DISSERTATIONES
ASTRONOMIAE
UNIVERSITATIS
TARTUENSIS
25

MOORITS MIHKEL MURU

Modeling the cosmic web with the Bisous method

DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS
25

DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS
25

MOORITS MIHKEL MURU

Modeling the cosmic web with the Bisous method

UNIVERSITY OF TARTU
Press
1632

This study was carried out at Tartu Observatory, University of Tartu, Estonia.

The Dissertation was admitted on September 28th, 2023, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Physics, and allowed for defense by the Council of the Institute of Physics, University of Tartu.

Supervisor: Dr. Elmo Tempel,
Tartu Observatory, UT,
Tõravere, Estonia

Opponent: Dr. Weiguang Cui,
Department of Theoretical Physics, UAM,
Madrid, Spain

Defense: November 10th, 2023, Tartu Observatory, UT, Estonia

ISSN 1406-0302 (print)
ISBN 978-9916-27-348-7 (print)
ISSN 2806-2132 (pdf)
ISBN 978-9916-27-349-4 (pdf)

University of Tartu Press
www.tyk.ee

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I **Muru, M. M.**, & Tempel, E. 2021, *Assessing the reliability of the Bisous filament finder*, Astronomy & Astrophysics, 649, A108

II **Muru, M. M.** 2023, *Detecting cosmic filamentary network with stochastic Bisous model*, Proceedings IAU Symposium, 16 (S362), 54-55

III **Muru, M. M.**, & Tempel, E. 2023, *Using photometric redshift data to improve the detection of galactic filaments with the Bisous model*, Astronomy & Astrophysics, 670

Other related publications of the dissertant:

IV Tempel, E., Tuvikene, T., **Muru, M. M.**, Stoica, R. S., Bensby, T., Chiappini, C., Christlieb, N., Cioni, M.-R. L., Comparat, J., Feltzing, S., Hook, I., Koch, A., Kordopatis, G. 2020, *An optimized tiling pattern for multiobject spectroscopic surveys: application to the 4MOST*, Monthly Notices of the Royal Astronomical Society, 497, 4626-4643

Author's contribution to the publications

The Author has made considerable contributions to the following original publications. The following list gives details on the author's work in each of the papers. The Roman numerals correspond to those in the list of publications.

Publication I. The author used a catalog of simulated galaxies to which he applied the Bisous model, analyzed the data, prepared the figures, and wrote the manuscript.

Publication II. The author used the output from the Bisous model from Publication I, extended the analysis, prepared the figures, and wrote the manuscript.

Publication III. Using a catalog of simulated galaxies, the author created mock data sets, applied the Bisous model using the new methodology, analyzed the data, prepared the figures, and wrote the manuscript.

Publication IV. The author prepared some of the figures and helped with the preparation of the manuscript.

INTRODUCTION

The grandiose goal of physics is to understand how the universe behaves. Physicists already know that the large and small scales behave differently and are currently described by different models that do not fit together well. For the smaller end of matter, we have quantum mechanics and the Standard Model of particle physics. The Standard Model of particle physics aims to describe all the particles and forces we know of, but we know it is incomplete. Astronomical observations are showing that there is excess mass that primarily interacts gravitationally and that has not yet been observed directly. Many theories of modified gravity have been proposed, but none can explain the multitude of observational phenomena that have been discovered. It is becoming more likely that dark matter consists of some unknown particles that are not accounted for by the Standard Model of particle physics. Unfortunately, the particle accelerators and other experiments built to study particles do not have enough power to study potential dark matter particles. Nor do any that are currently being planned. The only option to study dark matter and find the dark matter particle is through astronomical observations. If we want to go beyond the current Standard Model, then astronomy is the solution.

The current concordance theory of cosmology is the Λ Cold Dark Matter (ΛCDM) model. This model describes the universe with three major components — dark energy, cold dark matter, and ordinary matter — and assumes general relativity to be the correct theory of gravity. It is the simplest model that can explain all of the major observed properties of the universe: the homogeneity of cosmic microwave background radiation, the emergence of the large-scale structure in the galaxy distribution, the abundance of the lightest chemical elements, and the accelerating expansion of the universe. According to current understanding, the structure of the universe emerges out of a smooth and almost homogeneous distribution of matter that is seeded with small primordial density fluctuations. These fluctuations are characterized by the cosmic microwave background radiation. The initial density fluctuations evolved under the influence of gravity and the expansion of the universe (Peebles 1980). The large-scale structure has evolved into a complex interconnected web-like structure in a bottom-up sequence, also known as the Zel'Dovich pancake model, where matter first collapses into 2D sheets, then elongated filaments, and finally, fully collapsed clusters called nodes (Shandarin & Zeldovich 1989).

The first hints that the universe has a large-scale structure, not just randomly distributed galaxies, came in the second half of the last century (Jõeveer et al. 1977). An IAU symposium about the large-scale structure of the universe was held in Tallinn in 1977, where the Western and Soviet astronomers could meet and

discuss the emerging ideas of a unified structure of the universe. The idea was not initially widely accepted. One of the reasons was that there is too little redshift data for such conclusions (Einasto 2018). Another is that lots of the studies so far used 2D data, but filaments and the structure are more clearly seen in 3D data. This also created a demand for wide-field redshift surveys with high completeness to better map the galaxy distributions and the underlying structure. This was realized by the second Harvard Center for Astrophysics (CfA) redshift survey (de Lapparent et al. 1986). By 1996, the large-scale structure of the universe was much more widely accepted and got its current name, the cosmic web (Bond et al. 1996).

As our understanding of the cosmic web has improved, so have the methods to detect the large-scale structure and its elements. What started out as studying the galaxy number densities in the neighborhood has turned into a playground of much more sophisticated dynamical models, graph theory, and other mathematical tools. For example, this thesis uses a filament-finding method that utilizes marked point processes, Bayesian probability theory, and Markov Chain Monte Carlo methods.

With time, many new aspects and different properties of the cosmic web have been discovered. In general, galaxy clusters are much better understood than other cosmic web substructures, like filaments. The study of those structures is more empirical, for example, inference of properties from the observations and simulations. Detecting and studying the cosmic web network is challenging due to the complexity of the individual structures and the intrinsic multiscale nature (Cautun et al. 2014).

The advent of technology has made it possible to obtain more data with even higher precision. New surveys, such as 4MOST (de Jong et al. 2019), J-PAS (Benitez et al. 2014; Bonoli et al. 2020), DESI (Dey et al. 2019), Euclid (Sprenger et al. 2019), and LSST (Ivezić et al. 2019), are covering larger areas, observing the universe at higher redshifts and have higher completeness. This will improve our understanding of cosmic structures as we are able to study larger volumes with more detail. The simulations are also improving. There are large-scale magnetohydrodynamical simulations, such as IllustrisTNG (Springel et al. 2018) and MillenniumTNG (Pakmor et al. 2022), that capture the structure and evolution of the cosmic web, and also the connection between dark and baryonic matter[1] to study their influence on each other through gravity and various feedbacks. Constrained large box simulations, such as SLOW (Dolag et al. 2023a), are run to understand the large-scale structure in our own neighborhood and to differentiate local structure peculiarity from general trends. Simulations help to test cosmological theories and interpret the observations in the context

[1]Everything that is made of protons and neutrons is collectively named baryonic matter, for example, gas, stars, planets, and galaxies.

of a cosmological model. The dawn of the big data era and more computational power mean it is prime time for new cosmological discoveries. But to realize it, we need to use or develop tools suitable for cosmological big data. One of the goals of cosmologists is to build a rigorous mathematical framework that can be used to describe and study the cosmic web and the universe.

This thesis characterizes the edge cases of a cosmic web filament finder, the Bisous model, which specializes in using observational data and proposes a method to improve the detected filamentary network using data from photometric redshift surveys. The thesis is divided into five chapters. Chapter 1 gives a brief overview of the large-scale structure studies. Chapter 2 describes the Bisous filament finder in some detail. Then Chapter 3 presents the results from the published paper about the robustness of the Bisous model, and Chapter 4 describes the results from the paper about using photometric redshift data with the Bisous model. Chapter 5 summarizes the results, gives a short overview of planned improvements for the Bisous model, and some comments about the usefulness of marked point processes.

1 OVERVIEW OF THE COSMIC WEB AND ITS DETECTION

The cosmological large-scale structure is intimately connected to the contents and evolution of the universe, and therefore, the cosmic web incorporates invaluable information. Many decades have been devoted to the study of these cosmological structures, finding new physical relations, and constructing and improving theories based on new observations. Still, much exciting and challenging work remains to be done before we can feel confident that we fully understand the universe on these largest scales. This chapter gives a brief overview of why understanding the large-scale structure is important and how it is studied.

1.1 Cosmic web's importance in cosmology

Gravitational forces and the expansion of the universe mold the cosmic web. Therefore, studying the cosmic web reveals a lot about the matter that interacts gravitationally and the energy components of the universe that affect space around the matter. Thanks to the observable universe being so large and the speed of light being finite, we are also able to study the evolution of the cosmic web and, therefore the evolution of matter distribution and the effects of the energy components. Two of the largest questions in cosmology, or even in the whole physics research, are the nature and properties of dark matter and dark energy. The cosmic web is perfect for studying both of these questions.

The large-scale structure of the universe is often classified into four substructures. Figure 1.1 shows the intricate and interconnected cosmic web. Galaxies are clustered in overdense regions called the nodes of the web that are connected by elongated structures called filaments. Between the nodes and filaments are large underdense volumes called voids. These voids are surrounded by membranes called sheets. The web structure is dynamic and evolves with time. Filaments are the most dominant substructure of the cosmic web when looking at the force field inventory as they contain the most mass, and clusters, although the densest regions, only affect the close-by volumes (Cautun et al. 2014). Large, almost empty voids are a major part of the tidal field inventory because they take up the majority of the universe's volume. The general tendency is that matter first collapses into filaments, and then it accumulates in the nodes, making the clusters bigger and denser. In that process, the voids grow larger and less dense. Cautun et al. (2014) studies the migration of galaxies between cosmic web substructures. Analysis shows that filaments obtain their galaxies from voids and walls and mostly retain their mass, but the majority of the mass they lose is fed to the nodes.

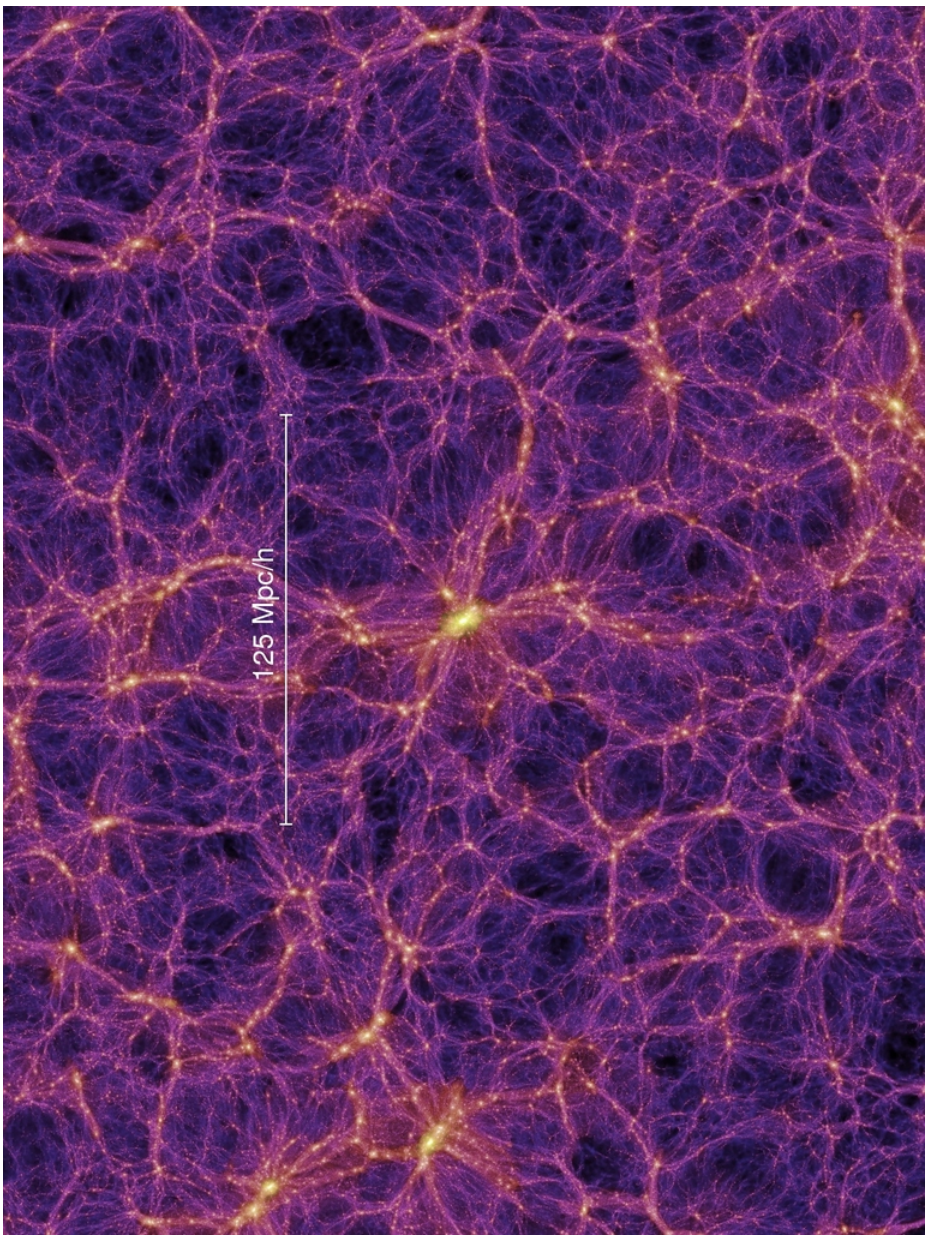

Figure 1.1: Dark matter cosmic web from the Millenium simulation. Lighter colors indicate higher densities. The filaments connect the clusters in the overdense nodes creating the web structure, and the sheets envelop the large underdense voids. Image credit: VIRGO consortium.

The dark energy component of the ΛCDM model accounts for the expansion of the universe. Since there are no direct ways to study dark energy, studying the evolution of the large-scale structure is one of the only ways we can infer new information about dark energy and the future of our universe. Voids, large volumes with very low mass densities, are one of the cleanest probes of cosmological parameters. Voids' structures, morphologies, dynamics, and counts reflect the nature of dark energy (Park & Lee 2007; Lavaux & Wandelt 2012; Pisani et al. 2015; Radinović et al. 2023).

Approximately 85% of matter in the universe is accounted for by dark matter (Planck Collaboration et al. 2020). That means that the cosmic web is mostly a dark matter structure. Simulations confirm that baryonic matter has almost no impact on the structure of the cosmic web (Cui et al. 2018, 2019). While studying the cosmic web, we are studying dark matter. This helps to differentiate between different dark matter models. For example, fuzzy, warm, and self-interacting dark matter each has been shown to form web-like structures in simulations that differ from ΛCDM cosmic web (Knebe et al. 2003; Rocha et al. 2013; Kuo et al. 2018; Dome et al. 2023). As no dark matter particle has been discovered, there are lots of modified gravity models giving different explanations for different observed phenomena. The cosmic web and especially the voids are one of the observable probes that can be used to discern between different gravity models (Zivick et al. 2015; Ho et al. 2018).

In addition to studying the things that influence the cosmic web, cosmologists are also interested in the cosmic web's influence on smaller scales and baryonic matter. Figure 1.2 shows three snapshots from a simulation that show the distributions of dark matter, gas, and stars in the same area and how well they imitate each other. The large-scale structure defines the environment where galaxies, groups, and clusters evolve and therefore influences their properties. For example, the global environment defined by the cosmic web affects the morphology and quenching of galaxies in different environments (Kuutma et al. 2017; Kraljic et al. 2018; Malavasi et al. 2022) and the concentration of diffuse gas along the cosmic web filaments (Tuominen et al. 2021; Galárraga-Espinosa et al. 2021). Large-scale structure affects the intrinsic alignment of galaxies (Tempel & Libeskind 2013; Ganeshaiah Veena et al. 2019), which is an important property for accurate weak lensing studies. In the Epoch of Reionization, the morphology of the ionization bubble network is intimately connected with the filaments and nodes of the cosmic web (Thélie et al. 2022).

