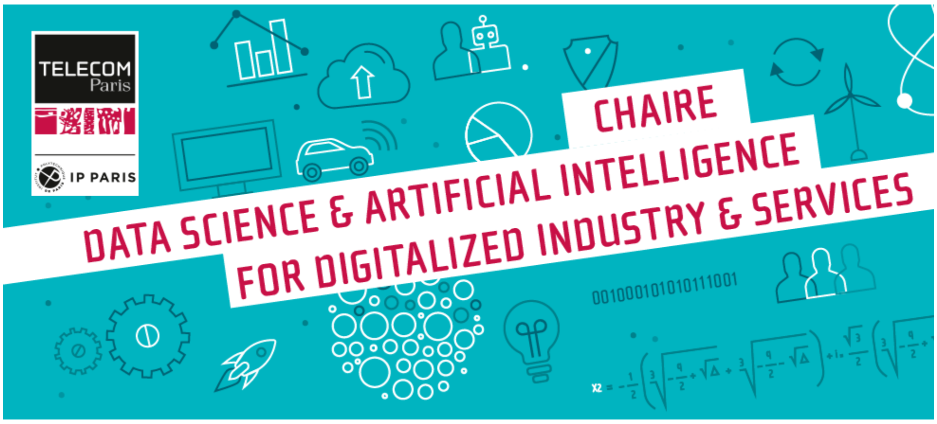

Paris, le 29/11/22

Offre de stage

Sujet : Modélisation et prévision de la dynamique de graphes

Topic : Modeling and forecasting the temporal evolution of graphs

Possibilité de poursuivre sur une thèse

La Chaire Data Science and Artificial Intelligence for Digitized Industry and Services

Portée par Florence d'Alché-Buc, enseignante-chercheur dans le département Image, Données, Signal de Télécom ParisTech, la chaire DSAI réunit cinq partenaires industriels : Airbus Defence & Space, Engie, Idemia, Safran et Valeo Finance. Son objectif général est de développer, en liaison étroite entre les Parties, une formation et une recherche de niveau international.

Ses quatre principaux axes de recherche sont :

1. Analyse et prévision de séries temporelles (Predictive Analytics on Time Series) ;

2. Exploitation de données hétérogènes, massives et partiellement étiquetées (Exploiting Large Scale and Heterogeneous, Partially Labelled Data) ;

3. Apprentissage pour une prise de décision robuste et fiable (Learning for Trusted and Robust Decision) ;

4. Apprentissage dans un environnement dynamique (Learning through Interactions with a Changing Environment).

Description du stage

Encadrement

Florence d'Alché (Télécom Paris), Charlotte Laclau (Télécom Paris), Rémi Flamary (Polytechnique) - Collaboration avec Gabriel Peyré (ENS).

Lieu et dates du stage

Telecom Paris, Place Marguerite Perey, 91120 Palaiseau (RER Massy-Palaiseau + bus91.06)

Starting in April 2023 / Date de début du stage : Avril 2023

Laboratory / équipe d'accueil

LTCI, département IDS, équipe S2A: Signal, Statistique, Apprentissage

Key-words / mots clefs

graphs, dynamics of graphs, graph-to-graph prediction, neural networks, kernel methods, optimal transport (Gromov-Wasserstein distance)

Topic / sujet

Graphs are ubiquitous in many fields such as economics, epidemiology and social networks where they allow to represent potentially complex interactions between a large number of components/Individuals. While static graph analysis or link and node classification (Vincent-Cuaz et al. 22) have attracted a lot of attention in the Machine Learning literature for two decades, the modeling of temporal evolution of graphs where both nodes and edges can vary as well as their labels, has been somehow overlooked these last years and has only emerged as a hot topic very recently. In this project, we consider that we observe the dynamics of graphs through graph snapshots not necessarily taken at regular intervals. The objective is to propose novel and well principled statistical learning techniques to model this temporal evolution for monitoring and forecasting purposes inspired by time-series modeling in Euclidean spaces (Chen et al. 2018). Choosing an appropriate graph representation and a dedicated loss function will be key to set up a principled approach that will be further implemented on simple to complex hypothesis spaces (RKHS, graph conv nets) with simplicity criteria in mind. In particular, a promising direction relies on taking a Gromov-Wasserstein distance as a loss and leveraging GW barycenters as initiated in (Motte et al. 2022) for supervised graph prediction. GW distance presents a number of appealing properties among which the ability to tackle graphs of different size, a case that inherently appears in graph dynamic modeling. Progresses on gradient flows, and more generally on the tight links between ODE and Neural ODEs (Chen et al. 2018) as well as their stochastic variant, are also encouraging and will be investigated.

Candidate Profile

Master 2: MVA, Master Data Science, any high-level master in Data Science or Machine Learning /Optimization

- Excellent level in statistical learning

- Background in dynamical modeling
- Notions of Optimal transport
- Very good level in Python programming
- Very good level in english

Candidatures

To be sent to florence.dalche@telecom-paris.fr,charlotte.laclau@telecom-paris.fr

- Curriculum Vitae
- Motivation letter / Lettre de motivation personnalisée expliquant l'intérêt du candidat sur le sujet (directement dans le corps du mail)
- Grades for the previous year / Relevés de notes des années précédentes
- At least one referee

Références

[1] Gabriel Peyré, Marco Cuturi, Justin Solomon: Gromov-Wasserstein Averaging of Kernel and Distance Matrices. ICML 2016

[2] Hu, W., Fey, M., Zitnik, M., Dong, Y., Ren, H., Liu, B., ... & Leskovec, J. (2020). Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, *33*, 22118-22133.

[3] Luc Brogat-Motte, Rémi Flamary, Céline Brouard, Juho Rousu, Florence d'Alché-Buc:

Learning to Predict Graphs with Fused Gromov-Wasserstein Barycenters. ICML 2022: 2321-2335

[4] Cédric Vincent-Cuaz, Rémi Flamary, Marco Corneli, Titouan Vayer, Nicolas Courty:

Template based Graph Neural Network with Optimal Transport Distances. CoRR abs/2205.15733 to appear in NeurIPS 2022 (2022)

[5] Kipf, T. N. et M. Welling . Semi-supervised classification with graph convolutional networks. ICLR 2017

[6] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud: Neural Ordinary Differential Equations. NeurIPS 2018: 6572-6583

[7] James Eells and Joseph H Sampson. Harmonic mappings of riemannian manifolds. American journal of mathematics, 86(1):109–160, 1964.

[8] Fillipo Santambrogio: {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences* pages 87–154 (2017)