Polyphorm: Structural Analysis of Cosmological Datasets via Interactive Physarum Polycephalum Visualization

Oskar Elek, Joseph N. Burchett, J. Xavier Prochaska, and Angus G. Forbes

Abstract—This paper introduces Polyphorm, an interactive visualization and model fitting tool that provides a novel approach for investigating cosmological datasets. Through a fast computational simulation method inspired by the behavior of Physarum polycephalum, an unicellular slime mold organism that efficiently forages for nutrients, astrophysicists are able to extrapolate from sparse datasets, such as galaxy maps archived in the Sloan Digital Sky Survey, and then use these extrapolations to inform analyses of a wide range of other data, such as spectroscopic observations captured by the Hubble Space Telescope. Researchers can interactively update the simulation by adjusting model parameters, and then investigate the resulting visual output to form hypotheses about the data. We describe details of Polyphorm’s simulation model and its interaction and visualization modalities, and we evaluate Polyphorm through three scientific use cases that demonstrate the effectiveness of our approach.

Index Terms—Astrophysics visualization, agent-based modeling, intergalactic media, Physarum polycephalum, Cosmic Web.

1 INTRODUCTION

Composed primarily of hot gas and plasma, the intergalactic medium (IGM) is theorized to contain a significant amount of the matter in the Universe that has so far eluded observation. Astrophysicists refer to this as “the missing baryon problem” [25], and the detection and localization of IGM remains an ongoing research effort. The best evidence that the IGM contains the unaccounted baryonic matter is provided by large-scale cosmological simulations, such as the Millenium [68], EAGLE [62] and Bolshoi-Planck [42] simulations. These simulations also unanimously agree on the expected distribution of the IGM: after billions of years, the combined effect of gravitational attraction and dark-energetic repulsion has shaped the universe as a vast network called the Cosmic Web. This quasi-fractal structure [65] consists of knots—galaxies and their clusters—interconnected by a complex network of filaments and separated by vast cells of empty space called voids. In addition, the entire structure is underlined by a scaffolding of dark matter. These simulations have been constructed to confirm the earlier findings of the Zeldovich framework [22, 33, 78], which on theoretical grounds predicts the emergence of these very structures—knots, filaments, voids—under the influence of the dominant force dynamics in the Universe.

Given the indirect evidence for its existence, astrophysicists are interested in determining the structure of the Cosmic Web. The intergalactic filaments consist primarily of diffuse non-luminous matter, and are extremely challenging to observe directly. Both topological [6] and geometric [19, 71] approaches using graphs as a core representation to
model the Cosmic Web have been proposed. However, most of the existing approaches focus on cataloging and classifying the filaments, and no method described to date has been able to sufficiently recover the full 3D density map of the intergalactic filaments. Given that the Cosmic Web is a heterogeneous multi-scale structure without an intrinsic topology, having an accurate estimate for its density distribution would benefit many tasks, such as a better understanding the transition from the galactic to intergalactic environments. We capture these trajectories as a density field in 3D space, interpreting would benefit many tasks, such as a better understanding the transition from the galactic to intergalactic environments. A schematic providing a high-level overview of the filamentary structure of the Cosmic Web is shown in Fig. 2.

In this paper, we introduce Monte Carlo Physarum Machine (MCPM), an agent-based virtual Physarum machine that algorithmically mimics the organism’s growth and is able to construct precise trajectories, and the Polyphorm interactive visualization software, which facilitates the interactive analysis of large cosmological datasets using MCPM. Relying on Physarum's ability to approximate optimal transport networks, we “feed” to its virtual counterpart galaxies and dark matter halos as 3D point attractants. The MCPM agents navigate these input points to find most probable paths between them, and the total aggregate of their trajectories then forms the filamentary network. We capture these trajectories as a density field in 3D space, interpreting it as a proxy of the intergalactic gas distribution that provides an estimate for the Cosmic Web. Using Polyphorm, astrophysicists are able to gain new insights into the composition and structure of the Universe.

The contributions of this paper are: a) an exposition of the MCPM model, b) an introduction to the Polyphorm software and a discussion of the visualization techniques that enable the application of MCPM to cosmological datasets and the analysis of the simulation outputs, and c) a series of scientific use cases that demonstrate the utility of Polyphorm in predicting the location and density of IGM filaments. In Sec. 2, we contextualize this work in terms of previous related work in both the visualization community and the astrophysics community. In Sec. 3, we present a taxonomy of the primary structural and visual analysis tasks enabled through our approach, as well as the design requirements for our software implementation. In Sec. 4, we present details of the MCPM, focusing on innovation in the agent propagation step, which is defined in terms of different probability distributions. In Sec. 5, we provide a summary of the visual analysis features available in the Polyphorm software. In Sec. 6, we evaluate our methods by demonstrating how the use of Polyphorm facilitates the investigation of the filamentary structure of the Cosmic Web. A schematic providing a high-level overview of Polyphorm is shown in Fig. 2.

2 RELATED WORKS

Cosmological simulations based on the cold dark matter paradigm predict that matter in the Universe is organized into a Cosmic Web, a series of interconnected filaments that are formed from dark matter comprised of low density IGM gas as well as galaxies and the circum-galactic medium (CGM) gas found within dark matter halos adjacent to these galaxies. In regions of the Universe nearer to the Milky Way, the distribution of galaxies mapped by large surveys visually indicate the Cosmic Web structure, and a study by Wakker et al. uses measurements from the Hubble Space Telescope to investigate one visually identified Cosmic Web filament and its imprint on H I absorption lines.

