Tracing Filaments

Ntsikelelo Luvuyo Tshangela

28 June 2025

1 Abstract

This paper investigates the *Missing Baryon Problem* by exploring how the Friends-of-Friends (FoF) algorithm can be applied to redshift surveys of the cosmic web to identify filamentary structures. These structures could offer clues about the properties of the intergalactic medium (IGM) and help uncover the location of the universe's missing baryons. According to current cosmological theories, the universe is composed of roughly 5% baryonic matter, 27% dark matter, and 68% dark energy. While at high-redshift ($z > 2$) observations account for nearly all the expected baryons, at $z = 0$ nearly half appear unaccounted for. This discrepancy may arise because a substantial portion ($40\% - 50\%$) of baryons resides in the IGM, distributed across different phases. Cen and Ostriker (2006) proposed that the majority of these missing baryons are likely located in the Warm-Hot Intergalactic Medium (WHIM), a tenuous plasma with densities 10 to 20 times the mean baryonic matter density ($4 \cdot 10^{-28} kg \cdot m^{-3}$) and temperatures around 10^7k radiating primarily in the soft X-ray regime. However, detecting this radiation is challenging, prompting researchers to explore alternative methods to observe the IGM and characterize the WHIM in hopes of resolving the Missing Baryon Problem. The goals set out for our experiment are to visualize the cosmic web using redshift survey called Two-degree-Field Galaxy Redshift Survey (2dFGRS), and then sub-divide the data into groups(clusters) and filaments.

2 Introduction

2.1 Scientific Background

According to the currently accepted cosmological models, the Big Bang occurred approximately 13.8 billion years ago. The early universe was extremely hot and nearly uniform, and was filled with blackbody radiation. As it expanded, it cooled enough for matter to form, and small deviations from uniformity began to emerge. These variations, reflected in the cosmic microwave background (CMB) as tiny fluctuations in temperature and density, are attributed primarily to dark matter (Simcoe, R. 2004). By the time matter and radiation decoupled, these

fluctuations had created regions with slightly higher matter density (Cen and Ostriker, 1999). Gravitational attraction in these regions drew in surrounding material, initiating the formation of large-scale structure. Over time, this gravitational clumping developed into the cosmic web and its substructures (Simcoe, R. 2004). Libeskind et al. (2017) describe the cosmic web as a vast arrangement of galaxies, intergalactic gas, and dark matter configured in a filamentary and sheet-like pattern, surrounding enormous, nearly empty voids. These over dense regions, where galaxies are concentrated, are connected by elongated filaments of dark matter and gas that act as pathways, directing matter toward denser hubs such as galaxy clusters and superclusters. Between these structures lie vast cosmic voids, distinguished by their extremely low matter density, typically with an over density of ≤ 0.8, as noted by Colberg et al. (2005). The cosmic web represents the largest gravitationally bound structure in the observable universe (Santiago-Bautista et al., 2020).

Measurements of galaxy cluster masses consistently reveal a discrepancy between the observable baryonic matter and the total inferred mass.This suggests that dark matter constitutes the majority of the cluster mass (Carroll & Ostlie, 2007). Most galaxies reside within dark matter halos, and the intergalactic medium (IGM) may also remain gravitationally bound due to the pervasive distribution of dark matter throughout the cloud.

The cosmic web's structure is shaped by galaxy groups and clusters, bound not by individual interactions but by their collective immersion in a shared, dark matter-dominated gravitational potential. This extended halo sustains the system, with groups typically containing fewer than 50 galaxies and clusters ranging from 50 to over a thousand (Carroll & Ostlie, 2007).

Clusters are classified as *rich* or *poor* based on galaxy count: rich clusters host hundreds to thousands of galaxies, while poor clusters contain relatively few (Carrol & Ostlie, 2007). Larger collections of clusters form *superclusters*, which are among the most massive structures in the universe (Porter & Raychaudhury, 2005). These superclusters mark the nodes of the cosmic web and are interconnected by vast filamentary bridges of dark matter and gas.

Everywhere in the cosmic web the intergalactic medium is present. The IGM refers to the sparse, diffuse material, primarily gas, that exists in the spaces between galaxies (Barkana & Loeb, 2007). According to Davé et al. (2018), the IGM has four phases, which depend on temperature and over density δ:

1. Diffused gas (photoionized gas): $\delta < 1000$, $T < 10^5$ K
2. condensed gas (Stars and cool galactic gas) : $\delta > 1000$, $T < 10^5$ K
3. Warm-Hot (The Warm-Hot Intergalactic Medium (WHIM)): $10^5 < T < 10^7$ K
4. Hot (Gas in galaxy clusters and large groups): $T > 10^7$ K

At high redshifts ($z > 2$), baryonic matter primarily exists as diffuse ionized gas due to the absence of significant cosmic structure (Davé et al. 2018). As

structure formation progresses, this gas undergoes shock heating, caused by infalling material entering high-density regions at high velocities, which converts kinetic energy into thermal energy and elevates the gas temperature. This mechanism gives rise to the warm–hot intergalactic medium (WHIM), characterized by temperatures in the range of 10^5–10^7K, leading to emission predominantly in the extreme ultraviolet (EUV) and soft X-ray bands. Simulations by Cen & Ostriker (2006) and Davé et al. (1999) estimate that approximately 40–50% of baryonic matter at the present epoch resides in this phase. However, much of it remains observationally elusive, as the gas is typically infalling and lacks sufficient thermal energy to emit strongly. Furthermore, as noted by Cen & Ostriker (1999), detection is complicated by foreground contamination: the Milky Way's interstellar medium emits strongly in soft X-rays, with an effective temperature near 10^6K, making it difficult to isolate the radiation. Additionally, neutral hydrogen within the Milky Way absorbs EUV radiation, further hindering observations of the WHIM from Earth (Cen & Ostriker, 1999).

Everything that we see in the universe is baryonic matter, the stars, planets, gas, and IGM. However, baryonic matter only accounts for a tiny fraction of the total mass in the universe. According to the ΛCMD model of the universe we know the universe is composed of 5% baryonic matter, 27% dark matter, and 68% dark energy. According to Davé et al. (2018), studies suggest that at high redshifts ($z > 2$), most of the baryons existed in a diffuse state, specifically as photoionized IGM. This material is observable as H I absorption lines in the spectra of distant quasars, and by using this method, the baryonic density can be inferred (Davé et al., 2018). Cen & Ostriker (1999) demonstrated that, although the total baryon density remains consistent with early universe predictions, a substantial fraction of baryons at redshift ($z = 0$) resides in a warm-hot, diffuse phase of the intergalactic medium, rendering them difficult to detect through conventional observational methods. So the big question is: where are the missing baryons?

How can this gas be observed? Cen et al. (1995) claim that the soft X-ray background can be detected by observing large structure features, as well as by the soft X-ray angular autocorrelation function. They further discuss other methods of detecting this gas, for instance:

1. Searching for shadows cast by nearby neutral hydrogen-rich galaxies using AXAF.

2. Simulating the existence of an X-ray absorption forest due to ionized oxygen (O VII 574 eV line) in the warm/hot temperature range (Hellsten, Gnedin, & Miralda-Escude, 1998).

3. Using vacuum UV absorption lines (O VI 1032A, 1038A doublets) to detect gas in the distant outskirts of galaxies (Cen et al., 1999).

