

Morpho-statistical characterization of the cosmic web using marked point processes

Radu S. Stoica

Université Lille 1 - Laboratoire Paul Painlevé
Observatoire de Paris - Institut de Mécanique Céleste et de Calcul des Éphémérides
email: radu.stoica@univ.lille1.fr

Abstract. The cosmic web is the intricate network of filaments outlined by the galaxies positions distribution in our Universe. One possible manner to break the complexity of such an elaborate geometrical structure is to assume it made of simple interacting objects. Under this hypothesis, the filamentary network can be considered as the realization of an object or a marked point process. These processes are probabilistic models dealing with configurations of random objects given by random points having random characteristics or marks. Here, the filamentary network is considered as the realization of such a process, with the objects being cylinders that align and connect in order to form the network. The paper presents the use of marked point processes to the detection and the characterization of the galactic filaments.

Keywords. methods: statistical, large-scale structure of universe

1. Introduction

The marked point processes were already used in cosmology to describe the galaxies distribution and the different correlations of their corresponding characteristics (Martinez & Saar 2002). The galaxies are not spread uniformly within the observed Universe. Their positions form an intricate network made of complex geometrical objects such as filaments, walls and clusters. During the last years, the filamentary network was studied under the assumptions that the filaments are made of small segments (cylinders) that connect and align forming the network. From a mathematical point of view, this hypothesis allows to consider the cosmic network as the realisation of a marked point process. Under this assumption, the network was detected and characterized from both, morphological and statistical, points of view (Stoica *et al.* 2005b; Stoica *et al.* 2007; Stoica *et al.* 2010; Tempel *et al.* 2013; Tempel *et al.* 2014).

2. Modelling the filamentary network

Let the observation window be a measure space (K, \mathcal{B}, ν) , with $K \subset \mathbb{R}^d$, \mathcal{B} the Borel σ -algebra and $0 < \nu(K) < \infty$ the Lebesgue measure. A point process in K is a finite random configuration of points leaving in K . To each point, marks or characteristics may be attached. For this purpose, the marks probability space (M, \mathcal{M}, ν_M) is considered. A marked point process is a random sequence $\mathbf{x} = \{x_n = (k_n, m_n)\}$ such that the points k_n are a point process in K and m_n are the marks corresponding for each k_n . A random configuration of cylinders is a marked point process given by the centres positions of the cylinders and their corresponding random geometrical shapes attached to each point position. For further studying and reading on marked point processes we recommend the excellent monographs (van Lieshout 2000; Møller and Waagepetersen 2004).

The most known marked point process is the stationary Poisson point process of unit intensity with i.i.d. marks. This process works as it follows : choose a number of points according to a Poisson law of intensity parameter $\nu(K)$, spread the chosen points uniformly in K and to each point attach an independent mark using the distribution ν_M . Due to the independence property, the Poisson point process exhibits no structure. Nevertheless, it is perfectly possible to build models taking into account objects interactions, hence forming structures. Such a process can be described by a probability density with respect to the reference measure given by the unit intensity Poisson point process. Here, the probability density of the filamentary network has the general expression

$$p(\mathbf{x}|\theta) = \frac{\exp[-U(\mathbf{x}|\theta)]}{Z(\theta)}, \quad (2.1)$$

where $Z(\theta)$ is the normalising constant, θ is the model parameters vector, and $U(\mathbf{x}|\theta)$ is the (Gibbs) energy function of the system.

Our main hypotheses are that the galaxies forming filaments group together inside rather small cylinders, and that these small cylinders may connect and align to build the network.

Following these ideas, the energy function in (2.1) can be written as the sum :

$$U(\mathbf{x}|\theta) = U_d(\mathbf{x}|\theta) + U_i(\mathbf{y}|\theta), \quad (2.2)$$

where $U_d(\mathbf{x}|\theta)$ is the data energy and $U_i(\mathbf{x}|\theta)$ is the interaction energy. The data energy is related to the first assumption, that is the position of the cylinders in the galaxy field, whereas the interaction energy is related to the second hypothesis, that is the alignment and the connection of the cylinders forming the filamentary pattern. The data term is constructed using different statistical tests to verify that the galaxies are aligned along the main axis of the cylinder and that from a statistical point of view, the number of the galaxies inside of the cylinder is higher than outside of it. The interaction term is given by the Bisous model, a marked point process able to simulate random configurations of connected objects. All the details can be found in (Tempel *et al.* 2014).

