

La matrice de pondérations spatiales ou matrice d'interaction

- ▶ **Connectivité spatiale et voisinage**
- ▶ **Les pondérations spatiales**
- ▶ **Matrices de contiguïté**
- ▶ **Matrices de pondérations spatiales généralisées**
- ▶ **Caractéristiques de la matrice de pondérations spatiales**
- ▶ **Opérateur décalage spatial**

Connectivité spatiale

► Pourquoi des pondérations spatiales ?

- Covariances spatiales non nulles : $\text{cov}(y_i, y_j) \neq 0 \quad \forall i, j$
- Elles représentent la connectivité ou les interactions spatiales

► Structurer la connectivité spatiale

- Quels sont les unités spatiales i et j qui interagissent ?
- N observations pour estimer $N(N-1)/2$ covariances \Rightarrow impossible

► Hypothèses d'interaction

- i et j interagissent s'ils sont voisins
- interaction décroissante en fonction de la distance
- interaction croissante en fonction de la longueur de la frontière commune et décroissante en fonction de la distance

► Nécessité de définir le concept de voisinage pertinent et d'imposer une structure sur le schéma d'interactions spatiales

- Quels sont les voisins de chaque unité spatiale ou localisation ?
- Définir l'ensemble des voisins $N(i)$ pour chaque unité spatiale ou localisation i
- Outils fondamental : matrice des pondérations spatiales notée W

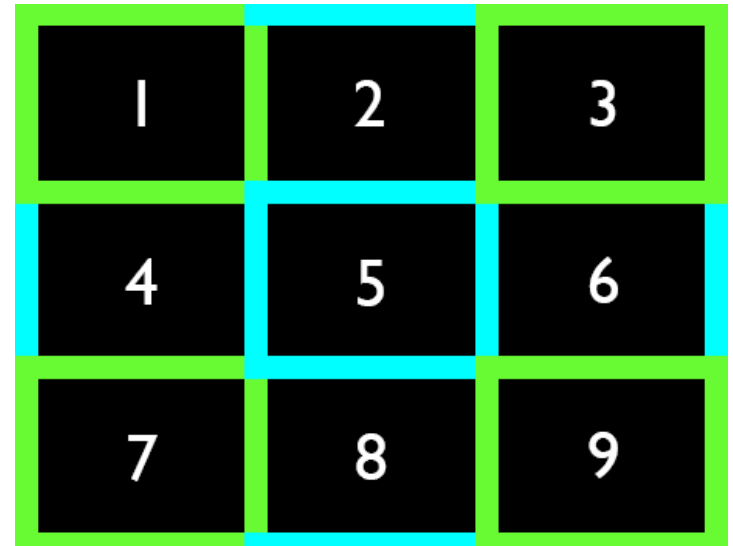
Voisinage et proximité

- ▶ **Contiguïté géographique ou cartographique – (SIG)**
 - Contiguïté = frontière commune, sommets communs
- ▶ **Distance physique (euclidienne / orthodromique)**
 - bandes de distance
 - plus proches voisins
 - fonction décroissante de la distance
- ▶ **Accessibilité**
 - distance temps
- ▶ **Distance socio-économique**
 - distance fondée sur des indicateurs socio-économiques
 - revenu, appartenance ethnique, structure industrielle, flux commerciaux, flux migratoires, etc.
 - problèmes d'endogénéité
 - Attention : distances nulles impossibles
 - ▶ $1/|z_i - z_j|$ n'est pas borné quand $z_i = z_j$

Contiguïté d'ordre 1

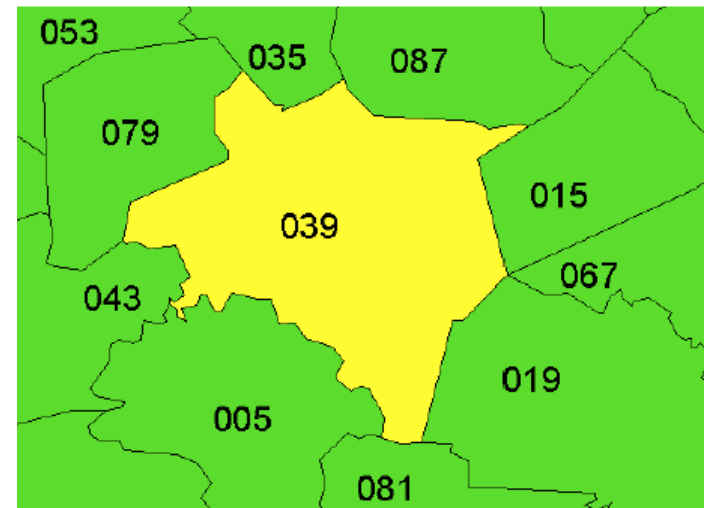
▶ Unités régulières

- tour
 - 2, 4, 6, 8
- fou
 - 1, 3, 7, 9
- reine
 - (2, 4, 6, 8) et (1, 3, 7, 9)



