# Extracting the Slime Mold Graph from the Cosmic Web

P. H. Barchi[a,b,*], J. N. Burchett[c], O. Elek[d], A. Forbes[d], J. X. Prochaska[c]

[a] *Lab for Computing and Applied Mathematics, National Institute for Space Research (INPE)*
*Av. dos Astronautas, 1.758 - Jardim da Granja, São José dos Campos - SP 12227-010, Brazil*
[b] *Physics Department, Brandeis University*
[c] *Department of Astronomy and Astrophysics, University of California, Santa Cruz (UCSC)*
[d] *Computational Media Department, University of California, Santa Cruz (UCSC)*

**Abstract**

The large-scale structure of the universe has an intricate graph structure. Although, the explicit network extracted from the galaxy distribution does not entirely reflect the cosmic web. One way to overcome this issue is to apply a method that transforms the discrete data from galaxies in a continuous density field. Physarum Polycephalum is a slime mold which typically finds the most efficient path between food locations over large areas. We employ Physarum Machines in galaxy distributions to obtain the density field, and, with this density field, we extract the graph of the large-scale structure. Graph is a simple and powerful structure which consists of vertices connected by edges. Two different groups of graph algorithms are hypergraph and community detection. Hypergraph algorithms cluster edges while community detection cluster vertices. By extracting the large-scale graph and applying these algorithms, we ultimately aim to (1) extract new relationships about abstract parameter space — another dimension for studying galaxy interactions; (2) employ hypergraph algorithms to produce filaments catalog; (3) apply community detection algorithms to detect galaxy clusters; (4) combine hypergraph and community detection to obtain a meta-graph with the rough large-structure. We plan to provide a filaments catalog as a main result of this work.

*Keywords:* methods: data analysis, mathematics of computing: discrete mathematics: graph theory, cosmology: large-scale structure of universe, computing methodologies: machine learning: machine learning approaches

## 1. Introduction

Large galaxy redshift surveys display a heterogeneous discrete distribution, which has a structure with two main components: galaxy filaments and voids. Galaxy filaments are large thread-like formations with gravitationally bound galaxies and gases that establish boundaries between large voids in the universe. Such structure of the large-scale structure of the universe is also known as cosmic web, because of its web-like appearance (Jõeveer et al., 1978; Bond et al., 1996). This salient structure is visually evident, however, it is quite challenging to define a mathematical structure from it. We appeal to graph theory to extract and study this network.

Graphs are mathematical structures used to model different types of relationships and processes. The basic structure of a graph is made up vertices which are connected by edges (Hazewinkel, 2013). Several algorithms can be employed to study and extract knowledge from the graph topology. One of them consists of obtaining a hypergraph from the original graph. A hypergraph is a generalization of a graph in which an edge can join any number of vertices. Edges are clustered by their similarity in the parameter space of the graph (Bretto, 2013). Another category of graph algorithms is community detection: a community is commonly defined as vertices which are densely connected with each other (Fortunato, 2010).

By extracting a graph from galaxy distributions only, the structure revealed does not seem natural, and also does not reveal the same appearance as we see in well-established simulations as (Illustris, Vogelsberger et al., 2014; Genel et al., 2014; Nelson et al., 2015) and (EAGLE, Schaye et al., 2014). One way to overcome this issue is to apply a bio-inspired method upon the discrete distribution of galaxies to obtain a more natural continuous density field of the universe.

Physarum Polycephalum (Physarum hereafter) is a slime mold, an unicellular protist that inhabits shady, cool and moist areas. Biological experiments have demonstrated Physarum's extraordinary intelligence to find the most efficient path among food locations over large areas and thus build efficient networks (Nakagaki et al., 2000; Oettmeier et al., 2017). Naturally, computer science researchers have been developing swarm intelligence algorithms to simulate Physarum's behaviour — also known as Physarum Machines (Sun, 2017). Burchett et al.; Elek et al. have been working on a 3D Physarum Machine adaptation for the large-scale structure of the universe.