4. Observing strong soft X-ray emission lines from highly ionized species, such as O VIII 653 eV (Jahoda et al., 1999).

5. Investigating EUV observations, which suggest that emission structures may align with or be superimposed on rich X-ray clusters, likely due to a warm/hot component (Mittaz, Lieu, & Lockman, 1998).

6. Searching for broad, relatively faint low-redshift $Ly\alpha$ clouds, which require more sensitive observations (Shull 1996, 1997).

7. detection of gas through cross correlating soft X-ray background (Refregier, Helfand, & McMahon 1997)

8. Sunyaev-Zeldovich effect (Refregier, Spergel, & Herbig 1998) with galaxies.

An active galactic nucleus (AGN) is a supermassive black hole powered by accretion that produces jets that are likely deflected by ram pressure from the surrounding intergalactic medium (IGM), as noted by Morris et al. (2022). They also discovered that the extent of jet bending is influenced by the density of the gas through which the AGN host galaxy is travelling, as well as the velocity of the AGN host galaxy relative to the IGM. Another cause for jet bending is that a group or cluster could have merged, resulting in the IGM becoming turbulent and providing the ram pressure necessary to bend the jets out to large distances from the cluster centre. It is for the last two reasons that we expect to find more bent galaxies in denser environments, giving us a way to detect the filaments. The study by Morris et al. (2022) found that, to a statistically significant degree, AGN with bent jets are more likely than their unbent counterparts to be part of groups with three or more members. Additionally, these AGN tend to reside in larger, denser, and more resource-rich environments. Because the IGM is difficult to measure directly, the number density of galaxies can serve as a proxy for IGM gas density. Assuming ram pressure is the dominant mechanism bending AGN jets, Morris et al. (2022) demonstrated that parameters of the WHIM can be estimated from AGN jet properties such as bending angle and jet velocity. So finding the groups of galaxies using algorithms like the Friends of friends algorithm can be crucial step in helping find how the distribution of gas in the IGM affect galaxy morphology and allow us to infer properties about the IGM which could possibly lead us to detecting more of the undiscovered WHIM.

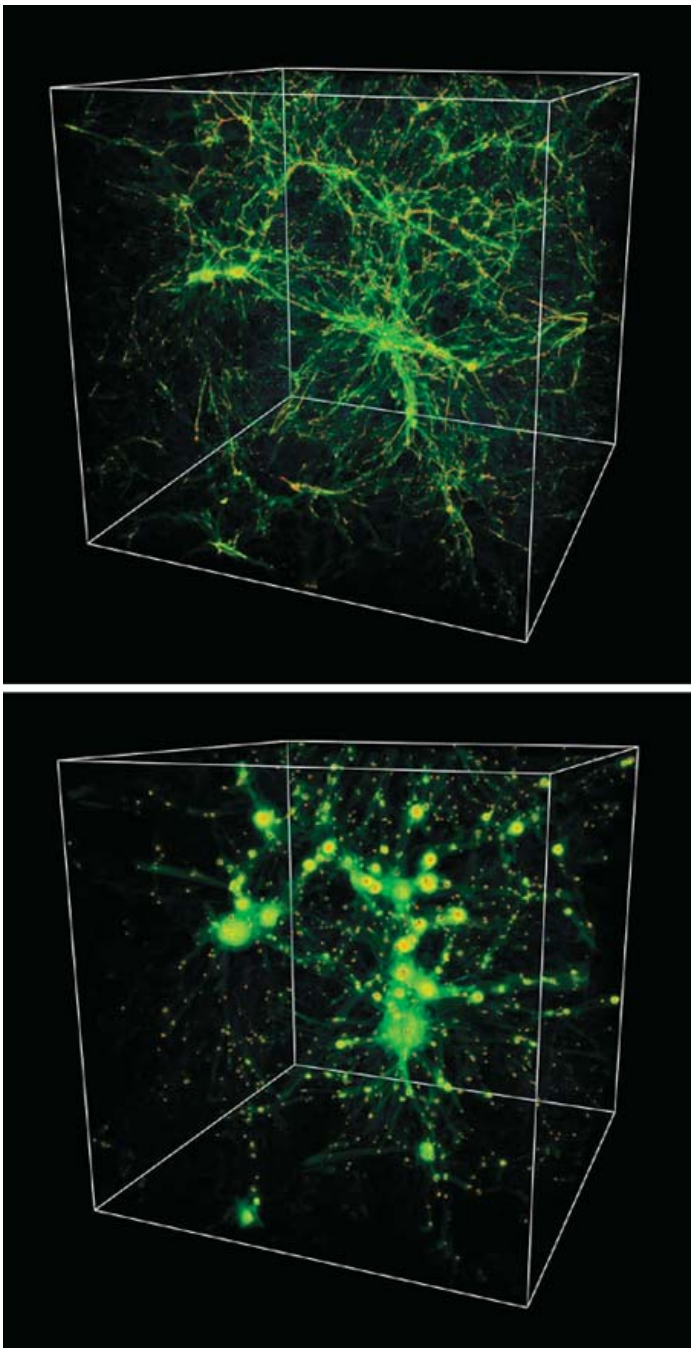

Figure 1: this image shows the cosmic web and its filaments

2.2 Technical Background

As mentioned, intergalactic gas is so tenuous, with densities 10 to 20 times below the mean baryonic matter density ($4 \cdot 10^{-28} kg \cdot m^{-3}$), that it emits no light of its own (Simcoe, 2004). Therefore, to observe it, we must detect it indirectly. To detect it we will be looking at how light from distant objects is affected by the IGM (Simcoe, 2004). Galactic distances are commonly determined by analysing the redshift of spectral lines, either emission from ionized gas or absorption from stellar atmospheres produced within the galaxy itself. These lines shift toward longer wavelengths due to the expansion of the universe, this new wavelength is determined by relativistic Doppler equation

$$\lambda_{observed} = \lambda_{rest} \sqrt{\frac{1 + \frac{v_r}{c}}{1 - \frac{v_r}{c}}}$$

where v_r is the radial velocity of light source, c is the speed of light and λ_{rest} is the wavelength measured in the same reference frame as source and $\lambda_{observed}$ is the measured wavelength. The degree of this redshift, denoted as (z), is defined as

$$z = \frac{\lambda_{observed} - \lambda_{rest}}{\lambda_{rest}}$$

Both formulas allow astronomers to estimate the galaxy's recessional velocity and then approximate its distance via Hubble's law.