▶ Unités irrégulières

- Frontière commune - tour
- Sommets communs (039 et 067)
- reine



Distance en géométrie plane (euclidienne)

► Définition

On appelle distance sur un ensemble E , l'application $d : E \times E$ dans \mathbb{R}^+ vérifiant les propriétés suivantes :

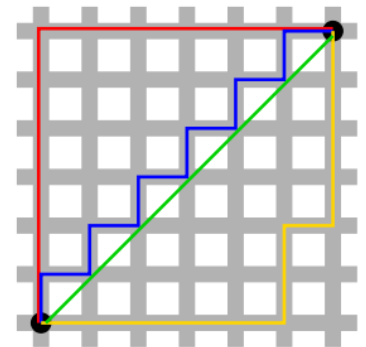
- $\forall x, y \in E, d(x, y) = d(y, x)$ symétrie
- $\forall x, y \in E, d(x, y) = 0 \Rightarrow x = y$ séparation
- $\forall x, y, z \in E, d(x, z) \leq d(x, y) + d(y, z)$ inégalité triangulaire

► **Distance de Manhattan** : $d(x, y) = \sum_{i=1}^N |x_i - y_i|$

► **Distance Euclidienne** : $d(x, y) = \sqrt{\sum_{i=1}^N |x_i - y_i|^2}$

► **Distance de Minkowski** : $d(x, y) = \sqrt[p]{\sum_{i=1}^N |x_i - y_i|^p}$

► **Distance de Tchebychev** : $d(x, y) = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^N |x_i - y_i|^p} = \max_i |x_i - y_i|$



Distance en géométrie sphérique



► Distance orthodromique ou géodésique ou distance du grand cercle

Distance à vol d'oiseau utilisée en navigation maritime et aéronautique

- Soit lat_i et $long_i$ la latitude et la longitude d'un point i et soit lat_j et $long_j$ la latitude et la longitude d'un point j , la distance sphérique entre i et j est donnée par :

$$d_{ij} = r \times \arccos \left[\sin lat_i \sin lat_j + \cos lat_i \cos lat_j \cos(long_i - long_j) \right]$$

où $r = 6371.01$ km est le rayon moyen de la terre

- Les latitudes et longitudes exprimées en degré, minute, seconde doivent être au préalable converties en degré décimal ($Sign \times (Deg + (Min + Sec / 60) / 60)$) puis en radian ($\times \pi / 180$)
- Autre formulation pour éviter les erreurs d'arrondis pour des distances faibles

$$d_{ij} = 2r \arcsin \sqrt{\sin^2 \left(\frac{lat_i - lat_j}{2} \right) + \cos lat_i \cos lat_j \sin^2 \left(\frac{long_i - long_j}{2} \right)}$$

Matrice de pondérations spatiales

► Définition

- Matrice carrée de dimension $(N \times N)$
de terme général w_{ij} non négatif, non stochastique, fini

