ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Hub-Aware Hybrid Search: Accelerating the Locally Aligned Ant Technique

Simone Vilardi[1], Reynier Peletier[2], Felipe Contreras[1,3] and Kerstin Bunte[1]

1- University of Groningen - Bernoulli Institute
for Mathematics, Computer Science and Artificial Intelligence

2- University of Groningen - Kapteyn Astronomical Institute

3- Instituto de Física y Astronomía, Universidad de Valparaíso

Abstract. Finding manifold structures in noisy and high-dimensional point clouds is a challenging but important problem. In astronomical observation survey and simulation data the detection of filaments, streams (1D), walls (2D) and clusters (3D) gives rise to deeper understanding of the evolution of our universe. The Locally Aligned Ant Technique (LAAT) uses biologically inspired agents to efficiently recover faint and multidimensional structures. However, very dense hubs (e.g. nodes or globular clusters) dominate the ants' activity, creating unnecessary computational overheads. In this paper we propose a two-stage solution. First a fast preprocessing step locates the hubs and replaces them with a tailored likelihood model. Subsequently, a mixed likelihood-pheromone strategy guides the ants to efficiently bridge the dense regions. We demonstrate improvements in detection efficiency and robustness of LAAT with synthetic and a large-scale astronomical N-body simulation of the cosmic web.

1 Introduction

To understand the large-scale structure of the Universe, it is paramount to identify filaments, streams, and clusters within the cosmic web, which are difficult to detect due to noise and high dimensionality in astronomical surveys and N-body simulations.[1–4] The 1-DREAM pipeline was developed to detect and model one-dimensional manifolds in such settings [5, 6], with LAAT as its first module for identifying likely manifold members via an ant-colony optimisation scheme combining local geometry and pheromone reinforcement [5]. However, real datasets often contain very dense hubs where ants rapidly accumulate, causing these clusters to dominate pheromone accumulation and subsequent thresholding. This wastes computation resources and hampers the desired one-dimensional manifold recovery. Recent extensions proposing a dynamic radius [7, 8] improved the noise handling and user-friendliness of LAAT, but the ants still spent a significant amount of steps (time) in dense regions. We present a revised LAAT that mitigates hub dominance[1] by: (1) identifying dense regions, (2) fitting a likelihood model to each region, (3) running ants on the combined likelihood-point cloud domain, and (4) introducing a modified transition probability incorporating those models to guide the ants towards fainter structures. These modifications retain the alignment-and-pheromone mechanism of LAAT, while avoiding the repeated exploration of dense hubs. This improves the recovery of filamentary manifolds in challenging cosmological datasets.

[1]code publicly available at `https://git.lwp.rug.nl/cs.projects/Hub-LAAT`

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 Methodology

LAAT guides stochastic agents across a dataset using pheromone deposition and evaporation schemes inspired by biology. Given a dataset $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N$, for each point, $\boldsymbol{x}_i \in \mathbb{R}^D$ we define a spherical neighbourhood $\mathcal{N}_r^{(i)}$ of radius r and compute local eigen-vectors and -values $\{(\boldsymbol{v}_d, \lambda)_d)\}_{d=1}^D$ within it. LAAT drives ants with two preferences: (i) Alignment of $(\boldsymbol{x}_j - \boldsymbol{x}_i)$ with $\boldsymbol{v}_d$ represented by angle $|\cos\alpha_d^{(i,j)}|$ and weighted by λ_d, thus $A^{(i,j)} = \sum_{d=1}^D \frac{|\cos\alpha_d^{(i,j)}|}{\sum_{d'}|\cos\alpha_{d'}^{(i,j)}|} \cdot \frac{\lambda_d}{\sum_{d'}\lambda_{d'}}$, and neighborhood relative $\bar{A}^{(i,j)} = \frac{A^{(i,j)}}{\sum_{j'\in\mathcal{N}_r^{(i)}} A^{(i,j')}}$. (ii) And relative accumulated pheromone $\bar{F}^j(t)$ at time t at point $\boldsymbol{x}_j \in \mathcal{N}_r^{(i)}$, that can evaporate on rarely visited points (see [5]). Both combined result in a movement preference from $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$ of an ant at time t with contributions controlled by $\kappa \in [0, 1]$:

$$V^{(i,j)}(t) = (1-\kappa)\bar{F}^j(t) + \kappa\bar{A}^{(i,j)}. \tag{1}$$

The transition probabilities are viewed as negative "energies"

$$P(j \mid i, t) = \exp\left(\beta\, V^{(i,j)}(t)\right) / \sum_{j'\in\mathcal{N}_r^{(i)}} \exp\left(\beta\, V^{(i,j')}(t)\right), \tag{2}$$

with inverse temperature β. Here j' indexes points in the local neighbourhood $\mathcal{N}_r^{(i)}$ of point i. To allow efficient parallelization with multiple ants, the pheromone is updated in epochs encompassing a number of ant steps. The evolutionary process is linear with the number of points, with a one time squared preprocessing cost including local PCA [5]. The presence of dense hubs induces two issues: (1) pheromone accumulates, making them local attractors and reducing visitations in fainter structures, and (2) structure members are detected by pheromone thresholding, which tends to keep a significant amount of points in these small dense regions, slowing down any subsequent analysis unnecessarily.

2.1 Hub identification and modelling

Our approach is designed for efficiency, identifying dense regions to reduce computational burden in the subsequent evolutionary computation. In the special case of $\kappa = 1$ in (1) the ants ignore the pheromone and transition probabilities are independent of t and stay constant with $P_{ij} = P(j|i)$. It can be considered a Markov Chain (MC) and studied as stationary distribution of the point cloud (several if isolated). It is independent of the starting point and will converge to a steady-state vector $\boldsymbol{\pi}$, the dominant left eigenvector of the transition matrix containing the visitation frequency of each data point. We compute this eigenvector using the Power method and use the thresholding strategy proposed in [7] to obtain high visitation values indicating dense regions. Concretely, for every particle $\boldsymbol{x}_i$ we sort its neighbours $\boldsymbol{x}_j \in \mathcal{N}_r^{(i)}$ in ascending order of visitation score $s_j = \pi_j$, obtain the minimum (maximum) value $s_{\min}$ ($s_{\max}$), and smooth the distribution by fitting a cubic spline $f(s)$. The first inflection point s_0 is found by differentiation and marks a steep change. A parameter η defines

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

a local threshold T_i, more aggressive or conservative:

$$T_i = \begin{cases} s_0 + (s_{\max} - s_0)\,\eta, & 1 \geq \eta \geq 0, \\ s_0 + (s_0 - s_{\min})\,\eta, & -1 \leq \eta < 0. \end{cases} \tag{3}$$

A minimum threshold is enforced as a fraction of the average visitation $T_i \geq \frac{\psi}{N}\sum_{j=1}^{N}\pi_j$. We then cluster high-visitation points via the friends-of-friends algorithm with linking length $l = b(\mathcal{V}/N)^{1/D}$, where $\mathcal{V}$ is the data volume (or surface) with free parameter $b \in [0.1, 1]$. Alternatively l can be chosen as LAAT neighbourhood radius r. We only retain hubs containing at least $N_{\min}$ points.

After detection we fit a likelihood model to each of the K hubs, forming the basis for the new hybrid ant-movement scheme. We model the distribution using a Bayesian Gaussian mixture with a Dirichlet process prior[2]. The fitted BGM yields a continuous likelihood field, which we divide into High-Likelihood Points (HLPs) or Low-Likelihood Points (LLP) to remove the dense central region and keep a transition region to lower density outskirts. Although hubs are general, we simplify using nested equal-volume Mahalanobis shells centred at the maximum likelihood point of the k's model $\boldsymbol{c}_k$ to approximate likelihood level sets. Let N_p denote the number of particles enclosed up to shell ℓ with radius r_ℓ, N_{hub} the total number of hub particles and the enclosed fraction $R = N_p/N_{\text{hub}}$. For growing shells $\mathcal{S}_\ell = \{\, \boldsymbol{x}_j : r_{\ell-1} < \|\boldsymbol{x_j} - \boldsymbol{c}_k\| \leq r_\ell \,\}$, we compute a stopping probability $p_{\text{stop}} = {(e^{\alpha/R} - 1)}/{(e^{1/R} - 1)}$ with $\alpha = \frac{R-0.5}{0.5}$, which increases as a larger fraction of the hub has been enclosed. A random draw $u \sim U(0,1)$ terminates the shell expansion when $u < p_{\text{stop}}$. The median likelihood of the final shell defines the hub threshold T_k separating HLP ($> T_k$) from LLP ($\leq T_k$). Removing the HLP from the data before exploration prevents any future ant visitation, reducing computational time and memory requirements.

