



# Projet : PageRank

1TR6

Guillaume Scotto

Thibault Pichel

P302

PageRank

# Démarche et choix réalisés

- Prise en main du sujet
- Recherche des structures et choix :
  - Matrice : tableau de vecteurs creux
  - Liens stockées dans un tableau dynamique à 2 dimensions
  - Tri par insertion, complexités ( $O(n)$ ,  $O(1)$ );
- Répartition des tâches de manière agile
- Développement des programmes et des tests

# Raffinage

R0 : Classer N pages selon l'algorithme PageRank

R1 : Calculer le vecteur  $P_i$  des poids réels de chaque page et classer les N pages en fonction, en utilisant une architecture peu coûteuse en mémoire

R2.1 : Lire le fichier contenant les couples (page source, page cible)

R2.2 : Calculer la matrice H

R2.3 : Effectuer I itérations de la relation du sujet pour calculer une valeur approchée de  $P_i$

R2.4 : Classer les pages du plus grand poids au plus faible

R3.1 : On lit le fichier et on enregistre les couples dans un tableau de dimension N par 2

R3.2.1 : On le tri par insertion et on supprimer les doublons

R3.2.2 : On entre les données dans la matrice creuse H

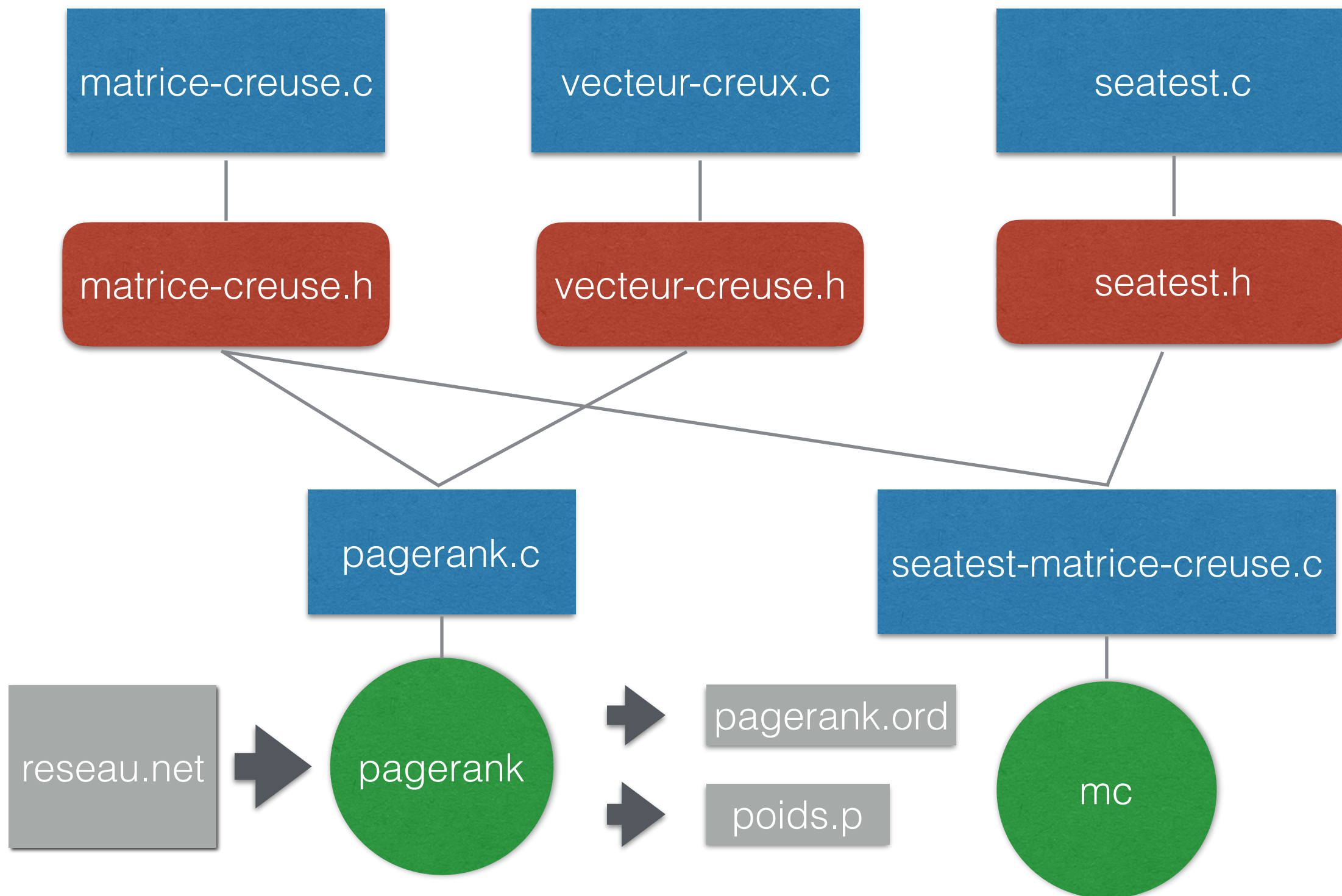
R3.3.1 : Initialiser  $P_{i0}$

R3.3.2 : On calcule la colonne i de G

R3.3.3 : On calcule la coordonnée i de  $P_i(k+1)$  qui vaut  $P_i(k)$  scalaire colonne\_i(G)

R3.4 : On tri les valeurs de  $P_i$

# Architecture du projet



# Principales sous-fonctions

- Initialisation, tri et remplissage de la matrice creuse
- Produit matriciel de  $P_i$  par  $G = f(H)$
- Fonctions de Tri
- Sous-fonctions pratiques : lecture du réseau, affichage de la matrice, lecture de la ligne de commande)

# Fonctions de lecture

Matrice H :

```
[3 , 0.500000] --> [4 , 0.500000] --> NULL
[2 , 0.333333] --> [3 , 0.500000] --> NULL
[4 , 0.500000] --> [5 , 1.000000] --> NULL
[0 , 0.500000] --> NULL
[0 , 0.500000] --> [2 , 0.333333] --> NULL
[2 , 0.333333] --> NULL
```

-->Lancement des itérations

Vecteur Pi approché

```
[0 , 0.000000] --> [1 , 0.000000] --> [2 , 0.000000] --> [3 , 0.000000] --> [4 , 0.000000] --> [5 , 0.000000] --> NULL
```

# Difficultés rencontrées

- Problème de temps du aux tests de vecteurs-creux
- Sens dans le produit matriciel et produit matriciel
- Lecture d'un fichier + problème “.net” ???
- Suppression des doublons

# Conclusion

- Projet très intéressant
- Première fois que l'on doit vraiment prendre en compte l'efficacité de nos programmes.
- Calcul des poids au fur et à mesure que les données sont stockées dynamiquement dans le tableau.
- Tableau dynamique à 2 dimensions
- Intérêt que ce soit une suite à notre projet précédent.
- Travail à deux enrichissant.