REPUBLIQUE DU CAMEROUN
Paix-Travail-Patrie
UNIVERSITE DE YAOUNDE 1
DEPARTEMENT
D'INFORMATIQUE
BP/P.O.Box 812
Yaounde-Cameroun

REPUBLIC OF CAMEROON
Peace-Work-Fatherland
UNIVERSITY OF YAOUNDE 1
COMPUTER SCIENCES
DEPARTMENT
BP/P.O.Box 812
Yaounde-Cameroun

Prédiction du risque de crédit à base de descripteurs issus de la modélisation des données en graphes

Noms et prénoms : Victor Nico DJIEMBOU TIENTCHEU
Matricule : 17T2051
Niveau : Master 2
Spécialité : Sciences de Données (DS)
Encadreur : Dr. Armel Jacques NZEKON NZEKO'O

Superviseur : Pr. Maurice TCHUENTE

Table des matières

1 Introduction

Le prêt est l'une des principales sources d'enrichissement des banques, mais est également une source de perte financière. Pour optimiser leurs profits, plusieurs travaux sur le crédit scoring ont longtemps considéré uniquement les attributs de description des prêts (caractéristiques de l'emprunteur, somme à prêter, ...) sans toutefois s'attarder sur une modélisation explicite des relations entre les emprunteurs.

Ceci peut être une limite, car les individus aux caractéristiques communes peuvent avoir les mêmes comportements de prêts et donc à partir des comportements connus d'un ensemble d'individus similaires à un individu cible, on peut déduire le comportement de ce dernier. Dans ce projet, il est donc question de modéliser les données d'une base de prêts bancaires par des graphes dont la définition des neouds et des arcs est suffisamment pertinente pour que les nouveaux descripteurs extraits de ses graphes contribuent fortement à la décision des modèles de d'apprentissage automatique pour la prédiction du Risque de prédiction du risque de crédit.

Nous allons pour ce faire déjà présenter dans la section 2 des généralités sur les graphes et les descripteurs extraits des graphes. Ensuite, nous présenterons dans la section 3 la mdélisation en graphes multicouches et les descripteurs extraites de ces graphes. Enfin, nous présenterons quelques expérimentations réalisées et les resultats obtenus à la section 4 et nous achèverons par une conclusion sur le travail fournis et quelques perspective dans la section 5.

2 Descripteurs extraits des graphes

2.1 Généralité et définition

Un graphe $G = \{V, E\}$ est une structure de données qui permet de modéliser les relations entre des entités. La modélisation d'un graphe repose sur deux notions, celle de nœuds V et celle d'arcs E. La notion de nœud est associée aux entités qui sont en relation et celle d'arc est associée à la nature de la relation d'une entité (nœud) avec une autre.

Après la construction d'un graphe, il est possible d'extraire des variétés de descripteurs liés soit aux nœuds et donc aux entités, soit à la relation entre les nœuds, et même à la topologie du graphe.

- **Descripteurs d'un nœud :** on peut citer les mesures de centralité qui estiment à quel point le nœud est incontournable dans la navigation dans le graphe. C'est le cas par exemple du PageRank[2] où initialement on attribue le même poids à chaque nœud, puis chaque nœud diffuse son poids à tous ses voisins directs proportionnellement aux poids des relations avec ses voisins. Le processus est répété jusqu'à ce que les poids des nœuds ne changent plus, ou alors jusqu'à ce qu'un nombre maximum d'étapes de diffusion soit atteint. Les nœuds aux poids les plus grands, sont les plus importants.
- **Descripteurs d'une relation entre deux nœuds :** on peut parler des mesures qui décrivent la relation entre deux nœuds à partir du nombre et de la longueur des plus courts chemins entre ces nœuds.
- **Descripteurs de la topologie du graphe :** il est possible d'extraire des communautés dans un graphe (sous-ensemble de nœuds densément connectés entre eux et faiblement connectés au reste du graphe).

