

Extraction de descripteurs des graphes

Cas de graphes Multicouches biparti

DJIEMBOU TIENTCHEU VICTOR NICO¹

¹*Université de Yaoundé I,
Faculté des Sciences,
Département d'Informatique,
Étudiant Master SD*

20 mars 2024



1 Graphes Multi-couches Biparti

- Définition
- Descripteurs
- Processus d'extraction des descripteurs

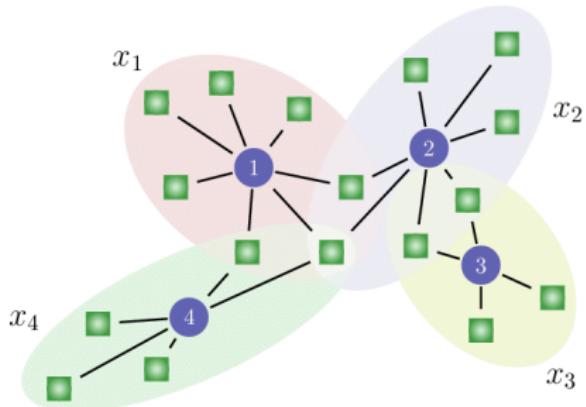
2 Application processus de l'extraction de descripteurs des graphes

- Données - Pré-traitement
- Construction d'un graphe multi-couche (MultiLayer Network (MLN)) biparti à k couches
- Extraction de descripteurs - Augmentation de données
 - Extraction de descripteurs du descripteur Degrée
 - Extraction du descripteur PageRank personnalisé
 - Augmentation de données

Graphes Multicouches Biparti

Définition

- **Noeud** : L'identificateur d'un emprunteur ou encore une modalité d'un attribut descriptif
- **Arête** : Matérialise qu'il existe une relation entre deux noeuds
- **Niveau ou Couche** : Ensemble d'emprunteur connecté à une modalité d'un attribut descriptif



Descripteurs

- **Centralité de degré**
- **Centralité de proximité**
- **Centralité d'intermédiairité**
- **Centralité de vecteur propre**
- **PageRank**

Descripteurs

Centralité de degré

Ensemble des utilisateurs ayant les mêmes caractéristiques.

PageRank

$$PR_{i+1} = \frac{1-d}{N} + d \times PR_i \times H \quad (1)$$

où :

PR_i est le PageRank à l'itération i,

PR_{i+1} est le PageRank à l'itération i+1,

d est le facteur d'amortissement (typiquement $d = 0,85$),

N est le nombre total de pages dans le réseau

H est la matrice de transition du réseau .

Descripteurs

PageRank Personnalisé

$$PR_{i+1} = \frac{1-d}{N} + d \times PR_i \times H \times P \quad (2)$$

où :

PR_i est le PageRank à l'itération i,

PR_{i+1} est le PageRank à l'itération i+1,

d est le facteur d'amortissement (typiquement $d = 0,85$),

N est le nombre total de pages dans le réseau

H est la matrice de transition du réseau .

P est le vecteur de probabilités de démarrage.

Extraction du descripteur du PageRank personnalisé

Algorithme PageRank personnalisé avec l'approche itérative

```
1 : procedure PageRank(d,P)
2 :     Initialise  $PR$  avec des valeurs égales pour chaque noeud
3 :     repeat
4 :          $PR_{new} = \frac{1-d}{N} + d \times PR \times H \times P$ 
5 :         Mettre à jour les valeurs de  $PR$  :  $PR \leftarrow PR_{new}$ 
6 :     until convergence
7 :     return  $PR$ 
8 : end procedure
```

Critères de convergence

- Stagnation des solutions (epsilon= 0.0001)
- Nombre d'itération
- ...

Extraction du descripteur de Degr e 

Algorithme d'extraction de la centralit  de degr e

```
1 : procedure Degree
2 :     Initialise  $DG$  de taille gale au nombre de noeud emprunteurs
3 :     for chaque noeud  $i$  do
4 :          $DG[i] \leftarrow$  nombre d'emprunteurs partageant les m mes informations
5 :     end for
6 :     return  $DG$ 
7 : end procedure
```

Données - Pré-traitement

Données Traditionnel

ID	person_home_ownership	loan_intent	loan_status
0	RENT	PERSONAL	1
1	OWN	EDUCATION	0
2	MORTGAGE	MEDICAL	1
3	RENT	MEDICAL	1
4	OWN	VENTURE	1
5	MORTGAGE	EDUCATION	0
6	RENT	MEDICAL	0
7	RENT	PERSONAL	1
8	OWN	HOMEIMPROVEMENT	1

