

# Projects

MSc in Statistics for Smart Data – Introduction to graph analysis and modeling

Julien Chiquet, November the 22, 2018

## Instruction

In each project, the trinome should read the journal paper and write a report on the general motivations, goals and methods used. You do not need to have a deep understanding of the technical aspects.

Then, you must illustrate these methods on a real world data set picked up on the internet. For each project, there exist an R package so you will not need to write much code<sup>1</sup>. The second part of your report will present the results of your analyses.

Your oral presentation will present your report, that should not exceed 10 pages long (but *well written*).

*Send me the report back at [julien.chiquet@inra.fr](mailto:julien.chiquet@inra.fr) before the 13th of December at midnight.*

## Web resources

### Network data

Pick up some network data (with less than 500/1000 nodes for your convenience!) to illustrate the method that you study, for instance in

- Network repository <http://networkrepository.com/>
- General network data: <http://www-personal.umich.edu/~mejn/netdata/>
- Ecological network database: <http://networkrepository.com/eco.php>
- SNAP database: <https://snap.stanford.edu/data/index.html>
- ... *feel free to use your own network data.*

## Project 1 - beyond binary edges

This paper presents several extensions of the Stochastic Block Model where edges are weighted with various distributions (Poisson and Gaussian for instance). It also shows how one can include external knowledge on top of the network structure, by means of covariates on the nodes of the graph. All the corresponding models are implemented in the package **blockmodels**. Use it to analyse some weighted network data and/or binary network with covariates.

---

<sup>1</sup>expect for making some fancy figures!

Journal paper. Mahendra Mariadassou, Stéphane Robin and Corinne Vacher. *Uncovering latent structure in valued graphs: a variational approach*. The Annals of Applied Statistics (2010): 715-742. <https://arxiv.org/pdf/1011.1813.pdf>

R package. <https://CRAN.R-project.org/package=blockmodels>, see also <https://arxiv.org/abs/1602.07587>

## Project 2 - dynamic SBM

This paper uses an extension of the Stochastic Block Model where memberships may vary across time in order to analyse an ecological network gathered in time. The corresponding model is implemented in the package **dynsbm**. Use it to analyse some time-varying network data.

Journal paper. Vincent Miele and Catherine Matias. *Revealing the hidden structure of dynamic ecological networks*. Royal Society Open Science 4.6 (2017): 170251. <http://rsos.royalsocietypublishing.org/content/4/6/170251>

R package. <https://CRAN.R-project.org/package=dynsbm>

## Project 3 - Multiplex SBM

When several relationships of various types can occur jointly between nodes, the data are represented by multiplex networks where more than one edge can exist between the nodes. The bivariate Bernoulli version of the multiplex SBM used in the following paper is implemented in the package **blockmodels**. Use it to analyse some multiplex network data.

Journal papers. Pierre Barbillon, Sophie Donnet, Emmanuel Lazega and Avner Bar-Hen. *Stochastic block models for multiplex networks: an application to a multilevel network of researchers*. Journal of the Royal Statistical Society: Serie A, 180: 295-314. (2017)

R package. <https://CRAN.R-project.org/package=blockmodels>, see also <https://arxiv.org/abs/1602.07587>