Le procédé qui consiste à construire un graphe et à calculer des descripteurs, permet d'apporter de nouvelles

informations pour enrichir la description des entités considérées en entrée d'un problème abordé

2.2 Modélisation existant pour extraire les descripteurs des graphes

De nombreux modélisation jusqu'ici ont déjà été pensé pour modéliser des graphes avec des base de prêt afin d'extraire des descripteurs pertinents.

S. Mario[3] pour capturer les relations financières entre entités intervenantes (emprunteurs, instituts financiers) dans le contexte d'une inter-coopération des institutions financières et des emprunteur va modéliser un graphes orienté représentant une micro-structure du réseau où chaque noeud représente un emprunteur ou une institution financier et les arêtes, le lien financier existant entre les noeuds. Il va extraire des mesures telles que la taille des k plus proche voisin, le dégrée d'un noeud, la distance de plus court chemin, le grade d'un noeud, le sous graphe maximal de distance minimal afin d'améliorer l'évaluation du risque de crédit avec des modèles explicable comme Support Vector Machine (SVM) et Logistic Regression (LR).

Xiujuan Xu et Al.[4] vont construire un graphe biparti pour représenter les informations d'historique de prêts d'institut bancaire. Ici chaque noeud matérialise soit un emprunteur soit un modalité caractéristique et une arête entre un noeud emprunteur et un noeud modalité signifie que l'emprunteur est décrite par cette modalité. Pour prendre en compte les relations complexe entre les emprunteurs, ils vont utiliser trois algorithmes d'analyse de liens basés sur le prétraitement de SVM : Hub Authority Ranking Applicants Algorithm (HARA), Hub-Avg ranking applicants Algorithm (HubAvgRA), Authority-Threshold Algorithm (ATkRA).

Ses deux approches font bien d'utiliser des techniques d'analyse de graphes pour capturer des descripteurs pertinents pouvant améliorer la prédiction du risque de crédit par des algorithmes d'apprentissage automatique, toutefois, elles ne considèrent pas que la relation entre emprunteurs peut être fortement influencé par un sous ensemble de caractéristiques ou niveaux. Hors hypothétiquement, si je suis en relation avec un individu c'est forcement au moins à cause d'un niveau de caractérisation en commun.

3 Descripteurs extraits des graphes multicouches

La lecture de l'article « Multilayer network analysis for improved credit risk prediction » de María Óskarsdóttir et Cristián Bravo[8], a été le point d'entrée pour l'usage des descripteurs issus de la modélisation en graphe, pour enrichir les données d'apprentissage des modèles classique de prédiction du risque de crédit.

En effet, les auteurs utilisent la modélisation par des graphes multicouches (multilayer network) où un emprunteur a autant de nœuds qu'il y a de dimensions qui le caractérisent, et dans chaque dimension, il est relié à des attributs qui le définissent suivant cette dimension. Ainsi, plus les emprunteurs sont similaires, plus ils sont proches dans le graphe multicouche. Une dimension peut par exemple être la localisation géographique ou encore le type d'activité exercé.

Les auteurs s'appuient sur l'idée selon laquelle, les prêts des emprunteurs qui ont un grand nombre de caractéristiques en commun (suivant l'ensemble des dimensions) doivent avoir de grandes probabilités d'être de la même classe.

Ainsi, il se pose les difficultés suivantes :
- Comment établir les relations entre les emprunteurs ?
- Comment déduire des caractéristiques à exploiter à partir de la nouvelle représentation ?

— Comment prédire la classe d'un prêt ?

3.1 Processus de construction du graphe

Pour établir les relations entre les emprunteurs, les auteurs construisent un graphe multicouche.

Pour construire un tel graphe, on fixe les dimensions considérées et les attributs associés à chacune de ces dimensions.

Dans le cas de l'article qui s'attarde sur les prêts agricoles, les deux dimensions choisis sont : la localisation géographique et les produits vendus par les agriculteurs.

Les attributs de la dimension localisation géographique peuvent être le district, l'arrondissement etc, et concernant la dimension produit, les attributs peuvent être les différents produits répertoriés.