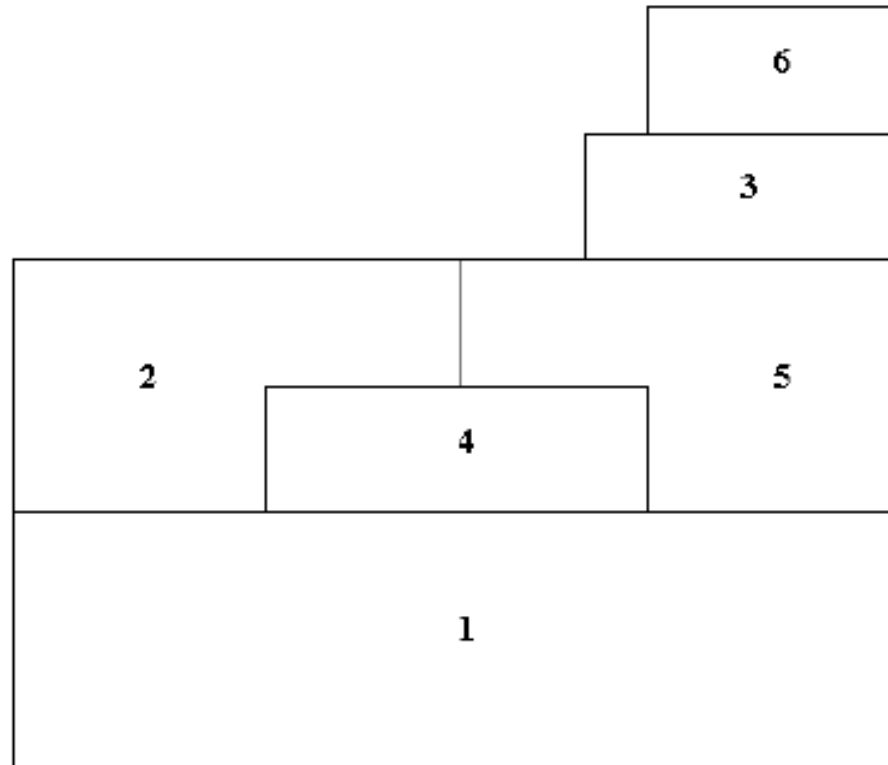
► Forme la plus simple : matrice binaire symétrique ou asymétrique

- $w_{ij} = 1$ si i et j sont « voisins »
par exemple $w_{ij} = 1$ si i et j sont contigus
si $d_{ij} < D$ distance critique (bandes de distance)
si i et j sont plus proches voisins
- $w_{ij} = 0$ sinon
- $w_{ij} = 0$ si $i = j$ par convention

► Standardisation en ligne

- moyenne pondérée des valeurs voisines
- $w^s_{ij} = w_{ij} / \sum_j w_{ij}$ tels que $\sum_j w^s_{ij} = 1$
- les coefficients d'autocorrélation spatiale sont alors comparables d'un échantillon à l'autre
- si la matrice était symétrique, elle ne l'est plus après standardisation

Exemple 1



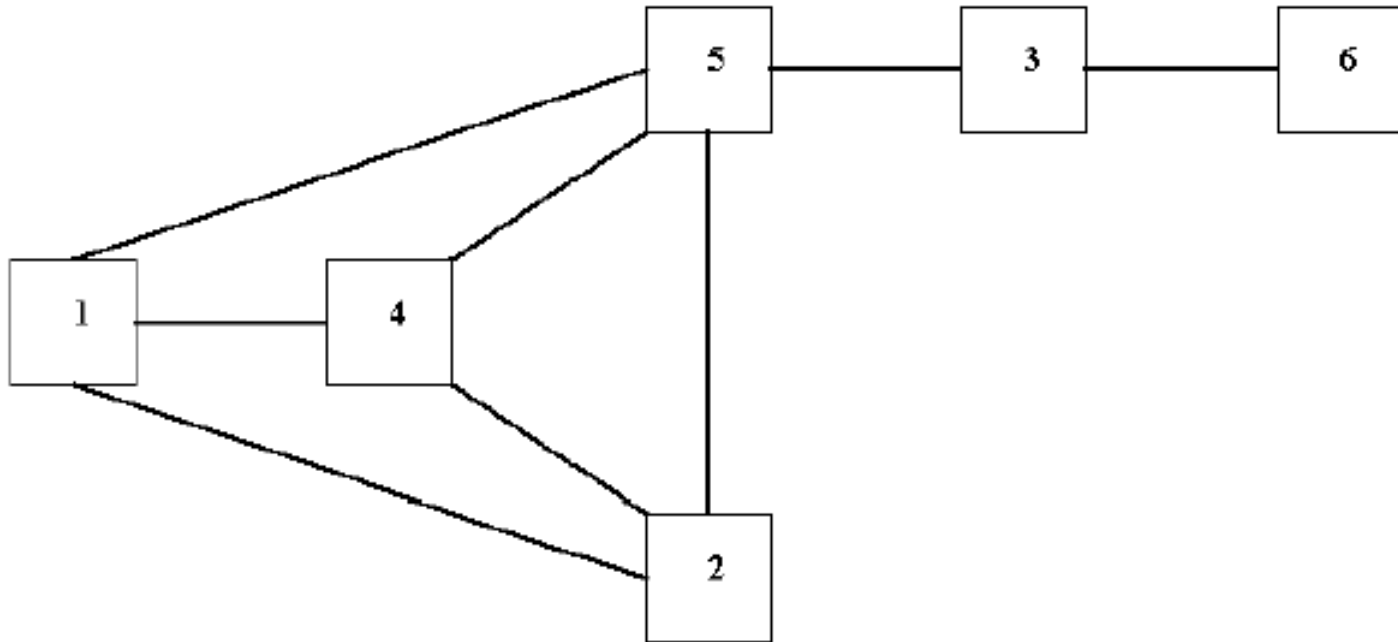
$N = 6$ unités spatiales
contiguïté représentée sur un plan ou une carte
contiguïté = frontière commune

