# GALAXY FILAMENT DETECTION USING THE QUALITY CANDY MODEL

*P. Gernez, X. Descombes, J. Zerubia*

Ariana Research Group, INRIA/I3S,
2004 route des lucioles - BP 93,
06902 Sophia-Antipolis Cedex, France
firstname.lastname@inria.fr

*E. Slezak, A. Bijaoui*

Observatoire de la Côte d'Azur,
Boite postale 4229,
06304 Nice Cedex 4, France
firstname.lastname@obs-nice.fr

## ABSTRACT

We propose to apply a marked point process to detect a galaxy filament network. From the "Quality Candy" model, initially developed for road network extraction in remotely sensed images, we adapt the data term to the filament detection. The optimization is realized by a simulated annealing using a Reversible Jump Markov Chain Monte Carlo algorithm. Results are presented on a numerical simulation and on an astronomical survey.

## 1. INTRODUCTION

Beyond three hundred billion light-years, when averaged over 100 Mpc, the visible cosmos can be seen as a gas of galaxies, uniformly distributed. At smaller spatial scales, astronomical observations as well as numerical simulations have shown that the repartition of the luminous matter in the Universe is not so homogeneous. Galaxies cluster within elongated large-scale structures, called filaments, and leave huge voids between those filaments. These filaments, which might only occupy 10% of the volume of the Universe, are organized in a complex three dimensional network often described as leading to a sponge-like or cell-like 3D topology. As shown in figure 1, such a filament is not a single structure with sharp edges, but instead a fuzzy set of points more or less scattered, which makes its detection difficult. Another difficulty in the detection process comes from the difference of spatial scales between sparse and prominent compact features. The gradual disappearance of structures with increasing distance results from the use of a magnitude-limited sample. The apparent luminosity of any object is fainter as distance increases, and only the few galaxies with the highest intrinsic luminosity are then included.

Up to now, there are only a few methods to extract the filamentary structure. The Minimal Spanning Tree (MST) method, firstly formalized by [1] has been then mostly used [2]. Recently, a method based on the "Candy" model [3] showed interesting results for the filament detection from a cosmological simulation [4]. In section 4.1, a comparison of our detection result with the MST and "Candy" detection results is presented.

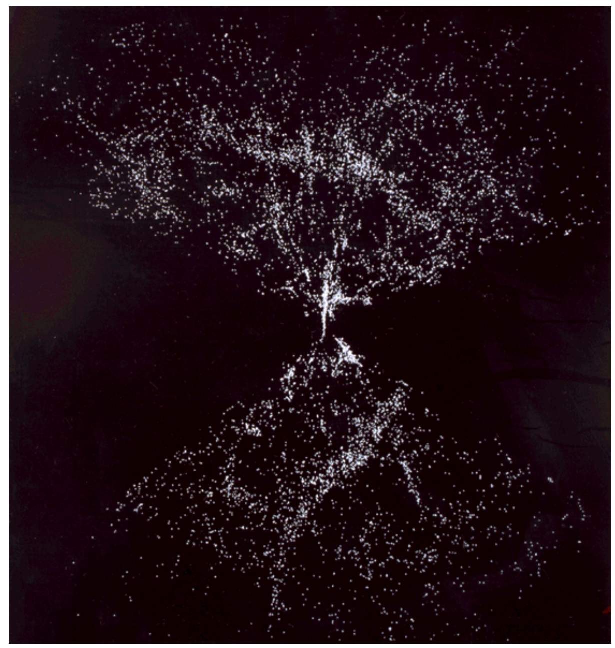

**Fig. 1**. Three dimensional view of galaxies up to 500 billion light-years, from the two CfA observation cones (credits: Center for Astrophysics, Harvard).

As the "Candy" model has been successfully adapted to extract cosmic filaments, we proposed to use the "Quality Candy" model [5] for the same application. Both models are based on a marked point process approach, whose efficacity has been shown for road network extraction in remote sensing [3, 5]. Marked point processes are shortly described in section 2. The proposed detection model is presented in section 3 and tests on 2-D galaxy maps are then shown in section 4.

.