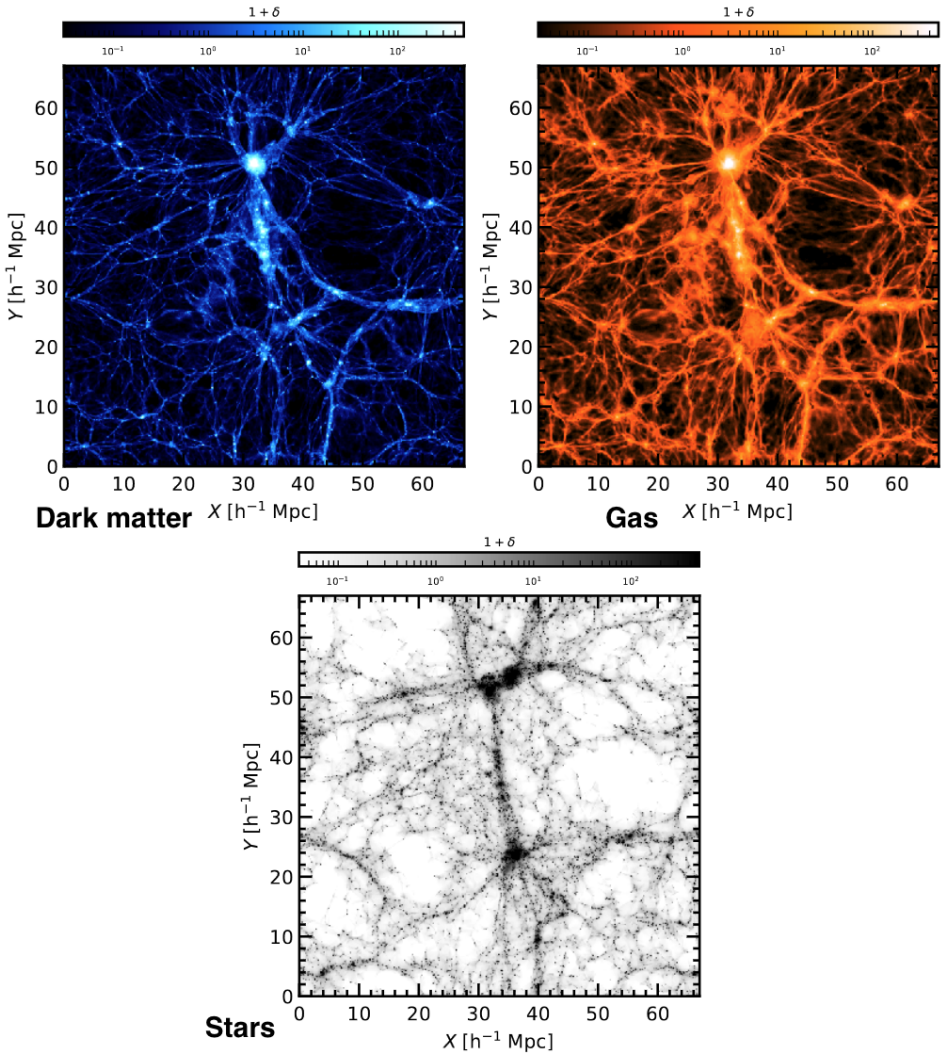

Figure 1.2: Dark matter, gas, and stellar density fields. The slices are taken from the EAGLE hydrodynamical simulation (Schaye et al. 2015). The high correlation between the density fields shows how the dark matter, gas, and stellar components' evolution is connected to each other. (Ganeshaiah Veena et al. 2019)

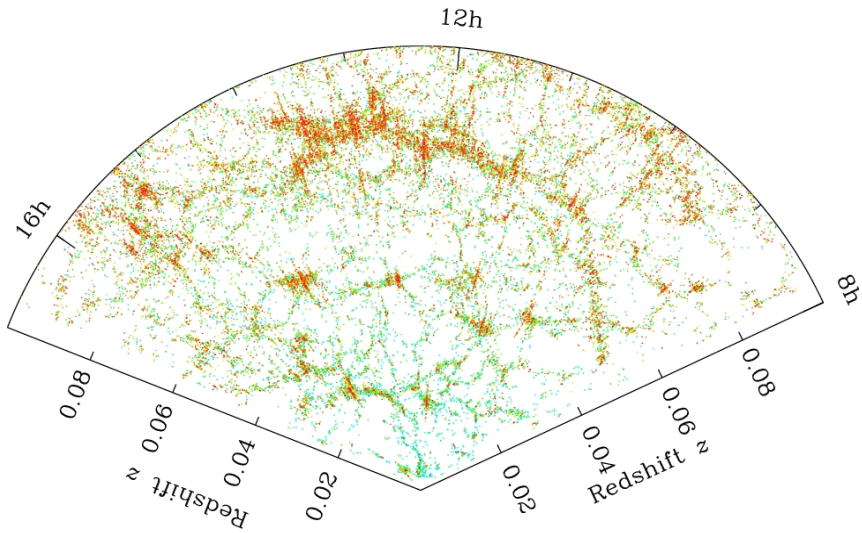

Figure 1.3: Galaxy distribution in the SDSS survey. Each dot represents a galaxy and the color its g-r color. The galaxy distribution follows the dark matter cosmic web and shows the filamentary network. Image Credit: M. Blanton and SDSS.

1.2 Methods to study the large-scale structure

There are two approaches to obtaining data for large-scale structure studies — observations and simulations. First, the obvious choice is to use the real data that we get from observations. Often, large-scale structure studies are statistical in nature and require large data sets. This creates the need for large surveys to map as much of our surroundings as possible, usually maps of galaxies or gas distribution by Ly-alpha or radio measurements.

One of the most widely used surveys in cosmology for galaxy distribution is the Sloan Digital Sky Survey (SDSS, York et al. 2000; Eisenstein et al. 2011). Figure 1.3 shows the SDSS galaxy catalog in redshift space. SDSS provides good completeness in the low redshift universe, which is very useful for group and filament detection, as well as for studying galaxy properties connected to the large-scale structure using photometric imaging. The Dark Energy Spectroscopic Instrument (DESI, Dey et al. 2019) is ongoing and complementary survey for the SDSS in the southern sky providing higher quality imaging and spectra for ten times more galaxies, around 35 million, and covering larger sky areas. For the evolution of galactic filaments, higher redshift surveys are used, such as the VIMOS (Visible Multi-Object Spectrograph) Public Extragalactic Redshift Survey (VIPERS, Guzzo et al. 2014), which probes a smaller patch in the sky but to a much greater depth, up to redshift $z = 1.2$. Many upcoming surveys are very exciting for large-scale structure studies. The 4-metre Multi-Object

Spectroscopic Telescope (4MOST, de Jong et al. 2019) is a spectroscopic survey mapping the whole southern sky with unprecedented galaxy number density and also includes some deeper subsurveys. Javalambre Physics of the Accelerating universe Astrophysical Survey (J-PAS, Benitez et al. 2014) is an upcoming photometric imaging survey that uses over 50 narrow-band filters and measures redshifts with uncertainties comparable to low-resolution spectroscopic surveys. This will produce very high galaxy number density data with high-quality imaging that will be invaluable to studying the connection between galaxy properties and large-scale structure. European Space Agency's Euclid mission maps a third of the whole sky with a space-based telescope and will produce large amounts of high-quality data for large-scale structure studies (Euclid Collaboration et al. 2022).

Although observations will always define our truth condition, i.e., any model or theory has to conform to observations, there are many downsides to using observations. As the cosmic web is mostly dark matter structure and we cannot observe dark matter, we can only trace the web using galaxies and gas. There are many observational effects and biases to account for, e.g., Malmquist bias, fingers-of-god, Kaiser effect (Kaiser 1987), galaxy bias (Kaiser 1984), selection biases, and correcting for these may be extremely difficult.

On the other hand, simulations have complete information, which means in addition to galaxy positions, there is information about dark matter, densities, velocities, etc. It is also easier to study the redshift evolution in simulations as one can study the same exact object at different stages of its evolution. But there are also large downsides to using simulations. The results are only as good as the simulation and its implementation of physical laws. Depending on the resolutions and the scale of interest, numerical effects may play a significant role. To study the large-scale structure, the simulation box has to be large enough for the web to form, which also increases the computation time and limits the resolution that can be used. One of the most popular simulations for cosmological studies is the magnetohydrodynamical large box simulation suite IllustrisTNG (Nelson et al. 2019). In addition to random initial condition cosmological simulations, there are constrained simulations that aim to recreate the correct local environment from observations. CLUES project has created many constrained simulations, e.g., HESTIA (Libeskind et al. 2020) for Local Group and SLOW (Dolag et al. 2023b) for cosmological box centered around Local Group.

Most approaches to large-scale structure research are geometrical and topological in nature. One of the oldest methods to study clustering is the power spectrum or two-point correlation function that was popularized in the second half of last century with the advent of fast digital computers by, for example, Peebles (1973). The two-point correlation function and higher-order versions are still commonly used to describe matter distribution and clustering. As galaxy surveys became larger and simulations more complex, a challenge emerged to investigate the different substructures of the cosmic web in addition to clustering. This led to

methods that are able to detect not just the clusters and voids but also the elongated filaments and plane-like sheets. The improvements in computational power and methods have influenced the approaches used in cosmology. For example, there are methods to reconstruct the data that we are not able to observe, like filling the zone of avoidance (Ganeshaiah Veena et al. 2023), and using simulations to study universes with different physics, for example, different gravity models. Simulations also enable the user to control the environment very precisely to separate the causes and effects of different environmental elements and objects.

1.3 Cosmic filaments

The most prominent feature of the cosmic web is the filamentary network connecting the nodes. Approximately 50% of matter in the universe is in the filaments, at high redshifts, most of the gas, and at low redshifts, about half of the warm gas is in filaments (Cautun et al. 2014). In simulations, we can see the dark matter's large-scale structure and the intricate web, but in observations, the cosmic web is traced only by galaxies and gas. Filaments are dynamically more complex than clusters as filaments are not yet collapsed structures and therefore, the virial theorem is not applicable to them.

This also makes it harder to define what a filament is, which is still an open question. Due to filaments' dynamic complexity and multiscale nature, it is very difficult to define a boundary for a filament. For different applications and studies, filament's definition might take many different forms. For example, for defining the width of a filament, one could use a fixed radius, a radius based on the galaxy or gas distribution, a radius based on density, velocities, tidal fields, or phase-space, or no radius at all, detecting only the spine of a filament. On the other hand, there is a question, how to measure the filament's length? Do filaments continue through the nodes? At what point does a filament merge into a cluster? In addition, how do we detect the spine or define the filament's orientation?

Even though filaments are still loosely defined, the connection and influence to the surroundings are significant. For example, Tempel & Libeskind (2013) shows that the spin of galaxies is aligned with the spine of the filaments but differently for elliptical and spiral galaxies. The halo spin alignment depends on the mass and evolves with time (Ganeshaiah Veena et al. 2018, 2019, 2021). Galaxy pairs are aligned with filaments (Tempel & Tamm 2015), and also, the satellite galaxies around the host are influenced by the host filament (Wang et al. 2020). This all ties the filaments strongly to the galaxy's evolution. Wang et al. (2021) show that filaments themselves are spinning as well about their axis, which are the largest known objects to rotate and have angular momentum. The number of filaments connected to a cluster is related to many of the cluster's properties, such as its mass, ellipticity, and star formation rate (Gouin et al. 2021; Galárraga-

Espinosa et al. 2023). Filaments are also important to solve the baryonic matter budget issue, the missing baryon problem. Tuominen et al. (2021) shows that missing baryons are in the form of warm-hot intergalactic medium/gas (WHIM) that populates the filamentary structure.

1.4 Detection of the large-scale structure

The absence of a definition for a filament has brought about an abundance of filament and cosmic web finders. Some methods require a more complete set of information and are more suitable for simulations, and some are developed to be applicable to observational data, such as galaxy distribution. It is important to keep in mind that the galaxies are biased tracers of the density distribution (Kaiser 1984). Galaxies overpopulate overdense regions compared to the less dense areas. This has to be taken into account when using galaxies as tracers for dark matter filaments.

Libeskind et al. (2018) gives an overview and comparison of different cosmic web and filaments finders. One large group of methods use the geometric information in the Hessian of the density, tidal, and velocity shear fields, such methods as T-web (Forero-Romero et al. 2009) and NEXUS+ (Cautun et al. 2013). These methods are based on the fact that the tidal force field molds the formation and dynamical evolution of filaments. The NEXUS+ algorithm is developed keeping in mind the multiscale nature of the cosmic web and is able to detect filaments on different scales (see Fig. 1.4). Some methods use the information of the six-dimensional phase-space and differential geometry to detect the cosmic web using the Caustic Skeleton model (Feldbrugge & van de Weygaert 2023). The advent of machine learning has also influenced filament detection. New methods using more complex and specialized algorithms to detect filaments are being developed, such as the manifold detecting toolbox 1-DREAM (Canducci et al. 2022).

To be able to detect filaments from the observation, another approach is needed as the data is low-dimensional, noisy, and incomplete. We are already able to construct the local velocity field from the observations (Courtois et al. 2023) and it will improve with new surveys producing higher accuracy and density data for larger volumes. Still, the catalog of distances is much more complete and accurate. Therefore, methods that are able to detect filaments using just the positions of galaxies have an advantage over the ones that need velocity or tidal fields. Even using only the galaxy position, there are many ways to detect the filamentary network. For example, DisPerSE uses the topology of the galaxy distribution to detect ridges and is able to map all the cosmic web substructures in noisy data (Sousbie 2011; Sousbie et al. 2011). Some methods take a straightforward approach and use the galaxy positions to build a graph and

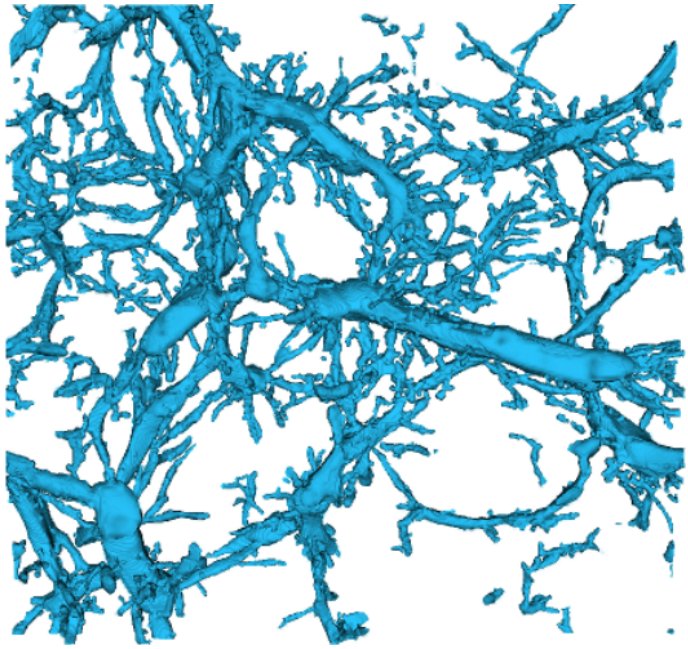

Figure 1.4: Filamentary network as detected by the NEXUS+ algorithm. The resulting map shows the multiscale nature of the cosmic web. (Cautun et al. 2014)

apply graph theory algorithms. Two such examples is a methods by Alpaslan et al. (2014) and Bonnaire et al. (2020) that apply the minimal spanning tree methodology to detect filaments from observations.

The Bisous model (see Chapter 2) utilizes Bayesian methodology to model the filamentary network using a marked point process and Markov chain Monte Carlo algorithms. The benefits of the method are that it uses directly the galaxy distribution without the need to construct a field or smooth it. Modeling the filamentary network provides a probability of detecting a filament and, thus, a way to estimate the uncertainty of the model. The Bisous model is very efficient at obtaining the orientation of the filament without overfitting and is well-suited for alignment studies. The SDSS galaxy catalog is one of the most commonly used in cosmological studies, and publicly available filament catalogs have been published by both Biosus (Tempel et al. 2014b) and DisPerSE (Malavasi et al. 2020).

The field of filament finders is still rapidly expanding with new ideas and approaches. With new wide-area and deep-field surveys and multimessenger astronomy, we will have more complete and higher dimensional data. This will let us produce a more complete catalog of filaments, that will lead to a better understanding of the filaments themselves, but also the connection between filaments, galaxies, clusters, and intergalactic medium.

2 BISOUS FILAMENT FINDER

The goal of the Bisous filament finder is to model the spines of the filaments in the cosmic web from the distribution of galaxies. This makes the Bisous model especially well-suited for alignment studies and modeling the cosmic web using observational data. The main idea of the Bisous model is to use the marked point process with interactions to model the network of chains of galaxies. The aim is to build a mathematical model that can be used to describe the cosmic web's filamentary network. The core mathematical model of the filament finder is explained and mathematically proven (Stoica et al. 2005b,a, 2007a,b; Stoica et al. 2010, 2014). The Bisous model is described with examples by Tempel et al. (2014b, 2016). An overview is given here to help understand the results of this thesis.

2.1 Bisous model

In observations, filaments look like chains of galaxies. The Bisous model uses a marked point process, where a mark denotes a cylinder's properties. More specifically, the mark contains the length, width, and orientation of a cylinder. These cylinders are the building blocks of the filaments. When the marks and points represent shapes or objects, the point process is called the object point process. Each configuration of objects has a defined probability density value

$$p(\mathbf{y} \mid \theta) = \alpha \exp[-U(\mathbf{y} \mid \theta)] \,, \tag{2.1}$$

where $\mathbf{y}$ is the configuration of cylinders, θ is the vector of model parameters, α the normalizing constant and $U(\mathbf{y} \mid \theta)$ the energy function of the system. The main assumptions of the model is that galaxies are grouped together along the filaments inside small cylinder segments that connect and align with other cylinders. These ideas are represented in the system energy in the following way.

$$U(\mathbf{y} \mid \theta) = U_{\mathrm{i}}(\mathbf{y} \mid \theta) + U_{\mathrm{d}}(\mathbf{y} \mid \theta) \,, \tag{2.2}$$

where $U_i(\mathbf{y} \mid \theta)$ represents the interaction energy describing the interconnectedness of the cylinder network and $U_{\mathrm{d}}(\mathbf{y} \mid \theta)$ is the data energy describing the positions of the cylinders in the context of the underlying galaxy distribution.

The interaction energy depends on the relative positions of the cylinders forming the network. To calculate the energy value, one has to count the cylinders that have 0, 1, and 2 connections to other cylinders. Cylinders are connected when they attract each other, do not form a repulsive pair, and are well aligned. Two cylinders attract each other when the distance between the endpoints is

smaller than the chosen attraction radius. Cylinders form a repulsive pair if their centers are closer than the minimum allowed distance and the cylinders are not orthogonal. Repulsive pairs are forbidden and correspond to $p(\mathbf{y} \mid \theta) = 0$. Distances and angles for attraction, alignment, and repulsion are predefined as model parameters. The interaction energy promotes the interconnectedness of the system by assigning lower energy values if there are more 1- and 2-connected cylinders if the number of cylinders is constant.

The data energy depends on the cylinder positions relative to the underlying galaxy distribution. It is motivated by the physical properties of filaments, mainly that filaments are at galaxy overdensities and along the galaxy chains. There are several properties that influence the data energy value. First, the galaxy distribution inside the cylinder should be close to uniform along the main axis. Second, the number density of galaxies inside the cylinder should be higher than in the vicinity around the cylinder. Third, galaxies inside the cylinder should be concentrated along the main axis. The data energy is the sum of the energy of each cylinder. The implementation of data energy makes sure the network of cylinders follows the galaxy distribution.

