Mem. S.A.It. Vol. 94, 131

**Memorie** della

# Swarm Intelligence-based Extraction and Manifold Crawling along the Large Scale Structure

P. Awad[1,2], R. Peletier[1], M. Canducci[3], R. Smith[4], A. Taghribi[2], M. Mohammadi[2], J. Shin[5], P. Tiňo[3], and K. Bunte[2]

[1] University of Groningen, Kapteyn Astronomical Institute, 9747 AD Groningen, The Netherlands e-mail: p.awad@rug.nl
[2] University of Groningen, Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence
[3] University of Birmingham, School of Computer Science, B15 1TT, Birmingham, United Kingdom
[4] Universidad Technica Frederico de Santa Maria, Avenida Vicuña Mackenna 3939, San Joaquín, Santiago
[5] Korea Astronomy and Space Science Institute (KASI) 776, Daedeok-daero, Yuseong-gu, Daejeon, 34055, South Korea

Received: 21-11-2022; Accepted: 22-02-2023

**Abstract.** The spatial distribution of matter on the mega-parsec scale of the Universe forms a complex and highly anisotropic pattern termed the Cosmic Web or the Large-Scale Structure. In the study of the Cosmic Web, several tools and methodologies have been developed to inspect the properties of its different environments i.e. clusters, filaments, walls, and voids. In this work, we show that the previously introduced framework 1-Dimensional Recovery, Extraction, and Analysis of Manifolds (1-DREAM) can analyze cosmological N-body simulation data of the Cosmic Web. 1-DREAM is a toolbox consisting of five Machine Learning algorithms that jointly serve the extraction and modelling of 1-dimensional structures in big data astronomical settings. After explaining the function of the different algorithms, we compare our toolbox with other methods which trace structures of the Cosmic Web. We show that 1-DREAM is able to split the network into its various environments with results comparable to the state-of-the-art methodologies. A comparison with the publicly available code DisPerSE demonstrates the ability of 1-DREAM to recover axes well aligned with the centers of cosmic filaments.

**Key words.** Cosmology: large-scale structure – Cosmology: simulations – Methods: data analysis

## 1. Introduction

Large observational surveys repeatedly confirm that matter in the Universe is distributed in the form of a highly anisotropic and interconnected network known as the Cosmic Web (York et al. 2000; Jones et al. 2009; Lambert et al. 2020). The exploration of this web provides great insight into the nature of gravity, the cosmological processes governing the formation and expansion of the Universe, as well as the properties of galaxies belonging to its various parts. To study this network, multiple (semi-)automated numerical algorithms have been developed in such a way that the Cosmic Web's different properties are taken into account. In the work of Awad et al. (2023), we focus on the toolbox 1-DREAM first introduced in Canducci et al. (2022). The toolbox consists of five main algorithms that extract and construct probabilistic models of 1-dimensional astronomical structures embedded in an otherwise scattered or "noisy" distribution of particles. In Awad et al. (2023), we study how 1-DREAM can be applied on cosmological N-body simulations of the Cosmic Web. We also compare 1-DREAM's performance with previously developed techniques, with emphasis on the comparison with the public code DisPerSE (Sousbie 2011).

This document will be structured as follows: we give an overview of the different methodologies and their functionalities in Section 2, then discuss the toolbox's performance when compared to state-of-the-art methodologies in Section 3. Finally, we give a brief summary and conclude our work in section 4.

## 2. Explanation of the algorithms

1-DREAM consists of five Machine Learning methodologies that operate on a given distribution of particles. We illustrate the functionality of each algorithm by applying it on a dark matter only Λ-CDM cosmological simulation generated using the GADGET-3 code (Springel 2005) and assuming $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.047$ and $h_0 = 0.684$. A subset of the simulation is shown in panel (a) of Figure 1 containing two main clusters connected together by a filament and several fainter filaments surrounding them. We briefly name the algorithms and list their intended functions here, then explain each one in greater detail afterwards:

1. LAAT: enhances the density contrast between structures of varying density within the Cosmic Web.
2. EM3A: collapses the positions of particles belonging to a structure onto that structure's central axis.
3. DimIndex: distinguishes the different cosmic environments (clusters, walls, filaments, and voids).
4. MMCrawling: partitions the data into a set of filaments represented by their central axes and the group of particles surrounding them.
5. SGTM: performs the subsequent probabilistic modelling of the distribution of matter within cosmic filaments.

We use the *Locally Aligned Ant Technique* (LAAT) to initiate a random walk through the subset and distribute an artificially defined quantity termed the "pheromone" in reliance on the concept of Ant Colony Optimization (Dorigo & Stützle 2004, ACO). Running LAAT allows the pheromone to be concentrated on regions of high density and regions aligned with locally detected manifolds or structures. The pheromone quantity therefore enhances the contrast between low-density and high-density regions, and can be used as a threshold parameter to separate regions of varying density. Applying this threshold on the subset of Figure 1 produces the result shown in panel (b). The outcome of this procedure isolates the regions of highest density and filters out the remaining points. Note that LAAT can be re-applied on the remaining points to extract the fainter structures.

