The persistent cosmic web and its filamentary structure: (I) Theory, Implementation[1] and (II) Illustrations[2]

T. Sousbie, C. Pichon and H. Kawahara

and summarised by : Utkarsh Detha

(Dated: March 6, 2019)

This paper provides the essential background for DisPerSe, a software used for the identification of large scale structure in the universe by working directly on discrete distribution of galaxies. The distinguishing feature of the method is the usage of persistence based identification to deal with the spurious identifications introduced by the shot noise.

I. MORSE THEORY

The goal of DisPerSe is to identify topological features in density fields described by the galaxy distributions. Morse theory which relates the geometrical and topological properties of a smooth scalar function, is a promising candidate for achieving the said goal. This theory is applicable on a special class of functions called Morse functions, which are essentially smooth functions, that have non-degenerate critical points (a k-order critical point is a critical point for which the Hessian matrix has k negative eigenvalues; non-degeneracy here means that there are only non-zero eigenvalues for the Hessian). The additional requirement for non-degeneracy leads to an important property for field lines: they do not intersect and that they only originate and terminate at critical points. For such Morse functions, it is possible to partition the space into manifolds: regions of space that are occupied by integral lines originating or terminating at their critical points. For example, the region of space reached by integral lines originating from a k-order critical point P, is called the (d-k) order ascending manifold of P, and the set of field lines reaching P define the k-order descending manifold, where d is the dimension of the space. Furthermore, the intersection of these manifolds defines regions that are covered by field lines that all share their origin and terminals. These regions are called Morse-Smale n-cells (where n is the dimension of the cell) and the set of all the Morse-Smale cells is collectively known as the Morse-Smale Complex.

A. Discrete Morse Theory

Real-life observations are almost exclusively discrete and thus we need a discrete analogue for Morse theory. Discrete Morse theory is defined over *simplicial complexes*. A ***k-simplex*** (σ_k) is the convex hull of $k+1$ affinely independent points, in other words it is the smallest possible solid with the given set of points for vertices. Moreover, *faces* of a simplex σ_k are defined as simplices that have vertices that are a subset of the vertices of σ_k which is in turn the *coface* of its faces. A simplicial complex is defined as a finite union of simplices such that any face of a simplex in the complex also belongs to the complex and the intersection of any two simplices is either empty or a simplex with dimension lower than or equal to the highest dimension simplex in the initial pair.

A function defined over this complex would assign a value to every simplex in the complex. The discrete analogue of a gradient field is defined by coupling simplices in gradient arrows: if a simplex σ_k has a lower valued cofacet α_{k+1} then they form an arrow $[\sigma_k, \alpha_{k+1}]$ or if σ_k has a higher valued facet β_{k-1} then the arrow would be $[\sigma_k, \beta_{k-1}]$. Now in analogy with a smooth Morse function, we define a Discrete Morse function as a function which assigns values to the simplices over the complex such that there is at most one small coface and at most one greater face for each simplex. This removes any ambiguity in the direction of the gradient at any simplex (analogous to a point in the continuous case). Moreover, a critical simplex is one which has no higher faces and no lower cofaces. Now, just like in the continuous case, we can define *field lines* as gradient arrows arranged end-to-end such that the simplices are in a decreasing order. The analogy can then be extended to define the ascending and descending manifolds and eventually, a *Discrete Morse Complex.*

B. Delaunay tessellation field estimator (DTFE)

To convert a distribution of points into a density field, we start by taking the Delaunay triangulation of the points. This divides the space into simplices with vertices being points from the distribution and such that the circum-hypersphere of each d-simplex includes no other points from the distribution. This gives us a simplicial complex. Now to assign the values of density, we attribute each point a density that is inversely proportional to the volume of the dual Voronoi cell of that point. Thus, for regions with fewer points, we get big Voronoi cells and therefore lower densities.

II. TOPOLOGICAL PERSISTENCE

Topological persistence is a way of establishing the importance of topological features in a function. The basic idea is to set a threshold and the points that have the value of the function greater than the threshold form what is called an *Excursion set*. As the value of the threshold is changed, new ***structures*** will form (or get destroyed) everytime a critical point enters the excursion set. The persistence is then defined as the difference between the values of the function for a pair of critical points that is responsible for the creation and destruction of a structure. Persistence is therefore an estimate of the *longevity* of the structure.

To formalize the idea of structures, we identify them as topological features, called *k-cycles*. For a 3-d function, as we decrease the threshold, we get isolated islands around the maxima. These are called *components* or 0-cycles. As the threshold goes lower, the islands merge together at the saddle points of type 1, destroying the isolated components and forming rings (around holes) and these are the 1-cycles. For lower thresholds, the holes get filled up at saddle points of type 2, destroying the 1-cycles and forming spherical shells (2-cycles) around minima. Eventually the 2-cycles also get filled up as we reach the global minimum.