This work employs the filament density map (extracted from galaxy halos given by the Sloan Digital Sky Survey, SDSS) by Burchett et al.; Elek et al. using a custom 3D adaptation of the Physarum Machine model. Using the Physarum density to define a spatial *throughput*, we build the basic architecture of the graph (vertices and edges). With this base graph, we plan to (1)

*Corresponding author
*Email address:* `paulobarchi@gmail.com` (P. H. Barchi)
*URL:* `paulobarchi.github.io` (P. H. Barchi)

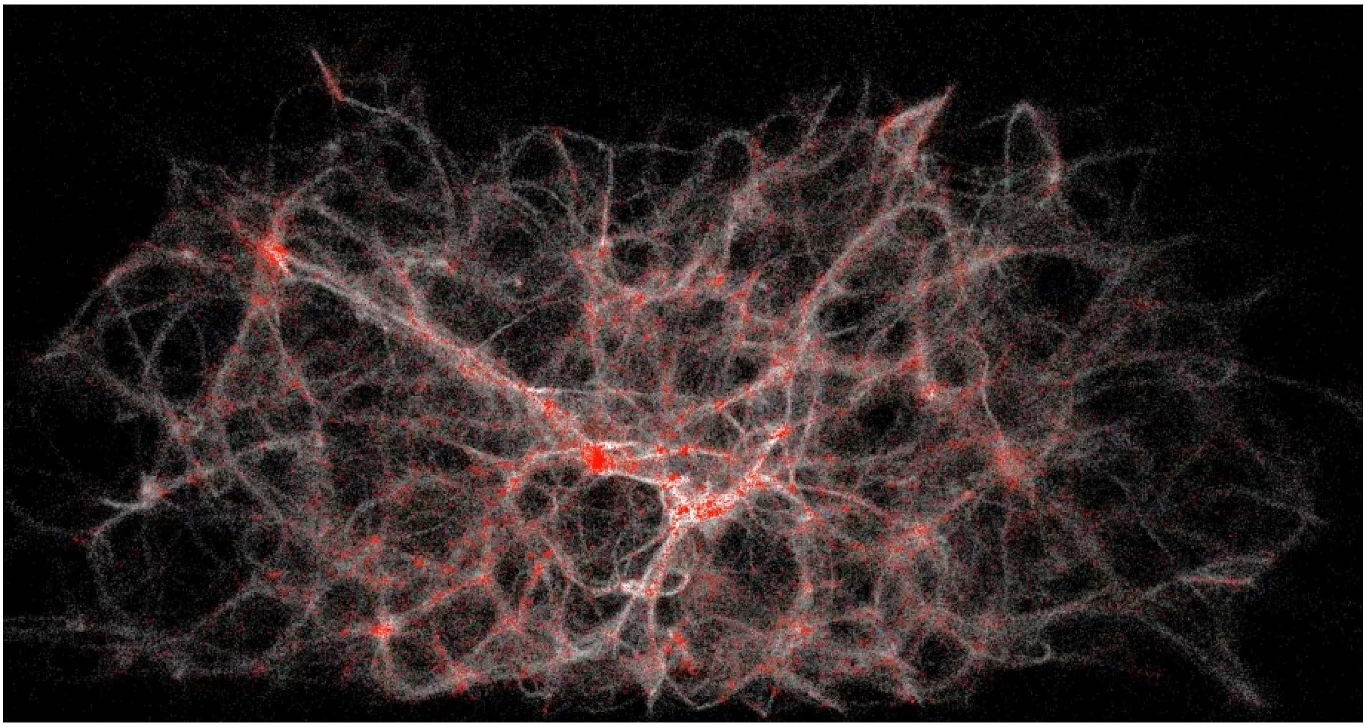


Figure 1: Snapshot of the Physarum fit in 3D (in white) on SDSS galaxies (in red) used to extract the graph.

extract new relationships about abstract parameter space – another dimension for studying galaxy interactions; (2) employ hypergraph algorithms to generate filaments catalog; (3) apply community detection algorithms to detect galaxy clusters; (4) combine hypergraph and community detection to obtain a meta-graph with the rough large-structure. We will provide a filaments catalog as a final product of this work.

This document is organized as follows: in Section 2, we briefly describe Physarum Machines; in Section 3, we explain how we extract the base graph the Physarum throughput; and, in Section 5, we summarize what we have achieved so far and future plans.

## 2. Physarum machines for large-scale structure finding

The many-headed slime mold (Physarum Polycephalum) is an unicellular protist, although it can be seen without microscope. Because of its efficiency to find the best path among food locations (its body morphology approximates an optimal transport network constrained by the available nutrient distribution), one of the aims of research with regard to Physarum is to understand and model mechanisms of information processing coupled to network topology and morphology. The intelligent behaviour showed to build such optimized networks have been exploited to solve different network problems (Nakagaki et al., 2000; Oettmeier et al., 2017) and have been inspiring swarm intelligence and cellular automata algorithms to simulate the Physarum behaviour as well (Sun, 2017). Jones (2010) presents a well-established 2D implementation of a Physarum Machine.