These energies define the probability density in Eq. 2.1. The probability density is used to optimize the model with simulations that are explained in the next section. The lower energy of the system means a higher probability density value, and the higher the probability, the better the model describes the cosmic web based on the input galaxy distribution. Mathematical definitions and detailed explanations of the probabilities and energies are given in Tempel et al. (2014b, 2016).

2.2 Simulation of the model

To model the cosmic web using Bisous, we need to sample from the joint probability density $p(\mathbf{y}, \theta) = p(\mathbf{y} \mid \theta)p(\theta)$. The sampling is done using an iterative Markov chain Monte Carlo (MCMC) algorithm. First, choosing values for the parameters in θ, from $p(\theta)$, and second, using Metropolis-Hastings (MH) algorithm to sample a marked point pattern from $p(\mathbf{y} \mid \theta)$. The MH algorithm uses three types of moves: birth (adding a cylinder), death (removing a cylinder), and change (modifying a cylinder's parameters). After a move is proposed, an acceptance probability is constructed based on the type of the move, the system energies of the previous configuration and the new configuration. For birth, the acceptance probability is

$$\min\left\{1, \frac{p_\mathrm{d}}{p_\mathrm{b}} \frac{d(\mathbf{y} \cup \{\zeta\}, \zeta)}{b(\mathbf{y}, \zeta)} \frac{p(\mathbf{y} \cup \zeta)}{p(\mathbf{y})}\right\}, \tag{2.3}$$

where $\mathbf{y}$ is the object configuration and ζ is the object being suggested, p_d and p_b are parameters of the model defining the general death and birth move probability,

function $d(\mathbf{y} \cup \{\zeta\}, \zeta)$ is the probability of removing object ζ from $\mathbf{y} \cup \zeta$, and function $b(\mathbf{y}, \zeta)$ the probability of adding the object ζ to the current configuration, models probability density $p(\mathbf{y} \mid \theta)$ is defined in Equation 2.1 and θ is omitted for brevity. For the death move, a similar but reciprocal acceptance probability is used. For change move, the model uses acceptance probability defined as

$$\min\left\{1, \frac{p(\mathbf{y} \setminus \{\zeta_{\text{old}}\} \cup \{\zeta_{\text{new}}\})}{p(\mathbf{y})}\right\}, \tag{2.4}$$

where $\mathbf{y} \setminus \{\zeta_{\text{old}}\} \cup \{\zeta_{\text{new}}\}$ is the configuration without the old object and with the new object. When a move is accepted, the simulation proceeds with the new object configuration, otherwise, the move is rejected, the simulation continues with the previous configuration, and a new move will be proposed. Details are given in Tempel et al. (2014b, 2016).

To maximize the $p(\mathbf{y}, \theta)$, the described MCMC sampling mechanism is integrated into a simulated annealing algorithm. The simulated annealing is a global optimization method that accelerates the convergence of the sampling. The main idea of the simulated annealing algorithm is iteratively sample from $p_n(\mathbf{y}, \theta) \propto [p(\mathbf{y} \mid \theta)p(\theta)]^{1/T_n}$, while the temperature T_n decreases toward zero. Higher temperatures increase the probabilities closer to 1, and low temperatures make any deviation from 1 more significant. This helps to escape local maxima in the beginning and also quickens the convergence process as the temperature lowers.

As the sampling uses randomness, it is useful to average over the results of many simulations to suppress as much noise from the MCMC as possible since many local minima are equally good solutions. A snapshot of cylinder configuration is called a realization. Usually, a single simulation produces 10 to 50 realizations after the model has mostly converged. The number of moves between consecutive realizations is chosen sufficiently high for the realizations to be almost uncorrelated. We also run many parallel simulations on the same data, usually 50 to 100 simulations. Figure 2.1 shows small sample boxes with a comparison between a single realization and a superposition of 1000 realizations. From these realizations, two different maps are constructed that are used for further inference. Visit map $\mathcal{L}(\mathbf{k})$ for a given point $\mathbf{k} = (x, y, z)$ estimates the probability that point $\mathbf{k}$ is touched by a cylinder in a randomly chosen realization. The visit map can also be used as an estimator for the filamentary pattern of the cosmic web. Orientation map $\mathcal{G}(\mathbf{k}, \boldsymbol{\omega})$ measures how well-aligned the cylinders are in different realizations at $\mathbf{k}$ in direction $\boldsymbol{\omega}$. The maximum value over all directions at each point of the orientation map is stored as the directional strength map. Often, the visit map value is enough to trace the filaments. Clusters also have high visit map values, but usually, the cylinders are weakly aligned, and this helps to separate clusters from filaments.

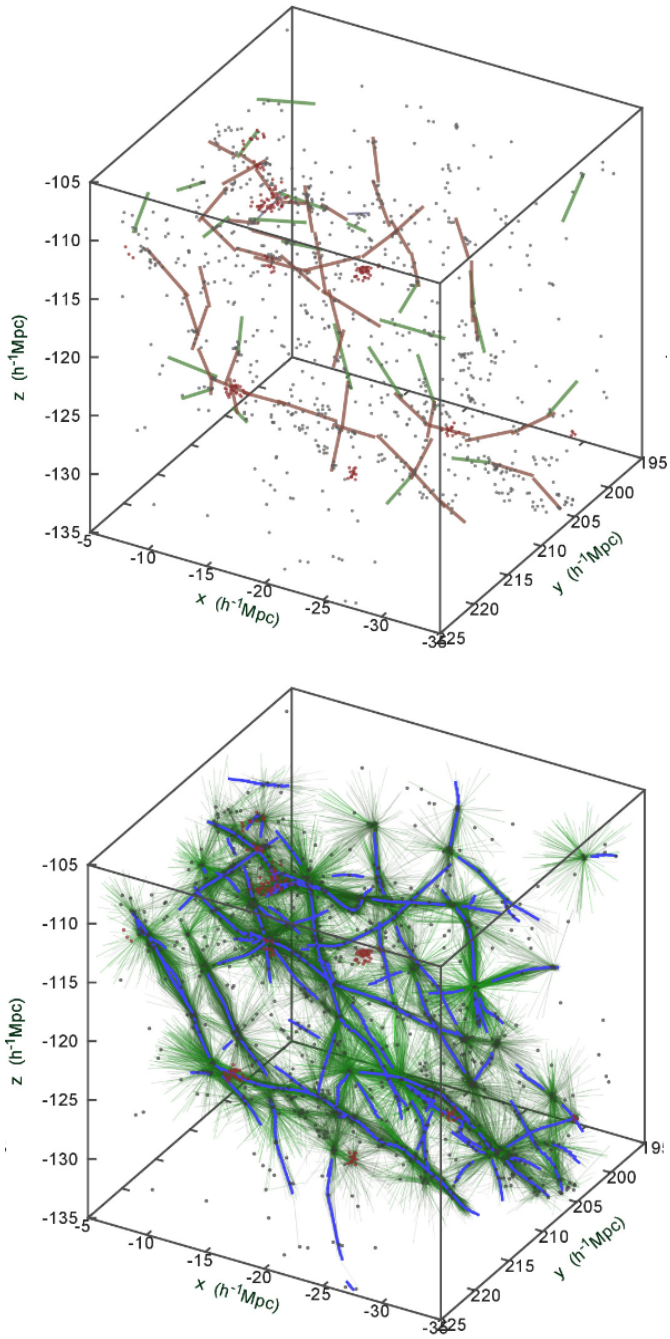

Figure 2.1: Bisous realizations in a small sample box. The points represent the galaxy distribution. The upper box shows the cylinder axes from a single realization. The lower box shows a superposition of 1000 realizations and the detected filament spines with thick blue lines. (Tempel et al. 2014b)

The visit and orientation maps are usually compiled from ~ 1000 realizations and a visit map threshold of 0.05 is used to suppress Poisson noise and maximize the usage of available information. Figure 2.2 demonstrates the aggregation of different sets of realizations. This illustrates how single realizations only contain some features of the filamentary pattern, but stacking many realizations smooths out differences. With a superposition of 200 realizations, there are already almost no differences between different sets of realizations. Muru & Tempel (2021) presents a study on the robustness of the Bisous model, which is also presented in Chapter 3. Figure 2.3 shows how different visit map and directional strength map thresholds affect the fraction of galaxies that reside in a filament. Visit map distribution has maximums both around 0 and 1. Directional strength distribution has a maximum close to 1. This means that most galaxies are clearly inside filaments or not, the in-between ones are much less common, and most filaments are well-aligned in different realizations. In Chapter 3 and 4, the analysis uses visit map threshold 0.05 and directional strength map threshold 0, unless explicitly stated otherwise.

2.3 Implementation of the Bisous model

The choice of the parameters for the model is a complex problem. Some of them are physically informed choices, such as cylinder radius, which is based on gas densities around filament spines, which drops sharply at around $1\,\mathrm{Mpc}$ (Galárraga-Espinosa et al. 2021; Tuominen et al. 2023). Some are chosen with trial and error, such as free parameters for the data and interaction energies, such that the resulting model visually resembles the cosmic web, and some minimize the computational time, such as the MCMC parameters. The time needed for the computations is usually long, from a few days to a month, depending on the complexity of the input data. But this is usually not a fatal factor for studies, as it is mostly necessary to run the Bisous only once. The model works best with moderate galaxy number density, such as in SDSS. If the density is significantly smaller, it is more difficult to find suitable places for objects as the underlying filamentary network is harder to detect. Fewer moves are accepted, and the simulation might take a significantly longer time.

After obtaining the visit map $\mathcal{L}$ and orientation map $\mathcal{G}$, we can extract the filaments from these maps. First, starting from the point where the visit map has the highest value, we follow the ridge defined by the high visit map value and the orientation defined by the orientation map until the visit map value or the orientation strength drops below the thresholds. After that, the spine coordinates are extracted while checking that there are no sharp turns or breaks in the spine. When the extraction is done, the area around the spine is masked out, and the process is restarted by finding the highest visit map value that is not yet masked

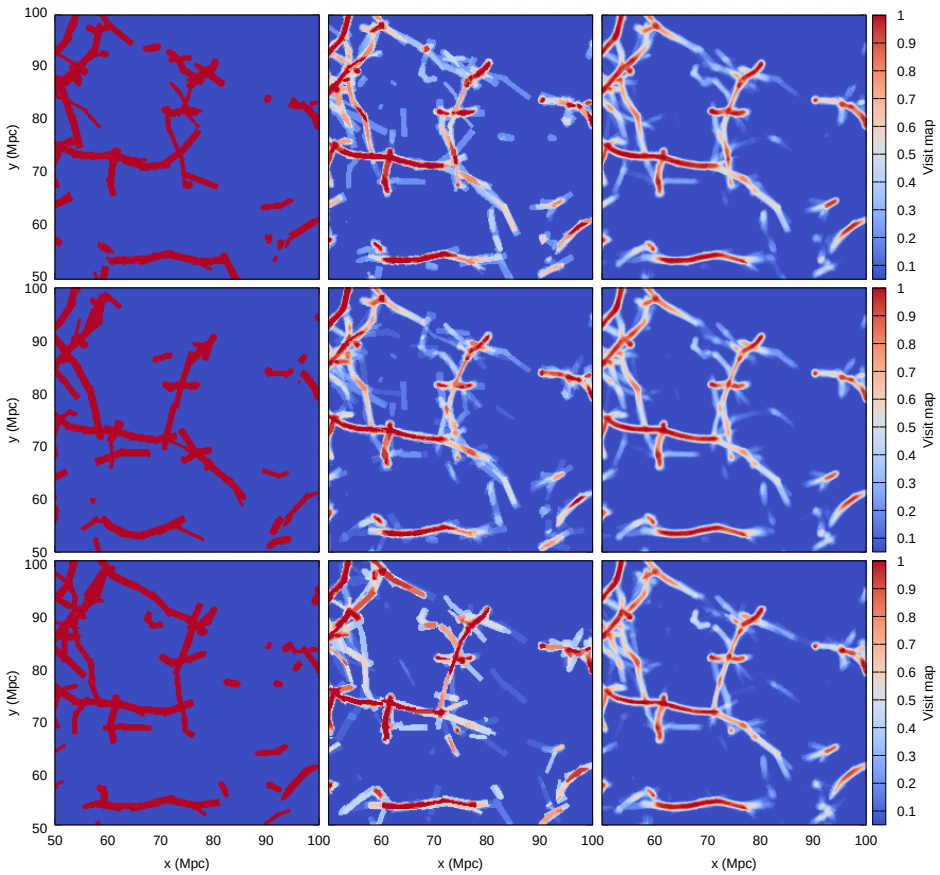

Figure 2.2: Maximum visit map values in a slice. The slice thickness is $10\,\mathrm{Mpc}$. The leftmost column has single realizations, the middle column stacks of 20, and the rightmost column stacks of 200 realizations. Each row has different sets of realizations for comparison. Single realizations only contain some features of the whole filamentary network. The more realizations are stacked, the smaller the differences between different sets. The rightmost panels only differ in minor detail. (Muru & Tempel 2021)

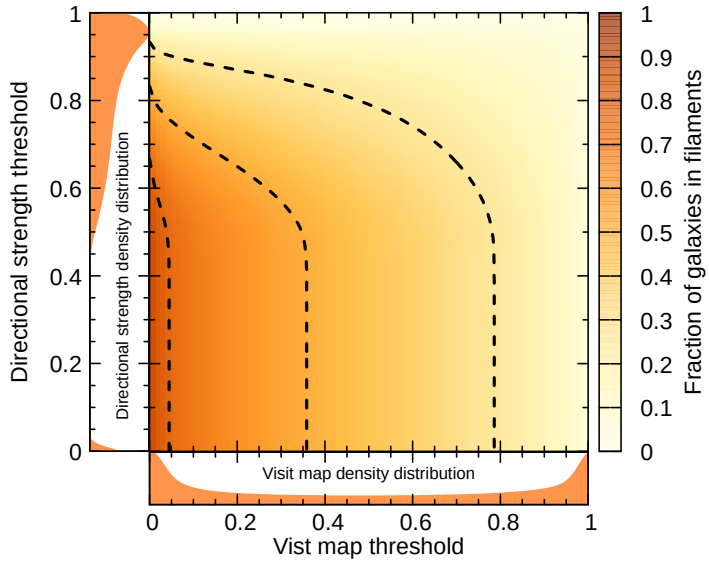

Figure 2.3: Fraction of galaxies in filaments as a function of visit and directional strength map thresholds. Dashed lines indicate isolines for the fractions of galaxies in filaments with values of (from left) 0.75, 0.5 and 0.25. Subplots on the left and below the main plot show distributions of directional strength and visit map values. (Muru & Tempel 2021)

out. This is repeated until all the volume, where the values are above thresholds, is masked out. This produces a catalog of filament spines. The lower box in Figure 2.1 shows an example of cylinders from 1000 simulations and the detected spines. These spines can be used to study, for example, alignment with filaments (Tempel & Libeskind 2013; Tempel & Tamm 2015; Ganeshaiah Veena et al. 2019) or different properties as a function of distance from the center of the filament (Kuutma et al. 2017; Tuominen et al. 2021). The filament radius is not defined in the catalog but is strongly influenced by the chosen cylinder radius in the model. This means that it is not straightforward to study the volume fill of the filaments and define whether a galaxy is inside or outside a filament, but a substitute for that is to use the visit map and a threshold to define the filament's extent in the space. Such an approach is utilized in the analysis of Muru & Tempel (2021, 2023).

The Bisous model is implemented in modern Fortran, and the source code is publicly available at a Github repository[1]. A single simulation of the Bisous model runs on a single core, and the memory usage is well-optimized and usually stays around a few hundred megabytes. To generate more of the independent realizations, usually, up to a hundred simulations are run in parallel. This offers a simple method to parallelize the calculations. A single simulation might take

[1]https://github.com/etempel/Bisous

from a few days up to a few weeks to generate enough realizations depending on the input data density and model parameters. An animation showcasing the Bisous model from the MCMC sampling to extracting the filaments and a fly-through of the SDSS filamentary network catalog is available at `https://www.aai.ee/~elmo/sdss-filaments/`. The implementation of the Bisous model was developed before this thesis project. This thesis involves extensive use and testing of the Bisous code to understand the algorithmic limitations of Bisous' mathematical model. The simulations were run in the High Performance Computing Center of the University of Tartu (University of Tartu 2018).

3 ROBUSTNESS OF THE BISOUS MODEL

As the Bisous model uses the MCMC sampling, it is inherently stochastic. This raises the question of how robust the results are and how it depends on the input data. This Chapter is based on Muru & Tempel (2021), where these questions were studied in depth. The aim is to study the model properties that depend on the input data and affect the robustness for a more accurate dissection of the results, mainly on observational data. However, for this study, we need more control and information on the data and therefore opted to use simulation data.

3.1 Data

The analysis in this paper is based on simulated mock data. The data originates from the MultiDark-Galaxies catalog from the MultiDark-Planck 2 simulation (MDPL2, Klypin et al. 2016) with SAG semi-analytic model for galaxies (Knebe et al. 2018). The MDPL2 is a dark matter-only simulation that uses ΛCDM model with Planck 2015 parameters: $\Omega_{\rm m} = 0.307$, $\Omega_{\rm B} = 0.048$, $\Omega_\Lambda = 0.693$, $\sigma_8 = 0.823$, $n_s = 0.96$, and $h = 0.678$ (Planck Collaboration et al. 2016). The simulation box size is $1000\,h^{-1}\,\mathrm{Mpc}$ ($1475.6\,\mathrm{Mpc}$) with 3840^3 particles and mass resolution of $m_{\rm p} = 1.51 \times 10^9\,h^{-1}\,M_\odot$. Taking into account the computational resources required by this study, we only used a smaller box with a size of $250\,\mathrm{Mpc}$, which is big enough sample for statistical analysis. We also used a magnitude cut of ≤ -18.5 in the SDSS *r* band. The final data contains $461\,551$ galaxies and has galaxy number density of $\rho_{\rm gal} = 0.0295\,\mathrm{Mpc}^{-3}$. Figure 3.1 shows the density field of the data after volume and magnitude cuts.