$$H_0 = vd$$

where $H_0 = 70 \frac{km}{s \cdot Mpc^{-1}}$ is Hubble's constant, v is the velocity of the object of interest, and d is the distance to the object. Large redshift surveys, such as the Two-degree-Field Galaxy Redshift Survey (2dFGRS), use these measurements to map the three-dimensional distribution of galaxies. By combining redshift data with their angular positions on the sky, we can trace the filamentary structure of the cosmic web and identify its substructures. Consider the situation when the light beam is emitted from the light source towards earth but encounters the IGM first. What occurs the IGM will either absorb or scatter certain wavelengths of light causing the light beam to lose brightness and certain wavelengths of light through the processes of scattering and absorption, and redshift. The loss of photons due to absorption or scattering is called extinction. In an absorption process, the photon ceases to exist and its energy is given up to the thermal energy of the gas. In a scattering process the photon continues on in a different direction. Both absorption and scattering can remove photons from a beam of light. In the case the IGM absorbs the light since it mostly made of hydrogen the wavelengths that are missing the spectrum will be those related to the energy levels of hydrogen. According to Simcoe (2004) because clouds in IGM are at different distances they will have different redshifts the spectrum from light source will have different emission lines below the hydrogen emission line, so using modern telescopes we are able to resolve the intergalactic medium into individual clouds. The intergalactic clouds are the filaments that

make up the cosmic web see Figure 1. A redshift survey is an observation of a section of the sky aimed at measuring the redshift of astronomical objects such as galaxies, galaxy clusters, or quasars. The redshift can be used to estimate the recessional velocity of the galaxy by using the relativistic Doppler effect the following equation can be derived:

$$v = c\frac{(1+z)^2 - 1}{(1+z)^2 + 1}$$

where c is the speed of light and z is the redshift of the galaxy of interest. The distance to a galaxy from Earth is approximated by applying Hubble's law

$$H_0 = vd$$

where $H_0 = 70\frac{km}{s \cdot Mpc^{-1}}$ is Hubble's constant, v is the velocity of the object of interest, and d is the distance to the object. By combining redshift with angular position data, a redshift survey maps the 3D distribution of matter within a field of the sky an example of this is the Two-degree-Field Galaxy Redshift Survey (2dFGRS) see Figure 2. These observations help us understand how matter is spread out across the universe. From the 2dFGRS we can see the filamentary structures and voids that make up the cosmic web. By combining the redshift data and the angular positions on the sky of each galaxy we can use a grouping algorithm to identify the substructures within the cosmic web.

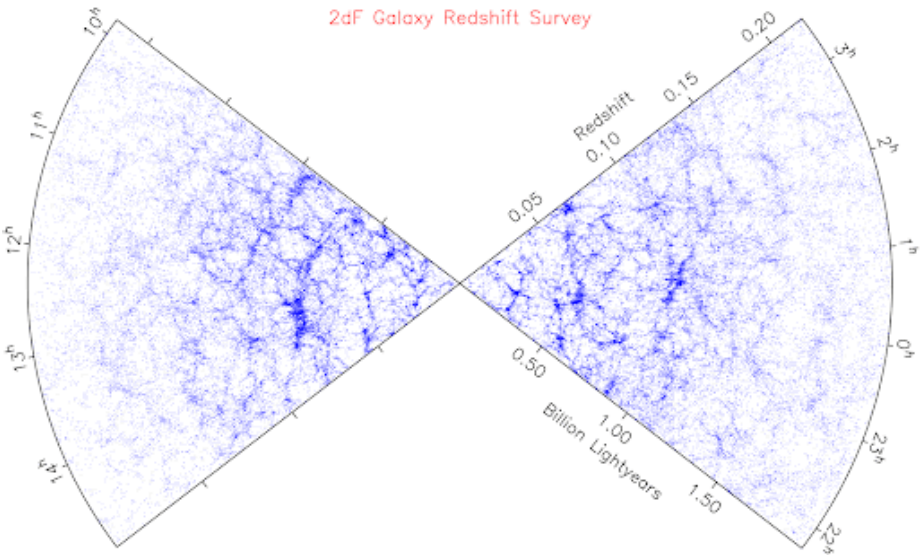

Figure 2: Two-degree-Field Galaxy Redshift Survey (2dFGRS) visualisation of the distribution of matter

3 Data description and analysis

3.1 Data description

3.1.1 Catalogue description

The dataset under investigation originates from the Two-degree-Field Galaxy Redshift Survey (2dFGRS), comprising 245, 591 galaxies. A 3D Cartesian coordinate system is adopted such that the z-axis points toward the celestial north pole, the y-axis aligns with RA = 18h, Dec = 0°, and the x-axis aligns with RA = 6h, Dec = 0°. A quick summary of the spatial distribution is summarised in Table 1.

property	x-coordinate [Mpc]	y-coordinate [Mpc]	z-coordinate [Mpc]
min	-3852.24	-2530.87	-2873.59
max	3540.14	2232.58	75.924
mean	74.411	27.038	-153.123
median	183.171	6.781	-123.598
standard deviation	405.543	180.463	154.75

Table 1: Statistical summary of distribution of galaxies

The mapped survey volume spans approximately 7092.38 Mpc along the x-axis, 4763.45 Mpc along the y-axis, and 2949.21 Mpc along the z-axis. The bimodal X-distribution of 2dFGRS consists of two galaxy concentrations: one at $x > 0$ and one at $x < 0$, see Figure 3.

In contrast, the graph of the distribution of galaxies in the y axis (Y-distribution) is positively skewed, so most of the galaxies will reside $y > 0$ quadrants see Figure 4.

In contrast, the graph of the distribution of galaxies in the z axis (Z-distribution) is negatively skewed, with most galaxies exhibiting $Z < 0$, indicating a preference for southern celestial latitudes Declination is less than zero i.e. ($\delta < 0°$). This asymmetry implies that the 2dFGRS catalogue samples more extensively in the southern celestial hemisphere Figure 5.

