

La structure des graphes d'échanges entre entreprises

S. Contassot-Vivier, N. Fatès, J. Guichon

23/03/2023

Présentation des données : Graphes terrains

Jeu de données fourni par Infocert, entreprise de facturation électronique italienne

Composition totale :

- Plus de 400 000 entreprises anonymisées
- Environ 4 000 000 de factures en 2019 et 2020

Sous-jeux de données :

- Jeu sur 1 an avec 60 000 entreprises et 2 000 000 factures
- Jeu sur 1 mois avec 40 000 entreprises et 125 000 factures

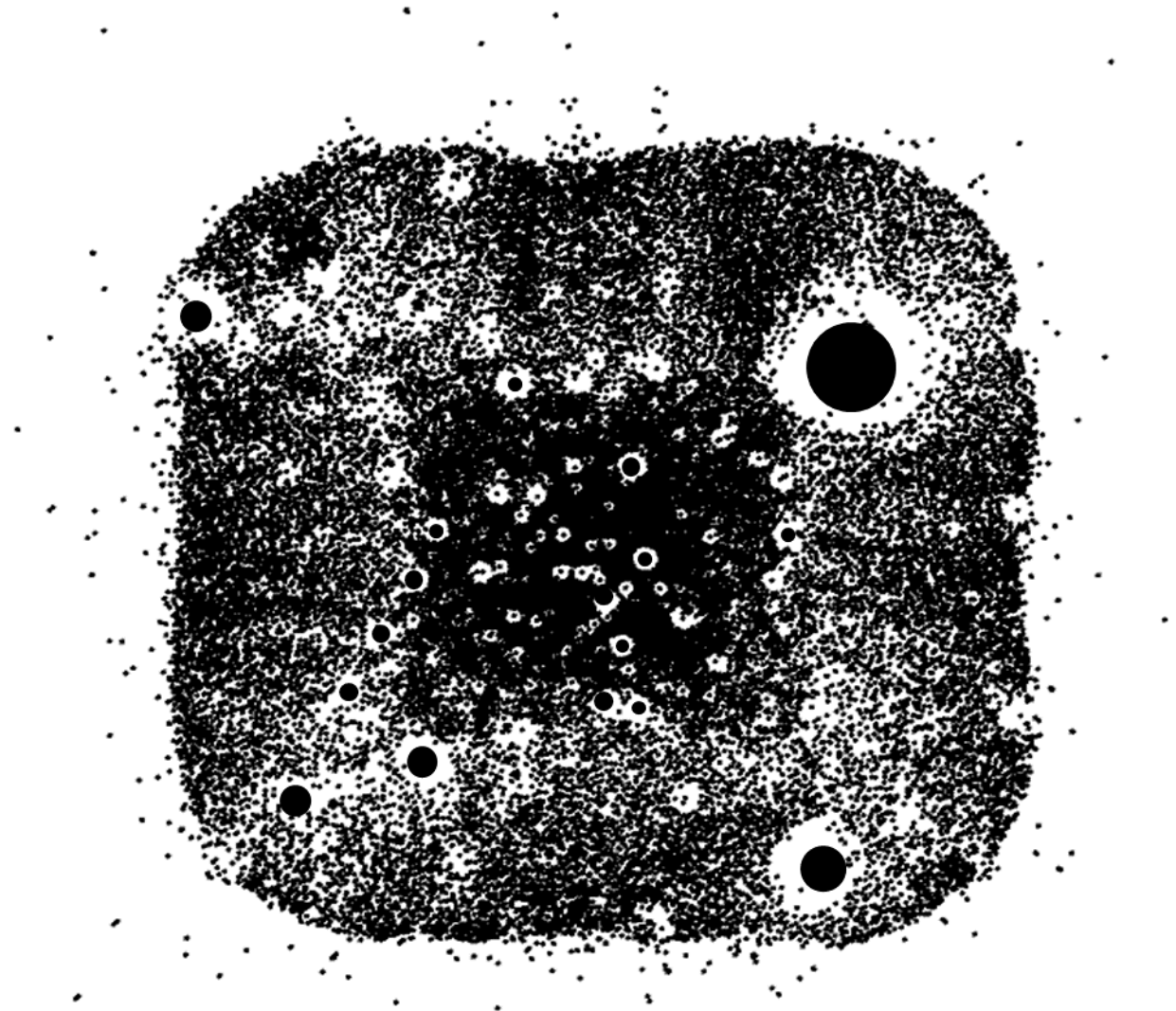
Multigraphes orientés pondérés jamais utilisés auparavant pour des études dans le cadre de la compensation de dettes

Graphe terrain : 1 mois

- 37 188 entreprises
- 124 665 factures

Limitations :