Neither the filaments nor their model parameters are known. Therefore, under the Bayesian framework, the model is completed with a prior density $p(\theta)$ for the model parameters. This allows to write the joint estimator of the filamentary pattern and the parameters as

$$\begin{aligned} (\hat{\mathbf{x}}, \hat{\theta}) &= \arg \max_{\Omega \times \Psi} p(\mathbf{x}, \theta) = \arg \max_{\Omega \times \Psi} p(\mathbf{x}|\theta)p(\theta) \\ &= \arg \min_{\Omega \times \Psi} \left\{ \frac{U_d(\mathbf{x}|\theta) + U_i(\mathbf{x}|\theta)}{Z(\theta)} + \frac{U_p(\theta)}{Z_p(\theta)} \right\}, \end{aligned} \quad (2.3)$$

with $p(\theta) = \exp[-U_p(\theta)]/Z_p(\theta)$ the prior law for the model parameters, Ω the cylinders configurations space and Ψ the model parameters space.

3. Simulation and statistical inference

The normalizing constant of the models (2.1) does not have always a precise closed analytical form. This implies that the simulation of the model should be done using elaborate methods. Several Markov chains Monte Carlo (MCMC) techniques are available to simulate marked point processes (van Lieshout 2000; Møller and Waagepetersen 2004). Here, we have adopted the Metropolis-Hastings algorithm with tailored to the model transition kernels (van Lieshout and Stoica 2003; Stoica *et al.* 2005a). The classical Metropolis-Hastings randomly adds and deletes a cylinder from a configuration.

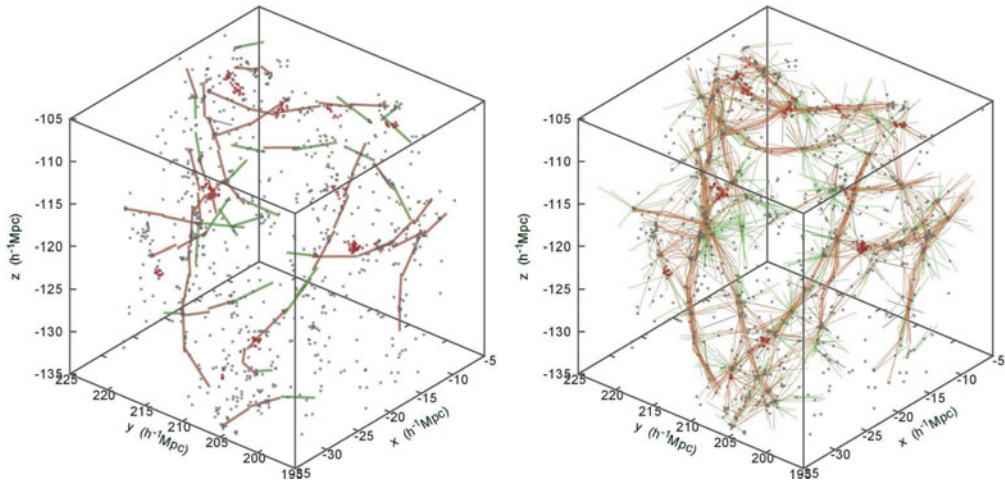


Figure 1. a) Detection result using the simulated annealing algorithm ; b) Superposition of several detection results.

This algorithm uses uniform proposals for its transition kernel, in order to guarantee the theoretical convergence to the desired equilibrium distribution. Nevertheless, whenever models exhibiting complex interactions are simulated, this convergence is very slow in practice. This is also the case for the filamentary network model. The solution is to build an algorithm using not only uniform proposals. More precisely, this means to not propose only uniformly spread cylinders, but also cylinders that tend to connect in order to form a network. As showed in (van Lieshout and Stoica 2003) such an algorithm has the required convergence properties.

