 1969). "Magnetic Poles and the Polarization Structure of Pulsar Radiation". In: *Astrophys. Lett.* 3, p. 225.

Radhakrishnan, V. and R. N. Manchester (Apr. 1969). "Detection of a Change of State in the Pulsar PSR 0833-45". In: *Nature* 222.5190, pp. 228–229. DOI: 10.1038/222228a0.

Radhakrishnan, V. and G. Srinivasan (Dec. 1982). "On the origin of the recently discovered ultra-rapid pulsar". In: *Current Science* 51, pp. 1096–1099.

Radhakrishnan, V. et al. (Feb. 1969). "Evidence in Support of a Rotational Model for the Pulsar PSR 0833-45". In: *Nature* 221.5179, pp. 443–446. DOI: 10.1038/221443a0.

Raihan, Tasneem (2017). "Predicting US Recessions: A Dynamic Time Warping Exercise in Economics". In: *International Political Economy: Investment & Finance eJournal*. URL: https://api.semanticscholar.org/CorpusID:32099625.

Ramaswamy, Ramakrishna (Jan. 2016). *D.D. Kosambi: Selected Works in Mathematics and Statistics*. Springer New Delhi. ISBN: 978-81-322-3674-0. DOI: 10.1007/978-81-322-3676-4.

Raschka, Sebastian (2015). *Python Machine Learning*. Packt Publishing. ISBN: 1783555130.

Rashidi, R., F. Z. Boroujeni, M. Soltanaghaei, et al. (2024). "Prediction of Influential Nodes in Social Networks Based on Local Communities and Users' Reaction Information". In: *Scientific Reports* 14, p. 15815. DOI: 10.1038/s41598-024-66277-6. URL: https://doi.org/10.1038/s41598-024-66277-6.

Ravi, V. et al. (2019). "A fast radio burst localized to a massive galaxy". In: *Nature* 572.7769, pp. 352–354. DOI: 10.1038/s41586-019-1389-7. URL: https://doi.org/10.1038/s41586-019-1389-7.

Ray, P. S. et al. (May 2012). "Radio Searches of Fermi LAT Sources and Blind Search Pulsars: The Fermi Pulsar Search Consortium". In: *arXiv e-prints*, arXiv:1205.3089, arXiv:1205.3089. DOI: 10.48550/arXiv.1205.3089. arXiv: 1205.3089 [astro-ph.HE].

Rencher, Alvin C. and William F. Christensen (2012). *Methods of multivariate analysis*. URL: http://www.amazon.de/Methods-Multivariate-Analysis-Probability-Statistics/dp/0470178965.

Richards, D. W. and J. M. Comella (1969). "The Period of Pulsar NP 0532". In: *Nature* 222.5193, pp. 551–552. DOI: 10.1038/222551a0. URL: https://doi.org/10.1038/222551a0.

Roberts, Mallory S. E. (Mar. 2013). "Surrounded by spiders! New black widows and redbacks in the Galactic field". In: *Neutron Stars and Pulsars: Challenges and Opportunities after 80 years*. Ed. by Joeri van Leeuwen. Vol. 291, pp. 127–132. DOI: 10.1017/S174392131202337X. arXiv: 1210.6903 [astro-ph.HE].

Roberts, Mallory S. E. et al. (Aug. 2018). "X-Ray and Optical Properties of Black Widows and Redbacks". In: *Pulsar Astrophysics the Next Fifty Years*. Ed. by P. Weltevrede et al. Vol. 337, pp. 43–46. DOI: 10.1017/S1743921318000480. arXiv: 1801.09903 [astro-ph.HE].

Romani, Roger W. and Kyle P. Watters (May 2010). "Constraining Pulsar Magnetosphere Geometry with γ-ray Light Curves". In: *ApJ* 714.1, pp. 810–824. DOI: 10.1088/0004-637X/714/1/810. arXiv: 0912.3845 [astro-ph.HE].

Romani, Roger W. and I. A. Yadigaroglu (Jan. 1995). "Gamma-Ray Pulsars: Emission Zones and Viewing Geometries". In: *ApJ* 438, p. 314. DOI: 10.1086/175076. arXiv: astro-ph/9401045 [astro-ph].

