

Structural Analysis of the Milky Way Galaxy A Computational Approach

A THESIS PRESENTED
BY
ANA-MARIA CONSTANTIN
TO
THE DEPARTMENTS OF ASTRONOMY AND COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF BACHELOR OF ARTS

HARVARD COLLEGE
CAMBRIDGE, MA
APRIL 2016

Structural Analysis of the Milky Way Galaxy

ABSTRACT

Mapping the 3D structure of our home galaxy has long been a challenge for the scientific community. Continuing efforts are being made for the study of the Milky Way, of its origins and evolution. There have been many attempts to mapping our galaxy and cosmic neighborhood, and while we have an understanding of its general shape, we lack precise measurements of objects situated in the galactic plane. The most exact method to date for mapping the spiral arms of the Milky Way is by identifying the position and length of bones - thin and elongated CO molecular clouds filaments. We believe these features are at the core of the spiral arms of our galaxy, and so identifying all of them would help us track the skeleton of the Milky Way in a more precise manner than previous methods. Having a map of existing bone filaments is also of interest for star formation researchers, since these filaments represent the core of the spiral arms of our Galaxy, and thus track regions of intense star formation. Previous methods for mapping the skeleton of the Milky Way have been led, and we already know the position of 11 such “bones”. We suggest using a new, computational approach for mapping other features alike, and implement a Minimum Spanning Tree (MST) algorithm to be ran over the Peretto & Fuller catalog of Infrared Dark Clouds. We believe that by treating the catalog as a Euclidean graph and running MST on this network will help us unveil new bone candidates situated in the galactic plane. By discovering new bone-like features with the minimum spanning tree algorithm, we hope to get a better intuition of where to conduct additional observations for the visual scanning of these bones. Proving the efficiency of running MST Prim’s algorithm on the Peretto & Fuller catalog is part of the larger effort of using computational techniques in order to enhance our 3D selection and 3D segmentation capabilities. These techniques can be applied not only in the context of astronomy, but also for medical research purposes, as part of the larger astronomical medicine project.

Contents

1	INTRODUCTION - WELCOME TO THE MILKY WAY!	8
1.1	Not For Tourists Guide to the Milky Way	8
1.2	Cosmic Neighborhood Overview	11
2	BONES OF THE GALAXY	20
2.1	Tracing the Skeleton of the Milky Way	21
2.2	The Importance of Bone Discoveries	21
2.3	The Story of Nessie	23
3	MINIMUM SPANNING TREE METHOD	27
3.1	Motivation	27
3.2	Theoretical Backdrop	28
4	RESULTS	34
5	DISCUSSION	46
5.1	Future Work	51
	REFERENCES	60
6	APPENDIX A	
	MOVING FORWARD - COMPUTATIONAL TECHNIQUES IN ASTRONOMY	61
7	APPENDIX B	63
8	APPENDIX C	72
9	APPENDIX D	77

Listing of figures

1.1.1	Map of the Milky Way Galaxy, William Herschel	10
1.1.2	Rotational Curve of the Milky Way Galaxy	12
1.2.1	New Spiral Arm in the Far Outer Galaxy	19
1.2.2	Core Structures of the Scutum Centaurus Arm	19
2.3.1	Nessie, Spitzer GLIMPSE mid-IR	24
4.0.1	Peretto & Fuller Catalog in Galactic Coordinates	35
4.0.2	Filament1 Prim MST	37
4.0.3	Filament1 Kruskal MST	37
4.0.4	Filament2 Prim MST	38
4.0.5	Filament2 Kruskal MST	38
4.0.6	Filament3 Prim MST	39
4.0.7	Filament3 Kruskal MST	39
4.0.8	Filament4 Prim MST	40
4.0.9	Filament4 Kruskal MST	40
4.0.10	Filament5 Prim MST	41
4.0.11	Filament5 Kruskal MST	41
4.0.12	Filament6 Prim MST	42
4.0.13	Filament6 Kruskal MST	42
4.0.14	Filament9 Prim MST	43
4.0.15	Filament9 Kruskal MST	43
4.0.16	Filament10 Prim MST	44
4.0.17	Filament10 Kruskal MST	44
5.0.1	All Generated MSTs in the Peretto & Fuller Catalog	50
5.0.2	Histogram of all Average Angles of Inclination	52
5.0.3	Average Angle of Inclination - Bones vs. All-MSTs	52

8.o.1 Filament1 GLIMPSE original image	73
8.o.2 Filament2 GLIMPSE original image	73
8.o.3 Filament3 GLIMPSE original image	74
8.o.4 Filament4 GLIMPSE original image	74
8.o.5 Filament5 GLIMPSE original image	75
8.o.6 Filament6 GLIMPSE original image	75
8.o.7 Filament9 GLIMPSE original image	76
8.o.8 Filament10 GLIMPSE original image	76

List of Tables

1.1.1	Distance to the Galactic Center, different indicators	11
2.3.1	Previously Confirmed Bone Candidates	26
5.0.1	Tree Diagnosis for Previously Confirmed Bones	49
7.0.1	MSTs detected in Filament ₁ region	64
7.0.2	MSTs detected in Filament ₂ region	65
7.0.3	MSTs detected in Filament ₃ region	66
7.0.4	MSTs detected in Filament ₄ region	67
7.0.5	MSTs detected in Filament ₅ region	68
7.0.6	MSTs detected in Filament ₆ region	69
7.0.7	MSTs detected in Filament ₉ region	70
7.0.8	MSTs detected in Filament ₁₀ region	71

Acknowledgments

I would like to thank my advisor, Professor Alyssa Goodman, for her continuous guidance and support throughout the development of my Senior Thesis. She introduced me to the world of scientific research during my Freshman Year, and has been my mentor throughout my undergraduate career. Alyssa has shaped my college experience in many ways: by getting me on board to teach astronomy to middle school students in the Boston area during my Freshman year, helping me learn about the importance of giving back and the efforts behind getting children interested in STEM; by showing me the power that lies behind combining education and technology; by introducing me to the world of astronomy software and helping me discover a passion for Computer Science, which resulted in me pursuing a Joint Concentration in Astrophysics and CS; by listening to me and giving valuable advice throughout my undergraduate years, from when I was a freshman deciding which Astronomy class to take next, to when I was a senior deciding on a full time post-graduation job. For these, and many more, I thank her.

Many thanks also to the other members of Alyssa's research group, and particularly to Catherine Zucker (Harvard PhD Student) for her continuous assistance and advice. Catherine has generously accepted to advise me in the process of writing my thesis, and she has been a role model to me throughout the past year. Thank you to the Astronomy 99 Advising Faculty, Professor David Charbonneau and Professor Alicia Soderberg. I would additionally like to thank the Harvard Department of Astronomy and the Harvard School of Engineering and Applied Sciences for teaching and training me over the course of the past three years.

1

Introduction - Welcome to the Milky Way!

1.1 NOT FOR TOURISTS GUIDE TO THE MILKY WAY

WHAT GREEK AND ROMAN MYTHOLOGIES as a splash of milk on the night sky, coming from the Gods, the Milky Way is the home galaxy of our Solar System. The “milky” part of its name was given due to its “misty patch” appearance as seen from Earth at night – the region of the sky containing the galactic plane (which in projection on our celestial sphere approximates a big circle) is seen as a dim glowing ring in powerful contrast with its dark surroundings, for which the human eye is unable to resolve individual stars.

Understanding our galaxy and our place in the universe has been a long pursued challenge for humans across the world – there are different legends describing the creation of the Milky Way in multiple cultures, and scientists worldwide have long been fascinated with the study of our galaxy. Through his telescopic observations, Galileo Galilei first suggested that the Milky Way is a large composite of individual stars, which can be observed even with the naked eye by looking at the night sky at an angle of approximately 62.6° above the celestial equator. Immanuel Kant and Thomas Wright have later argued that our galaxy is a disk-shaped stellar collection, and that our Sun and Solar System lie inside the plane of this disk. William Herschel first created a map of the Milky Way in 1785 (see **Fig. 1.1.1**), by counting all the stars and analyzing their spatial distribution as he could observe from Earth, to conclude that stars are located in a big disk formation. The 20th century then brought additional discoveries to the field of astronomy, with Edward Shapley calculating our location inside the Milky Way in 1920. By using Cepheid variable stars as distance indicators and applying the Period-Luminosity Law discovered by Henrietta Leavitt, Edward Shapley estimated distances to nearby globular clusters, which he has tracked in 683 regions of the sky. By plotting the found distances, he calculated that the Sun lies approximately two-thirds of the way out from the galactic center.

Today we believe the Milky Way is 13.21 billion years old, and that we are located at a distance of 8kpc (calculated using different methods, see summary in **Table 1.1.1** below) from the Galactic Center. A wealth of new techniques and data analysis methods for space exploration have been developed since the early scientific efforts mentioned previously, yet we still do not have a precise 3D map of our galaxy. There are multiple advantages to

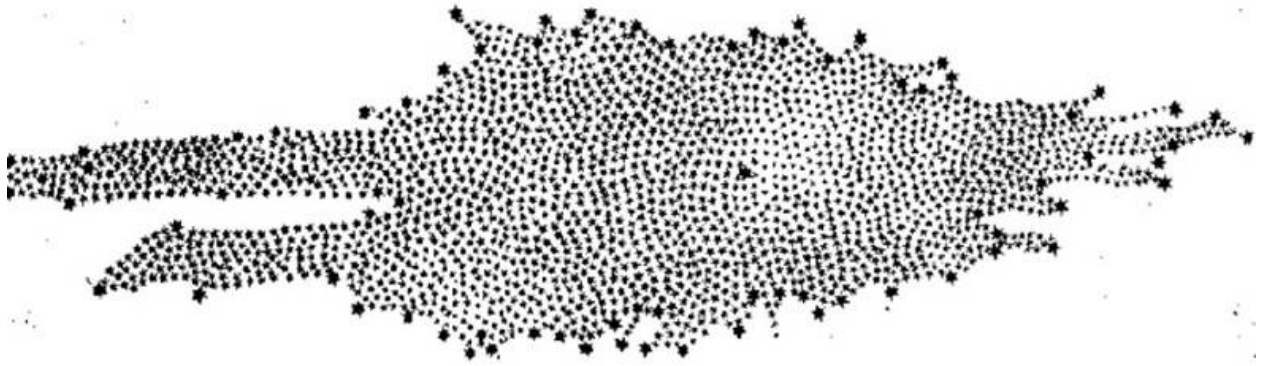


Figure 1.1.1: Map of the Milky Way Galaxy, as designed by William Herschel (Yerkes Observatory).

building one, ranging from the large spatial resolution of observations conducted in our galactic neighborhood (which allows us to collect detailed information about specific types of objects), to the possibility of discovering and testing theories on the Milky Way which we could expand for the future study of other galaxies.

Unfortunately we encounter a number of challenges when trying to create an accurate 3D map of the Milky Way, mainly due to our location in a vantage point of the galactic plane. One of the most popular methods for mapping our galaxy has been the one of rotational curves (see **Fig. 1.1.2** below), in which we infer the velocity of an object and convert it into position – this method gives results which are highly volatile, since the environment in which we are observing velocities is a medium filled with varying potentials and winds coming from supernova remnants, stellar winds, and HII regions ([19] Heyer et al 2015).

RR Lyrae Variables	8.0 ± 0.5 kpc
Globular Clusters	8.0 ± 0.8
Miras	7.9 ± 1.0
Cepheids	8.0 ± 0.5
OB Stars	9.1 ± 1.0
HI and HII regions	8.1 ± 0.5
H ₂ O proper motion	7.2 ± 0.7
OH/ IR stars	8.1 ± 1.1
Nearby stars/ Oort constants	8.9 ± 1.0
X-Ray Sources	7.4 ± 1.0
Planetary Nebulae	7.6 ± 0.7
Red Giants	7.9 ± 1.0
Sgr A* Proper Motions	7.7 ± 0.9
Weighted Average	8.0 ± 0.5

Table 1.1.1: Distance to the Galactic Center using different indicators (data inferred from Reid et al. (1993))

1.2 COSMIC NEIGHBORHOOD OVERVIEW

1.2.1 METHODS AND TRACERS

Multiple efforts to map the cosmic neighborhood of our Solar System have been conducted thus far, and the rotational velocity curve method has been exploited by [8] Dame et al 2001 to create a catalog of CO molecular clouds, by [37] Shane 1972 in order to map HI regions inside the Milky Way, and by [4] Reid et al 2014 to analyze the results of BeSSeL (the Bar and Spiral Structure Legacy Survey) for studying the spiral structure and kinematics of the galaxy. Numerous surveys have also been conducted such that we have more detailed observations of the Milky Way – in this paper, we are particularly interested in analyzing results coming from surveys monitoring CO and HI molecular clouds. Given the large distribution of CO and HI molecules inside the galaxy, studying the molecular ISM (Interstellar Medium) and measuring distances to different molecular clouds not only

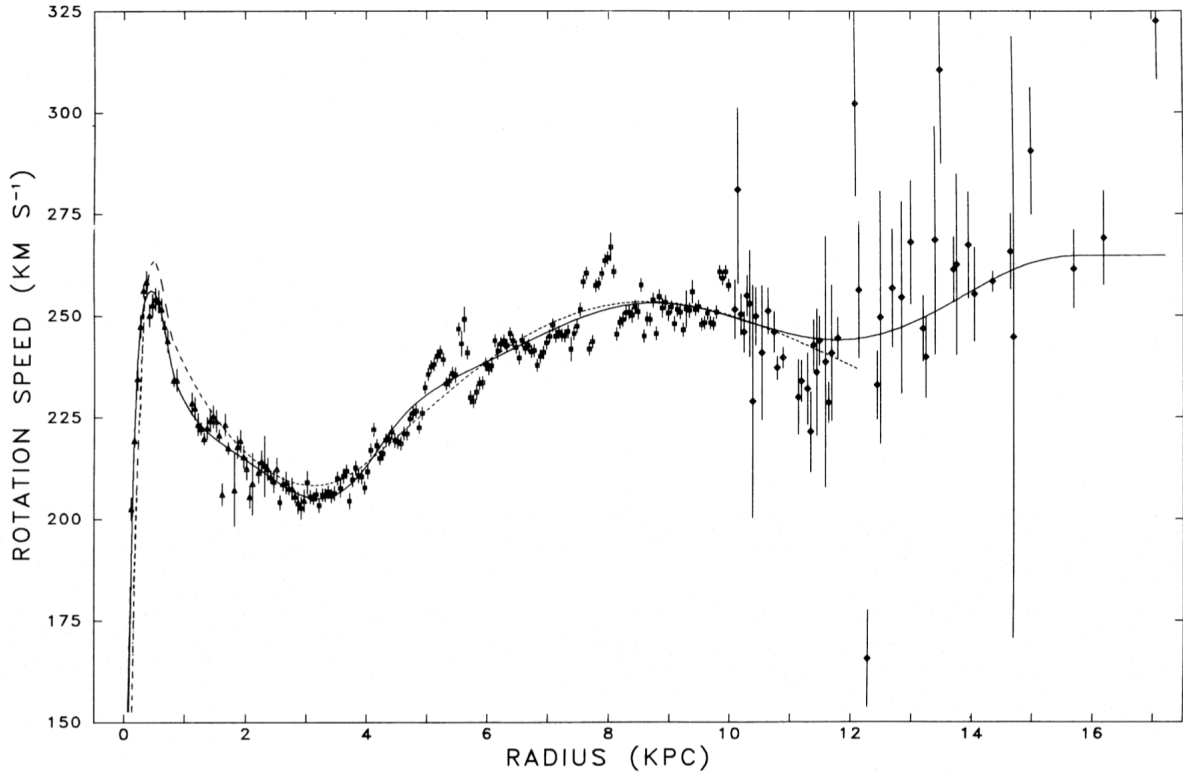


Figure 1.1.2: Rotational Velocity Curve of the Milky Way Galaxy (courtesy of [5].

helps us in tracing a map of the Milky Way, but also into possessing a better understanding of molecular star-forming regions and of molecular circumstellar disks of newborn stars. Additionally, observations of molecular clouds are important for studying Galactic properties at multiple wavelengths: CO clouds absorb in the X-ray domain ([40] Wang Yu 1995), emit in the Gamma-Ray space ([1] Ackermann et al. 2012), and they can be observed under Optical wavelengths as dark nebulae ([23] Lombardi et al. 2011), while their dust components emit in IR (**Planck Collaboration XXIII et al. 2014**) and their internal star forming regions represent strong sources of radio emissions.

1.2.2 DOWNTOWN AREA – SURVEYING THE GALACTIC PLANE

Most of the mass of the galaxy is comprised within the galactic plane, and the majority of our observations are also coming from this region. We define the galactic plane region as the part of the sky contained within $\pm 5^\circ$ in galactic latitude. There are multiple surveys of molecular clouds concentrated in this range of latitudes – the most recent and accurate one was developed by [8] Dame et al (2001), mapping the interstellar carbon monoxide (CO), also known as the best tracer for the largely invisible molecular hydrogen. H_2 is of particular interest to astronomers since most of the mass of molecular clouds is concentrated in the form of these molecules. While the [8] Dame et al (2001) catalog includes data observed uniformly across the galactic plane, in general the first quadrant of the galaxy has been more extensively studied than other parts. [19] Heyer and Dame (2015) mention there was a 20 year gap until observations conducted in the third quadrant reached a level of accuracy of better than 0.5° angular resolution.

1.2.3 SUBURBIA – HIGH-LATITUDE GALACTIC SURVEYS

Observations outside of the galactic plane are even more difficult to conduct, since we lack the bright background which allows us to spot molecular clouds due to their absorption properties. [18] Hartmann et al. (1998) mapped the Northern hemisphere at latitudes $b > 30^\circ$, while [24] Magnani et al (2000) mapped the Southern equivalent of $b < 30^\circ$, both of them affiliated with the Harvard Center for Astrophysics. More recent observations were conducted by the *ESA Planck Mission* – while targeting the Cosmic Microwave

Background, 6 of the 9 bands of data collection for Planck have been able to reach precisions as close to the ones calculated by Magnani et al (2000). Even though Planck's data is one of the newest efforts in the field, we view the results it collected as inaccurate since they do not contain kinematic information, and are likely affected by the Sunyaev-Zel'dovich effect as well.

1.2.4 OVERALL DISTRIBUTION OF CO INSIDE THE MILKY WAY

It has been challenging to draw a model for the overall distribution of molecular clouds inside the Milky Way, since the rotational velocity method is introducing uncertainties due to the kinematic distance approach. Not only are we measuring velocities in an environment that is highly exposed to stellar winds and fronts coming from supernova remnants (which induce non-circular motions), but we are also observing multiple objects along the same line of sight (which makes it difficult to narrow down a specific molecular component and assign to it a particular velocity/distance).

When analyzing the distribution of CO clouds in the galaxy, we are more interested in the GMCs located inside the galactic plane than in the ones situated in the halo. We learn from the surveys described above that CO molecular clouds are more ubiquitous within $\pm 5^\circ$ of galactic latitude, and from the [19] Heyer and Dame (2015) study that there are also differences in the intrinsic CO emission properties which come as a consequence of the cloud location. Thus, clouds which are situated in the outer Galaxy are ~ 7 times fainter than the ones from the galactic plane, and those in the inner Galaxy can be up to ~ 60 times fainter than the ones close to the galactic center. This was first suggested by [9] Digel

et al. (1996), who explained the low emissivity of CO in the outer galaxy by suggesting a correlation between CO emission and gas temperatures and metallicities. [27] Narayanan et al (2012) prove that outer galactic clouds have lower temperatures and metallicities. Another explanation of this effect, suggested by [19] Heyer and Dame (2015), is that environmental conditions at high galactic latitudes enforce lower mass surface densities on the molecular clouds situated in those locations.

For purposes of this thesis, we are interested in tracking the distribution of a particular subset of CO molecular clouds located within the galactic plane, hoping to gain a better understanding of the fundamental structures of our galaxy. In particular, we want to use existing data from surveys and catalogs in order to automatize the process of mapping our galactic neighborhood. We are focused on understanding what are the core features of the Milky Way galaxy, and whether they can be revealed through a computational approach.

1.2.5 PROPERTIES AND STRUCTURE OF THE GALAXY – AS KNOWN TODAY

Due to our position in a vantage point of the galactic plane, it is difficult to conduct observations of our own galaxy. We can thus infer some of the properties of the Milky Way by analyzing similar spiral galaxies (preferably face-on and perpendicular to our line of sight such that we can observe the spiral structure) and drawing comparisons between them. By following this approach, we learn that there is a strong correlation between the distribution of CO molecular clouds and the structure of a galaxy. [28] Schinnerer et al. (2013) conduct interferometric observations of the M51 Galaxy in order to quantitatively describe the correlation between the distribution of CO molecular clouds and other tracers of the

interstellar medium and of star-forming regions. By using data supplied through the *Plateau de Bure Interferometer Arcsecond Whirlpool Survey*, they prove that molecular clouds and regions of star formation are distributed as a function of the underlying gravitational potential.

The connection between molecular clouds and the spiral structure of galaxies thus makes them an interesting area of research for purposes of this thesis. In addition to the above mentioned surveys, astronomers have also analyzed the CO molecular cloud components of the Milky Way on a case by case method. Several initiatives were launched also to resolve for the uncertainties induced through the kinematic distance method, among them the *Very Large Array Galactic Plane Survey* (which observes in the first quadrant of the galaxy) and the *Southern Galactic Plane Survey* (which covers the fourth quadrant). Molecular clouds found in these two quadrants have been particularly useful in tracing down the spiral structure of our galaxy: [26] Dame et al. (1986), [38] Solomon Rivolo (1989), and [33] Roman-Duval et al. (2010) succeeded in mapping the Sagittarius and Scutum arms located in the first quadrant, while [16] Grabelsky et al. (1988), and [14] Garcia et al. (2014) did so for the Carina, Centaurus and Norma arms in the fourth quadrant.

While these results help us define the shape and location of the spiral arms to a certain degree, they are highly ambiguous since the data they are calculated from suffers two transformations from the real galactic distribution to an XY plane, and then back to the real, initial one. This approach is called the Double Gaussian method and it is not a popular one among scientists since it is possible that through the double transformation of coordinates some of the details of the spiral arm shapes are lost. [6] Combes (1991) proves this for the

case of the M_{51} Galaxy.

The question of precisely mapping the spiral arms of the Milky Way has been a long studied one, and a continuous challenge for the astronomy community. To this day, we do not know if we are located in a spiral arm, or in a spur between two arms. Astronomers have been able to pick the major spiral features through CO observations, and yet we still have difficulties in defining and mapping the precise location of the core of these spiral arms.

The most exact calculations to date for the spiral structures of the galaxy still involve uncertainties of ~ 10 distance from the galactic center. These results have been obtained by tracing and analyzing distributions of CO molecular clouds from existing catalogs: we were thus able to identify and approximately locate some of the most prominent features of the galaxy. The shape of the Sagittarius arm is described by [41]Wu et al. (2014), the Scutum arm by [36]Sato et al. (2014), the Outer Spiral Arm was explored by both [17]Hachisuka et al. (2009), and [35]Sanna et al. (2011), the Local Arm is the subject of [42]Xu et al. (2013), the Perseus one of [43]Zhang et al. (2013), while the Expanding 3-kpc Arm is featured by [34]Sanna et al. (2009). In addition to these, [7]Dame et al. (2011) claim to have found an additional spiral arm (**Fig 1.2.1** below) in the far outer galaxy, also through the method of tracing CO molecular agglomerates. This feature was found beyond the Outer Arm in the first quadrant, at a distance of ~ 15 kpc from the galactic center, and we believe it is a continuation of the Scutum – Centaurus Arm in the outer galaxy space, in addition to being a symmetric counterpart of the Perseus Arm.

Most of the face-on views we have of the galaxy are not accurate since we cannot identify the location of the observed spiral shapes with precision. As noted in [44] Zucker et al.