The second algorithm is *Evolutionary Manifold Alignment Aware Agents* (EM3A/EM3A+) which operates again by initiating a random walk through the data and moving the particles closer to their local manifold. Running this algorithm allows the particles to align along the central axis of the structures within the data. The outcome of this step when applied on the results of panel (b) in Figure 1 is what is drawn in black in the following panel (c). The particles shown in black are the new positions of the subset particles. The new positions demonstrate an

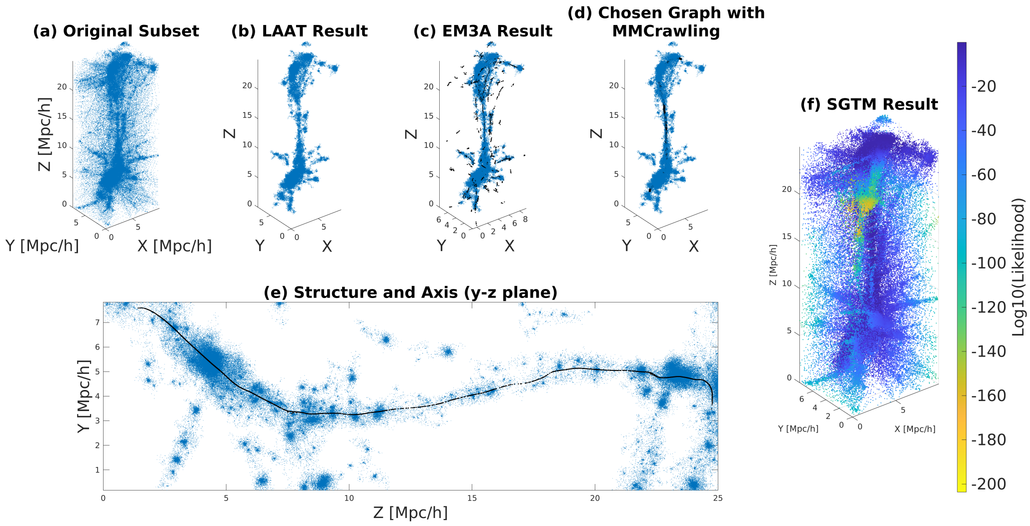


**Fig. 1.** Panel (a) shows a subset of the data containing two clusters connected together by a large filament and a distribution of surrounding particles. Panel (b) shows the significant structures extracted by LAAT. Panel (c) shows the central axes found by EM3A+. A different view of the central axis of the largest structure is also shown in panel (e). Panel (d) shows MMCrawling's graph representation (nodes and edges) of the longest recovered axis of panel (c). Panel (f) shows the result of modeling this structure's particle distribution using SGTM. The color-scheme corresponds to the calculated likelihood of each particle to belong to the structure where regions in blue correspond to high likelihoods. *Adapted from Awad et al. (2023).*

alignment with the central axis of the manifold they belong to. Panel (e) of the same figure portrays a projection on the $y - z$ plane of the resulting axis (in black) running through the largest structure i.e. the filament and two clusters (in blue).

The third and fourth algorithms, MMCrawling and Stream-Generative Topographic Mapping (SGTM), comprise the modelling function of the 1-DREAM toolbox. MMCrawling separates the data into a collection of central axes represented by graphs (nodes connected by edges), and the collection of particles surrounding them. The axis that runs through the main structure in Figure 1 is shown in panel (d) in black. SGTM then takes the results of MMCrawling and creates a model of the structure consisting of a collection of Gaussian probability distributions (Gaussian Mixture Model) that best fits the distribution of particles forming the structure. We then sample the position of the particles from the probabilistic model, and thus attribute to each particle a likelihood that it belongs to the modelled structure. The result is portrayed in panel (f) where regions of higher likelihood to belong to the structure are shown with darker colors. The purpose of the probabilistic modelling is to relax the idea that a structure begins and ends at a definite position. The model replaces this concept with a probability to belong to a structure instead. The line connecting centers of the trained Gaussian distributions can also be used as an axis to "crawl" along the structure and measure various physical properties pertaining to it.

Finally, Dimensionality Index (DimIndex) is the fifth algorithm of 1-DREAM, and is used to separate the particles of the cosmological simulation between those belonging to clusters, filaments, walls, and voids. This is performed by calculating the local dimensionality of the manifolds that the particles belong to. Since the separation of the environments is possible using several other Cosmic Web tracing algorithms, this provides grounds for comparison of performance with these methodologies. Further details about the comparison are discussed in the next section.