2.2 New ant transition probabilities

All remaining data points $\boldsymbol{x}_i$ carry likelihood information for the K modelled hubs $\mathcal{L}_k^{(i)}$ with $k = 1 \ldots K$. The friends-of-friends clustering provides a crisp assignment so each point has at most one likelihood value (extension to fuzzy memberships are possible). For points that do not belong to any hub the original LAAT transition probabilities (2) are used. When the current point $\boldsymbol{x}_i$ has a valid hub assignment k, we replace the old transition using the hub's LLPs as follows. For a neighbour $\boldsymbol{x}_j \in \mathcal{N}_r^{(i)}$ we define the likelihood difference term $\Delta_k^{(i,j)} = \left|\mathcal{L}_k^{(i)} - \mathcal{L}_k^{(j)}\right|$, and the likelihood-based transition probability becomes

$$P(j \mid i) = \frac{\Delta_k^{(i,j)}\, \mathcal{L}_k^{(j)}/T_k}{\sum_{j' \in \mathcal{N}_r^{(i)}} \Delta_k^{(i,j')}\, \mathcal{L}_k^{(j')}/T_k}. \tag{4}$$

This probability favours target neighbours that both differ strongly in likelihood from $\boldsymbol{x}_i$ and lie in high-likelihood regions, thus directing ants towards sharp likelihood gradients. This transition alone would steer the ants to surround the

[2] `BayesianGaussianMixture` in `https://github.com/scikit-learn/scikit-learn`.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

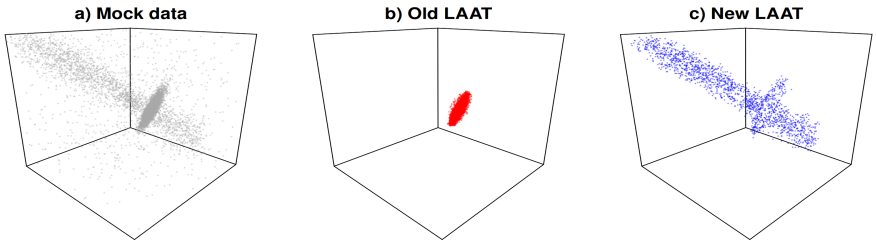

Fig. 1: a) Mock data. b) and c) Old and New LAAT outputs.

cavity. To allow faster exploration, we propose a double jump strategy with controlled repulsion. For the target $\boldsymbol{x}_j$ from $\boldsymbol{x}_i$ (4) we compute the Euclidean distance δ_{ij}. With a probability $P_{\text{2nd}}(j) = \exp(-\delta_{ij}/r)$ a 2nd jump is triggered to a random point $\boldsymbol{x}_{j'}$ in the LLP set of model k, effectively expanding the neighbourhood $\mathcal{N}_r^{(i)}$ to the model shell LLP_k. A longe-range repulsion follows the 2nd jump to increase the chance to leave the hub. Concretely, a shell is formed centred at $\boldsymbol{x}_{j'}$ with potential target points $\boldsymbol{x}_m$ at a distance $\gamma_{\min} r \leq \delta_{j'm} \leq \gamma_{\max} r$. The parameters $\gamma_{\min}$ and $\gamma_{\max}$ control the repulsion range relative to the LAAT parameter radius r. The transition probability for repulsion $P_{\text{rep}} = \exp(-\mathcal{L}_k^{(m)}/\mathcal{L}_k^{(j')})$, favours moves towards lower-likelihood points, that connect the hub to fainter structures, effectively enabling the ants to leave the hub.