(2015), different calculations involving CO clouds and the kinematic distance approach give different locations for the core areas of the galactic arms. **Fig 1.2.2** below shows how different measurements of the core structures inside the Scutum – Centaurus Arm lead to different results, even though all papers analyzing these region have used the same scientific method throughout their research process.

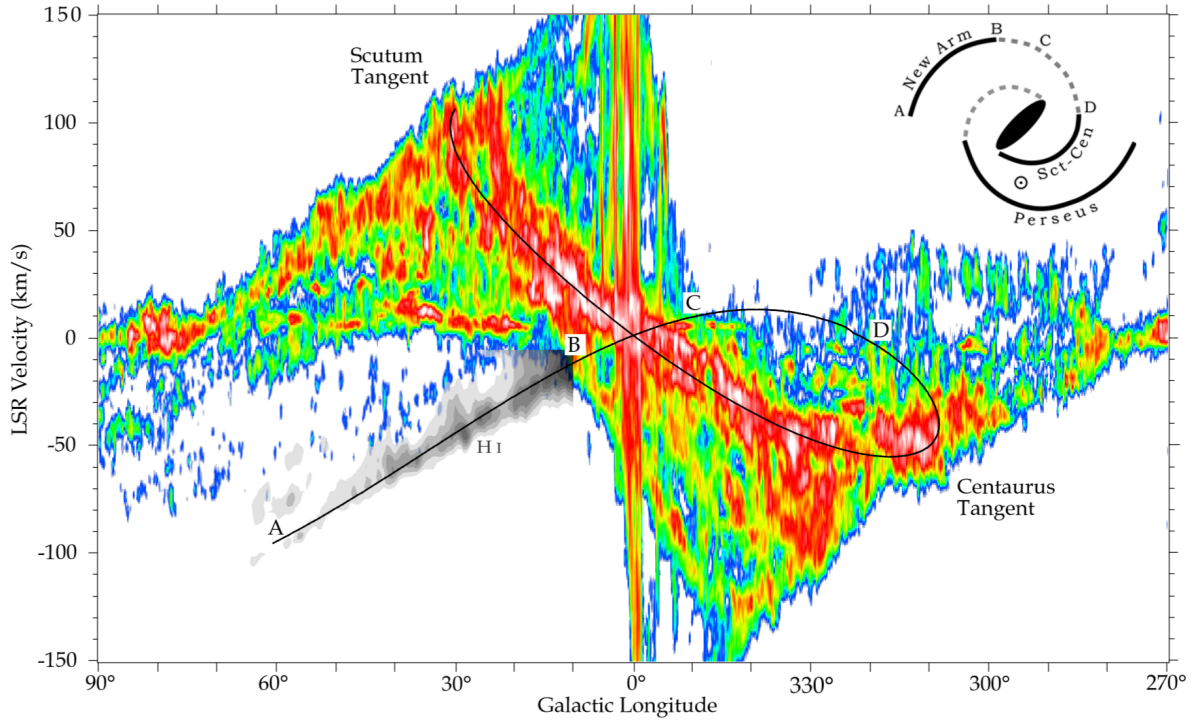


Figure 1.2.1: Spiral Arm found by [7] Dame et al (2011).

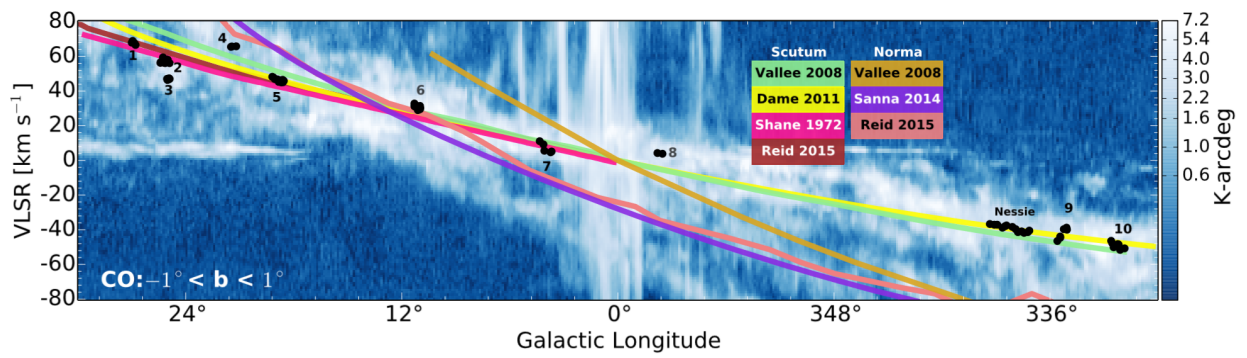


Figure 1.2.2: Scutum Centaurus Arm, picture credit to [44] Zucker et al. (2015).

2

Bones of the Galaxy

ALTHOUGH MULTIPLE RESEARCH EFFORTS have been conducted in order to map the 3D structure of the Milky Way we are still lacking a precise and accurate view of our galaxy, and we still have many fundamental questions about its structure to be yet answered. To this day, it is not fully understood what causes the spiral shape of the Milky Way, how many major spiral arms our galaxy has, what is the precise location of these arms and their internal structure, or what is the curvature of a given arm.

Accurately mapping the Milky Way remains an on-going challenge for the astronomical

community, mainly because of our position inside the plane of the galaxy, and thus difficulty in creating a face-on view based on current observations.

2.1 **TRACING THE SKELETON OF THE MILKY WAY**

Recently made discoveries indicate similarities between the fundamental structures of known spiral arms, and the presence of long filaments of Infrared Dark Clouds (IRDC) at the core of these arms. [15] Goodman et al. (2014) launch a conversation on the “skeleton” of the Milky Way, and announce the discovery of a first “bone” of our galaxy, a highly elongated IRDC which lies in the mid-galactic plane and within the Scutum-Centaurus spiral arm. Further observing and exploring the physical properties of this IRDC filament (named Nessie by [21] Jackson et al.(2010)) helps us define the characteristics of a new class of objects called “Bones” of the galaxy, through which we will be able to describe the structure of the Milky Way in a more precise way than has ever been done before.

2.2 **THE IMPORTANCE OF BONE DISCOVERIES**

Much of our current understanding of the internal structure of the Milky Way has been done by comparing our galaxy with external ones, particularly with those which can be observed face-on from Earth. Our location inside the galactic plane makes it difficult for us to trace the precise coordinates of other objects hosted within the Milky Way, and yet several methods have been developed for approximating (to a certain degree) a 3D galactic map. Rotational velocity curves proved to be particularly useful since they combine the rotational pattern of the Milky Way with line-of-sight velocities in order to obtain the

distance to a given object. Applying this method to existing spectroscopic surveys allowed us to map the location of most of the Milky Way's molecular, ionized, and atomic gases. In particular, the identifying probes for calculating distances within our galaxy have been spectroscopic observations of HI and CO, which are found in abundance in this environment. We have thus been able to create catalogs such as the [8] Dame et al. (2001) for CO molecular clouds or [37] Shane (1972) for HI regions, however the distances measured through this method are inaccurate at times, due to the fact that HI and CO are frequently found in astronomical observations, and thus multiple traces of this kind can overlap on our line-of-sight and lead to erroneous calculations. Other methods which have been explored include trigonometric measurements of parallaxes and proper motion of masers (such as the *BeSSeL survey* analyzed by [4] Reid et al. (2014)). Similar initiatives have been led on the *Bolocam Galactic Plane Survey*, on the *Millimeter Astronomy Legacy Team 90 GHz Survey*, on the *H₂O Southern Galactic Plane Survey*, and on the *ATLASGAL* follow-up spectral line survey. All these efforts were led in order to calculate the spectroscopic velocities of dense gases found in molecular clouds and in high-mass star forming regions.

Alternative approaches to the rotational velocity curve method have included the observation of young open clusters and of Classical Cepheids as distance indicators. While these two methods cannot predict accurate distances within the galactic plane, they are excellent for proving that the arms of the Milky Way follow a spiral shape. In particular, [39] Vazquez et al. (2008) used this method in order to confirm that Cygnus, the Outer Arm of the Galaxy (located in the third quadrant) has a spiral design, while [25] Majaess et

al. (2009) proved the same result for the Sagittarius-Carina major arm.

All the above practices predict that the shape of our galaxy is a spiral one and further approximate curvatures for the Cygnus, Perseus, and Sagittarius-Carina arms. While initial calculations predicted that the Milky Way is a galaxy with two spiral arms ([20] Jackson et al. (2008), [12] Francis Anderson (2009), [10] Dobbs Burkert (2012)), more recent inquiries indicate the presence of four arms ([32] Reid et al. (2009), [2] Bobylev Bajkova (2014), [11] Urquhart et al. (2013)). Thus, in addition to the three arms mentioned above, we have been able to trace the core location of the Scutum-Centaurus Arm through the observation of Nessie, a long IRDC filament found in the mid-galactic plane which is the first of its kind object discovered to be a “bone” of our galaxy.

We turn to the method of observing Nessie-like Infrared Dark Clouds (IRDCs) since their density is high enough that they can be observed against other objects lying in the galactic plane, by following extinction patterns in the mid-infrared range. Our location slightly above the galactic plane (re-calculated since the IAU established the current system of galactic coordinates in 1959) offers us valuable insight into observing Nessie-like features and into tracking IRDC filaments – “bones”. This is powerful because observing bones can not only help us track the skeleton of the Milky Way, but also analyze regions which will potentially become spots of high-mass formation.

2.3 THE STORY OF NESSIE

First discovered by [21] Jackson et al (2010), Nessie was presented as the first known bone of the Galaxy by [15] Goodman et al (2014). This thin and highly dense filamentary cloud

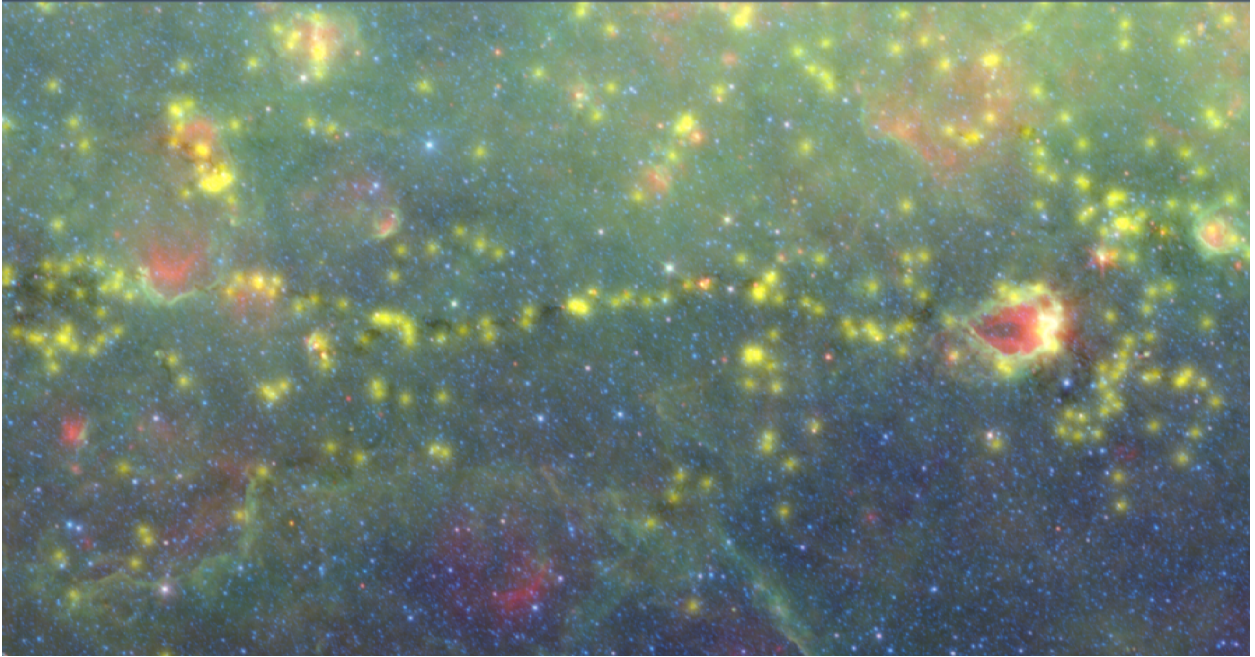


Figure 2.3.1: Nessie, Spitzer GLIMPSE mid-IR view (figure adapted from [15] Goodman et al. (2014).

(see Fig. **Fig. 2.3.1**) spans a length of 38 degrees and its position follows the core of the Scutum-Centaurus arm of the Milky Way. Like all IRDCs, Nessie describes a dusty area of the sky which is highly dense and which can be observed against the bright mid-infrared background of the galactic plane, mainly as a long dark extinction feature.

The discovery of Nessie as a bone of the Milky Way motivated researchers to further investigate for similar features along the plane of the galaxy, such that we can precisely trace the position of all existing spiral arms. [44]Zucker et al (2015) extends the search for bones inside the galaxy and announce the discovery of 10 additional Nessie-like filaments. Their method is based on the visual exploration of the Milky Way in l , b , and v space, and on overlaying the predicted positions of the four arms of the galaxy (Scutum-Centaurus, Carina-Sagittarius, Norma, and Perseus) on images belonging to the Spitzer GLIMPSE

Survey in WorldWide Telescope. Their set of criteria for whether a found filament is a bone of the galaxy or not is as follows:

1. Largely continuous mid-infrared extinction feature
2. Parallel to the Galactic Plane, to within 30deg
3. Within 20pc of the physical Galactic mid-plane, assuming a flat galaxy
4. Within 10 km/s of the global-log spiral fit to any Milky Way arm
5. No abrupt shifts in velocity (of more than $3 \text{ km}/(\text{s} * 10\text{pc})$) within extinction feature
6. Projected aspect ratio $> 50: 1$.

According to the criteria above, the method used by [44]Zucker et al (2015) extended the current number of known bones by a number of 10, with detailed information about each of them being summarized in Table **Table 2.3.1** below.

Name of Filament	Referenced Name	Length ($^{\circ}$)	Range in Galactic Long.	Range in Galactic Lat.
BCo26.94-0.30	Filament 1	0.16	26.86 - 27.02	[0.2, -0.8]
BCo25.24-0.45	Filament 2	0.63	24.608 - 25.555	[0.05, -0.95]
BCo24.95-0.17	Filament 3	0.18	24.608 - 24.959	[0.33, -0.67]
BCo21.25-0.15	Filament 4	0.2	21.15 - 21.35	[0.35, -0.65]
BCo18.88-0.09	Filament 5	0.7	18.53 - 19.23	[0.41, -0.59]
BCo11.13-0.12	Filament 6	0.38	10.94 - 11.32	[0.38, -0.62]
BCo04.14-0.02	Filament 7	0.69	3.795 - 4.485	[0.48, -0.52]
BC357.62-0.33	Filament 8	0.4	357.42 - 357.82	[0.17, -0.83]
BC335.31-0.29	Filament 9	0.6	335.01 - 335.61	[0.21, -0.79]
BC332.21-0.04	Filament 10	0.9	331.76 - 332.66	[0.46, -0.54]

Table 2.3.1: Previously Confirmed Bone Candidates (data adapted from [44] Zucker et al. (2015))

This thesis focuses on exploring “smart” methods for the discovery of bone filaments of the Milky Way, with the goal of extending the current list of bones to areas that cannot be easily explored through the visual methods which are currently being used.

In particular, we focus on applying a Minimum Spanning Tree algorithm to the clouds cataloged by [13] Peretto Fuller (2009), guided by an overall strategy of connecting dots. [15] Goodman et al. (2014) suggest this method as a future approach for the discovery of bones which cannot be observed visually, and argue that by plotting the Peretto Fuller catalog of 11,000 high-density clouds, Nessie and other thin long IRDCs appear as easily connectable sources. Since Nessie itself is comprised of approximately 100 clouds present in the [13] Peretto Fuller catalog, we hope that analyzing this dataset will reveal additional structures that can fit the criteria for being considered bones of the Milky Way.

3

Minimum Spanning Tree Method

3.1 MOTIVATION

THIS CONNECT-THE-DOTS APPROACH was first tested by [30] Gutermuth et al (2009) in the context of cluster core extraction and basic structural analysis of Young Open Clusters included in the Spitzer Survey, which are found within 1 kpc of the Sun.

Based on these results, [15] Goodman et al (2014) suggest using a similar approach for

identifying long, thin dense filaments in the [13] Peretto & Fuller catalog such that we can discover more Nessie-like structures which lie in the plane of the galaxy, and we build a more complete mapping of the skeleton of the Milky Way.

Minimum Spanning Tree algorithms can be applied on weighted, undirected graphs with the purpose of connecting its vertices together into a path that minimizes the sum of total weights on its edges. Briefly put, it would calculate the minimum path between several points belonging to a given graph, from a weighing perspective. In order to apply it to our current problem, we thus need to treat the Peretto & Fuller catalog as a connected, weighted, undirected graph, with each vertex/ node corresponding to a particular cloud.

3.2 THEORETICAL BACKDROP

3.2.1 GRAPH THEORY OVERVIEW

The Minimum Spanning Tree (MST) algorithm can be applied to connected, weighted graphs consisting of unordered pairs of vertices. The properties of such a graph are thus:

- $G(V, E)$ connected, unordered
- V = set of vertices
- $|E| = \mathcal{O}(V^2)$, $|E| \leq V \times V$

A spanning tree of G will be a subset $T \subseteq E$ such that T is acyclic and it connects all vertices of G . For a connected and acyclic $G(V, E)$, a spanning tree will thus satisfy the relation $|E| = |V| - 1$.

3.2.2 WHAT IS A MINIMUM SPANNING TREE

A graph can have multiple spanning trees, yet the minimum spanning tree will be the one connecting all points while minimizing the total weight of its component edges. Thus, if we consider an edge $(u, v) \in E$, the cost $w(u, v)$ describes the weight associated with that edge. The total weight of a graph's MST will thus be:

$$w(T) = \min(\sum_{(u,v) \in T} w(u, v))$$

Minimum Spanning Trees are constructed by consecutively adding edges of minimum weight to some existing subset S of an MST. The Cut Property of Minimum Spanning Trees guarantees that this method outputs a tree of minimum weight for the graph.

Cut Property of MSTs

If $S \in T$, with T being an MST for the initial graph $G(V, E)$, if $U \in V$ is a subset such that no edge in S crosses between U and $(V - U)$, then for a given edge e with one endpoint in U and one endpoint in $(V - U)$, $T + e = T'$, T will also be an MST for $G(V, E)$ given that e has minimum weight among all edges connecting the two subsets of vertices.

3.2.3 GREEDY APPROACH, PRIM'S & KRUSKAL'S ALGORITHM

The above Cut Property indicates that we can build Minimum Spanning Trees through a Greedy Approach, therefore choosing a locally optimal solution in order to reach a global optimal one in the end. This method is explained in the [31] Cormen, Leiserson, Rivest and Stein book "Introduction to Algorithms": by first initializing A as a subset of some minimum spanning tree of $G(V, E)$, at each iteration we keep adding additional edges

(u, v) to A such that $A \cup (u, v)$ will also be a subset of a minimum spanning tree of $G(V, E)$.

The pseudocode for this method can thus be summarized through the following steps:

Greedy-MST(G):

1. $A = \emptyset$
2. *while* ($A \neq \text{spanning tree of } G$):
3. *find* $e(u, v)$ *of minimum weight to connect* A *to* $(G - A)$
4. $A = A \cup e(u, v)$
5. *return* A

All edges $e(u, v)$ which can be added in step (3) above are called safe edges, and different algorithms have been developed in order to identify them. A safe edge can only be selected from among multiple light edges, where a light edge that satisfies a property is defined as the minimum weight edge which satisfies that property.

In particular, the theorem which needs to remain valid throughout the implementation of a minimum spanning tree algorithm is the following, as explained by [31] Cormen et al. in their “Introduction to Algorithms” (Theorem 23.1):

$G(V, E)$ is a connected, weighted, undirected graph, and A is a subset of E such that $A \in \text{some MST of } G$. For a cut $(S, V - S)$ of G (such that no edge in A crosses this cut), if $e(u, v)$ is a light edge crossing $(S, V - S)$ then $e(u, v)$ is a safe edge for A .

Prim's Algorithm for MST Detection:

I implemented Prim's Algorithm for identifying minimum spanning trees present in the Peretto & Fuller catalog of Infrared Dark Clouds. This algorithm is a particular case of the Greedy method explained above, and it introduces a method for identifying safe edges in step 3 of *Greedy* — $MST(G)$.

Prim's Algorithm functions by identifying shortest paths in a given graph, which is convenient for our case of detecting MSTs in a collection of points situated in the galactic plane, since we are treating the [13] Peretto & Fuller catalog as a graph of connected, undirected vertices which are united by edges whose weights are equivalent to the Euclidean distance (in angular units). Prim's Algorithm is initialized by taking the graph G and a root node r of the MST as inputs, and placing all other vertices outside of the MST in a min-priority queue Q . We track each of these vertices via two attributes: $v.key$ (in which we store the value of the min-weight edge connecting v to the MST), and $v.parent$ (through which we keep track of the parents of v inside the MST). We consider $v.key = \infty$ for any node which is not connected to the MST.

The pseudocode for this method can thus be summarized through the following steps, as described by [31] Cormen et al (1991).:

1. ***def PRIM-MST (G,r):***
2. ***for*** each $u \in V$:
3. $u.key = \infty$
4. $u.parent = NULL$

```

5.       $r.key = 0$ 

6.       $Q = V$ 

7.      while ( $Q \neq \emptyset$ ):

8.           $u = \text{min} - \text{pop}(Q)$ 

9.          for each  $v \in u.neighbors$ :

10.             if  $v \in Q \ \&\& \ w(u, v) < v.key$ :

11.                  $v.parent = u$ 

12.                  $v.key = w(u, v)$ 

```

Kruskal's Algorithm for MST Detection:

Unlike Prim's Algorithm, the Kruskal approach is initialized by treating the graph as a forest with unconnected vertices, which will eventually become tied to one another by the time algorithm terminates. At any step, the algorithm is looking for a safe edge to add to the forest, such that the edge is of minimum weight, and does not create a cycle inside the forest.

```

1.      def KRUSKAL-MST ( $G, r$ ):

2.           $A = \emptyset$ 

3.          for each  $u \in V$ :

4.              MAKE-SET( $u$ )

5.          sort edges non-decreasingly by weight  $w$ 

```

6. **for** each $e(v, u) \in E$:
7. **if** $FIND - SET(v) \neq FIND - SET(u)$
8. $A = A \cup e(v, u)$
9. $UNION(v, u)$
10. **return** A

Joseph Kruskal developed his algorithm in 1956 ([22]), and Robert C. Prim designed his solution in 1957. Both approaches solved the problem of interconnecting the nodes of a network while minimizing the shortest possible sum of lengths of its direct links. In his initial paper, [29]Prim (1957) has proved his Shortest Connection Network algorithm by using an example involving a network of cities across the United States and assigning each edge between different cities a weight equal to their corresponding Euclidean distance. In a similar fashion, we are applying this algorithm in order to find shortest connection networks among the [13]Peretto & Fuller catalog of Infrared Dark Clouds situated in the plane of the Galaxy, where we are considering each cloud to be a node in the network, and the edges between clouds to have weights equal to the Euclidean distance between their connecting vertices.

4

Results

THE PERETTO & FULLER CATALOG, plotted in galactic coordinates before applying the MST Algorithm, can be found in **Fig. 4.0.1**. I treated this collection of points in 2D as a Euclidean graph of unconnected nodes, and I connected those vertices which are separated by an angular distance less than a given threshold. I experimented with thresholds less than 0.5° , and by running my algorithm for thresholds ranging from 0.05° to 0.5° I concluded that an optimal threshold is 0.1° .