Dans le graphe multicouche :
- chaque emprunteur a autant de nœuds qu'il y a de couches considérées
- les nœuds de chaque emprunteur sont tous reliés les uns aux autres
- chaque attribut d'une dimension a un nœud associé
- si un emprunteur est décrit par un attribut dans une dimension donnée, alors le nœud emprunteur de cette dimension est relié au nœud attribut associé
- la navigation d'une couche à une autre se fait en passant par les nœuds emprunteurs des différentes couches

3.2 Extraction des nouveaux descripteurs du graphe multicouche

Lorsque le graphe multicouche est construit, les nouveaux descripteurs du prêt sont calculés suite à des applications du PageRank Personnalisé sur le graphe résultat.

Les auteurs proposent 03 façons différentes de calculer les nouveaux descripteurs :
- **Intra-influence :** le PageRank Personnalisé est initialisé de manière à favoriser les relations intra-couche dans le processus de diffusion.
- **Inter-influence :** le PageRank Personnalisé est initialisé de manière à favoriser les relations inter-couche dans le processus de diffusion.
- **Influence-combinée :** le PageRank Personnalisé ne favorise pas un type de relation.

Un graphe multicouche M, ayant N nœuds, et L couches, correspond à une représentation de dimension N x N x L x L, ceci peut être résumé en une matrice carrée (N x L) x (N x L).

Dans l'article, les auteurs considèrent deux dimensions pour décrire les emprunteurs dans le graphe multicouche, à savoir la localité et les produits vendus par ces derniers.

Considérons un cas où nous avons 4 emprunteurs (des fermiers), 2 localités (localité des fermiers) et 3 produits (produits agricoles vendus par les fermiers). Dans ce cas de figure, nous avons 3 couches (Emprunteur, Localité et Produit), et nous avons 9 nœuds (4 nœuds emprunteurs + 2 nœuds localités + 3 nœuds produits), et donc la matrice carrée qui permet de représenter le graphe multicouche est de taille (9 x 3) x (9 x 3)

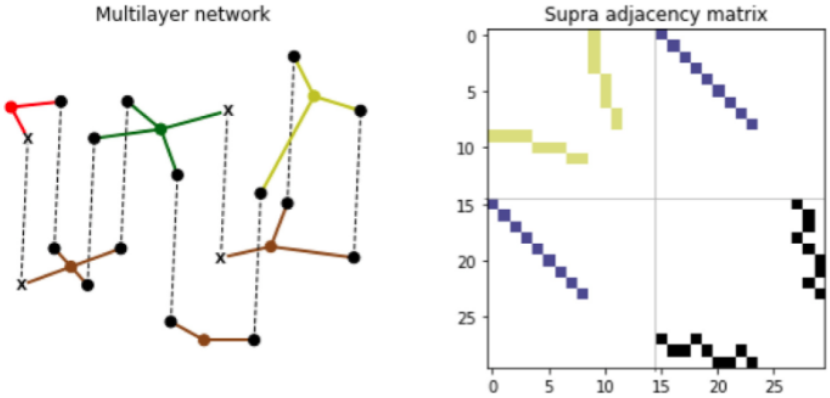

FIGURE 1 – A gauche on a un exemple de graphe multicouche à deux couches, et à droite on a la représentation de ce graphe sous forme de matrice. Ce graphe contient 9 emprunteurs (nœuds noirs), 3 nœuds de la dimension localité des emprunteurs (nœuds marron) et 3 nœuds de la dimension produits (vert, rouge, jaune). Les relations inter- couches sont matérialisées par des traits interrompus et existent uniquement entre les nœuds emprunteurs qui représentent le même emprunteur dans les différentes couches. Les relations intra-couche sont matérialisés par les autres types de trait (trait marron dans la dimension localité et les autres couleurs dans la dimension produit).

