

# Détection des problèmes de différenciation géographique dans le cadre des données carroyées

Vianney Costemalle

Division des méthodes et référentiels géographiques

SMS, 24 juin 2019

# Au croisement de la géographie et de la confidentialité

**Un exemple introductif**

**Un problème de combinatoire**

**Application aux données carroyées de Filosofi 2015**

**Limites et autres**

# Plan

**Un exemple introductif**

Un problème de combinatoire

Application aux données carroyées de Filosofi 2015

Limites et autres

# Le secret statistique

## Secret primaire



# Le secret statistique

Secret primaire



Secret secondaire



# Le secret statistique

Secret primaire



Secret secondaire

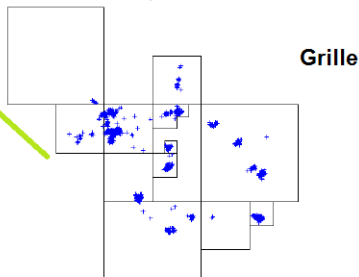
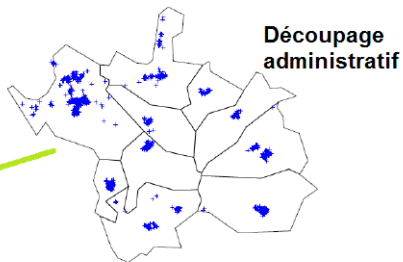


seuil de confidentialité

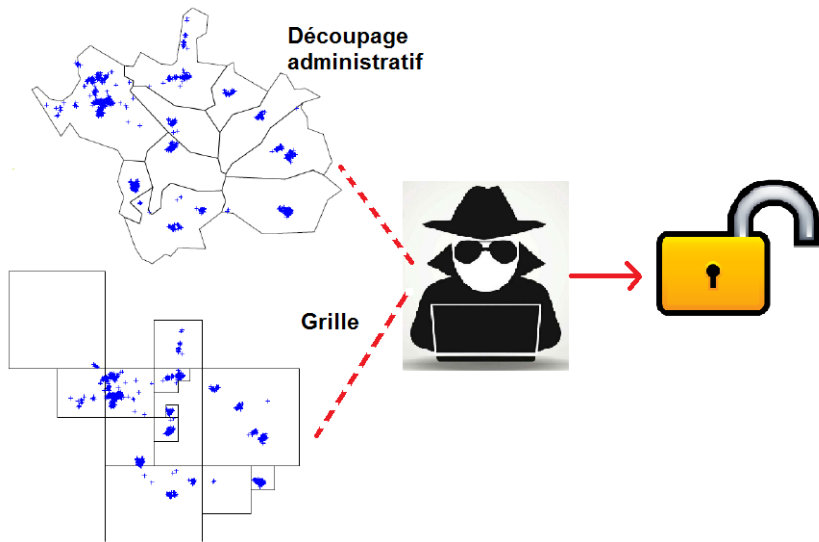


(11 dans les exemples)

## 2 nomenclatures de diffusion

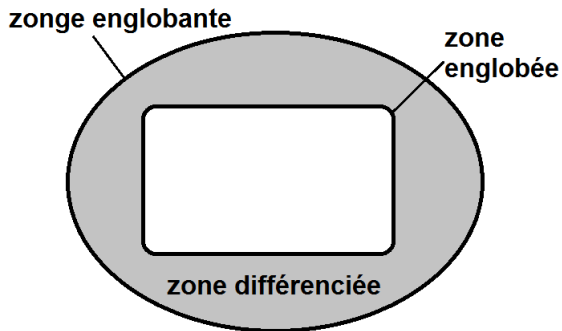


# Rupture du secret secondaire





# Principe de la différenciation géographique



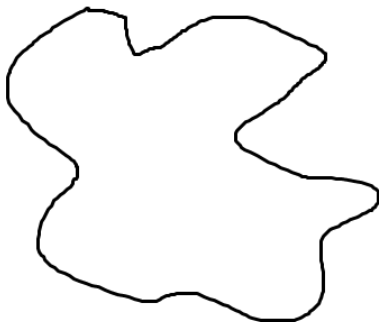
$$\overbrace{X_1 + X_2 + \dots + X_k + \dots + X_n}^{\text{zone englobante}}$$

$X_1 + X_2 + \dots + X_k$        $+ \dots + X_n$

zone englobée      zone différenciée

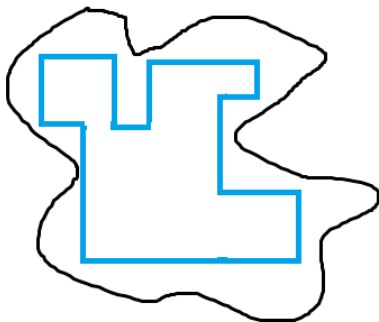
# Différenciation interne et externe

Un ensemble de communes ...