Basically, each Physarum particle has defined a location $(x, y)$, a movement direction a sensor direction and a sensor distance. The food sources emit signals which can be sensed by the particles, and the particles steer to reach the food locations. The particles leave a deposit while moving, which can diffuse to its neighbors in the 2D grid, and then decay, if such path is not used frequently by all slime particles. We use the 3D Physarum Machine implementation from Burchett et al.; Elek et al. for obtaining the most visually accurate fit of the Cosmic Web using SDSS galaxies as food locations (see Figure 1). We release a swarm counting millions of slime mold agents into the same 3D space which "food data" (galaxy halo locations, in this work) is defined in. These data emit a signal that can be sensed by slime mold cells which will then find efficient paths between food locations (see Burchett et al.; Elek et al., for more details).

## 3. Translating Physarum density field to graph: semantics

Physarum Machines simulate the behavior of a large number of Physarum Polycephalum cells, which leave a trace of their path constituting a continuous density field in the defined space of the problem. We interpret the discrete output of the algorithm as the throughput set of the snapshot that builds up such density field.

We obtain a density field from the Cosmic Web by applying Physarum Machines on a three-dimensional space which also has a discrete distribution of galaxies. How to obtain a mathematical structure from the density field created by Physarum? Among different possibilities, we use a graph to define such structure. The idea is to map density points and/or regions as vertices and their relationships as edges. Given a successfully graph extraction that maps the filamentary structure of the large-scale structure, we can perform topology studies and apply a variety of graph algorithms. Here, we focus on community detection and hypergraph, which we explain in the following subsection.

### 3.1. *Communities and hypergraph*

Graphs representing natural systems are often complex networks with order coexisting with disorder. Within such structures, generally there is an underlying community structure. Community detection is currently one of the most popular types of graph algorithms (Fortunato, 2010). Among all available methodologies to detect communities, this work focuses on the hierarchical optimization of modularity by Blondel et al. (2008), where vertices are assigned to clusters iterativelly, which results in a hierarchical representation of the vertices in the graph (see Blondel et al., 2008, for further details). By detecting communities and translating the coordinates back to real data, we aim to study galaxies relationships within such communities, and, possibly detect galaxy clusters and/or galactic halos as well.

A hypergraph is a graph where the edges can connect any number of vertices — these are called hyperedges (Bretto, 2013). While community detection algorithms cluster vertices, hypergraphs create hyperedges by clustering edges from the base graph. With hypergraphs, we can study if the hyperedges found in this work are related to filaments in the large-scale structure.

## 4. Extracting the graph from Physarum density field

The data used from the Physarum Machine fit contains 3D coordinates $(x, y, z)$ and throughput values. This constitutes a data cube with dimensions $360 \times 360 \times 360$ and more than 17 million non-zero values. We normalize the throughput values and we constrain the dataset to have throughput $\geq 0.9$ to obtain a reduced dataset. Figure 2 presents the current methodology for this work. We explain each of the steps in the following subsections. The code for this project is available at: `https://github.com/paulobarchi/SlimeGraphExtraction` (work in progress).

### 4.1. *Using KDTree for storing distances*

One of the major challenges is to calculate the thoughput distances along the data cube to decide whether or not to assign a edge between them. We apply KDTree to efficiently obtain and store the distances from the data cube. KDTree is a space-partitioning data structure which organizes points in a k-dimensional space ($k = 3$ in this case) (Bentley, 1975; Berg, de et al., 2008). This step uses the whole dataset (without downsampling) because we will need the distances among all throughput values to assign weights for the edges further up.

### 4.2. *Partitioning, downsampling and creating vertices*

A well defined fit of Physarum Machines for the large-scale structure requires millions of cells. Ideally, we would build an initial base graph where every data point is a vertex, then apply graph algorithms upon it. However, as we consider 3D positions and throughput values as vertex properties, and the distance among them to define edges, it is not computationally feasible to build up the initial graph with the whole dataset. First we split the data cube in $n_{sb}$ sub-cubes. For the first experiments, we empirically define $n_{sb} = 27$ since we significantly reduce the computational cost and preserve meaningful structures in smaller scales. We uniformly downsample the dataset (from $\approx 10^7$ to $\approx 10^5$ points) and generate the initial graph with the resulting sample as vertices.