Since we work in the discrete case, we need a discrete analogue of the above-mentioned ideas. For a simplicial complex K, we can apply the idea of an excursion set by studying the evolution of its ***filtration***, which is a sequence of subcomplexes such that $\emptyset = K^0 \subseteq K^1 \subseteq ...K^N = K$ and $K^{i+1} = K^i \cup \delta^i$ (δ^i is a different set of simplices of K). Here, as the threshold changes, new simplices enter the filtration and the persistence pairs the critical simplices (σ_a, σ_b) that create and destroy a given feature, their corresponding persistence being defined by $\rho_D(\sigma_a) - \rho_D(\sigma_b)$, D is the dimension of the complex.

A. Topological simplification

It is always possible to locally modify the function so as to cancel non-persistent pairs, thereby eliminating topological noise. If one modifies the function such that the values of the function at critical points that form a persistence pair and are neighbours, are interchanged, then both points will cease to be critical points. Thus we would have eliminated that topological feature from the Morse complex, without affecting the rest of the features. This is the central idea behind topological simplification.

B. Filtering Poisson noise

It is not possible to use the DMC computed over a distribution directly, the main reason being that there is a huge number of spurious detections that are caused by the sampling noise. Mainly due to the scale-free nature of DTFE, it is very sensitive to Poisson noise. Now, consider a persistence pair $q_k = [\sigma_{k+1}, \sigma_k]$ then the persistence ratio is given by $r(q_k) = \frac{\rho_D(\sigma_{k+1})}{\rho_D(\sigma_k)}$. Also note that $P_k(r_0)$ is the cumulative probability that a persistence pair with a ratio $r \geq r_0$ exists in the Delaunay tessellation of a random discrete Poisson distribution. It is then convenient to express the importance of a persistence pair as the significance, (expressed in units of σ in analogy with Gaussian distribution)

$$S(r(q_k)) = \text{Erfc}^{-1}\left(\frac{P_k(r(q_k)) + 1}{2}\right)$$

where Erfc is the error function.

To estimate $P_k(r)$, a Monte carlo simulation was used, measuring $P_k(r_0)$ as the fraction of persistence pairs of order k with the persistence ratio $r \geq r_0$ in a Poisson sample. From these results, fits for the cumulative probabilities of the different orders of persistence pairs were obtained (they can be found in the original paper [1]: eq (5) - eq (9)). The probabilities are plotted w.r.t. the persistence ratio (figure 1).

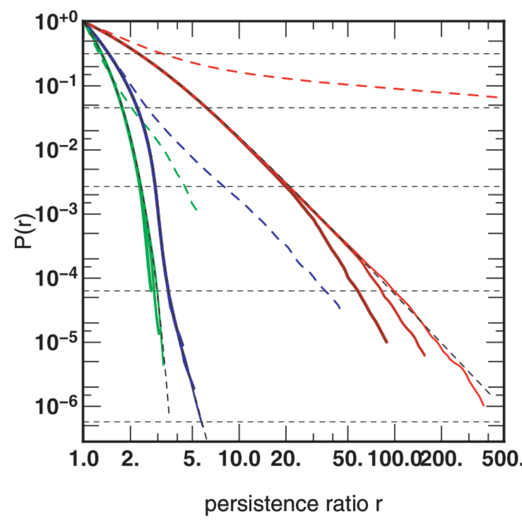

FIG. 1. The cumulative probability $P_k(r)$ vs r for a 3D scale-free Gaussian random field (coloured plain curves) and in a 50 h^{-1} Mpc dark matter cosmological simulation (coloured dashed curves). The red, blue and green curves correspond to (maxima,1-saddle), (1-saddle,2-saddle) and (2-saddle,minima) pairs, respectively. The different shades, from darker to lighter, correspond to 64^3-, 128^3- and 192^3-particle realizations, respectively. The black-dashed curves show fits to the Gaussian case while the horizontal dashed lines correspond to different significance levels, ranging from $S = 1\sigma$ to 5σ.

An important thing to note here is that the CDF of the $k = 1$ persistence pairs undergoes a transition near the significance level of 3.5σ, this is expected because of the nature of DTFE itself: due to the larger size of the Voronoi cells of minima, the number of these minima is naturally much lower than other critical points. This leads to the probability distribution function being biased towards higher densities. For a 2-d distribution obtained by an N-body simulation, figure 2 shows the DMC with no cancelling of persistence pairs, and a DMC with a significance level of 4σ. Simply by visual comparison of the DMC's with the original distribution, one can tell that the 4σ DMC traces the filaments very well, without many spurious filaments.