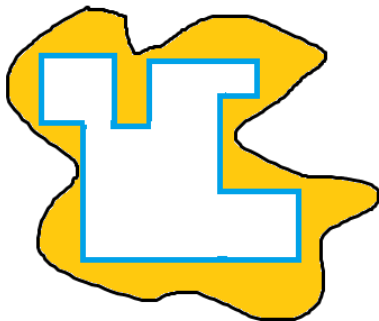
# Différenciation interne et externe

les carreaux internes ...



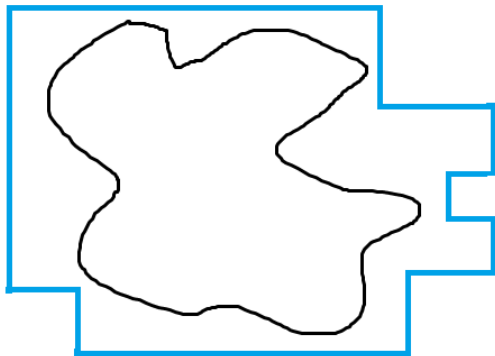
# Différenciation interne et externe

différenciation "interne"



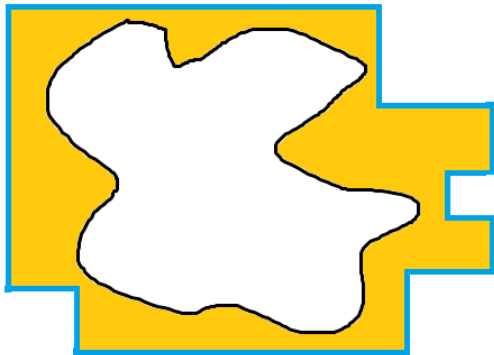
# Différenciation interne et externe

les carreaux internes + les carreaux à cheval ...

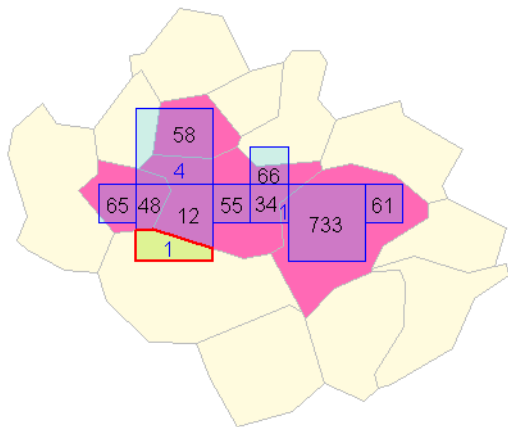


# Différenciation interne et externe

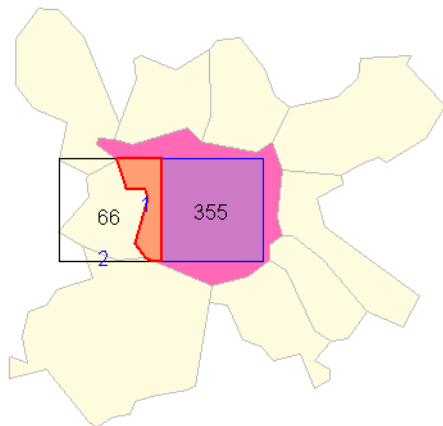
différenciation "externe"



## D'autres exemples réels ...

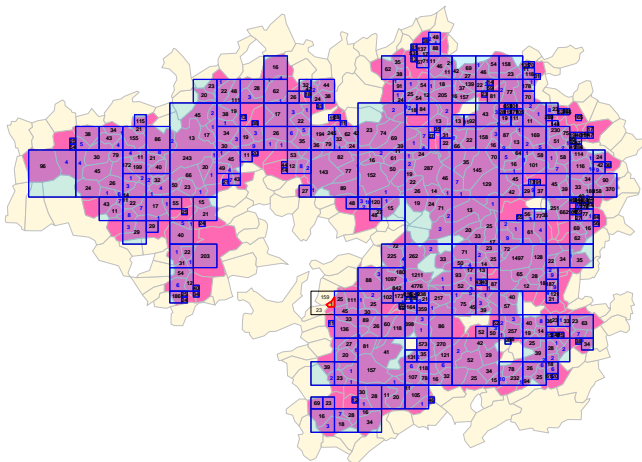


... simple ...