- Seulement Business to Business
- Taille réduite mais bon point de départ

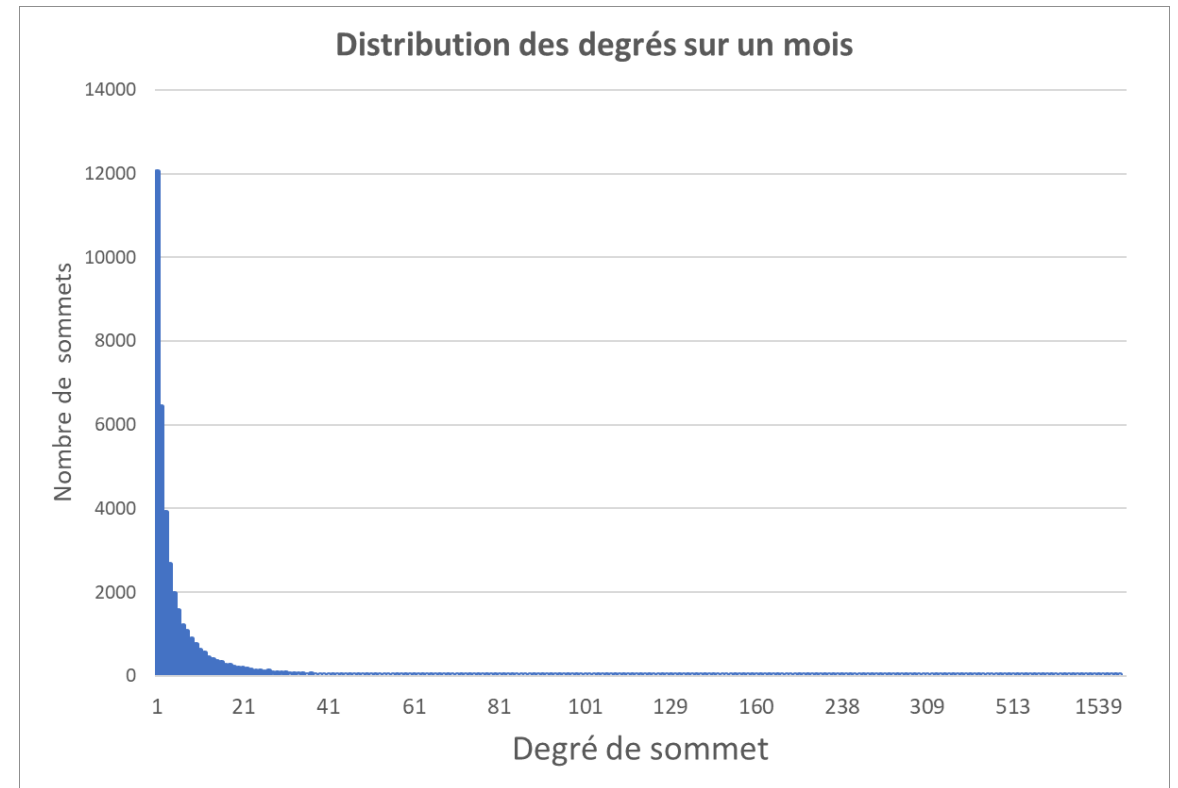


Visualisation du graphe terrain de un mois dans Gephi

Présentation des données : Graphes terrains

Une propriété particulière de ces graphes :

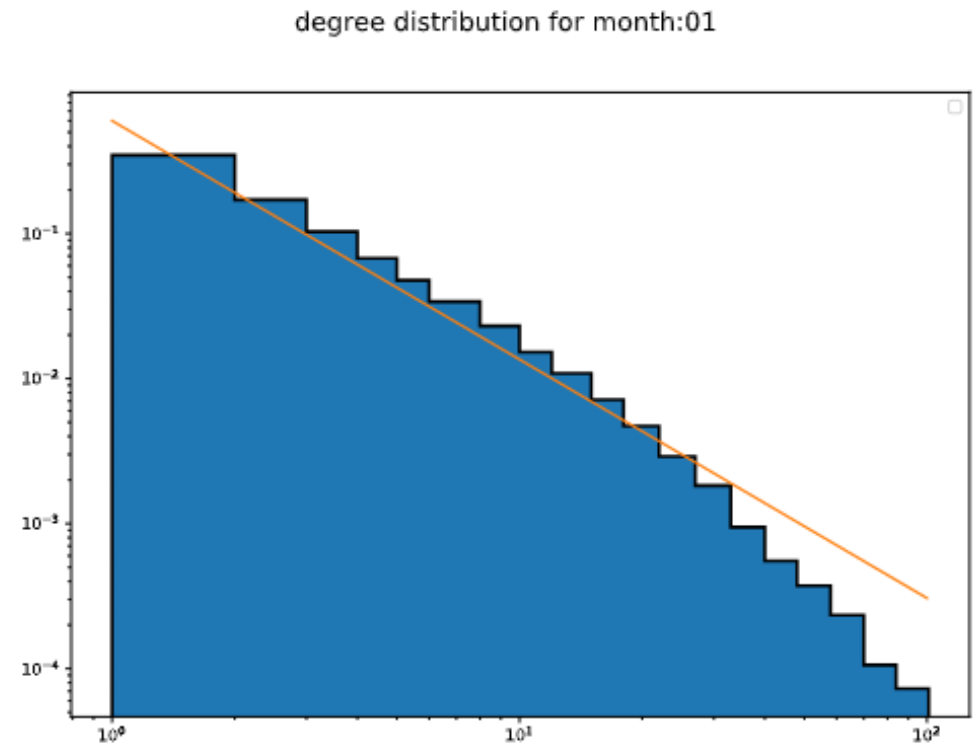
- Lois de puissance $P(k) = k^\gamma$ sur :
 - La distribution des degrés
 - La distribution du poids des arcs



Présentation des données : Graphes terrains

Une propriété particulière de ces graphes :

- Lois de puissance $P(k) = k^\gamma$ sur :
 - La distribution des degrés
 - La distribution du poids des arcs



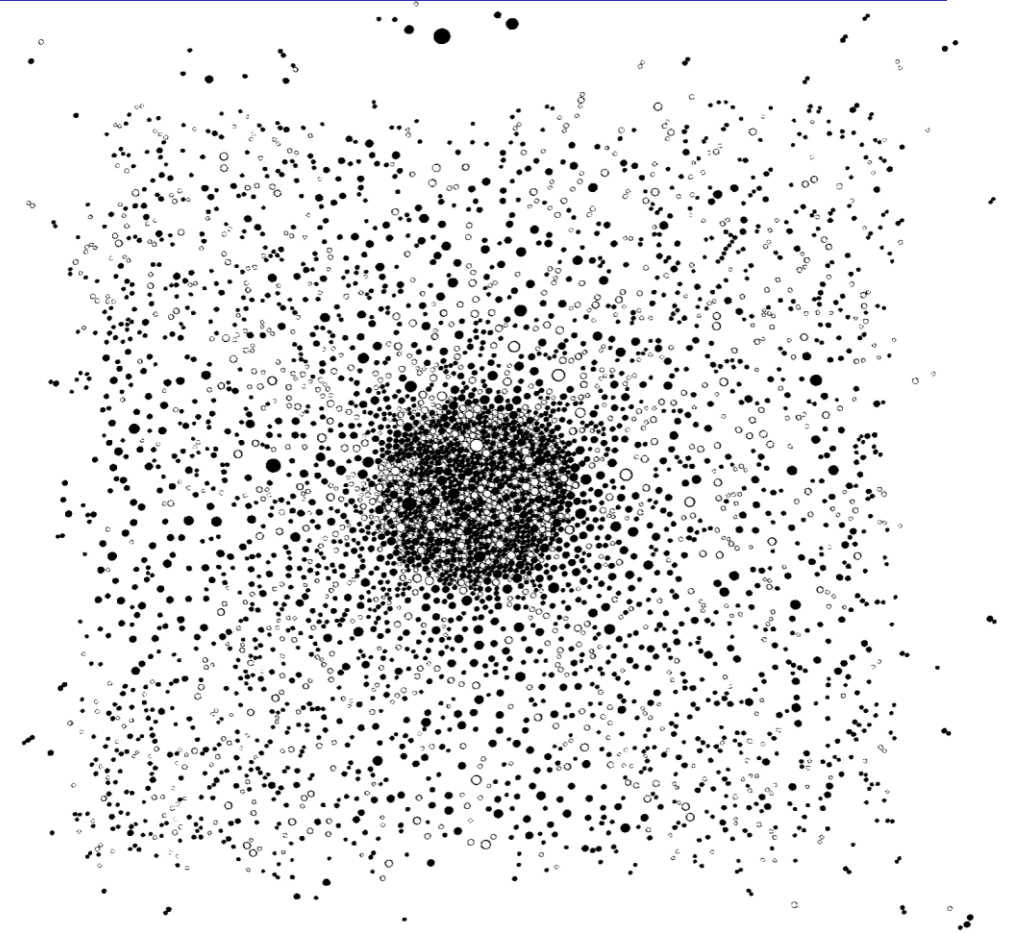
Présentation des données : Graphes d'imitation

Graphes générés avec des algorithmes réalisés par Arthur Rousseau pendant un stage

Première modélisation des graphes terrains en suivant des lois de puissance sur la distribution des degrés.