These cosmological simulations attempt to model physical processes spanning several orders of magnitude in scale and use empirical observations to validate the models' accuracy. While the models are tuned to match certain aspects of the observed galaxy population, other observable quantities may serve as pure predictions (consequences of the physics being tuned to match certain data). Thus, as new measurements become available, astrophysicists can refine the simulations, leading to a better understanding of the physical processes in galaxy formation. For example, the EAGLE project presents hydrodynamical simulations that model the formation and co-evolution of galaxies and black holes with the largest scale structures of the Universe. The FIRE project also produces simulations of galaxy formation, but focuses on resolving the interstellar medium of individual galaxies. The Hardware/Hybrid Accelerated Cosmology Code (HACC) framework includes models of baryonic matter as well as active galactic nuclei associated with violent bursts of energy from supermassive black holes, and has been used to illustrate how this energy is imparted to the surrounding gas and affects subsequent structure formation.

Kapferer and Riser articulate the importance of interactive visualization to help researchers understand the coherences in cosmological simulations, and a range of frameworks have been developed to support astrophysics visualization. For example, Woodring et al. introduce new analysis features implemented within ParaView to analyze simulation data, and the Aladin Sky Atlas enables users to add annotation markers to observational image data catalogs. More recently, the Open Space “astrographics” system facilitates the interactive display of integrated data from multiple sources, supporting research and science communication. A number of tools present celestial objects as volumes within a 3D view, and Pomarède et al. make use of images, videos, and derived isosurface structures within a 3D representation to show galaxy position, density fields, gravitational potential, and velocity shear tensors. Visual analytics software has been developed to explore simulation data that models the evolution of the Universe. For example, Almyrde and Forbes introduce an interactive web application to visualize “traces” of dark matter halos, and an application developed by Scherzinger et al. provides 2D and 3D views to support the analysis of halo substructures and hierarchies.

To the best of our knowledge, no previous techniques explore cosmological data using a virtual Physarum machine. However, there are a range of works that utilize models of Physarum polycephalum to make sense of complex systems. Efficient virtual Physarum machines have been developed that utilize cellular automata and formal logic, and an approach introduced by Jones uses a large number of agents to generate emergent structures. This latter approach is similar in some ways to agent-based models that model crowds and flocks of birds, but differs in that it uses a continuous representation of the agents density for navigation. In addition to the successful application to “classic” problems such as road network optimization, the agent-based computational approach is suitable for abstract and interactive tasks, such as B-spline fitting, convex and concave hull approximation, and robot navigation. The design of Polyphorm is also inspired by an expressive model — itself based on Jones’ pattern formation explorations — created by the artist Sage Jenson as a means for producing intriguing, complex animations.

The structure-finding approach of Polyphorm provides a coupling between the reconstruction of the 3D density field and its visualization, enabling researchers to interact with the model both visually and
through numerical measures simultaneously. Since the underlying model (see Sec. 3) runs interactively on a desktop computer with a dedicated GPU, the fitting process takes at most a few minutes, allowing for interventions, whereas previous filament finding methods can take significantly longer and require access to high performance computing systems. Contrary to approaches that produce either graph-based reconstructions or classification masks [43], Polyphorm constructs a continuous network-like geometric structure in which densities are defined at every point in the 3D volume, and where the derived density field may be validated visually against the input data distribution and weights. The availability of such density data greatly expands the range of questions we can now ask about the intergalactic medium [15, 16].

3 TASK TAXONOMY

Based on our review of the literature and on extensive conversations with astrophysicists researching intergalactic environments and interested in understanding the structure and composition of the Cosmic Web, we define the main structural analysis tasks that our model supports. Additionally, to facilitate structural analysis, we further define a set of visual analysis tasks that support the investigation and interpretation of the output generated via MPCM.

3.1 Structural Analysis Tasks

S1: Generate density fields. A core task for astrophysicists investigating dark matter filaments and the intergalactic medium that permeates them is to gain insight into the structure of the Cosmic Web. The primary output of Polyphorm is a 3D density field that corresponds to (or “reconstructs”) the input data—a set of points in 3D Euclidean space, each optionally assigned a weight—in a meaningful way. In the astronomical context, that conforms to the rules of optimal transport. While there are many approaches to finding higher-order structures in point data, many of them are based on discrete, graph-like representations [10]. Though these representations enable easy navigation and organization, they can obscure important aspects of some astrophysics datasets.

In the scientific use cases described in Sec. 3.1 a key dataset is the Bolshoi-Planck simulation, which explores dark matter dynamics in a cubic region of roughly 200 Mpc in size. (For reference, the Milky Way is 0.03 Mpc in diameter). Out of this massive dataset, a set of about 2M dark matter halos has been extracted and are used as the basis for our fitting. The presence of ground-truth densities enables the calculated output in our model in terms of mapping to the canonical “cosmic overdensity” quantity. The other key dataset is the catalog of observed galaxies mapped by the Sloan Digital Sky Survey (SDSS). Querying observations from the range of cosmic redshifts $z = 0.0138 - 0.0318$ yields approximately 37k galaxies, again suitably represented by 3D weighted points across the 200 Mpc region of interest. In contrast to the simulated data, the observations are known to be incomplete, and lack a ground-truth density map. By generating reasonable density fields, astrophysicists can better understand the structure of the Cosmic Web.