To study how the input data density affects the results, I created two sets of samples with different densities, one set with magnitude cuts from -21 to -18.5 mag, and the other set with densities matching the first set, but they are random samples rather than magnitude cuts. Figure 3.2 compares the galaxy number densities of the samples to SDSS galaxy number density as a function of distance. The highest-density sample, mag1, is a bit above the highest values in the SDSS catalog, and the lowest-density sample, mag6, has a similar density to SDSS at $400\,\mathrm{Mpc}$. I ran the Bisous model on all of the 12 samples.

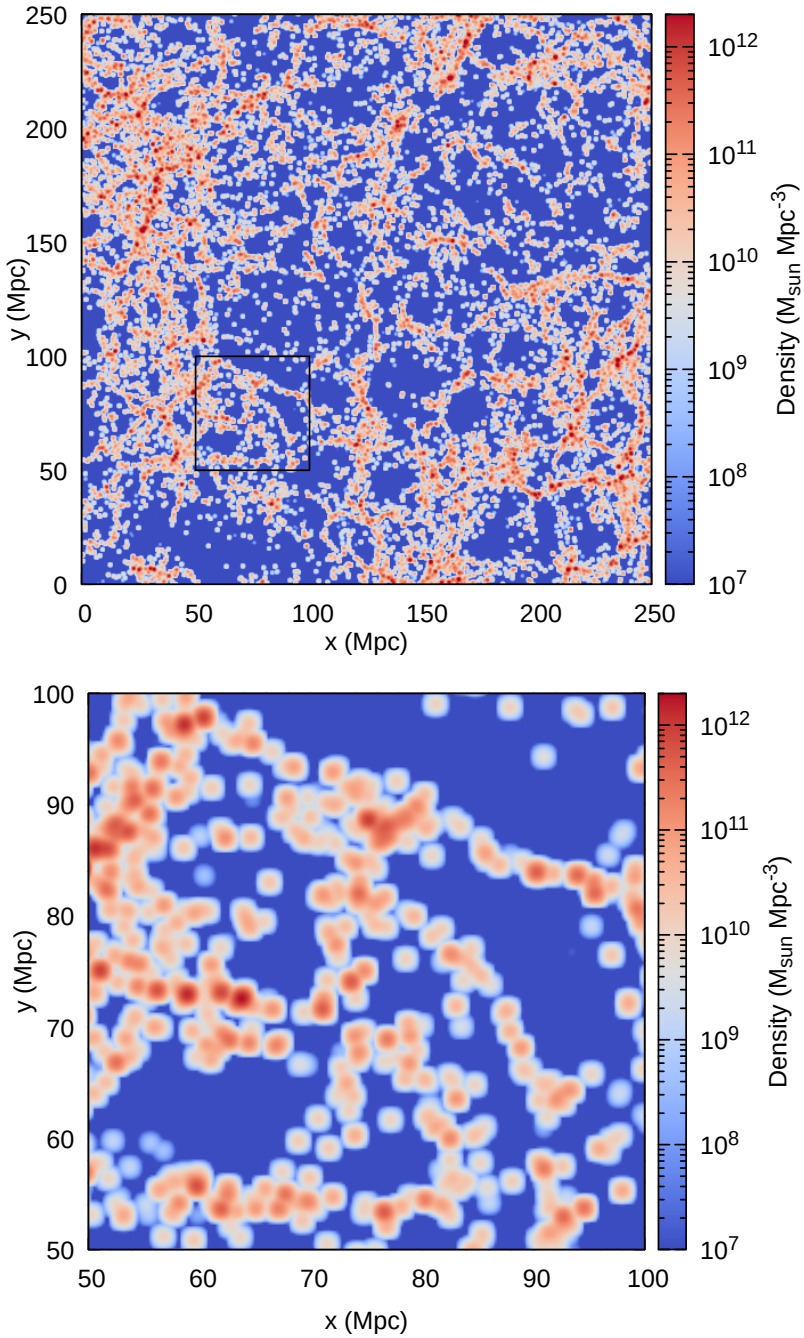

Figure 3.1: Mass density field in a slice after volume and magnitude cuts. The slice thickness is $10\,\mathrm{Mpc}$. Smoothing uses a Gaussian kernel with a width of $0.7\,\mathrm{Mpc}$. The lower panel shows a magnification of the area marked with a black square on the upper panel. (Muru & Tempel 2021)

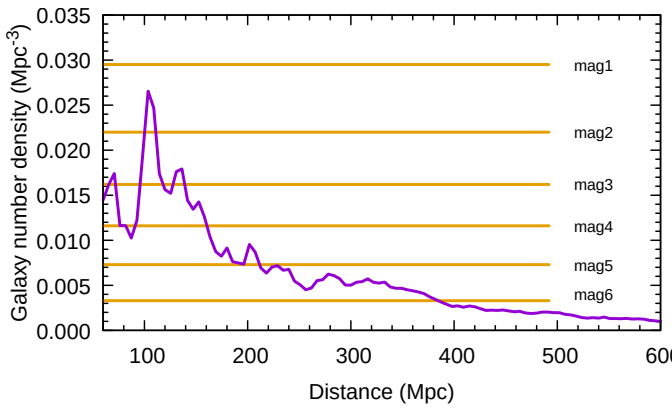

Figure 3.2: Galaxy number density for different samples. For comparison, the purple line indicates the number density of SDSS galaxies as a function of distance. (Muru & Tempel 2021)

3.2 Galaxy number density of the input data

The galaxy number density of flux-limited surveys changes significantly with distance. For a given wide-field survey, the selection of targets is usually done so that the galaxy number density is homogeneous in different areas of the sky. However, when comparing different surveys, the densities often differ, and the effects on the resulting filamentary network should be taken into account for a proper comparison. Figure 3.2 shows the galaxy number density of SDSS as a function of distance. Tempel et al. (2014b) states that the Bisous filaments catalog presented in that work is representative until about $240\,\mathrm{Mpc}$. This chapter presents the analysis of how the input galaxy number density affects the resulting filamentary network when using the Bisous method.

The lower the galaxy number density, the fewer filaments Bisous is able to detect. Figure 3.3 compares the volume fill of filaments using samples with different number densities and fractions of galaxies in filaments while counting all galaxies in the data or only the galaxies in the subsamples that were used to model the cosmic web. This analysis is done with both magnitude cut subsamples and random cut subsamples. Both plots show that, when using magnitude cuts, the fraction of galaxies in filaments and filament volume fill have slightly higher values. This indicates that galaxies in filaments are generally more luminous. When comparing the fractions of galaxies in filaments for all galaxies (solid lines) and only subsample galaxies (dashed lines), we can see how much more sharply the all galaxies fraction falls. This shows that when estimating the fraction of galaxies in filaments based on the galaxies we see in the observation, we overestimate the true fraction of galaxies in filaments.

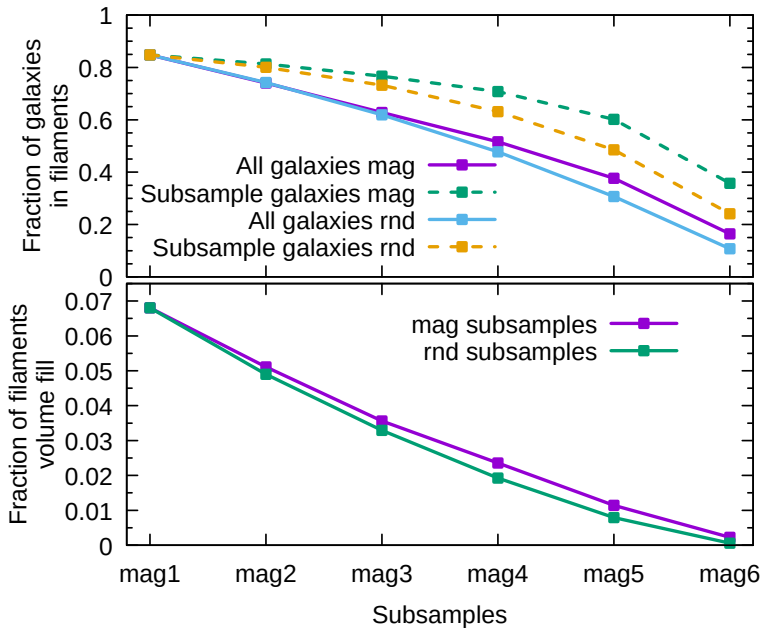

Figure 3.3: Fraction of galaxies inside filaments and filaments volume fill for different subsets. Solid lines show two different subsamples, where mag are made with magnitude cuts and rnd have matching number densities, but random galaxies. Dashed lines on the upper plot are calculated using only galaxies present in the subsample. In contrast, solid lines include all galaxies, i.e., also galaxies that were not visible to the filament finder. (Muru & Tempel 2021)

Muru & Tempel (2021) studies the false detections of filaments when comparing lower number density subsamples to higher ones. They find, when classifying the galaxies, that there are less than 3% of false positive classifications. When comparing the volume elements in a grid, the false positive detection goes up to 15%. When using lower input densities, the false negative ratio increases, but the false positives do not. This means that when studying observational data, the galaxies that are classified as filament galaxies are almost all true filament galaxies, even when the galaxy number density is low. But when defining some volume areas as filaments or not, about 15% of these filaments are spurious and do not exist in more complete data. This accuracy is quite high for a filament finder in comparison to how different results different filament finders produce on the same inputs (Libeskind et al. 2018). Figure 3.4 shows these effects visually. The areas that have high visit map values in lower-density subsamples also have high values in higher-density subsamples. There are very few exceptions. Figure 3.4 also shows how the detected filamentary network depends on the input data density and that the Bisous filament finder is conservative with its detection, which guarantees a lower number of false positive detections and the robustness of the detected structure.

3.3 Stochasticity in the results

As the Bisous model uses stochasticity for model optimization, it is important to study how this affects the results. This is something that affects the results regardless of the input number density and should be studied independently. For this, I applied the model with the same configuration on the same data 200 times and studied the differences in the output.

I used both visit and directional strength values to classify whether a galaxy is in a filament or not. Each galaxy has 200 classifications, which can be used to calculate the probability that a galaxy is in a filament based on its visit and directional strength values. Muru & Tempel (2021) finds that when using the usual visit map threshold of 0.05, about 68% galaxies were classified the same, either in a filament or not, 90% of the runs, i.e., the uncertainty of classification is less than 10%. Other 32% of the galaxies have uncertainty larger than 10%. The galaxies with uncertain classification are mostly situated in very weak and low population filaments or cluster outskirts as the model does not include clusters to the filamentary network, and it is difficult to define exactly where the filament merges with the cluster. About 38% of the galaxies are classified exactly the same in every run, half of those were always in filaments, and the other half were never in filaments in any of the runs.

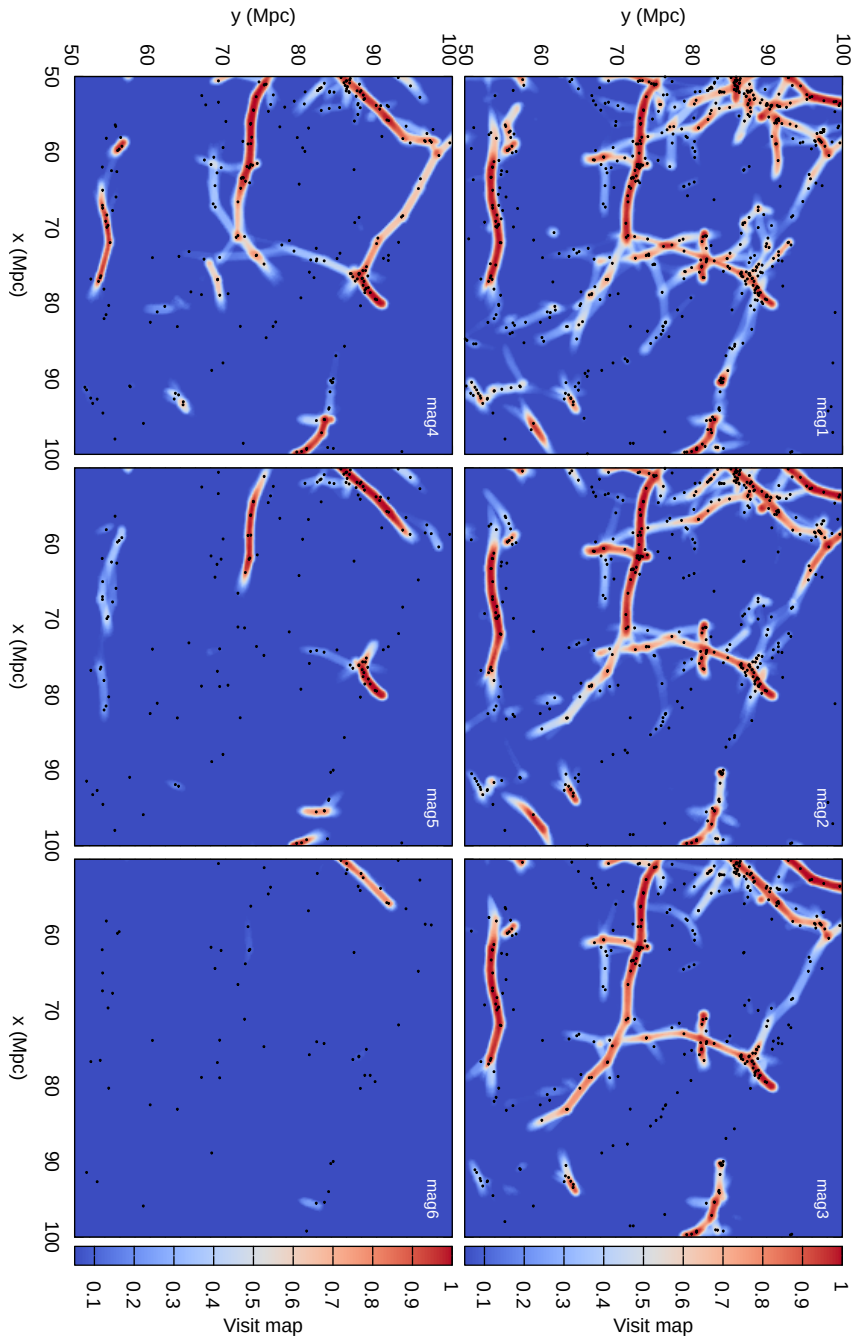

Figure 3.4: Maximum visit map values in a slice for different subsamples. The slice thickness is $10\,\mathrm{Mpc}$. The galaxies that are used to calculate the visit maps are shown as black dots. This is a $50\,\mathrm{Mpc} \times 50\,\mathrm{Mpc} \times 10\,\mathrm{Mpc}$ cut from the whole box. (Muru & Tempel 2021)

A different approach showed that when using only visit map values to compare different runs, then galaxies with a visit value higher than 0.1 were almost always in filaments. Correlation analysis showed that the Pearson correlation coefficient between the visit values of a galaxy in the 200 runs was 0.98. This indicates a very strong correlation between runs and that the model has been run until close to convergence, thus minimizing the stochasticity. For the results to converge completely, one would have to run the model infinitely long, which is practically unfeasible. These results are useful when trying to interpret the reliability of results from a single run of the model.

4 IMPROVING THE COSMIC WEB FILAMENT DETECTION WITH PHOTOMETRIC REDSHIFT DATA

In the previous chapter, we saw how the galaxy number density impacts how well the Bisous model is able to trace the filaments. To detect the large-scale structure of the universe using the Bisous model or other methods that use the positions of the galaxies, it is necessary to know the positions of the galaxies with relatively small errors. This is why only spectroscopic measurements have been used so far. The redshift estimate from the specter is significantly more accurate than the photometric measurements, where we only have as many data points for the redshift estimation as there are different filters.

Even though photometric surveys have magnitudes more data and better galaxy number density than spectroscopic surveys, so far, the data has found little use for large-scale structure detection purposes. So far, the most common use is to use high-precision photometric redshift surveys separately in high redshift surveys where spectroscopic redshifts are not available (Malavasi et al. 2017) and detecting 2D filaments in different redshift bins (Laigle et al. 2018).

The main problem is the significant uncertainty in the redshift estimates. This disfigures all the detected structures and elongates them along the line of sight or makes them undetectable. Elongation is a significant problem when studying the shapes or orientations of large-scale structure elements. Figure 4.1 compares mock data sets where I added errors along the line of sight to mimic data with photometric redshifts. The left panel shows data with spectroscopic redshift estimates and the other two data with a small error when compared to photometric surveys. The J-PAS survey (Benitez et al. 2014; Bonoli et al. 2020) is estimated to achieve redshift measurements with uncertainty approximately $10\,\mathrm{Mpc}$. For example, for SDSS, the photometric redshift uncertainties are about ten times worse (Beck et al. 2016). From the rightmost panel, it is evident that it is very difficult to detect the large-scale structure elements confidently. The Bisous model has been used to study alignments between filaments and galaxy spin (Tempel & Libeskind 2013; Tempel et al. 2014a), galaxy pairs (Tempel & Tamm 2015), and satellite planes around galaxies (Wang et al. 2020). Therefore, knowing the unbiased orientations, or at least biases affecting them, is important for many science questions.