### 4.3. *Defining edges and their weights*

A graph without edges has no information about the relationship between vertices. We define the edges based on the distance among the vertices. We explore different values as the maximum distance threshold ($m$) to define edges and set $m = 1$ for preliminary graphs. Initially, the weight of each edge is defined by the distance between the vertices. In the next step, we query the KDTree to extract all the original throughput values in the path of each edge. We empirically define that the throughput is in the path if the distance from it to the edge is less than or equal 0.3. Given this defined throughput path, we integrate all these values to assign the weight of the edges. At this point, we have a graph with defined vertices, edges and weights.

### 4.4. *Preliminary Graphs*

In this subsection, we present the preliminary graphs and communities obtained in experiments performed during the *Kavli Summer School in Astrophysics 2019: Machine Learning in the era of large astronomical surveys*. We explored different number of sub-cubes ($n_{sc}$) in the partition steps and report here experiments with $n_{sc} = 27$. These experiments does not involve downsampling neither integration of throughput path, since we are currently implementing these steps in the current system. Central sub-cubes have a highly dense structure, while the corners of the data cube are basically empty.

Figure 3 presents sub-cubes and their graphs shifting along the original data cube. Shifting is applied in all three axes. Figures 3a, 3b, and 3c are shifted by 10, 12, and 14 units, respectively. Each of these figures shows the density field with all the throughput values in the left column, the graph extracted in the middle column, and the resulting graph from the community detection algorithm in the right column. For the first two columns, the rows just present different angles for better visualization. With lower shifts — closer to the center — more complex and denser structures are shown, as expected. With higher shifts, simpler graphs are extracted as sparsity takes over.

## 5. Summary

This document describes the first steps towards the extraction of graph structures from 3D Physarum Machine applied on the large-scale structure of the universe. We summarize this work as follows:

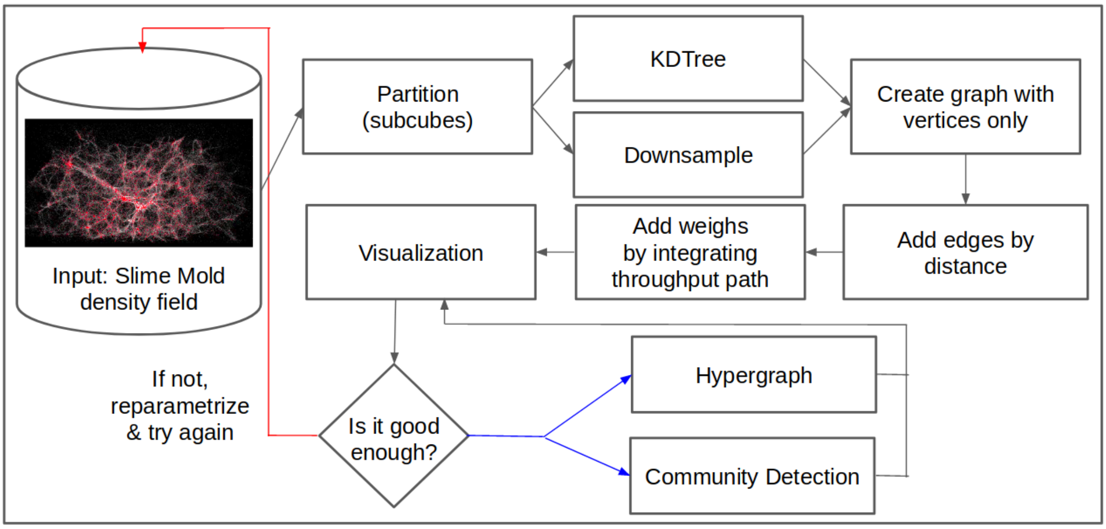



Figure 2: Illustrative sketch of the methodology for the graph extraction from the Physarum fit.

- Physarum Machines can efficiently find paths among a discrete galaxy distribution in a defined space and produce a density field (Cosmic Web).
- We propose and implement a methodology to extract graphs from density fields.
- Preliminary results obtained during the *Kavli Summer School in Astrophysics 2019: Machine Learning in the era of large astronomical surveys* are shown in Subsection 4.4.

We also highlight work in progress topics and next steps in the following items:

- We are currently developing the method to integrate the throughput between two vertices to assign the definitive edge weight.
- By applying hypergraph algorithms, we ultimately aim to provide a filaments catalog.
- We aim to compare our filaments catalog with the catalog provided by Tempel et al. (2014).
- We aim to explore communities in the graph with different community detection algorithms and analyse if they translate to either galaxy clusters, galactic halos or other galaxy interactions.
- The Physarum fit reveals hidden channels that are not present in galaxy distributions only. By studying the graph topology, we can analyse galaxy interactions in an abstract parameter space that was not available before.