S2: Identify anisotropic structures. Once an estimated density field is available, astrophysicists can use it to identify anisotropic structures of the Cosmic Web. Such structures are expected to correspond to the dark matter filaments and should in turn comprise the gaseous filaments of the intergalactic medium. The presence of these structures is predicted by simulations [43, 63], resulting mainly from the gravitational collapse of matter on the Universal timescales. Unfortunately, observational evidence of these structures is extremely challenging to obtain due to the low particle density of these filamentary structures. One notable exception are spectroscopic measurements parametrized by redshift (distance from Earth), which systematically show absorption signatures even in regions of space where galactic halos and other visible agglomerations of matter are not present. The key ingredient in finding these structures is a model capable of consistently reconstructing such structures from sparse positional data.

S3: Detect features across different scales. A task related to S2 is detecting geometric features across several orders of magnitude, in terms of both spatial scale and density. Estimations of the scale of the Cosmic Web filaments range between 1 and 100 Mpc. Smaller filaments contain less matter, so the reconstructing model needs to generate a density field with several orders in dynamic range. This enables researchers to simulate both dense and sparse regions of the Universe.

S4: Support structural transfer between relevant datasets. Another important task is supporting structural transfer between input point clouds. The idea is to use structures detected in one dataset as a prior estimator for the model parameters when fitting to another dataset. This task becomes useful with the availability of simulated data, as structures detected in a dense simulated dataset (especially with ground-truth density available) can bootstrap the fitting to observation data. Doing so is expected to impact the robustness of the model when fitting to potentially incomplete data, as well as the confidence in such a fit.

3.2 Visual Analysis Tasks

Complementing the structural analysis tasks, we identify a number of important tasks that support the visual analysis of the model. These tasks facilitate the understanding of the input data distribution, as well as provide diagnostics of the Cosmic Web estimate, including its global and local structural properties, alignment with the input data, and allocation of matter density. Visual feedback is especially important here, since the solution to the Cosmic Web reconstruction problem is heavily under-determined. That is, we cannot rely only on numerical fitting, since the loss function that would define the fitness of the solution is not initially known.

V1: Interactive exploration of the input data and MPCM simulation. The core task here is to visualize both the input data and the resulting Cosmic Web estimate. Interacting with the 3D rendering is necessary due to the supervised nature of the reconstruction methodology, and researchers need to be able to freely rotate, pan, and zoom into the data and the solution. An essential part of this task is the ability to slice the 3D data, given the presence of many overlapping filaments and the general complexity of the estimated structures.

V2: Interactive tuning of the model parameters. Another requirement for supervised fitting is to expose the model parameters to the researcher. By tuning them, changes to the estimated solution can be made interactively. In conjunction with V1, this makes it possible to carry out a single fitting session within minutes. Providing an integrated feedback loop enables the researcher to quickly iterate on the solution, and accelerates the subsequent analysis and validation. Naturally, the employed reconstruction model has to be responsive enough to support this.

V3: Provide multiple rendering modalities. Both point-based rendering and volume rendering modalities are necessary to analyze cosmological datasets and to support filament reconstruction. Point-based rendering is used to visualize the raw input data (a set of points in 3D space) as well as the discrete agents of the MPCM reconstruction algorithm (described in Sec. 3.1). Volume rendering supports a range of different views, including:

- the 3D volumetric footprint of the input data, essentially a kernel point-density estimate, or deposit, used to show the overall impact of the inputs, including the impact of their weights (i.e., masses);
- a 3D density of the reconstructed estimate of the Cosmic Web itself, or trace, used to examine the geometric structures and density distribution of the solution without the visual bias of the input data;
- a combined 3D view of the deposit+trace, used to visually examine the alignment between them and to check for any data not covered by the estimate.

V4: Statistical summarization of the model fit. Alongside the visual analysis of the Cosmic Web reconstruction, researchers find it useful to view a concise set of statistics summarizing the solution. The quantities that have proven to be most valuable when determining the quality
of the solution are: the overall energy (fitness) of the reconstruction and its temporal evolution during a fitting session; a density histogram across the entire 3D domain, to indicate any suspicious discontinuities or inexplicable increases in the number of modes in the distribution; a histogram of densities specifically at the input data locations, to understand whether the estimated density field is proportional to the input data weights; and the number of null-density points, i.e., the input data points missed by the reconstruction (ideally, this will approach zero, as the fit should cover all the data).

V5: Interactive highlighting of transition regions. Of particular interest to astrophysicists is understanding the transition regions of the Cosmic Web, the subset of 3D space on the boundary of the circum-galactic medium (CGM) surrounding galaxies and the intergalactic medium (IGM) between galaxies. While models exist for describing the CGM and for predicting the statistical properties of the IGM [73], the transition region has so far eluded deeper understanding. Visualization can facilitate the investigation of these regions, such as through using density-dependent color coding and interactive thresholding.

V6: Inclusion of proxy objects. Assisting with the fitting is the ability to use visual aids, such as sightlines or distance markers. These proxy objects need to be consistently superimposed against the volumetric data to help navigation in the 3D space, a task that can be challenging due to the complex structures obtained during the reconstruction.