## 2. MARKED POINT PROCESSES

The network of filaments is modeled by a marked point process, that is to say a random set of objects whose number of data points is also a random variable [6, 7] . The objects of this process are segments described by three random variables corresponding to their midpoint, their length and their orientation. The segment distribution is simulated by a density probability. For a uniform distribution, we use a Poisson process. In order to find the segment configuration $\mathbf{x}$ that better fits the filamentary network, we define a density probability $f(\mathbf{x})$ which takes into account the interactions between segments. $f$ is given by a Gibbs point process: $f \propto \exp(-U)$. The configuration of segments composing the filament network is estimated by the minimum of the energy $U$ of the system. $U$ has two components: the prior term $U_P$ forces the segment configuration to be a network and the data term $U_D$ helps this network to best fit the data. The estimate of $\hat{\mathbf{x}} = \arg\min U(\mathbf{x})$ is obtained by means of a simulated annealing algorithm. This algorithm iteratively samples the law $[f(\mathbf{x})]^{\frac{1}{T}}$ while slowly decreasing the temperature $T$. At high temperature, a lot of configurations are explored. When the temperature goes down to zero, the configuration of minimal energy is reached, assuming that a geometrical cooling scheme $T_{k+1} = cT_k$ with $c \in [0.99, 1]$ is sufficient. The probability density $f$ is simulated through a reversible jump Metropolis-Hastings dynamics sampling [8, 9, 10]. Basically, this dynamics drives the system to the minimal state by means of a set of segment perturbations: birth, death, translation, rotation and dilation. From an initial configuration $\mathbf{x}_0$, the algorithm is, at step t:

1. Compute the energy $U_{T_t}$ of the configuration $\mathbf{x}_t$,
2. Propose a new configuration $\mathbf{y}_t$, obtained by a perturbation of $\mathbf{x}_t$,
3. Evaluate the Green acceptance ratio $R(T_t)$,
4. Move to $\mathbf{y}_t$ with a probability equal $min\big(1, R(T_t)\big)$,
5. Decrease the temperature $T_t$.

We do not describe the computation of the Green ratio and of the perturbation proposal, because the algorithm has been presented in previous papers (see [11, 12])

## 3. THE GALAXY FILAMENT DETECTION MODEL

### 3.1. The prior energy

$U_P$ is the prior term of the energy. It takes into account the geometrical constraints of the network: slow curvature and good crossing points between the segments. As described in [13, 14], the network structure is obtained by penalizing segments which are not connected. The curvature constraint is optimized by quality functions with respect to the connection angles and the orientation between the segments. The overlaps between segments are forbidden in order to have neat crossing points.

### 3.2. The data energy

As shown in picture 1, filaments of galaxies are locally overdense clouds of points elongated along a principal direction. When the algorithm proposes a segment, it computes its data energy term to decide whether it reasonably suits a filament or not. $U_D$ is computed with respect to the circular neighbourhood depicted in figure 2. Three parameters are then defined: $\rho = n/r^2$ as a density term, $c = |\overline{y}|/r$ as a centering term and $a = \sum d^2(y,s)/d^2(x,s)$ as an elongating term, where $n$ is the number of galaxies is the neighbourhood, $r$ is the length of segment S, $d$ is the Euclidian distance and $x, y$ are respectively the longitudinal and latitudinal coordinates from the center of S. The energy of segment S is then given by:

$$U_D(S) = -\omega_1 g_1(\rho) - \omega_2 g_2(c) - \omega_3 g_3(a)$$

where the $\omega_i$ are weighting constants and the $g_i$ are quality functions. These quality functions are simple thresholding functions, whose threshold values have been experimentally determined.

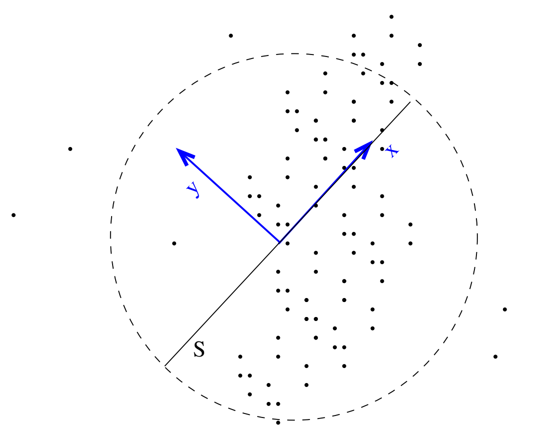


**Fig. 2**. Segment fitting to the local data distribution. Black points represent galaxies.

## 4. RESULTS

The "Quality Candy" model has been developed for 2D data (remotely sensed images). In order to reduce the computing time, its adaptation to the detection of cosmic filaments has also been performed for 2D astronomical catalogues.