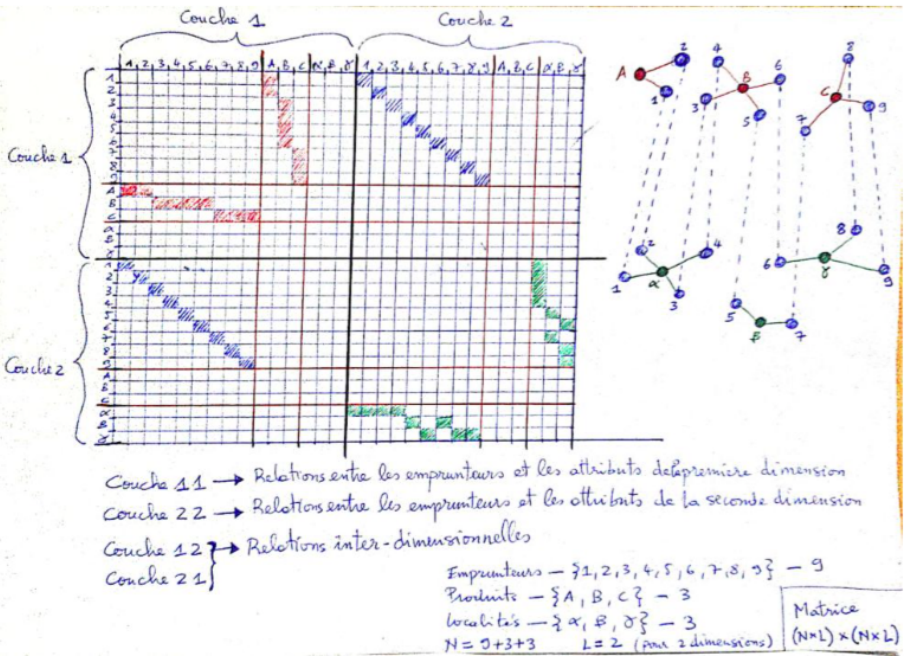

FIGURE 2 – Représente le même graphe multicouche précédent, mais avec des détails supplémentaires sur le procédé de construction du graphe et de la matrice associée. Un emprunteur (fermier) est relié à sa localité et aux produits agricoles qu'il commercialise.

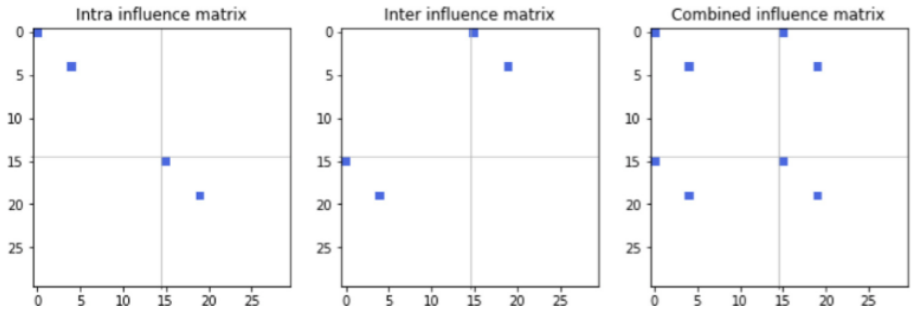

FIGURE 3 – Les trois scénarios considérés par les auteurs pour calculer les nouveaux descripteurs des prêts par l'application du PageRank Personnalisé sur le graphe multicouche.

Notons que d'autres descripteurs sont considérés dans le graphe multicouche, à l'exemple de :

- Nombre de nœuds emprunteurs qui vendent les mêmes produits que l'emprunteur cible
- Nombre de nœuds emprunteurs défaillants qui vendent les mêmes produits que l'emprunteur cible
- Nombre de nœuds emprunteurs de la même localité que l'emprunteur cible
- Nombre de nœuds emprunteurs défaillants de la même localité que l'emprunteur cible
- Nombre d'emprunteurs de la même localité que l'emprunteur cible et qui vendent les mêmes produits que lui
- Nombre d'emprunteurs défaillants de la même localité que l'emprunteur cible et qui vendent les mêmes produits que lui

3.3 Prédiction du risque de crédit avec les nouveaux descripteurs

Pour procéder à la prédiction du risque de crédit avec les modèles classiques d'apprentissage automatique, les descripteurs présents dans le jeu de données, et les nouveaux descripteurs extraits des graphes, sont utilisés comme données d'apprentissage des modèles classiques choisis (Régression logistique et XGBoost) pour la prédiction des risques de crédit.