Roughgarden, T. (2019). *Algorithms Illuminated (Part 3): Greedy Algorithms and Dynamic Programming*. Soundlikeyourself Publishing, LLC.

Roughgarden, Tim and Éva Tardos (Feb. 2002). "How Bad is Selfish Routing?" In: *Journal of the ACM* 49. DOI: 10.1145/506147.506153.

Sabidussi, Gert (1966). "The Centrality Index of a Graph". In: *Psychometrika* 31.4, 581–603. DOI: 10.1007/BF02289527.

Saitou, Kei et al. (Aug. 2009). "Suzaku X-Ray Study of an Anomalous Source XSS J12270-4859". In: *PASJ* 61, p. L13. DOI: 10.1093/pasj/61.4.L13. arXiv: 0904.4427 [astro-ph.HE].

Sakoe, H. and S. Chiba (1978). "Dynamic programming algorithm optimization for spoken word recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1, pp. 43–49. DOI: 10.1109/TASSP.1978.1163055.

Samara, E. et al. (Mar. 2022). "Dynamic Time Warping as a Means of Assessing Solar Wind Time Series". In: *ApJ* 927.2, 187, p. 187. DOI: 10.3847/1538-4357/ac4af6. arXiv: 2109.07873 [astro-ph.SR].

Sánchez, Néstor, Emilio J. Alfaro, and Fátima López-Martínez (Apr. 2018). "A method for determining the radius of an open cluster from stellar proper motions". In: *MNRAS* 475.3, pp. 4122–4132. DOI: 10.1093/mnras/sty088. arXiv: 1801.03804 [astro-ph.GA].

Sanna, A. et al. (Oct. 2017). "On the timing properties of SAX J1808.4-3658 during its 2015 outburst". In: *MNRAS* 471.1, pp. 463–477. DOI: 10.1093/mnras/stx1588. arXiv: 1706.06554 [astro-ph.HE].

Sart, Doruk et al. (2010). "Accelerating Dynamic Time Warping Subsequence Search with GPUs and FPGAs". In: *2010 IEEE International Conference on Data Mining*, pp. 1001–1006. DOI: 10.1109/ICDM.2010.21.

Scholz, P. et al. (Dec. 2016). "The Repeating Fast Radio Burst FRB 121102: Multi-wavelength Observations and Additional Bursts". In: *ApJ* 833.2, 177, p. 177. DOI: 10.3847/1538-4357/833/2/177. arXiv: 1603.08880 [astro-ph.HE].

Schreier, E. et al. (Mar. 1972). "Evidence for the Binary Nature of Centaurus X-3 from UHURU X-Ray Observations." In: *ApJ* 172, p. L79. DOI: 10.1086/180896.

Sempena, Samsu, Nur Ulfa Maulidevi, and Peb Ruswono Aryan (2011). "Human action recognition using Dynamic Time Warping". In: *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, pp. 1–5. DOI: 10.1109/ICEEI.2011.6021605.

Sen, Bidisha et al. (July 2024). "The Orbit and Companion of PSR J1622-0315: Variable Asymmetry and a Massive Neutron Star". In: *arXiv e-prints*, arXiv:2407.10800, arXiv:2407.10800. DOI: 10.48550/arXiv.2407.10800. arXiv: 2407.10800 [astro-ph.HE].

Shapiro, Stuart L. and Saul A. Teukolsky (1983). *Black holes, white dwarfs and neutron stars. The physics of compact objects*. DOI: 10.1002/9783527617661.

Shields, Rob (2012). "Cultural Topology: The Seven Bridges of Königsburg, 1736". In: *Theory, Culture & Society* 29.4-5, pp. 43–57. DOI: 10.1177/0263276412451161. eprint: https://doi.org/10.1177/0263276412451161. URL: https://doi.org/10.1177/0263276412451161.

Shklovsky, I. S. (Apr. 1967). "On the Nature of the Source of X-Ray Emission of Sco XR-1." In: *ApJ* 148, p. L1. DOI: 10.1086/180001.