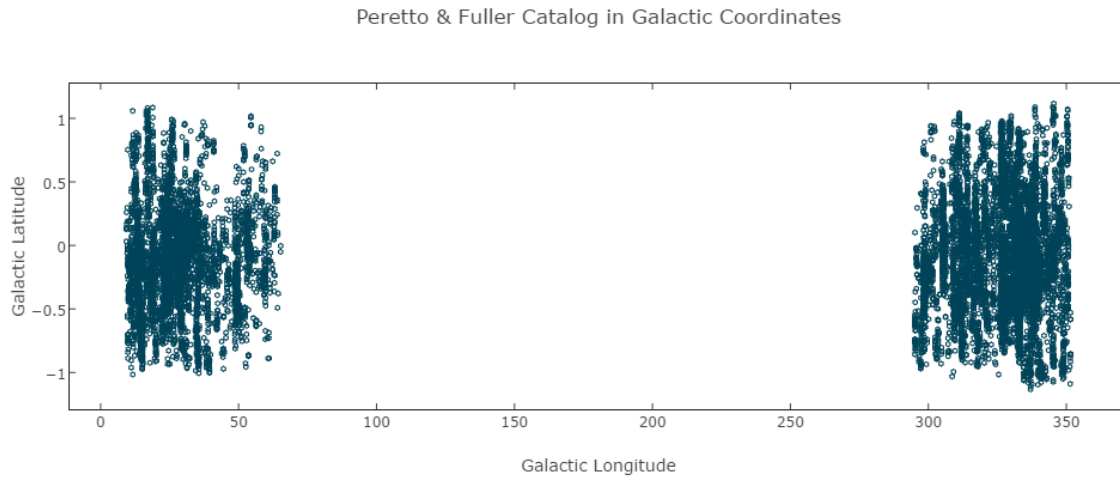


Figure 4.0.1: Peretto&Fuller Catalog plotted in Galactic Coordinates.

I created edges between the points separated by a distance below this threshold and thus obtained a Euclidean, weighted, undirected graph. I then ran both Prim's and Kruskal's algorithm for Minimum Spanning Tree detection on the resulting graph. To summarize, the overall strategy of my approach can be explained through the steps below:

1. Treat Peretto & Fuller catalog as a Euclidean graph
2. Connect points if they are closer than a given threshold: 0.1 degrees
3. Apply MST on resulting Euclidean graph
4. Mark filament positions with boxes and check if tree shape matches
5. Repeat above procedure for both Prim & Kruskal
6. Run Tree Diagnostics on results

In order to test the accuracy of this method, I applied the MST algorithm on the zoomed-in regions defined by Nessie and the previously studied bone candidates (as defined in [44]Zucker et al (2015)). All of these filaments can be found within $10 < |l| < 60^\circ$ Galactic Longitude, and within $|b| < 1^\circ$ Galactic Latitude. Their known positions and lengths are the ones which I have explained in **Table 2.3.1**.

I used these coordinates when plotting my resulted Minimum Spanning Trees using both Prim and Kruskal Algorithms. The images obtained can be viewed below, for the Nessie region and for the already known bone candidates. We are interested in seeing if the outputted MSTs match the positions of the bone candidates (marked with boxes in the overlaid figures). I am including below both the original GLIMPSE files (as presented by [44]Zucker et al (2015), see **Appendix C**), and similar ones with overlaid Minimum Spanning Trees on top (obtained through both Prim and Kruskal methods). We observe that the shape of resulting minimum spanning trees matches the extinction features present in the original images.

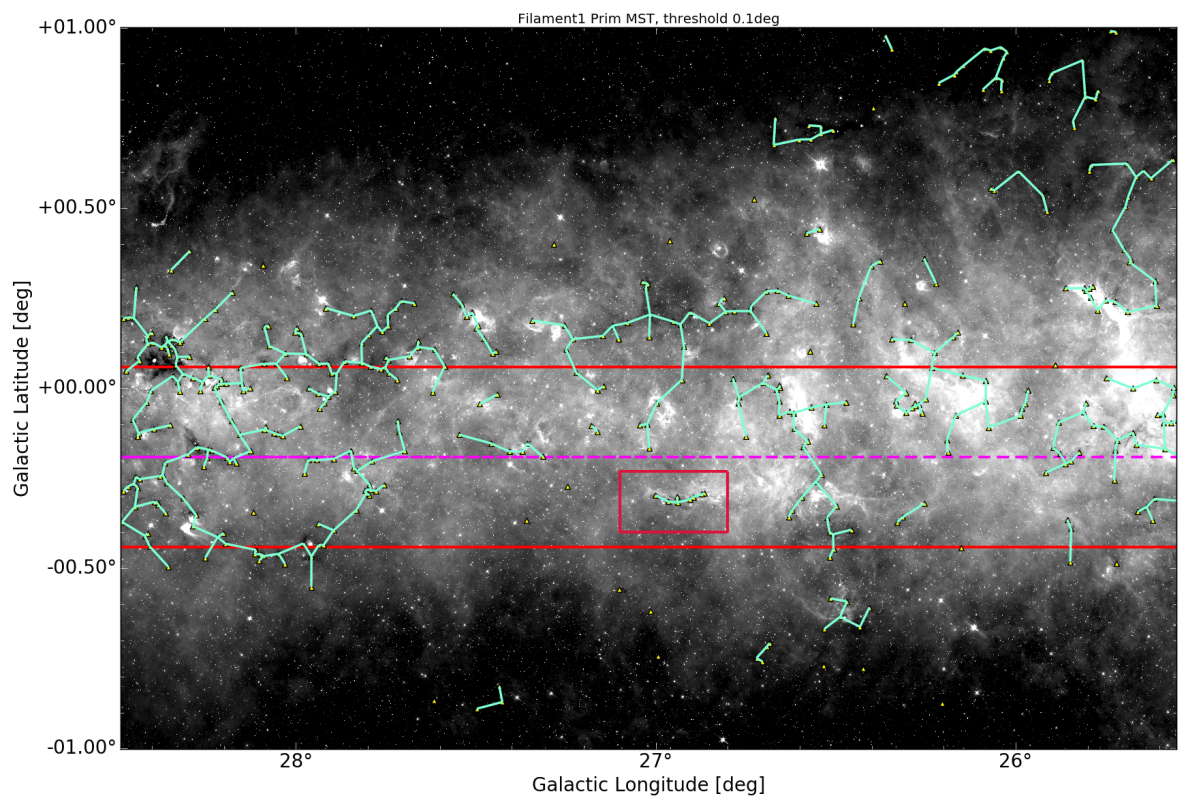


Figure 4.0.2: Filament1 Prim MST.

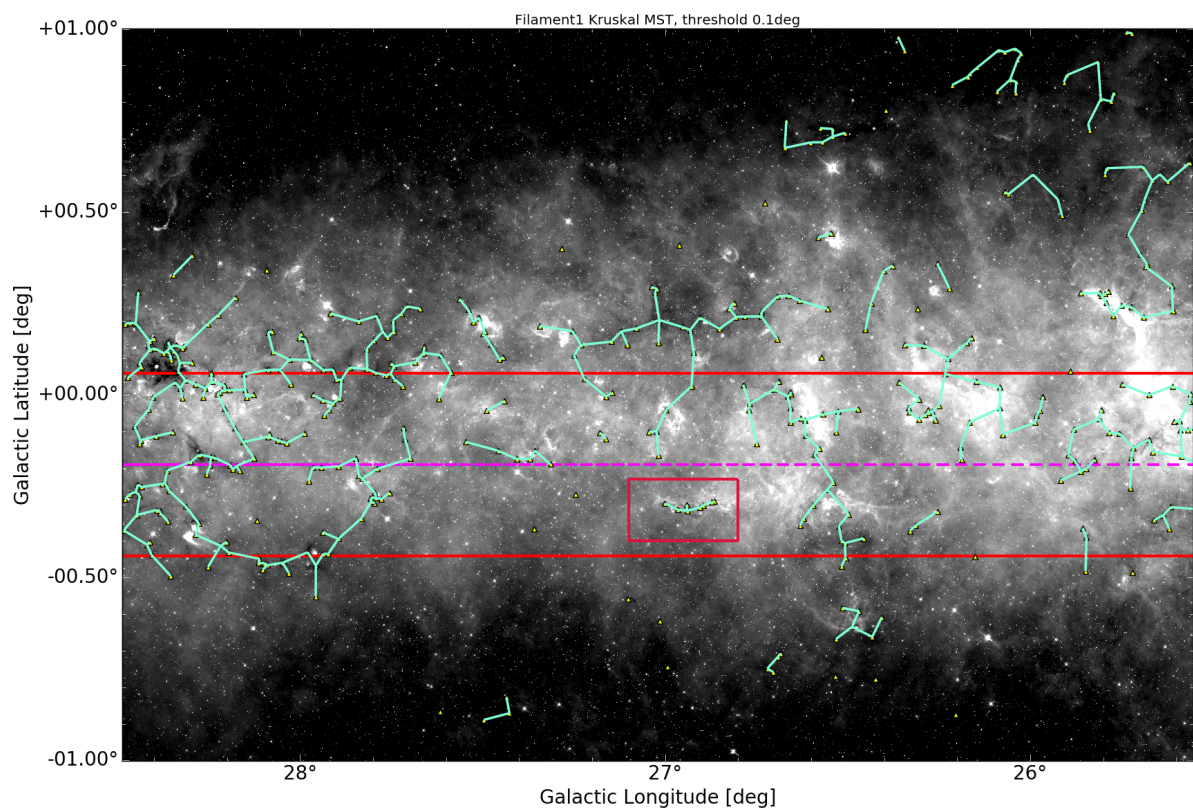


Figure 4.0.3: Filament1 Kruskal MST.

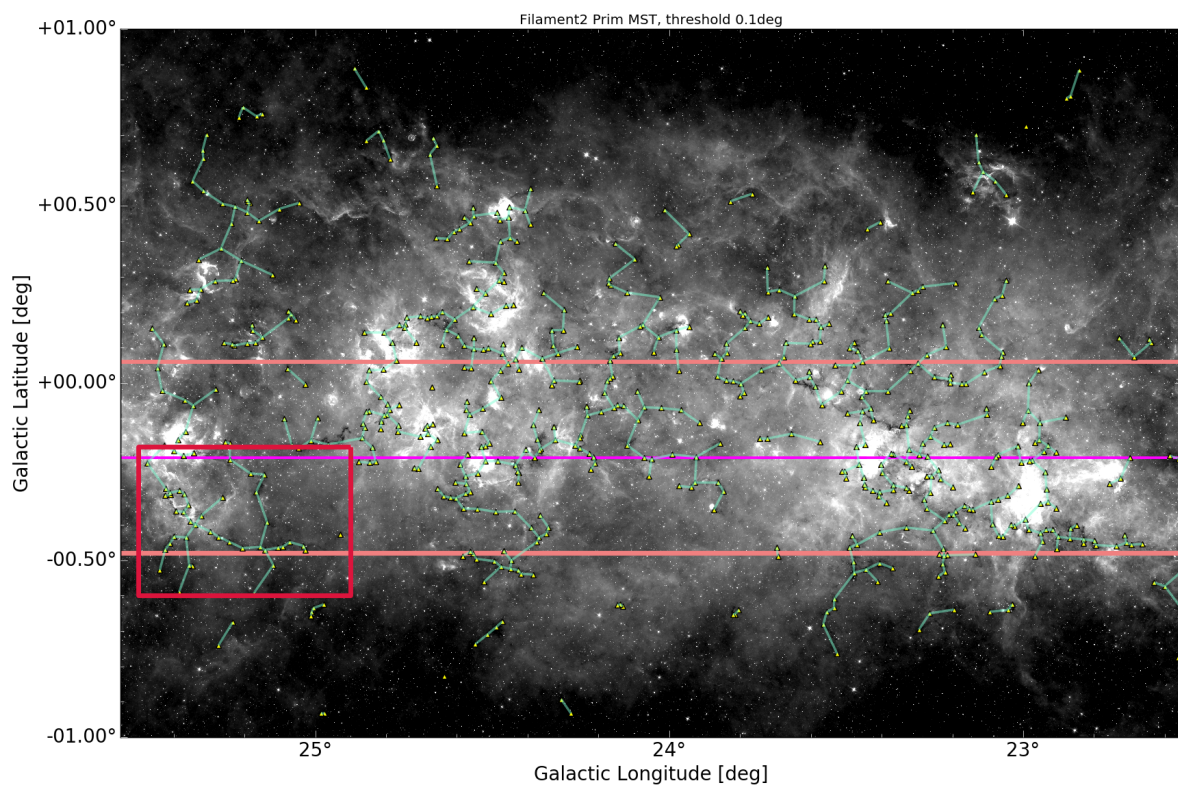


Figure 4.0.4: Filament2 Prim MST.

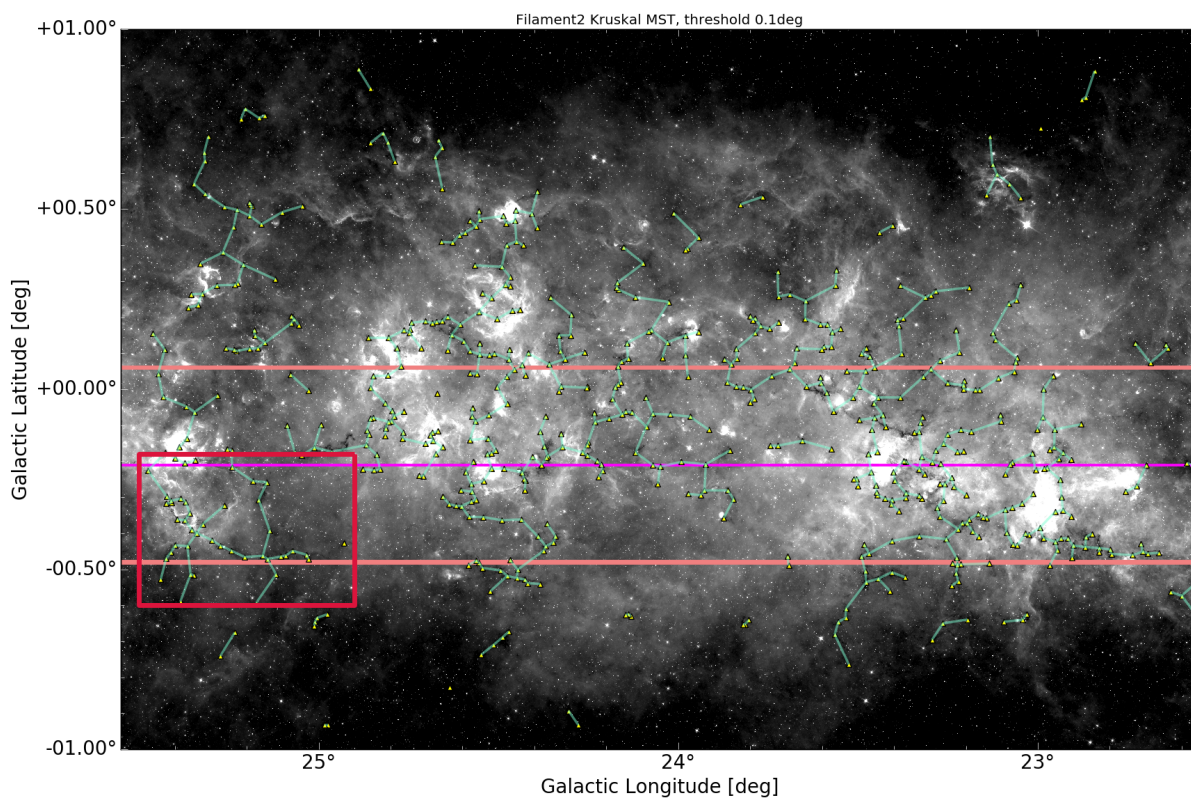


Figure 4.0.5: Filament2 Kruskal MST.

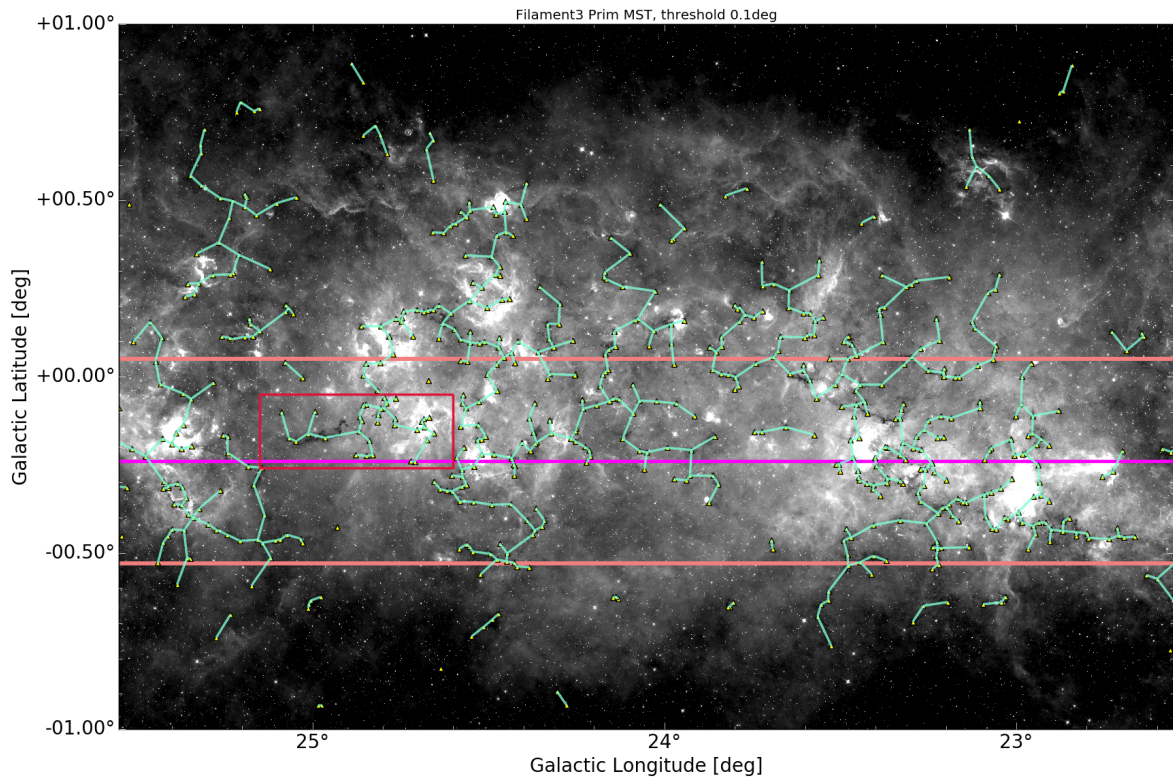


Figure 4.0.6: Filament3 Prim MST.

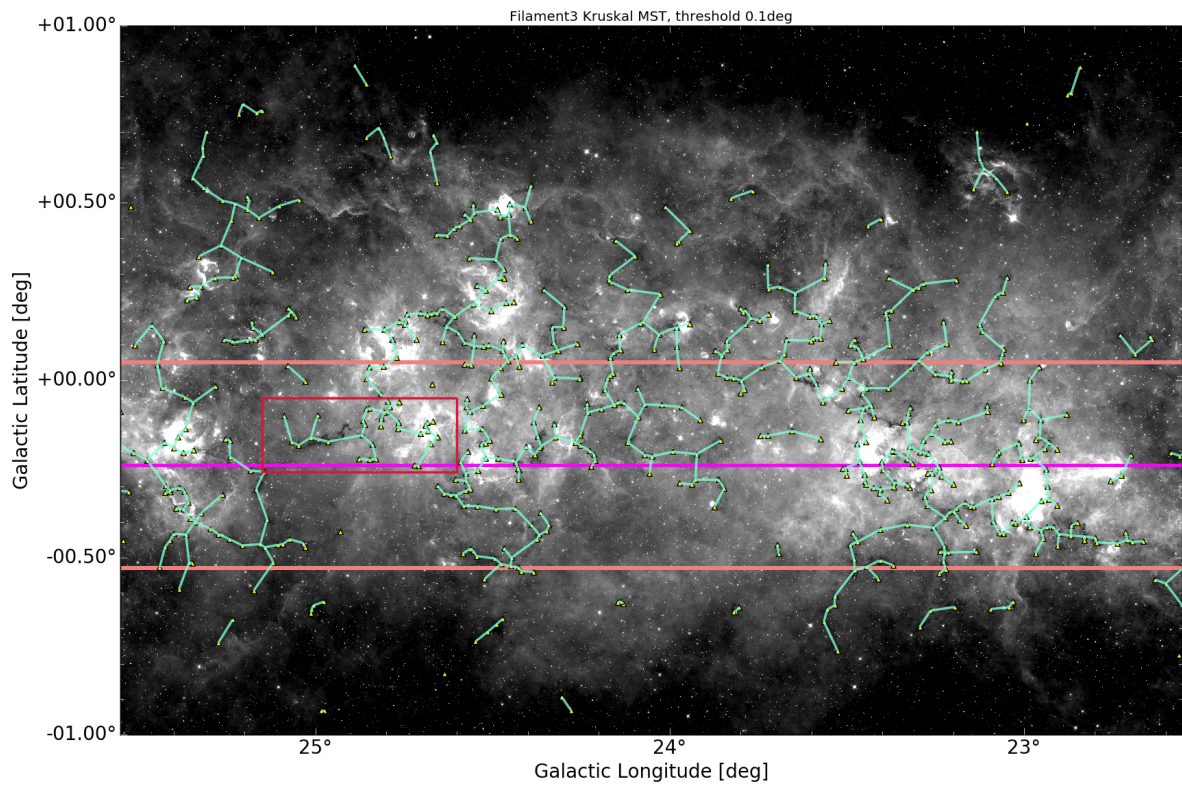


Figure 4.0.7: Filament3 Kruskal MST.

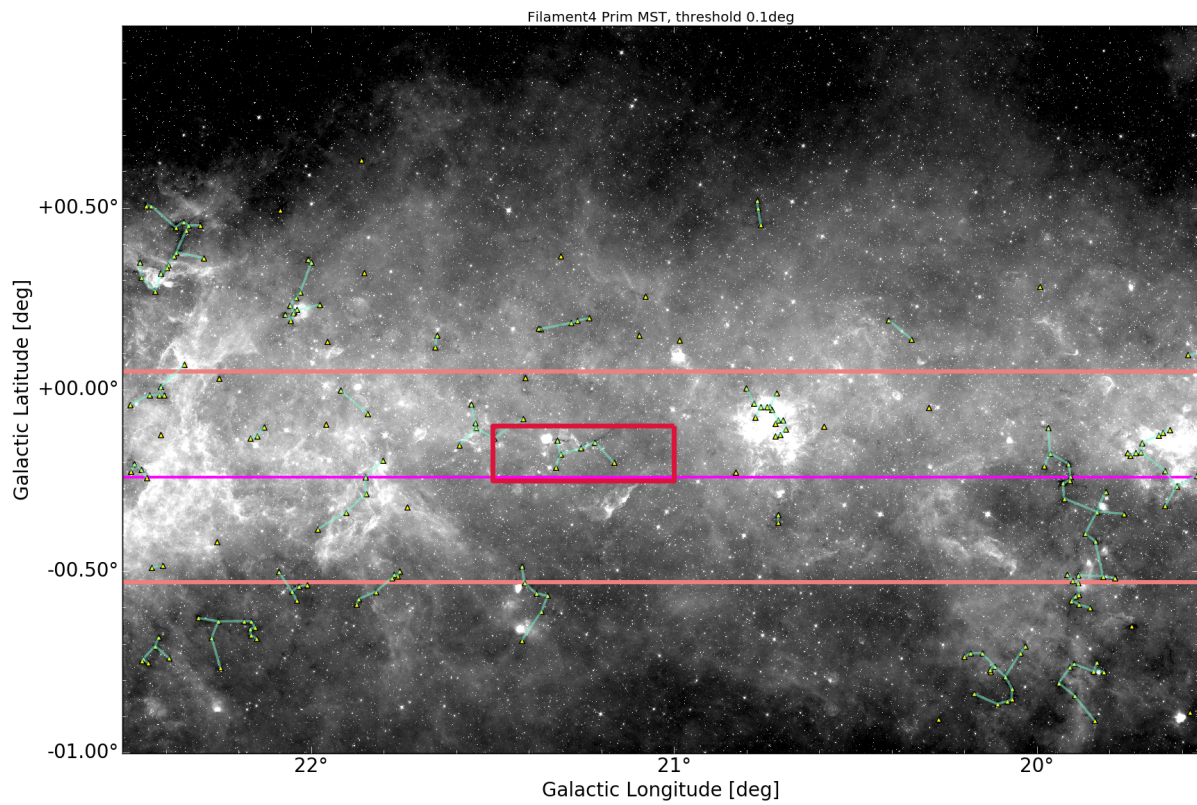


Figure 4.0.8: Filament4 Prim MST.