Une fois que ces modèles sont construits, ces derniers sont utilisés pour prédire les classes des prêts du jeu de test. Dans l'article, les comparaisons des performances des modèles avant et après l'insertion des nouveaux descripteurs, montrent que les nouveaux descripteurs améliorent la qualité des prédictions.

Par ailleurs, les analyses sur l'explicabilité de ces modèles ont montré que les nouveaux descripteurs étaient parmi ceux qui contribuent le plus à la prise de décision des modèles de Régression logistique et XGBoost.

3.4 Limites et perspectives au travail sur le graphe multicouche

María Óskarsdóttir et Cristián Bravo[8] dans leur article ont proposé une modélisation des données des prêts en graphe multicouche. Cette approche a pour fort de s'apparenter à la réalité de la vie mais suscite encore des critiques qui définissent en réalité des limites á leur modélisation.

- Tous les attributs ne sont pas considérés dans le graphe multicouche.
- Le choix des attributs catégorielles à considéré comme couche du graphe multicouche est fait de façon arbitraire. Il serait intéressant de proposer un protocole qui permet de faire des choix pertinents.

— Les applications du PageRank Personnalisé sur le graphe multicouche ne sont pas suffisamment personnalisées, car le niveau de personnalisation reste à l'échelle intra et inter collectif (pour tous les nœuds concernés). Cependant, on pourrait avoir 03 applications différentes du PageRank Personnalisé pour chacun des prêts. Dans cette optique, seuls les nœuds qui sont liés au prêt courant peuvent recevoir le valeur 1.

Ses critiques nous permettent d'émettre ses possible extensions :

— Exploiter un graphe qui prend en compte tous les attributs descriptifs des prêts. On peut par exemple avoir un nœud pour chacune des modalités possibles de chaque attribut. Ensuite, relier tous les nœuds du même prêt ou alors incrémenter les poids des arcs qui relient tous les nœuds des modalités d'un prêt. Ensuite, appliquer le PageRank Personnalisé par chaque prêt, sur le graphe résultat, afin de ressortir avec de nouveaux descripteurs du prêt à l'exemple de l'estimer la classe de ce prêt par le PageRank.
— Proposer un protocole qui permet de choisir efficacement les attributs catégoriels à considérer dans le graphe multicouche.
— Personnaliser davantage les exécutions du PageRank Personnalisé de manière à affecter la valeur 1, uniquement aux nœuds associés au prêt courant pour lequel on calcul les valeurs des descripteurs.
— Faire ce travail sur plusieurs jeux de données

4 Expérimentation et résultats

4.1 Application des descripteurs du graphe multicouche sur le jeu de données AFB (Afriland First Bank)

Pour expérimenter les concepts appris, il est nécessaire de choisir des attributs qualitatifs qui vont représenter les dimensions (couches) du graphe multicouche. Nous avons donc recensé les attributs catégoriels du jeu de données AFB de Afriland First Bank.

Lorsqu'on ignore la classe des prêts, les attributs catégoriels de ce jeu de données sont :

— **Type / Motif :** le type de prêt ou motif du prêt bancaire
— **Fonction :** le métier ou l'occupation de l'emprunteur
— **Civilité :** civilité de l'emprunteur (Monsieur, Madame, Mademoiselle)
— **Statut matrimonial :** statut matrimonial de l'emprunteur (Célibataire, Marié, Divorcé)

Après avoir recensé les attributs catégoriels, nous avons choisi d'implémenter trois graphes multicouches à deux couches. Le premier graphe multicouche nommé ici **MLN1** est construit à partir des attributs **Fonction & Civilité**, le second graphe **MLN2** est construit à partir des attributs **Fonction & Statut-Matrimonial** et enfin, le troisième graphe **MLN3** est construit à partir des attributs **Fonction & Motif**.