V7: Exporting the reconstructed solution. Once satisfied with the estimated Cosmic Web structure, researchers need the output data available in a format that supports further investigation using astrophysics analysis packages, such as yt [74], pyigm [55], and linetools [56]. The output data includes a 3D table of scalar density values packed in a binary file, and associated metadata describing the parametrization of the reconstruction model and the spatial orientation of the estimate.

4 SIMULATION AND MODEL FITTING

The Monte Carlo Physarum Machine (MCPM) model enables the structural analysis tasks defined above. It serves as an effective structural interpolator for cosmological datasets, which on the scales where the Cosmic Web exhibits distinct features (between 1 and 100 Mpc) can be represented as a weighted set of points in 3D space. MCPM is a dynamic model defined by a set of priors: a set of probability density functions and their parameters obtained from domain-specific knowledge and fitting to training data. Once configured, the model is fitted to the input data, resulting in a continuous geometric structure which we interpret as a transport network connecting the data points. Rather than a graph, this geometry is represented as a 3D density field stored as a regular lattice. An overview of MCPM outputs and related visualization modalities is provided in Fig. 1.

MCPM has both a discrete and a continuous component. The discrete component is an ensemble of particle-like agents that can freely navigate the simulation domain, representing different nuclei of the virtual organism. The continuous component is a 3D scalar lattice representing a concentration of a marker (also referred to as a “chemoattractant” in the biological context) that facilitates communication between the agents and the data. The model’s behavior is based on a feedback loop cycling between these two components, executed in two alternating steps, obtaining approximations from an iterated composition of simple, local rules (which are illustrated in Fig. 3).

The propagation step is executed in parallel for each of the agents, which are the model’s device for exploring the simulation domain. Each agent’s state is represented by a position and a movement direction, which are stochastically updated (Fig. 3c) to navigate through the deposit field (referred to as the “trail” in [55], see Fig. 3, gray). The deposit field is stored as a 3D lattice of scalar values, representing the marker emitted by both the input data points as well as the agents. The deposit guides the agents towards the input data points, as the agents move with higher likelihood to the regions with greater deposit values. In addition to the deposit, we also maintain a scalar trace field (the green cells in Fig. 3d) which records the agents’ equilibrium spatio-temporal density, but which does not participate in their decision making about which direction to move in. Successive propagation steps define the agents trajectories, which follow the structure of the deposit field, and are recorded in the trace field. While each individual agent draws a seemingly chaotic path (Fig. 4 left), a superposition of many agents averaged over a sufficient time window results in a smooth and converged structure (Fig. 4 right).

The input data points are also represented by agents, but are of a special type which fixes their position in space and enables them continually emit an amount of deposit proportional to their weight. We found this to be the most consistent way of handling the inputs, as the relative structural impact of the agents and the data simply becomes a question of tuning the amount of deposit emitted by each type.

The agent propagation step is further defined in terms of three probability distributions, the configuration of which can be interactively tuned in runtime. These are:

- \( P_{\text{dir}} \) for sampling the agents’ directional decisions during the sensing and movement phases, defined on the unit sphere relatively to the current direction of propagation.
- \( P_{\text{dist}} \) for sampling the agents’ distance decisions during the sensing and movement phases, defined in positive real numbers along the current propagating direction.
- \( P_{\text{branch}} \) for making the binary decision whether the agent should take a newly sampled direction or remain on its current course.

![Fig. 3. Visual depiction of the low-level sampling decisions performed by each MCPM agent per single propagation step. All three steps (a)–(c) are stochastic, generating a random walk (d) which constitutes the agent’s trajectory guided by the data-emitted deposit (grey field).](image-url)

![Fig. 4. Superposition of agent trajectories in 3D space. While individual agents follow seemingly chaotic paths, the order emerges from a statistical aggregate of millions of them. That aggregate—the trace field—is the reconstructed output of MCPM.](image-url)
In order to apply a Physarum model to the analysis of cosmological data, MCPM extends the pioneering work of Jones [35] in several ways. First, in MCPM, all the decisions made in the agent propagation step are stochastic. This makes the model highly customizable, as prior domain-specific knowledge can be encoded by probability density functions. Second, we extend Jones’s model from 2D to 3D, taking advantage of the probabilistic form, and enabling analysis of higher-dimensional datasets. Third, we add a new data modality—the trace field—that represents the equilibrated agent density, decoupled from the algorithms signaling mechanism.

In summary, MCPM solves tasks S1–S4 by functioning as a structural interpolator for input data. When configured and parametrized appropriately (typical settings are listed in Table 1), it reliably converges to transport networks that effectively connect neighboring data points in space. The key is that the agents are drawn towards large ‘pools’ of deposit when \( P_{\text{branch}} \) is defined such that it preferentially selects those directions where more deposit is detected. Agents are therefore attracted towards data in their vicinity (thus interconnecting them), as well as other nearby agents (thus reinforcing existing pathways). The shape characteristics of the network (such as the curvature of the paths)

### Table 1. Default parameter values used in the MCPM fitting.

| Parameter           | Value       | Units   |
|---------------------|-------------|---------|
| sense_angle         | 20          | deg     |
| sense_distance      | 2.5         | Mpc     |
| move_angle          | 10          | deg     |
| move_distance       | 0.1         | Mpc     |
| agent Deposit       | 0.005       |         |
| data Deposit        | 10          |         |
| agent Deposit       | 0.005       |         |
| data Deposit        | 10          |         |

The relaxation step ensures that the simulation eventually reaches an equilibrium. The deposit field is spatially diffused by a small isotropic kernel, and then attenuated (i.e., each cell in the deposit field is multiplied by a value \( < 1 \)). The trace field is also attenuated but does not diffuse in order to preserve geometric features in the agents’ distribution, which the trace ultimately represents. The simulation reaches its equilibrium when the amount of deposit and trace injected into the respective fields in the propagation step equals the amount removed by the attenuation in the relaxation step, typically within 100s of iterations.