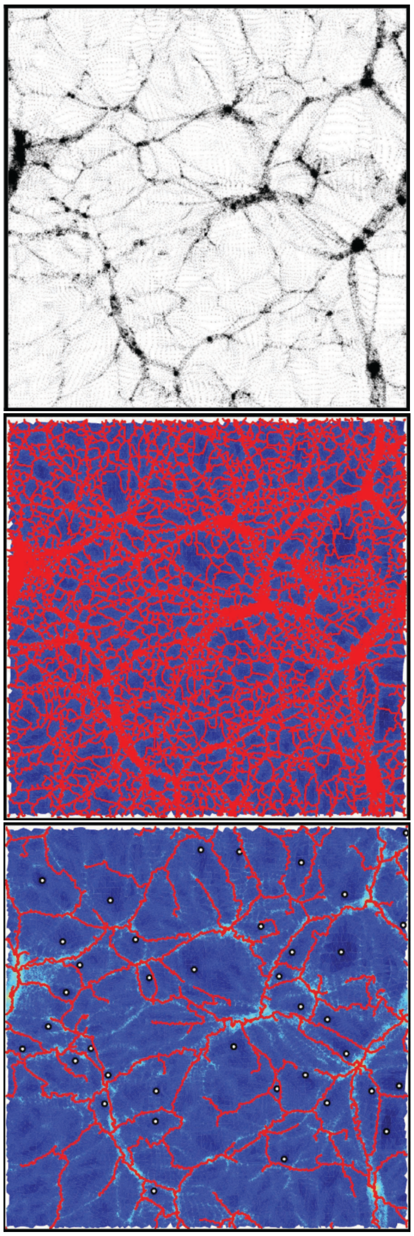

FIG. 2. The first image shows the original 2-d distribution and the second and third images show the 0 and 4σ cases, respectively.

It is, however, crucial to determine the appropriate significance level. If one chooses a level that is too low, then obviously there will be many spurious detections, while a significance level that is too high will ignore the weaker filaments (and other structures). To determine the appropriate significance level for the persistence pairs of various orders, it is useful to see the plot of number of k-order persistence pairs vs the signigicance levels (figure 3) for three scenarios: first is a cosmological simulation, second is the superposition of the same simulation and a random distribution of particles, and third is purely a random distribution of particles. It can be seen that when the significance threshold is above 3σ, the shape of the currve for the persistence pairs of type P_1 and P_2 (i.e. [2-saddle point,1-saddle point] pairs, green curves, and [1-saddle point,maxima] pairs, red curves, respectively) in $S_N^{2\times128}$ and S^{128} become similar. This suggests that the main source of persistence pairs is the original distribution and few structures arise out of the noise. For the case of type 0 persistence pairs though (i.e. [minima,2-saddle point] pairs), we see that $S_N^{2\times128}$ follows the behaviour of S_R^{128} more closely. Even so, the number of persistence pairs from the $S_N^{2\times128}$ is always higher. The reason for this behaviour is that the number of minima coming from physical processes is relatively low compared to the minima introduced by the noise alone: which is in turn caused by the scale-free nature of DTFE, as stated before. This tendency of the 0-order persistence pairs warns us against choosing a significance that is too high, if one wants to select a complete DMC.

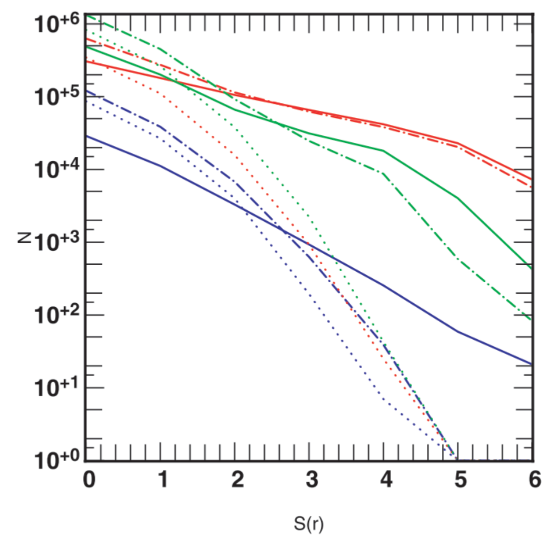

FIG. 3. Number of persistence pairs of type k as a function of the significance threshold $S_k(r)$ (in units of σ) in a $250h^{-1}$ Mpc large ΛCDM simulation downsampled to 128^3 particles, S^{128} (filled curves), the same distribution with 128^3 additional randomly located particles, $S_N^{2\times128}$ (dot-dashed curve) and a random distribution of particles within the same volume, S_R^{128} (dotted curves). The blue, green and red colours correspond to persistence pairs of type 0, 1 and 2, respectively.

III. CONCLUSION

It can be seen that the persistence based detection of cosmological structures is an effective strategy to eliminate spurious detections. To re-confirm the detected structures, Sousbie et. al. [2] also compare the detection of maxima (which correspond to dark matter haloes) to previously determined haloes based on the FOF algorithm, to excellent agreement. Moreover, to further cement the ability of this technique, an *optically faint* cluster which was expected at the junction of several filaments by DisPerSe was indeed observed by the X-ray satellite *Subaru.* Thus, Soubie et. al. have demonstrated an elegant technique to detect cosmological strutures, of every order.

[1] Sousbie T., *The persistent cosmic web and its filamentary structure- I. Theory and Implementations*, 2011, MNRAS, in press (doi:10.1111/j.1365- 2966.2011.18394.x)

[2] Sousbie T., Pichon C. and Kawahara H., *The persistent cosmic web and its filamentary structure- II. Illustrations*, 2011, MNRAS, in press (doi: 10.1111/j.1365-2966.2011.18395.x)