Les attributs extraits des ces différents graphes multicouches sont énumérés comme suit :

4.1.1 Attributs du MLN1 - Fonction & Civilité

— **MLN_fonction_degré :** le nombre d'emprunteur avec la même fonction
— **MLN_civilité_degré :** le nombre d'emprunteur avec la même civilité
— **MLN_fonction_et_civilité_degré :** le nombre d'emprunteur avec la même fonction et civilité
— **MLN_bipart_intra_fonction_civilité :** le score PageRank maximal entre le noeud de l'emprunt de couche civilité et fonction lorsque seul les noeuds intra (modalités civilité et fonction) sont

inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_civilité :** le score PageRank maximal entre le noeud de l'emprunt de couche civilité et fonction lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_civilité :** le score PageRank maximal entre le noeud de l'emprunt de couche civilité et fonction
- **MLN_bipart_intra_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt lorsque seul les noeuds intra (modalités de fonction) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_max :** le score PageRank maximal de la fonction associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt
- **MLN_bipart_intra_civilité_max :** le score PageRank maximal du noeud de civilité associé à un emprunt lorsque seul les noeuds inter (modalités de civilité) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_civilité_max :** le score PageRank maximal du noeud de civilité associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_civilité_max :** le score PageRank maximal du noeud de civilité associé à un emprunt

4.1.2 Attributs du MLN2 - Fonction & Statut-Matrimonial

- **MLN_fonction_degré :** le nombre d'emprunteur avec la même fonction
- **MLN_sit_matrim_degré :** le nombre d'emprunteur avec le même statut matrimonial
- **MLN_fonction_et_sit_matrim_degré :** le nombre d'emprunteur avec les mêmes fonctions et statut matrimonial
- **MLN_bipart_intra_fonction_sit_matrim :** le score PageRank maximal entre le noeud de l'emprunt de couche sit_matrim et fonction lorsque seul les noeuds intra (modalités sit_matrim et fonction) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_sit_matrim :** le score PageRank maximal entre le noeud de l'emprunt de couche sit_matrim et fonction lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_sit_matrim :** le score PageRank maximal entre le noeud de l'emprunt de couche sit_matrim et fonction
- **MLN_bipart_intra_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt lorsque seul les noeuds intra (modalités de fonction) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_max :** le score PageRank maximal de la fonction associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt
- **MLN_bipart_intra_sit_matrim_max :** le score PageRank maximal du noeud de la sit_matrim associé à un emprunt lorsque seul les noeuds inter (modalités de sit_matrim) sont inclus dans le

vecteur de personnalisation du PageRank
- **MLN_bipart_inter_sit_matrim_max :** le score PageRank maximal du noeud de la sit_matrim associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_sit_matrim_max :** le score PageRank maximal du noeud de la sit_matrim

4.1.3 Attributs du MLN3 - Fonction & Motif

- **MLN_fonction_degré :** le nombre d'emprunteur avec la même fonction
- **MLN_motif_degré :** le nombre d'emprunteur avec le même statut matrimonial
- **MLN_fonction_et_motif_degré :** le nombre d'emprunteur avec les mêmes fonctions et statut matrimonial
- **MLN_bipart_intra_fonction_motif :** le score PageRank maximal entre le noeud de l'emprunt de couche motif et fonction lorsque seul les noeuds intra (modalités motif et fonction) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_motif :** le score PageRank maximal entre le noeud de l'emprunt de couche motif et fonction lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_motif :** le score PageRank maximal entre le noeud de l'emprunt de couche motif et fonction
- **MLN_bipart_intra_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt lorsque seul les noeuds intra (modalités de fonction) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_fonction_max :** le score PageRank maximal de la fonction associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_fonction_max :** le score PageRank maximal du noeud de la fonction associé à un emprunt
- **MLN_bipart_intra_motif_max :** le score PageRank maximal du noeud du motif associé à un emprunt lorsque seul les noeuds inter (modalités de motif) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_motif_max :** le score PageRank maximal du noeud du motif associé à un emprunt lorsque seul les noeuds inter (emprunt ou emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_motif_max :** le score PageRank maximal du noeud du motif associé à un emprunt

4.1.4 Mise en œuvre : intégration des attributs extraits des graphes multicouches dans le processus de prédiction du risque de crédit

Nous avons considéré cinq modèles classiques de l'apprentissage automatique pour la prédiction du risque de crédit : Arbre de décision, Forêt Aléatoire d'arbre de décision (Random Forest), Régression logistique, XGBoost et SVM.