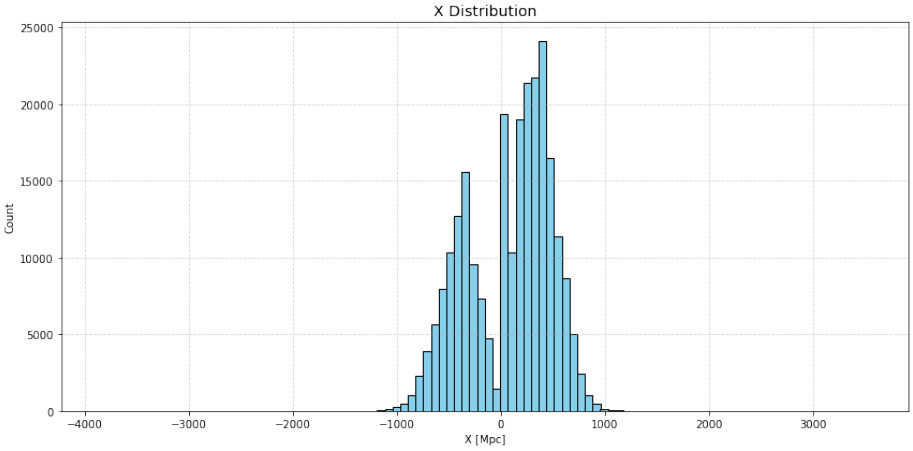

Figure 3: The distribution of galaxies in the 2dFGRS in the x-axis

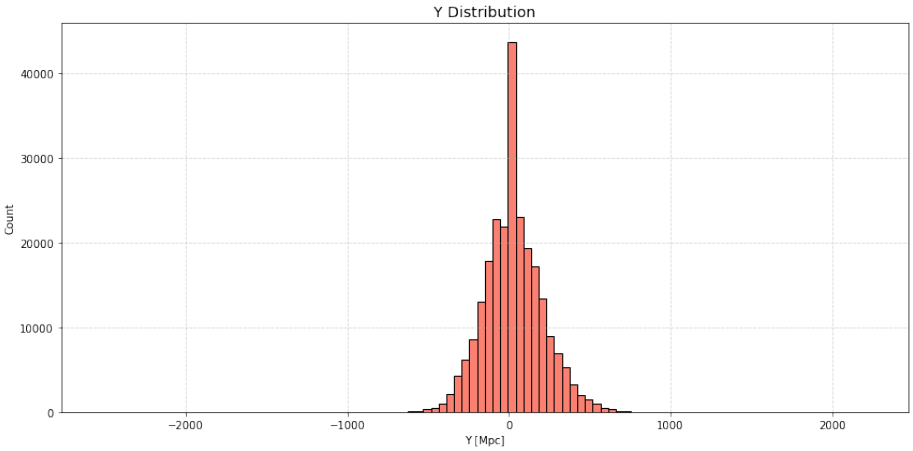

Figure 4: The distribution of galaxies in the 2dFGRS in the y-axis

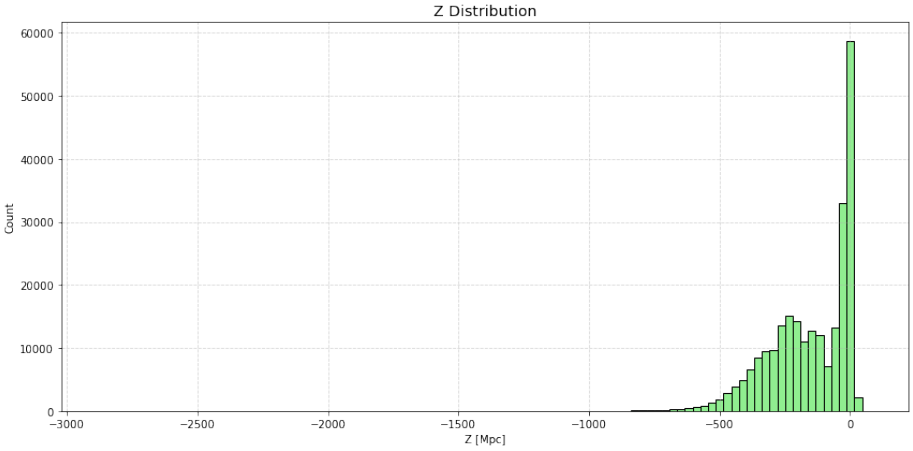

Figure 5: The distribution of galaxies in the 2dFGRS in the z-axis

3.2 Preparation of Data

After retrieving the data from Vizier, the initial step in processing involves transforming the right ascension (RA) and declination (Dec) values from hms (hours, minutes, seconds) and dms (degrees, minutes, seconds) formats into strictly numerical hours and degrees, respectively. This conversion is necessary to ensure consistency in astronomical calculations and is accomplished using the Astropy library, particularly the modules SkyCoord, units, and coordinates.

Since RA values are conventionally expressed in hours, which range from $0h$ to $24h$**,** an adjustment is required for certain cases. Specifically, galaxies with RA greater than 12 hours undergo a correction where 24 hours is subtracted from the original value. This adjustment is crucial in maintaining continuity across the celestial meridian, ensuring that the RA remains properly mapped within the standard coordinate system and avoiding misinterpretations in spatial analyses.

Once these coordinate transformations are completed, the next stage involves correcting galaxy velocities to account for multiple influences affecting their motion relative to the cosmic expansion (Botzler et al., 2002). The raw heliocentric velocities, those measured relative to the Sun do not fully represent the galaxies' true motion within the universe, as local gravitational effects alter their apparent velocity. To obtain cosmologically meaningful velocities, two primary corrections are implemented:

1. **Correction for Solar Motion Relative to the Local Group**: A fundamental adjustment is made to compensate for the Sun's motion within the Local Group. Due to the Sun's peculiar velocity, its movement introduces small distortions when measuring the redshift of galaxies relative to the broader cosmic expansion. Consequently, galaxies with corrected velocities below $30km \cdot s^{-1}$ are assigned an indicative velocity of $300km \cdot s^{-1}$.

This reassignment prevents artificial singularities at $0km \cdot s^{-1}$ a value that would imply no motion, which is physically unrealistic within the framework of large-scale cosmic dynamics. By incorporating this correction, redshift measurements become more accurate, allowing for improved calculations of distances and clustering properties.

2. **Virgocentric Flow Correction:** In addition to the solar motion correction, a second adjustment is made to account for the Virgocentric flow, the systematic motion of the Local Group toward the Virgo Cluster, driven by the cluster's significant gravitational attraction. The Virgo Cluster, being the nearest large gravitational structure, exerts a considerable influence on surrounding galaxies. The correction assumes an infall velocity V_{in} of $300km \cdot s^{-1}$ which modifies the observed velocities accordingly. This adjustment is essential because neglecting Virgo's gravitational pull would result in overestimated distances and inaccurate velocity profiles of galaxies.

After applying these corrections, a velocity filter is imposed:

- Only galaxies with corrected recessional velocities below $12,000km \cdot s^{-1}$ are included in the cluster search.
- Let the Hubble constant be $H_0 = 70km \cdot s^{-1}Mpc^{-1}$ which is consistent with Geller & Huchra (1983).
- The final corrected recessional velocity is determined using the equation:

$$v_{corrected}(z) = v(z) - v_v - v_s$$

where:

- $v(z)$ is the observed velocity of the galaxy,
- v_v is the Virgocentric velocity correction,
- v_s is the solar motion correction.

3.2.1 Data Visualization & Galaxy Cluster Identification

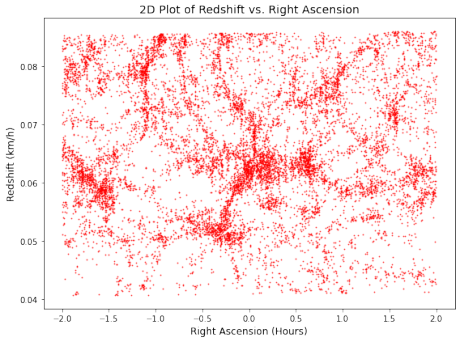

Figure 6: redshift-Ra Projection of The Pisces-Cetus Supercluster spherical coordinates

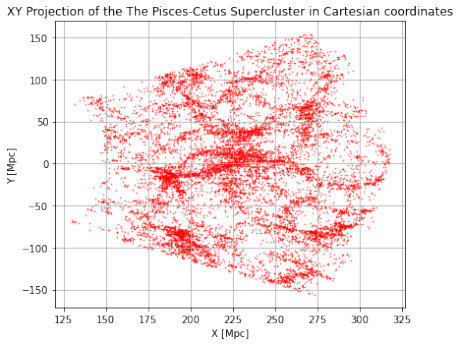

Figure 7: XY Projection of The Pisces-Cetus Supercluster cartesian coordinates

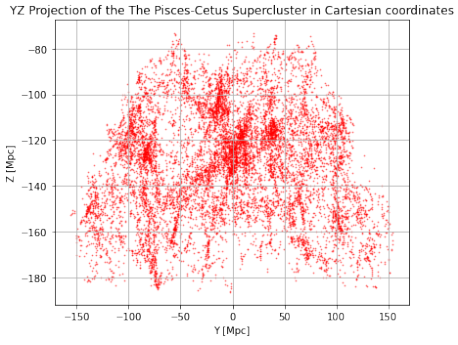

Figure 8: YZ Projection of The Pisces-Cetus Supercluster cartesian coordinates

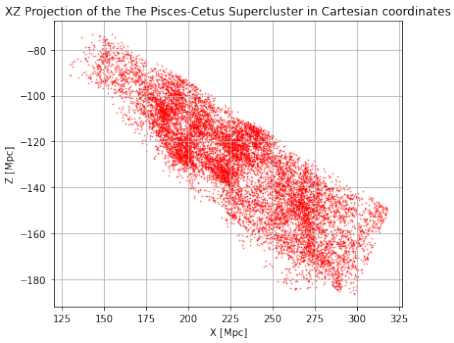

Figure 9: XZ Projection of The Pisces-Cetus Supercluster cartesian coordinates

Once all corrections and transformations are completed, the dataset is plotted to visually inspect galaxy distributions and identify potential clusters look at figure 6, 7 and 8. This step allows for the extraction of boundary conditions necessary for group-finding algorithms that categorize galaxies based on proximity and shared velocity patterns. Through this visualization, the supercluster region is identified with specific constraints:

- RA range: Between 23h and 1h, marking the region where dense galaxy clustering is evident in figure 6.
- Redshift range: Corresponding to radial velocities between $12,000 km \cdot s^{-1}$ and $24,000 km \cdot s^{-1}$, defining the velocity domain of the supercluster in figure 6.
- Cartesian coordinates we get x has a range (175Mpc, 300Mpc) and y has a range (−50Mpc, 100Mpc). figure 7 y is in the range (−50Mpc to 50Mpc) and in the z direction(−140Mpc, −90Mpc) check figure 7 and 8.

3.3 Data Analysis

The Friends-of-Friends (FoF) algorithm is employed to analyse a set of galaxies based on their angular positions on the sky and redshifts, with the objective of identifying gravitationally bound structures within the cosmic web traced by observed galaxies. This algorithm operates through an iterative process, outlined as follows:

1. **Galaxy Selection and Filtering:** The initial step involves refining the galaxy set by ensuring that galaxies are not excessively separated in terms of radial distance from Earth. Each subset is associated with a key redshift value, z_n, which is defined recursively as: $z_n = z_{n-1} + \Delta z$ where Δz is carefully chosen to optimize the volume encompassed by the cross-sectional slice corresponding to (z_n).

2. **Categorization of Galaxies:** From each subset, a galaxy is selected and classified into one of two categories: groups, which is a set that consists of multiple galaxies gravitationally bound together, or isolated groups, which is a set that contains either two bound galaxies or a single, associated galaxy.

3. **Group Formation and Merging:** Subsequently, the algorithm examines the newly identified groups to determine whether additional substructures can be formed. In instances where multiple groups exhibit potential connectivity, they merged into larger structures, thereby refining the hierarchal representation of galaxy clusters

Through this iterative approach, the FoF algorithm facilitates the identification and characterization of cosmic structures, thereby contributing to a more comprehensive understanding of the large-scale distribution of matter in the universe.

The categorization of galaxies works as follows. Recall that we define a group as three or more galaxies that are gravitationally bound, and anything less is considered an isolated group. Two crucial factors are considered when determining whether two galaxies are gravitationally bound: their relative separation distance D_{ij} and their relative velocity V_{ij}. If the distance between two galaxies, D_L, is small enough such that the mutual gravitational pull keeps the galaxies bound, they can form a group; otherwise, their gravitational force is too weak to bind them together. Similarly, if the galaxies are moving too fast relative to each other, they will not be gravitationally bound. Although we might not know the transverse velocity, we do know the radial velocity; if that velocity exceeds $\frac{V_L}{2}$, then the two galaxies are not gravitationally bound.

There are two parameters, D_L and V_L, called linking lengths, that preserve these conditions. When two galaxies satisfy these binding conditions, they form a group. For instance, consider one galaxy as the reference (Galaxy 1). By identifying all companion galaxies (friend galaxies) that meet the criteria with respect to Galaxy 1, a group is constructed. Next, another galaxy from within this group is compared with galaxies that are not yet part of the group. If these additional galaxies satisfy the binding conditions with any galaxy in the reference group, they are then incorporated into the group associated with Galaxy 1. By repeating this process, you will eventually discover a group of galaxies that are gravitationally bound to one another.

This method is essential because it helps us identify regions with a high density of galaxy clusters, which, in turn, assists in identifying the filaments of the cosmic web.

The extended Friends-of-Friends (FOF) algorithm developed by Botzler et al. (2002), as presented in their Figure 10, formed the basis of our implementation.

Figure 10: Flow chart of the original Friend of Friends algorithm (Botzler et al. 2002)

3.3.1 Preprocessing and Initial Conditions

The analysis begins by initializing three dictionaries:

- **groups**: stores identified galaxy clusters,
- **unified_groups**: consolidates large-scale structures,

- **isolated_groups**: contains galaxies that do not meet clustering thresholds.

Redshift values are iterated in discrete steps of $\Delta z = 5 \times 10^{-4}$, beginning from z_{min} and advancing until z_{max}. For each redshift step, the corresponding velocity v_{ini} is computed using a redshift-to-velocity transformation, ensuring consistent reference values for clustering criteria.

3.3.2 Galaxy Selection within Redshift Bins

At each redshift iteration, galaxies are selected based on their velocity proximity to v_{ini}. Specifically, galaxies with velocity deviations within $1000km/s$ of v_{ini} form the initial sample group sample_galaxies_z_ini. The number of galaxies in each subset is tracked to ensure enough data for reliable clustering.

3.3.3 Group Identification and Connectivity Search

For each galaxy in the selected sample linking parameters are computed:

- **Velocity linking length (V_L)**, setting the threshold for velocity-based clustering.

$$V_L = V_0 R$$

- **Spatial linking length (D_L)**, determining the maximum projected separation for group membership.

$$D_L = D_0 R$$

Where ($D_0 = 0.52$) Mpc and ($V_0 = 600$) km/s.

According to Botzler et al. (2002) the scaling factor (R) plays a critical role by compensating for variations in the sampling of the luminosity function. In galaxy surveys, observational limitations, especially at fainter magnitudes, lead to an incomplete sampling of the galaxy population. Without correcting for this bias, regions where fainter galaxies are missed would appear artificially sparse (Schechter, 1975). To address this, the linking lengths are scaled according to Botzler et al. (2002).

$$R = \left(\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{\lim}} \Phi(M) dM}\right)^{-1/3}$$

Here, the numerator represents the cumulative density of galaxies brighter than a given absolute magnitude M_{ij}, and the denominator is the cumulative density up to the limiting absolute magnitude M_{lim} of the survey (Botzler et al., 2002). This ratio adjusts the linking lengths to account for regions where the detection of galaxies is incomplete, thereby mitigating selection bias. It ensures that the group-finding algorithm applies a consistent criterion across varying

depths of observation. The Schechter luminosity function, which describes the distribution of galaxy luminosities, is given by

$$\Phi(M) = \frac{ln(10)}{2.5}\Phi^* 10^{0.4(\alpha+1)(M-M^*)} \exp[-10^{0.4(M-M^*)}]$$

with the parameters $\alpha = -1.30$, $M^* = -19.40$ mag, and $\Phi^* = 0.0143 Mpc^{-3}$. This well-established formulation provides the expected density of galaxies as a function of absolute magnitude. The limiting absolute magnitude $M_{\rm lim}$ is calculated as

$$M_{lim} = m_{lim} - 25 - 5\log(\frac{V_{fid}}{H_0})$$

where $V_{fid} = 1000 km{\cdot}s^{-1}$ is a fiducial velocity, and m_{lim} is the limiting apparent magnitude of the survey. Similarly, the absolute magnitude for a galaxy, M_{ij}, is determined by

$$M_{ij} = m_{\rm lim} - 25 - 5\ln(\frac{cz_{ini}}{H_0})$$

A potential group is then identified when a galaxy's velocity differs from the initial reference velocity v_{ini} by no more than $\frac{V_L}{2}$. This condition ensures that only galaxies with similar velocities, indicating of a true physical association, are grouped together.