In order to apply a Physarum model to the analysis of cosmological data, MCPM extends the pioneering work of Jones [35] in several ways. First, in MCPM, all the decisions made in the agent propagation step are stochastic. This makes the model highly customizable, as prior domain-specific knowledge can be encoded by probability density functions. Second, we extend Jones’s model from 2D to 3D, taking advantage of the probabilistic form, and enabling analysis of higher-dimensional datasets. Third, we add a new data modality—the trace field—that represents the equilibrated agent density, decoupled from the algorithms signaling mechanism.

In summary, MCPM solves tasks S1–S4 by functioning as a structural interpolator for input data. When configured and parametrized appropriately (typical settings are listed in Table 1), it reliably converges to transport networks that effectively connect neighboring data points in space. The key is that the agents are drawn towards large ‘pools’ of deposit when \( P_{\text{branch}} \) is defined such that it preferentially selects those directions where more deposit is detected. Agents are therefore attracted towards data in their vicinity (thus interconnecting them), as well as other nearby agents (thus reinforcing existing pathways). The shape characteristics of the network (such as the curvature of the paths)
which point the researcher can continue to use Polyphorm. The interactivity of both the MCPM and the visualization enables a responsive feedback loop, reducing the time it takes to obtain a satisfactory result. At any point in time, based on the current simulation state, any structural or visual parameter can be adjusted until a satisfactory result is produced, at which point the researcher can continue to use Polyphorm to investigate the reconstruction, or can export the reconstruction for further analysis.

5 Visual Analysis

Polyphorm is comprised of two integrated components: the simulation engine and the visualization interface. A researcher initializes the fitting with initial structural parameters (determined, e.g., from prior fitting runs), and then observes the progress of the fitting. At any point in time, based on the current simulation state, any structural or visual parameter can be adjusted until a satisfactory result is produced, at which point the researcher can continue to use Polyphorm to investigate the reconstruction, or can export the reconstruction for further analysis.

An example of a fitting session is depicted in Fig. 5, which shows the reconstruction of the Cosmic Web from 37.6k observed galaxies from the SDSS corpus. Although a full estimate takes about 600 MCPM iterations to produce (just over 1 minute on our development machine), a significant part of the reconstruction is revealed in a fraction of that time (at around 100 iterations), allowing for quick readjustments of the model parameters without having to wait for a fully converged estimate.

The interactive solution from hours—typical for other Cosmic Web reconstruction methods—to minutes. This provides a critical advantage when tackling a problem like this, where the solution is under-determined from the data and human supervision is required. Below, using the 37.6k galaxies extracted from the SDSS corpus, we describe the main components of the Polyphorm interface that contribute to the reconstruction process, where each component is aligned with one or more of the the visual analysis tasks described in Sec. 3.2.

Upon startup, Polyphorm displays the particle rendering mode (V1). The input dataset is by default visualized as red particles, while the MCPM simulation agents are white (users can customize the color mapping as desired). In our SDSS example, the simulation agents outnumber the data about 300-fold (see Fig. 9). The structural and visual parameters are exposed in the top-left part of the viewport, and each of them can be adjusted interactively (V2). A visualization option available in all the rendering modalities that allows the researcher to slice the dataset along any 3 axes (V1), and in the bottom-right corner, a display indicates the current trimming state. A statistical summary of the fit (V4) is provided in the bottom-left corner, showing a per-galaxy trace histogram, the energy plot of the fit plus its evolution during fitting, and the number of null-density data points. In Fig. 6 we can see from the energy plot that the simulation has already converged to an equilibrium, with the MCPM agents tracing the Cosmic Web estimate. At this point the simulation can be exported as a 3D density field for further analysis (V7).

5.1 Trace and Deposit Rendering Modes

To visualize the resulting fits (V1), Polyphorm uses two main views (V3): trace-only (Fig. 8 top) and combined deposit+trace (Fig. 8 bottom). Recall that the deposit represents the volumetric footprint of the input data and that the trace is the Cosmic Web density field estimated by the MCPM simulation. Both of these views are rendered via a standard volume rendering technique, using the emission+absorption tom). The trace is rendered by mapping the density field values to a customizable color map (akin to the color maps used in Matplotlib and similar packages), which therefore serves as a transfer function. In the trace-only view, we use a simple color gradient, by default starting with red at low trace values and ending in white for high values. The transparency of the medium is likewise scaled by the trace values, so that empty spaces (‘voids’) do not occlude the filament structure. The purpose of this view is to examine the geometry of the Cosmic Web without being biased by the input data distribution.

