

Classification des coupes, premier regard sur la betweenness

Stage Casser des Graphes

Louis Milhaud

March 19, 2024

Complex Networks - LIP6

Outline

Classification des coupes

analyse des fréquences

Classification: critères

Classification: méthode

Betweenness vs fréquence

Classification des coupes

20 arêtes les plus coupées avec $\epsilon = 0.01$



20 arêtes les plus coupées avec $\epsilon = 0.03$



20 arêtes les plus coupées avec $\epsilon = 0.1$



critère d'intersection

Soit c_1 et c_2 deux coupes, $c_1 \sim_\epsilon c_2$ lorsque:

$$|c_1 \cap c_2| \geq \epsilon \cdot |c_1| \quad \epsilon \in [0, 1]$$

critère de voisinage

Soit c_1 et c_2 deux coupes, $c_1 \sim_k c_2$ lorsque:

$$\forall e_1 \in c_1 \exists e_2 \in c_2, d(e_1, e_2) \leq k \quad k \in \mathbb{N}$$

ce qui a été testé

- intersection fonctionne rapidement, problème de précision
- voisinage trop lent (python)
- critères mixtes toujours trop lent et peu d'améliorations

méthode du représentant

Une coupe arbitraire est la représentante d'une classe, le critère est appliqué par rapport à elle.

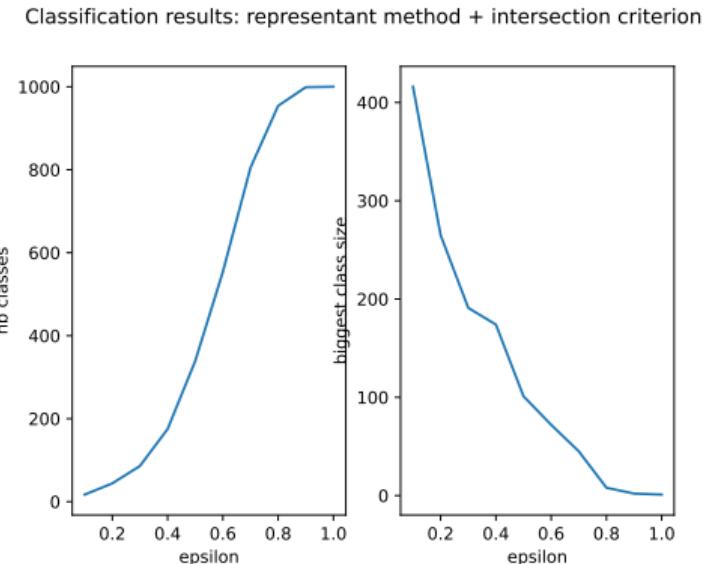
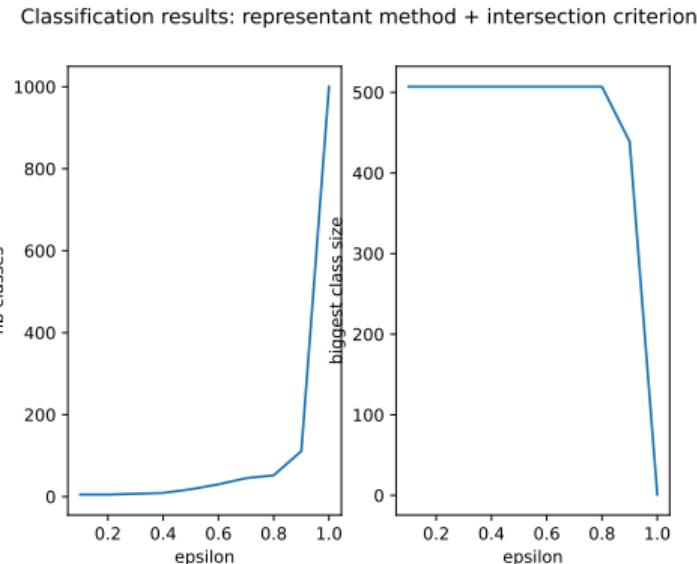
Avantages:

- classe réelle (faite à partir de caractéristiques observées)
- simple à appliquer

Inconvénient:

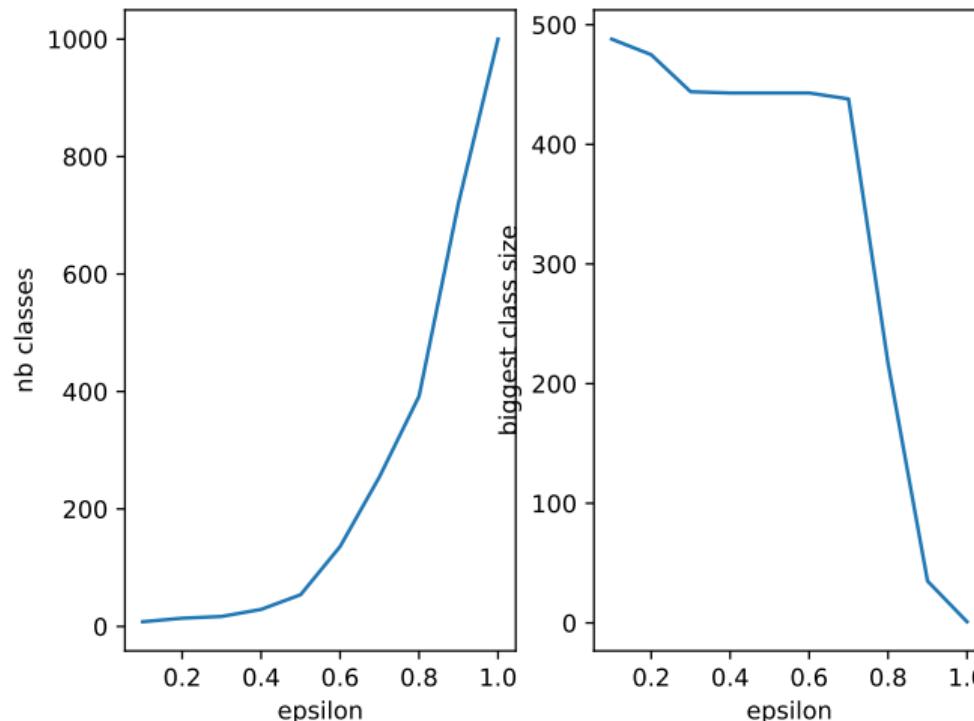
- on ne prend pas le meilleur représentant
- trop restrictif ?

