PhD Thesis position

Clusterization models on random graphs with applications to social networks and image processing

For more information or to apply, please contact:

Konstantin AVRACHENKOV (Inria, Neo Team, and Université Côte d'Azur)
https://www-sop.inria.fr/members/Konstantin.Avratchenkov/me.html
email: Konstantin.Avratchenkov@inria.fr

and

Josiane ZERUBIA (Inria, Ayana Team, and Université Côte d'Azur)
http://www-sop.inria.fr/members/Josiane.Zerubia/index-eng.html
email: Josiane.Zerubia@inria.fr

Application:

Please email a full application to both Konstantin Avrachenkov and Josiane Zerubia, indicating "PhD Thesis position" in the e-mail subject line. The application should contain:

- CV including publication list
- publications in pdf (if any)
- minimum 2 reference letters
- motivation letter demonstrating academic strengths and related experience to this position (in particular, mention your competences w.r.t. graph theory, probability, statistical mechanics and image processing).

Abstract:

The Ising model and the Potts model are two renown examples of Gibbsian models in statistical physics. The Potts model is a generalization of the Ising model when a spin can take more than two values. There are other more general Gibbsian models that are used in physics and image processing. The properties of the Gibbsian models have been extensively studied on the lattice and on the complete graph. However, much less is known about the properties of the Gibbsian models on other types of graphs.

Many natural and technological systems can be modeled by Random Geometric Graphs. Random geometric graphs are random graphs where the nodes are embedded into a metric space and the probability of link creation depends on (typically decreases with) the distance between the nodes. Furthermore, many network systems have clustering structure. For instance, in a social network people interact more often with the other people who live in the same geographical region or with colleagues from the same institution. This phenomenon can be modeled by a Soft Geometric Block Model (SGBM). In SGBM a link is more likely to be created between the nodes belonging to the same cluster. SGBM is just one way to introduce heterogeneity into the standard geometric random graph model. However, there can be other ways to introduce heterogeneity, *e.g.*, by changing the density function of the nodes.

The goals of the present thesis proposal are multiple: Glauber dynamics and simulated annealing have been used to detect clusters. This method based on the mean-field analysis indicate high efficiency in the homogeneous case. Is this still true in the inhomogeneous one? Other classification methods, 'spectral clustering', using linear algebra tools have shown impressive results. Are these results (theoretical and practical) still valid in the heterogeneous case? Then, it is expected to apply the developed methods to social networks and image processing clustering problems.

Let us expand a bit more about the usefulness of the framework based on these ideas for image processing. It can be useful to group neighboring pixels into areas having homogeneity properties (often called “superpixels”). To do so, we propose to develop graph network methods quite different from stochastic ones. Same ideas based on stochastic calculations allow to extract line networks (*e.g.* roads extraction in remote sensing images or galaxy filament detection in astronomical images). One advantage of the proposed framework is to be able to deal with both 2D and 3D images. Another advantage is to improve computational time. One example with astronomical images: At very coarse scale, the Cosmos seems to be a gas of galaxies uniformly distributed. This is not the case if we take a closer look. Galaxies cluster within elongated structures called filaments (*cf.* Fig. 1 below). This could be a good example of the use of RGGs to image processing.

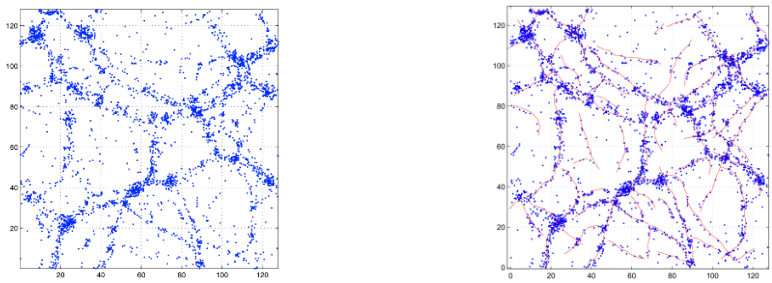

Figure 1. (Left) The galaxy distribution and (Right) An example of filaments' structure generated by the "Candy" model developed by Radu Stoica et al. *in 2004.*

Candidate profile:

We encourage applications from outstanding candidates with strong academic background (particularly in graph theory, probability and statistical mechanics). An experience in machine learning, in particular in classification, will be appreciated as well as an experience in image processing. The candidate must know how to program in Python and/or Matlab.
At Inria and UCA, we seek to increase the number of women in areas where they are under-represented and therefore explicitly encourage women to apply. Furthermore, we are committed to increasing the number of individuals with disabilities in our workforce and therefore encourage applications from such qualified individuals.

Indemnity:

PhD Thesis salary:

- around 1600~1700 EUR per month (located at Inria and Université Côte d'Azur in Sophia-Antipolis on the French Riviera).
- trips and stays for national and international conferences may be funded independently of the PhD Thesis salary.

Duration:

3 years

Starting date:

October 2023