The estimator in (2.3) is computed using the simulated annealing algorithm. This global optimization technique iteratively samples from $p(\mathbf{x}, \theta)^{1/T}$ while the temperature parameter T goes slowly to 0. The output of the algorithm converges in law towards the uniform distribution over the configuration sub-space made of the solutions of (2.3). The sampling mechanism at fixed temperature is a tailored Metropolis-Hastings. In order to ensure the algorithm convergence, the cooling schedule of the temperature should be slow (Stoica *et al.* 2005a). The solution given by the simulated annealing is not unique. Therefore, we are interested in averaging the obtained solution. This can be done using visit maps or level sets (Heinrich *et al.* 2012). Fig. 1 shows a detection result of filaments in a sample of cosmological catalogue. The superposition of several results is shown as well. Based on these results, visit maps and spines can be computed for the filaments (Tempel *et al.* 2014). All these together gave a rather robust detector of filaments.

A configuration of cylinders is made of cylinders that are connected at both extremities, the double connected cylinders, that are connected at one extremity only, the single connected cylinders, and that are not connected at all, the free cylinders. The sufficient statistics of the model are given by the number of these three types of cylinders, respectively. These statistics determine entirely the model, together with the mark distribution and the different interactions ranges. Furthermore, these statistics allow to characterize different galaxy catalogues from a morphological point of view. The authors in (Stoica *et al.* 2010) compare mock catalogues with real observation. It was stated, that the real filaments are longer and less fragmented. The sufficient statistics were also used to test the general method. This test was based on a bootstrap procedure, and it certified that the method does not detect filaments by chance, but only because these filamentary

structures do really exist in the observed data (Stoica *et al.* 2007). Once in possession of a map of filaments, interesting cosmological questions can be answered, as well. As an example, we mention (Tempel *et al.* 2013), where the authors found correlations between the filaments orientations and the spin alignment for spiral and elliptical galaxies.

4. Conclusions and perspectives

The different components of our model are simple and intuitive, allowing a visual definition and construction of a filament. The simplicity characteristic is adopted on purpose. Since, the objects we are looking for are not observed, the simpler hypotheses should be formulated and tested first. Under these circumstances, our results are comparable with all the other methods. All these techniques use a local gradient or score for the filaments together with a smoothing or regularization procedure. The strong points of our approach are the versatility of the model and its probabilistic general framework. Clearly, the model can be and should be improved. If more knowledge is available concerning the local definition of a filament and the general topology of the network, this knowledge must be integrated in the model. The probabilistic framework allows the computation of integrals, average quantities and characteristics of the filaments, leading to a reliable description of one of the most intriguing patterns in our observed Universe.

Acknowledgements

This paper is based on very recent work done together with E. Saar, E. Tempel, V. J. Martinez, L. J. Liivamägi and G. Castellan. E. Tempel provided also the Fig. 1.

References

- P. Heinrich, R. S. Stoica, & V. C. Tran. *Spatial Statistics*, 2:47–61, 2012.
- M. N. M. van Lieshout. Imperial College Press, London, 2000.
- M. N. M. van Lieshout & R. S. Stoica. *Statistica Neerlandica*, 57:1–30, 2003.
- V. J. Martinez & E. Saar. Chapman and Hall, 2002.
- J. Møller & R. P. Waagepetersen. Chapman and Hall/CRC, Boca Raton, 2004.
- R. S. Stoica, P. Gregori, & J. Mateu. *Stochastic Processes and their Applications*, 115:1860–1882, 2005.
- R. S. Stoica, V. J. Martinez, J. Mateu, & E. Saar. *Astronomy and Astrophysics*, 434:423–432, 2005.
- R. S. Stoica, V. J. Martinez, & E. Saar. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 55:189–205, 2007.
- R. S. Stoica, V. J. Martinez, & E. Saar. *Astronomy and Astrophysics*, 510(A38):1–12, 2010.
- E. Tempel, R. S. Stoica, & E. Saar. *Monthly Notices of the Royal Astronomical Society*, 428:1827–1836, 2013.
- E. Tempel, R. S. Stoica, E. Saar, V. J. Martinez, L. J. Liivamägi, & G. Castellan. *Monthly Notices of the Royal Astronomical Society*, 438: 3465–3482, 2014.