The combined deposit+trace view is rendered using the same methodology as the trace-only view. In this view, however, the deposit (white...
halos) is superimposed with the trace (purple filaments). The purpose of this view is to visually verify that the trace is distributed evenly and that it aligns well with the input data.

5.2 Trace Highlighting and Segmentation

Complementing the global visual inspection of the trace, Polyphorm implements additional visualization modalities: density segmentation and highlighting (V3, V5). The main purpose of the segmentation mode is to split the reconstructed trace distribution into three mutually exclusive density intervals, as shown in Fig. 7. Here, the threshold values for the segmentation have been determined by our astrophysicist collaborators, relying on the calibrated mapping between the trace values and the cosmic overdensity, a well understood quantity in cosmology (see Sec. 4).

To highlight a particular narrow density interval in the trace distribution, Polyphorm also implements thresholding in the deposit-trace modality. In Fig. 9 we use a trace density value that is known to correspond to the medium density on the outskirts of galactic halos and denser filaments. This highlighting can be used to verify that this spatial assumption is found in our reconstruction of the Cosmic Web.

5.3 Axis-aligned Trimming

An important interactive component of Polyphorm is the axis-aligned trimming (or ‘slicing’) available across all of the visualization modalities (V1, V3). Having the ability to slice the data in different directions is invaluable, since spatial orientation in the full 3D dataset can be quite a challenge. Fig. 10 demonstrates this functionality by isolating a thin vertical slice from the SDSS dataset (about 10% of the total width). Regardless of the employed rendering modality, we can observe structures that may be occluded when viewing the reconstruction in its entirety.

5.4 Proxy Visuals

Polyphorm also includes custom visualization aids (V6), and is designed so that additional proxy objects can be added as needed for particular datasets. These help to orient the user when navigating large 3D datasets and to better understand or to emphasize spatial relationships in the data. For example, we can add spherical shells to signify distances from Earth measured in the Cosmic redshift values. These translucent proxy objects make it apparent that the data is defined in spherical coordinates, and simplify the comparison of redshift distances of various parts in the reconstruction. Additionally, we can add one or more sightlines connecting Earth to distant objects, such as quasars or fast radio bursts (FRBs). In Use Case 3 (Sec. 6.3), astrophysicists use Polyphorm to measure the amount of dispersion in the signal arriving from a particular FRB. According to their hypothesis, this dispersion is due to interactions with the Cosmic Web. By visualizing the FRB sightline (see Fig. 12), astrophysicists can more easily reason about its interactions with the Cosmic Web.

5.5 Implementation Details

Polyphorm was written in C++ (for the CPU code) and DirectX’s HLSL (for the GPU code). All performance-critical steps are parallelized and implemented as HLSL compute shaders, including MCPM’s propagation and relaxation steps, as well as the particle and volume rendering modes. All data structures in Polyphorm are ordered, with emphasis on fast access in both sequential and random modes. Specifically, the input data points (galaxies or dark matter halos) and MCPM agents are stored in linear arrays, using fp32 to store their positions, weights, and, in the case of agents, directions. The deposit and trace fields are stored in regular 3D grids with fp16 voxels for efficiency reasons (as they represent the primary memory bottleneck).

Table 2 lists performance figures for Polyphorm using the three datasets analyzed in the Scientific Use Cases below: Bolshoi-Planck (large dataset, Sec. 6.1), SDSS (medium dataset, Sec. 6.2) and FRB (small dataset, Sec. 6.3). All results were obtained using a desktop computer outfitted with an Intel Core i9-9900K CPU and an NVIDIA TitanX GPU. The simulation and visualization performance scales approximately linearly with the number of voxels, as expected, and sub-linearly with the number of agents (e.g., using 100M agents instead of 10M incurs only a 2x slowdown). Although using higher numbers of agents has only a small impact on the model’s spatial details, this does improve the model’s effective resolution by reducing the Monte Carlo noise levels. Using 10M agents has proven sufficient across all our use cases.

6 Scientific Use Cases

6.1 Use Case 1: Establishing a mapping between the MCPM density estimate and cosmic overdensity

Polyphorm was used to establish a complete, calibrated map of the dark matter density field (traced by the locations of galaxies) in a sufficiently representative region of space in order to study the distribution of IGM gas relative to the dark matter structures [16]. As current consensus estimates that the sizes of IGM filaments range between a few to dozens of Mpc, our reconstruction needed to encompass a region of space larger than this. Similarly, an effective spatial resolution is needed to represent
scales as small as the extent of typical galactic halos, between 0.1 and 1 Mpc.

The trace produced by the MCPM model defines a 3D density field \( \rho_{\text{phys}}(S1) \). To calibrate these values to the more cosmologically meaningful cosmic matter density term \( \rho_m \), our collaborators used Polyphorm to fit the MCPM model to the halo catalog produced by the ROCKSTAR algorithm \( V \) for the Bolshoi-Planck (BP) simulation at the \( z = 0 \) redshift snapshot. The resulting fit is shown in Fig. 4 in the different visual modalities provided by the Polyphorm visualization engine \( V3 \). As per requirements defined by the structural tasks \( S2 \) and \( S3 \), the reconstruction yields the expected anisotropical filamentary structures across multiple scales in space and density (see Fig. 11). The reconstruction also has excellent alignment with the 4804 dark matter halos in the BP dataset, and only a negligible number of halos \((< 0.1\%) \) are not included in the reconstruction. The spatial requirements were likewise fulfilled: the reconstruction spans a cubic volume 170 Mpc in each side, with a density field resolution of 1024\(^3 \), i.e., the minimum represented feature size was 0.166 Mpc.