Pour chacun des graphes multicouches considérés (MLN1, MLN2, et MLN3), chaque modèle de prédiction est appliqué 04 fois. Chacune des applications du modèle diffère de l'autre par l'ensemble d'attributs

descripteurs considérés :

- **Classic** : les attributs considérés sont tous ceux fournis avec le jeu de données
- **Classic + MLNi** : on considère tous les attributs du jeu de données et on ajoute les douze autres attributs extraits du graphe multicouches **MLNi**
- **Classic − AMLNi** : on considère une partie des attributs fournis avec le jeu de données. Ceux qui sont liés aux dimensions du graphe multicouches **MLNi** sont ignorés. Par exemple, pour le cas MLN1, les attributs relatifs à Fonction et à Civilités seront complètement écartés de la phase d'apprentissage
- **Classic − AMLNi + MLNi** : on écarte les attributs relatifs aux dimensions du graphe multicouche MLNi, et on intégre les douze attributs extraits du graphe multicouche **MLNi**

Nous pouvons ainsi évaluer l'impact des attributs choisis dans le graphe multicouches associés à leur représentation standard fourni dans le jeu de données **(Classic + MLNi)**, sans leur représentation standard **(Classic − AMLNi + MLNi)**, et enfin évaluer l'impact de leur absence des données d'apprentissage des modèles **(Classic − MLNi)**.

4.2 Application du PageRank personnalisé pour chaque prêt

4.3 Choix du nombre de couche

Pour expérimenter l'impact du nombreux de couches sur la qualité des descripteurs choisis, nous avons construire un graphe multicouche à 1 seule couche ou un seul attribut catégorielle (MLN 1) et un autre graphe multicouche avec autant de couche qu'il y'a d'attributs de type dans l'ensemble de données (MLN All).

4.3.1 MLN 1

Comme mise en hypothèse, il peut s'avérer que les relations entre les emprunts soient essentiellement porter sur un seul niveau, caractéristique ou couche. Alors, pour verifier cela, nous avons modélisé nos données sous forme de graphes multicouche biparti à une seule couche.

Pour chaque attribut catégoriel du jeu de données, nous construisons ces graphes et extrayons les informations suivante où **case_k** représente le nom de l'attribut qui sert à construire le MLN 1 :

- **MLN_case_k_degré :** le nombre d'emprunteur avec la même case_k
- **MLN_bipart_intra_case_k :** le score PageRank du noeud de l'emprunt de couche case_k lorsque seul les noeuds intra (modalités motif) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_case_k :** le score PageRank du noeud de l'emprunt de couche case_k lorsque seul les noeuds intra (noeuds emprunt) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_combine_case_k :** le score PageRank du noeud de l'emprunt de couche case_k
- **MLN_bipart_intra_max_case_k :** le score PageRank du noeud de la modalité de couche case_k lorsque seul les noeuds intra (modalités motif) sont inclus dans le vecteur de personnalisation du PageRank
- **MLN_bipart_inter_max_case_k :** le score PageRank du noeud de la modalité de couche case_k lorsque seul les noeuds intra (noeuds emprunt) sont inclus dans le vecteur de personnalisation du PageRank

— **MLN_bipart_combine_max_case_k :** le score PageRank du noeud de la modalité de couche case_k
— **MLN_bipart_ultra_case_k :** le score PageRank du noeud de l'emprunt de couche case_k lorsque seul les noeuds associés à l'emprunt (un seul emprunteur) sont inclus dans le vecteur de personnalisation du PageRank
— **MLN_bipart_ultra_max_case_k :** le score PageRank du noeud de la modalité de couche case_k lorsque seul les noeuds associés à l'emprunt (un seul emprunteur) sont inclus dans le vecteur de personnalisation du PageRank