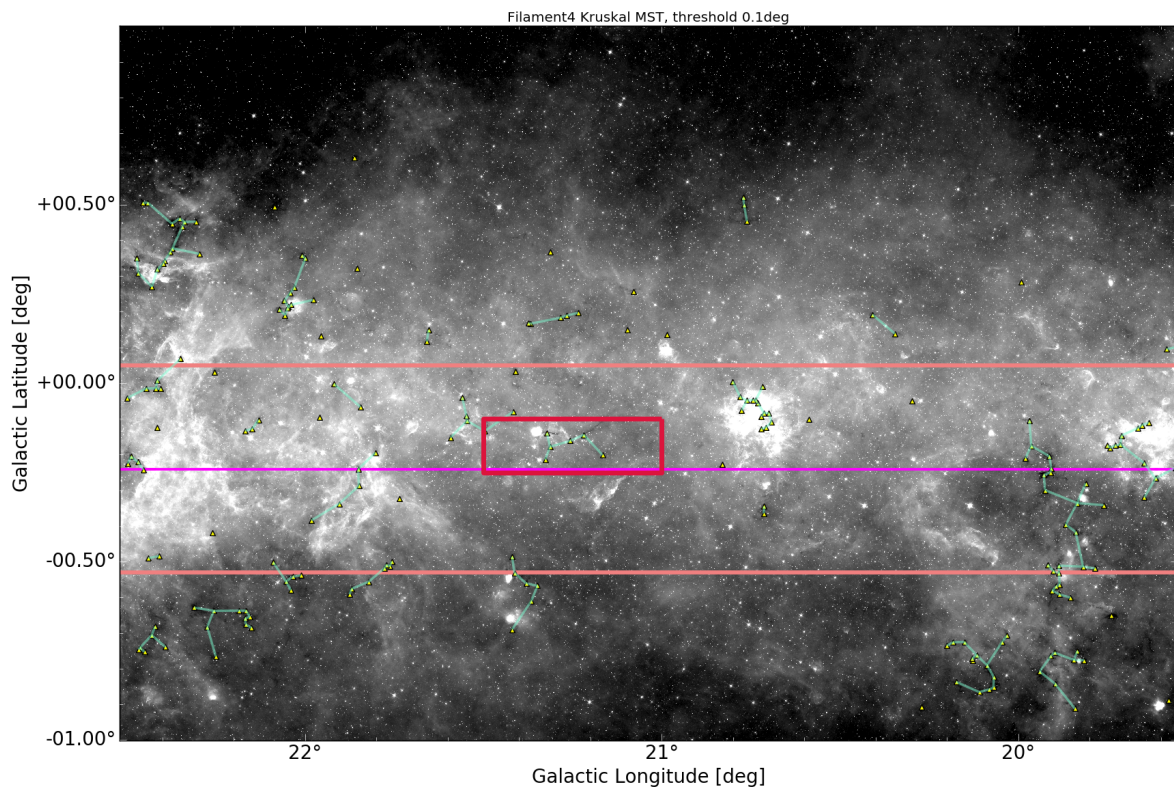


Figure 4.0.9: Filament4 Kruskal MST.

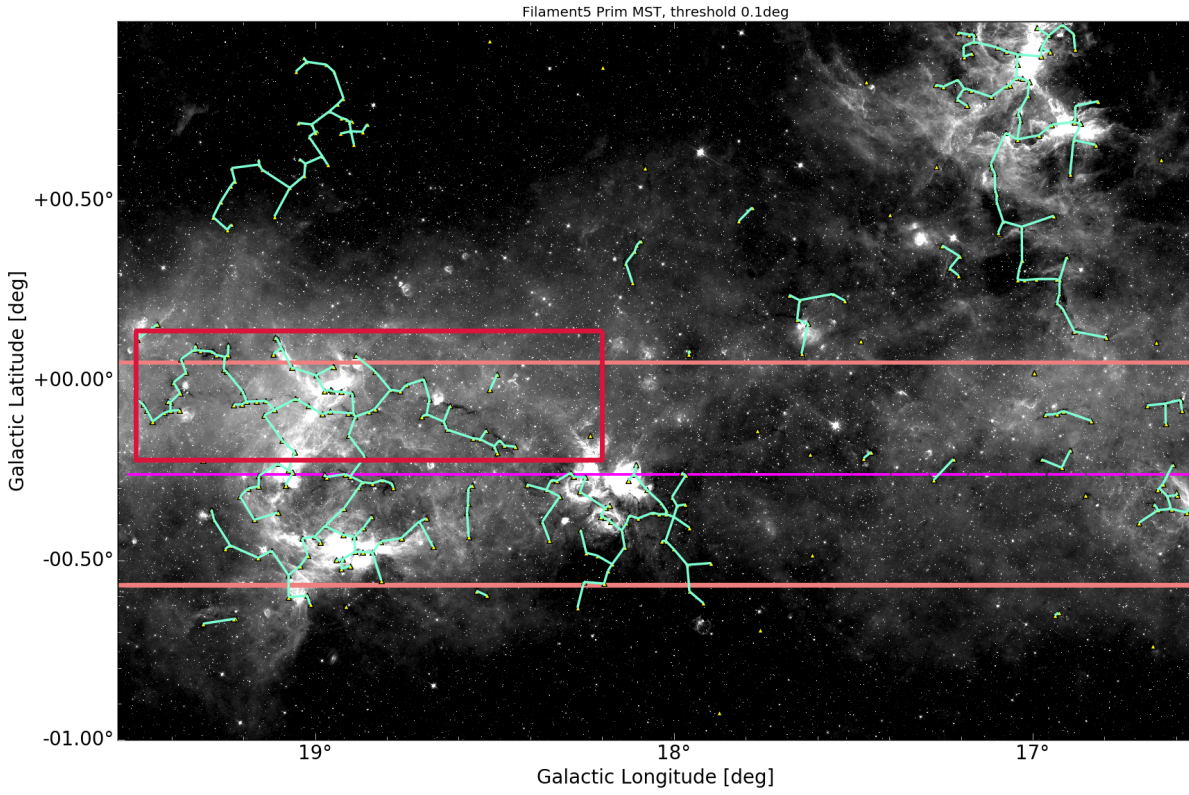


Figure 4.0.10: Filament5 Prim MST.

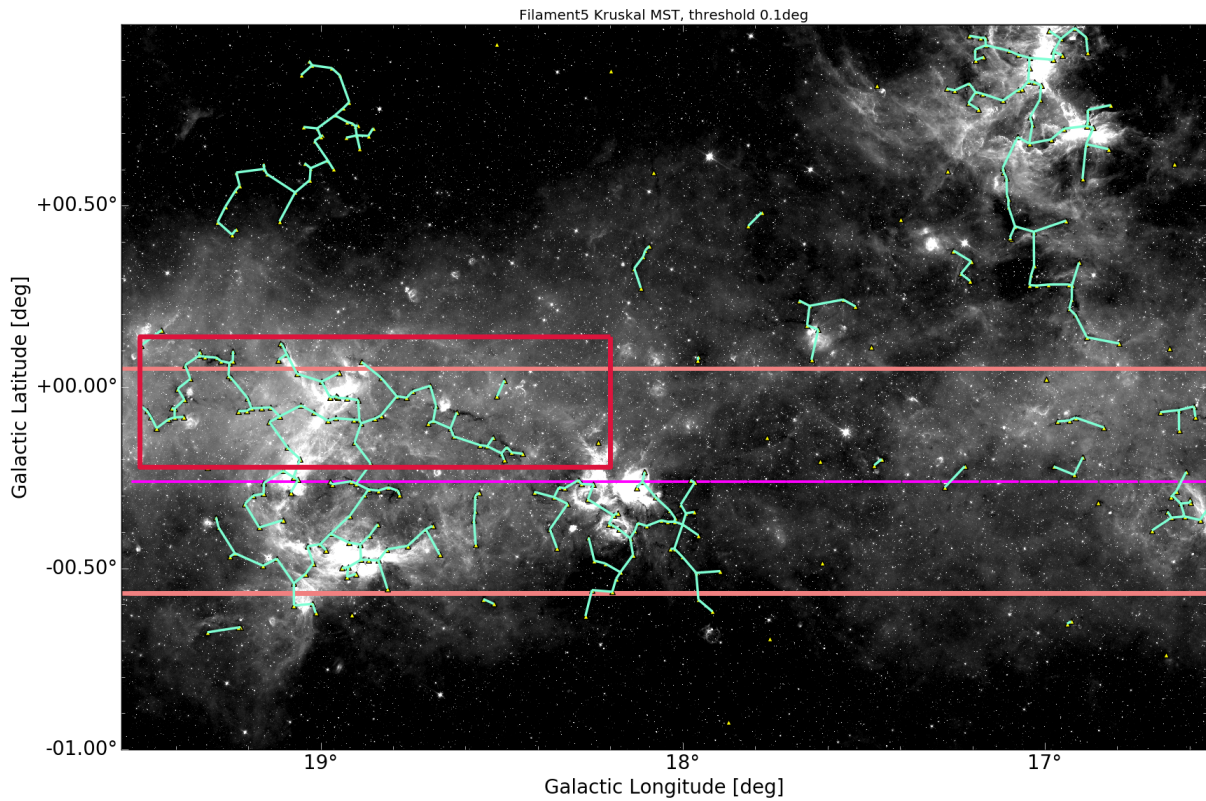


Figure 4.0.11: Filament5 Kruskal MST.

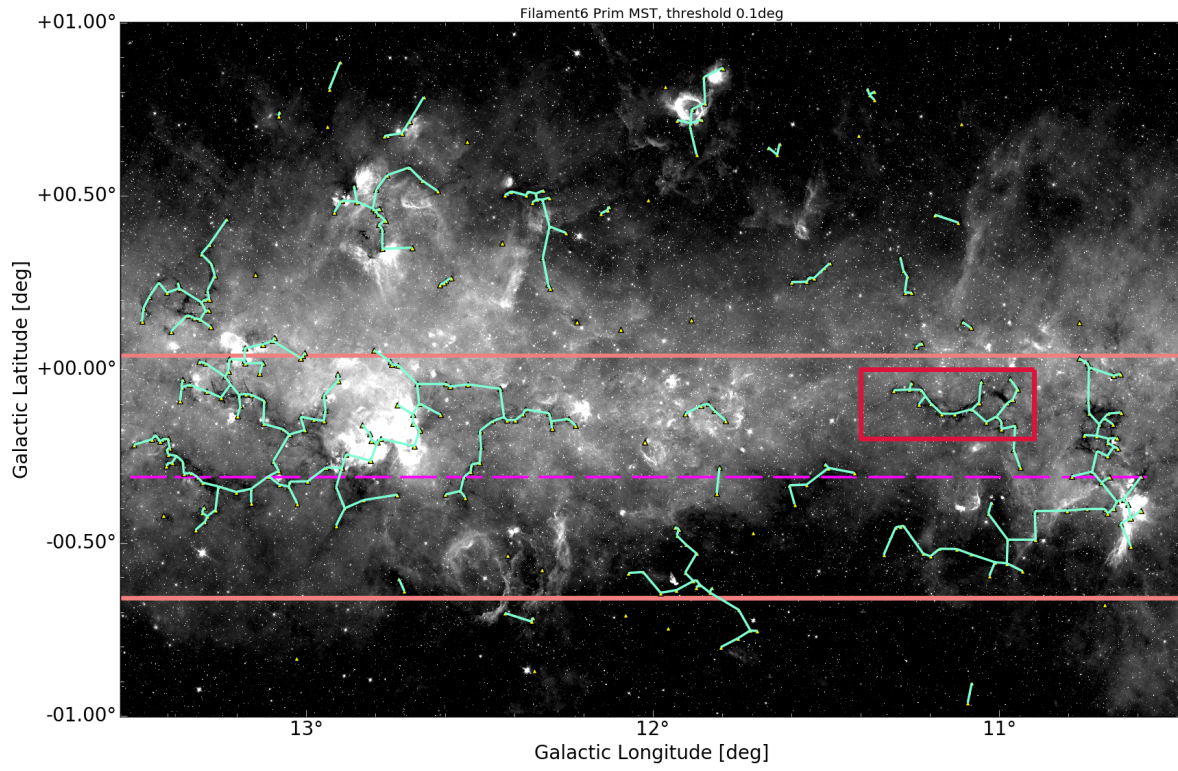


Figure 4.0.12: Filament6 Prim MST.

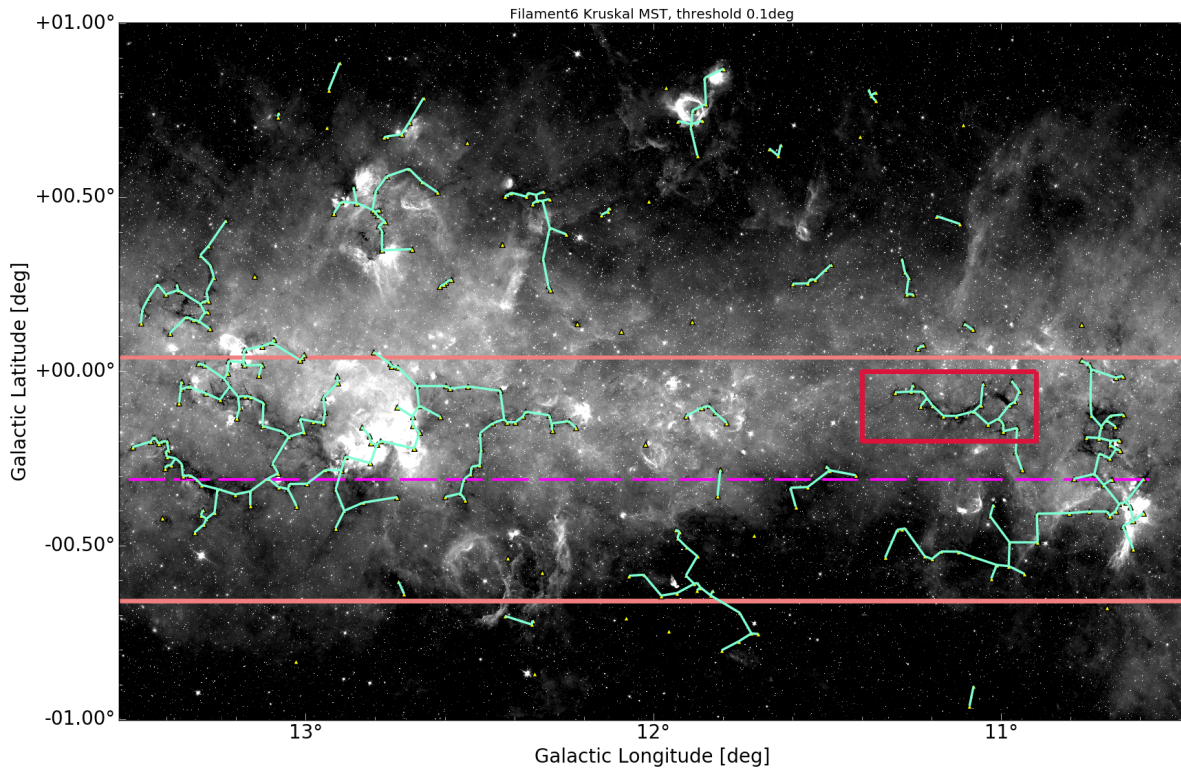


Figure 4.0.13: Filament6 Kruskal MST.

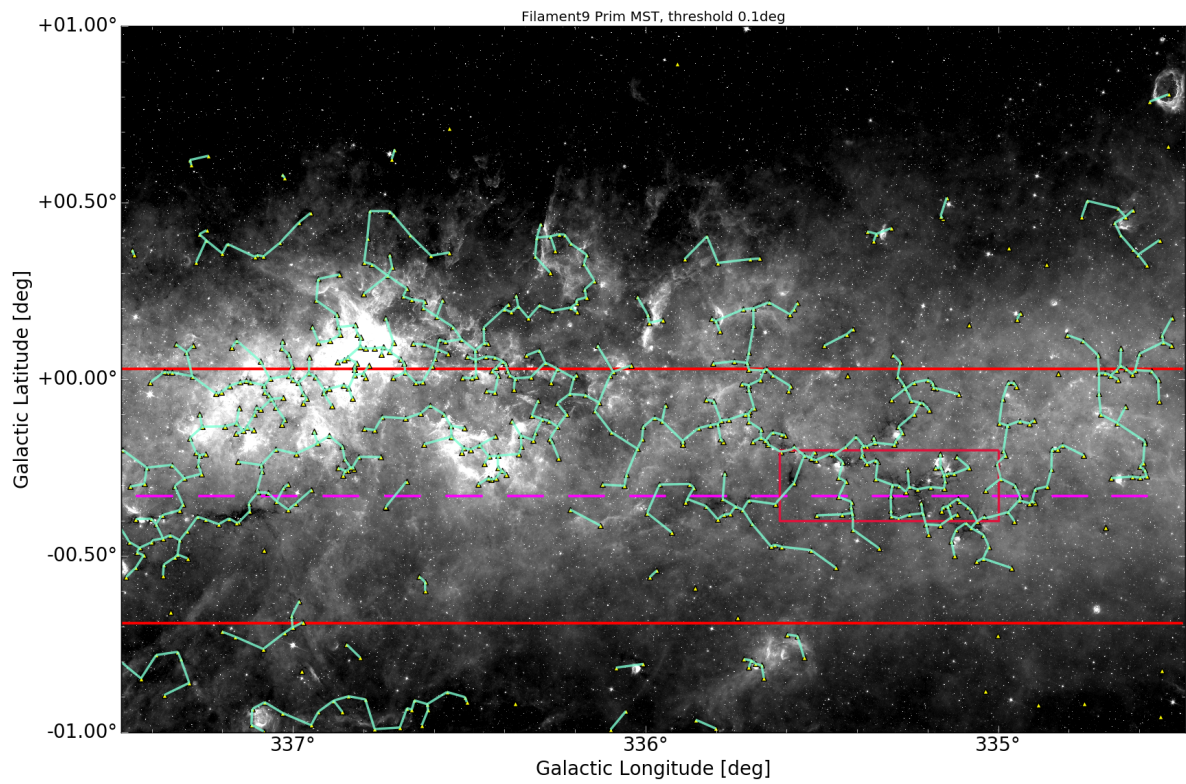


Figure 4.0.14: Filament9 Prim MST.

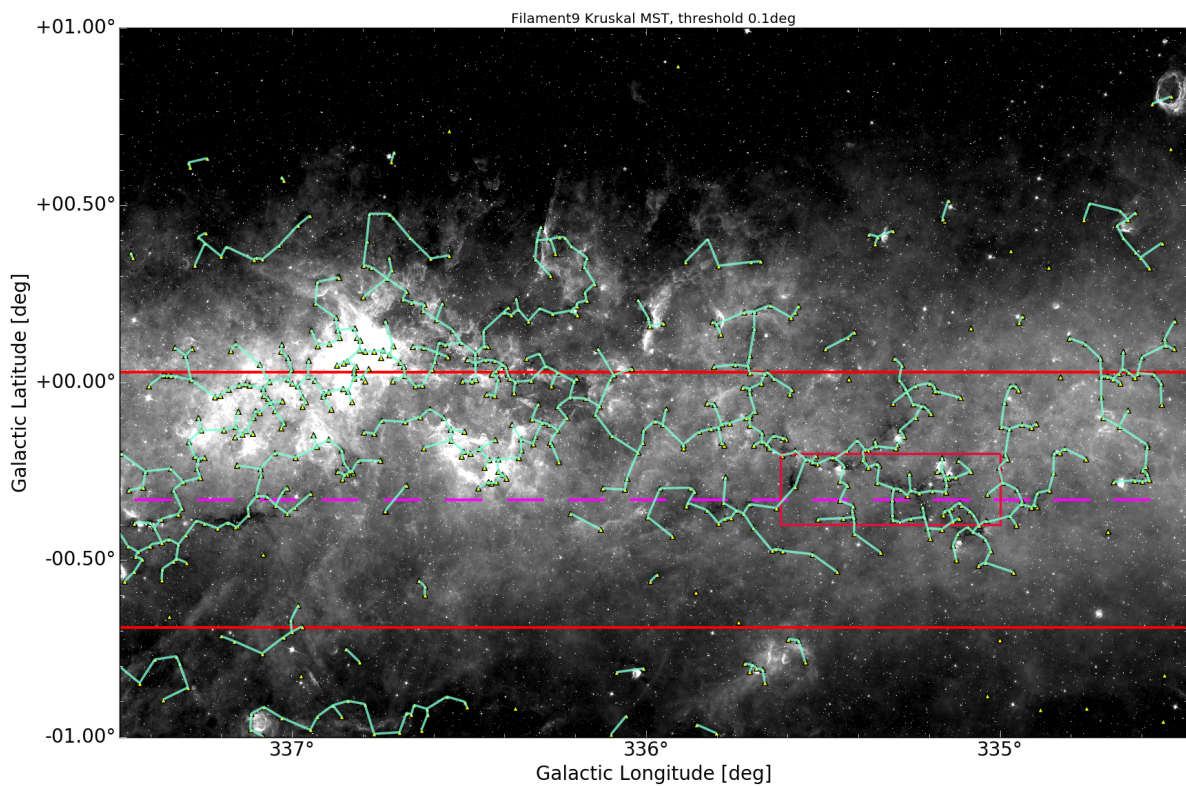


Figure 4.0.15: Filament9 Kruskal MST.

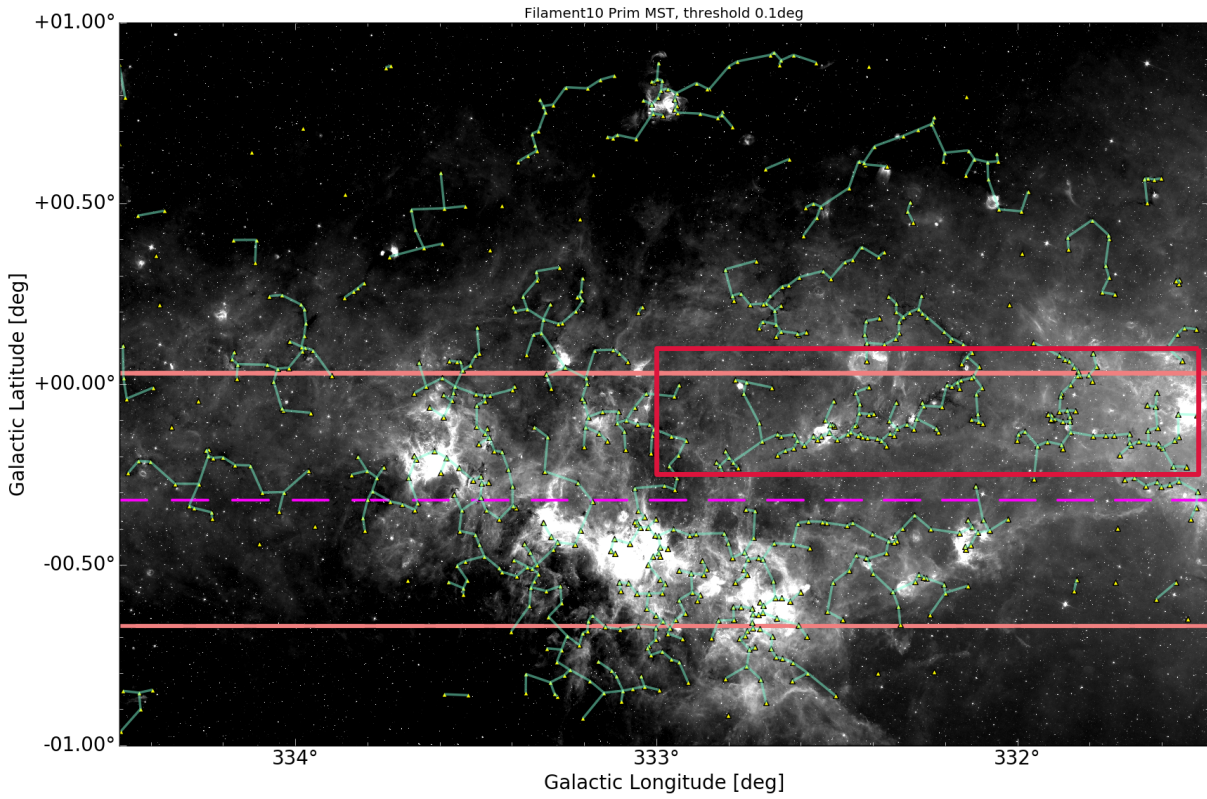


Figure 4.0.16: Filament10 Prim MST.