Matrice de contiguïté du premier ordre

zones	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	1	1	0
3	0	0	0	0	1	1
4	1	1	0	0	1	0
5	1	1	1	1	0	0
6	0	0	1	0	0	0

$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Exemple 2



exemple pour $N = 6$
contiguïté représentée sur un graphe
contiguïté = nœuds connectés

Matrices de contiguïté d'ordre supérieur

▶ Définition récursive

- j est contigu à i à l'ordre p :
 - j est contigu à k à l'ordre 1
 - k est contigu à i à l'ordre $p-1$
 - i et j ne sont pas contigus à un ordre inférieur
- i et j sont contigus à l'ordre p si p est le nombre minimal de frontières à traverser pour aller de i à j

▶ Circularité et redondance

- puissance d'une matrice de pondérations spatiales
 - ▶ approche standard invalide
- Éliminer la circularité et la redondance
 - ▶ Représentation des pondérations spatiales par un réseau ou un graphe
 - ▶ Algorithme de Dijkstra modifié pour identifier le nombre de pas entre plus proches voisins (Anselin et Smirnov, 1996)
 - ▶ Nombre de pas = ordre de contiguïté

zones	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	1	1	0
3	0	0	0	0	1	1
4	1	1	0	0	1	0
5	1	1	1	1	0	0
6	0	0	1	0	0	0

Matrice de contiguïté du premier ordre

3	2	1	2	2	0
2	3	1	2	2	0
1	1	2	1	0	0
2	2	1	3	2	0
2	2	0	2	4	1
0	0	0	0	1	1

Matrice de contiguïté du premier ordre au carré

zones	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	1	0	0	0
3	1	1	0	1	0	0
4	0	0	1	0	0	0
5	0	0	0	0	0	1
6	0	0	0	0	1	0

Matrice de contiguïté du second ordre correcte

-1	1	2	1	1	3
1	-1	2	1	1	3
2	2	-1	2	1	1
1	1	2	-1	1	3
1	1	1	1	-1	2
3	3	1	3	2	-1

Matrice des ordres de contiguïté

Autres matrices de pondérations spatiales binaires

- ▶ Matrices de bandes de distance (symétrique)
 - $w_{ij} = 1$ si $d_{ij} < D$ seuil de distance ou distance critique
où d_{ij} est la distance entre i et j
 - $w_{ij} = 0$ sinon
 - $w_{ij} = 0$ si $i = j$ par convention

- ▶ Matrices des k plus proches voisins (asymétrique)
 - $w_{ij} = 1$ si i et j sont plus proche voisin
 - $w_{ij} = 0$ sinon
 - $w_{ij} = 0$ si $i = j$ par convention
 - chaque unité territoriale i a exactement k voisins quelle que soit la distance qui les séparent de i
 - implique une distorsion de l'espace
 - n'est pas nécessairement symétrique

Matrice de pondérations spatiales généralisées

- ▶ Matrices basées sur une fonction décroissante de la distance
 - ▶ Les pondérations sont données par $w_{ij} = f(d_{ij})$ où f est une fonction décroissante de la distance, par exemple :

$$f(d_{ij}) = 1 / d_{ij} \text{ ou } f(d_{ij}) = 1 / d_{ij}^2 \text{ ou } f(d_{ij}) = e^{-2d_{ij}}$$

- ▶ Les pondérations peuvent contenir des paramètres
 - ▶ exemple : $w_{ij} = (d_{ij})^{-\alpha}$
 - ▶ les paramètres sont estimés à partir des données ou fixés a priori
 - en pratique : $\alpha = 1$ (distance inverse) ou 2 (modèles de gravité)
- ▶ Pondérations asymétriques de Cliff-Ord
 - w_{ij} reflète les interactions spatiales potentielles entre i et j
 - $w_{ij} = [d_{ij}]^{-\alpha} \cdot [b_{ij}]^{\beta}$
où :
 d_{ij} est la distance entre i et j
 b_{ij} est la part de la longueur de la frontière commune entre i et j par rapport au périmètre de i