Table – Un extrait du jeu de données Crédit Risk du benchmark

Données - Pré-traitement

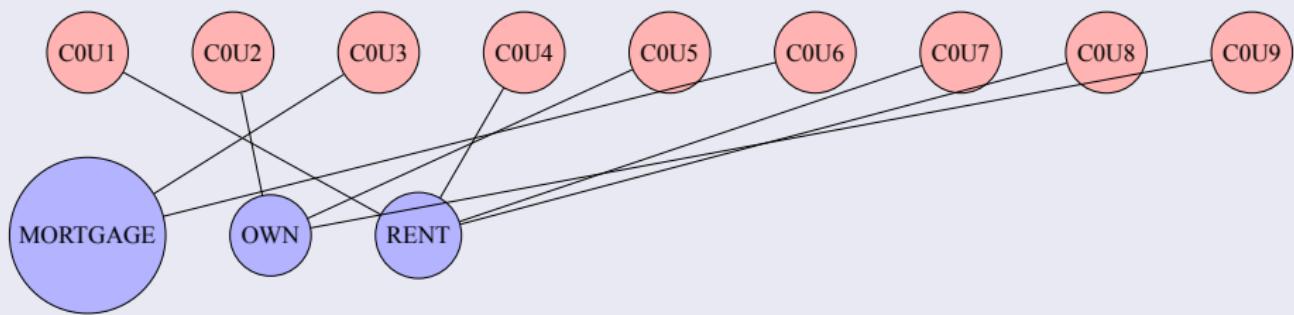
Données Binarisé

ID	RENT	OWN	MORTGAGE	PERSONAL	EDUCATION	MEDICAL	VENTURE	HOMEIMPROVEMENT	loan_status
0	1	0	0	1	0	0	0	0	1
1	0	1	0	0	1	0	0	0	0
2	0	0	1	0	0	1	0	0	1
3	1	0	0	0	0	1	0	0	1
4	0	1	0	0	0	0	1	0	1
5	1	0	1	0	1	0	0	0	0
6	1	0	0	0	0	1	0	0	0
7	1	0	0	1	0	0	0	0	1
8	0	1	0	0	0	0	0	1	1

Table – Version OHE de l'extrait du jeu de données Crédit Risk du benchmark

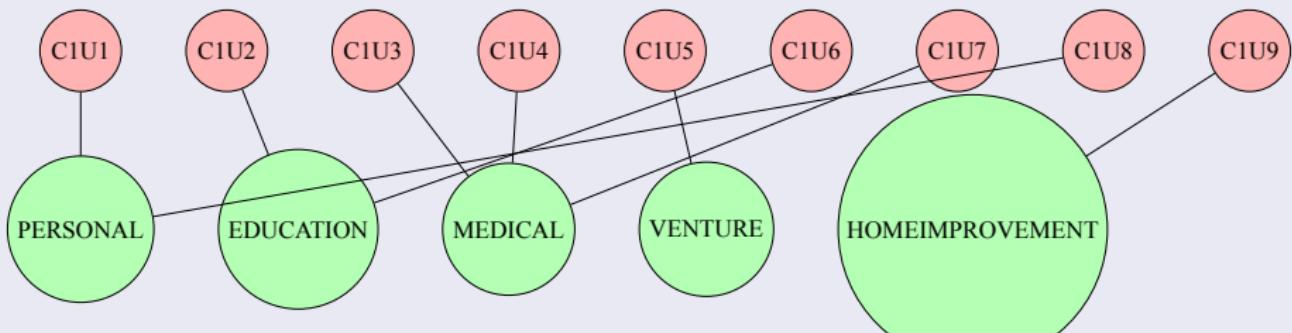
Construction d'un graphe multi-couches biparti à k couches