Limite :

- Imitation de certains attributs seulement



Visualisation d'un graphe d'imitation de 4137 nœuds et 16 681 arrêtes

Présentation des données : Graphes invariants d'échelle

Graphes utilisés pour modéliser de très nombreux réseaux dont des réseaux sociaux, vivants, Internet, etc.

■ Invariance d'échelle :

Une fonction g est dite invariante d'échelle s'il existe une fonction φ telle que pour tout x et y :

$$\frac{g(x)}{g(y)} = \varphi\left(\frac{x}{y}\right)$$

Alors, il existe une constante C et un exposant γ tels que : $g(x) = Cx^\gamma$.

En effet on a $\frac{g(x)}{g(y)} = \frac{Cx^\gamma}{Cy^\gamma} = \left(\frac{x}{y}\right)^\gamma$ donc $\varphi(x) = x^\gamma$ convient

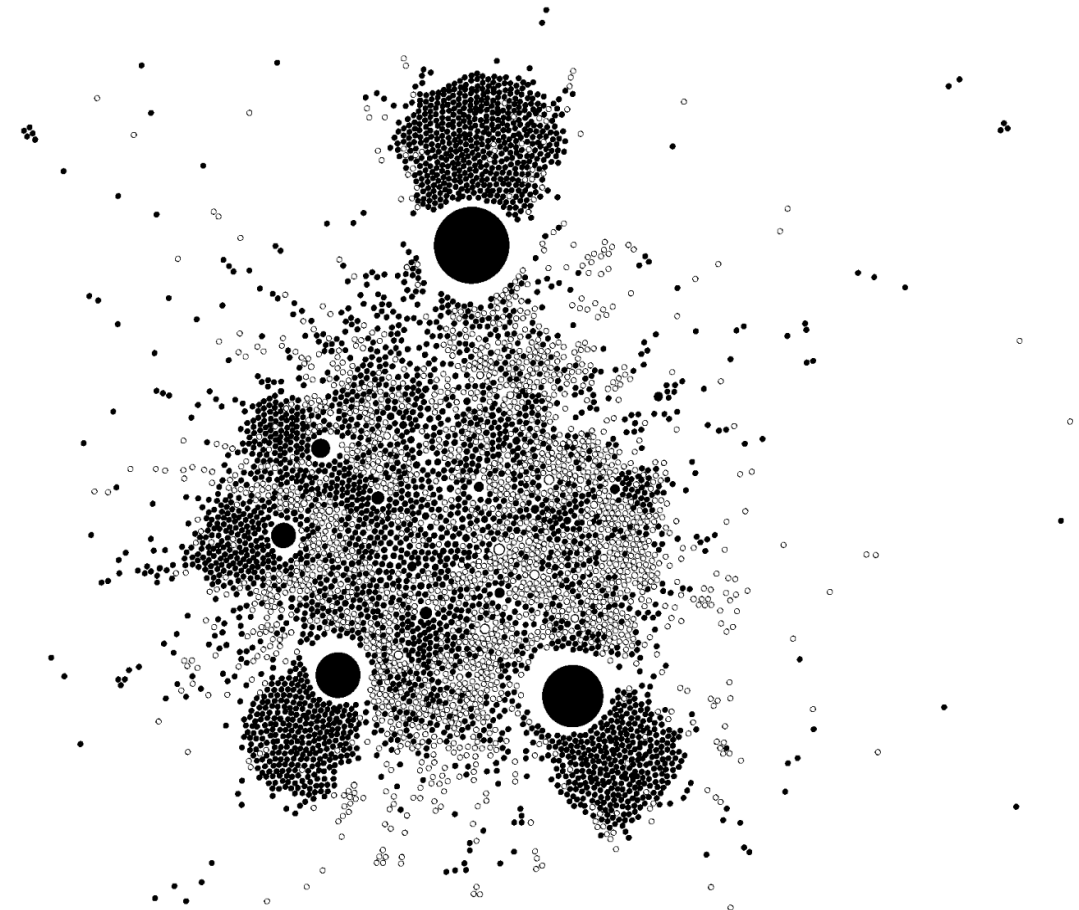


Présentation des données : Graphes invariants d'échelle

Plusieurs modèles de génération pour ces graphes dont celui défini par B.Bollobás et al. dans leur article “Directed scale-free graphs”, SODA, 2003

Possibilité de jouer sur les paramètres pour générer des réseaux précis.

On peut approcher l'exposant de la loi de puissance par recherche empirique.



Visualisation d'un graphe invariant d'échelle de 5000 nœuds et 14 856 arrêtes

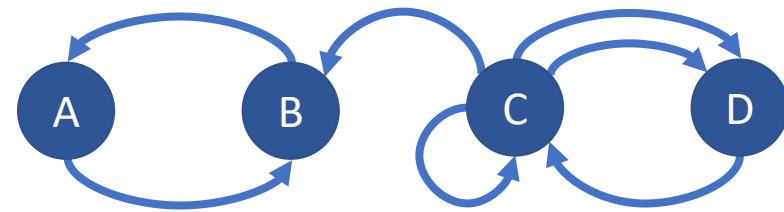
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



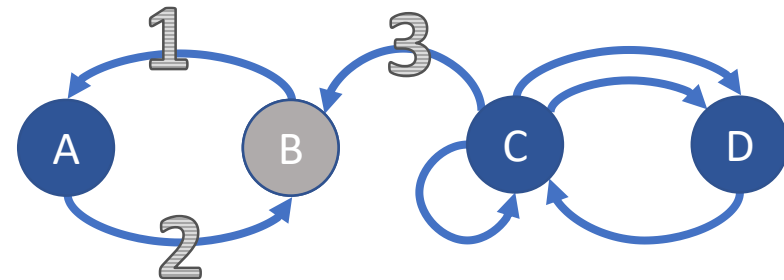
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