3.3.4 Friends-of-Friends Expansion

Once an initial group candidate is selected, the FoF algorithm iteratively expands the group by evaluating additional galaxies based on two criteria:

1. Spatial proximity: The angular separation between the new galaxy and existing group members is calculated using the angular_separation function. The physical projected distance D_{ij} is determined based on redshift scaling.
$$D_{ij} = 2\sin(\frac{\theta_{ij}}{2})D(z_{ini}) < \frac{D_L}{2}$$

2. Velocity compatibility: The galaxy's velocity deviation is checked using the formula
$$V_{ij} = |v_i - v_j| < \frac{D_L}{2}$$
where $\frac{V_L}{2}$ is a predefined threshold , ensuring coherence in the group's motion.

If both conditions are satisfied, the galaxy is added to the group. The search continues iteratively, with each newly added galaxy prompting an additional check for potential members. The expansion process repeats until no further galaxies meet the linking criteria.

3.3.5 Validation and Classification of Groups

Upon completing the group formation, its membership count is evaluated:

- If the number of galaxies in the final group is $\geq N_{min} = 3$, it is recorded in dictionary named groups, using the term z_{ini} to designate a valid cluster. Here, N_{min} represents the minimum number of galaxies required for a group to not be considered isolated.
- If the group contains fewer than N_{min}, it is designated as isolated and stored in a dictionary named isolated_groups, using the term z_{ini}. This classification prevents undersized groups from influencing the clustering analysis while still retaining data on small, dispersed galaxy sets.

3.3.6 Incrementing Redshift and Iterating the Process

After processing the current redshift bin, z_ini is incremented by Δz, advancing to the next redshift slice. The iteration continues until the maximum limit z_max is reached, ensuring a complete scan through the dataset.

By systematically applying velocity and spatial linking criteria, refining groups via the friends-of-friends algorithm, and filtering statistically weak associations, this method ensures robust identification of galaxy structures. The resulting dataset provides a foundation for analyzing the dynamics, clustering behavior, and large-scale structure formation of galaxies. Access my analysis on Github using the hyperlink: https://github.com/LuvuyoBhele/Tracing-Filaments

4 Results & Discussion

A total of 2104 galaxies have been classified into distinct groups. Among these, 91 groups contain more than three galaxies, while 41 have been identified as lone groups (groups with one or two members). Thus, the total number of classified groups identified is 132. Among the groups with more than three members, the most common group size is three, the minimum threshold for classification as a group. The largest group contains 572 galaxies, and only one such group exists. The data is found in Table 2 and the summary of the group size with groups of more than 3 galaxies distribution is presented in Table 3.

Figures 11 and 13 the reveal three distinct structural sections. The most prominent feature is a vertical arrangement of groups, dominated by the largest group and several smaller ones, which together form a dividing structure between two separate group clusters. Also take note that there are also groups that are not considered isolated groups that are not in close proximity to the other groups.

Figures 7, 8 and 9 illustrate the structure of the supercluster in the absence of the identified groups. The darkest parts resemble the supercluster. In contrast, Figures 14, 15 and 16 reveal how some the identified groups trace the structure

of the supercluster. This behaviour aligns with our theoretical expectation that galaxies are embedded within filaments. While some of the grouped galaxies appear not to belong to the main supercluster structure, the overall distribution of groups broadly follows the shape of the supercluster, even though not all regions are uniformly populated.

When viewed from the XZ projection, it's difficult to draw definitive conclusions since Figures 8 and 14 lack sufficient detail; however, certain properties can be extrapolated from the other projections.

group size	frequency	group size	frequency	group size	frequency	group size	frequency
1	29	11	2	24	1	136	1
2	12	12	1	27	1	282	1
3	22	13	2	29	1	572	1
4	11	14	2	34	1		
5	6	15	2	36	1		
6	7	16	1	46	1		
7	6	19	1	49	1		
8	3	20	4	51	1		
9	4	23	5	94	1		

Table 2: Table showing the number of groups of galaxies that were identified by FoF algrithm

Property	Group size
mean	22.1
standard deviation	67.4
min	3
max	572
median	6

Table 3: Statistical summary of groups with 3 or more galaxies

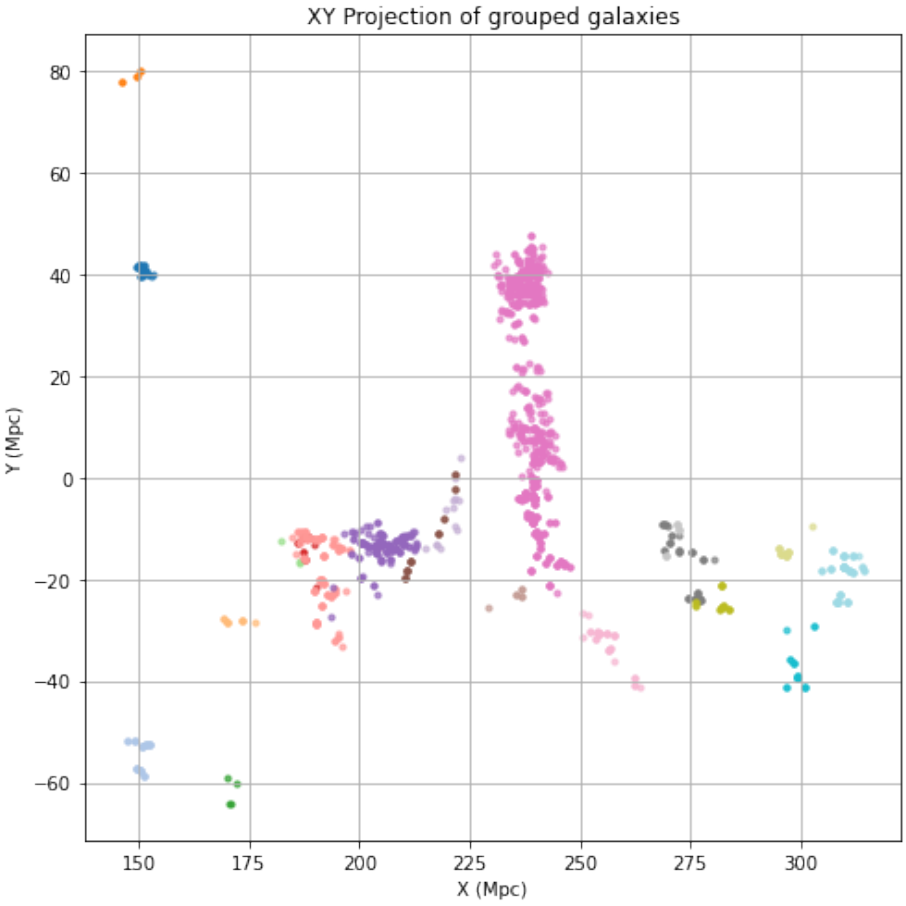

Figure 11: XY Projection of Groups galaxies in The Pisces-Cetus Supercluster in cartesian coordinates