Shlens, Jonathon (Apr. 2014). "A Tutorial on Principal Component Analysis". In: *arXiv e-prints*, arXiv:1404.1100, arXiv:1404.1100. arXiv: 1404.1100 [cs.LG].

Sigurdsson, Steinn et al. (July 2003). "A Young White Dwarf Companion to Pulsar B1620-26: Evidence for Early Planet Formation". In: *Science* 301.5630, pp. 193–196. DOI: 10.1126/science.1086326. arXiv: astro-ph/0307339 [astro-ph].

Simon, Olivier B. et al. (Oct. 2021). "A novel approach to understanding Parkinsonian cognitive decline using minimum spanning trees, edge cutting, and magnetoencephalography". In: *Scientific Reports* 11, 19704, p. 19704. DOI: 10.1038/s41598-021-99167-2.

Smerlak, Matteo et al. (July 2015). "Mapping Systemic Risk: Critical Degree and Failures Distribution in Financial Networks". In: *PLoS ONE* 10.7, e0130948. DOI: 10.1371/journal.pone.0130948. arXiv: 1402.4783 [q-fin.GN].

Smith, D. A. et al. (Jan. 2019). "Searching a Thousand Radio Pulsars for Gamma-Ray Emission". In: *ApJ* 871.1, 78, p. 78. DOI: 10.3847/1538-4357/aaf57d. arXiv: 1812.00719 [astro-ph.HE].

Smith, D. A. et al. (Dec. 2023). "The Third Fermi Large Area Telescope Catalog of Gamma-Ray Pulsars". In: *ApJ* 958.2, 191, p. 191. DOI: 10.3847/1538-4357/acee67. arXiv: 2307.11132 [astro-ph.HE].

Sobey, C. et al. (May 2022). "Searching for pulsars associated with polarised point sources using LOFAR: Initial discoveries from the TULIPP project". In: *Astronomy & Astrophysics* 661, A87. ISSN: 1432-0746. DOI: 10.1051/0004-6361/202142636. URL: http://dx.doi.org/10.1051/0004-6361/202142636.

Spitkovsky, Anatoly (Sept. 2006). "Time-dependent Force-free Pulsar Magnetospheres: Axisymmetric and Oblique Rotators". In: *ApJ* 648.1, pp. L51–L54. DOI: 10.1086/507518. arXiv: astro-ph/0603147 [astro-ph].

Spitler, L. G. et al. (Mar. 2016). "A repeating fast radio burst". In: *Nature* 531.7593, pp. 202–205. DOI: 10.1038/nature17168. arXiv: 1603.00581 [astro-ph.HE].

Staelin, David H. and Edward C. Reifenstein (1968). "Pulsating Radio Sources near the Crab Nebula". In: *Science* 162.3861, pp. 1481–1483. DOI: 10.1126/science.162.3861.1481. URL: https://doi.org/10.1126/science.162.3861.1481.

Stappers, B. W. et al. (July 2014). "A State Change in the Missing Link Binary Pulsar System PSR J1023+0038". In: *ApJ* 790.1, 39, p. 39. DOI: 10.1088/0004-637X/790/1/39. arXiv: 1311.7506 [astro-ph.HE].

Strader, Jay et al. (2019). "Optical Spectroscopy and Demographics of Redback Millisecond Pulsar Binaries". In: *The Astrophysical Journal* 872.1, p. 42. DOI: 10.3847/1538-4357/aafbaa. URL: https://dx.doi.org/10.3847/1538-4357/aafbaa.

Strader, Jay et al. (Feb. 2019). "Optical Spectroscopy and Demographics of Redback Millisecond Pulsar Binaries". In: *ApJ* 872.1, 42, p. 42. DOI: 10.3847/1538-4357/aafbaa. arXiv: 1812.04626 [astro-ph.HE].

Strüder, L. et al. (Jan. 2001). "The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera". In: *A&A* 365, pp. L18–L26. DOI: 10.1051/0004-6361:20000066.

Suleimanov, V. et al. (Nov. 2010). "Radiative properties of highly magnetized isolated neutron star surfaces and approximate treatment of absorption features in their spectra". In: *A&A* 522, A111, A111. DOI: 10.1051/0004-6361/200913641. arXiv: 1006.3292 [astro-ph.HE].