Muru & Tempel (2023) proposes a simple method to incorporate data with photometric redshift in the filament-finding process with the Bisous model. This could improve the reliability of the detected filamentary network and increase the areas where Bisous can detect filaments. I first studied the photometric redshift data separately and then used mixed samples of spectroscopic and photometric

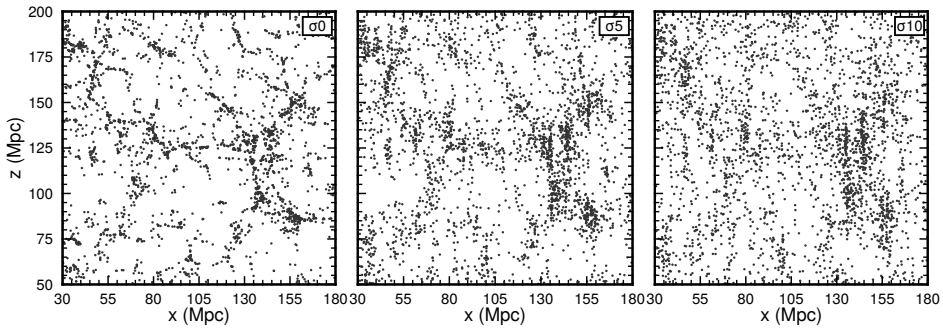

Figure 4.1: Galaxy distributions in different mock catalogs. The left panel shows the galaxy positions taken from a simulation. In the context of this analysis, this emulates a spectroscopic redshift survey, where the distances are very precise compared to photometric redshifts. The middle panel has added errors to the line of sight component (z-axis) with Gaussian random distribution with a standard deviation of $\sigma = 5\,\mathrm{Mpc}$ and the right panel with $\sigma = 10\,\mathrm{Mpc}$. (Muru & Tempel 2023)

redshifts in different fractions. The latter would help to bolster the galaxy number densities of spectroscopic surveys where it would be too low to detect the filaments reliably. Mainly to extend the area to higher redshifts. The aim is to be able to mitigate the error originating from the large uncertainty in the redshift estimate. The current chapter explains the methodology used to include data with high uncertainty in one axis with the Bisous model, then presents the mock data used for testing, and finally, the resulting filamentary networks and their reliability.

4.1 Methodology

The method to make use of photometric redshift data presented here is simplistic and serves as a proof of concept that photometric redshift data can be used beneficially for large-scale structure detection purposes. Bisous requires the galaxy data to be in comoving space and uses a cartesian coordinates system. This means that the used photometric data also has to be converted into comoving space and transformed into a cartesian coordinate system. The method requires the galaxy positions and probability density functions for the distance estimates. For simplicity, I used Gaussian distribution for the mock probability density functions. In practice, the distribution doesn't need analytical representation and can be different for every galaxy.

Usually, the Bisous model is run with $N_\mathrm{R} = 50$ to 100 parallel processes. Normally, these parallel processes use the same parameters and the same input, but to use the information on the probability density of distance estimates, I generated separate inputs for each run. Each input file is generated using the

galaxy positions and choosing random distance values based on the probability densities associated with the galaxies. This means that in each Bisous run, every galaxy has a new random position. For this analysis, I used $N_{\mathrm{R}} = 80$, meaning that 80 different input catalogs were generated from the initial one using the uncertainties for the distance measure. Each of the 80 Bisous runs generates hundreds of realizations of the cylinder configuration as explained in Chapter 2. These realizations are used to construct the visit and orientation maps for filament detection.

The idea behind this method is that randomly choosing galaxy positions based on the probability density enables us to use the whole information about the distance instead of just the mean value. In some runs, the galaxy is closer to its true position and produces more realizations with similar cylinder pattern and, therefore, will have higher visit and orientation map values. In other runs, where the galaxy position was chosen further away from the true position, it will produce fewer realizations with the same cylinder configuration and will not produce filaments. As seen from Chapter 3, the Bisous filament finder is rather conservative with its detections.

The biggest question is whether this simple method will produce a more reliable filamentary network map than the usual way with spectroscopic redshift data. I also explored a way to mix spectroscopic and photometric redshift data. In this case, the galaxies with spectroscopic redshifts are considered to have exact distances and have the same position in every Bisous run. When using these mixed input data, the question is whether adding the photometric redshift data improves the results compared to just using the spectroscopic data or makes it more difficult to find good cylinder configurations for the Bisous model.

4.2 Mock photometric redshift data

To analyze the method presented in the previous section, I created mock data sets with different uncertainties for the photometric redshift data and also mixed data sets with different fractions of spectroscopic and photometric redshift data. I used the same simulation as in Chapter 3. To imitate photometric redshift uncertainties, I added probability densities to the distances. I chose Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$ for the probability density functions for simplicity. Different data sets used Gaussian distributions with different standard deviation values ranging from $\sigma = 1\,\mathrm{Mpc}$ to $10\,\mathrm{Mpc}$, in total, six full photometric data sets. The highest value $\sigma = 10\,\mathrm{Mpc}$ was chosen because, at this point, the Bisous model already had a hard time detecting any filaments and started to take too many computational resources. For mixed samples, I chose two probability density functions for the photometric galaxies, $\sigma = 5\,\mathrm{Mpc}$ and $10\,\mathrm{Mpc}$, and proportions for the photometric and spectroscopic data mix ranging from 10% to 50% of

Table 4.1: Photometric distance uncertainties and percentage of spectroscopic galaxies in each sample. The distance uncertainties column shows the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ used to generate uncertainties for distances of photometric galaxies. The last column shows the percentage of the brightest galaxies with spectroscopic distances, i.e., exact distances. Samples that do not have galaxies with photometric distances are indicated with an em dash (—) in the second column. (Muru & Tempel 2023)

Name	Distance uncertainties	Spectroscopic distances
σ0	—	100%
σ1	$\mathcal{N}(0\,\mathrm{Mpc}, (1\,\mathrm{Mpc})^2)$	0%
σ2	$\mathcal{N}(0\,\mathrm{Mpc}, (2\,\mathrm{Mpc})^2)$	0%
σ3	$\mathcal{N}(0\,\mathrm{Mpc}, (3\,\mathrm{Mpc})^2)$	0%
σ5	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	0%
σ7	$\mathcal{N}(0\,\mathrm{Mpc}, (7\,\mathrm{Mpc})^2)$	0%
σ10	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	0%
σ5s50	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	50%
σ5s40	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	40%
σ5s30	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	30%
σ5s20	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	20%
σ5s10	$\mathcal{N}(0\,\mathrm{Mpc}, (5\,\mathrm{Mpc})^2)$	10%
σ10s50	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	50%
σ10s40	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	40%
σ10s30	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	30%
σ10s20	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	20%
σ10s10	$\mathcal{N}(0\,\mathrm{Mpc}, (10\,\mathrm{Mpc})^2)$	10%
s50	—	50%
s40	—	40%
s30	—	30%

spectroscopic redshift data in the sample. The galaxies with spectroscopic redshift estimates, i.e., exact distance, are always the brightest in the sample. In addition, I used spectroscopic-only samples with different galaxy number densities for comparison.

Table 4.1 lists all samples used for the analysis. The naming of samples encodes both the size of the photometrical distance uncertainty and the fraction of spectroscopic distances. The name σXXsYY gives the standard deviation $\sigma = \mathrm{XX}$ of the Gaussian distribution and the fraction of spectroscopic galaxies in the sample as YY%. All the samples have the same galaxy number density $\rho_{\mathrm{gal}} = 0.0162\,\mathrm{Mpc}$, except for the last three in Table 4.1. This galaxy number density was chosen as it is similar to the SDSS galaxy number density in close-by volume. The samples named sYY are spectroscopic-only and contain only YY% of galaxies, which all have exact distances. These are used for comparison for the mixed and photometric-only samples. The s30 sample is the lowest galaxy

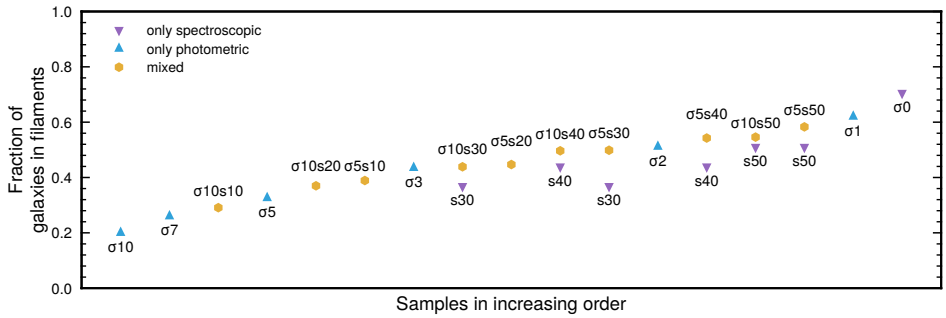

Figure 4.2: Fraction of galaxies in filaments for different spectroscopic-only, photometric-only, and mixed samples. The samples are ordered so that the y-axis values of photometric-only and mixed samples are in increasing order. The spectroscopic-only samples are used for reference values to show the increase in the fraction of galaxies in filaments for mixed samples. The samples are explained in Table 4.1. (Muru & Tempel 2023)

number density sample used because using even lower densities would yield almost no filaments, as shown in Chapter 3.

4.3 Bisous model with and without photometric redshift data

As there is no ground truth or consensus on how to define a filament, we need to use more general statistics to compare the different filamentary networks. One statistic that is also easy to infer for observational results is the fraction of galaxies in filaments. This statistic is also used for the robustness study in Chapter 3. The $\sigma 0$ sample result serves as a fiducial result as in the filaments the Bisous would detect from the dark matter distribution given the galaxies with no distance uncertainties. These can be considered as the true Bisous filaments for this data, but not as a general truth.

Figure 4.2 presents the fractions of galaxies in filaments for all the samples used in the analysis. As expected, the samples with very small uncertainties, unobtainable in photometric surveys, give similar results to the fiducial sample. When using photometric-only samples with larger uncertainties, such as $5\,\mathrm{Mpc}$ to $10\,\mathrm{Mpc}$, the amount of galaxies in filaments drops very low, meaning only the strongest filaments are detected and other galaxies that should also be in filaments are not. Using the mixed samples significantly boosts the fraction of galaxies in filaments when compared to photometric-only samples. Looking from the other side, using mixed samples is also an improvement over using just the spectroscopic samples. Figure 4.2 shows a comparison of the mixed samples and using only the spectroscopic galaxies in the mixed samples. We see that even with $\sigma = 10\,\mathrm{Mpc}$ photometric data, the mixed samples show better results when

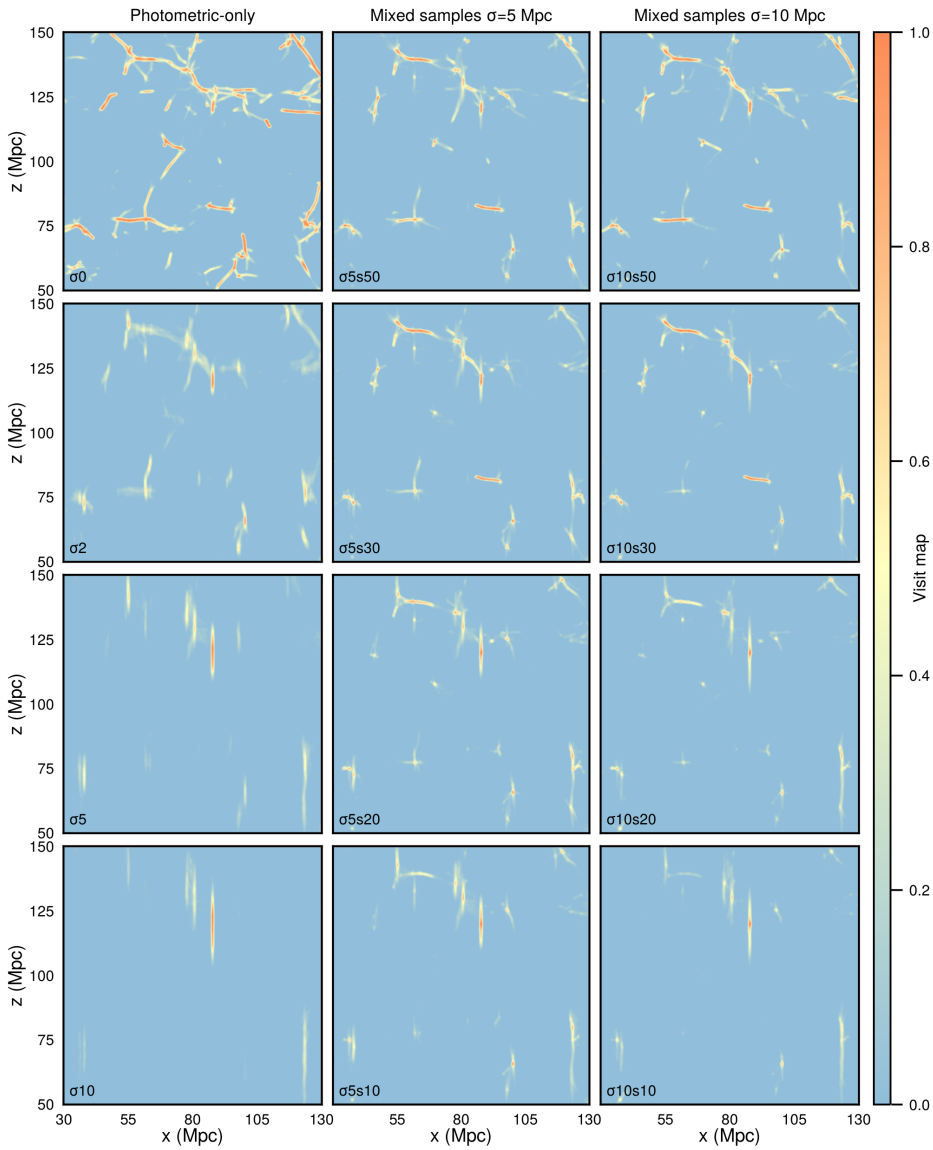

Figure 4.3: Projections of maximum visit map values in slices obtained from the Bisous model using different samples. Only a smaller $100\,\mathrm{Mpc} \times 100\,\mathrm{Mpc}$ area is shown for visual clarity. The thickness of the slice is $10\,\mathrm{Mpc}$. The vertical axis (z) is parallel to the axis of the distance uncertainties, i.e., it emulates the line of sight. The leftmost column shows samples with only photometric galaxies, the middle column shows mixed samples with medium uncertainties ($\sigma = 5\,\mathrm{Mpc}$), and the rightmost column shows mixed samples with larger uncertainties ($\sigma = 10\,\mathrm{Mpc}$). Different rows in the leftmost column have different photometric uncertainties, and the middle and the rightmost column have different fractions of the brightest galaxies as spectroscopic galaxies. (Muru & Tempel 2023)

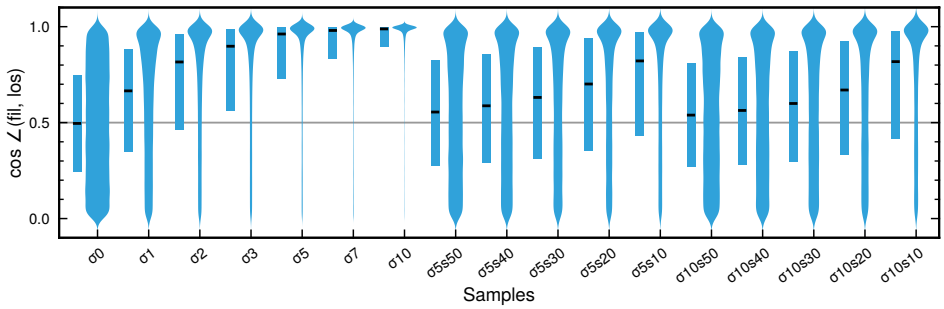

Figure 4.4: Distributions of the cosine of the angle between filament spines (fil) and the line of sight (los). For each sample, there are two plots. The left one is a bar plot of the quartiles of the distribution, where the black crossbar indicates the second quartile (the median). The right plot is a violin plot that shows the density curve of the distribution. The horizontal grey line indicates the median value for a uniform distribution. The closer the distribution gets to value 1, the more filaments are parallel to the line of sight (z-axis in other plots). (Muru & Tempel 2023)

comparing the fractions of galaxies in filaments.

Figure 4.3 shows the filaments in different samples. As low fractions of galaxies in filaments predict, samples with high distance uncertainty and little to no spectroscopic distances have significantly fewer filaments. In some samples, the Bisous model is only able to detect a few separate filaments, not the connected web. With higher distance uncertainties, mostly filaments parallel to the line of sight remain. Using mixed samples improves the detection of filaments with brighter galaxies but still suffers from orientation bias. When studying different alignments to the filaments, it is very important to know the true orientation of the filaments or at least describe the bias affecting the orientation.

To characterize the orientation bias when using galaxies with distance uncertainties, I studied the angle between the filament spines and the line of sight along which the distance uncertainty is spread. The base assumption is that the filaments should have no preferred orientation and be isotropically distributed. That means that the absolute value of the cosine of the angle between the filament spine and any fixed orientation should be a uniform distribution. These distributions are shown in Figure 4.4. The fiducial sample $\sigma 0$ shows a uniform distribution as expected. The photometric-only samples show that the distribution of cosines quickly deteriorates as the distance uncertainty increases, leaving almost only filaments parallel to the line of sight. Adding the spectroscopic galaxies to the mix significantly weakens the bias. Even though bias remains and filaments parallel to the line of sight are more common, using spectroscopic galaxies also makes the method usable for the alignment analysis when considering the bias.

