

Sujet de stage de M2 / Master's Degree Internship 2022-2023

Intitulé Prédiction d'attributs spatiaux au moyen de réseaux convolutifs sur graphes
Title *Prediction of Spatial Attributes Based on Graph Convolutional Networks*

Encadrants [Vincent Labatut](mailto:vincent.labatut@univ-avignon.fr) (LIA) <vincent.labatut@univ-avignon.fr>
Advisors [Margot Ferrand](#) (CIHAM)

Lieu / Location [Laboratoire Informatique d'Avignon](#), Avignon, France

Contexte. Dans le cadre de son doctorat, Margot Ferrand a extrait un *graphe de confronts* à partir d'une collection d'actes notariés produits à Avignon au moyen âge (à l'époque des papes, pour être plus précis). Dans ce graphe spatial, les sommets représentent des biens, rues, et autres points de repère géographiques, et les arêtes correspondent à leurs positions relatives (i.e. v_1 est au nord/sud/est/ouest de v_2). Pour certains de ces sommets, la position GPS est connue, mais pas pour d'autres. Actuellement, nous interpolons ces coordonnées simplement en moyennant celles des sommets voisins, mais il serait préférable d'explorer des méthodes plus élaborées afin d'améliorer ces approximations.

Objectif. Ce stage vise à tester sur ces données historiques le pouvoir prédictif des réseaux de neurones convolutifs sur graphes. Ces réseaux ([Graph Neural Networks](#)) constituent une généralisation des réseaux de neurones, destinée à être appliquée à des données prenant la forme de [graphes](#) (par opposition aux données numériques tabulaires classiques). Dans le cadre de ce stage, nous nous intéressons à la version *convolutive* de cette méthode (*Graph Convolutional Networks*), qui permet notamment de prédire des valeurs manquantes dans un graphe incomplètement décrit par des attributs attachés à ses sommets ou à ses arêtes.

Organisation. Le ou la stagiaire devra effectuer une brève revue de la littérature concernant les GCN, ainsi que des bibliothèques disponibles pour les appliquer. Il ou elle devra également comparer les performances obtenues à celles produites via des méthodes plus standards (comme par exemple l'interpolation basée sur le voisinage). Un travail de modélisation sera également au programme, visant à élaborer un modèle générateur de données artificielles dans le but d'évaluer la qualité des prédictions réalisées.

Profil du candidat. L'étudiant-e doit être en dernière année de diplôme d'ingénieur ou en M2 d'informatique (ou domaine proche). Il ou elle doit posséder des notions de programmation, maîtriser les méthodes standard d'apprentissage automatique, et si possible avoir une expérience dans l'utilisation de bibliothèques de réseaux de neurones, ainsi qu'en graphes.

Le montant de la rémunération est standard (≈ 550 €/mois), et le stage a lieu de Février à Juillet. Pour postuler, merci d'envoyer par email à Vincent Labatut les trois documents suivants : 1) CV ; 2) relevé de notes (licence et master) ; 3) lettre de motivation. **Votre lettre doit obligatoirement être écrite pour ce stage-ci : toute lettre générique invalidera la candidature.**

Context. During her PhD, Margot Ferrand has extracted a so-called *confront graph* based on a collection of medieval legal documents written in Avignon under the popes. In this spatial graphs, vertices represent real estate, streets and other geographical landmarks, and edges correspond to their relative positions (i.e. v_1 is north/south/east/west of v_2). The GPS position of some of these vertices is known, but that is not the case for all of them. We currently interpolate these missing coordinates using a simple method consisting in averaging the position of the neighboring vertices, but we want to explore more advanced methods in order to improve these approximations.

Goal. This internship aims at assessing, on these historical data, the predictive abilities of [Graph Neural Networks](#). They constitute a generalization of neural networks, designed to handle [graphs](#) instead of the usual numeric tabular data. In the context of this internship, we are interested in the *convolutional* version of this approach (*Graph Convolutional Networks*), which is able to predict missing values in graphs incompletely described by attributes attached to its vertices and/or edges.

Organization. The intern will first do a short survey of the literature related to GCNs, as well as of the libraries available to apply these methods. Next, he or she will compare the performances obtained through GCNs with those coming from more traditional methods (such as the neighborhood-based interpolation mentioned before). Finally, he or she will also work on a model able to generate realistic artificial data, in order to assess the quality of the predictions.

Required profile. The student must be in last year of engineering program, or in 2nd year of master's degree, in Computer Science or a related field. He or she must be able to code, be familiar with machine learning method, and possibly have already worked with standard neural networks libraries, and with graphs.

Application. The salary matches French regulation (≈ 550 €/month). The internship duration is 6 months (Feb-Jul). To apply, please send an email to Vincent Labatut, with the three following documents: 1) CV; 2) academic transcript (BSc & MSc); 3) motivation letter. **This letter must absolutely be written specifically for this internship: generic letters will be ignored.**

Mots-clés Histoire, Réseaux complexes, Réseaux de neurones, Apprentissage automatique
Keywords History, Complex networks, Neural networks, Machine learning