... ou complexe !



# Plan

Un exemple introductif

**Un problème de combinatoire**

Application aux données carroyées de Filosofi 2015

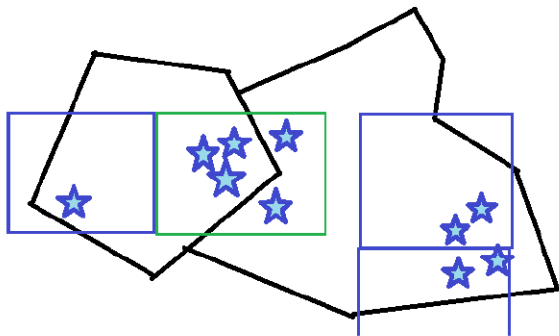
Limites et autres

# Questions

1. Y-a-t-il beaucoup de cas de différenciations menant à une rupture de la confidentialité ?
2. Comment détecter tous les cas problématiques possibles ?

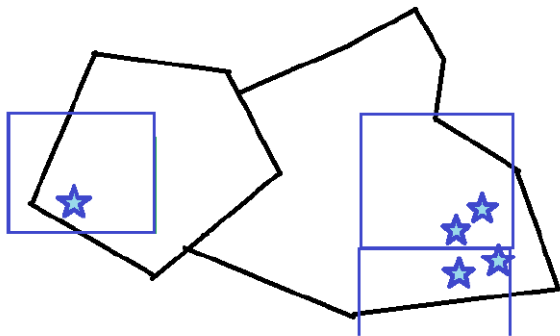
# Communes contiguës

Communes contiguës

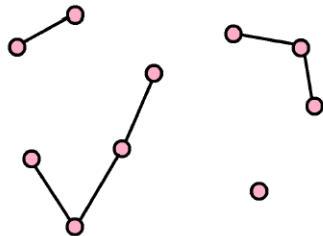
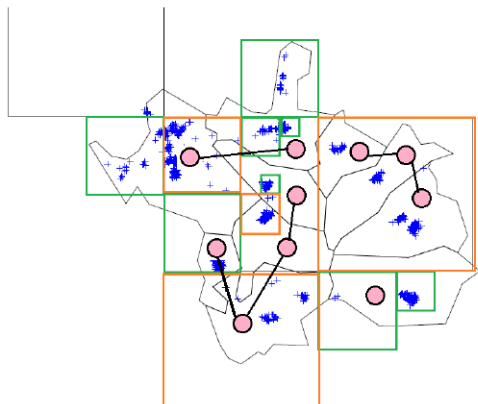