Sun, Wan-Peng et al. (Feb. 2025). "Exploring the Key Features of Repeating Fast Radio Bursts with Machine Learning". In: *ApJ* 980.2, 185, p. 185. DOI: 10.3847/1538-4357/adad6a. arXiv: 2409.11173 [astro-ph.HE].

Suvorova, Sofia, Jade Powell, and Andrew Melatos (June 2019). "Reconstructing gravitational wave core-collapse supernova signals with dynamic time warping". In: *Phys. Rev. D* 99.12, 123012, p. 123012. DOI: 10.1103/PhysRevD.99.123012. arXiv: 1901.02535 [astro-ph.IM].

Swiggum, J. K. et al. (Feb. 2023). "The Green Bank North Celestial Cap Survey. VII. 12 New Pulsar Timing Solutions". In: *The Astrophysical Journal* 944.2, p. 154. ISSN: 1538-4357. DOI: 10.3847/1538-4357/acb43f. URL: http://dx.doi.org/10.3847/1538-4357/acb43f.

Swihart, Samuel J. et al. (Dec. 2022). "A New Flaring Black Widow Candidate and Demographics of Black Widow Millisecond Pulsars in the Galactic Field". In: *ApJ* 941.2, 199, p. 199. DOI: 10.3847/1538-4357/aca2ac. arXiv: 2210.16295 [astro-ph.HE].

Tananbaum, H. et al. (June 1972). "Discovery of a Periodic Pulsating Binary X-Ray Source in Hercules from UHURU". In: *ApJ* 174, p. L143. DOI: 10.1086/180968.

Tarjan, Robert Endre (1983). *Data structures and network algorithms*. USA: Society for Industrial and Applied Mathematics. ISBN: 0898711878.

Tauris, T. M., N. Langer, and M. Kramer (Sept. 2012). "Formation of millisecond pulsars with CO white dwarf companions - II. Accretion, spin-up, true ages and comparison to MSPs with He white dwarf companions". In: *MNRAS* 425.3, pp. 1601–1627. DOI: 10.1111/j.1365-2966.2012.21446.x. arXiv: 1206.1862 [astro-ph.SR].

Tauris, T. M. and E. P. J. van den Heuvel (2006). "Formation and evolution of compact stellar X-ray sources". In: *Compact stellar X-ray sources*. Ed. by Walter H. G. Lewin and Michiel van der Klis. Vol. 39, pp. 623–665. DOI: 10.48550/arXiv.astro-ph/0303456.

Tauris, T. M. et al. (Sept. 2017). "Formation of Double Neutron Star Systems". In: *ApJ* 846.2, 170, p. 170. DOI: 10.3847/1538-4357/aa7e89. arXiv: 1706.09438 [astro-ph.HE].

Tauris, Thomas M. and Gerrit J. Savonije (Oct. 1999). "Formation of millisecond pulsars. I. Evolution of low-mass X-ray binaries with P_orb> 2 days". In: *A&A* 350, pp. 928–944. DOI: 10.48550/arXiv.astro-ph/9909147. arXiv: astro-ph/9909147 [astro-ph].

Tauris, Thomas M. and Edward P. J. van den Heuvel (2023). *Physics of Binary Star Evolution. From Stars to X-ray Binaries and Gravitational Wave Sources*. DOI: 10.48550/arXiv.2305.09388.

Tchekhovskoy, A., A. Spitkovsky, and J. G. Li (Aug. 2013). "Time-dependent 3D magnetohydrodynamic pulsar magnetospheres: oblique rotators." In: *MNRAS* 435, pp. L1–L5. DOI: 10.1093/mnrasl/slt076. arXiv: 1211.2803 [astro-ph.HE].

Ten Holt, Gineke A, Marcel JT Reinders, and Emile A Hendriks (2007). "Multi-dimensional dynamic time warping for gesture recognition". In: *Thirteenth annual conference of the Advanced School for Computing and Imaging*. Vol. 300, p. 1.

The Bokeh team (2022). *Bokeh documentation*. https://docs.bokeh.org/en/latest/. [Online; accessed April 1, 2022].