MLN à 1 couche de person_home_ownership



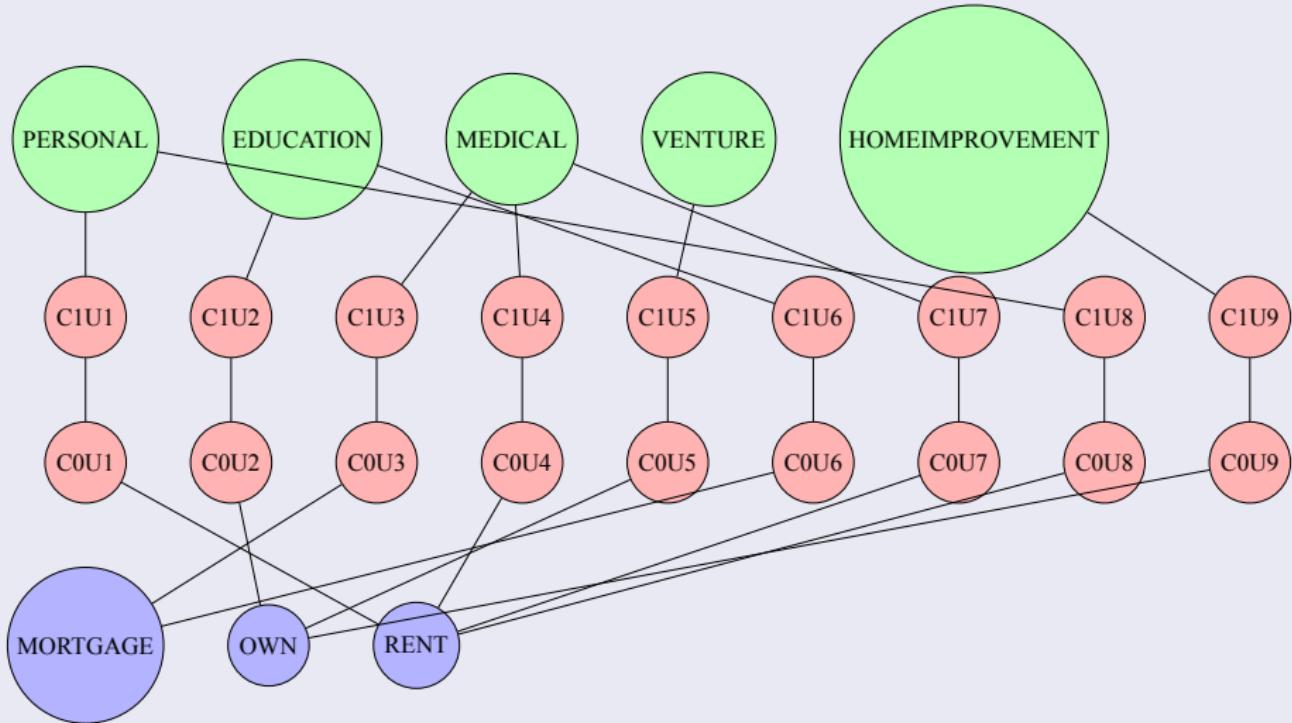
Construction d'un graphe multi-couches à k couches

MLN à 1 couche de loan_intent



Construction d'un graphe multi-couches à k couches

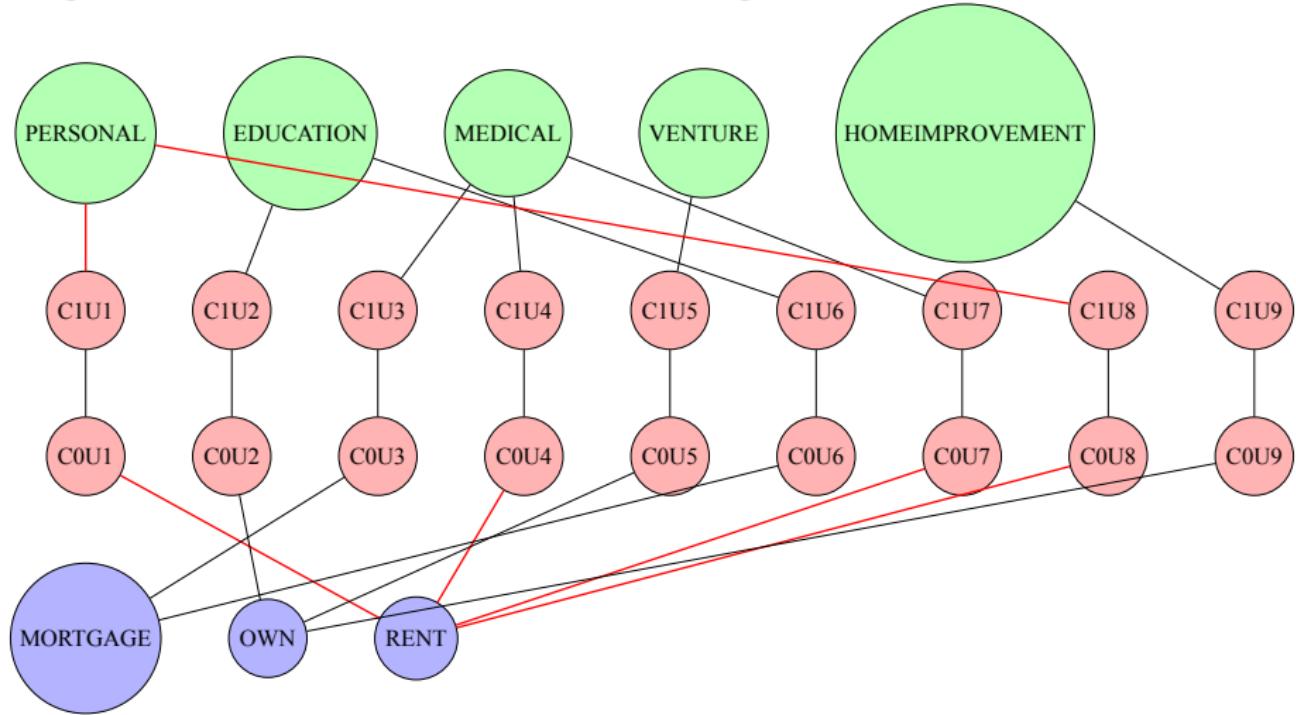
MLN à 2 couches de loan_intent et de person_home_ownership