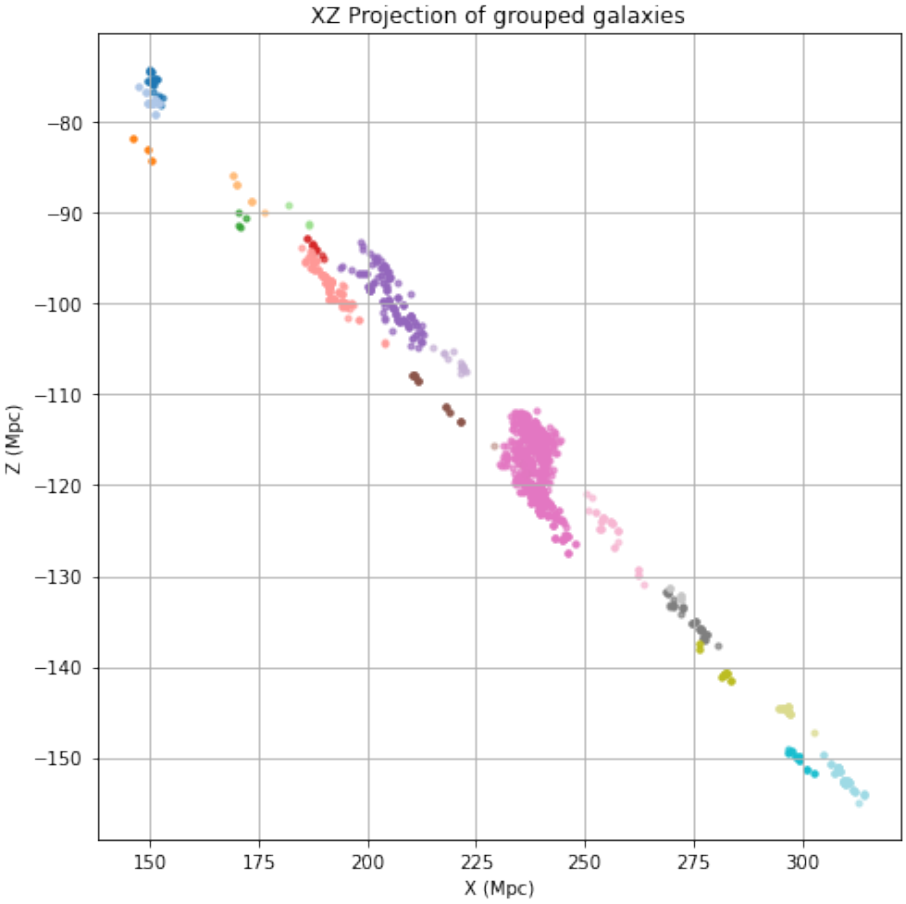

Figure 12: XZ Projection of Groups galaxies in The Pisces-Cetus Supercluster in cartesian coordinates

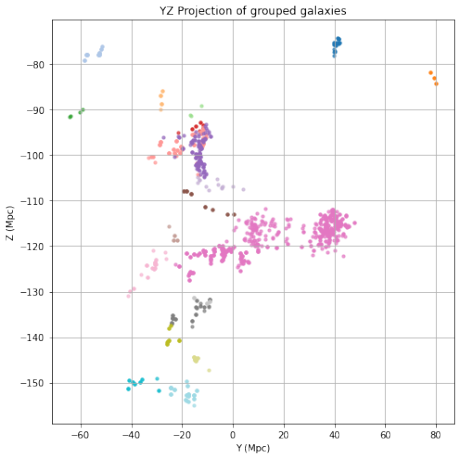

Figure 13: YZ Projection of Groups galaxies in The Pisces-Cetus Supercluster in cartesian coordinates

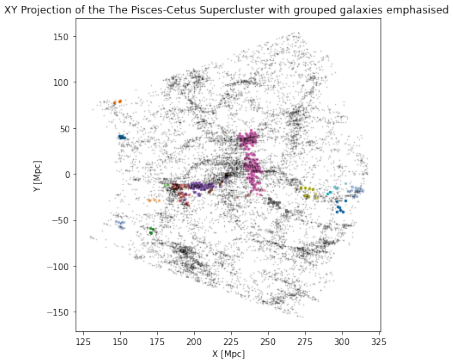

Figure 14: XY Projection of The Pisces-Cetus Supercluster with the groups highlighted in cartesian coordinates

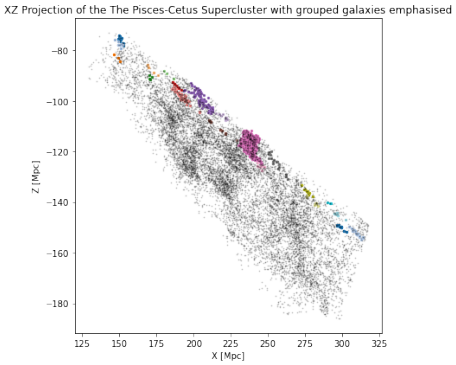

Figure 15: XZ Projectin of The Pisces-Cetus Supercluster with the groups highlighted in cartesian coordinates

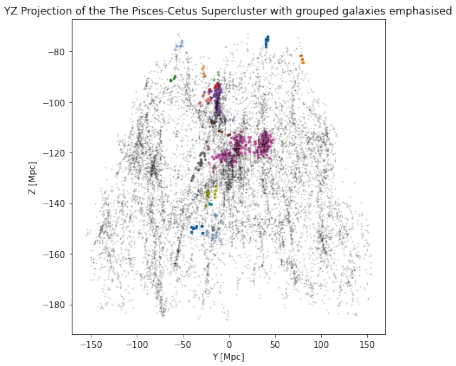

Figure 16: YZ Projection of The Pisces-Cetus Supercluster with the groups highlighted in cartesian coordinates