The Chime/FRB Collaboration et al. (Apr. 2023). "CHIME/FRB Discovery of 25 Repeating Fast Radio Burst Sources". In: *ApJ* 947.2, 83, p. 83. DOI: 10.3847/1538-4357/acc6c1. arXiv: 2301.08762 [astro-ph.HE].

The graphviz team (2022). *graphviz documentation*. https://pypi.org/project/graphviz/. [Online; accessed April 1, 2022].

Thornton, D. et al. (July 2013). "A Population of Fast Radio Bursts at Cosmological Distances". In: *Science* 341.6141, pp. 53–56. DOI: 10.1126/science.1236789. arXiv: 1307.1628 [astro-ph.HE].

Thorsett, S. E., Z. Arzoumanian, and J. H. Taylor (July 1993). "PSR B1620-26: A Binary Radio Pulsar with a Planetary Companion?" In: *ApJ* 412, p. L33. DOI: 10.1086/186933.

Thorsett, S. E. et al. (Oct. 1999). "The Triple Pulsar System PSR B1620-26 in M4". In: *ApJ* 523.2, pp. 763–770. DOI: 10.1086/307771. arXiv: astro-ph/9903227 [astro-ph].

Tiengo, Andrea et al. (2013). "A variable absorption feature in the X-ray spectrum of a magnetar". In: *Nature* 500.7462, pp. 312–314. DOI: 10.1038/nature12386. URL: https://doi.org/10.1038/nature12386.

Tolman, Richard C. (1939). "Static Solutions of Einstein's Field Equations for Spheres of Fluid". In: *Phys. Rev.* 55 (4), pp. 364–373. DOI: 10.1103/PhysRev.55.364. URL: https://link.aps.org/doi/10.1103/PhysRev.55.364.

Torres, D. F. et al. (May 2014). "Time-dependent modeling of TeV-detected, young pulsar wind nebulae". In: *Journal of High Energy Astrophysics* 1, pp. 31–62. DOI: 10.1016/j.jheap.2014.02.001. arXiv: 1402.5485 [astro-ph.HE].

Torres, Diego F. (2018). "Order parameters for the high-energy spectra of pulsars". In: *Nature Astronomy* 2.3, pp. 247–256. DOI: 10.1038/s41550-018-0384-5. URL: https://doi.org/10.1038/s41550-018-0384-5.

Torres, Diego F. et al. (Feb. 2017). "A Search for Transitions between States in Redbacks and Black Widows Using Seven Years of Fermi-LAT Observations". In: *ApJ* 836.1, 68, p. 68. DOI: 10.3847/1538-4357/836/1/68. arXiv: 1612.07083 [astro-ph.HE].

Torres, Diego F. et al. (Nov. 2019). "Synchrocurvature modelling of the multifrequency non-thermal emission of pulsars". In: *MNRAS* 489.4, pp. 5494–5512. DOI: 10.1093/mnras/stz2403. arXiv: 1908.11574 [astro-ph.HE].

Truemper, J. et al. (Feb. 1978). "Evidence for strong cyclotron line emission in the hard X-ray spectrum of Hercules X-1." In: *ApJ* 219, pp. L105–L110. DOI: 10.1086/182617.

Tsuruta, Sachiko and A. G. W. Cameron (1966). "Rotation of Neutron Stars". In: *Nature* 211.5047, pp. 356–357. DOI: 10.1038/211356a0. URL: https://doi.org/10.1038/211356a0.

Tukey, John W. (1977). *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.

Turchetti, Simone (2023). "Representing noise: stacked plots and the contrasting diplomatic ambitions of radio astronomy and post-punk". In: *The British Journal for the History of Science* 56.2, 225–245. DOI: 10.1017/S0007087423000122.

Turner, M. J. L. et al. (Jan. 2001). "The European Photon Imaging Camera on XMM-Newton: The MOS cameras". In: *A&A* 365, pp. L27–L35. DOI: 10.1051/0004-6361:20000087. arXiv: astro-ph/0011498 [astro-ph].

van Haaften, L. M. et al. (May 2012). "Formation of the planet around the millisecond pulsar J1719-1438". In: *A&A* 541, A22, A22. DOI: 10.1051/0004-6361/201218798. arXiv: 1203.2919 [astro-ph.SR].