Extraction de descripteurs du descripteur Degr e

$$Degree_{U_i} = |\{U_i : \exists(U_i, k_i)\}| - 1 \quad (3)$$

tel que k_i soit une des modalit s associ es   l'emprunteur U_i dans les k couches



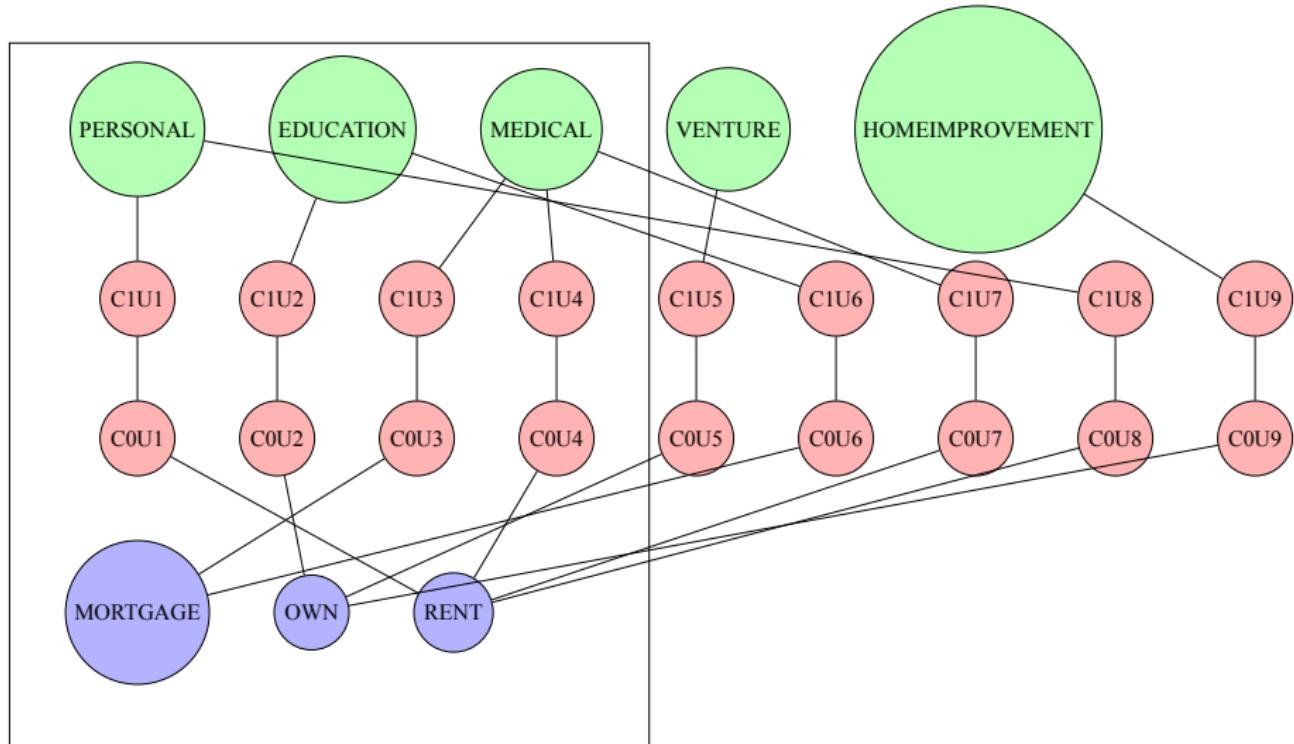
Extraction de descripteurs du descripteur Degr e

