

## Modeling and inference of the persistence of information on social networks

**Keywords :** Social networks, topic modeling, multivariate long-range dependence

**Context :** Social networks and medias in general create a huge quantity of information which may differ according to the location (countries, areas, cities.....) and the time periods. A natural question is to identify which main topics are persistent in a corpus of documents as tweets, websites or scientific papers. The aim of the project is to take into account the specificities of data as similarities between different regions or countries as well as the time stamp of the document...This question has been already addressed in several papers (see for e.g. [1]) and several models have been proposed to summarize the temporal evolution (see for e.g. [2]).

**Challenges :** We aim at complementing these works studying spatio-temporal persistence in textual data. Using dynamic topic modeling [3], we can modeled in real-time the content evolution of a corpus. Our goal will be to identify which topics are persistent in a corpus, taking into account both spatial and temporal information. The part simulation and inference will be designed using Monte Carlo methods [6,7] whereas persistence will be measured using multivariate long range dependence [4].

### Bibliography

- [1] S. ASUR, B. A. HUBERMAN, G. SZABO, C. WANG. Trends in social media: persistence and decay. *In ICWSM. (2011)*.
- [2] Y. WANG, E. AGICHTAIN, M. BENZI. TM-LDA: efficient online modeling of latent topic transitions in social media. *Proc. of the 18th ACM SIGKDD. ACM (2012)*.
- [3] D. BLEI, J. D. LAFFERTY. Dynamic topic models. Proceedings of the 23rd international conference on Machine learning. ACM, (2006).
- [4] S. KECHAGIAS, V. PIPIRAS. Definitions and representations of multivariate long-range dependent time series. *JTSA 36.1 1-25 (2015)*.
- [5] M. LI, X. WANG, K. GAO, S. ZHANG. A survey on information diffusion in online social networks: Models and methods. *Information 8, no. 4: 118 (2017)*.
- [6] G. WINKLER, Image analysis, random fields and MCMC methods, *Springer (2003)*
- [7] R. S. STOICA, A. PHILIPPE, P. GREGORI, J. MATEU. ABC Shadow algorithm: a tool for statistical analysis of spatial patterns. *Stat. Comp., 27(5) : 1225-1238, (2017)*

**Duration:** 3 years (full time position). **Starting date:** October, 2019

**Supervisors:** This thesis will be cosupervised by M. Clausel and R. Stoica both full Professors at IECL (Nancy, France):

**M. Clausel :** <https://sites.google.com/site/marianneclausel/>

e-mail address : [marianne.clausel@univ-lorraine.fr](mailto:marianne.clausel@univ-lorraine.fr)

**R. Stoica :** <https://sites.google.com/site/radustefanstoica/>

**Contacts :** [marianne.clausel@univ-lorraine.fr](mailto:marianne.clausel@univ-lorraine.fr), [radu-stefan.stoica@univ-lorraine.fr](mailto:radu-stefan.stoica@univ-lorraine.fr)

**Working Environment:** The PhD candidate will work between the Probability and Statistic team of the IECL lab which is a leading institutions, respectively in Mathematics in France. The lab is a located at Nancy, France. This subject is part of the OLKI project

(<http://lue.univ-lorraine.fr/fr/open-language-and-knowledge-citizens-olki>) of the programm Lorraine Université d'Excellence.

**Location :** Nancy, which is the capital of Lorraine in France, with excellent train connection to Luxembourg (1h30) and Paris (1h30).