3 Experiments and Discussion

We compare the new algorithm with the original LAAT on a synthetic dataset with known ground truth components, to measure component recovery directly. Then we study the new parameters sensitivity using a second cosmic-web-like mock dataset and demonstrate its scalability on a $50^3\,\text{Mpc}^3/h$ cosmic web N-body simulation cube containing 2.8×10^5 particles. Throughout the experiments, we fix $\beta = 10$ in the MC transition probabilities tolerance $\varepsilon = 0.1$ and a maximum of 100 steps in the Power Iteration. We model each Bayesian Gaussian mixture with maximal 10 components, full covariances and a uniform concentration prior. We set ant-transition parameters $\gamma_{\min} = 1.5$ and $\gamma_{\max} = 2.5$.

Synthetic data: the sample contains 8×10^3 particles: 64% in the dense hub, 26% in the filament, and 10% noise. For the old method we use 100 ants, 1000 epochs, and 10000 steps with $r = 0.25$. For the new one, the new hyperparameters (ψ, η, b) are set to $(0.005, -0.2, r)$. To ensure a fair comparison, we estimated, by wall-clock CPU timing, that the new preprocessing steps weight around 40 epochs of 10000 steps on this dataset. As a consequence, we leave the number of ants and steps unchanged, and reduce the number of epochs to 960. Results are shown in Figure 1 and Table 1 reporting the recovery fractions of each component and demonstrate the significant improvement in filamentary component recovery achieved by the new algorithm.

Table 1: Component recovery.

	OLD	**NEW**
Dense hub	88%	5%
Filament	5%	72%
Noise	1%	3%

Parameter analysis: the algorithm behaviour is primarily governed by the preprocessing parameters, with all others fixed due to their weaker empirical impact.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

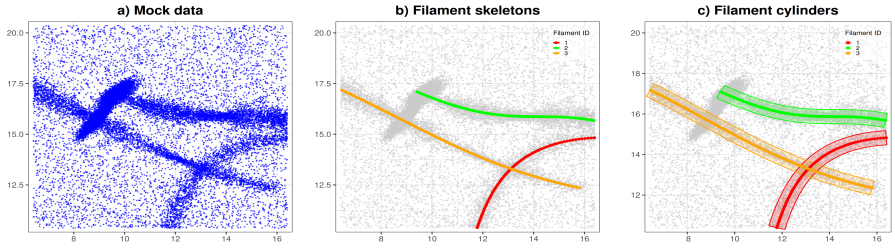

Fig. 2: a) Mock data. b) Filament skeletons. c) Tubes to select candidate filament points for the analysis.

The parameter b controls the typical spatial extent of detected hubs, while ψ and η regulate the partition into dense and non-dense points, balancing speed and completeness. Lower values classify more points as dense, risking removal of faint structures, whereas higher values make hub detection stricter and limit hub usage. To study this trade-off, we use a synthetic, cosmic web-like point cloud comprising two high-density Gaussian nodes, three filaments of varying length, shape and density, and uniform background noise (totalling around $2.4 \cdot 10^4$ points). We perform a sensitivity analysis of the parameters similar to [5]. Specifically, we alter $\psi \in [0.0005, 0.5]$ and $\eta \in [-1, 1]$, and fix the remaining parameters to 100 ants, 100 epochs, 5000 steps, $r = 0.5$, and $b = 0.2$. Figure 2 shows the dataset with filament skeletons and tube regions to select points at a fixed radial distance ($r = 0.35$), corresponding to $\sim 70\%$ of all filament points. As in Taghribi et al. [5] we evaluate the recovery of the filament using the average Hausdorff distance (AHD) [9]. We focus specifically on filaments with comparisons made between the union of true filament points within the tubes in Figure 2 and LAAT-selected points constrained to the same regions (Table 1).

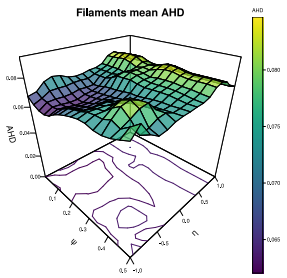

Fig. 3: Global mean AHD as a function of the hyperparameters ψ and η.