# Communes contiguës

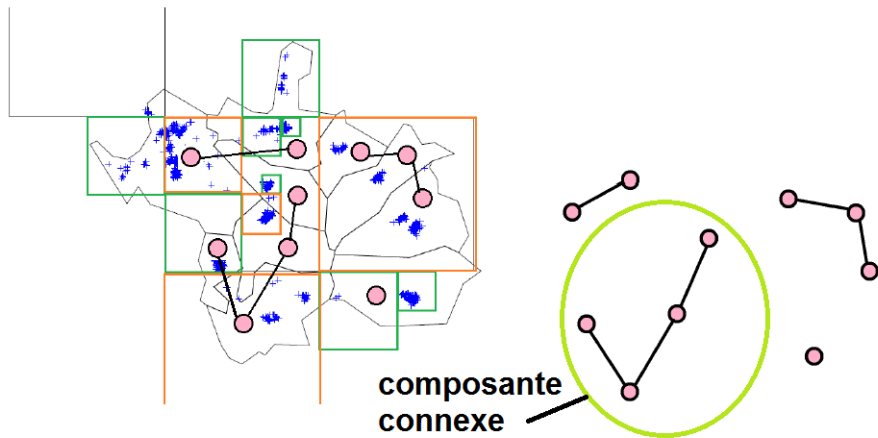
Communes non contiguës



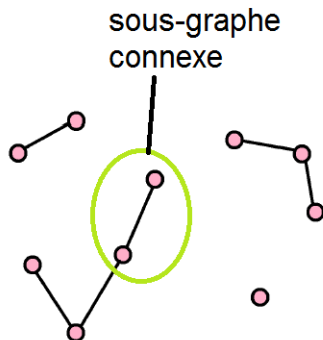
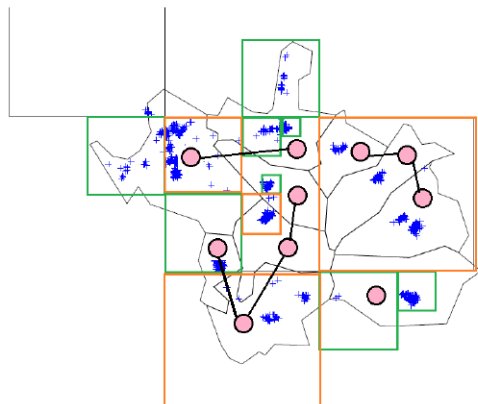
# Le graphe des communes contiguës



# Le graphe des communes contiguës

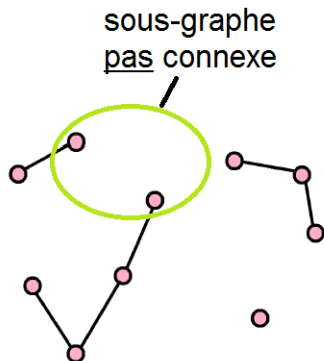
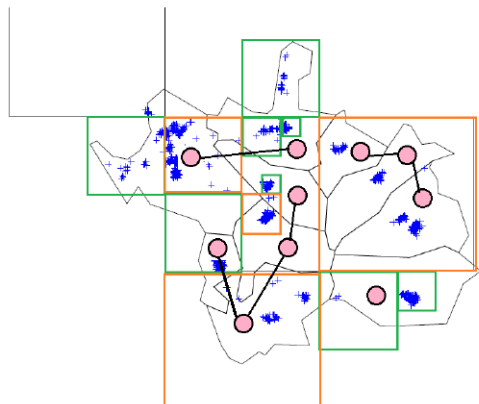


# Le graphe des communes contiguës





# Le graphe des communes contiguës

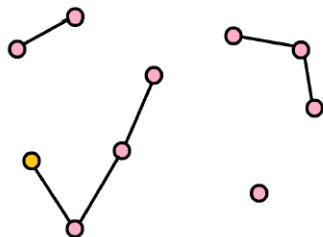
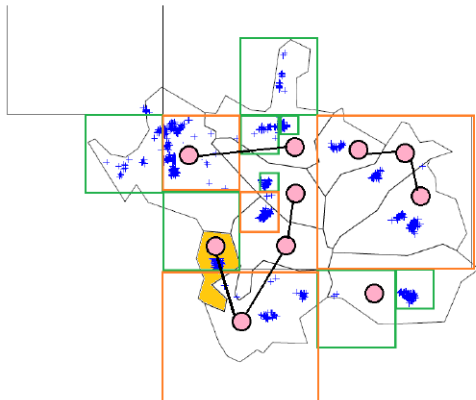


# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes

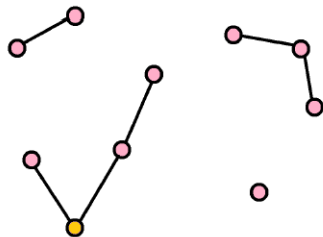
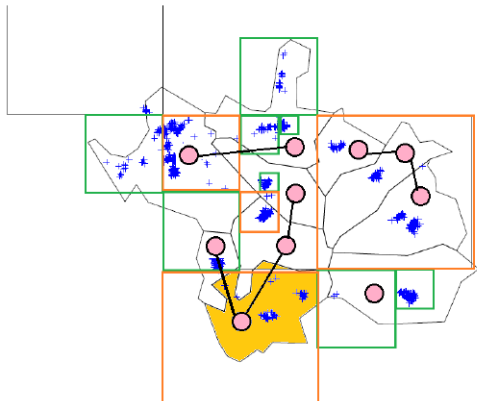
# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