Matrice de pondérations spatiales généralisées

- ▶ Pondérations économiques et sociales (Case, Conley)
 - structure en blocs : effet pays
 - ▶ $w_{ij} = 1$ pour tout i et j appartenant au même bloc et 0 sinon
- ▶ Pondérations fonction décroissante d'une distance économique
 - $w_{ij} = 1 / (|r_i - r_j| + 1)$
 - ▶ avec par exemple r_i l'emploi total de l'unité territoriale i

Construire des matrices de pondérations spatiales dans GeoDa (Anselin, 2003)

Matrices de contiguïté

CREATING WEIGHTS

Input File (*.shp) |

Save output as |

Select an ID variable for the weights file <Rec_Num: 1,2,3,...,N>

Contiguity Weight

Queen Contiguity The order of contiguity 1

Rook Contiguity Include all the lower orders

Distance Weight

Select distance metric <Euclidean Distance>

Variable for x-coordinates <X-Centroids>

Variable for y-coordinates <Y-Centroids>

Threshold Distance 0.00

Cut-off point

Inverse Distance Standardize

Power 1

k-Nearest Neighbors # of neighbors 4

Save weights in GeoDa Legacy format

Create Reset Close

CREATING WEIGHTS

Input File (*.shp) C:\columbus\columbus.shp

Save output as C:\columbus\columbus.gal

Select an ID variable for the weights file POLYID

Contiguity Weight

Queen Contiguity The order of contiguity 1

Rook Contiguity Include all the lower orders

Create Reset Close

1. Menu Tools – Weights – Create
2. Entrer le nom du projet extension .shp
3. Entrer le nom du fichier pour la matrice de contiguïté, extension .gal
4. Entrer la variable codant les unités spatiales : polyd
5. Choisir le type de contiguïté : rook (frontières communes / queen frontières et sommets communs
6. Choisir l'ordre de contiguïté
7. Inclure ou non les contiguïtés d'ordre inférieur

Construire des matrices de pondérations spatiales dans GeoDa (Anselin, 2003)

```
0 49 columbus POLYID
1 2
5 4
2 2
8 5
3 5
15 11 9 6 8
4 7
12 7 15 11 9 6 1
5 9
25 16 12 7 26 22 10 1 2
6 8
26 22 10 15 11 8 3 4
7 7
19 16 18 12 11 5 4
8 8
16 15 9 6 3 2 14 13
9 10
28 27 25 16 20 17 11 8 3 4
10 11
35 33 27 32 40 23 26 22 15 6 5
11 11
25 24 18 14 13 15 9 6 3 4 7
12 8
25 24 15 19 18 5 4 7
```

Matrices de contiguïté

columbusCR.gal est un fichier ascii pouvant se lire avec un éditeur de texte

Structure du fichier

Taille de l'échantillon – nom du projet – nom de la variable de codage

Unité 1 : 2 unités contigües (5 et 4)

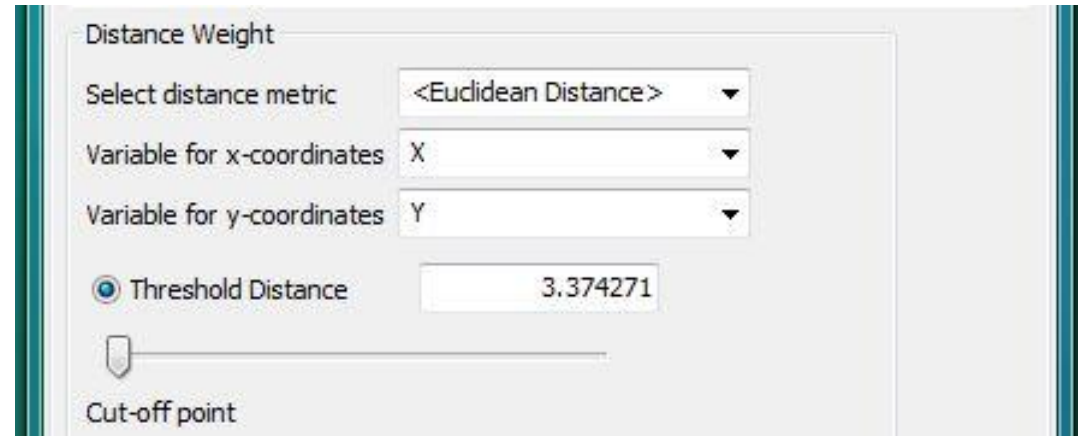
Unité 2 : 2 unités contigües (8 et 5)

Unité 3 : 3 unités contigües (15, 11, 9, 6, 8)

etc.