## Acknowledgements


This work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors (especially P. H. Barchi) thank the financial support from The Kavli Foundation, the National Science Foundation (NSF) and University of California, Santa Cruz (UCSC) for sponsoring the Kavli Summer Program in Astrophysics, where and when P. H. Barchi have met the other authors and started working on this project. P. H. Barchi also thank Marcelle Soares-Santos and Brandeis University for hosting him in USA and travel funding.


## References


Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Commun. ACM 18, 509–517. doi:10.1145/361002.361007.

Berg, de, M., Cheong, O., Kreveld, van, M., Overmars, M., 2008. Computational geometry : algorithms and applications. 3rd ed ed., Springer. doi:10.1007/978-3-540-77974-2_5.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008, P10008. doi:10.1088/1742-5468/2008/10/p10008.

Bond, J.R., Kofman, L., Pogosyan, D., 1996. How filaments of galaxies are woven into the cosmic web. Nature 380, 603–606. doi:10.1038/380603a0.

Bretto, A., 2013. Hypergraph theory. An introduction. Mathematical Engineering. Cham: Springer .

Burchett, J., Elek, O., Forbes, A., et al., . in prep .

Elek, O., Burchett, J., Forbes, A., et al., . in prep .

Fortunato, S., 2010. Community detection in graphs. Physics reports 486, 75–174.

Genel, S., Vogelsberger, M., Springel, V., Sijacki, D., Nelson, D., Snyder, G., Rodriguez-Gomez, V., Torrey, P., Hernquist, L., 2014. Introducing the illustris project: the evolution of galaxy populations across cosmic time. Monthly Notices of the Royal Astronomical Society 445, 175–200.

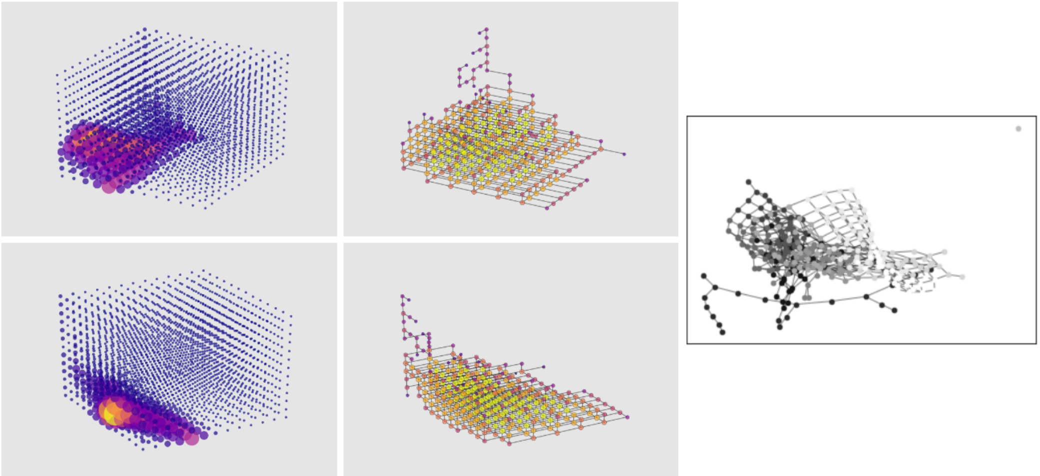
Hazewinkel, M., 2013. Encyclopaedia of mathematics: C an updated and annotated translation of the soviet mathematical encyclopaedia. volume 2. Springer Science & Business Media.

Jõeveer, M., Einasto, J., Tago, E., 1978. Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere. MNRAS 185, 357–370. doi:10.1093/mnras/185.2.357.