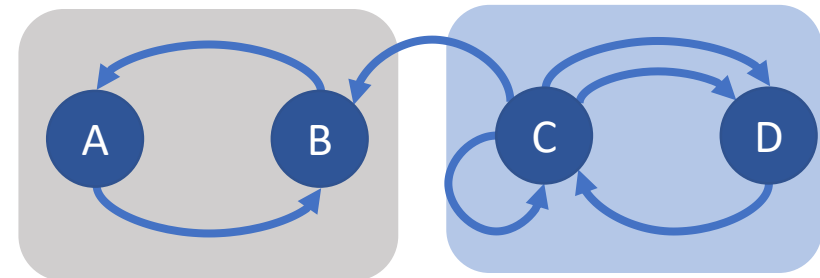
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



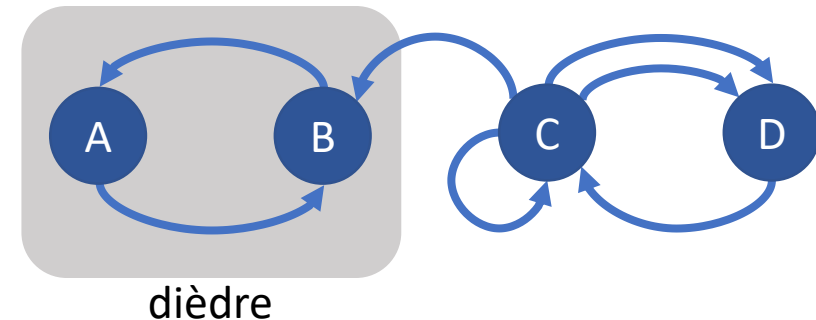
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



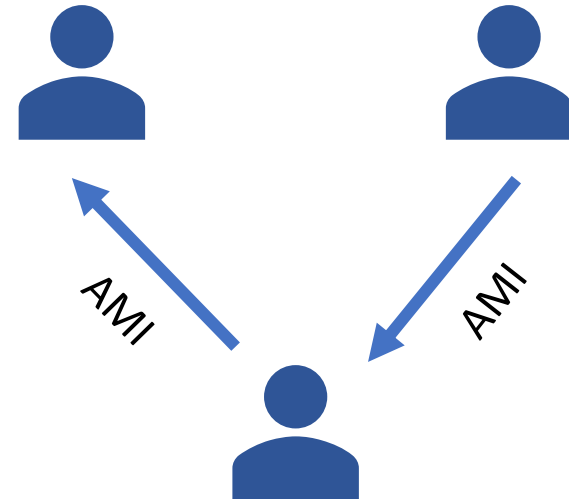
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



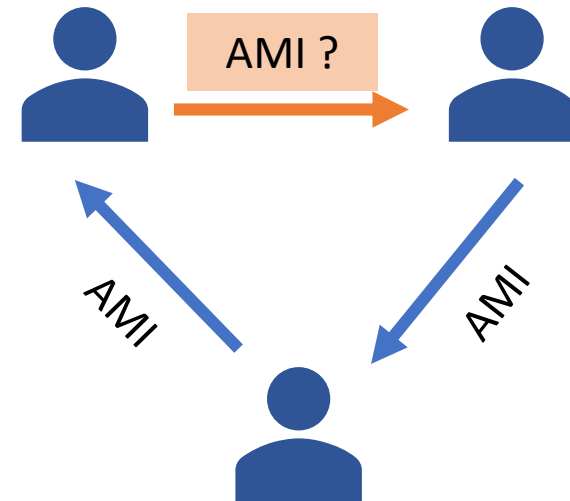
Analyse classique des graphes

Objectifs:

- Classer nos graphes dans des familles connues si possible
- Récupérer les attributs principaux des graphes terrain et les comparer à des graphes théoriques

Attributs observés :

- Degrés et leur répartition
- Composantes connexes
- Proportion de triangles et dièdres
- Coefficient de clustering



Exploration communautaire des graphes

Comment peut-on faire des regroupements sur nos graphes pour extraire des sous-graphes intéressants ?

Y a-t-il des structures particulières ?

Travail sur des Multigraphes orientés pondérés donc inhabituels

Essai de plusieurs méthodes :

- Réordonnancement de matrices
- Communautarisation par modularité

Exploration communautaire : Matrices d'adjacence

Matrice d'adjacence A du graphe G :

- Taille $n \times n$ (n : nombre de sommets)
- A_{ij} est le nombre d'arcs du sommet i au sommet j .

Première méthode :

- Réorganisation de matrices
 - Objectif : faire ressortir des formes
cf. M.Behrisch et al., « Matrix Reordering Methods for Table and Network Visualization », Wiley, 2016.

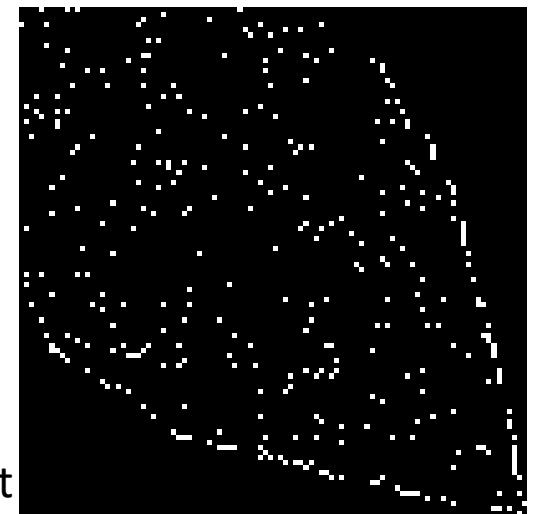
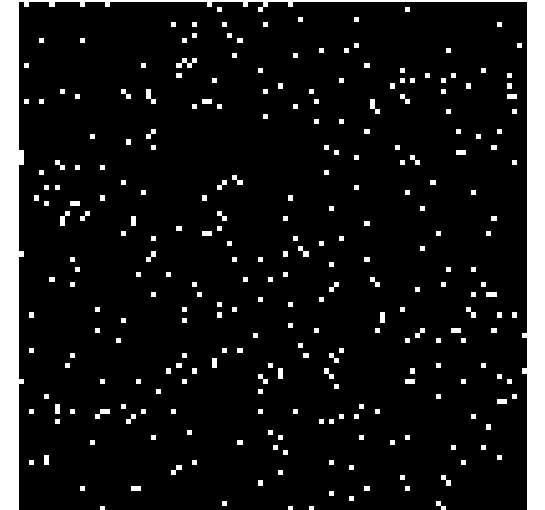
Exploration communautaire : Matrices d'adjacence

- Utilisation d'un algorithme pour réduire la bande passante de la matrice

→ Resserrer les liens autour de la diagonale.