Van Rossum, Guido and Fred L Drake Jr (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vaughan, S. and A. C. Fabian (Mar. 2004). "A long hard look at MCG-6-30-15 with XMM-Newton- II. Detailed EPIC analysis and modelling". In: *MNRAS* 348.4, pp. 1415–1438. DOI: 10.1111/j.1365-2966.2004.07456.x. arXiv: astro-ph/0311473 [astro-ph].

Vidal Ruiz, Enrique, Francisco Casacuberta Nolla, and Hector Rulot Segovia (1985). "Is the DTW "distance" really a metric? An algorithm reducing the number of DTW comparisons in isolated word recognition". In: *Speech Communication* 4.4, pp. 333–344. ISSN: 0167-6393. DOI: https://doi.org/10.1016/0167-6393(85)90058-5. URL: https://www.sciencedirect.com/science/article/pii/0167639385900585.

Viganò, D. et al. (Sept. 2013). "Unifying the observational diversity of isolated neutron stars via magneto-thermal evolution models". In: *MNRAS* 434.1, pp. 123–141. DOI: 10.1093/mnras/stt1008. arXiv: 1306.2156 [astro-ph.SR].

Viganò, Daniele, Diego F. Torres, and Jonatan Martín (Nov. 2015). "A systematic synchro-curvature modelling of pulsar γ-ray spectra unveils hidden trends". In: *MNRAS* 453.3, pp. 2599–2621. DOI: 10.1093/mnras/stv1582. arXiv: 1507.04021 [astro-ph.HE].

Viganò, Daniele et al. (Mar. 2015a). "An assessment of the pulsar outer gap model - I. Assumptions, uncertainties, and implications on the gap size and the accelerating field". In: *MNRAS* 447.3, pp. 2631–2648. DOI: 10.1093/mnras/stu2564. arXiv: 1412.1289 [astro-ph.HE].

Viganò, Daniele et al. (Mar. 2015b). "An assessment of the pulsar outer gap model - II. Implications for the predicted γ-ray spectra". In: *MNRAS* 447.3, pp. 2649–2657. DOI: 10.1093/mnras/stu2565. arXiv: 1412.1290 [astro-ph.HE].

— (Feb. 2015c). "Compact formulae, dynamics and radiation of charged particles under synchro-curvature losses". In: *MNRAS* 447.2, pp. 1164–1172. DOI: 10.1093/mnras/stu2456. arXiv: 1411.5836 [astro-ph.HE].

Vintsyuk, T. K. (1968). "Speech discrimination by dynamic programming". In: *Cybernetics* 4.1, pp. 52–57. ISSN: 1573-8337. DOI: 10.1007/BF01074755. URL: https://doi.org/10.1007/BF01074755.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Vlachos, Michalis, George Kollios, and Dimitrios Gunopulos (Feb. 2002). "Discovering similar multidimensional trajectories". In: pp. 673–684. ISBN: 0-7695-1531-2. DOI: 10.1109/ICDE.2002.994784.

Vleeschower, L. et al. (May 2024). "Discoveries and timing of pulsars in M62". In: *MNRAS* 530.2, pp. 1436–1456. DOI: 10.1093/mnras/stae816. arXiv: 2403.12137 [astro-ph.HE].

Vohl, D., J. van Leeuwen, and Y. Maan (July 2024). "Topology of Pulsar Profiles (ToPP). I. Graph theory method and classification of the EPN". In: *A&A* 687, A113, A113. DOI: 10.1051/0004-6361/202348621. arXiv: 2311.09201 [astro-ph.HE].

Wall, J. V. and C. R. Jenkins (2003). *Practical Statistics for Astronomers*. Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press.

wannesm et al. (Jan. 2022). *wannesm/dtaidistance: v2.3.5*. Version v2.3.5. DOI: 10.5281/zenodo.5901139. URL: https://doi.org/10.5281/zenodo.5901139.