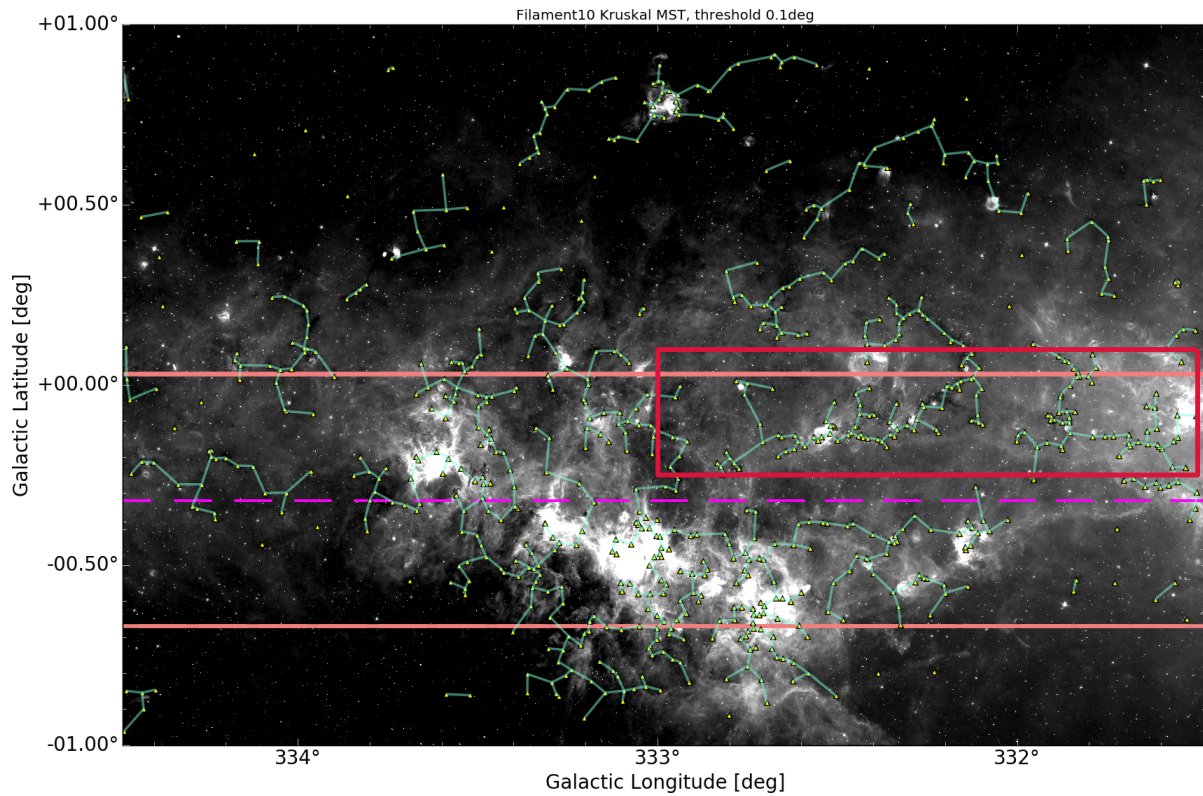


Figure 4.0.17: Filament10 Kruskal MST.

I conclude that the MST algorithm is a good approximation of the shape of previously found bone candidates, and that further testing on other regions of the Peretto & Fuller catalog needs to be done to see if other bones can be found through this method. For testing purposes, I ran the MST algorithm on the entire Peretto & Fuller catalog, in order to see how many minimum spanning trees we find in the entire catalog, and what are the parameters defining them.

I found a total of 795 trees in the Peretto Fuller catalog, and I saved the coordinates of the centers of these trees, alongside with the parameters describing them: number of nodes, tree size, tree length, tree longest path, average connectivity degree, average clustering coefficient, density of tree, length on x-axis, length on y-axis, length ratio, and an average inclination label. When saving this information, I have also flagged the already confirmed bones in the data. I will motivate choosing these parameters for tree diagnostics in Chapter 5, and show what are the ones most relevant to defining a minimum spanning tree as a bone.

5

Discussion

WE ANALYZE THE IMAGES obtained in Chapter 4) and perform tree diagnostics on the MSTs corresponding to the 10 found filaments by [44]Zucker et al. (2015). By analyzing these MSTs and deciding what are the parameters which describe them and which set them apart from other MSTs in the image, we want to be able to automatically distinguish minimum spanning trees of CO molecular clouds which match the location of bones of the galaxy.

Ultimately, we have the goal of performing tree diagnostics on the entire minimum

spanning forest that results from running Prim’s algorithm on the [13] Peretto & Fuller catalog, such that we identify potential areas of interest for the exploration of additional bones of the Milky Way. Once we identify potential regions of interest in the catalog and additional bone candidates, we can proceed with performing data analytics in these areas by using previous methods as well, such as the visual method in $l - b - v$ space suggested by [44] Zucker et al. (2015). The MST approach is thus valuable since it can give us a better intuition with regards to locations on the sky where we are most likely to find additional bones, and eventually map the entire skeleton of the galaxy.

For purposes of performing tree diagnostics, we suggest using the parameters below in order to describe the shape of a given tree and its likelihood for identifying a bone of the galaxy. We motivate choosing these parameters not only based on common properties of already identified filaments, but also based on the study of social networking graphs and of community detection algorithms [3].

- **Number of nodes** – total number of vertices in a tree, or in our case total number of clouds connected inside a filament
- **Tree Size** – total number of edges in a tree, always equal to $N - 1$, where N is the total number of nodes
- **Tree Total Length** – sum of weights of all edges inside a given tree
- **Longest Path Between Two Vertices** – sum of weights along the longest path connecting two nodes inside a given tree

- **Average Degree** – average connectivity degree of nodes inside a tree. In graph theory, the degree of a node is the total number of edges incident on that node.
- **Average Clustering Coefficient** – measuring the degree to which nodes inside a graph are clustering together. Always zero for the case of a spanning tree (inside which there are no cycles)
- **Density of Tree** – the density of undirected graphs is calculated using the formula $\rho = \frac{2m}{n(n-1)}$, where m is the number of edges, and n is the number of nodes.
- **Length on x-axis** – spanning degree in longitude, calculated using $\Delta x = x_{max} - x_{min}$, where x is the longitude coordinate of a given node inside the tree.
- **Length on y-axis** – spanning degree in latitude, calculated using $\Delta y = y_{max} - y_{min}$, where y is the latitude coordinate of a given node inside the tree.
- **Length Ratio** - $\Delta x / \Delta y$, showing how features are more elongated on the x -axis than on the y -axis
- **Average Inclination Angle** – average value of the slope calculated for each tree edge

We calculated the above parameters for all found MSTs in the regions of interest of our previously confirmed bones, and also on the entire list of 795 minimum spanning trees found present in the [13] Peretto & Fuller catalog. We include the resulting MSTs, alongside their parameter values, in Appendix B. Below I present **Table 5.0.1** containing diagnosis information for the previously confirmed bone candidates.

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament1	8	7	0.161	6	1.142	0	0.25	0.135	0.025	5.383	4.706
Filament2	55	54	2.309	30	1.018	0	0.036	0.446	0.749	0.595	2.339
Filament3	227	226	8.414	79	1.004	0	0.008	1.241	1.109	1.119	-3.031
Filament4	6	5	0.257	4	1.2	0	0.333	0.161	0.075	2.144	2.973
Filament5	130	129	4.954	46	1.007	0	0.015	1.056	0.743	1.419	4.034
Filament6	56	55	2.411	33	1.018	0	0.035	0.743	0.626	1.186	-2.114
Filament9	523	522	18.008	168	1.001	0	0.003	2.731	1.036	2.634	-0.004
Filament10	356	355	13.325	99	1.002	0	0.005	1.818	1.249	1.455	1.480
Filament10	101	100	2.833	56	1.01	0	0.019	0.738	0.494	1.493	-1.921

Table 5.0.1: Tree Diagnosis for Previously Confirmed Bones

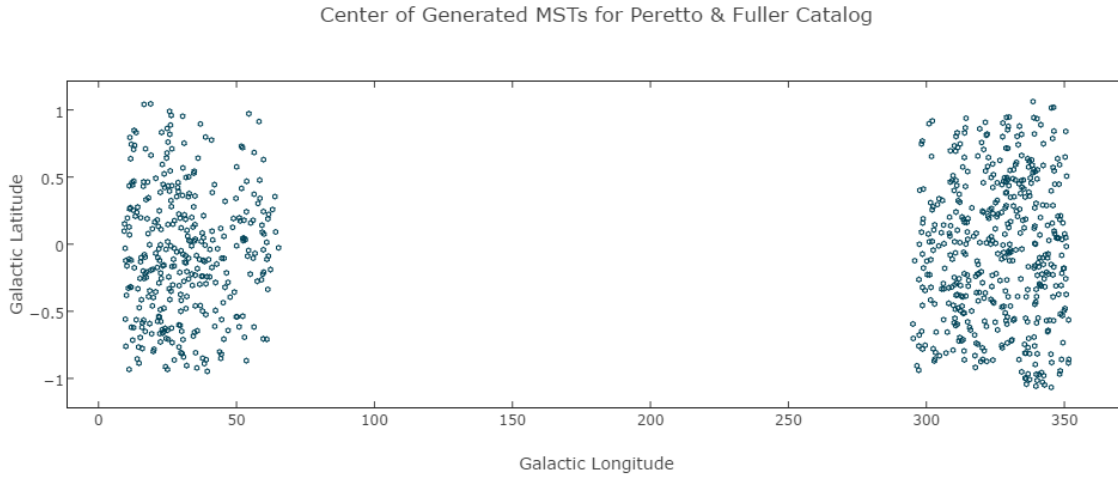


Figure 5.0.1: All Generated MSTs in the Peretto & Fuller Catalog.

Out of the above parameters, we identify the *Length Ratio* and the *Average Angle of Inclination* as the two most important ones. By looking at the Angle of Inclination, we can easily rule out MST candidates who are not bones - since we are only interested in filaments that are horizontal and thus approximately parallel to the Galactic Plane.

A histogram showing the distribution of average angles of inclination for all generated minimum spanning trees is shown below, alongside with the histogram of angles of inclination for previously identified bone candidates (see **Figures 5.0.2** and **5.0.3**).

I am also including below a graph showing the scatter plot of the center of all generated minimum spanning trees in the Peretto & Fuller catalog. This can be compared with the galactic coordinate representation of the Peretto & Fuller catalog in Chapter 4,

Figure 5.0.1.

5.1 FUTURE WORK

For the continuation of this project, I suggest:

1. adding velocity to the initial graph of position - position IRDCs, in order to obtain 3D minimum spanning trees
2. run machine learning regression and classification methods on the tree diagnosis table of 795 found filaments in the Peretto & Fuller catalog
3. use MST method alongside visual detection in order to investigate areas of interest and validate new bone findings.

By following the guidelines of point 1) above, we can eliminate the uncertainties induced by the case in which two points are considered to be close to each other on the sky because of their angular separation, instead of physical distance. Using velocity information and rotational curve, we can have a better intuition of the initial physical separation of different points belonging to the Euclidean graph.

Option 2) is helpful for the "smart" analysis of the data contained in the the tree diagnosis tables. I suggest trying both linear regression and a support vector classifier in order to analyze this data and decide on bones vs. non-bones candidates within the set of generated minimum spanning trees. For the regression method, a score of 1 can be assigned to previously confirmed bones, and a score of 0 to minimum spanning trees which can be rejected as bones by following criteria of the "average angle of inclination". For example,

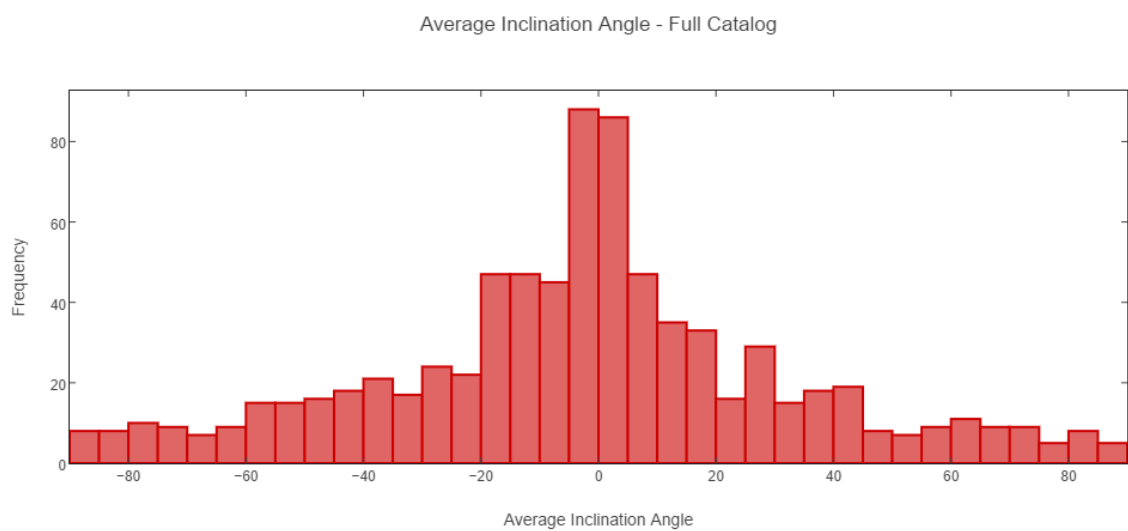


Figure 5.0.2: Average Angles of Inclination for all generated MSTs in the Peretto & Fuller Catalog.

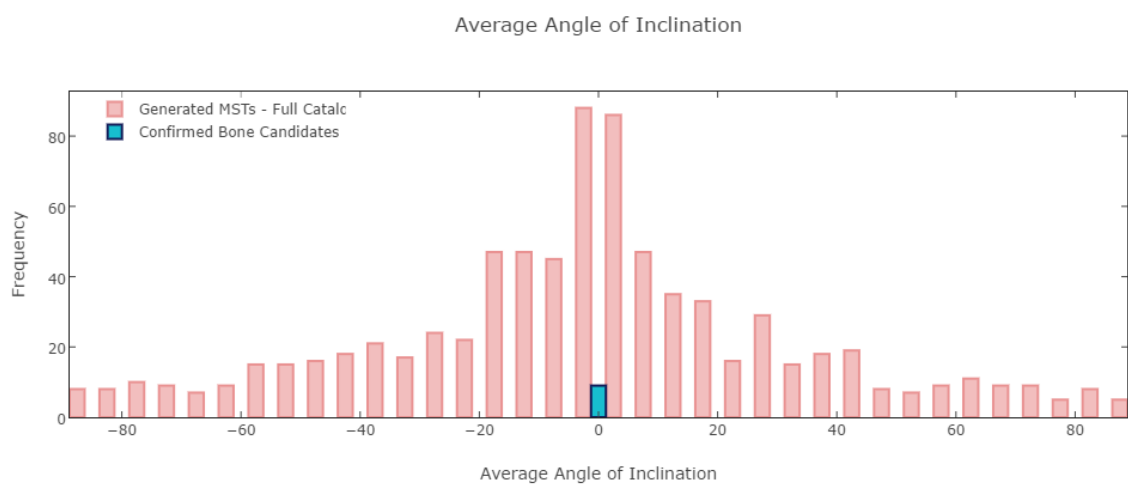


Figure 5.0.3: Average Angle of Inclination - Bones vs. All-MSTs.

MSTs which are perpendicular to the plane of the galaxy can be ruled out from the beginning from being bones. Training a regression model on such a dataset, then testing it on the entire 795 generated MSTs will help predict bone likelihoods for all MST candidates.

The Support Vector Classifier is likely going to give meaningful results too, since it is versatile and it works particularly well on data sets which have a high number of dimensions (in our case 10 different parameters describing a minimum spanning tree). Additionally, SVMs come with the option of defining different kernel functions for the decision-making process, thus helping the user fine-tune results for optimal performance.

Lastly, as I mention in option 3), I suggest using all the above practices in order to form a better intuition of where to search visually for bones in the galactic plane. Any new-found bones through the MST method should also be validated through visual detection.

References

- [1] M. Ackermann, M. Ajello, A. Allafort, E. Antolini, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, R. D. Blandford, E. D. Bloom, E. Bonamente, A. W. Borgland, E. Bottacini, T. J. Brandt, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, S. Buson, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, C. Cecchi, A. Chekhtman, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, F. D’Ammando, A. de Angelis, F. de Palma, C. D. Dermer, E. do Couto e Silva, P. S. Drell, A. Drlica-Wagner, T. Enoto, L. Falletti, C. Favuzzi, S. J. Fegan, E. C. Ferrara, W. B. Focke, Y. Fukazawa, Y. Fukui, P. Fusco, F. Gargano, D. Gasparri, S. Germani, N. Giglietto, F. Giordano, M. Giroletti, T. Glanzman, G. Godfrey, S. Guiriec, D. Hadasch, Y. Hanabata, A. K. Harding, M. Hayashida, K. Hayashi, D. Horan, X. Hou, R. E. Hughes, M. S. Jackson, G. Jóhannesson, A. S. Johnson, T. Kamae, H. Katagiri, J. Kataoka, M. Kerr, J. Knödlse, M. Kuss, J. Lande, S. Larsson, S.-H. Lee, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, K. Makishima, M. N. Mazziotta, J. Mehault, W. Mitthumsiri, A. A. Moiseev, C. Monte, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, T. Nakamori, M. Naumann-Godo, S. Nishino, J. P. Norris, E. Nuss, M. Ohno, T. Ohsugi, A. Okumura, M. Orienti, E. Orlando, J. F. Ormes, M. Ozaki, D. Paneque, J. H. Panetta, D. Parent, V. Pelassa, M. Pesce-Rollins, M. Pierbattista, F. Piron, G. Pivato, T. A. Porter, S. Rainò, M. Razzano, A. Reimer, O. Reimer, M. Roth, H. F.-W. Sadrozinski, C. Sgrò, E. J. Siskind, G. Spandre, P. Spinelli, A. W. Strong, H. Takahashi, T. Takahashi, T. Tanaka, J. G. Thayer, J. B. Thayer, O. Tibolla, M. Tinivella, D. F. Torres, A. Tramacere, E. Troja, Y. Uchiyama, T. L. Usher, J. Vandenbroucke, V. Vasileiou, G. Vianello, V. Vitale, A. P. Waite, P. Wang, B. L. Winer, K. S. Wood, Z. Yang, and S. Zimmer. GAMMA-RAY OBSERVATIONS OF THE ORION MOLECULAR CLOUDS WITH THE FERMI LARGE AREA TELESCOPE. *ApJ*, 756(1):4, aug 2012. doi: 10.1088/0004-637x/756/1/4. URL <http://dx.doi.org/10.1088/0004-637X/756/1/4>.
- [2] V. V. Bobylev, A. V. Mosenkov, A. T. Bajkova, and G. A. Gontcharov. Investigation of the Galactic bar based on photometry and stellar proper motions. *Astron. Lett.*, 40

- (2-3):86–94, feb 2014. doi: 10.1134/s1063773714030037. URL <http://dx.doi.org/10.1134/s1063773714030037>.
- [3] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? In *Proceedings of the 23rd international conference on World wide web - WWW '14*. Association for Computing Machinery (ACM), 2014. doi: 10.1145/2566486.2567997. URL <http://dx.doi.org/10.1145/2566486.2567997>.
- [4] Y. K. Choi, K. Hachisuka, M. J. Reid, Y. Xu, A. Brunthaler, K. M. Menten, and T. M. Dame. TRIGONOMETRIC PARALLAXES OF STAR FORMING REGIONS IN THE PERSEUS SPIRAL ARM. *ApJ*, 790(2):99, jul 2014. doi: 10.1088/0004-637x/790/2/99. URL <http://dx.doi.org/10.1088/0004-637x/790/2/99>.
- [5] D. P. Clemens. Massachusetts-Stony Brook Galactic plane CO survey - The Galactic disk rotation curve. *ApJ*, 295:422, aug 1985. doi: 10.1086/163386. URL <http://dx.doi.org/10.1086/163386>.
- [6] Françoise Combes. Distribution of CO in the Milky Way. *Annual Review of Astronomy and Astrophysics*, 29(1):195–237, sep 1991. doi: 10.1146/annurev.aa.29.090191.001211. URL <http://dx.doi.org/10.1146/annurev.aa.29.090191.001211>.
- [7] T. M. Dame and P. Thaddeus. A MOLECULAR SPIRAL ARM IN THE FAR OUTER GALAXY. *ApJ*, 734(1):L24, may 2011. doi: 10.1088/2041-8205/734/1/L24. URL <http://dx.doi.org/10.1088/2041-8205/734/1/L24>.
- [8] T. M. Dame, Dap Hartmann, and P. Thaddeus. The Milky Way in Molecular Clouds: A New Complete CO Survey. *ApJ*, 547(2):792–813, feb 2001. doi: 10.1086/318388. URL <http://dx.doi.org/10.1086/318388>.
- [9] S. W. Digel, I. A. Grenier, A. Heithausen, S. D. Hunter, and P. Thaddeus. Diffuse High-Energy Gamma-Ray Emission beyond the Solar Circle: The Cepheus and Polaris Flares and the Perseus Arm. *ApJ*, 463:609, jun 1996. doi: 10.1086/177275. URL <http://dx.doi.org/10.1086/177275>.
- [10] C. L. Dobbs and A. Burkert. The myth of the molecular ring. *Monthly Notices of the Royal Astronomical Society*, 421(4):2940–2946, mar 2012. doi:

- 10.1111/j.1365-2966.2012.20515.x. URL
<http://dx.doi.org/10.1111/j.1365-2966.2012.20515.x>.
- [11] D. J. Eden, T. J. T. Moore, L. K. Morgan, M. A. Thompson, and J. S. Urquhart. Star formation in Galactic spiral arms and the interarm regions. *Monthly Notices of the Royal Astronomical Society*, 431(2):1587–1595, mar 2013. doi: 10.1093/mnras/stt279. URL <http://dx.doi.org/10.1093/mnras/stt279>.
 - [12] C. Francis and E. Anderson. Galactic spiral structure. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2111):3425–3446, aug 2009. doi: 10.1098/rspa.2009.0036. URL <http://dx.doi.org/10.1098/rspa.2009.0036>.
 - [13] G. A. Fuller and N. Peretto. A New Comprehensive Catalogue of Infrared Dark Clouds. *Proc. IAU*, 5(H15):796, nov 2009. doi: 10.1017/S1743921310011749. URL <http://dx.doi.org/10.1017/S1743921310011749>.
 - [14] P. García, L. Bronfman, Lars-Åke Nyman, T. M. Dame, and A. Luna. GIANT MOLECULAR CLOUDS AND MASSIVE STAR FORMATION IN THE SOUTHERN MILKY WAY. *ApJS*, 212(1):2, apr 2014. doi: 10.1088/0067-0049/212/1/2. URL <http://dx.doi.org/10.1088/0067-0049/212/1/2>.
 - [15] Alyssa A. Goodman, João Alves, Christopher N. Beaumont, Robert A. Benjamin, Michelle A. Borkin, Andreas Burkert, Thomas M. Dame, James Jackson, Jens Kauffmann, Thomas Robitaille, and Rowan J. Smith. THE BONES OF THE MILKY WAY. *ApJ*, 797(1):53, nov 2014. doi: 10.1088/0004-637x/797/1/53. URL <http://dx.doi.org/10.1088/0004-637x/797/1/53>.
 - [16] D. A. Grabelsky, R. S. Cohen, L. Bronfman, and P. Thaddeus. Molecular clouds in the Carina arm - The largest objects associated regions of star formation, and the Carina arm in the Galaxy. *ApJ*, 331:181, aug 1988. doi: 10.1086/166548. URL <http://dx.doi.org/10.1086/166548>.
 - [17] K. Hachisuka, Y. K. Choi, M. J. Reid, A. Brunthaler, K. M. Menten, A. Sanna, and T. M. Dame. PARALLAXES OF STAR-FORMING REGIONS IN THE OUTER SPIRAL ARM OF THE MILKY WAY. *ApJ*, 800(1):2, feb 2015. doi: 10.1088/0004-637x/800/1/2. URL <http://dx.doi.org/10.1088/0004-637x/800/1/2>.

