

# TP Advanced Data Mining Graphs for PR & Mining

JY Ramel – 2020

*Le compte-rendu de TP (fichier zip avec sources et fichier contenant les réponses aux questions) à déposer sur Celene (bien préciser les noms des élèves associés à chaque rendu).*

L'objectif de ce TP réalisé sous Jupyter Notebook est de mettre en place des techniques de Graph Matching et de Graph Classification en utilisant la librairie NetworkX.

## 1. Données étudiées

Le dataset exploité pour ce TP est la base Letters. Cette base fait partie de [IAM Graph Database Repository](#) et est aussi proposée par le [IAPR TC15 resources](#).

Elle constitue souvent un "TOY EXAMPLE" pour expérimenter des méthodes d'appariement et classification de graphes.

## 2. Codes et données mis à disposition

- Le fichier GMnotebooks.zip disponible sur Celene contient les éléments suivants :

- GraphMatching.ipynb → notebook principal comportant le code de la partie Graph Matching à compléter
- GraphClassif.ipynb → notebook comportant le code de la partie Graph Classification à compléter (en partie facultatif)
- KPPV.ipynb → Fonctions éventuellement utiles pour mettre en place la classification de graphes
- Bac\_a\_sable\_NetworkX.ipynb → exemples de codes exploitant NetworkX pour l'analyse de graphes

- Letters.zip → Dataset Letters du repository IAM.

## 3. Travail à réaliser

Démarche :

- Comprendre et compléter le notebook GraphMatching.ipynb
  - Lecture des fichiers de la base Letters
  - Manipulation des Graphes Letters avec NetworkX
  - Proposition de mesures de similarité et dissimilarité entre nœuds
  - Graph Matching par appariement des nœuds
  - Visualisation et analyse des résultats d'appariement
  - Exploitation de la méthode GED disponible dans NetworkX
- Comprendre et compléter le notebook GraphClassification.ipynb (**en partie facultatif**)
  - Lecture des fichiers de la base Letters et répartition en différents sous-ensembles
  - Exploitation des méthodes mises en place dans la partie GM pour mesurer des similarités /dissimilarités entre Graphes
  - Mise en place d'une fonction de Graph Embedding adaptée à la base Letter
  - Exploitation du classifieur KPPV (ou autre) pour classifier les graphes de Letters
  - Mise en place d'une fonction de Graph Embedding plus génériques (capable de fonctionner sur différentes bases de graphes). On pourra par exemple s'inspirer des techniques de caractérisation de nœuds et arcs utilisés en Graph Mining