- Limites :

- Difficultés d'interprétation surtout sur de grands graphes
- Formation de groupes assez faible
- Peu utilisé pour modéliser des réseaux
- Perte d'information sur le multi-arrêtes



Matrice d'adjacence d'un graphe aléatoire de 100 sommets et son réordonnancement

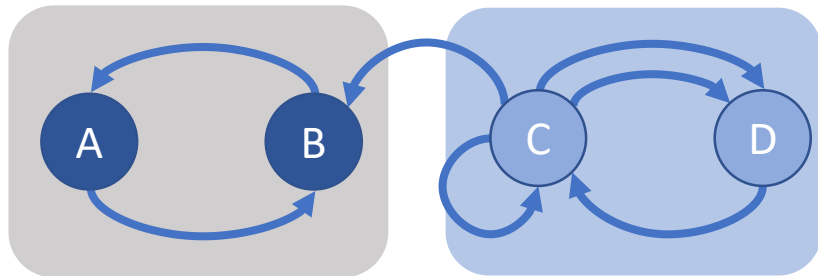
Exploration communautaire par modularité

Au lieu de réordonner on décide de partitionner le graphe :

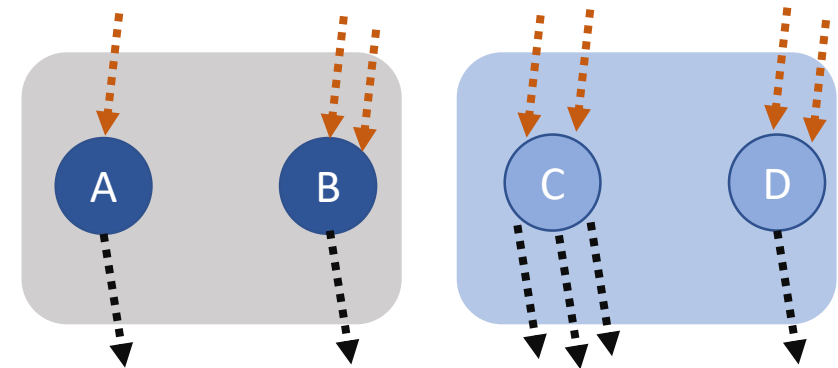
Mais comment juger de la qualité d'une partition ?

→ Modularité utilisée dans les graphes classiques pour mesurer la qualité

$$Q \propto \sum_{s \in S} (\text{nb liens dans } s - \text{nb liens attendus dans } s)$$



VS



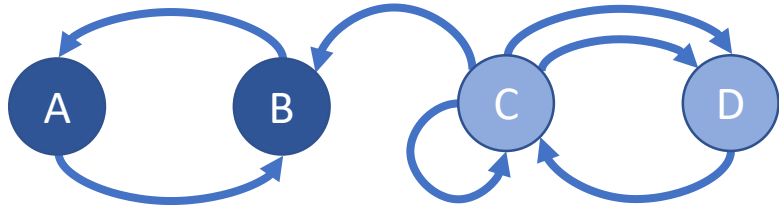
Exploration communautaire par modularité

On peut alors définir la formule suivante que l'on a adapté a notre cas particulier :

$$Q(G, S) = \frac{1}{2m} \sum_{s \in S} \sum_{i \in s} \sum_{j \in s} \left[A_{ij} - \frac{k_i^+ k_j^-}{2m} \right]$$

- G : graphe
- S : partitionnement
- s : communauté
- i et j : sommets
- A_{ij} : valeur de la matrice d'adjacence en (i, j)
- m : nombre d'arcs
- d_i^+ le degré sortant de i
- d_j^- le degré entrant de j

Exploration communautaire par modularité



$$S = \{ \{A; B\}; \{C; D\} \} \quad m = 7$$

$A =$

	A	B	C	D
A	0	1	0	0
B	1	0	0	0
C	0	1	1	2
D	0	0	1	0

	ds	de
A	1	1
B	1	2
C	4	2
D	1	2

$$Q(G, S) = \frac{1}{14} \left(\begin{aligned} & \left[\left(0 - \frac{1 * 1}{14} \right) + \left(1 - \frac{1 * 2}{14} \right) \right] \\ & + \left[\left(1 - \frac{1 * 1}{14} \right) + \left(0 - \frac{1 * 2}{14} \right) \right] \\ & + \left[\left(1 - \frac{4 * 2}{14} \right) + \left(2 - \frac{4 * 2}{14} \right) \right] \\ & + \left[\left(1 - \frac{1 * 2}{14} \right) + \left(0 - \frac{1 * 2}{14} \right) \right] \end{aligned} \right)$$

$$Q(G, S) = 0.296$$

Exploration communautaire par modularité

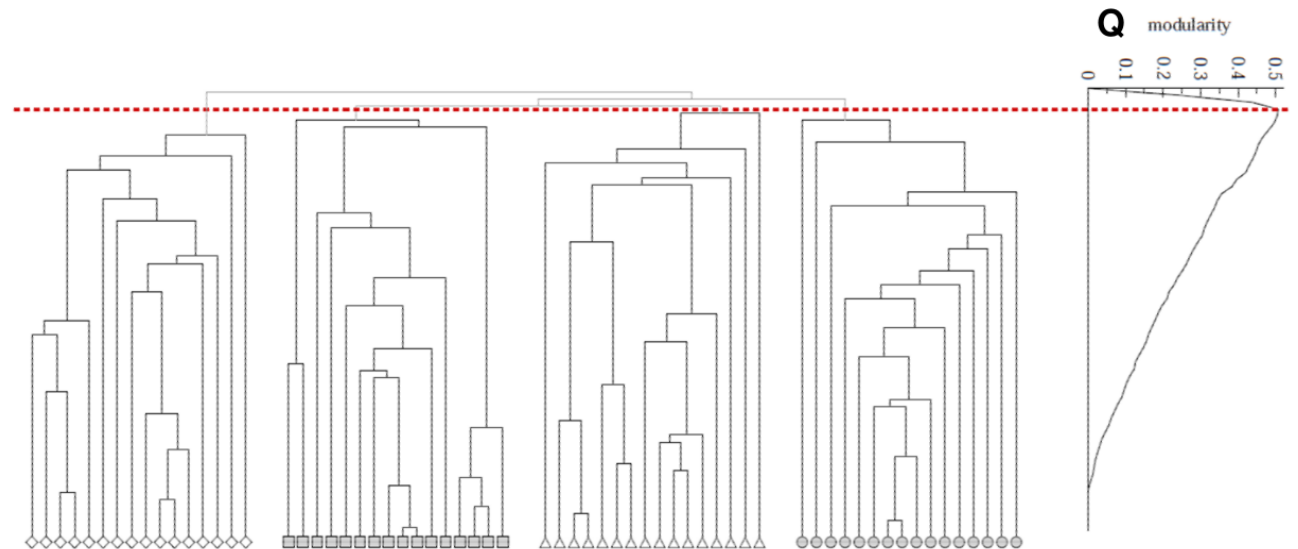
Impossible de tester les 2^n combinaisons de communautés...