- [18] Dap Hartmann, Loris Magnani, and Patrick Thaddeus. A Survey of High-Latitude Molecular Gas in the Northern Galactic Hemisphere. *ApJ*, 492(1):205–212, jan 1998. doi: 10.1086/305019. URL <http://dx.doi.org/10.1086/305019>.
- [19] Mark Heyer and T.M. Dame. Molecular Clouds in the Milky Way. *Annual Review of Astronomy and Astrophysics*, 53(1):583–629, aug 2015. doi: 10.1146/annurev-astro-082214-122324. URL <http://dx.doi.org/10.1146/annurev-astro-082214-122324>.
- [20] James M. Jackson, Susanna C. Finn, Jill M. Rathborne, Edward T. Chambers, and Robert Simon. The Galactic Distribution of Infrared Dark Clouds. *ApJ*, 680(1): 349–361, jun 2008. doi: 10.1086/587539. URL <http://dx.doi.org/10.1086/587539>.
- [21] James M. Jackson, Susanna C. Finn, Edward T. Chambers, Jill M. Rathborne, and Robert Simon. THE “NESSIE” NEBULA: CLUSTER FORMATION IN A FILAMENTARY INFRARED DARK CLOUD. *ApJ*, 719(2):L185–L189, aug 2010. doi: 10.1088/2041-8205/719/2/l185. URL <http://dx.doi.org/10.1088/2041-8205/719/2/l185>.
- [22] Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7(1):48–48, jan 1956. doi: 10.1090/s0002-9939-1956-0078686-7. URL <http://dx.doi.org/10.1090/s0002-9939-1956-0078686-7>.
- [23] M. Lombardi, J. Alves, and C. J. Lada. 2MASS wide field extinction maps. *Astronomy & Astrophysics*, 535:A16, oct 2011. doi: 10.1051/0004-6361/201116915. URL <http://dx.doi.org/10.1051/0004-6361/201116915>.
- [24] Loris Magnani, Dap Hartmann, Sharon L. Holcomb, Laura E. Smith, and Patrick Thaddeus. A Survey of High-Latitude Molecular Gas in the Southern Galactic Hemisphere. *ApJ*, 535(1):167–175, may 2000. doi: 10.1086/308841. URL <http://dx.doi.org/10.1086/308841>.
- [25] D. J. Majaess, D. G. Turner, and D. J. Lane. Characteristics of the Galaxy according to Cepheids. *Monthly Notices of the Royal Astronomical Society*, 398(1):263–270, sep 2009. doi: 10.1111/j.1365-2966.2009.15096.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2009.15096.x>.

- [26] P. C. Myers, T. M. Dame, P. Thaddeus, R. S. Cohen, R. F. Silverberg, E. Dwek, and M. G. Hauser. Molecular clouds and star formation in the inner galaxy - A comparison of CO H II and far-infrared surveys. *ApJ*, 301:398, feb 1986. doi: 10.1086/163909. URL <http://dx.doi.org/10.1086/163909>.
- [27] Desika Narayanan, Mark R. Krumholz, Eve C. Ostriker, and Lars Hernquist. A general model for the CO-H₂ conversion factor in galaxies with applications to the star formation law. *Monthly Notices of the Royal Astronomical Society*, 421(4): 3127–3146, mar 2012. doi: 10.1111/j.1365-2966.2012.20536.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2012.20536.x>.
- [28] Jérôme Pety, Eva Schinnerer, Adam K. Leroy, Annie Hughes, Sharon E. Meidt, Dario Colombo, Gaelle Dumas, Santiago García-Burillo, Karl F. Schuster, Carsten Kramer, Clare L. Dobbs, and Todd A. Thompson. THE PLATEAU DE BURE 30 m ARCSECOND WHIRLPOOL SURVEY REVEALS A THICK DISK OF DIFFUSE MOLECULAR GAS IN THE M₅₁ GALAXY. *ApJ*, 779(1):43, nov 2013. doi: 10.1088/0004-637x/779/1/43. URL <http://dx.doi.org/10.1088/0004-637x/779/1/43>.
- [29] R. C. Prim. Shortest Connection Networks And Some Generalizations. *Bell System Technical Journal*, 36(6):1389–1401, nov 1957. doi: 10.1002/j.1538-7305.1957.tb01515.x. URL <http://dx.doi.org/10.1002/j.1538-7305.1957.tb01515.x>.
- [30] Sarah E. Ragan, Edwin A. Bergin, and Robert A. Gutermuth. DETECTION OF STRUCTURE IN INFRARED-DARK CLOUDS WITH SPITZER : CHARACTERIZING STAR FORMATION IN THE MOLECULAR RING. *ApJ*, 698(1):324–349, may 2009. doi: 10.1088/0004-637x/698/1/324. URL <http://dx.doi.org/10.1088/0004-637x/698/1/324>.
- [31] V.J. Rayward-Smith, Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. Introduction to Algorithms. *The Journal of the Operational Research Society*, 42(9): 816, sep 1991. doi: 10.2307/2583667. URL <http://dx.doi.org/10.2307/2583667>.
- [32] M. J. Reid, K. M. Menten, X. W. Zheng, A. Brunthaler, L. Moscadelli, Y. Xu, B. Zhang, M. Sato, M. Honma, T. Hirota, K. Hachisuka, Y. K. Choi, G. A. Moellenbrock, and A. Bartkiewicz. TRIGONOMETRIC PARALLAXES OF MASSIVE STAR-FORMING REGIONS. VI. GALACTIC STRUCTURE, FUNDAMENTAL

- PARAMETERS, AND NONCIRCULAR MOTIONS. *ApJ*, 700(1):137–148, jun 2009. doi: 10.1088/0004-637x/700/1/137. URL <http://dx.doi.org/10.1088/0004-637x/700/1/137>.
- [33] Julia Roman-Duval, James M. Jackson, Mark Heyer, Jill Rathborne, and Robert Simon. PHYSICAL PROPERTIES AND GALACTIC DISTRIBUTION OF MOLECULAR CLOUDS IDENTIFIED IN THE GALACTIC RING SURVEY. *ApJ*, 723(1):492–507, oct 2010. doi: 10.1088/0004-637x/723/1/492. URL <http://dx.doi.org/10.1088/0004-637x/723/1/492>.
- [34] A. Sanna, M. J. Reid, L. Moscadelli, T. M. Dame, K. M. Menten, A. Brunthaler, X. W. Zheng, and Y. Xu. TRIGONOMETRIC PARALLAXES OF MASSIVE STAR-FORMING REGIONS. VII. G9.620.20 AND THE EXPANDING 3 kpc ARM. *ApJ*, 706(1):464–470, nov 2009. doi: 10.1088/0004-637x/706/1/464. URL <http://dx.doi.org/10.1088/0004-637x/706/1/464>.
- [35] A. Sanna, M. J. Reid, T. M. Dame, K. M. Menten, A. Brunthaler, L. Moscadelli, X. W. Zheng, and Y. Xu. TRIGONOMETRIC PARALLAXES OF MASSIVE STAR-FORMING REGIONS. IX. THE OUTER ARM IN THE FIRST QUADRANT. *ApJ*, 745(1):82, dec 2011. doi: 10.1088/0004-637x/745/1/82. URL <http://dx.doi.org/10.1088/0004-637x/745/1/82>.
- [36] M. Sato, Y. W. Wu, K. Immer, B. Zhang, A. Sanna, M. J. Reid, T. M. Dame, A. Brunthaler, and K. M. Menten. TRIGONOMETRIC PARALLAXES OF STAR FORMING REGIONS IN THE SCUTUM SPIRAL ARM. *ApJ*, 793(2):72, sep 2014. doi: 10.1088/0004-637x/793/2/72. URL <http://dx.doi.org/10.1088/0004-637x/793/2/72>.
- [37] W. W. Shane. Neutral Hydrogen in an Interior Region of the Galaxy; the Longitude Interval 220 to 42g. III. The Scutum Arm. *AA*, 16:118, 01 1972. doi: 10.1086/163386. URL <http://adsabs.harvard.edu/abs/1972A%26A...16...118S>.
- [38] P. M. Solomon and A. R. Rivolo. A face-on view of the first galactic quadrant in molecular clouds. *ApJ*, 339:919, apr 1989. doi: 10.1086/167345. URL <http://dx.doi.org/10.1086/167345>.
- [39] Ruben A. Vazquez, Jorge May, Giovanni Carraro, Leonardo Bronfman, Andre Moitinho, and Gustavo Baume. Spiral Structure in the Outer Galactic Disk. I. The Third Galactic Quadrant. *ApJ*, 672(2):930–939, jan 2008. doi: 10.1086/524003. URL <http://dx.doi.org/10.1086/524003>.

- [40] Q. Daniel Wang and Ka Chun Yu. Shadowing the soft x-ray background by infrared cirrus: A study of selected regions. *The Astronomical Journal*, 109:698, feb 1995. doi: 10.1086/117312. URL <http://dx.doi.org/10.1086/117312>.
- [41] Y. W. Wu, M. Sato, M. J. Reid, L. Moscadelli, B. Zhang, Y. Xu, A. Brunthaler, K. M. Menten, T. M. Dame, and X. W. Zheng. Trigonometric parallaxes of star-forming regions in the Sagittarius spiral arm. *Astronomy & Astrophysics*, 566:A17, jun 2014. doi: 10.1051/0004-6361/201322765. URL <http://dx.doi.org/10.1051/0004-6361/201322765>.
- [42] Y. Xu, J. J. Li, M. J. Reid, K. M. Menten, X. W. Zheng, A. Brunthaler, L. Moscadelli, T. M. Dame, and B. Zhang. ON THE NATURE OF THE LOCAL SPIRAL ARM OF THE MILKY WAY. *ApJ*, 769(1):15, apr 2013. doi: 10.1088/0004-637x/769/1/15. URL <http://dx.doi.org/10.1088/0004-637x/769/1/15>.
- [43] B. Zhang, M. J. Reid, K. M. Menten, X. W. Zheng, A. Brunthaler, T. M. Dame, and Y. Xu. PARALLAXES FOR W₄₉N AND G048.600.02: DISTANT STAR FORMING REGIONS IN THE PERSEUS SPIRAL ARM. *ApJ*, 775(1):79, sep 2013. doi: 10.1088/0004-637x/775/1/79. URL <http://dx.doi.org/10.1088/0004-637x/775/1/79>.
- [44] Catherine Zucker, Cara Battersby, and Alyssa Goodman. THE SKELETON OF THE MILKY WAY. *ApJ*, 815(1):23, dec 2015. doi: 10.1088/0004-637x/815/1/23. URL <http://dx.doi.org/10.1088/0004-637x/815/1/23>.

6

Appendix A

Moving Forward - Computational Techniques in Astronomy

THE METHOD OF USING MINIMUM SPANNING TREES in order to analyze the structure of the galactic plane fits into the larger category of applying computational techniques to astronomical data. I believe there are many advantages to using "smart", computational approaches to the field of astronomy research - more efficient calculations, sophisticated techniques for data analysis, greater capacity for advanced simulations of physical phenomena, and the power of revealing hidden trends behind collected observational data, to name just a few. Some of the unsolved problems in computational astronomy remain 3D Selection and 3D Segmentation. Solving the problem of finding fundamental structures within the plane of the Galaxy by using a Minimum Spanning Tree approach is a gateway towards developing finer 3D Selection and 3D Segmentation techniques.

Ultimately, with better capabilities for 3D Selection and 3D Segmentation, interacting with volumetric data will be more efficient and straightforward (directly in the 3D space, instead of combining together slices of 2D images) which can have major breakthroughs not only in astronomy (identifying structures within the Universe, as seen at large scales), but also in other fields. The Astronomical Medicine area falls under this category, and continuing to develop sophisticated technologies for the exploration of astronomy data strongly influences the medicine field as well, given the similarity between identifying tumors in radiology volumetric data and identifying shapes and structures in the physical universe.

7

Appendix B

THE GENERATED MSTs FROM THE SKY REGIONS OF PREVIOUSLY IDENTIFIED BONE candidates, alongside their parameters obtained by performing tree diagnostics according to the directions discussed in Chapter 5. In the tables below, I am highlighting the entries which correspond to previously confirmed bones.

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament1.1	183	182	6.263	66	1.005	0	0.010	0.895	0.832	1.076	0.699
Filament1.2	32	31	1.312	19	1.032	0	0.062	0.329	0.506	0.650	-1.899
Filament1.3	37	36	1.840	20	1.027	0	0.054	0.788	0.460	1.711	-2.007
Filament1.4	4	3	0.074	3	1.333	0	0.5	0.036	0.050	0.709	-23.417
Filament1.5	13	12	0.405	11	1.083	0	0.153	1.190	0.121	1.562	-29.863
Filament1.6	28	27	1.251	17	1.037	0	0.071	0.393	0.333	1.180	-10.736
Filament1.7	28	27	1.259	19	1.037	0	0.071	0.404	0.284	1.422	-4.856
Filament1.8	25	24	1.058	19	1.041	0	0.08	0.299	0.422	0.708	0.982
Filament1.9	2	1	0.016	1	2	0	1	0.015	0.028	5.480	10.339
Filament1.10	5	4	0.238	4	1.25	0	0.4	0.156	0.112	1.394	27.822
Filament1.11	2	1	0.023	1	2	0	1	0.016	0.015	1.061	43.298
Filament1.12	6	5	0.269	4	1.2	0	0.333	0.125	0.084	1.482	-25.986
Filament1.13	4	3	0.198	3	1.333	0	0.5	0.075	0.175	0.430	-58.371
Filament1.14	8	7	0.301	6	1.142	0	0.25	0.164	0.074	2.210	-7.872
Filament1.15	2	1	0.041	1	2	0	1	0.017	0.037	0.469	64.846
Filament1.16	5	4	0.187	3	1.25	0	0.4	0.133	0.059	2.25	-1.142
Filament1.17	2	1	0.074	1	2	0	1	0.052	0.052	0.987	-45.366
Filament1.18	6	5	0.267	5	1.2	0	0.333	0.231	0.061	3.787	3.967
Filament1.19	8	7	0.161	6	1.142	0	0.25	0.135	0.025	5.383	4.706
Filament1.20	9	8	0.363	6	1.125	0	0.222	0.136	0.187	0.732	4.611
Filament1.21	9	8	0.236	8	1.125	0	0.222	0.118	0.165	0.718	22.901
Filament1.22	3	2	0.091	2	1.5	0	0.666	0.074	0.052	1.436	-37.764
Filament1.23	3	2	0.118	2	1.5	0	0.666	0.010	0.116	0.088	-10.310
Filament1.24	3	2	0.116	2	1.5	0	0.666	0.070	0.061	1.147	32.270
Filament1.25	2	1	0.055	1	2	0	1	0.048	0.026	1.846	-28.433
Filament1.26	2	1	0.037	1	2	0	1	0.035	0.011	3.00	-18.430
Filament1.27	2	1	0.076	1	2	0	1	0.033	0.068	0.481	64.27

Table 7.0.1: Tree Diagnosis for MSTs in the sky region of Filament 1, 27 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament2.1	227	226	8.414	79	1.004	0	0.008	1.241	1.109	1.119	-3.031
Filament2.2	55	54	2.309	30	1.018	0	0.036	0.446	0.749	0.595	2.339
Filament2.3	26	25	1.137	15	1.04	0	0.076	0.317	0.474	0.667	-10.412
Filament2.4	219	218	7.922	74	1.004	0	0.009	1.209	1.095	1.104	-0.787
Filament2.5	4	3	0.146	3	1.333	0	0.5	0.068	0.101	0.675	-12.391
Filament2.6	12	11	0.303	8	1.090	0	0.166	0.202	0.096	2.103	10.324
Filament2.7	4	3	0.053	3	1.333	0	0.5	0.036	0.031	1.148	-54.00
Filament2.8	2	1	0.065	1	2	0	1	0.049	0.044	1.113	41.925
Filament2.9	3	2	0.124	2	1.5	0	0.666	0.098	0.056	1.744	-32.669
Filament2.10	3	2	0.089	2	1.5	0	0.666	0.035	0.079	0.446	-50.641
Filament2.11	2	1	0.048	1	2	0	1	0.027	0.040	0.694	55.237
Filament2.12	6	5	0.265	4	1.2	0	0.333	0.094	0.169	0.555	35.138
Filament2.13	4	3	0.095	3	1.333	0	0.5	0.066	0.028	2.346	-19.294
Filament2.14	2	1	0.025	1	2	0	1	0.003	0.025	0.121	83.068
Filament2.15	4	3	0.173	3	1.333	0	0.5	0.169	0.022	7.627	1.882
Filament2.16	3	2	0.063	2	1.5	0	0.666	0.054	0.021	2.570	-8.206
Filament2.17	4	3	0.146	3	1.333	0	0.5	0.019	0.135	0.146	29.865
Filament2.18	4	3	0.129	3	1.333	0	0.5	0.066	0.078	0.847	31.578
Filament2.19	3	2	0.099	2	1.5	0	0.666	0.057	0.074	0.781	-46.099
Filament2.20	4	3	0.099	3	1.333	0	0.5	0.076	0.063	1.198	-40.183
Filament2.21	3	2	0.071	2	1.5	0	0.666	0.063	0.020	3.015	-29.289
Filament2.22	2	1	0.065	1	2	0	1	0.061	0.021	2.892	-19.07
Filament2.23	2	1	0.062	1	2	0	1	0.032	0.053	0.610	58.59
Filament2.24	2	1	0.009	1	2	0	1	0.009	0.002	49.9	-1.14
Filament2.25	2	1	0.042	1	2	0	1	0.037	0.019	1.898	-27.776
Filament2.26	4	3	0.142	3	1.333	0	0.5	0.090	0.053	1.705	22.389
Filament2.27	4	3	0.186	2	1.333	0	0.5	0.082	0.113	0.727	9.972
Filament2.28	2	1	0.076	1	2	0	1	0.040	0.065	0.616	-58.337
Filament2.29	3	2	0.021	2	1.5	0	0.666	0.016	0.012	1.334	-24.995
Filament2.30	3	2	0.017	2	1.5	0	0.666	0.015	0.006	2.418	13.088

Table 7.0.2: Tree Diagnosis for MSTs in the sky region of Filament 2, 30 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament3.1	219	218	7.922	74	1.004	0	0.009	1.209	1.095	1.104	-0.787
Filament3.2	227	226	8.414	79	1.004	0	0.008	1.241	1.109	1.119	-3.031
Filament3.3	4	3	0.146	3	1.333	0	0.5	0.068	0.101	0.675	-12.391
Filament3.4	25	24	1.058	19	1.041	0	0.08	0.299	0.422	0.708	0.982
Filament3.5	84	83	3.685	34	1.012	0	0.023	0.888	0.749	1.185	-0.323
Filament3.6	5	4	0.187	3	1.25	0	0.4	0.133	0.059	2.250	-1.142
Filament3.7	26	25	1.137	15	1.04	0	0.076	0.317	0.474	0.667	-10.412
Filament3.8	4	3	0.173	3	1.333	0	0.5	0.169	0.022	7.627	1.882
Filament3.9	12	11	0.303	8	1.090	0	0.166	0.202	0.096	2.103	10.324
Filament3.10	2	1	0.065	1	2	0	1	0.049	0.044	1.113	41.925
Filament3.11	4	3	0.146	3	1.333	0	0.5	0.019	0.135	0.146	29.856
Filament3.12	6	5	0.149	3	1.2	0	0.333	0.075	0.071	1.060	-17.827
Filament3.13	17	16	0.626	13	1.065	0	0.117	0.233	0.235	0.988	-17.638
Filament3.14	10	9	0.351	8	1.111	0	0.2	0.160	0.137	1.173	19.219
Filament3.15	7	6	0.242	5	1.166	0	0.285	0.174	0.109	1.594	-18.122
Filament3.16	2	1	0.076	1	2	0	1	0.040	0.065	0.616	-58.337
Filament3.17	9	8	0.217	7	1.125	0	0.222	0.064	0.167	0.383	-30.939
Filament3.18	3	2	0.089	2	1.5	0	0.666	0.035	0.079	0.446	-50.641
Filament3.19	9	8	0.363	6	1.125	0	0.222	0.136	0.187	0.732	4.611
Filament3.20	4	3	0.129	3	1.333	0	0.5	0.066	0.078	0.847	31.578
Filament3.21	2	1	0.042	1	2	0	1	0.037	0.019	1.898	-27.776
Filament3.22	5	4	0.138	3	1.25	0	0.4	0.077	0.077	0.998	18.657
Filament3.23	4	3	0.095	3	1.33	0	0.5	0.066	0.028	2.346	-19.294
Filament3.24	3	2	0.141	2	1.5	0	0.666	0.079	0.112	0.707	60.104
Filament3.25	4	3	0.099	3	1.33	0	0.5	0.076	0.063	1.198	-40.183
Filament3.26	3	2	0.071	2	1.5	0	0.666	0.063	0.020	3.015	-29.289
Filament3.27	6	5	0.265	4	1.2	0	0.333	0.094	0.169	0.555	35.138
Filament3.28	3	2	0.051	2	1.5	0	0.66	0.037	0.030	1.237	-35.103
Filament3.29	3	2	0.017	2	1.5	0	0.66	0.015	0.006	2.418	13.088
Filament3.30	2	1	0.048	1	2	0	1	0.027	0.040	0.694	55.237
Filament3.31	3	2	0.099	2	1.5	0	0.666	0.05	0.074	0.781	-46.099
Filament3.32	4	3	0.142	3	1.33	0	0.5	0.090	0.053	1.705	22.389
Filament3.33	3	2	0.124	2	1.5	0	0.666	0.098	0.056	1.744	-32.669
Filament3.34	4	3	0.053	3	1.33	0	0.5	0.036	0.031	1.148	-54.000
Filament3.35	2	1	0.009	1	2	0	1	0.009	0.0002	49.9	-1.148
Filament3.36	2	1	0.062	1	2	0	1	0.032	0.053	0.610	58.59
Filament3.37	3	2	0.118	2	1.5	0	0.66	0.010	0.116	0.088	-10.31
Filament3.38	3	2	0.021	2	1.5	0	0.666	0.016	0.012	1.334	-24.995
Filament3.39	2	1	0.065	1	2	0	1	0.061	0.021	2.892	-19.070
Filament3.40	4	3	0.073	3	1.333	0	0.5	0.021	0.069	0.305	-70.728
Filament3.41	4	3	0.186	2	1.333	0	0.5	0.082	0.113	0.727	9.972
Filament3.42	7	6	0.181	5	1.166	0	0.285	0.132	0.050	2.623	4.763
Filament3.43	2	1	0.016	1	2	0	1	0.015	0.002	5.48	10.339
Filament3.44	2	1	0.030	1	2	0	1	0.029	0.006	4.804	-11.756
Filament3.45	2	1	0.025	1	2	0	1	0.0031	0.0255	0.121	83.068