Degr e de chaque emprunteur

- $DG_{U1} = |\{U8, U1\} \cap \{U8, U1, U7, U4\}| - 1 = 1$
- $DG_{U2} = |\{U6, U2\} \cap \{U5, U2, U9\}| - 1 = 0$
- $DG_{U3} = |\{U7, U3, U4\} \cap \{U6, U3\}| - 1 = 0$
- $DG_{U4} = |\{U7, U4, U3\} \cap \{U8, U1, U7, U4\}| - 1 = 1$
- $DG_{U5} = |\{U5\} \cap \{U5, U2\}| - 1 = 0$
- $DG_{U6} = |\{U2, U6\} \cap \{U6, U3\}| - 1 = 0$
- $DG_{U7} = |\{U7, U3, U4\} \cap \{U8, U1, U7, U4\}| - 1 = 1$
- $DG_{U8} = |\{U8, U1\} \cap \{U8, U1, U7, U4\}| - 1 = 1$
- $DG_{U9} = |\{U9\} \cap \{U5, U2, U9\}| - 1 = 0$

$$DG = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Extraction du descripteur PageRank personnalisé



Extraction du descripteur PageRank personnalisé

Variables et Paramètres

$$P_{U1} = \begin{bmatrix} PERSONAL : 1 \\ EDUCATION : 0 \\ MEDICAL : 0 \\ C1U1 : 1 \\ C1U2 : 0 \\ C1U3 : 0 \\ C1U4 : 0 \\ MORTGAGE : 0 \\ OWN : 0 \\ RENT : 1 \\ C0U1 : 1 \\ C0U2 : 0 \\ C0U3 : 0 \\ C0U4 : 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}; PR_0 = \frac{1}{14} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}; d = 0, 85; \text{convergence à 2 itérations.}$$

Extraction du descripteur PageRank personnalisé

Variables et Paramètres

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \text{ La matrice de transition}$$

Extraction du descripteur PageRank personnalisé

Itération 1 U1

$$PR = \frac{3}{280} + \begin{bmatrix} \frac{17}{280} \\ \frac{17}{280} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{17}{280} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{25} \\ \frac{1}{100} \\ \frac{1}{100} \\ \frac{1}{10} \\ \frac{1}{100} \end{bmatrix}$$

Extraction du descripteur PageRank personnalisé

Itération 2 – U1

$$PR = \frac{3}{280} + \begin{bmatrix} \frac{17}{500} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{17}{200} \\ \frac{17}{200} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{119}{2000} \\ \frac{119}{2000} \\ \frac{17}{2000} \\ \frac{17}{2000} \\ \frac{17}{2000} \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{21}{400} \\ \frac{1}{100} \\ \frac{1}{100} \\ \frac{1}{100} \\ \frac{207}{2000} \\ \frac{1}{100} \end{bmatrix}$$

Augmentation de données

Données Augmenté de descripteurs extraits de graphes

ID	RENT	OWN	MORTGAGE	PERSONAL	EDUCATION	MEDICAL	VENTURE	HOMEIMPROVEMENT	Degree	PPR	loan_status
0	1	0	0	1	0	0	0	0	1	$\frac{207}{2000}$	1
1	0	1	0	0	1	0	0	0	0	0	0
2	0	0	1	0	0	1	0	0	0	0	1
3	1	0	0	0	0	1	0	0	1	1	1
4	0	1	0	0	0	0	1	0	0	0	1
5	1	0	1	0	1	0	0	0	0	0	0
6	1	0	0	0	0	1	0	0	1	1	0
7	1	0	0	1	0	0	0	0	1	1	1
8	0	1	0	0	0	0	0	1	0	0	1

Table – Données augmentées

Conclusion

En résumé

Merci de votre attention

Victor DJIEMBOU¹

¹ *Université de Yaoundé I,
Faculté des Sciences,
Département d'Informatique,
Étudiant Master SD*