→ Besoin de construire automatiquement ou en faisant moins de calculs

→ Algorithme de construction des communautés :

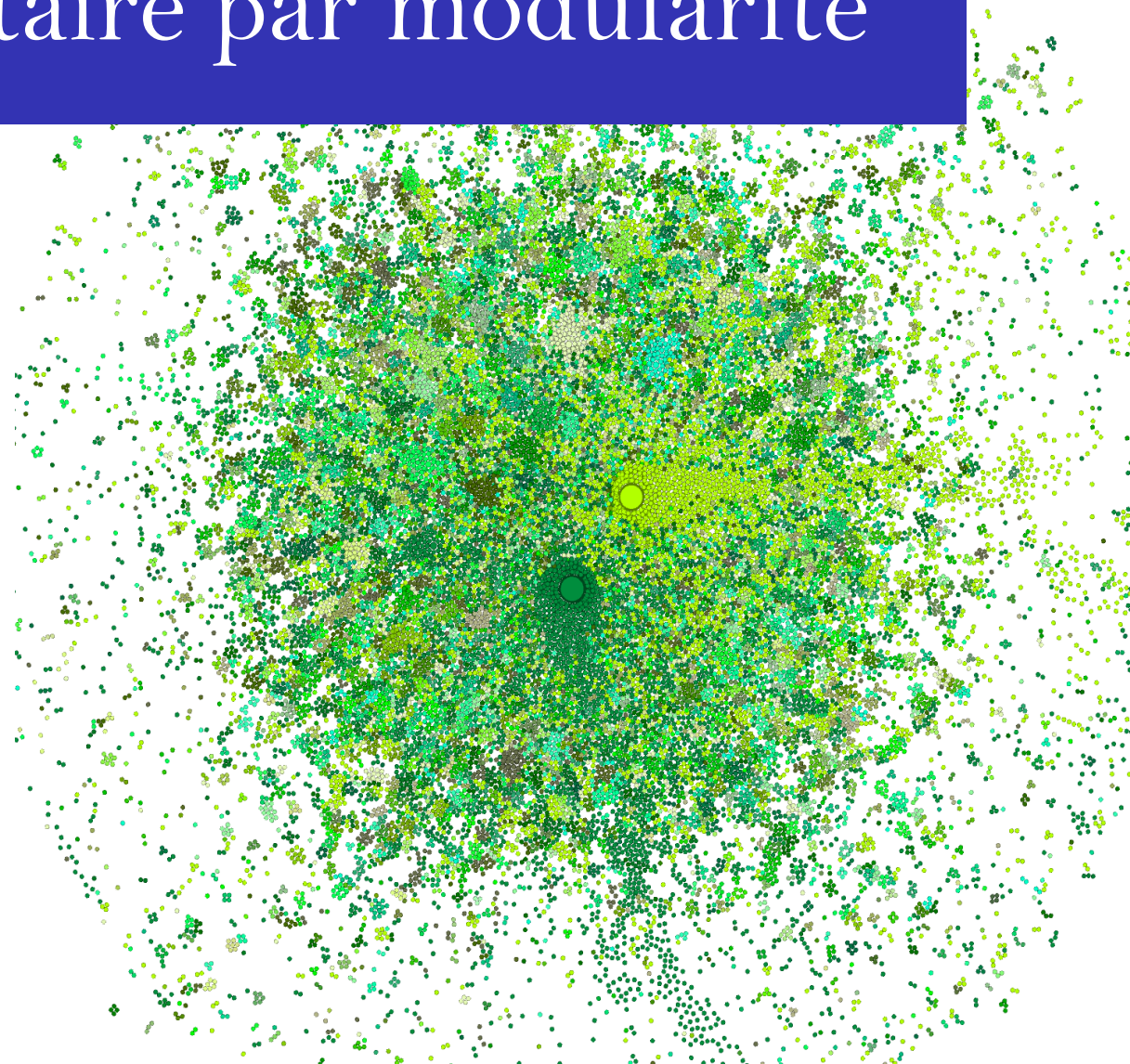
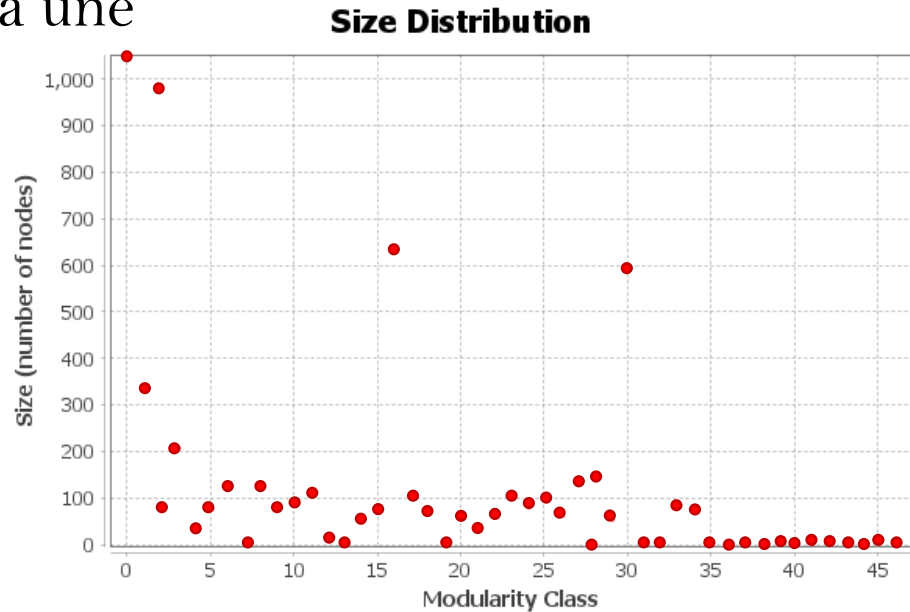
Chaque nœud est seul dans sa communauté au départ puis on fusionne les deux communautés qui améliorent le mieux la modularité jusqu'à ce que l'on puisse plus l'améliorer.