As shown in Figure 3, increasing either ψ or η makes hub detection more restrictive, reducing filament completeness and increasing the average Hausdorff distance. The best overall recovery for this dataset is obtained for $(\psi, \eta) = (0.1, -1)$, which corresponds to the lowest mean AHD.

cosmic web: The new Hub-LAAT significantly improves scenarios involving multiple dense regions, as they guide the ants more efficiently, accelerating and substantially improving the overall exploration process, as illustrated in Figure 4. Panels (b) and (c) show, the top 10^5 particles with the highest pheromone values obtained with the old and the new LAAT algorithms, using identical settings (100 ants, 100 epochs, 10^4 steps and $r = 1$). Hub-LAAT hyperparameters (ψ, η, b) were set to $(0.005, -0.6, r)$. Panel (d) displays the top 10^5 particles from Hub-LAAT (c) with an additional number of particles equalling those removed during preprocessing ($\sim 3.7 \times 10^4$).

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

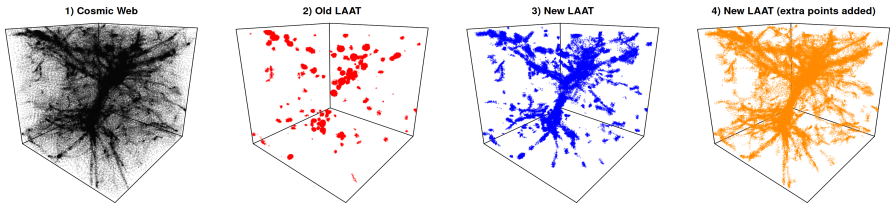

Fig. 4: a) cosmic web data. b) and c) Top 10^5 particles with highest pheromone levels from original and new LAAT. (d) Top 10^5 particles plus $\sim 3.7{\cdot}10^4$ particles (matching the number new LAAT preprocessing removed).

Lacking ground truth, the evaluation is necessarily qualitative.

4 Conclusions and future work

In this paper, we propose a two-stage modification to LAAT: fast hub detection with a tailored likelihood model, followed by mixed likelihood-pheromone ant transition probabilities. These extensions mitigate hub-induced overconcentration, reduce computational and memory overhead, accelerate the recovery of filaments and streams, and improve robustness in noisy, high-dimensional cosmological datasets. In future work we will develop a temporal version of the methodology to model the dynamic evolution of galaxy structures over time.

Acknowledgments

Funded by the European Union (MSCA EDUCADO, GA 101119830). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] J. R. Bond, L. Kofman, and D. Pogosyan. “How filaments of galaxies are woven into the cosmic web”. In: *Nat.* 380.6575 (1996), pp. 603–606.

[2] M. Cautun et al. “Evolution of the cosmic web”. In: *MNRAS* 441.4 (2014), pp. 2923–2973. DOI: 10.1093/mnras/stu768.

[3] V. Springel. “The cosmological simulation code GADGET-2”. In: *MNRAS* 364.4 (2005), pp. 1105–1134. DOI: 10.1111/j.1365-2966.2005.09655.x.

[4] J. Schaye et al. “The EAGLE project: simulating the evolution and assembly of galaxies and their environments”. In: *MNRAS* 446.1 (2015), pp. 521–554.

[5] A. Taghribi et al. “LAAT: Locally Aligned Ant Technique for Discovering Multiple Faint Low Dimensional Structures of Varying Density”. In: *IEEE Trans. Knowl. Data Eng.* 35.6 (2023), pp. 6014–6027. DOI: 10.1109/TKDE.2022.3177368.

[6] M. Canducci et al. “1-DREAM: 1D Recovery, Extraction and Analysis of Manifolds in noisy environments”. In: *Astron. Comput.* 41 (2022), p. 100658.

[7] F. Contreras, K. Bunte, and R. Peletier. “Improved LAAT strategy to recover manifolds embedded in strong noise”. In: *31st ESANN*. 2023, pp. 71–76.

[8] F. Contreras, K. Bunte, and R. Peletier. “Adaptive Locally Aligned Ant Technique for Manifold Detection and Denoising”. In: *33rd ESANN*. 2025, pp. 111–116.

[9] M.-P. Dubuisson and A. Jain. “A modified Hausdorff distance for object matching”. In: *Proc. of the 12th Int. Conf. Pattern Recognit.* Vol. 1. 1994, 566–568 vol.1. DOI: 10.1109/ICPR.1994.576361.