Construire des matrices de pondérations spatiales dans GeoDa (Anselin, 2003)

Matrices de bandes de distance



The screenshot shows the 'Distance Weight' dialog box in GeoDa. It includes the following settings:

- Select distance metric: <Euclidean Distance>
- Variable for x-coordinates: X
- Variable for y-coordinates: Y
- Threshold Distance: 3.374271
- Cut-off point: (indicated by a slider icon)

Distance critique minimal : distance telle que chaque unité ait au moins un voisin : ici 3.37427138

Matrices des k plus proches voisins



The screenshot shows the 'k-Nearest Neighbors' dialog box in GeoDa. It includes the following settings:

- k-Nearest Neighbors: # of neighbors 4

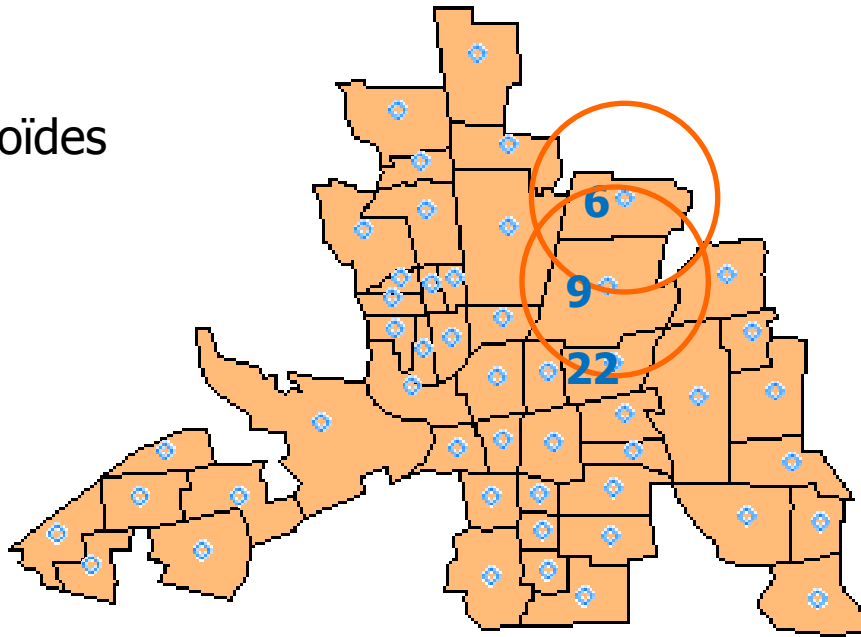
Construire des matrices de pondérations spatiales dans GeoDa (Anselin, 2003)

Matrices de bandes de distance

columbusBD.gwt est un fichier ascii pouvant se lire avec un éditeur de texte

D = 3.37427138 distance critique minimal assurant la connectivité de 6 à 9

◇ centroïdes



1 49	columbus POLYID
1 3	3.06471892
2 4	2.05766992
3 1	3.06471892
3 5	3.18567092
4 2	2.05766992
4 8	1.82214248
5 3	3.18567092
5 15	3.11578202
5 11	2.74874579
6 9	3.37427138
7 8	3.36340393
7 13	3.27844353
8 4	1.82214248
8 7	3.36340393
8 14	3.26377117
8 13	2.43180655
8 11	2.66602658
8 12	2.62383339
9 6	3.37427138
9 22	2.46099473
10 17	2.18689232

d_{ij}

Construire des matrices de pondérations spatiales dans GeoDa (Anselin, 2003)

Matrices des 4 plus proches voisins

columbusNN.gwt est un fichier ascii pouvant se lire avec un éditeur de texte

Chaque unité a exactement 4 voisins

-2 49 columbus POLYID

1 3 3.06471892

1 2 3.60117989

1 4 4.22995217

1 8 5.75306099

2 4 2.05766992

2 1 3.60117989

2 8 3.82844808

2 3 4.36806717

3 1 3.06471892

3 5 3.18567092

3 4 3.38496676

3 8 3.97152504

4 8 1.82214248

4 2 2.05766992

4 3 3.38496676

4 7 3.78307993

5 11 2.74874579

5 15 3.11578202

5 3 3.18567092

5 8 3.37551359

6 9 3.37427138

6 5 3.95297513

6 3 4.3651923

6 10 4.804083

d_{ij}

Caractéristiques de la matrice de pondérations spatiales

▶ Mesures de la connectivité spatiale globale

- pourcentage de pondérations spatiales non nulles dans W
⇒ matrices creuses (sparse matrix)
- pondération moyenne
- nombre moyen de liens
- valeurs propres min et max