Table 7.0.3: Tree Diagnosis for MSTs in the sky region of Filament 3, 45 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament4.1	5	4	0.278	4	1.25	0	0.4	0.179	0.189	0.947	-7.140
Filament4.2	2	1	0.098	1	2	0	1	0.074	0.064	1.155	40.867
Filament4.3	7	6	0.164	6	1.166	0	0.285	0.118	0.092	1.270	-39.814
Filament4.4	14	13	0.299	10	1.076	0	0.142	0.108	0.132	0.815	3.013
Filament4.5	8	7	0.326	6	1.142	0	0.25	0.125	0.161	0.772	-8.194
Filament4.6	24	23	0.989	17	1.043	0	0.083	0.219	0.495	0.442	-15.141
Filament4.7	12	11	0.414	7	1.090	0	0.166	0.190	0.210	0.906	-14.006
Filament4.8	5	4	0.124	3	1.25	0	0.4	0.063	0.069	0.916	-17.819
Filament4.9	6	5	0.257	4	1.2	0	0.333	0.161	0.075	2.144	2.973
Filament4.10	3	2	0.067	2	1.5	0	0.666	0.010	0.066	1.157	81.792
Filament4.11	6	5	0.286	4	1.2	0	0.333	0.176	0.111	1.581	21.92
Filament4.12	2	1	0.032	1	2	0	1	0.005	0.032	0.16	-80.873
Filament4.13	14	13	0.443	9	1.076	0	0.142	0.169	0.158	1.067	-4.488
Filament4.14	6	5	0.149	3	1.2	0	0.333	0.075	0.071	1.060	-17.872
Filament4.15	4	3	0.073	3	1.333	0	0.5	0.044	0.037	1.197	9.183
Filament4.16	16	15	0.555	13	1.066	0	0.125	0.174	0.235	0.741	-16.886
Filament4.17	5	4	0.138	3	1.25	0	0.4	0.077	0.077	0.998	-18.657
Filament4.18	2	1	0.030	1	2	0	1	0.029	0.006	4.804	-11.756
Filament1.19	6	5	0.210	4	1.2	0	0.33	0.148	0.109	1.354	-28.450
Filament4.20	6	5	0.265	5	1.2	0	0.333	0.071	0.204	0.349	2.153
Filament4.21	11	10	0.288	8	1.1	0	0.181	0.095	0.167	0.572	-21.279
Filament4.22	2	1	0.007	1	2	0	1	0.003	0.006	0.589	59.46
Filament4.23	5	4	0.143	4	1.25	0	0.4	0.139	0.03	4.597	-8.690
Filament4.24	10	9	0.351	8	1.11	0	0.2	0.16	0.137	1.173	19.219
Filament4.25	3	2	0.051	2	1.5	0	0.66	0.037	0.030	1.237	-35.103
Filament4.26	2	1	0.019	1	2	0	1	0.001	0.019	0.063	-86.377
Filament4.27	2	1	0.008	1	2	0	1	0.063	0.053	1.187	40.08

Table 7.0.4: Tree Diagnosis for MSTs in the sky region of Filament 4, 27 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament _{5.1}	39	38	1.414	22	1.026	0	0.051	0.429	0.478	0.897	5.025
Filament _{5.2}	14	13	0.427	7	1.076	0	0.142	0.185	0.161	1.147	1.053
Filament _{5.3}	130	129	4.954	46	1.007	0	0.015	1.056	0.743	1.419	4.034
Filament _{5.4}	3	2	0.092	2	1.5	0	0.666	0.091	0.014	6.337	-9.938
Filament _{5.5}	2	1	0.048	1	2	0	1	0.020	0.044	0.461	-65.209
Filament _{5.6}	53	52	2.013	26	1.01	0	0.037	0.513	0.397	1.292	-1.815
Filament _{5.7}	71	70	2.783	35	1.014	0	0.028	0.476	0.870	0.547	-3.655
Filament _{5.8}	4	3	0.133	3	1.333	0	0.5	0.048	0.083	0.585	6.203
Filament _{5.9}	2	1	0.080	1	2	0	1	0.055	0.058	0.949	-46.476
Filament _{5.10}	4	3	0.135	3	1.333	0	0.5	0.129	0.027	4.666	-3.505
Filament _{5.11}	3	2	0.115	2	1.5	0	0.66	0.080	0.048	1.652	-20.755
Filament _{5.12}	2	1	0.009	1	2	0	1	0.001	0.009	0.138	-82.128
Filament _{5.13}	5	4	0.180	3	1.25	0	0.4	0.097	0.066	1.464	37.491
Filament _{5.14}	2	1	0.071	1	2	0	1	0.057	0.042	1.345	-36.621
Filament _{5.15}	2	2	0.031	1	2	0	1	0.028	0.013	2.160	24.84
Filament _{5.16}	5	4	0.149	4	1.25	0.	0.4	0.014	0.143	0.098	-31.427
Filament _{5.17}	7	6	0.338	5	1.166	0	0.285	0.156	0.166	0.939	-15.616
Filament _{5.18}	2	1	0.053	1	2	0	1	0.037 7 0.037	0.988	-45.321	
Filament _{5.19}	5	4	0.132	4	1.25	0	0.4	0.041	0.11	0.35	-24.587
Filament _{5.20}	2	1	0.027	1	2	0	1	0.020	0.017	1.196	-39.884
Filament _{5.21}	2	1	0.013	1	2	0	1	0.011	0.006	1.839	-28.530

Table 7.0.5: Tree Diagnosis for MSTs in the sky region of Filament 5, 21 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament6.1	127	126	4.798	50	1.007	0	0.015	1.275	0.554	2.301	-0.759
Filament6.2	24	23	0.800	15	1.043	0	0.083	0.244	0.323	0.755	-0.326
Filament6.3	2	1	0.011	1	2	0	1	0.003	0.010	0.358	-70.291
Filament6.4	3	2	0.092	2	1.5	0	0.666	0.080	0.024	3.311	-26.494
Filament6.5	15	14	0.510	10	1.071	0	0.133	0.176	0.282	0.624	7.677
Filament6.6	6	5	0.277	4	1.2	0	0.33	0.187	0.116	1.609	18.75
Filament6.7	6	5	0.135	5	1.2	0	0.33	0.11	0.05	2.043	-26.991
Filament6.8	56	55	2.411	33	1.018	0	0.035	0.743	0.626	1.186	-2.114
Filament6.9	4	3	0.054	2	1.33	0	0.5	0.02	0.031	0.647	29.771
Filament6.10	5	4	0.125	4	1.25	0	0.4	0.032	0.102	0.312	13.922
Filament6.11	23	22	0.832	14	1.045	0	0.086	0.365	0.256	1.426	-12.232
Filament6.12	18	17	0.705	12	1.058	0	0.11	0.297	0.234	1.266	0.835
Filament6.13	5	4	0.173	4	1.25	0	0.4	0.114	0.113	1.005	-35.150
Filament6.14	2	1	0.038	1	2	0	1	0.015	0.035	0.449	65.778
Filament6.15	21	20	0.803	13	1.05	0	0.095	0.371	0.345	1.074	2.151
Filament6.16	9	8	0.384	6	1.125	0	0.22	0.131	0.25	0.525	-6.901
Filament6.17	5	4	0.144	4	1.25	0	0.4	0.117	0.049	2.390	11.452
Filament6.18	2	1	0.019	1	2	0	1	0.017	0.00963	1.782	-29.286
Filament6.19	4	3	0.038	3	1.333	0	0.5	0.0321	0.020	1.595	-33.478
Filament6.20	2	1	0.071	1	2	0	1	0.067	0.022	2.987	18.508
Filament6.21	2	1	0.074	1	2	0	1	0.007	0.073	0.104	-84.056
Filament6.22	2	1	0.057	1	2	0	1	0.011	0.056	0.205	-78.410
Filament6.23	2	1	0.028	1	2	0	1	0.023	0.016	1.420	35.143
Filament6.24	3	2	0.033	2	1.5	0	0.666	0.023	0.020	1.145	-48.169
Filament6.25	2	1	0.084	1	2	0	1	0.031	0.078	0.408	-67.788
Filament6.26	3	2	0.066	2	1.5	0	0.66	0.032	0.032	1.019	-19.069

Table 7.0.6: Tree Diagnosis for MSTs in the sky region of Filament 6, 26 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament9.1	523	522	18.008	168	1.001	0	0.003	2.731	1.036	2.634	-0.004
Filament9.2	2	1	0.097	1	2	0	1	0.085	0.045	1.884	27.950
Filament9.3	8	7	0.342	6	1.142	0	0.25	0.228	0.136	1.675	-4.242
Filament9.4	12	11	0.503	10	1.090	0	0.166	0.324	0.14	2.307	-18.691
Filament9.5	3	2	0.094	2	1.5	0	0.666	0.081	0.047	1.718	-33.472
Filament9.6	5	4	0.291	4	1.25	0	0.4	0.213	0.075	2.830	3.016
Filament9.7	22	21	0.995	18	1.047	0	0.090	0.610	0.111	5.468	-2.734
Filament9.8	3	2	0.047	2	1.5	0	0.666	0.016	0.040	0.413	65.247
Filament9.9	34	33	1.252	23	1.030	0	0.058	0.297	0.454	0.655	-0.029
Filament9.10	11	10	0.459	8	1.1	0	0.181	0.220	0.182	1.208	9.660
Filament9.11	6	5	0.11	4	1.2	0	0.33	0.056	0.053	1.054	14.821
Filament9.12	2	1	0.014	1	2	0	1	0.004	0.014	0.329	71.761
Filament9.13	7	6	0.399	6	1.166	0	0.285	0.202	0.122	1.652	14.436
Filament9.14	2	1	0.053	1	2	0	1	0.038	0.037	1.033	44.054
Filament9.15	4	3	0.093	2	1.333	0	0.5	0.067	0.037	1.821	-25.197
Filament9.16	2	1	0.057	1	2	0	1	0.053	0.020	2.573	-21.23
Filament9.17	2	1	0.067	1	2	0	1	0.043	0.052	0.829	-50.30
Filament9.18	3	2	0.093	2	1.5	0	0.66	0.047	0.064	0.743	40.558
Filament9.19	2	1	0.019	1	2	0	1	0.009	0.017	0.535	-61.842
Filament9.20	3	2	0.063	2	1.5	0	0.666	0.017	0.058	0.292	-11.723
Filament9.21	2	1	0.078	1	2	0	1	0.077	0.010	7.474	-7.619
Filament9.22	4	3	0.119	3	1.333	0	0.5	0.071	0.072	0.978	41.661
Filament9.23	2	1	0.012	1	2	0	1	0.007	0.009	0.778	52.100
Filament9.24	3	2	0.068	2	1.5	0	0.666	0.052	0.026	1.998	29.716
Filament9.25	2	1	0.028	1	2	0	1	0.008	0.027	0.307	-72.911
Filament9.26	3	2	0.096	2	1.5	0	0.666	0.059	0.075	0.782	10.484
Filament9.27	2	1	0.0624	1	2	0	1	0.057	0.024	2.292	23.562
Filament9.28	2	1	0.028	1	2	0	1	0.020	0.019	1.063	-43.240

Table 7.0.7: Tree Diagnosis for MSTs in the sky region of Filament 9, 28 detected candidates

Filament No.	No. of Nodes	Tree Size	Tree Length	Tree Longest Path	Average Degree	Clustering Coefficient	Tree Density	Length on X-axis	Length on Y-axis	Length Ratio	Avg Angle of Inclination
Filament10.1	356	355	13.325	99	1.002	0	0.005	1.818	1.249	1.455	1.480
Filament10.2	101	100	2.833	56	1.01	0	0.019	0.738	0.494	1.493	-1.921
Filament10.3	23	22	0.865	14	1.045	0	0.086	0.276	0.326	0.846	-1.127
Filament10.4	21	20	0.898	12	1.05	0	0.095	0.498	0.189	2.631	5.632
Filament10.5	30	29	0.98	17	1.034	0	0.066	0.440	0.248	1.773	0.767
Filament10.6	37	36	1.158	19	1.027	0	0.054	0.578	0.24	2.411	-11.763
Filament10.7	14	13	0.454	12	1.076	0	0.142	0.267	0.237	1.125	1.347
Filament10.8	28	27	1.243	17	1.034	0	0.071	0.622	0.328	1.893	-9.566
Filament10.9	68	67	2.025	34	1.014	0	0.029	0.499	0.489	1.020	4.017
Filament10.10	5	4	0.112	4	1.25	0	0.4	0.045	0.068	0.661	18.948
Filament10.11	2	1	0.072	1	2	0	1	0.055	0.046	1.194	-39.933
Filament10.12	8	7	0.511	5	1.142	0	0.25	0.210	0.232	0.903	27.583
Filament10.13	2	1	0.012	1	2	0	1	0.009	0.008	1.110	-41.99
Filament10.14	2	1	0.015	1	2	0	1	0.014	0.004	3.335	-16.688
Filament10.15	5	4	1.25	0	0.4	0.08	0.09	0.867	-11.873		
Filament10.16	2	1	0.074	1	2	0	1	0.073	0.013	5.392	-10.505
Filament10.17	3	2	0.060	2	1.5	0	0.666	0.015	0.055	0.276	-3.998
Filament10.18	2	1	0.065	1	2	0	1	0.065	0.002	24.867	2.302
Filament10.19	4	3	0.227	3	1.33	0	0.5	0.083	0.147	0.568	49.11
Filament10.20	3	2	0.019	2	1.5	0	0.666	0.008	0.010	0.756	-51.608
Filament10.21	5	4	0.209	3	1.25	0	0.4	0.087	0.114	0.761	7.524
Filament10.22	9	8	0.379	8	1.125	0	0.22	0.131	0.205	0.639	13.879
Filament10.23	4	3	0.068	3	1.333	0	0.5	0.054	0.040	1.349	-36.765
Filament10.24	2	1	0.031	1	2	0	1	0.006	0.031	0.191	-79.137
Filament10.25	2	1	0.070	1	2	0	1	0.065	0.027	2.374	-22.834
Filament10.26	3	2	0.125	2	1.5	0	0.666	0.063	0.061	1.018	-43.749
Filament10.27	2	1	0.019	1	2	0	1	0.010	0.016	0.643	-57.236
Filament10.28	2	1	0.093	1	2	0	1	0.016	0.092	0.181	79.707

Table 7.0.8: Tree Diagnosis for MSTs in the sky region of Filament 10, 28 detected candidates

8

Appendix C

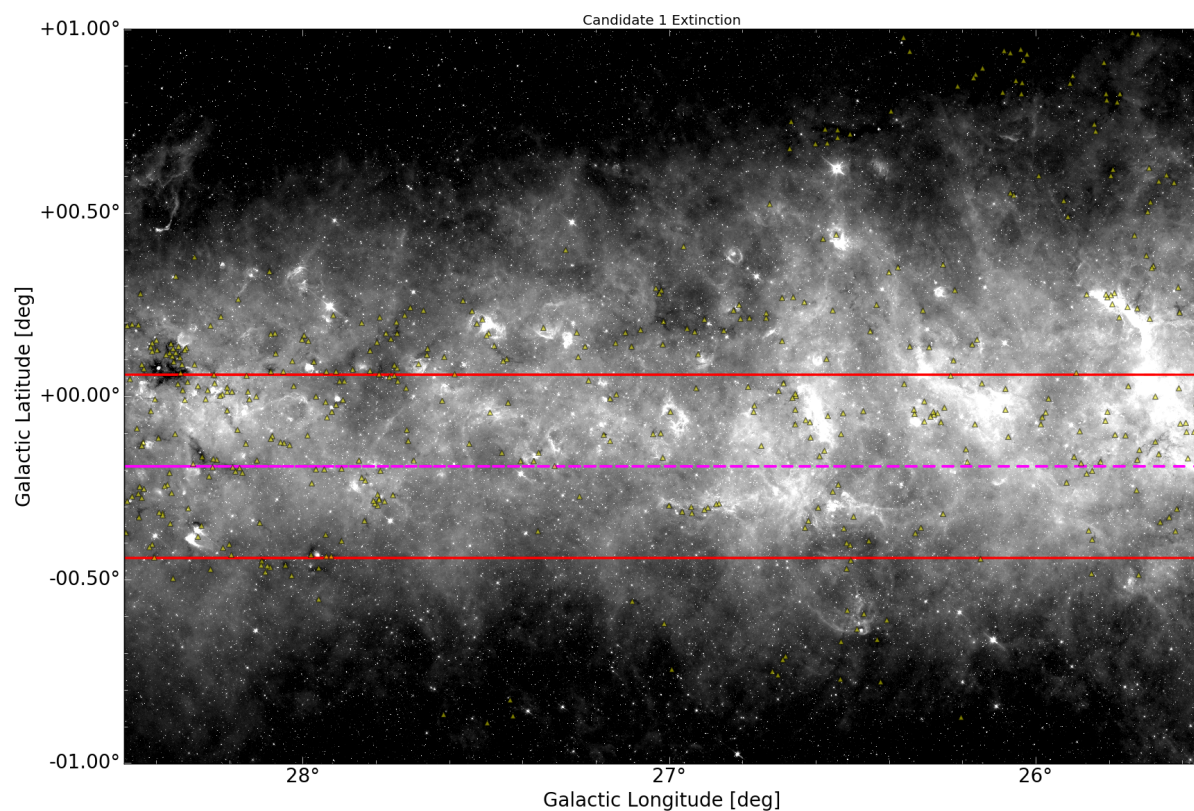


Figure 8.0.1: Filament1 GLIMPSE image, with overlaid Peretto Fuller clouds.

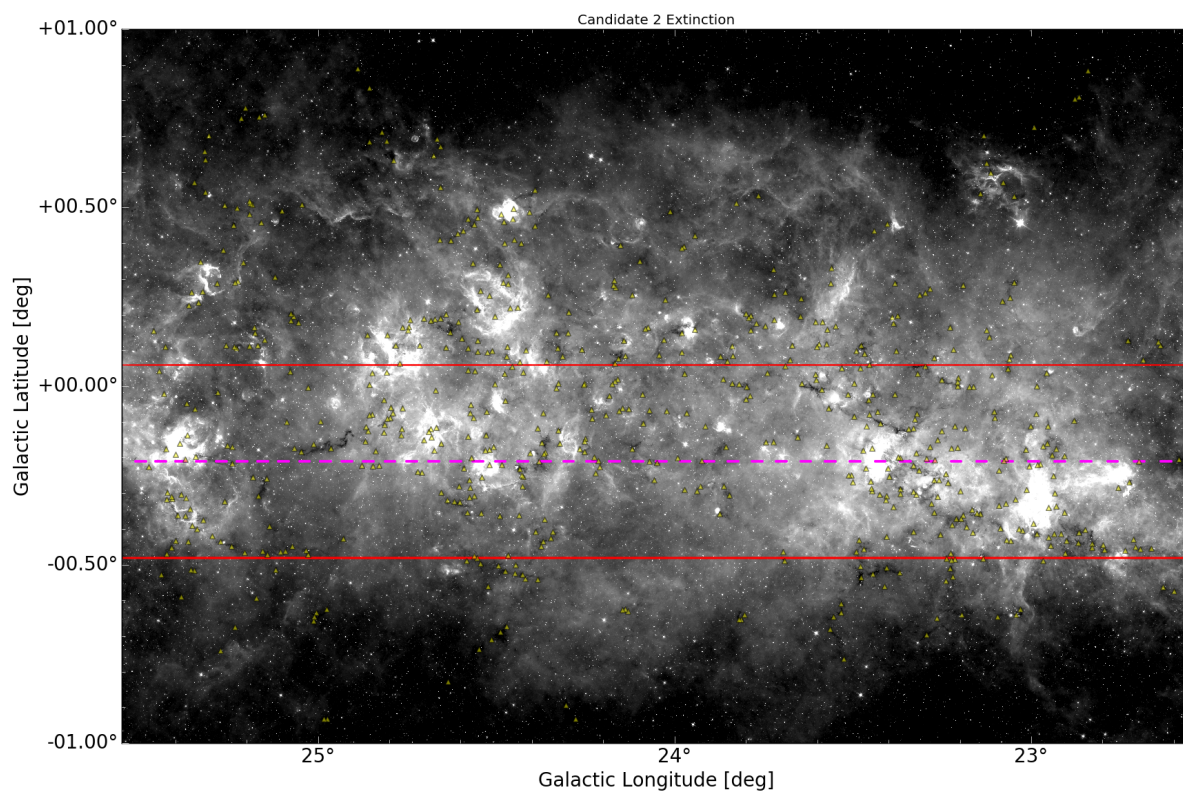


Figure 8.0.2: Filament2 GLIMPSE image, with overlaid Peretto Fuller clouds.

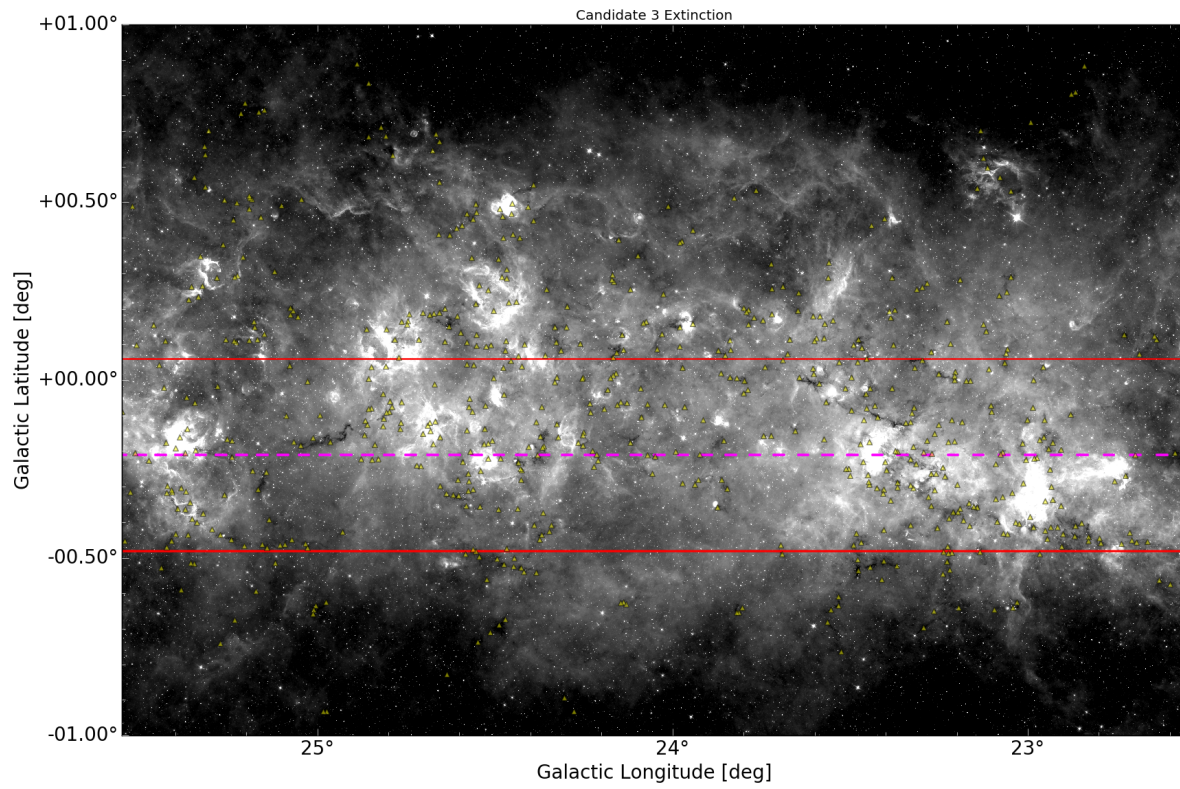


Figure 8.0.3: Filament3 GLIMPSE image, with overlaid Peretto Fuller clouds.