If we consider the $\sigma 0$ run the ground truth for our data, we can study the

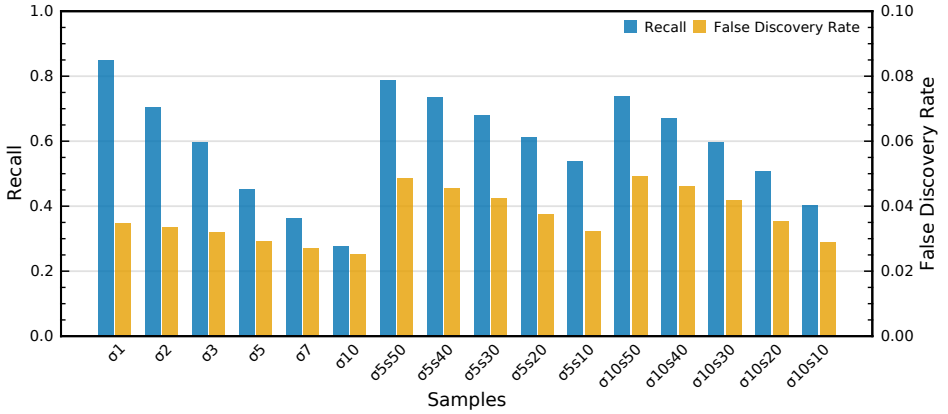

Figure 4.5: Recall and false discovery rates for photometric and mixed samples. All the samples have the same total number of galaxies. The false discovery rate uses the secondary vertical axis on the right side of the plot. Including spectroscopic galaxies improves recall but also increases false discovery rates. (Muru & Tempel 2023)

correctness of other runs. I used the statistical classification scheme commonly used in machine learning applications known as the confusion matrix. The fiducial run defines actual labels, and another selected run the predicted labels. Labels, in this case, show whether a galaxy resides in a filament. Therefore, each predicted label can take one of four values in the confusion matrix: true positive, true negative, false positive, or false negative. Using these, we can define different statistics, I used the recall

$$\text{Recall}_s = \frac{\text{TP}_s}{\text{P}_{\sigma 0}} \,, \tag{4.1}$$

where TP_s is the number of true positive values in the sample s, and $\text{P}_{\sigma 0}$ the number of positive values in the fiducial sample $\sigma 0$;
and the false discovery rate,

$$\text{False discovery rate}_s = \frac{\text{FP}_s}{\text{P}_s} \,, \tag{4.2}$$

where FP_s is the number of false positive values in the sample s, and P_s is the number of all positive values in the sample s, which includes both the true positive and false positive values. The recall describes how well the model can find the correct filaments, and the false discovery rate shows what fraction of the detected filaments are incorrectly detected, i.e., not true filaments. The recall and false discovery rates for the photometric-only and mixed samples are shown in Figure 4.5. As seen from Chapter 3, Bisous model detections are rather conservative, and even with the photometric redshift estimates, the false discovery rates are all below 5%. This is an important result, as even if the model cannot detect all or

Table 4.2: Comparison of Bisous model results with and without using the method described in Sect. 4.1. PB in front of the sample name indicates that the results are obtained with the plain Bisous model. Recall and false discovery rate are defined by Equations 4.1 and 4.2. (Muru & Tempel 2023)

Sample	gal in fil[a]	Recall	FDR[b]
PB $\sigma5$	0.217	0.296	0.037
$\sigma5$	0.329	0.452	0.029
PB $\sigma5$s30	0.398	0.541	0.039
$\sigma5$s30	0.501	0.679	0.042
PB $\sigma10$	0.127	0.172	0.039
$\sigma10$	0.200	0.276	0.025
PB $\sigma10$s30	0.348	0.473	0.039
$\sigma10$s30	0.441	0.598	0.042

[a] Fraction of galaxies in filaments; [b] false discovery rate

most of the filaments, at least the ones it detects are likely to be true filaments. The recall drops with increasing distance uncertainties, but using mixed samples slows it down. When using large uncertainties ($\sigma = 10\,\mathrm{Mpc}$) and 40% of the galaxies have spectroscopic distance estimates, then still about 70% of filaments are detected when compared against knowing all the data without uncertainties.

Lastly, there remains a question of whether the method of sampling the distances from the probability distributions improves anything, or would applying the Bisous model on given galaxy positions without taking the uncertainties into account work as well. I applied the Bisous model to the data without sampling the distances to answer this. I will call these runs plain Bisous (PB) runs. Table 4.2 includes four different samples, and each was used as Bisous input with and without the sampling method. The table lists fractions of galaxies in filaments, recalls, and false discovery rates of these eight runs. The fraction of galaxies in filaments is higher for runs using the sampling method for every sample, which shows that the model finds more filaments this way. This is also confirmed by the recall values, showing that when compared to the fiducial run $\sigma0$ the results using the sampling method find more correct filaments. When using the sampling method, the false discovery rates show a slight increase in false detections for the mixed samples. The differences are small and come from the fact that those samples detect more filaments overall and, therefore, some more incorrect ones. In conclusion, the sampling method substantially improves the detected filamentary network compared to the plain Bisous runs.

5 DISCUSSION AND CONCLUSIONS

The large-scale structure of the universe called the cosmic web affects the evolution of groups and galaxies, and also carries information on the evolution of the universe. It is paramount that we study the large-scale structure and be able to map it using observations. The detection of the filamentary network from the observations is difficult because the information we are able to obtain is limited and ambiguous. For example, the galaxy distribution is a biased tracer for the matter density, and observational effects obfuscate the measured data, which produces errors in the analysis. The sky area that is covered with observations with good enough data is also limited. The Bisous model aims to fill the niche of detecting filaments from observations.

The aims of this thesis are to test the limits of the Bisous in environments with low galaxy number density, study the robustness of the detected filaments and the effects of the stochasticity in the model, and study the prospect of using data with photometric redshifts to improve the detected filamentary networks. Understanding how Bisous works in low galaxy number density environments and how the stochastic behavior of the model affects the results is essential in confining the uncertainties in the resulting filamentary network. Having a better understanding of the uncertainties is critical to making informed conclusions from the model's results. For example, when studying the population of galaxies in filaments and the model is only able to detect the brightest filaments, the results will be biased, and this has to be taken into account. For mapping purposes, it is self-explanatory that the positions of galaxies should have smaller uncertainties or of the same magnitude as the structure we are trying to map. This is why photometric redshift data has not been commonly used so far for large-scale structure mapping as the uncertainties are sometimes hundreds of times larger than the filaments themselves. But there are magnitudes more photometric redshift data (Beck et al. 2016) and it is correlated with filaments (Kruuse et al. 2019). Therefore, it should be possible to make use of it. This doctoral work is about these ideas, analysis, and findings.

The Bisous model has been used to compile a filaments catalog for the SDSS sample (Tempel et al. 2014b) and is well-suited for observational data. Still, the analysis done for this thesis uses data from dark matter-only simulations with semi-analytical galaxies (MultiDark-Galaxies, Klypin et al. 2016). This provided us with a catalog without observational biases and known completeness. More complete information on the data made it possible to study the properties and effects of the Bisous model and also enabled us to compare how well the filaments trace the dark matter structure of the cosmic web.

The main results of this thesis are:

1. The effectiveness of the Bisous model in finding the filaments depends strongly on the galaxy number density. The lower the number density, the fewer filaments are detected (cf. Fig 3.3 and 3.4), but the model produces very few false positive detections (cf. Sect 3.2). When using an incomplete galaxy catalog for the Bisous model, more than 97% of the galaxies are correctly classified, and more than 85% of the volume is correctly classified as a filament (cf. Article I Fig 8).

2. The filamentary networks from the Bisous model are very robust. Comparing the results from 200 Bisous runs, I found a correlation of $R = 0.98$ (cf. Sect 3.3). This shows that even when the model is built on stochastic frameworks, it is run until near convergence, and the results do not change much from run to run. These results help to estimate the uncertainty in the filamentary network modeled by the Bisous.

3. I present a proof of concept that photometric redshift data, i.e., data with large uncertainty for distance, can be used to improve filament detection with the Bisous model. Catalogs of mixed data with spectroscopic and photometric redshifts show improvements over using only the spectroscopic redshift galaxies.

4. Using photometric redshift data introduces an orientation bias for the filaments, where filaments parallel to the line of sight are more common (cf. Fig 4.4). Using mixed samples alleviates this bias but not entirely, and it should still be taken into account when studying alignments.

5. The recall and false discovery rate depends on the fraction of spectroscopic galaxies and the size of the distance uncertainty. The false discovery rates were below 5% for every sample analyzed and recalls reached 70-80% even with significant distance uncertainties and at least 40% of spectroscopic redshifts in the mixed sample (cf. Fig 4.5). This shows again the conservative nature of the Bisous model, only producing small amounts of false positive results.

All of the current photometric surveys still have unusably large redshift uncertainties. For example, SDSS photometric catalogs have uncertainties ten times larger than the ones analyzed in this work (Beck et al. 2016). Fortunately, the next generation of photometric surveys is already here. The Javalambre Physics of the Accelerating Universe Astrophysics Survey (J-PAS, Benitez et al. 2014; Bonoli et al. 2020) is an upcoming photometric survey that uses 59 filters and aims to achieve a new level of accuracy for the redshifts. They estimate to produce redshifts for galaxies in the redshift range of $0.1 < z < 1.2$ with the

precision of $\sigma_z = 0.003(1+z)$. This would mean that at redshift $z = 0.1$ the precision would be $\sigma_z = 0.003 \times 1.1 \approx 14\,\mathrm{Mpc}$, which is of the same order of magnitude as the $\sigma 10$ samples in this analysis. Combining this with the SDSS data to create the mixed samples of spectroscopic and photometric redshift data, we could extend the filamentary networks detected to higher redshifts than with only SDSS data. The J-PAS is set to start observations at the end of 2023 or early 2024, and we plan to apply the Bisous model to the data once available. In addition to the SDSS, the J-PAS data can be combined with other spectroscopic surveys in the northern hemisphere, for example, the Dark Energy Spectroscopic Instrument (DESI) Bright Galaxy Survey (BGS) (Dey et al. 2019; Ruiz-Macias et al. 2021).

5.1 Future improvements for the Bisous model

Although the Bisous model has already been used for compiling filament catalogs from observational data and a plethora of scientific discoveries, there are many ways and plans to add improvements to the model. Many of the Bisous model's parameters have been manually tuned to produce a connected filamentary network and to help with the convergence. These parameters have to be recalibrated for every new data set, which involves testing. This is time- and resource-consuming and requires specific knowledge about the Bisous model. In an ideal case, most of it would be done automatically. One option is to implement the Approximate Bayesian Computation (ABC) Shadow algorithm (Stoica et al. 2015). This algorithm enables us to sample the model parameters from the posterior densities, which would increase the convergence of the model and eliminate the need to propose prior densities for the parameters. It is specifically developed for spatial pattern models similar to the Bisous model.

Currently, the model only detects the most prominent part of the cosmic web, the filamentary network. This is somewhat limiting because it requires filaments to be connected to other filaments to create a network, but often filaments are connected to large clusters called nodes in the context of the large-scale structure of the universe. These nodes appear as discontinuities in the Bisous filamentary network. For a more complete view of the cosmic web, the nodes (clusters) and filaments should be detected simultaneously, or the model should be able to take a galaxy group catalog as an input and use it to detect filaments. We plan to merge the grouping algorithm with the Bisous model to generate a catalog of the complete cosmic web.

Using the groups and clusters as connection nodes for the filaments is especially important when applying the Bisous model on the next-generation survey data that have magnitudes higher galaxy number density than, for example, SDSS, for which the current Bisous model has been developed. The galaxy

number densities studied in Article I and presented in Section 3.2 are similar to SDSS galaxy number densities, as the current model is not optimized to work with densities much higher. In these cases, one of the improvements to the model could be that a grouping algorithm is run on the data, and groups are aggregated to a single data point which the Bisous model uses to define the model's data energy. Grouping is also important to suppress redshift space distortions, such as Fingers of God. This or some equivalent approach is needed to apply the Bisous model to the upcoming 4-metre Multi-Object Spectroscopic Telescope (4MOST, de Jong et al. 2019) data. The 4MOST Wide-Area VISTA Extragalactic Survey (WAVES, Driver et al. 2019) will produce a denser and deeper galaxy catalog than any currently available. For example, the galaxy number density will be 10 to 100 times higher than in the SDSS. We plan to apply the Bisous on this new and intriguing data set, but the model needs to be upgraded beforehand.

5.2 Usefulness of the marked point processes

The Bisous model is based on the marked point process that is specifically designed for filament extraction from galaxy distribution in the universe. However, marked point processes are a universal tool for spatial analysis, and different models can be developed to detect or infer different things from spatial data. There are a few examples from astronomy where marked point processes have been utilized.

To be more specific, the marks in the Bisous model are cylinders, and the process is sometimes named object point process, meaning the mark represents an object, for example, a shape. A similar model has been used for group detection by Tempel et al. (2018). The interaction and data energies had to be changed to solve the correct optimization problem, but one of the more obvious changes was the new object for the mark. In the case of filaments, we are looking for elongated structures that we want to build link-by-link. For this, the cylinder is an optimal shape. To detect the groups, roughly spherical collapsed structures, a sphere would be more sensible. But because of the redshift space distortions, an elongated sphere is chosen, as we expect the groups to be stretched out along the line of sight.

A marked point process is also being used to optimize the observing strategy of the 5-year 4MOST survey. Tempel et al. (2020) presents a model to optimize the tiling pattern, which describes all the pointings of the telescope. This mainly depends on the target distribution and the shape and size of the telescope's field of view, which is a hexagon. For this purpose, the mark is chosen to be a hexagon. Similarly to data energy for filaments and group detection, this model defines energy that depends on how well the hexagon distribution (the point process realization) covers the targets. In addition, the model has to take into account

the target properties, for example, requested exposure time, sky conditions, and overhead associated with the pointings. This all complicates the model in the sense that it is more difficult to come up with suitable energy definitions and marks, but it is organically feasible with object point processes.

Another application of object point processes is minimizing the bias of measured peculiar velocities. Bias-minimized peculiar velocities enable us, for example, to test the Hubble constant value and produce more accurate constrained simulations of the local universe. Sorce et al. (2023) apply the marked point process and Bayesian inference to minimize the Malmquist bias by modeling and perturbing the distribution of peculiar velocities.

In conclusion, building models with a marked point process is a useful and versatile tool to have in one's toolset. Finding suitable energy definitions and marks can take time and effort, but marked point processes are a powerful tool that can help study complex problems. In comparison with neural networks, marked point processes are not black boxes, you know exactly what and why it is doing something and can be used on different kinds of data without needing to retrain. Marked point processes have a large field of applications in astronomy and elsewhere.

5.3 Concluding remarks

The scientific area of cosmic web studies and large-scale structure detection is very active and competitive. The definition of a filament still eludes us, because of its complex and dynamic nature. Filament finders use different approaches to detect the structures most suitable to their study case. New filament finders are published to take advantage of new methods and data, and previous ones are continuously improved to keep up. This thesis contributes to the development and analysis of the Bisous model to be ready for next-generation cosmological surveys. As the observations get more precise and accurate, and cover larger volumes, simulations improve the resolution, size, and implemented physics, and detection methods are being perfected, so will our understanding of the universe and its structure steadily become clearer.

REFERENCES

Alpaslan, M., Robotham, A. S. G., Driver, S., et al. 2014, *Galaxy And Mass Assembly (GAMA): the large-scale structure of galaxies and comparison to mock universes*, MNRAS, 438, 177

Beck, R., Dobos, L., Budavári, T., Szalay, A. S., & Csabai, I. 2016, *Photometric redshifts for the SDSS Data Release 12*, MNRAS, 460, 1371

Benitez, N., Dupke, R., Moles, M., et al. 2014, *J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey*, arXiv e-prints, arXiv:1403.5237

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *How filaments of galaxies are woven into the cosmic web*, Nature, 380, 603

Bonnaire, T., Aghanim, N., Decelle, A., & Douspis, M. 2020, *T-ReX: a graph-based filament detection method*, A&A, 637, A18

Bonoli, S., Marín-Franch, A., Varela, J., et al. 2020, *The miniJPAS survey: a preview of the Universe in 56 colours*, arXiv e-prints, arXiv:2007.01910

Canducci, M., Awad, P., Taghribi, A., et al. 2022, *1-DREAM: 1D Recovery, Extraction and Analysis of Manifolds in noisy environments*, Astronomy and Computing, 41, 100658

Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, *NEXUS: tracing the cosmic web connection*, MNRAS, 429, 1286

Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, *Evolution of the cosmic web*, MNRAS, 441, 2923

Courtois, H. M., Dupuy, A., Guinet, D., et al. 2023, *Gravity in the local Universe: Density and velocity fields using CosmicFlows-4*, A&A, 670, L15

Cui, W., Knebe, A., Libeskind, N. I., et al. 2019, *The large-scale environment from cosmological simulations II: The redshift evolution and distributions of baryons*, MNRAS, 485, 2367

Cui, W., Knebe, A., Yepes, G., et al. 2018, *The large-scale environment from cosmological simulations - I. The baryonic cosmic web*, MNRAS, 473, 68

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, *4MOST: Project overview and information for the First Call for Proposals*, The Messenger, 175, 3

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, *A Slice of the Universe*, ApJ, 302, L1

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *Overview of the DESI Legacy Imaging Surveys*, AJ, 157, 168

Dolag, K., Sorce, J. G., Pilipenko, S., et al. 2023a, *Simulating the LOcal Web (SLOW): I. Anomalies in the local density field*, arXiv e-prints, arXiv:2302.10960

Dolag, K., Sorce, J. G., Pilipenko, S., et al. 2023b, *Simulating the LOcal Web (SLOW): I. Anomalies in the local density field*, arXiv e-prints, arXiv:2302.10960

Dome, T., Fialkov, A., Sartorio, N., & Mocz, P. 2023, *Cosmic Web Dissection in Fuzzy Dark Matter Cosmologies*, arXiv e-prints, arXiv:2301.09762

Driver, S. P., Liske, J., Davies, L. J. M., et al. 2019, *4MOST Consortium Survey 7: Wide-Area VISTA Extragalactic Survey (WAVES)*, The Messenger, 175, 46

Einasto, J. 2018, *Cosmology Paradigm Changes*, ARA&A, 56, 1

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems*, AJ, 142, 72

Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022, *Euclid preparation. I. The Euclid Wide Survey*, A&A, 662, A112

Feldbrugge, J. & van de Weygaert, R. 2023, *Cosmic web & caustic skeleton: non-linear constrained realizations - 2D case studies*, Journal of Cosmology and Astroparticle Physics, 2023, 058

Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, *A dynamical classification of the cosmic web*, MNRAS, 396, 1815

Galárraga-Espinosa, D., Aghanim, N., Langer, M., & Tanimura, H. 2021, *Properties of gas phases around cosmic filaments at z = 0 in the IllustrisTNG simulation*, A&A, 649, A117

Galárraga-Espinosa, D., Garaldi, E., & Kauffmann, G. 2023, *Flows around galaxies. I. The dependence of galaxy connectivity on cosmic environments and effects on the star formation rate*, A&A, 671, A160

Ganeshaiah Veena, P., Cautun, M., Tempel, E., van de Weygaert, R., & Frenk, C. S. 2019, *The Cosmic Ballet II: spin alignment of galaxies and haloes with large-scale filaments in the EAGLE simulation*, MNRAS, 487, 1607