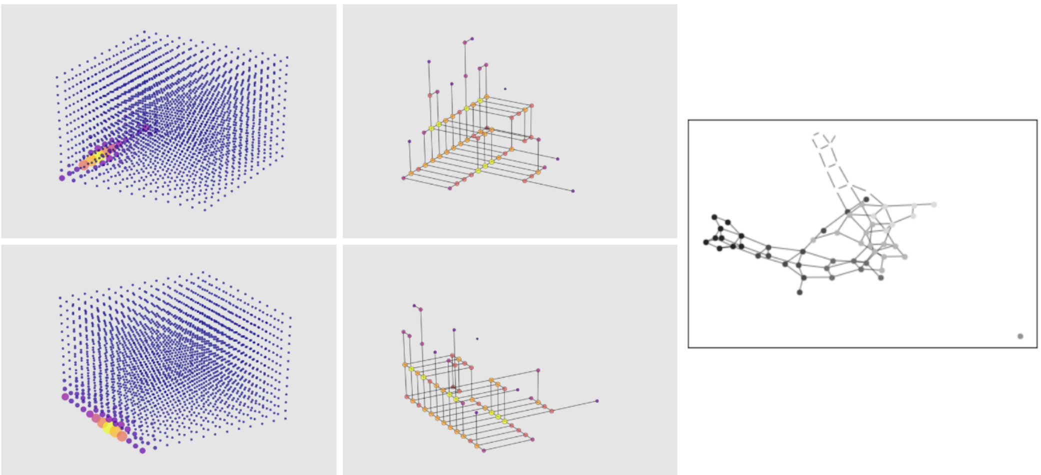
Jones, J., 2010. Characteristics of pattern formation and evolution in approximations of physarum transport networks. Artificial Life 16, 127–153. doi:10.1162/artl.2010.16.2.16202.

Nakagaki, T., Yamada, H., Tóth, Á., 2000. Maze-solving by an amoeboid organism. Nature 407, 470.

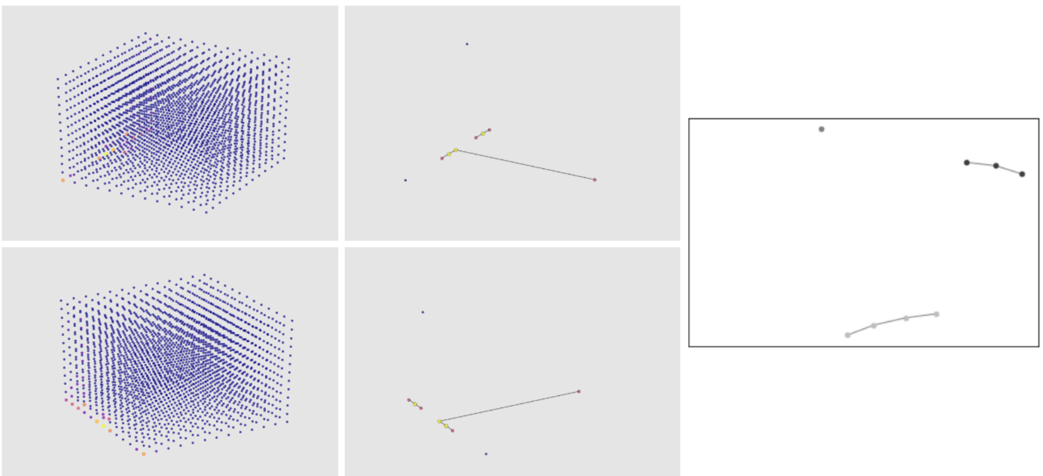
Nelson, D., Pillepich, A., Genel, S., Vogelsberger, M., Springel, V., Torrey, P., Rodriguez-Gomez, V., Sijacki, D., Snyder, G.F., Griffen, B., et al., 2015.



(a) $shift = 10$



(b) $shift = 12$



(c) $shift = 14$

Figure 3: Plots of density fields (first column), graphs (middle column) and community graphs (third column) extracted along the original data cube. Different rows present the same content from a different angle. See the explanation in 4.4 for further explanation.

The illustris simulation: Public data release. Astronomy and Computing 13, 12–37.
Oettmeier, C., Brix, K., Dbereiner, H.G., 2017. Physarum polycephalum—a new take on a classic model system. Journal of Physics D: Applied Physics 50, 413001. doi:10.1088/1361-6463/aa8699.
Schaye, J., Crain, R.A., Bower, R.G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C.S., McCarthy, I., Helly, J.C., et al., 2014. The eagle project: simulating the evolution and assembly of galaxies and their environments. Monthly Notices of the Royal Astronomical Society 446, 521–554.
Sun, Y., 2017. Physarum-inspired Network Optimization: A Review. arXiv e-prints arXiv:1712.02910.
Tempel, E., Stoica, R.S., Martnez, V.J., Liivamgi, L.J., Castellan, G., Saar, E., 2014. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. Monthly Notices of the Royal Astronomical Society 438, 3465–3482. doi:10.1093/mnras/stt2454.
Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., Hernquist, L., 2014. Introducing the illustris project: simulating the coevolution of dark and visible matter in the universe. Monthly Notices of the Royal Astronomical Society 444, 1518–1547.