Wheaton, Wm. A. et al. (1979). "An absorption feature in the spectrum of the pulsed hard X-ray flux from 4U0115 + 63". In: *Nature* 282.5736, pp. 240–243. DOI: 10.1038/282240a0. URL: https://doi.org/10.1038/282240a0.

Wheeler, John Archibald (Jan. 1966). "Superdense Stars". In: *ARA&A* 4, p. 393. DOI: 10.1146/annurev.aa.04.090166.002141.

Whitney, C. A. (Mar. 1983). "Principal components analysis of spectral data. I. Methodology for spectral classification." In: *A&AS* 51, pp. 443–461.

Wijnands, Rudy and Michiel van der Klis (1998). "A millisecond pulsar in an X-ray binary system". In: *Nature* 394.6691, pp. 344–346. DOI: 10.1038/28557. URL: https://doi.org/10.1038/28557.

Wilson, R. J. (2010). *Introduction to graph theory*. Pearson.

Winther, R. G. (2018). *Phylogenetic Inference, Selection Theory, and History of Science Selected Papers of A. W. F. Edwards with Commentaries*. Cambridge University Press. DOI: 10.1017/9781316276259.003.

Witten, I. H. et al. (2017). *Data mining : practical machine learning tools and techniques*. Amsterdam; London: Morgan Kaufmann. ISBN: 9780128042915 0128042915. URL: http://www.worldcat.org/title/data-mining-practical-machine-learning-tools-and-techniques/oclc/972378340&referer=brief_results.

Wolfe, Douglas A. (2012). "Nonparametrics: Statistical Methods Based on Ranks and Its Impact on the Field of Nonparametric Statistics". In: *Selected Works of E. L. Lehmann*. Ed. by Javier Rojo. Boston, MA: Springer US, pp. 1101–1110. ISBN: 978-1-4614-1412-4. DOI: 10.1007/978-1-4614-1412-4_96. URL: https://doi.org/10.1007/978-1-4614-1412-4_96.

Wolszczan, A. and D. A. Frail (Jan. 1992). "A planetary system around the millisecond pulsar PSR1257 + 12". In: *Nature* 355.6356, pp. 145–147. DOI: 10.1038/355145a0.

Wolszczan, A. et al. (Sept. 2000). "A 25.3 Day Periodicity in the Timing of the Pulsar PSR B1257+12: A Planet or a Heliospheric Propagation Effect?" In: *ApJ* 540.1, pp. L41–L44. DOI: 10.1086/312872.

Wolszczan, Alexander (Apr. 1994). "Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257+12". In: *Science* 264.5158, pp. 538–542. DOI: 10.1126/science.264.5158.538.

Woltjer, L. (Oct. 1964). "X-Rays and Type I Supernova Remnants." In: *ApJ* 140, pp. 1309–1313. DOI: 10.1086/148028.

Xiao, Di, FaYin Wang, and ZiGao Dai (Apr. 2021). "The physics of fast radio bursts". In: *Science China Physics, Mechanics, and Astronomy* 64.4, 249501, p. 249501. DOI: 10.1007/s11433-020-1661-7. arXiv: 2101.04907 [astro-ph.HE].

Xiao, Di, Fayin Wang, and Zigao Dai (2022). "Fast Radio Bursts". In: *Handbook of X-ray and Gamma-ray Astrophysics*, 28, p. 28. DOI: 10.1007/978-981-16-4544-0_128-1.

Xu, H. et al. (Sept. 2022). "A fast radio burst source at a complex magnetized site in a barred galaxy". In: *Nature* 609.7928, pp. 685–688. DOI: 10.1038/s41586-022-05071-8. arXiv: 2111.11764 [astro-ph.HE].

Yadolah, Dodge (2008). "Kolmogorov–Smirnov Test". In: *The Concise Encyclopedia of Statistics*. New York, NY: Springer New York, pp. 283–287. ISBN: 978-0-387-32833-1. DOI: 10.1007/978-0-387-32833-1_214. URL: https://doi.org/10.1007/978-0-387-32833-1_214.

Yakovlev, Dmitrii G et al. (Mar. 2013). "Lev Landau and the concept of neutron stars". In: *Physics-Uspekhi* 56.3, 289–295. ISSN: 1468-4780. DOI: 10.3367/ufne.0183.201303f.0307. URL: http://dx.doi.org/10.3367/UFNe.0183.201303f.0307.