Ganeshaiah Veena, P., Cautun, M., van de Weygaert, R., Tempel, E., & Frenk, C. S. 2021, *Cosmic Ballet III: Halo spin evolution in the cosmic web*, MNRAS, 503, 2280

Ganeshaiah Veena, P., Cautun, M., van de Weygaert, R., et al. 2018, *The Cosmic Ballet: spin and shape alignments of haloes in the cosmic web*, MNRAS, 481, 414

Ganeshaiah Veena, P., Lilow, R., & Nusser, A. 2023, *Large-scale density and velocity field reconstructions with neural networks*, MNRAS, 522, 5291

Gouin, C., Bonnaire, T., & Aghanim, N. 2021, *Shape and connectivity of groups and clusters: Effect of the dynamical state and accretion history*, A&A, 651, A56

Guzzo, L., Scodeggio, M., Garilli, B., et al. 2014, *The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at 0.5 < z < 1.2*, A&A, 566, A108

Ho, A., Gronke, M., Falck, B., & Mota, D. F. 2018, *Probing modified gravity in cosmic filaments*, A&A, 619, A122

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, ApJ, 873, 111

Jõeveer, M., Einasto, J., & Tago, M. 1977, *The cell structure of the Universe*, Tartu Astrofüüs. Obs. Preprint, 1, A

Kaiser, N. 1984, *On the spatial correlations of Abell clusters.*, ApJ, 284, L9

Kaiser, N. 1987, *Clustering in real space and in redshift space*, MNRAS, 227, 1

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, *MultiDark simulations: the story of dark matter halo concentrations and density profiles*, MNRAS, 457, 4340

Knebe, A., Devriendt, J. E. G., Gibson, B. K., & Silk, J. 2003, *Top-down fragmentation of a warm dark matter filament*, MNRAS, 345, 1285

Knebe, A., Stoppacher, D., Prada, F., et al. 2018, *MULTIDARK-GALAXIES: data release and first results*, MNRAS, 474, 5206

Kraljic, K., Arnouts, S., Pichon, C., et al. 2018, *Galaxy evolution in the metric of the cosmic web*, MNRAS, 474, 547

Kruuse, M., Tempel, E., Kipper, R., & Stoica, R. S. 2019, *Photometric redshift galaxies as tracers of the filamentary network*, A&A, 625, A130

Kuo, J.-L., Lattanzi, M., Cheung, K., & Valle, J. W. F. 2018, *Decaying warm dark matter and structure formation*, Journal of Cosmology and Astroparticle Physics, 2018, 026

Kuutma, T., Tamm, A., & Tempel, E. 2017, *From voids to filaments: environmental transformations of galaxies in the SDSS*, A&A, 600, L6

Laigle, C., Pichon, C., Arnouts, S., et al. 2018, *COSMOS2015 photometric redshifts probe the impact of filaments on galaxy properties*, MNRAS, 474, 5437

Lavaux, G. & Wandelt, B. D. 2012, *Precision Cosmography with Stacked Voids*, ApJ, 754, 109

Libeskind, N. I., Carlesi, E., Grand, R. J. J., et al. 2020, *The HESTIA project: simulations of the Local Group*, MNRAS, 498, 2968

Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2018, *Tracing the cosmic web*, MNRAS, 473, 1195

Malavasi, N., Aghanim, N., Douspis, M., Tanimura, H., & Bonjean, V. 2020, *Characterising filaments in the SDSS volume from the galaxy distribution*, A&A, 642, A19

Malavasi, N., Arnouts, S., Vibert, D., et al. 2017, *The VIMOS Public Extragalactic Redshift Survey (VIPERS): galaxy segregation inside filaments at z = 0.7*, MNRAS, 465, 3817

Malavasi, N., Langer, M., Aghanim, N., Galárraga-Espinosa, D., & Gouin, C. 2022, *Relative effect of nodes and filaments of the cosmic web on the quenching of galaxies and the orientation of their spin*, A&A, 658, A113

Muru, M. M. & Tempel, E. 2021, *Assessing the reliability of the Bisous filament finder*, A&A, 649, A108

Muru, M. M. & Tempel, E. 2023, *Using photometric redshift data to improve the detection of galactic filaments with the Bisous model*, A&A, 670, A77

Nelson, D., Springel, V., Pillepich, A., et al. 2019, *The IllustrisTNG simulations: public data release*, Computational Astrophysics and Cosmology, 6, 2

Pakmor, R., Springel, V., Coles, J. P., et al. 2022, *The MillenniumTNG Project: The hydrodynamical full physics simulation and a first look at its galaxy clusters*, arXiv e-prints, arXiv:2210.10060

Park, D. & Lee, J. 2007, *Void Ellipticity Distribution as a Probe of Cosmology*, Phys. Rev. Lett., 98, 081301

Peebles, P. J. E. 1973, *Statistical Analysis of Catalogs of Extragalactic Objects. I. Theory*, ApJ, 185, 413

Peebles, P. J. E. 1980, The large-scale structure of the universe

Pisani, A., Sutter, P. M., Hamaus, N., et al. 2015, *Counting voids to probe dark energy*, Phys. Rev. D, 92, 083531

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *Planck 2015 results. XIII. Cosmological parameters*, A&A, 594, A13

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *Planck 2018 results. VI. Cosmological parameters*, A&A, 641, A6

Radinović, S., Nadathur, S., Winther, H. A., et al. 2023, *Euclid: Cosmology forecasts from the void-galaxy cross-correlation function with reconstruction*, arXiv e-prints, arXiv:2302.05302

Rocha, M., Peter, A. H. G., Bullock, J. S., et al. 2013, *Cosmological simulations with self-interacting dark matter - I. Constant-density cores and substructure*, MNRAS, 430, 81

Ruiz-Macias, O., Zarrouk, P., Cole, S., et al. 2021, *Characterizing the target selection pipeline for the Dark Energy Spectroscopic Instrument Bright Galaxy Survey*, MNRAS, 502, 4328

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*, MNRAS, 446, 521

Shandarin, S. F. & Zeldovich, Y. B. 1989, *The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium*, Reviews of Modern Physics, 61, 185

Sorce, J. G., Stoica, R. S., & Tempel, E. 2023, *Statistically bias-minimized peculiar velocity catalogs from Gibbs point processes and Bayesian inference*, arXiv e-prints, arXiv:2309.03945

Sousbie, T. 2011, *The persistent cosmic web and its filamentary structure - I. Theory and implementation*, MNRAS, 414, 350

Sousbie, T., Pichon, C., & Kawahara, H. 2011, *The persistent cosmic web and its filamentary structure - II. Illustrations*, MNRAS, 414, 384

Sprenger, T., Archidiacono, M., Brinckmann, T., Clesse, S., & Lesgourgues, J. 2019, *Cosmology in the era of Euclid and the Square Kilometre Array*, Journal of Cosmology and Astroparticle Physics, 2019, 047

Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *First results from the IllustrisTNG simulations: matter and galaxy clustering*, MNRAS, 475, 676

Stoica, R., Gay, E., & Kretzschmar, A. 2007a, *Cluster Pattern Detection in Spatial Data Based on Monte Carlo Inference*, Biometrical journal. Biometrische Zeitschrift, 49, 505

Stoica, R., Gregori, P., & Mateu, J. 2005a, *Simulated annealing and object point processes: Tools for analysis of spatial patterns*, Stochastic Processes and their Applications, 115, 1860

Stoica, R., Martínez, V., & Saar, E. 2007b, *A three-dimensional object point process for detection of cosmic filaments*, Journal of the Royal Statistical Society Series C, 56, 459

Stoica, R. S., Martínez, V. J., & Saar, E. 2010, *Filaments in observed and mock galaxy catalogues*, A&A, 510, A38

Stoica, R. S., Martínez, V. J., Mateu, J., & Saar, E. 2005b, *Detection of cosmic filaments using the Candy model*, A&A, 434, 423–432

Stoica, R. S., Philippe, A., Gregori, P., & Mateu, J. 2015, *ABC Shadow algorithm: a tool for statistical analysis of spatial patterns*, arXiv e-prints, arXiv:1507.04228

Stoica, R. S., Tempel, E., Liivamägi, L. J., Castellan, G., & Saar, E. 2014, *Spatial Patterns Analysis in Cosmology based on Marked Point Processes*, in EAS Publications Series, Vol. 66, EAS Publications Series, ed. D. Fraix-Burnet & D. Valls-Gabaud, 197

Tempel, E., Kruuse, M., Kipper, R., et al. 2018, *Bayesian group finder based on marked point processes. Method and feasibility study using the 2MRS data set*, A&A, 618, A81

Tempel, E. & Libeskind, N. I. 2013, *Galaxy Spin Alignment in Filaments and Sheets: Observational Evidence*, ApJ, 775, L42

Tempel, E., Libeskind, N. I., Hoffman, Y., Liivamägi, L. J., & Tamm, A. 2014a, *Orientation of cosmic web filaments with respect to the underlying velocity field*, MNRAS, 437, L11

Tempel, E., Stoica, R. S., Kipper, R., & Saar, E. 2016, *Bisous model-Detecting filamentary patterns in point processes*, Astronomy and Computing, 16, 17

Tempel, E., Stoica, R. S., Martínez, V. J., et al. 2014b, *Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS*, MNRAS, 438, 3465

Tempel, E. & Tamm, A. 2015, *Galaxy pairs align with Galactic filaments*, A&A, 576, L5

Tempel, E., Tuvikene, T., Muru, M. M., et al. 2020, *An optimized tiling pattern for multiobject spectroscopic surveys: application to the 4MOST survey*, MNRAS, 497, 4626

Thélie, E., Aubert, D., Gillet, N., & Ocvirk, P. 2022, *First look at the topology of reionisation redshifts in models of the epoch of reionisation*, A&A, 658, A139

Tuominen, T., Nevalainen, J., Heinämäki, P., et al. 2023, *Cosmic metal invaders: Intergalactic O VII as a tracer of the warm-hot intergalactic medium within cosmic filaments in the EAGLE simulation*, A&A, 671, A103

Tuominen, T., Nevalainen, J., Tempel, E., et al. 2021, *An EAGLE view of the missing baryons*, A&A, 646, A156

University of Tartu. 2018, *UT Rocket*

Wang, P., Libeskind, N. I., Tempel, E., Kang, X., & Guo, Q. 2021, *Possible observational evidence for cosmic filament spin*, Nature Astronomy, 5, 839

Wang, P., Libeskind, N. I., Tempel, E., et al. 2020, *The Alignment of Satellite Systems with Cosmic Filaments in the SDSS DR12*, ApJ, 900, 129

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *The Sloan Digital Sky Survey: Technical Summary*, AJ, 120, 1579

Zivick, P., Sutter, P. M., Wandelt, B. D., Li, B., & Lam, T. Y. 2015, *Using cosmic voids to distinguish f(R) gravity in future galaxy surveys*, MNRAS, 451, 4215

SUMMARY IN ESTONIAN

Universumi kärgstruktuuri kaardistamine Bisous mudeliga

Kosmoloogilised vaatlusprojektid on näidanud, et galaktikate jaotus universumis moodustab keeruka kärgstruktuuri. Selle kärgstruktuuri omaduste uurimine annab võimaluse kontrollida kosmoloogiliste mudelite korrektsust. Näiteks mõjutavad tumeaine ja gravitatsiooni omadused, milline kärgstruktuur tekib. Tehes arvutisimulatsioone saame välistada selliste parameetrite ja omadustega mudeleid, mis ei suuda taasluua struktuure, mida nähakse vaatlustest. Kärgstruktuur koosneb galaktikaparvedest, neid ühendavatest galaktikate ahelatest ehk filamentidest, mis moodustavad kärgstruktuurile omase võrgustiku, ja suurtest peaaegu tühjadest aladest ehk tühimikest. Filamendid hõlmavad enamust universumis olevast massist, kogudes seda ümbritsevast väiksema tihedusega aladelt ja transportides seda võrgustiku sõlmpunktidesse — galaktikaparvedesse. Kärgstruktuuri dünaamika ja mitmekesisus avaldavad mõju ka galaktikate arengule. Paljud galaktika omadused, nagu näiteks nende mass, struktuur, gaasirikkus, asend ruumis ja tähetekke intensiivsus, sõltuvad nende kosmoloogilisest keskkonnast, mille moodustavad kärgstruktuuri filamendid. Võimekus tuvastada ja uurida filamente on tähtis nii galaktika suurusskaaladel toimuvate protsesside kirjeldamiseks ja mõistmiseks kui ka universumit kirjeldava kosmoloogilise mudeli täiendamiseks ja kontrollimiseks.

Bisous meetod on välja töötatud filamentide võrgustiku mudeldamiseks vaatlusandmete põhjal. See meetod kasutab punktprotsesse ja Bayesi statistikat võrgustiku mudeli konstrueerimiseks. Kärgstruktuuri on vaatluste abil keeruline tuvastada, sest see on põhiliselt tumeaine struktuur ja kuna tumeaine ei kiirga ega neela valgust, siis saab filamente tuvastada ainult galaktikate või gaasi jaotuse järgi. Bisous mudel otsib tumeaine võrgustikku galaktikate jaotuse põhjal, ehitades ahelaid ehk filamente üles lülihaaval. Lülideks on silindrilised objektid, mis otsapidi ühendatult moodustavad pikad filamendid. Lülide asetust mudelis optimeeritakse Monte Carlo algoritmiga, et lülid järgiks galaktikate jaotust ja samas moodustaksid pikki ühendatud ahelaid. Selle tulemuseks on omavahel ühendatud kosmiline võrgustik, mis jäljendab tumeaine kärgstruktuuri. Käesoleva doktoritöö eesmärkideks on testida, karakteriseerida ja täiendada Bisous mudelit, et kvantitatiivselt kirjeldada mudeli usaldusväärsust ja täpsust ning tõsta mudeli efektiivsust.

Kuna Bisous mudel on stohhastiline ehk kasutab juhuslikkust, on oluline uurida tulemuste varieeruvust, et iseloomustada mudeli enda statistilist määramatust. See määramatus tuleneb praktilisest piirangust, et simulatsioone ei ole võimalik jooksutada kuni matemaatilise koondumiseni, mistõttu eri simulatsioonide

tulemused erinevad üksteisest. Seda määramatust annab vähendada sellega, et filamentide võrgustikku simuleeritakse palju kordi ja mudeli väljundiks on nende simulatsioonide keskmine. Keskmistamine minimeerib tulemuste Poissoni müra ja vähendab varieeruvust. Varieeruvuse uurimiseks mudelleerisin kärgstruktuuri 200 korda samadel andmetel ja tulemuste võrdlus näitas väga tugevat korrelatsiooni ($R = 0{,}98$). Tugev korrelatsioon näitab, et mudeli väljund on praktiliselt koondunud, stohhastilistest meetoditest tulenev müra on minimaalne ja tulemused samade andmete ja parameetritega on väga väikese varieeruvusega.

Spektroskoopilistel vaatlusprojektidel on tavaliselt määratud mingi alampiir, millest väiksema näiva heledusega objekte ei vaadelda, sest piisava kvaliteediga spektrite mõõtmiseks kulub ebamõistlikult palju vaatlusaega. Kuna objekti heledus väheneb kaugusega, siis ka selliste vaatlusprojektide vaadeldud objektide tihedus väheneb kaugusega, sest kauguse suurenedes nende objektide heledused kahanevad alla määratud alampiiri. Seega kaugemal on näha ainult suure absoluutse heledusega objekte. See seos on Bisous mudeli seisukohalt tähtis, sest mudeli tulemused sõltuvad galaktikate arvtihedusest sisendandmetes — mida vähem on vaatlustes tuvastatud galaktikaid, seda keerulisem on leida filamente ja seeläbi ka mudelleerida tumeaine kosmilist võrgustikku. Näiteks SDSS galaktikate kataloogis langeb detekteeritud filamentide arvtihedus oluliselt paarisaja megaparseki kaugusel, kuigi kõige kaugemad tuvastatud galaktikad on kordades kaugemal.

Erinevate vaatluste põhjal saadud Bisous mudeli tulemuste tõlgendamiseks ja võrdlemiseks on tähtis teada, kuidas filamentide arvtihedus sõltub sisendandmete galaktikate arvtihedusest. Selleks, et seda sõltuvust uurida, võtsin aluseks tumeaine simulatsiooni ja tekitasin sellest kaksteist andmestikku, millel kõigil on sama tumeaine kärgstruktuur, aga erinevad valimid galaktikatest, et uurida erinevaid arvtihedusi. Nende põhjal arvutasin sõltuvuse galaktikate arvtihedusest kahele statistikule — filamentidega täidetud ruumi osakaal ja galaktikate osakaal, mis asub filamentides. Need statistikud kirjeldavad tuvastatud filamentaarse võrgustiku täielikkust. Analüüs näitas, et kuigi väiksema galaktikate arvtiheduse korral suudab mudel tuvastada vähem filamente, on nende seas väga vähe valepositiivseid filamente, mida täielikumate andmete korral siiski ei tuvastata filamentidena. See tähendab, et kui suurematel kaugustel ei suuda vaatlused tuvastada suuremat osa galaktikatest ja andmete arvtihedus on väike, siis ka Bisous filamendid on ebatäielikud ehk ei kajasta täielikult tumeaine jaotust, kuid sisaldavad vähe valepositiivseid filamente. See on filamentide tuvastamisel väga kasulik omadus ja suurendab tulemuste usaldusväärsust.