### 4.1. Detection from a numerical simulation

We first tested our method on the numerical simulation kindly provided by the authors of [4]. As shown in figures 3, 4 and 5, both "Candy" and "Quality Candy" algorithms enable large improvement with respect to the MST method which basically simply connects all points of the map. The network extracted by the "Quality Candy" model, though partially incomplete, presents an overall structure in adequation with the description of the cosmic filaments made in section 1, that is to say, a network of filaments of varying sizes surrounding huge voids. The network extracted by the "Candy" model highlights more filaments, but some of them seem to be overdetected.

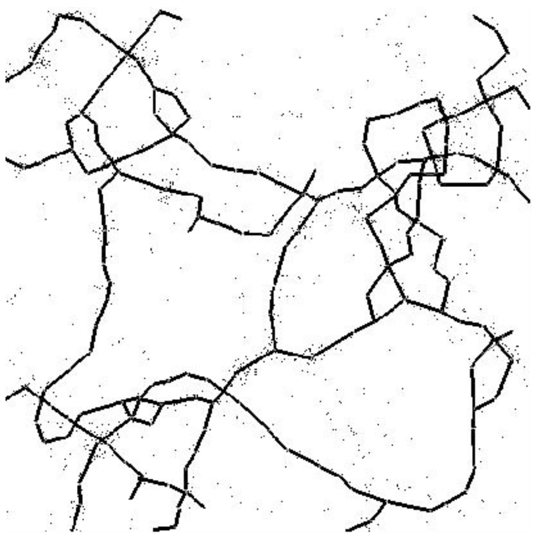

**Fig. 3**. Network extracted by the "Quality Candy" model (CPU time: 20mn).

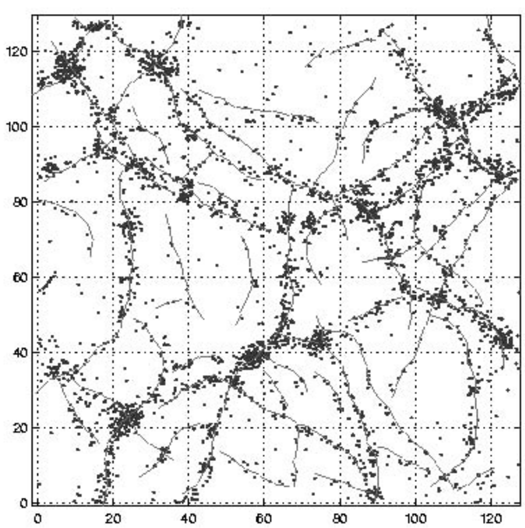


**Fig. 4**. Network extracted by the "Candy" model [4].

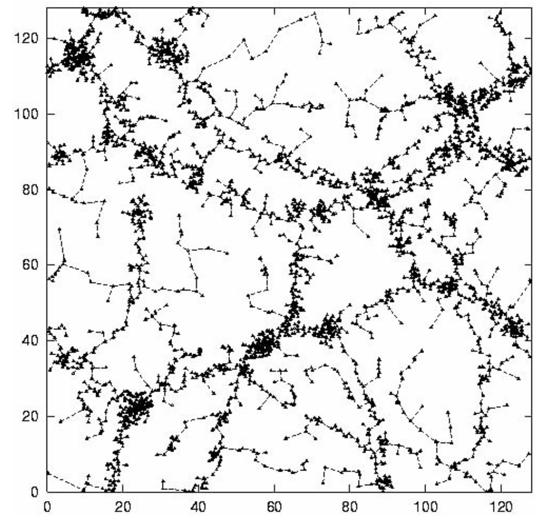


**Fig. 5**. Network extracted by the MST model [4].

### 4.2. Detection from an astronomical survey

The CfA redshift catalogue contains the coordinates of 18,000 galaxies covering a part of the Universe. From this catalogue a thin slice of the universe of about 1000 galaxies was extracted, with a right ascension between 8 and 17 hours, a declination between 26.5 and 32.5 degrees and a recession velocity lower than 15000 km.s$^{-1}$. The spatial distribution of this subset of galaxies is plotted in Figure 6. The result of the extraction with the "Quality Candy" model shows a very interesting filament network. One can recognize the famous "Great Wall" filament crossing the whole catalogue as well as the circular filaments surrounding void regions.