Yan, Zhen et al. (May 2013). "Very long baseline interferometry astrometry of PSR B1257+12, a pulsar with a planetary system". In: *Monthly Notices of the Royal Astronomical Society* 433.1, pp. 162–169. ISSN: 0035-8711. DOI: 10.1093/mnras/stt712. eprint: https://academic.oup.com/mnras/article-pdf/433/1/162/18710422/stt712.pdf. URL: https://doi.org/10.1093/mnras/stt712.

Yang, Yuan-Pei and Bing Zhang (Oct. 2021). "Fast Radio Bursts and Their High-energy Counterparts from Magnetar Magnetospheres". In: *ApJ* 919.2, 89, p. 89. DOI: 10.3847/1538-4357/ac14b5. arXiv: 2104.01925 [astro-ph.HE].

Yap, Y. X. Jane, Albert K. H. Kong, and Kwan-Lok Li (Sept. 2023). "A Light Redback Companion of PSR J1622-0315 and Irradiation Power in Spider Systems". In: *ApJ* 955.1, 21, p. 21. DOI: 10.3847/1538-4357/acea82. arXiv: 2307.13482 [astro-ph.HE].

Yi, Byoung-Kee and Christos Faloutsos (2000). "Fast Time Sequence Indexing for Arbitrary Lp Norms". In: *Proceedings of the 26th International Conference on Very Large Data Bases*. VLDB '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 385–394. ISBN: 1558607153.

Yin, Dejiang et al. (July 2024). "FAST Discovery of Eight Isolated Millisecond Pulsars in NGC 6517". In: *ApJ* 969.1, L7, p. L7. DOI: 10.3847/2041-8213/ad534e. arXiv: 2405.18228 [astro-ph.HE].

Zavlin, V. E. et al. (2000). "Discovery of 424 Millisecond Pulsations from the Radio-quiet Neutron Star in the Supernova Remnant PKS 1209–51/52". In: *The Astrophysical Journal* 540.1, p. L25. DOI: 10.1086/312866. URL: https://dx.doi.org/10.1086/312866.

Zhang, Bin-Bin, Bing Zhang, and Alberto J. Castro-Tirado (Apr. 2016). "Central Engine Memory of Gamma-Ray Bursts and Soft Gamma-Ray Repeaters". In: *ApJ*

820.2, L32, p. L32. DOI: 10.3847/2041-8205/820/2/L32. arXiv: 1603.01381 [astro-ph.HE].

Zhang, Bing (Nov. 2020). "The physical mechanisms of fast radio bursts". In: *Nature* 587.7832, pp. 45–53. DOI: 10.1038/s41586-020-2828-1. arXiv: 2011.03500 [astro-ph.HE].

— (July 2023). "The physics of fast radio bursts". In: *Reviews of Modern Physics* 95.3, 035005, p. 035005. DOI: 10.1103/RevModPhys.95.035005. arXiv: 2212.03972 [astro-ph.HE].

Zhang, Wei et al. (Nov. 2024). "Analysis of the possible detection of the pulsar wind nebulae of PSR J1208-6238, J1341-6220, J1838-0537, and J1844-0346". In: *A&A* 691, A332, A332. DOI: 10.1051/0004-6361/202348741. arXiv: 2410.21038 [astro-ph.HE].

Zhong, Shu-Qing et al. (July 2020). "On the Distance of SGR 1935+2154 Associated with FRB 200428 and Hosted in SNR G57.2+0.8". In: *ApJ* 898.1, L5, p. L5. DOI: 10.3847/2041-8213/aba262. arXiv: 2005.11109 [astro-ph.HE].

Zhu-Ge, Jia-Ming, Jia-Wei Luo, and Bing Zhang (Feb. 2023). "Machine learning classification of CHIME fast radio bursts - II. Unsupervised methods". In: *MNRAS* 519.2, pp. 1823–1836. DOI: 10.1093/mnras/stac3599. arXiv: 2210.02471 [astro-ph.HE].