Selline sõltuvus galaktikate arvtihedusest viib küsimuseni, kas ja kuidas me võiksime tihedust millegi arvelt suurendada. Tavaliselt kasutatakse filamentide tuvastamisks ainult spektroskoopiliste punanihetega andmeid, sest kauguse hinnangud on mitu korda paremad kui fotomeetriliste punanihetega andmete puhul ja ilma täpse kauguse hinnanguta on väga keeruline kolmemõõtmelist

struktuuri uurida. Teisest küljest on fotomeetriliste punanihetega andmeid sadu kordi rohkem kui spektroskoopilistega. Fotomeetriliste punanihetega andmete kasutuskõlblikkuse katsetamiseks rakendasin ja uurisin lihtsakoelist meetodit, mis hõlbustab suure kauguse määramatusega andmete kasutamist. Nagu eeldatud, näitavad tulemused mitmeid puudujääke selliste andmete kasutamise puhul. Valepositiivsete tuvastuste sagedus (*false discovery rate*) ja tõeliste filamentide tuvastamisvõime (*recall*) analüüs näitab, et kuigi valepositiivsete tulemuste osakaalud ei sõltu eriti kauguse määramatuse suurusest ja on alati väikesed, siis mida suurem on kauguse määramatus, seda vähem suudab mudel tõelisi filamente tuvastada. Suurema määramatuse korral mudel lihtsalt väljastab vähem filamente. Lisaks suureneb filamentide telgede ja vaatesihi vaheline korrelatsioon sisendandmete suurema kauguse määramatuse korral. See on mudeli arvutuslik iseärasus, mitte füüsikaline omadus, sest vaatesihid ja filamentide teljed ei tohiks olla omavahel korreleeritud. Kui kasutada fotomeetrilisi punanihke andmeid, siis suudab Bisous tuvastada ainult kõige tihedamini asustatud filamente piki vaatesihti.

Selle asemel, et kasutada ainult fotomeetriliste punanihetega andmeid, saame kasutada mõlemaid fotomeetriliste ja spektroskoopiliste punanihetega andmeid korraga ehk täiendatud andmeid. See tähendab, et kasutaksime andmeid taevaalade kohta, mis on kaetud nii fotomeetriliste kui ka spektroskoopiliste vaatlustega. Analüüs näitab, et täiendatud andmete kasutamine parandab oluliselt filamentide tuvastamist, ei suurenda märgatavalt valepositiivseid tulemusi ning vähendab vaatesihi ja filamentide telgede vahelist korrelatsiooni. Lisaks näitavad täiendatud andmed paremaid tulemusi kui ainult spektroskoopiliste punanihetega andmed. Järelikult on kasulik kasutada täiendatud andmeid, siis kui spektroskoopiliste punanihetega andmete arvtihedus ja filamentide avastamise efektiivsus on väikesed. Bisous mudeli tulemuste parandamiseks on siiski vaja suhteliselt väikeste määramatustega kauguseid ja praegused fotomeetrilised vaatlused ei suuda selliseid määramatusi saavutada. Õnneks peagi alustav J-PASi uuring võimaldab arvutada fotomeetrilisi punanihkeid sobiva määramatusega. J-PASi saaks ühendada juba olemasolevate SDSSi või DESI BGSi vaatlustega, et saada käesolevas töös kasutatud andmetega sarnaseid täiendatud andmestikke.

Tulevased vaatlusprojektid, mis katavad suuremaid taevaalasid ja koguvad andmeid suurema galaktikate arvtihedusega, nagu peagi alustav 4MOST, suurendavad oluliselt kärgstruktuuri uurimise võimalusi. Märgatavalt paremad andmestikud toovad kindlasti kaasa uusi avastusi, eriti uurimissuundades, mida praegu vaevavad suured statistilised määramatused, mis on tingitud liiga vähestest andmepunktidest. Ka Bisous mudel oleks praegusel kujul nende uute andmestikega töötades ebaefektiivne ja vajab täiendamist, et rakendada uutele andmestikele tõhusamaid lähenemisi. Käesolevas töös esitatud meetod fotomeetriliste punanihetega andmete kasutamiseks on kõigest esialgne katsetus, mis näitas positiivseid tulemusi. Selleks, et Bisous mudelit päris fotomeetrilistel

vaatlusandmetel rakendada, on samuti mõistlik meetodit uuendada ja tõhustada. Kuigi need uuendused ei mahtunud käesolevasse doktoritöösse, on need siiski plaanis lähitulevikus ellu viia. Seega on Bisous mudeli arendamine plaanis kahel rindel: suure arvtihedusega andmestike tõhusam kasutamine ning kärgstruktuuri mudeldamine fotomeetriliste punanihetega andmete abil.

ACKNOWLEDGEMENTS

It has been an adventure with many ups and downs. I am immensely grateful to my supervisor Elmo, who has been a mentor and a role model on my path to becoming a scientist. Your attention to detail, and deep understanding and knowledge of cosmology, observations, and so much more have been an immense asset on this journey. I feel lucky to have had the chance to study and practice science under your supervision.

The whole cosmology department at Tartu Observatory has been an important part of this adventure. Your kindness and helpfulness have supported me on the way, and the environment in and around the observatory is nothing but encouraging. Thank you, Taavi, Rain, Antti, María, Shishir, Juhan, Peeter, Indrek, Jukka, Jaan, Jaan, Maret, Enn, and Teet. Special thanks to the younger staff for journeying with me and providing guidance and company, namely Punya, Toni, Maarja, Heleri, Rando, Madis, Joosep, Kate, Mousumi, and Jad.

I can't leave out my biggest supporters, without whom this all would have been impossible — institutional research funding PRG1006 of the Estonian Ministry of Education and Research, Centre of Excellence "Dark Side of the Universe" (TK133), and National Science Foundation under Grants No. NSF PHY-1748958 and PHY-2309135. Thank you for your resourcefulness. Being able to participate in numerous conferences and meetings has definitely benefitted my career. Special thanks to the organizers of the Tartu-Tuorla meetings throughout the years. In addition, I would like to mention the scientific saunas in Tõravere and the hackathons with the productive and fun discussions. I hope these traditions will continue.

Thank you to my friends from high school and university, Karl, Henri, Tambet, Andri, Jaanus, Martin, Markus, Gutnar, Grete, Venno, Rauno, and Rimmo, for reminding me to relax and have fun on a regular basis. I hope the dungeoneering, water poloing, chessing, and hiking never ends. You have been invaluable companions, and I try to return the favor.

I want to thank my parents, Kai and Toomas, without whom I couldn't have even started. My sisters and brothers, Miina, Laurits, Marta, Morten, and also Maarika and Jane-Liis, for always being ready to support me and my family. And lastly, my wife Merilyn and children Magnus, Märten, and Maia, without whom this thesis would have been finished much earlier. I love you, and thank you for coming on the adventure with me (even though most of you didn't have a choice).

PUBLICATIONS

CURRICULUM VITAE

Personal data

Name	Moorits Mihkel Muru
Date and place of birth	July 25, 1993, Tallinn, Estonia
Citizenship	Estonian
Current employment	UT Tartu Observatory
Address	University of Tartu, Tartu Observatory Observatooriumi 1, Tõravere 61602, Tartu county, Estonia
Phone	(+372) 5303 5647
E-mail	moorits.mihkel.muru@ut.ee

Education

2000 – 2009	Tartu Descartes Lyceum
2009 – 2012	Hugo Treffner Gymnasium
2012 – 2016	University of Tartu, undergraduate student, BSc 2016 (physics)
2016 – 2018	University of Tartu, graduate student, MSc 2018 (physics)
2018 – 2023	University of Tartu, PhD student

Employment

2014 – 2018	Tartu Observatory, Technician
2018 – 2018	University of Tartu, Tartu Observatory, Technician
2019 – . . .	University of Tartu, Tartu Observatory, Junior Research Fellow in Astronomy

Professional training

2015	Summer school "Introduction to Cosmology" Kielce, Poland, 10 days
2016	Winter school "Nordic Optical Telescope school" Tuorla, Finland, 10 days

2019	Summer school "CSC Summer School in High-performance computing" Solvalla, Finland, 10 days
2021	Virtual summer school "Statistics for Astronomers" Center for Astrostatistics at the Pennsylvania State University, 5 days
2023	Science program "The Cosmic Web: Connecting Galaxies to Cosmology at High and Low Redshift" Kavli Institute for Theoretical Physics (KITP), UC Santa Barbara, California, USA, 4 weeks

Conference presentations

2018	Conference "The large scale properties of the universe as a whole" Tuorla, Finland, *Oral presentation*: "Looking for WHIM in SDSS"
2019	Conference "Einasto's Profile" Tartu, Estonia, *Oral presentation*: "Testing the limitations of Bisous model"
2020	Orsay – Tartu Workshop Online, *Oral presentation*: "The Bisous filament finder systematics"
2021	Conference "Interaction of the cosmic matter" Turku, Finland, *Oral presentation*: "Could photometric data be useful for the detection of large-scale structure?"

2021	Conference "IAU Symposium 362: The predictive power of computational astrophysics as a discovery tool" Online, *Poster presentation*: "Detecting cosmic filamentary network with stochastic Bisous model"
2022	Conference "Galaxy Dynamics and Beyond" Laulasmaa, Estonia, *Oral presentation*: "Filaments in J-PAS data"
2022	CLUES collaboration meeting Madrid, Spain, *Oral presentation*: "Detecting galactic filaments from observations"
2022	Conference "21st Annual Conference of the International Association for Mathematical Geosciences" Nancy, France, *Poster presentation*: "Detecting large-scale structure of the Universe using photometric data with the Bisous model"
2023	CLUES collaboration meeting Munich, Germany, *Oral presentation*: "Improved grouping algorithm and group mass estimation"
2023	EAS annual meeting Kraków, Poland, *Oral presentation*: "Using photometric redshifts to boost the filament detection with Bisous model"

Language skills

Estonian	native
English	fluent
French	basic

Honours and Awards

2014	Helle Randam scholarship (Dharma)
2015	Rotalia Foundation scholarship
2017	Estonian Physics Society's award
2020	E. Öpik scholarship (Tartu Observatory)

Fields of research

Large scale structure of the universe,
cosmic filaments in observational data.

Publications

1. **Muru, M. M.** (2023). *Detecting cosmic filamentary network with stochastic Bisous model*. Proceedings IAU Symposium, 16 (S362), 54-55
2. **Muru, M. M.**; Tempel, E. (2023). *Using photometric redshift data to improve the detection of galactic filaments with the Bisous model*. Astronomy & Astrophysics, 670, A77
3. Laur, J.; Tempel, E.; Tamm, A.; Kipper, R.; Liivamägi, L. J.; Hernán-Caballero, A.; **Muru, M. M.**; Chaves-Montero, J.; Díaz-García, L. A.; Turner, S.; Tuvikene, T.; Queiroz, C.; Bom, C. R.; Fernández-Ontiveros, J. A.; González Delgado, R. M.; Civera, T.; Abramo, R.; Alcaniz, J.; Benítez, N.; Bonoli, S. ... Ramió, H. V. (2022). *TOPz: Photometric redshifts for J-PAS*. Astronomy & Astrophysics, 668, A8
4. **Muru, M. M.**; Tempel, Elmo (2021). *Assessing the reliability of the Bisous filament finder*. Astronomy & Astrophysics, 649, A108
5. Tempel, E.; Tuvikene, T.; **Muru, M. M.**; Stoica, R. S.; Bensby, T.; Chiappini, C.; Christlieb, N.; Cioni, M.-R. L.; Comparat, J.; Feltzing, S.; Hook, I.; Koch, A.; Kordopatis, G.; Krumpe, M.; Loveday, J.; Minchev, I.; Norberg, P.; Roukema, B. F.; Sorce, J. G.; Storm, J.; Swann, E.; Taylor, E. N.; Traven, G.; Walcher, C. J.; de Jong, R. S. (2020). *An optimized tiling pattern for multiobject spectroscopic surveys: application to the 4MOST survey*. Monthly Notices of the Royal Astronomical Society, 497, 4626-4643

ELULOOKIRJELDUS

Isikuandmed

Nimi	Moorits Mihkel Muru
Sünniaeg ja -koht	25. juuli 1993, Tallinn, Eesti
Kodakondsus	Eesti
Praegune töökoht	TÜ Tartu observatoorium
Aadress	Tartu Ülikooli Tartu observatoorium Observatooriumi 1, Tõravere 61602, Tartumaa
Telefon	(+372) 5303 5647
E-mail	moorits.mihkel.muru@ut.ee

Haridus

2000 – 2009	Tartu Descartes'i Lütseum
2009 – 2012	Hugo Treffneri Gümnaasium
2012 – 2016	Tartu Ülikool, üliõpilane, BSc (füüsika) 2016
2016 – 2018	Tartu Ülikool, magistrant, MSc (füüsika) 2018
2018 – 2023	Tartu Ülikool, doktorant

Teenistuskäik

2014 – 2018	Tartu observatoorium, tehnik
2018 – 2018	Tartu Ülikool, Tartu observatoorium, tehnik
2019 – . . .	Tartu Ülikool, Tartu observatoorium, astronoomia nooremteadur

Täiendkoolitus

2015	Suvekool „Introduction to Cosmology“ Kielce, Poola, 10 päeva
2016	Talvekool „Nordic Optical Telescope school“ Tuorla, Soome, 10 päeva

2019	Suvekool „CSC Summer School in High-performance computing“ Solvalla, Soome, 10 päeva
2021	Virtuaalne suvekool „Statistics for Astronomers“ Center for Astrostatistics at the Pennsylvania State University, 5 päeva
2023	Teadusprogramm „The Cosmic Web: Connecting Galaxies to Cosmology at High and Low Redshift“ Kavli Institute for Theoretical Physics (KITP), UC Santa Barbara, California, USA, 4 nädalat

Konverentside ettekanded

2018	Konverents „The large scale properties of the universe as a whole“ Tuorla, Soome, *Suuline ettekanne*: „Looking for WHIM in SDSS“
2019	Konverents „Einasto's Profile“ Tartu, Eesti, *Suuline ettekanne*: „Testing the limitations of Bisous model“
2020	Orsay – Tartu kohtumine Veebipõhine, *Suuline ettekanne*: „The Bisous filament finder systematics“
2021	Konverents „Interaction of the cosmic matter“ Turu, Soome, *Suuline ettekanne*: „Could photometric data be useful for the detection of large-scale structure?“

2021	Konverents „"IAU Symposium 362: The predictive power of computational astrophysics as a discovery tool“ Veebipõhine, *Posterettekanne*: „Detecting cosmic filamentary network with stochastic Bisous model“
2022	Konverents „Galaxy Dynamics and Beyond“ Laulasmaa, Eesti, *Suuline ettekanne*: „Filaments in J-PAS data“
2022	CLUES collaboration meeting Madrid, Hispaania, *Suuline ettekanne*: „Detecting galactic filaments from observations“
2022	Konverents „21st Annual Conference of the International Association for Mathematical Geosciences“ Nancy, Prantsusmaa, *Posterettekanne*: „Detecting large-scale structure of the Universe using photometric data with the Bisous model“
2023	CLUES collaboration meeting München, Saksamaa, *Suuline ettekanne*: „Improved grouping algorithm and group mass estimation“
2023	EAS annual meeting Kraków, Poola, *Suuline ettekanne*: „Using photometric redshifts to boost the filament detection with Bisous model“

Keelteoskus

eesti keel	emakeel
inglise keel	väga hea
prantuse keel	algtase

Autasud ja stipendiumid

2014	Helle Randami stipendium (Dharma)
2015	Rotalia Foundation stipendium
2017	Eesti Füüsika Seltsi tunnustus
2020	E. Öpiku stipendium (Tartu observatoorium)

Peamised uurimissuunad

Universumi suuremastaabiline kärgstruktuur,
galaktilised filamendid vaatlusandmestikes.

DISSERTATIONES ASTRONOMIAE UNIVERSITATIS TARTUENSIS

1. **Tõnu Viik.** Numerical realizations of analytical methods in theory of radiative transfer. Tartu, 1991.
2. **Enn Saar.** Geometry of the large scale structure of the Universe. Tartu, 1991.
3. **Maret Einasto.** Morphological and luminosity segregation of galaxies. Tartu, 1991.
4. **Urmas Haud.** Dark Matter in galaxies. Tartu, 1991.
5. **Eugene A. Ustinov.** Inverse problems of radiative transfer in sounding of planetary atmospheres. Tartu, 1992.
6. **Peeter Tenjes.** Models of regular galaxies. Tartu, 1993.
7. **Ivar Suisalu.** Simulation of the evolution of large scale structure elements with adaptive multigrid method. Tartu, 1995.
8. **Teimuraz Shvelidze.** Automated quantitative spectral classification of stars by means of objective prism spectra: the method and applications. Tartu, 1999.
9. **Jelena Gerškevitš.** Formation and evolution of binary systems with compact objects. Tartu, 2002.
10. **Ivan Suhhonenko**. Large-scale motions in the universe. Tartu, 2003.
11. **Antti Tamm.** Structure of distant disk galaxies. Tartu, 2006.
12. **Vladislav-Veniamin Pustynski.** Modeling the reflection effect in pre-cataclysmic binary systems. Tartu, 2007.
13. **Anna Aret.** Evolutionary separation of mercury isotopes in atmospheres of chemically peculiar stars. Tartu, 2009.
14. **Mari Burmeister.** Characteristics of the hot components of symbiotic stars. Tartu, 2010.
15. **Elmo Tempel.** Tracing galaxy evolution by their present-day luminosity function. Tartu, 2011.
16. **Anti Hirv.** Estimation of time delays from light curves of gravitationally lensed quasars. Tartu, 2011.
17. **Rain Kipper**. Galaxy modelling: dynamical methods and applications. Tartu, 2016, 134 p.
18. **Lauri Juhan Liivamägi**. Properties and spatial distribution of galaxy superclusters. Tartu, 2017, 185 p.
19. **Jaan Laur**. Variability survey of massive stars in Milky Way star clusters. Tartu, 2017, 183 p.
20. **Boris Zhivkov Deshev.** On the coevolution of galaxies and their host clusters. Tartu, 2019, 199 p.
21. **Tiina Liimets**. Nebulosities and jets from outbursting evolved stars. Tartu, 2019, 207 p.
22. **Maarja Bussov**. Clustering Analysis for Astrophysical Structures. Tartu, 2020, 113 p.

23. **Teet Kuutma**. Galaxies and non-galactic baryons in cosmic filaments. Tartu, 2021, 123 p.
24. **Toni Tuominen**. Missing Baryons and the Large-Scale Structure of the Universe. Tartu, 2023, 121 p.