The calibration to cosmic matter density \( \rho_m \) then involved dividing the \( \rho_{\text{phys}} \) values into a histogram with bins of width 0.05 in the log\( _2 \) domain, and spatially cross-matching cells in each \( \rho_{\text{phys}} \) bin with a reference \( \rho_m \) field, which is part of the BP corpus. This reference 3D matter density field was derived by smoothing the particle mass density in the BP volume over 0.25 Mpc scales, similar to the resolution scale of our Polyphorm data product. The actual mapping was then obtained via the running median method \( V \) on the cross-matched densities \( \rho_{\text{phys}} \) and \( \rho_m \), and subsequent normalization by the mean matter density to obtain a relative quantity, cosmic overdensity \( \rho_{\text{phys}} / \rho_m \). It is important to note that while a more precise map could have been obtained by fitting to this reference \( \rho_m \) field, this was explicitly avoided since fitting to sparse point data (akin to the observational data where only galaxy locations are known) was the key task \( S1 \) of Polyphorm.

The MCPM trace and BP matter densities are well correlated for \( \rho_m \geq 1 \), which was the regime of interest for our astrophysicist collaborators. The use of Polyphorm was a critical ingredient in establishing a first-of-its-kind, continuous spatial map between a synthetic computational model (MCPM) and a canonical cosmological quantity, the cosmic overdensity \( \rho_{\text{phys}} / \rho_m \).

6.2 Use Case 2: Reconstructing the Cosmic Web from SDSS galaxy observations for IGAM analysis

Based on the results of Use Case 1, along with the fact that the MCPM model yields a 3D density distribution \( \rho_{\text{phys}}(S1) \), it now becomes possible to analyze astrophysical observables in context with the large scale structure. The MCPM trace density field data product readily enables analyzing correlations between, e.g., galaxies’ star formation rates and their local environmental density. This may be achieved in a straightforward manner because a local MCPM density value can be assigned to each galaxy based solely on its spatial coordinates. Polyphorm was used to track the density of atomic hydrogen (\( H I \)) as a tracer of the IGM, testing the long-held assumption that the spatial distribution of \( H I \) follows that of the overall Cosmic Web structure. To map out the predicted structure, our collaborators used 37.6k galaxies from the Sloan Digital Sky Survey (SDSS) dataset, spanning across 280 Mpc of space at redshifts \( z = 0.0138 – 0.318 \). (This is also the dataset used for Figs. 5–10 referenced in Sec. 5.)

To obtain a robust mapping to the SDSS data (which are relatively sparse), visual analysis tasks \( V1 – V4 \) were employed. The structural transfer \( S4 \) facilitated by the prior fitting to the BP data was key to estimating the MCPM parameters for this fitting. The primary technique for studying \( H I \) in the IGM is quasar absorption line spectroscopy, in which light from distant quasars serves as background radiation sources \( S1, S6, S7 \). Elements and ions, such as \( H I \), leave imprints on the background light in the form of spectral absorption features (or lines) at the specific locations they pass through. To detect and measure the line-of-sight density of the IGM gas, 304 spectroscopic sightlines of distant background quasars piercing the MCPM-mapped SDSS region were obtained from the Hubble Space Telescope archival database \( S9 \). Our astrophysicist collaborators preprocessed the spectra from this database to search for the \( H I \) \( Ly\alpha \) absorption signature by parametrizing each pixel by its distance from Earth along the line of sight. To avoid pollution of the absorption signal, they masked the specific wavelength ranges corresponding to the signatures of the Milky Way’s inner interstellar medium. These ‘cleaned’ spectra were then back-projected into 3D space and correlated with the Cosmic Web estimate provided by Polyphorm \( V7 \) and mapped from \( \rho_{\text{phys}} \) to \( \rho_m \). Binning the resulting pairs of log \( \rho_m \) density + \( Ly\alpha \) absorption averaged over all 346 sightlines, our collaborators found a clear relationship between these two quantities \( S16 \). They observed a statistically significant increase of the detected \( Ly\alpha \) absorption coincident with the cosmic overdensity increasing over the mean (log \( \rho_m / \rho_m > 0 \)). This relation demonstrates that Polyphorm is capable of producing a viable predictor for the \( H I \) distribution in the density regime \( 0 \leq \log \rho_m / \rho_m \leq 3 \). Interactively highlighting transition regions \( V5 \) provided the verification that this density regime corresponds to the circumgalactic environments as expected.

A leveling off of \( H I \) at log \( \rho_m / \rho_m \geq 2 \) and downsloving observed at log \( \rho_{\text{phys}} / \rho_m > 3 \) is counterintuitive, and the astrophysics community is still debating the cause of this behavior. The initial hypothesis presented by our collaborators was that, in this high-density regime, the increasing overall density is offset by the neutral \( H I \) becoming ionized due to shock heating processes and/or excess radiation from nearby galaxies. That is, the gas probed by the quasar sightlines contains fewer neutral atoms and therefore absorbs less in \( Ly\alpha \) transition. Polyphorm facilitated the visual analysis of the three main density regimes \( V3, V5 \), including: voids, the vast under-dense pockets in the Cosmic Web where no \( H I \) signature is detected; filaments, where we observe a monotonic increase in the absorption with increasing cosmic density; and high-density shock-heated regions, where the increasing absorption signature is eventually suppressed. These results provide strong evidence towards the conjecture that the bulk of the IGM resides in the Cosmic Web filaments \( S16 \).