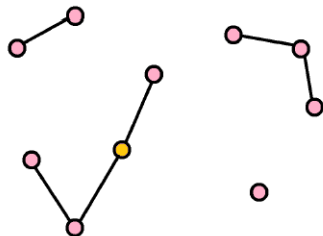
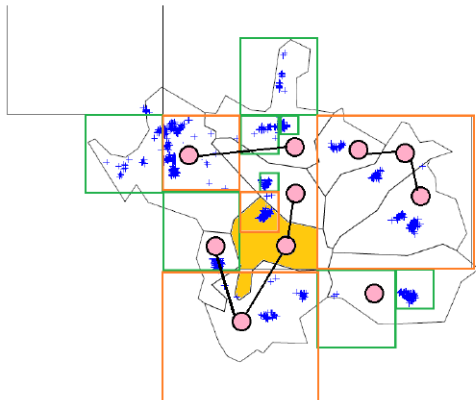
# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



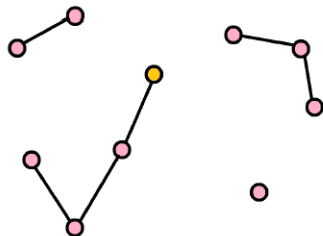
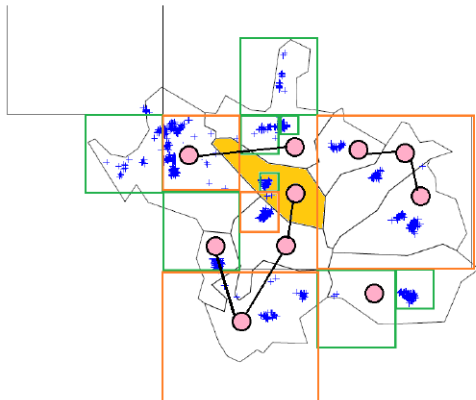
# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



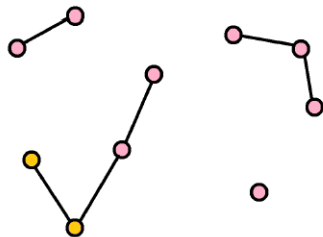
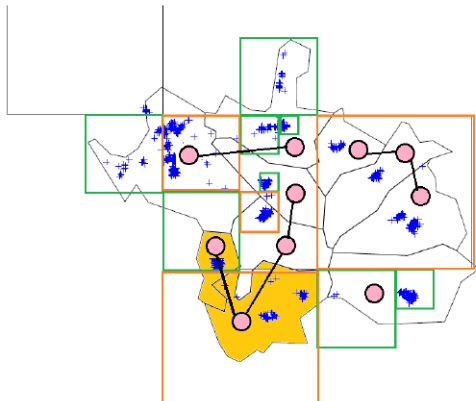
# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



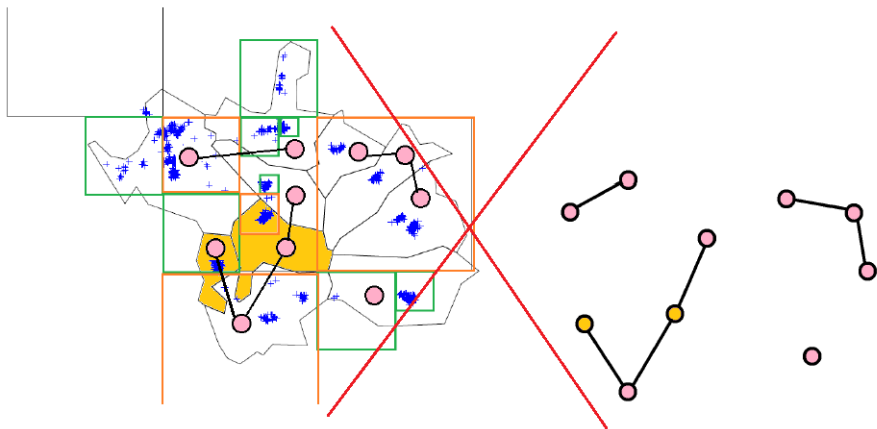
# Exploration du graphe

Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



# Exploration du graphe

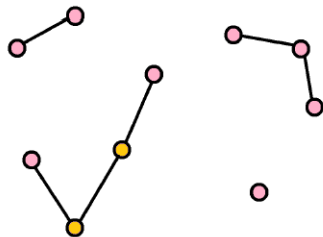
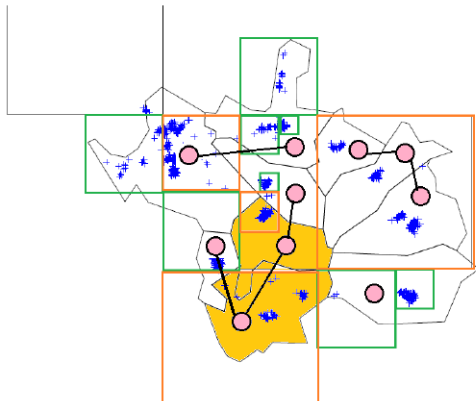
Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes





# Exploration du graphe

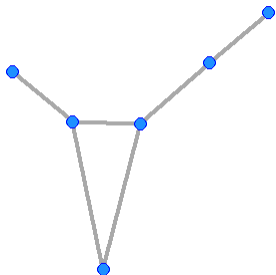
Calculer la différenciation interne et la différenciation externe sur  
l'ensemble des sous-graphes connexes



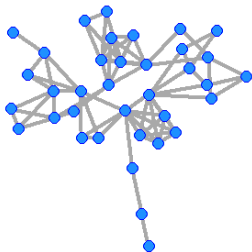
# Des composantes connexes plus ou moins grandes

327 composantes connexes

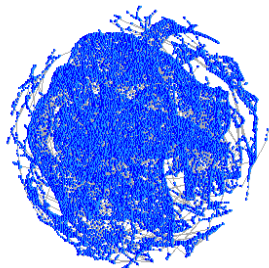
299 petites composantes  
connexes



27 moyennes  
composantes connexes



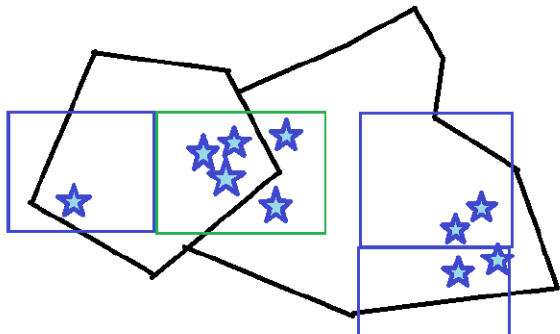
1 grande composante  
connexe



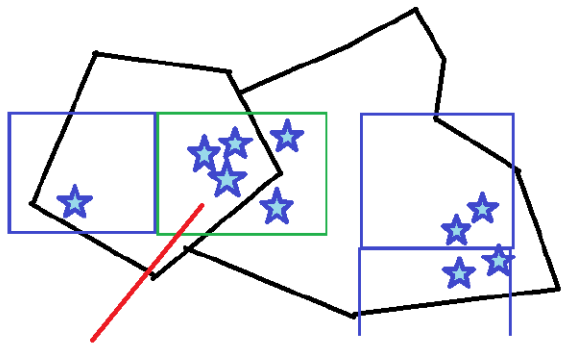
# Simplifier le graphe

1. Fusionner deux noeuds en un seul
2. Découper en sous-graphe

# Fusionner deux noeuds

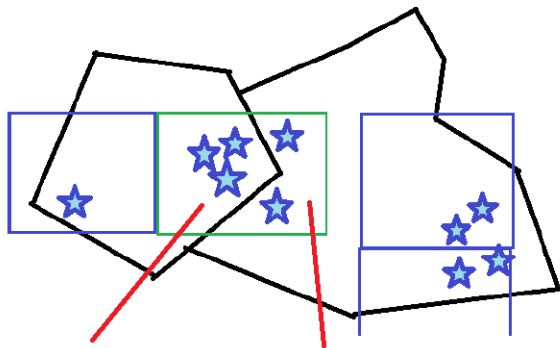


## Fusionner deux noeuds



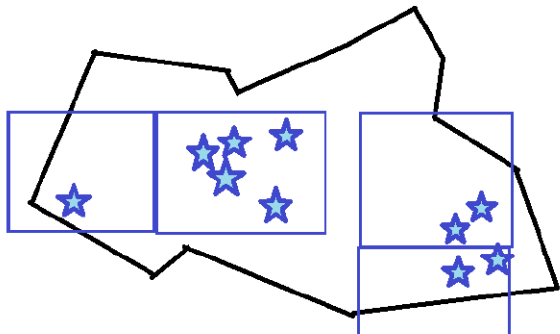
$\geq$  seuil ?