■ Résultat : partitionnement avec la meilleure modularité



Exploration communautaire par modularité

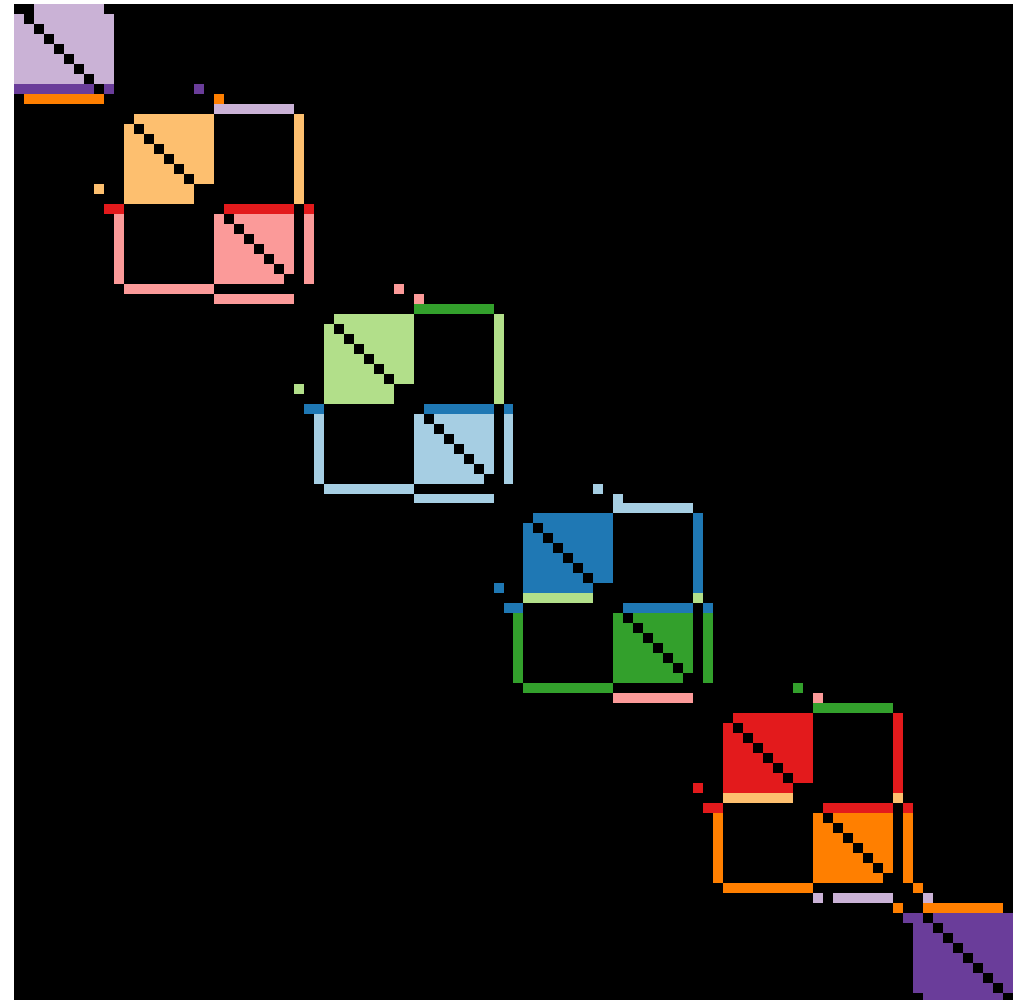
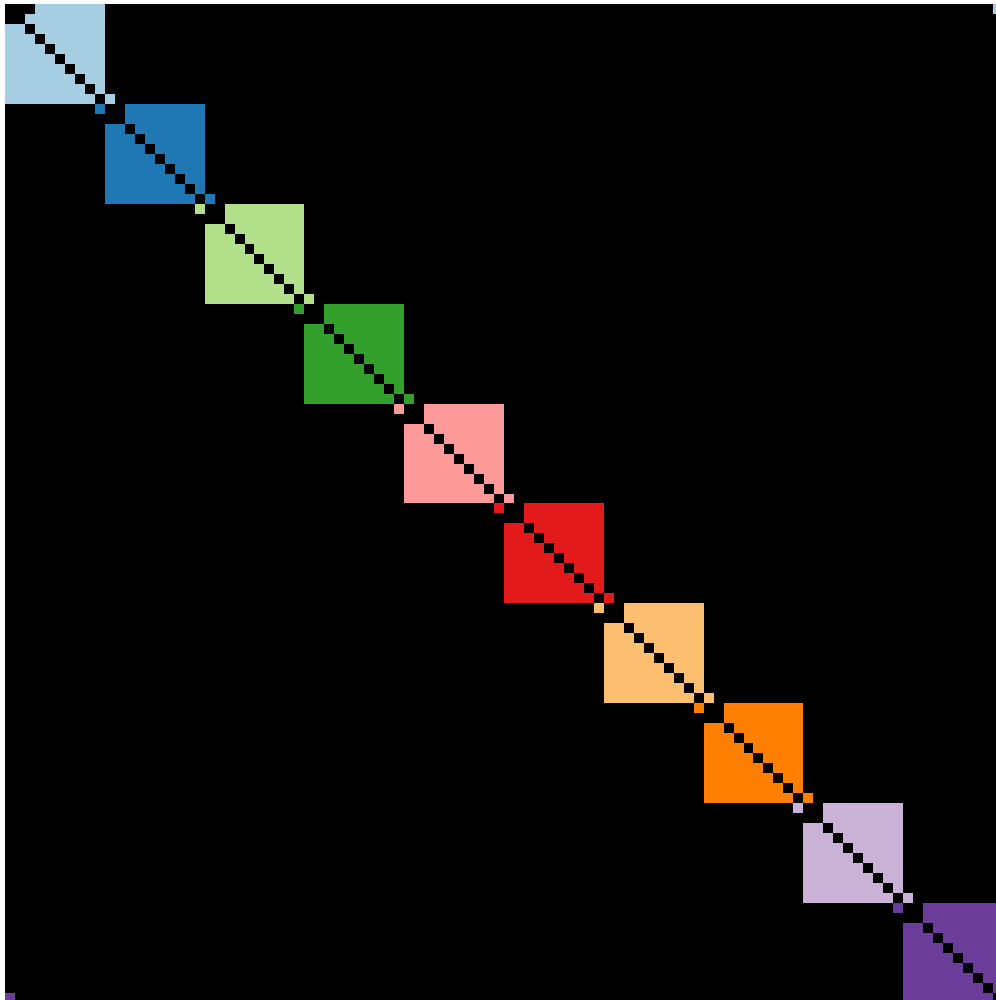
- Utilité :
 - Création de communautés d'acteurs liés
 - Coloration de nos graphes selon la partition obtenue
 - Etudes des grandes communautés une à une