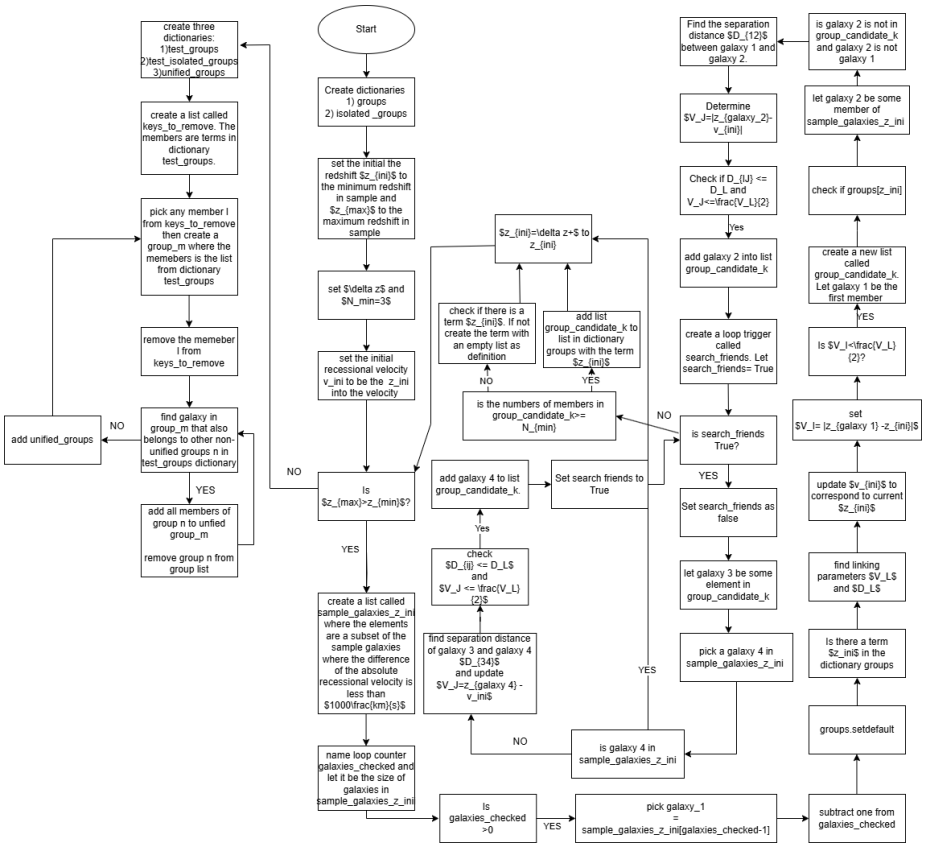

Figure 17: Modified version of Friends of Friends algorithm that was implemented

5 Conclusion

In this study, we used data from the 2dF Galaxy Redshift Survey together with an iterative Friends-of-Friends (FoF) clustering method to map large-scale cosmic structures. We corrected raw right ascension, declination, and redshift for solar motion and the Virgocentric flow, then filtered by velocity. Finally, we used Astropy to convert spherical coordinates into Cartesian coordinates, giving us precise 3D positions for every galaxy.

The FoF algorithm identified 132 galaxy groups: 91 contain more than three galaxies, while 41 are isolated systems. When we plotted those 91 richer groups in Cartesian space, they traced out the filamentary outline of a supercluster. This confirms that a supercluster is really a collection of smaller galaxy groups, in line with hierarchical-formation theory.

Our results also set the stage for tackling the Missing Baryon problem, which notes that a large fraction about $40\% - 50\%$ of normal matter is unaccounted

for. Cen and Ostriker (2006, 1999) predict these "missing" baryons live in the warm-hot intergalactic medium (WHIM), which is hard to observe directly. Next, we'll use FoF to pinpoint galaxy filaments, cross-match active galactic nuclei in those filaments with radio catalogues , and study how the intergalactic medium shapes AGN morphology. From those changes, we can estimate the WHIM's density and temperature. Finally, we'll apply this method to different filaments across the cosmic web to find and study the missing baryons in the WHIM.

6 Bibliography

1. . Davé, R., Cen, R., Ostriker, J.P., Bryan, G.L., Hernquist, L., Katz, N., Weinberg, D.H., Norman, M.L., and O'Shea, B. 2018. Baryons in the warm-hot intergalactic medium. *Astrophysics and Space Journal* , Available at: arXiv:astro-ph/0007217v1 [Accessed 28 March 2025].

2. Cen, R. and Ostriker, J.P., 1999. Where are the baryons? The Astrophysical Journal, 514(1), pp.1-6.

3. Kuchner, U., Aragón-Salamanca, A., Rost, A., Pearce, F.R., Gray, M.E., Cui, W., Knebe, A., Rasia, E., & Yepes, G. (2021). Cosmic filaments in galaxy cluster outskirts: quantifying finding filaments in redshift space. *Monthly Notices of the Royal Astronomical Society*, *503*(2), 2065–2076. https://doi.org/10.1093/mnras/stab567..

4. Porter, S. C. and Raychaudhury, S., 'The Pisces-Cetus Supercluster: A remarkable filament of galaxies in the 2dF Galaxy Redshift Survey and Sloan Digital Sky surveys', *Monthly Notices of the Royal Astronomical Society*, 000 (2005), 1–11.

5. Morris, M. E., Wilcots, E., Hooper, E., and Heinz, S., 'How Does Environment Affect the Morphology of Radio AGN?', *The Astronomical Journal*, 163 (2022), 280, doi:10.3847/1538-3881/ac66db.

6. Serra, AL & Diaferio, A 2013, 'Identification of members in the central and outer regions of galaxy clusters', *The Astrophysical Journal*, vol. 768, no. 2, p. 116, doi:10.1088/0004-637X/768/2/116.

7. Huchra, JP & Geller, MJ 1982, 'Groups of galaxies I. Nearby groups', *The Astrophysical Journal*, vol. 257, pp. 423–437.

8. Botzler, CS, Snigula, J, Bender, R & Hopp, U 2002, 'Finding structures in photometric redshift galaxy surveys: An extended friends-of-friends algorithm', *Monthly Notices of the Royal Astronomical Society*, vol. 000, pp. 1–15.

9. Dannerbauer, H., van Kampen, E., Afonso, J., Andreani, P., Arrigoni Battaia, F., Bertoldi, F., Casey, C. & Chen, C. 2019, 'Mapping Galaxy

Clusters in the Distant Universe', *Astro2020 Science White Paper*, vol. 1903, pp. 06238. Available at: arXiv:1903.06238.

10. Santiago-Bautista, I., Caretta, C.A., Bravo-Alfaro, H., Pointecouteau, E. & Andernach, H. 2020, 'Identification of filamentary structures in the environment of superclusters of galaxies in the Local Universe', *Astronomy & Astrophysics*, vol. 637, A31, doi:10.1051/0004-6361/201936397.

11. Mullis, C.R., Henry, J.P., Gioia, I.M., Bo¨hringer, H., Briel, U.G., Voges, W. and Huchra, J.P., 2001. The North Ecliptic Pole Supercluster. The Astrophysical Journal, **553(1)**, pp.113–120. Available at: https://doi.org/10.1086/320496 [Accessed 21 May 2025].

12. Carrol, B.W. and Ostile, D.A. (2014) An introduction to modern astrophysics. 2nd edn. Harlow: Pearson Education Limited.

13. Colberg, J.M., Sheth, R.K., Diaferio, A., Gao, L., and Yoshida, N. (2005). Voids in ΛCDM universe. Monthly Notices of the Royal Astronomical Society, 360, 216-226. doi:10.1111/j.1365-2966.2005.09064.x

14. Cen, R. and Ostriker, J.P(2006) 'Where are the Baryons? II. Feedback Effects", The Astrophysical Journal, 650(2), p. 560. doi:10.1086/506505.

15. Schechter, P. (1975) The Luminosity Function for Galaxies and the Clustering of Galaxies. PhD dissertation, California Institute of Technology. doi: 10.7907/08Z1-4A82. Available at: https://resolver.caltech.edu/CaltechETD: 08282008-090523 [Accessed: 28 June 2025]

7 NOTES

The creation, refinement, and editing of this report were supported by Microsoft Copilot, which was used solely to enhance clarity and structure. No content was generated by AI; its use was limited to linguistic and stylistic improvement.