## Fusionner deux noeuds

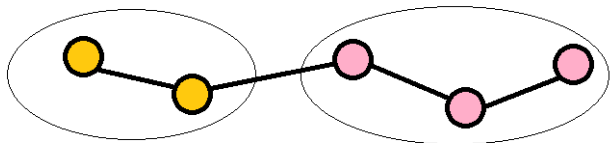


>= seuil ?

# Fusionner deux noeuds



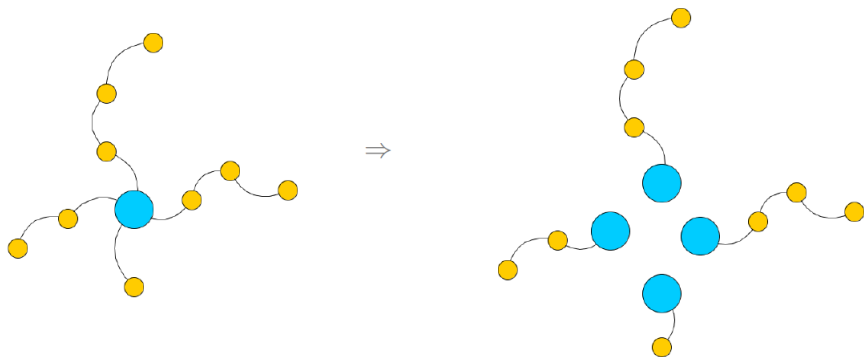
# Découper en sous-graphes



Différenciation interne = Différenciation externe



# Découper en sous-graphes



# Plan

Un exemple introductif

Un problème de combinatoire

**Application aux données carroyées de Filosofi 2015**

Limites et autres

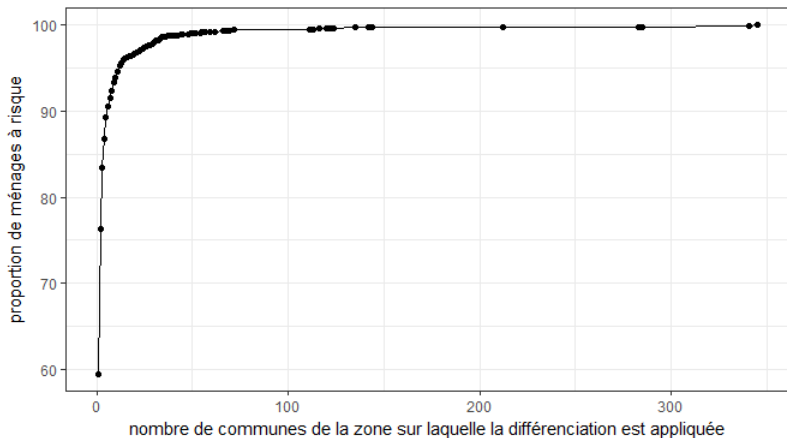
# Récapitulatif

- ▶ **Données** : Filosofi 2015  
35000 communes et 150000 carreaux  
135000 croisements carreaux-communes dont 52000 comportent moins de 11 ménages
- ▶ **Représentation sous forme de graphe** : 327 composantes connexes  
1 composante avec 33000 noeuds
- ▶ **Simplification du graphe**  
Fusion des noeuds  
Séparation du graphe
- ▶ **Recherche exhaustive** : sur les composantes connexes après simplification (programme d'Arlindo Dos Santos (DSAU) et François Sémécurbe (SSP Lab))

# Les ménages à risque

<b>nb. de ménages dans la zone différenciée</b>	<b>nb de différ - enciations internes</b>	<b>nb de différ - enciations externes</b>	<b>nb. de ménages à risque</b>
1	417	351	768
2	210	219	858
3	155	143	894
4	97	87	736
5	91	97	940
6	78	79	942
7	73	75	1036
8	73	65	1104
9	79	73	1368
10	91	65	1560
total	1364	1254	10206

# Taille des zones englobantes ou englobées



Localisation des ménages à risque

# Plan

Un exemple introductif

Un problème de combinatoire

Application aux données carroyées de Filosofi 2015

**Limites et autres**

# Limites et autres

- ▶ pas de généralisation à trois zonages ou plus se recoupant
- ▶ il faut connaître le nombre d'observations à chaque intersection carreau-commune pour appliquer la méthode
- ▶ dans le cas des données géographiques, il est possible pour un utilisateur extérieur d'avoir une estimation du nombre d'observations aux intersections
- ▶ méthode peut être appliquée à des données non géographiques : diffusion de données selon 2 nomenclatures différentes avec seuil de confidentialité