▶ Mesures locales

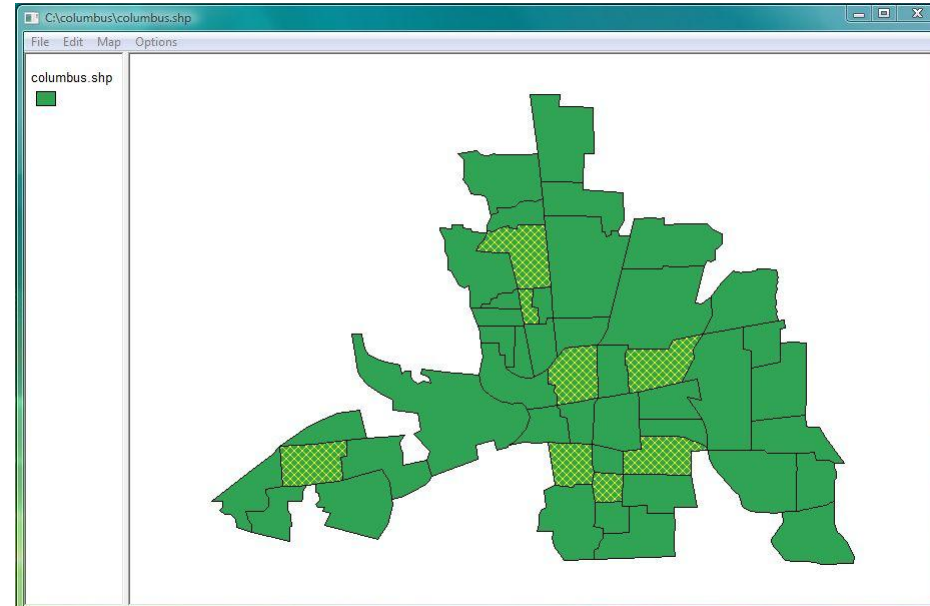
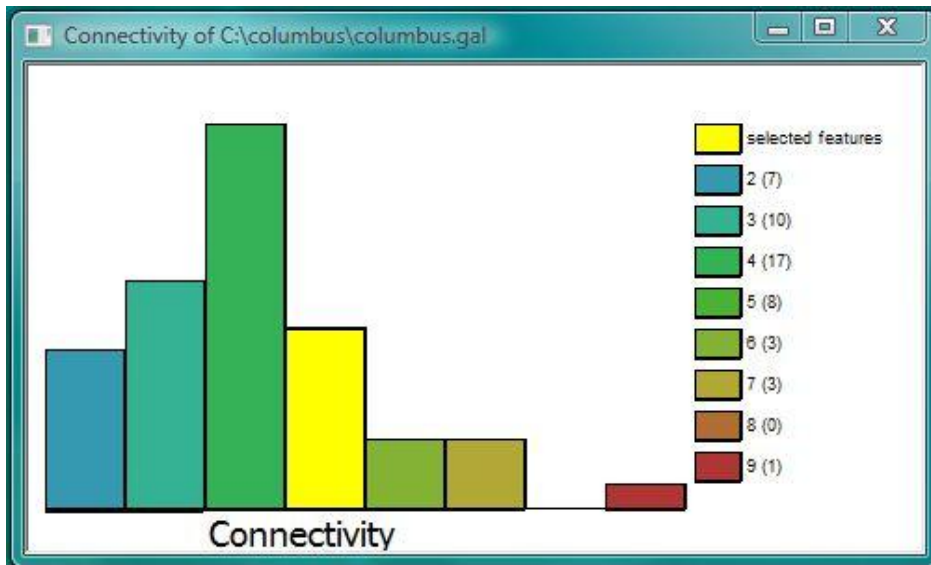
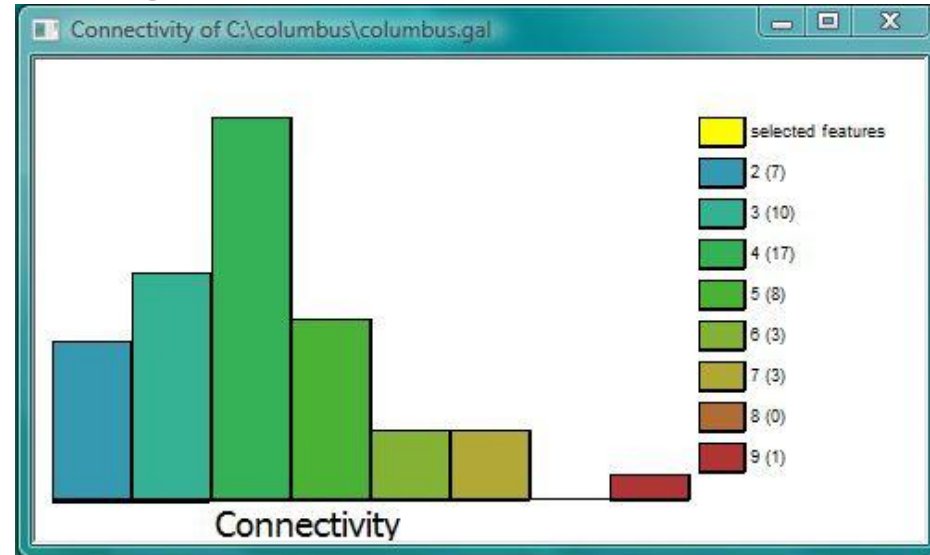
- observation la plus ou la moins connectée
- observations non connectées = îles (ligne et colonne nulles dans W)
- structure matricielle spécifique : bloc-diagonale