6.3 Use Case 3: Inferring the ionized IGM contribution to the dispersion measure of FRB 190608
Polyphorm was used to investigate the filamentary structure of the Cosmic Web in a region along a Fast Radio Burst (FRB) sightline, for which observed data indicate an unusually large dispersion measure compared to the cosmic average at its redshift, as well as an unusually large rotation measurement and pulse width. By producing a 3D map of ionized gas in the filaments, our collaborators were able to compute a dispersion measure contribution from matter outside halos, leading to the first detailed end-to-end study of matter along an FRB sightline \( S66 \).
In so doing, our collaborators were able to “disentangle” the Cosmic Web by correlating the dispersion measure of FRBs with the distributions of foreground galaxy halos. This dispersion measure (DM) is the path integral of the electron density $n_e$ weighted by the scale factor $(1+z)^{-1}$: $DM \equiv \int n_e ds/(1+z)$. That is, these FRB measurements are sensitive to all of the ionized gas along the sightline, and therefore can be used to trace the otherwise invisible plasma surrounding and in-between galaxy halos. The DM is an additive quantity, and for an FRB it may be split into individual contributions of intervening, ionized gas reservoirs, including a contribution from the Milky Way, from the host galaxy (where the FRB originates), and from the Cosmic Web (i.e., the intervening halos made up of CGM gas found within dark matter halos adjacent to galaxies and IGM gas found between galaxy halos). Since there are no known galaxies directly intercepted by the line of sight between the Milky Way and the host galaxy, FRB 190608 lends itself to an investigation of how filaments in the Cosmic Web affect the DM. The FRB sightline falls within a region that serendipitously is covered by SDSS, providing the opportunity to consider the cosmic gas residing within the underlying large-scale structure. Theoretical models predict shock-heated gas within the Cosmic Web as a natural consequence of structure formation, and FRBs offer one of the most promising paths forward in detecting this elusive material.

The MCPM methodology produces a continuous 3D density field defined even relatively far away from galaxies on Mpc scales, and traces anisotropic filamentary structures on both large and small scales. Using the SDSS galaxy distribution within 400 arcmin of the FRB sightline (1111 galaxies with mass $10^{10}$ $M_\odot$ or higher) as input to Polyphorm, our astrophysicist collaborators used MCPM reconstruction to map the large-scale structure intercepted by the FRB sightline. The resulting 3D map enabled astrophysicists to identify and characterize the density of filaments between Earth and FRB 190608 via a similar calibration procedure as that described in Use Case 1. With this density estimate, the contributions of the ionized gas within the Cosmic Web to the observed DM could be calculated. The reconstruction of the structure intercepted by the FRB sightline is depicted in Fig. 12. Visual analysis features that were especially important for this use case were the ability to interactively observe the reconstructed Cosmic Web in different modalities and to focus on particular slices of the dataset and filament reconstruction, as well as the ability to include visual annotations and to export the reconstructed density field for subsequent analysis of the observed DM along the FRB sightline.

### 7 Conclusion and Future Work

In this paper, we introduced Polyphorm, an interactive visualization tool that enables the application of the MCPM model to large cosmological datasets, and provided three scientific use cases in which the structural and visual analysis tasks enabled by Polyphorm led directly to new insights into composition and structure of the Cosmic Web. We will continue to use Polyphorm to explore additional astrophysics datasets in the future, and look for alternative methods to validate the MCPM model in other cosmological contexts, using both observed and simulated data. We plan to further investigate the mathematical foundations for our stochastic model and to include additional analysis methods in future iterations of Polyphorm, such as topological analyses for graph extraction, which is directly applicable to cataloging Cosmic Web filaments, and which may expand the range of structural tasks our software could support. Though our visualization and interaction modalities were effective in facilitating the scientific use cases we described, incorporating a richer palette of rendering techniques could provide additional insight into the input data and filament reconstructions, supporting further explorations of the transitional regions in the Cosmic Web.

The fitting of MCPM is a dynamic process with rate of convergence following a Pareto-like distribution, which is expected from a stochastic Monte Carlo simulation. In addition, the model is continuous in space and time, including the discretization of agent trajectories (the scale of which is now related to sampling density). We believe that this makes MCPM suitable for seamless fitting of time-dependent data. If the temporal evolution of the dataset is matched with the model dynamics, the agents can be made to travel in space-time with minimal distortion, which could provide an effective way to represent the temporal evolution of the Universe. Future work will explore extending Polyphorm to incorporate dynamic data.

*Polyphorm* is available via our open source GitHub project repository at [https://github.com/CreativeCodingLab/Polyphorm](https://github.com/CreativeCodingLab/Polyphorm) along with source code, detailed instructions on how to load in and analyze custom datasets, a video overview, and additional documentation.

### Acknowledgments

This research was supported by NASA through award HST-AR #15009 from the Space Telescope Science Institute. Hardware was provided by the NVIDIA GPU Grant program. Special thanks to Jan Ivanecký for his help in developing an initial prototype of the software.