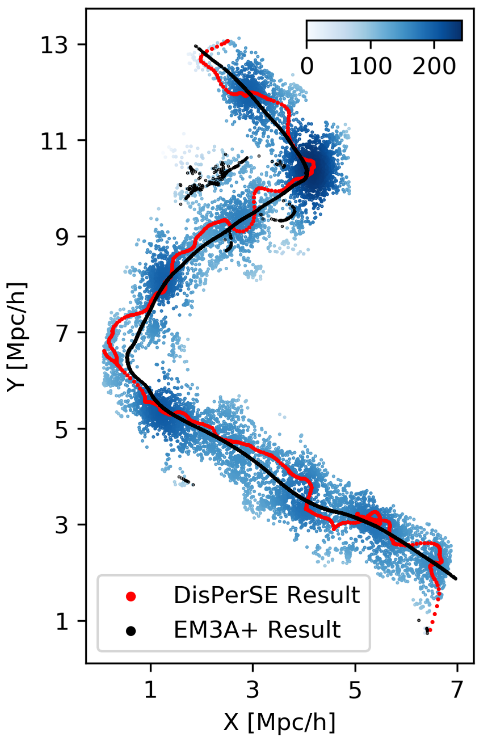


**Fig. 2.** Comparison between the axis retrieved by DisPerSE (red) and that retrieved by EM3A+ (black) on a simulated cosmic filament. The filament is shown in blue and projected onto the $x-y$ plane. Darker blue regions represent areas of high density, in contrast with light blue regions. *Adapted from Awad et al. (2023).*

## 3. Comparison with other tools

The work of Libeskind et al. (2018) has provided a cosmological Dark Matter-only N-body data set on which a collection of Cosmic Web tracing algorithms attempt to separate the Cosmic Web's different environments. Since our toolbox's DimIndex algorithm performs a similar functionality, we produce our classification of the Dark Matter particles in the simulation between those belonging to clusters, filaments, walls, and voids. This then allows us to perform the same statistical procedures followed in Libeskind et al. (2018), and therefore compare the results we get to those obtained by other methodologies. Through comparing our calculated probability density functions, halo mass functions, and volume/mass fractions as a function of the environment, we find that 1-DREAM is capable of splitting the different environments of the Cosmic Web with results comparable to the well known methods such as CLASSIC (Kitaura & Angulo 2012), ORIGAMI (Falck et al. 2012), and MMF (Aragón-Calvo et al. 2007).

A more detailed comparison is performed with the code DisPerSE Sousbie (2011). We extract a filament from the cosmological simulation and trace its central axis using our toolbox's EM3A+ and using DisPerSE. We then compare the axes retrieved by both tools. In Figure 2, the extracted filament is shown in blue where regions in darker blue are denser than regions in lighter blue. The axis retrieved by EM3A+ is shown in black, while the axis retrieved by DisPerSE is shown in red. We observe that the axis retrieved by DisPerSE contains a lot of twists and bends whereas EM3A+'s axis follows the center of the filament more straightforwardly. For studies that look at properties of galaxies as a function of distance to the cosmic filaments, it is important to have a good reference to the structure. Therefore, axes that contain a lot of bends are not ideal for such studies, and therefore EM3A+ would be a more applicable tool in this case. If however, what is required is to have axes that trace the local densities of the filaments, then DisPerSE would be the better tool of choice.

## 4. Conclusion

We introduce the toolbox 1-DREAM as a novel set of methodologies capable of studying filaments within cosmological N-body simulations of the Cosmic Web. The toolbox consists of five novel Machine Learning methodologies that have been grouped together in Canducci et al. (2022) and serve the extraction and probabilistic modeling of structures within N-body simulations. We demonstrate the different functionalities of the algorithms on an example filament extracted from the cosmological simulation. We also apply 1-DREAM on the data sets provided in Libeskind et al. (2018), and find that our toolbox is capable of separating the different environments of the Cosmic Web in a comparable manner to state-of-the-art methodologies such as CLASSIC (Kitaura & Angulo 2012), ORIGAMI (Falck et al. 2012), and MMF (Aragón-Calvo et al. 2007). Finally, a more detailed comparison is performed with the code DisPerSE (Sousbie 2011) where we show that 1-DREAM retrieves central axes more aligned with the center of the simulation's filaments. Given this analysis, we propose the application of 1-DREAM on studies of local and global properties of cosmic filaments on simulation data sets and available observational data including surveys such as the 2MASS Redshift Survey (2MRS) (Macri et al. 2019; Lambert et al. 2020).

*Acknowledgements.* This work is supported by the DSSC Doctoral Training Programme of the University of Groningen. We thank the students Joel During and Jeroen Klooster for making the algorithms suggested in this work more user-friendly.

## References

Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007, A&A, 474, 315

Awad, P., Peletier, R., Canducci, M., et al. 2023, Monthly Notices of the Royal Astronomical Society, 520, 4517

Canducci, M., Awad, P., Taghribi, A., et al. 2022, Astronomy and Computing, 41, 100658

Dorigo, M. & Stützle, T. 2004, Ant colony optimization (MIT Press)

Falck, B. L., Neyrinck, M. C., & Szalay, A. S. 2012, ApJ, 754, 126

Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683

Kitaura, F.-S. & Angulo, R. E. 2012, MNRAS, 425, 2443

Lambert, T. S., Kraan-Korteweg, R. C., Jarrett, T. H., & Macri, L. M. 2020, MNRAS, 497, 2954

Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2018, MNRAS, 473, 1195

Macri, L. M., Kraan-Korteweg, R. C., Lambert, T., et al. 2019, ApJS, 245, 6

Sousbie, T. 2011, MNRAS, 414, 350

Springel, V. 2005, MNRAS, 364, 1105

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579