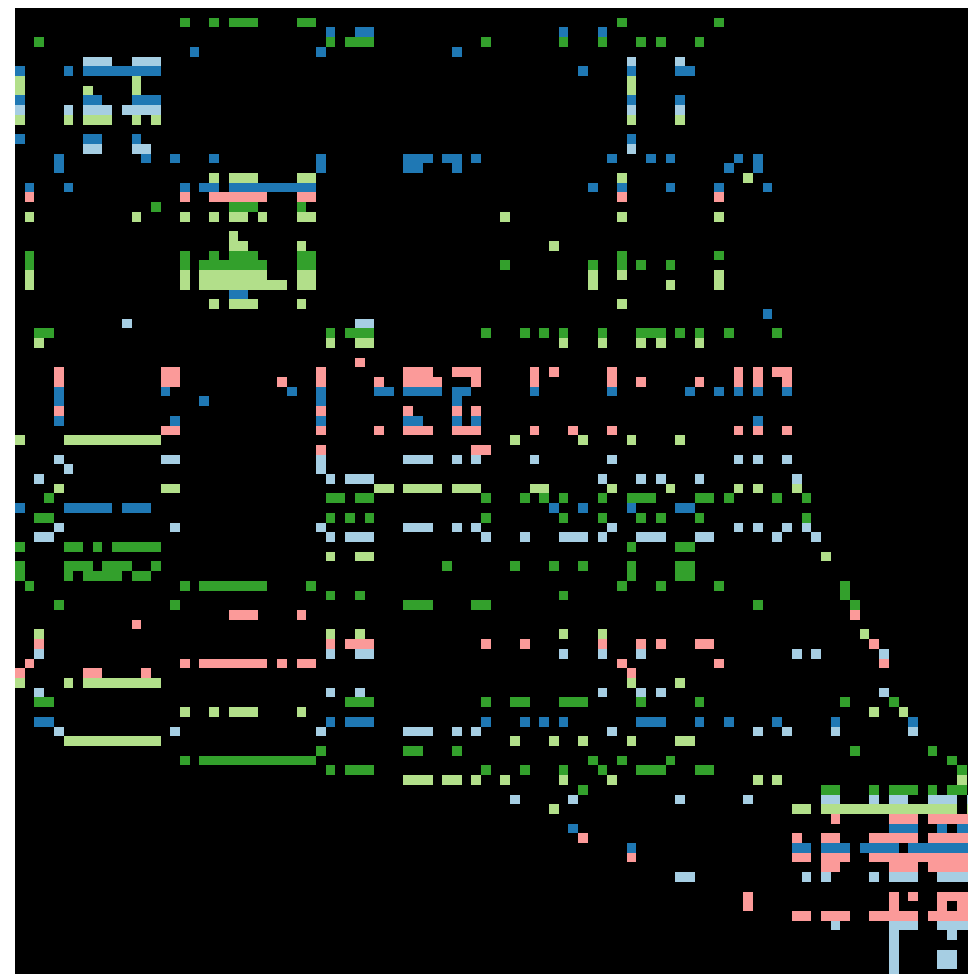
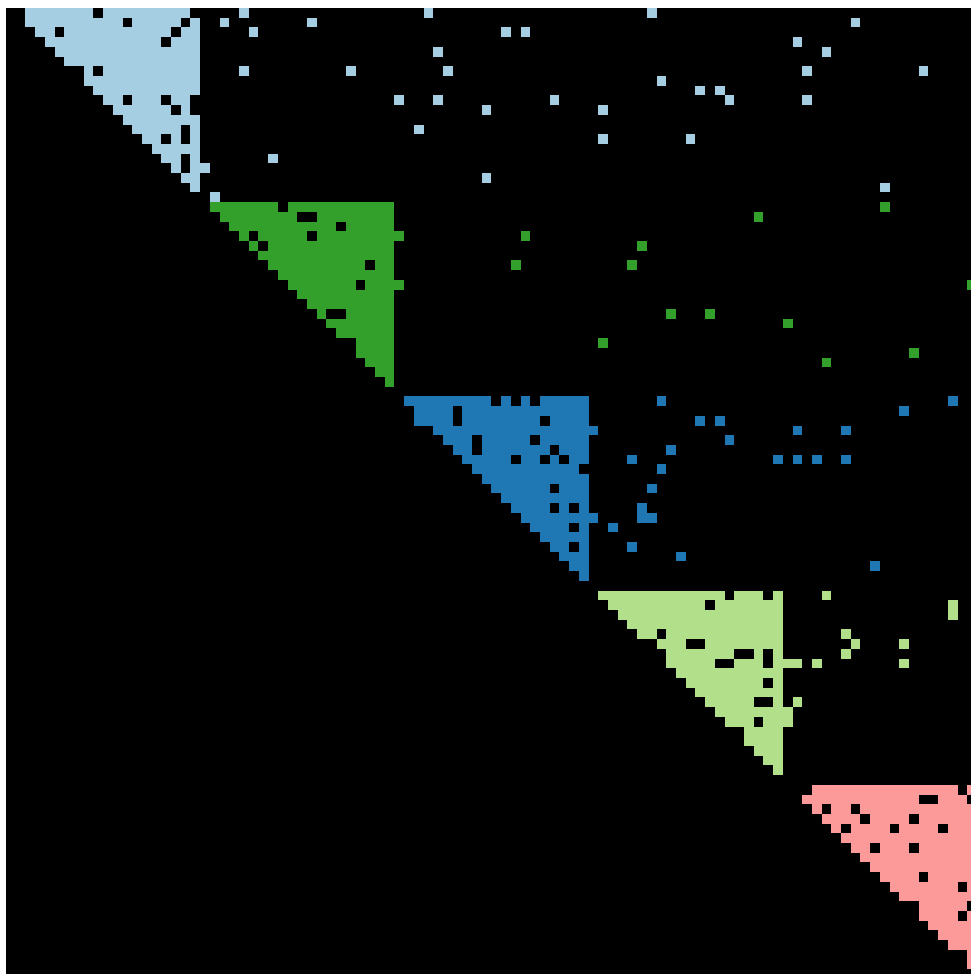
Perspectives

- Traitement des dettes à un niveau local
- Affinage de la création de communautés spécialisée pour nos graphes
- Réordonnancement des matrices par communautés
- Extraction et traitements à échelle communautaire

Homme des cavernes et réordonnancement



Homme des cavernes et réordonnancement



Autres types de visualisation

■ Dissuasion de HUB

■ Univers communautaire