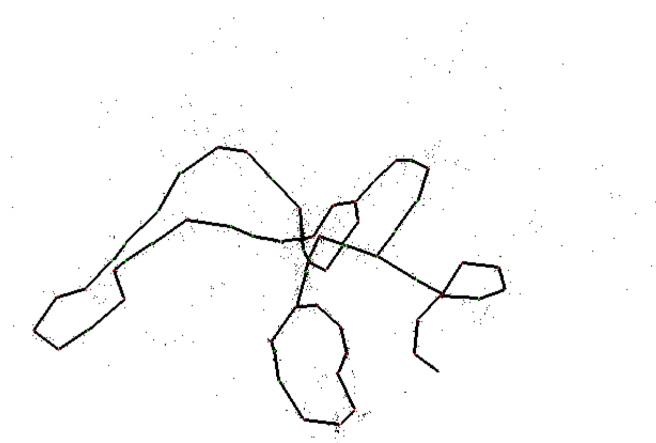

**Fig. 6**. Extracted filament network from a galaxy map of the CfA catalogue.

## 5. CONCLUSION AND FUTURE PROSPECTS

As shown in this paper, the "Quality Candy" model gives promising results for detecting a robust network of filaments in the spatial distribution of galaxies. It is the first time such a method is used to process an observational astronomical survey. Beyond the final tuning of the algorithm required for ending up with a fully satisfactory result from the astronomical point of view, a future step is to adapt our model to three dimensional data. Though theoretically possible, as both the prior and data terms of the energy are extendable to one more dimension, it requires to re-write the whole implementation of the algorithm. Eventually, it would also be interesting to achieve the estimation of the model parameters in order to make the detection fully automatic.

# Acknowledgments

The authors would like to thank Dr. R. Stoica for interesting discussions, Prof. J. Mateu and Prof. V. Martinez for providing the data shown in Fig. 4 and 5. Part of this work has been supported by an INRIA COLORS grant (FIGARO). .

## 6. REFERENCES

[1] J.D. Barrow, D.H. Sonoda, and S.P. Bhavsar, "Minimal spanning tree, filaments and galaxy clustering," *Mon. Not. R. Astr. Soc.*, vol. 216, pp. 17–35, 1985.

[2] V.J. Martinez and E. Saar, *Statistics of the Galaxy Distribution*, Chapman and Hall/CRC, Boca Raton, 2002.

[3] R. Stoica, X. Descombes, and J. Zerubia, "A Markov point process for road extraction in remotely sensed images," Tech. Rep., INRIA, 2000.

[4] R. Stoica, V.J. Martinez, J. Mateu, and E. Saar, "Detection of cosmic filaments using the Candy model," *Astronomy and Astrophysics*, vol. 434, pp. 423–432, 2005.

[5] C. Lacoste, X. Descombes, and J. Zerubia, "A comparative study of point processes for line network extraction in remote sensing," Tech. Rep., INRIA, 2002.

[6] M.N.M van Lieshout, *Markov Point Processes and their applications*, Imperal College Press, London, 2000.

[7] X. Descombes, "Méthodes stochastiques en analyse d'image: des champs de markov aux processus ponctuels marqués," Habilitation à diriger des Recherches, Université de Nice-Sophia Antipolis, France, 2004.

[8] M. Metropolis, A. Rosenbluth, R. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

[9] W.K Hastings, "Monte Carlo sampling using Markov chains and their applications," *Biometrica*, vol. 57(1), pp. 97–109, 1970.

[10] P.J Green, "Reversible jump Monte Carlo Markov chain computation and Bayesian model determination," *Biometrica*, vol. 57, pp. 97–109, 1995.

[11] M. Ortner, X. Descombes, and J. Zerubia, "A reversible jump MCMC sampler for building detection in image processing," in *Monte Carlo Methods and Quasi-Monte Carlo Methods*, Juan les Pins (France), june 2004.

[12] R. Stoica, X. Descombes, and J. Zerubia, "A Gibbs point process for road extraction from remotely sensed images," *IJCV*, vol. 57, no. 2, pp. 121–136, 2004.

[13] C. Lacoste, *Extraction de réseaux linéiques à partir d'images satellitaires et aériennes par processus ponctuels marqués*, Ph.D. thesis, Université de Nice-Sophia Antipolis, 2004.

[14] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE Trans. Pattern Analysis ans Machine Intelligence*, vol. 434, no. 10, pp. 1568–1579, 2005.