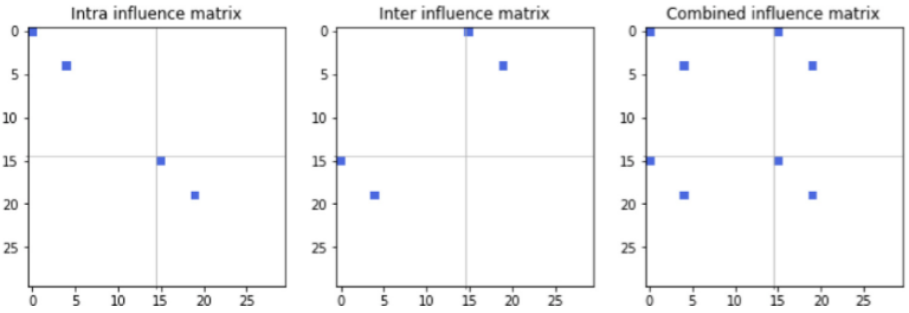

FIGURE 4 – impact des descripteurs issues des MLN 1 sur les métriques de modèles d'apprentissage automatique

Nous remarquons, que l'analyse des emprunts essentiellement permet déjà d'améliorer les métriques de classification.

4.3.2 MLN All

Nous expérimentons dans cette partie, l'impact d'une modélisation des emprunts avec autant de couches qu'il en a de données catégorielles sur l'évaluation du risque de crédit financier. Chaque couche dans cette modélisation met en relation chaque emprunt à sa modalité de la caractéristique de la couche et les différente couche communiquent entre elles par le biais des noeuds matérialisant l'emprunt. De l'extraction de métrique dans ce graphe, nous pouvons visualiser la figure suivante.

4.4 Choix des attributs à considérer

4.5 Les meilleurs performances

5 Conclusion et perspectives

Références

[1] D. A. Vega-Oliveros, P. S. Gomes, E. E. Milios, and L. Berton, "A multi-centrality index for graph-based keyword extraction," *Information Processing & Management*, vol. 56, no. 6, pp. 4–5, 2019.

[2] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of modern physics*, vol. 87, no. 3, pp. 925—-979, 2015.

[3] S. Mario, "Représentations graphiques des portefeuilles de crédit et leur analyse," *Revue européenne des sciences économiques et de gestion*, no. 1, pp. 23–28, 2021.

[4] X. Xu, C. Zhou, and Z. Wang, "Credit scoring algorithm based on link analysis ranking with support vector machine," *Expert systems with Applications*, vol. 36, no. 2, pp. 2625–2632, 2009.

[5] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk : a systemic review," *Neural Computing and Applications*, vol. 34, no. 17, pp. 14 327–14 339, 2022.

[6] M. Yıldırım, F. Y. Okay, and S. Özdemir, "Big data analytics for default prediction using graph theory," *Expert Systems with Applications*, vol. 176, p. 114840, 2021.

[7] C. Wang, F. Yu, Z. Zhang, and J. Zhang, "Multiview graph learning for small-and medium-sized enterprises' credit risk assessment in supply chain finance," *Complexity*, vol. 2021, pp. 1–13, 2021.

[8] M. Oskarsdottir and C. Bravo, "Multilayer network analysis for improved credit risk prediction," *Omega*, vol. 105, p. 102520, 2021.

[9] S. U. Rehman, A. U. Khan, and S. Fong, "Graph mining : A survey of graph mining techniques," in *Seventh International Conference on Digital Information Management (ICDIM 2012)*. IEEE, 2012, pp. 88–92.

[10] L. F. P. Tous, "Le crédit," Jan. 2023. [Online]. Available : https://www.lafinancepourtous.com/decryptages/marches-financiers/acteurs-de-la-finance/banque/la-banque-a-quoi-ca-sert/le-credit/

[11] ——, "La crise des subprimes (2007-2008)," Jan. 2024. [Online]. Available : https://www.lafinancepourtous.com/juniors/lyceens/les-crises/la-crise-des-subprimes-2007-2008/