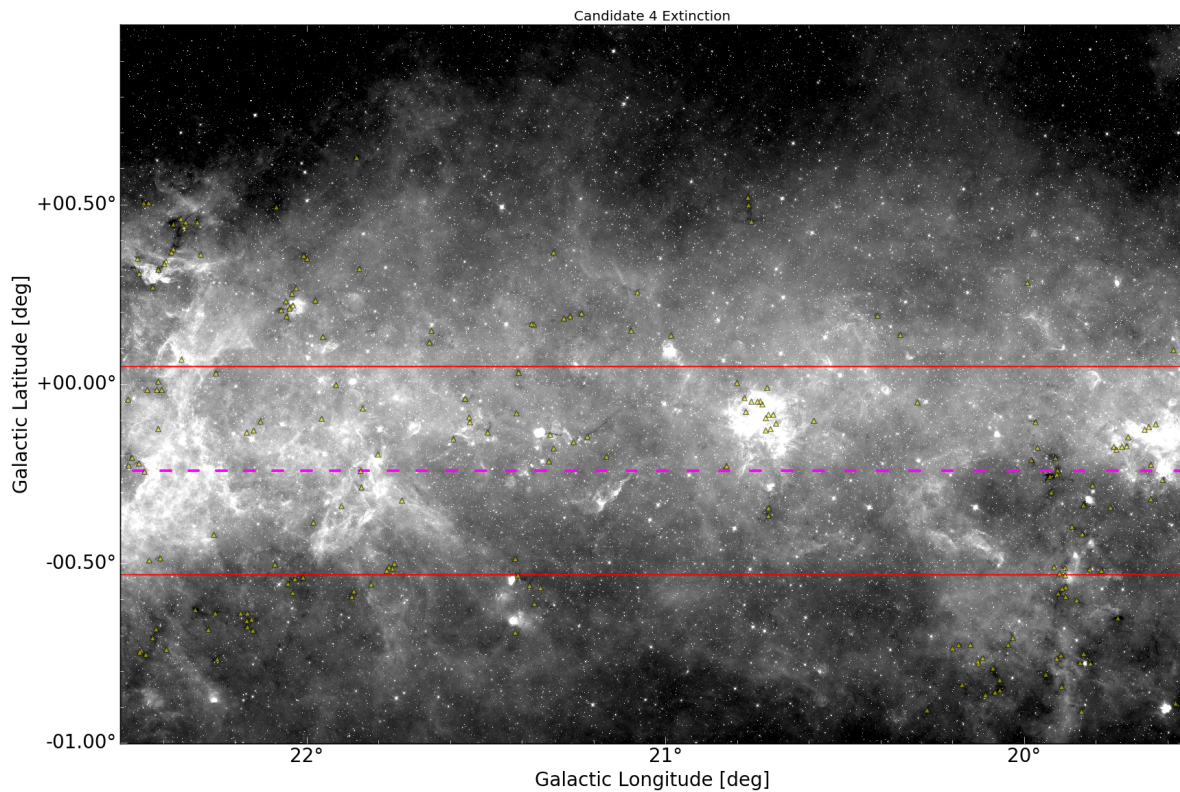


Figure 8.0.4: Filament4 GLIMPSE image, with overlaid Peretto Fuller clouds.

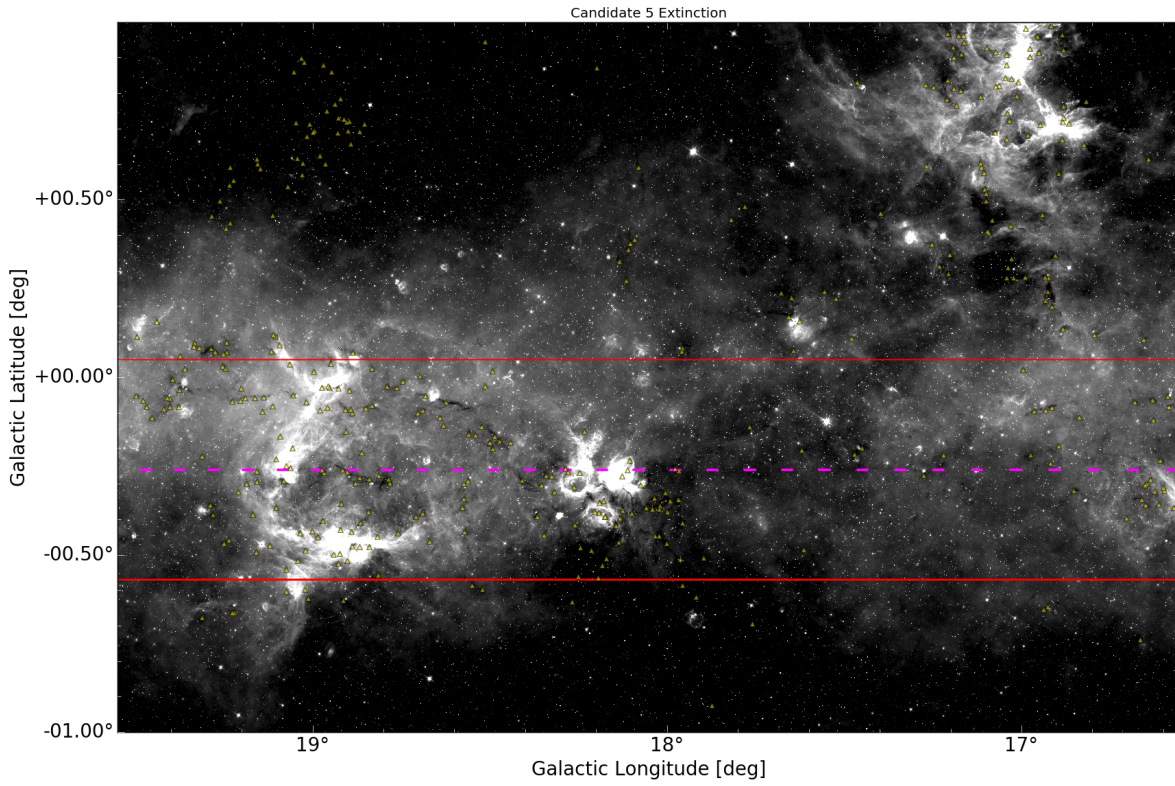


Figure 8.0.5: Filament5 GLIMPSE image, with overlaid Peretto Fuller clouds.

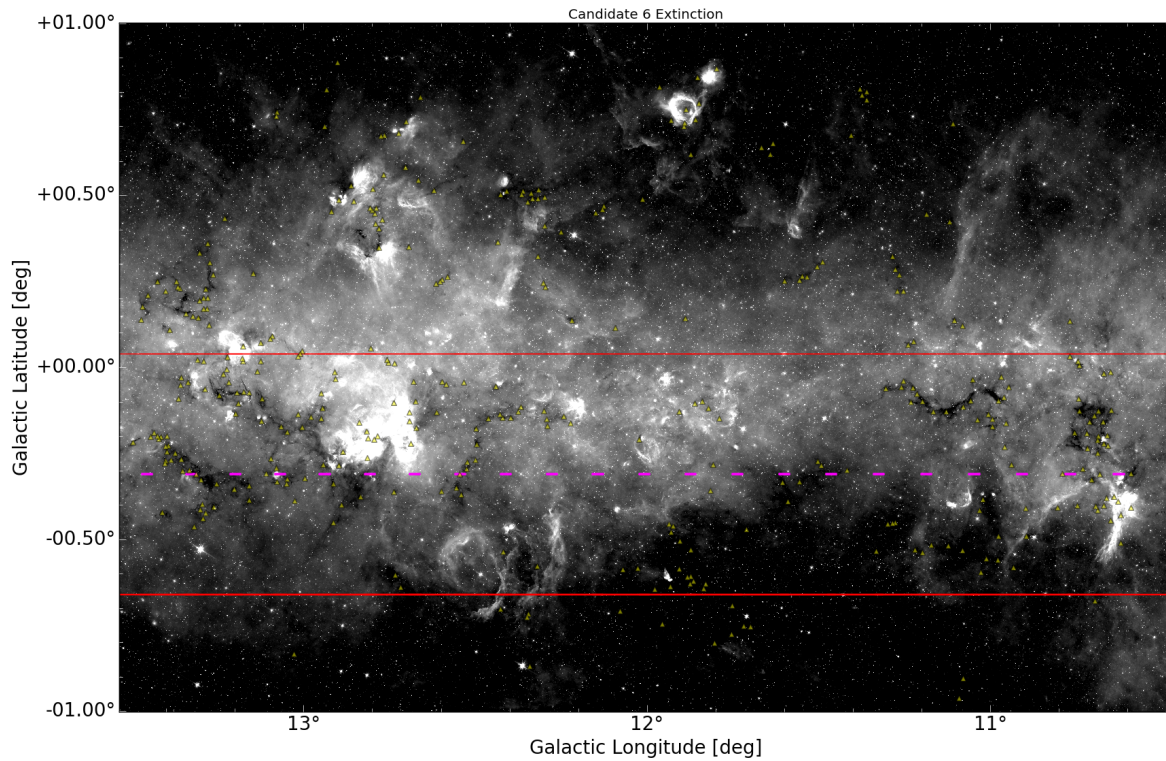


Figure 8.0.6: Filament6 GLIMPSE image, with overlaid Peretto Fuller clouds.

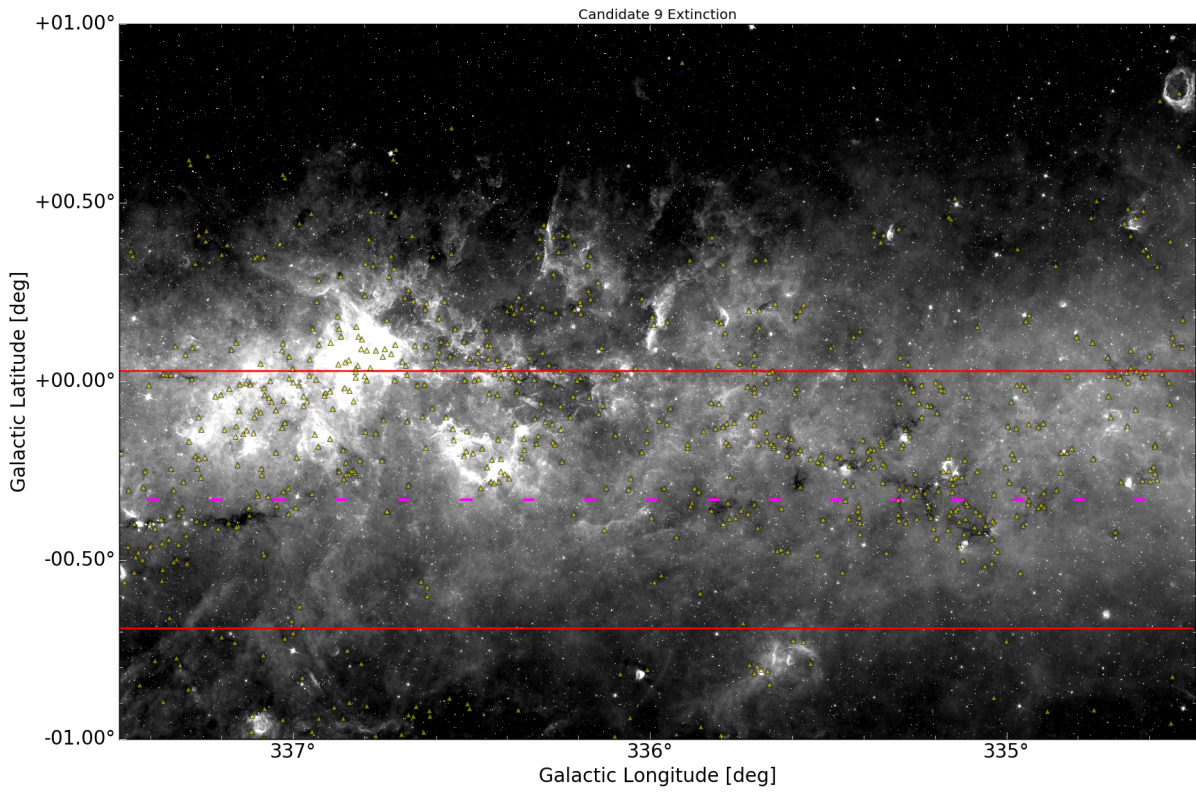


Figure 8.0.7: Filament9 GLIMPSE image, with overlaid Peretto Fuller clouds.

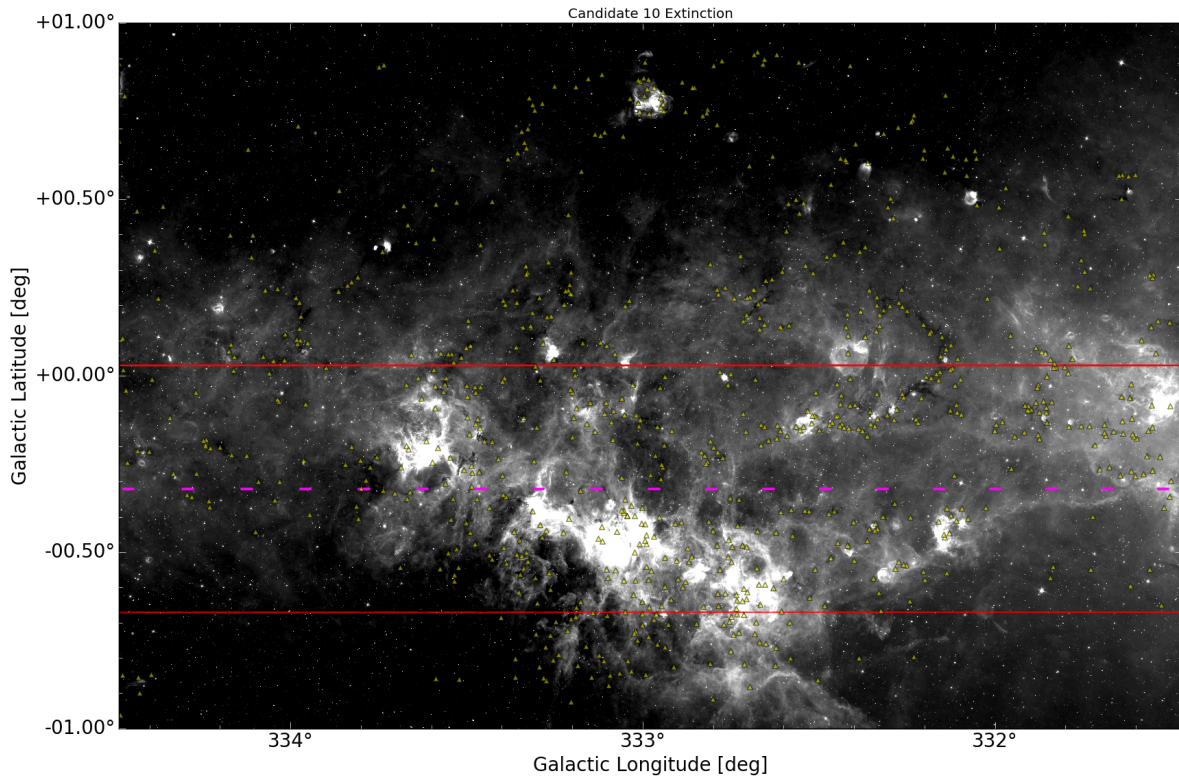


Figure 8.0.8: Filament10 GLIMPSE image, with overlaid Peretto Fuller clouds.

9

Appendix D

I AM INCLUDING BELOW THE PYTHON CORE SCRIPT which I have been using for generating and drawing Minimum Spanning Trees. The example below plots a GLIMPSE file with overlaid IRDCs from the Peretto & Fuller catalog, according to the region of *Filament1*. I am including all my other python scripts and necessary files in the github account situated at <https://github.com/amconstantin/SeniorThesis.git>.

```

import networkx as nx

import matplotlib.pyplot as plt

import csv

import numpy as np

import matplotlib as mpl

import aplpy

from ds9norm import DS9Normalize

import matplotlib.patches as mpatches

import matplotlib.lines as mlines

import math

import decimal

from astropy.table import Table

import xlwt

from xlwt import *

# path to file containing galactic coordinates of Peretto & Fuller
flb = open('C:\Users\UR Device Program\Documents\Thesis/cloudfitsLB.txt')

#create arrays for storing coordinates of molecular clouds

lat = []

lng = []

```

```

N = 0

#create networkx Graph, to be populated
#with points from Peretto & Fuller catalog
G = nx.Graph()

for line in flb:
    columns = line.split(", ")

    if ((float(columns[0]) >= 25.5) and (float(columns[0]) <= 28.5) and
        (float(columns[1])) <= 1 and (float(columns[1]) >= (-1))):
        lat.append(float(columns[0]))
        lng.append(float(columns[1]))
        G.add_node(str(N), posxy = (lat[-1], lng[-1]))
        N += 1

positions = nx.get_node_attributes(G, 'posxy')

# build edges of graph, within a certain distance threshold
for n,d in G.nodes_iter(data=True):
    xn, yn = positions[n]
    for m,g in G.nodes_iter(data=True):
        xm, ym = positions[m]

```

```

dist = (math.sqrt(math.pow((xm - xn),2) + math.pow((ym - yn),2)))

if (dist <= 0.1):

    G.add_edge(n, m, weight = dist)


myfig = plt.figure(figsize=(20,20))

fig=aplpy.FITSFigure("C:\Users\UR Device Program\OneDrive\Documents\
Senior Year\Spring Semester\Senior Thesis\Extinction Filaments Data\
Candid1_cropped_final.fits",figure=myfig, zorder = 0)

norm = DS9Normalize(stretch='sqrt', clip_hi=99,clip_lo=1, contrast=1.56, bi
norm.update_clip(fig._data)

fig.show_grayscale()

fig.image.set_norm(norm)

fig.show_grayscale(vmin=0, vmax=145)

fig.image.set_norm(norm)


fig.tick_labels.set_font(family='sans-serif', size='large')
fig.tick_labels.set_xformat("dd")
fig.tick_labels.set_yformat("d.dd")
fig.axis_labels.set_font(family='sans-serif',size='large')
fig.axis_labels.set_xtext("Galactic Longitude [deg]")
fig.axis_labels.set_ytext("Galactic Latitude [deg]")
plt.title("Filament1 Kruskal MST, threshold 0.1deg")

```

```

fig.axis_labels.set_font(size=20)
fig.tick_labels.set_font(size=20)

scutumdata=Table.read("C:\Users\UR Device Program\OneDrive\Documents\
Senior Year\Spring Semester\Senior Thesis\
Extinction Filaments Data\Scutum_final.txt",
format='ascii', delimiter='\t', guess=False)
ws = (scutumdata['long_final'] > 25.5) & (scutumdata['long_final'] < 28.5)
arml=scutumdata['long_final'][ws]
num_rows=arml.size

cmap=plt.cm.get_cmap("gist_rainbow")
cmap.set_under(color="k")
cmap.set_over(color="k")
norm=mpl.colors.Normalize(clip=False,vmin=65,vmax=85)
m=plt.cm.ScalarMappable(cmap=cmap,norm=norm)

#Filament 1 coordinates
upperb=[0.06]*num_rows
armb=[-0.19]*num_rows
lowerb=[-0.44]*num_rows
armv = scutumdata['Vlsr'][ws]

```

```

xbox = [26.8, 27.1, 27.1, 26.8, 26.8]
ybox = [-0.23, -0.23, -0.4, -0.4, -0.23]
box = [np.vstack((xbox, ybox))]

fig.show_rectangles(arml, armb, 0.03, 0.002, color = "magenta", cmap = cmap, norm=
fig.show_rectangles(arml, upperb, 0.3, 0.001, color="red", cmap = cmap, norm=
fig.show_rectangles(arml, lowerb, 0.3, 0.001, color="red", cmap = cmap, norm=
fig.show_markers(lat, lng, s = 30, alpha = 1, zorder = 2, marker = '^', c = 'ye
fig.show_lines(box, color = "crimson", linewidth = 3)

# using Kruskal's algorithm
T = nx.minimum_spanning_tree(G)

# using Prim's algorithm
# T = nx.prim_mst(G)

#print edges of the MST and overlay them on top of GLIMPSE file
pos_mst = nx.get_node_attributes(T, 'posxy')
for each in sorted(T.edges(data=True)):
    x1, y1 = pos_mst[each[0]]
    x2, y2 = pos_mst[each[1]]

```

```

mst_long = [x1, x2]
mst_lat = [y1, y2]
edge = [np.vstack((mst_long, mst_lat))]
fig.show_lines(edge, color = 'aquamarine', linewidth = 2.5)

# Tree Diagnosis

# number of trees in the graph
ntrees = nx.number_connected_components(T)
print "number of trees = " + str(ntrees)

#generate lists of individual trees
tree_list = nx.connected_component_subgraphs(T)

i = 1

w = Workbook()
ws = w.add_sheet('Filament1')

for each in tree_list:
    #number of nodes
    nnodes = each.number_of_nodes()

    #size
    tree_size = each.size()

```

```

#tree total length

tree_length = each.size(weight = 'weight')


#tree diameter

#calculated in number of vertices which must be visited in order to travel

diam = nx.diameter(each)


#average degree

avg_deg = float(nnodes) / tree_size


#clustering coefficient

cc = nx.clustering(each)

avg_cc = sum(cc.values()) / len(cc)


#tree density

#for undirected graphs, tree density =  $2m / (n*(n-1))$ , m = # of edges, n = #

rho = nx.density(each)


#node centralities

#measuring length/ width ratio of tree

pos_tree = nx.get_node_attributes(each, 'posxy')

```

```

x_val = []
y_val = []
for n,d in each.nodes_iter(data=True):
    xn, yn = pos_tree[n]
    x_val.append(xn)
    y_val.append(yn)
x_max = max(x_val)
x_min = min(x_val)
y_max = max(y_val)
y_min = min(y_val)
if (x_min <= 26.94 <= x_max) and (y_min <= -0.3 <= y_max):
    bone_likelihood = 1
else:
    bone_likelihood = 0
delta_x = x_max - x_min
delta_y = y_max - y_min
length_ratio = delta_x / delta_y

#average inclination angle
angle = []
for line in sorted(each.edges(data=True)):
    x1, y1 = pos_tree[line[0]]

```

```

        x2, y2 = pos_tree[line[1]]

        line_long = x2 - x1

        line_lat = y2 - y1

        angle.append(math.degrees(math.atan(line_lat/ line_long)))

    avg_angle = np.mean(angle)

tree = [nnodes, tree_size, tree_length, diam, avg_deg, avg_cc, rho, delta]

    for col, col_value in enumerate(tree):

        ws.write(i, col, col_value)

    i += 1

w.save('tree1.xls')

plt.savefig("Filament1 Kruskal 0.1 MST.png")

plt.show()

```

Colophon

THIS THESIS WAS TYPESET
using \LaTeX , originally
developed by Leslie

Lamport and based on Donald
Knuth's \TeX . The body text is set in 11
point Arno Pro, designed by Robert
Slimbach in the style of book types
from the Aldine Press in Venice, and
issued by Adobe in 2007. A template,
which can be used to format a PhD
thesis with this look and feel, has been
released under the permissive MIT
(X11) license, and can be found online
at github.com/suchow/ or from the
author at suchow@post.harvard.edu.