méthode du représentant avec critère d'intersection, choix du ϵ



méthode du représentant avec critère d'intersection, choix du ϵ

Classification results: representant method + intersection criterion

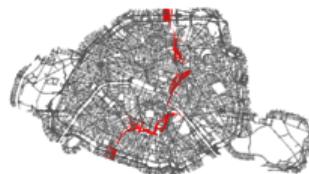


méthode du représentant avec critère d'intersection, visualisation

classe de taille 443



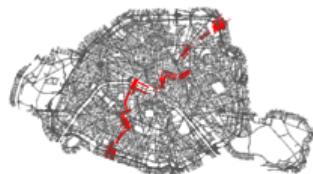
classe de taille 22



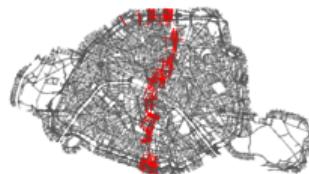
classe de taille 70



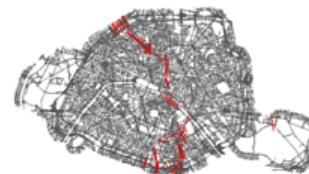
classe de taille 33



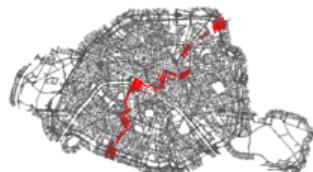
classe de taille 133



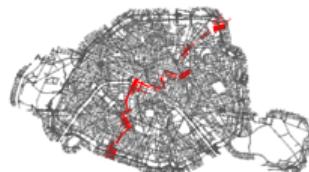
classe de taille 36



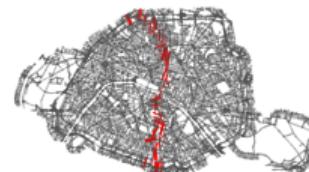
classe de taille 45



classe de taille 18



classe de taille 18



méthodes prédéfinies

On définit à l'avance les classes, on classe en fonction des critères prédéfinis.

Avantages:

- On paramétrise le niveau de contrainte et précision des classes

Inconvénients:

- On peut masquer / oublier une classe

Pas essayé car besoin de données géométriques (pour l'instant tout est fait sur le graphe KaHIP)

méthode du meilleur représentant

Méthode du représentant mais on itère récursivement dans la classe pour prendre le meilleur représentant.

Avantage:

- Prend un bon représentant
- n'oublie aucun éléments

Inconvénients:

- Unique classe ou aucune classe (inutile)

méthode de division itérative

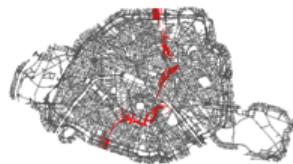
On choisit k représentants différents d'au moins t .

On classe les éléments restants par rapport à leur proximité avec le représentant

- + rapide $O(2kn)$ (linéaire)
- + toutes les coupes sont classées
- + les représentants sont forcément différents
- peut être pas les meilleurs représentants
- classé à la proximité (le meilleur n'est pas forcément bon)

Résultats avec proximité intersection, 4 classes

classe de taille 23



classe de taille 480



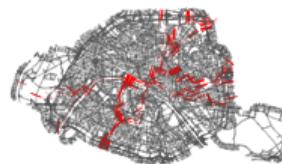
classe de taille 73



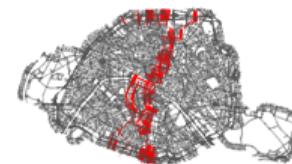
classe de taille 30



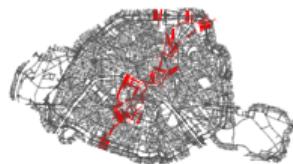
classe de taille 36



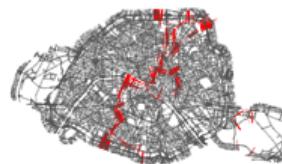
classe de taille 206



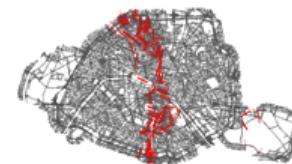
classe de taille 36



classe de taille 83



classe de taille 42



Betweenness vs fréquence

Corrélation des deux mesures

- $\epsilon = 0.01$: Correlation: 0.1
- $\epsilon = 0.03$: Correlation: 0.06
- $\epsilon = 0.1$: Correlation: 0.03