Caractéristiques des matrices de pondérations spatiales

Ville de Columbus (Ohio – USA)

Connectivité : Matrice de contiguïté - tour (frontière commune)

Classe sélectionnée :
8 unités spatiales ont chacune 5 unités spatiales contiguës



Matrice de pondérations spatiales standardisée en lignes

► Motivation

- moyenne pondérée des valeurs voisines
- les paramètres spatiaux sont comparables
- pour les matrices fondées sur la distance c'est la distance relative et non la distance absolue qui compte

- Soit \mathbf{W} la matrice de pondérations spatiales de terme général w_{ij} et soit w_{ij}^s le terme général de la matrice standardisée en ligne, alors on a :

$$w_{ij}^s = w_{ij} / \sum_j w_{ij}$$
$$\Rightarrow \sum_j w_{ij}^s = 1$$

Matrices de pondérations spatiales fondées sur la distance

- ▶ Forme générale de la matrice standardisée \mathbf{W} :

$$w_{ij}(k) = w_{ij}^*(k) / \sum_j w_{ij}^*(k) \quad \begin{cases} w_{ij}^*(k) = 0 \text{ si } i = j \\ w_{ij}^*(k) = f(d_{ij}) \text{ si } d_{ij} \leq D(k) \\ w_{ij}^*(k) = 0 \text{ si } d_{ij} > D(k) \end{cases} \quad k = 1, \dots, 4$$

- où f est une fonction décroissante en d_{ij} , par exemple :

$$f(d_{ij}) = 1/d_{ij} \text{ ou } f(d_{ij}) = 1/d_{ij}^2 \text{ ou } f(d_{ij}) = e^{-2d_{ij}}$$

- où d_{ij} est la distance sphérique entre les centroïdes des régions i et j
- paramètres de seuil :
D(1) = Q(1), D(2) = Me, D(3) = Q(3) et D(4) = Max (pas de seuil)

Matrices des k plus proches voisins

- ▶ Forme générale de la matrice standardisée \mathbf{W} :

$$w_{ij}(k) = w_{ij}^*(k) / \sum_j w_{ij}^*(k) \quad \begin{cases} w_{ij}^*(k) = 0 \text{ si } i = j \\ w_{ij}^*(k) = 1 \text{ si } d_{ij} \leq d_i(k) \\ w_{ij}^*(k) = 0 \text{ si } d_{ij} > d_i(k) \end{cases} \quad k = 10, 15, 20, 25$$

- où d_{ij} est la distance sphérique entre centroïdes des unités spatiales i and j
- $d_i(k)$ est la plus petite distance d'ordre k entre les centroïdes des unités spatiales i et j de manière à ce que chaque unité spatiale ait exactement k voisins

Opérateur décalage spatial (1)

► Décalage spatial

- n'est pas la contrepartie directe de l'opérateur décalage temporel, i.e. opérateur retard
- séries temporelles : $L_k y_t = y_{t-k}$
- séries spatiales : quels j sont-ils décalés de k de la localisation i ?
- sur des treillis réguliers : est, ouest, nord, sud
 - $(i-1, j), (i+1, j), (i, j-1), (i, j+1)$
- arbitraire pour des treillis irréguliers
 - les observations ont différents nombres de voisins

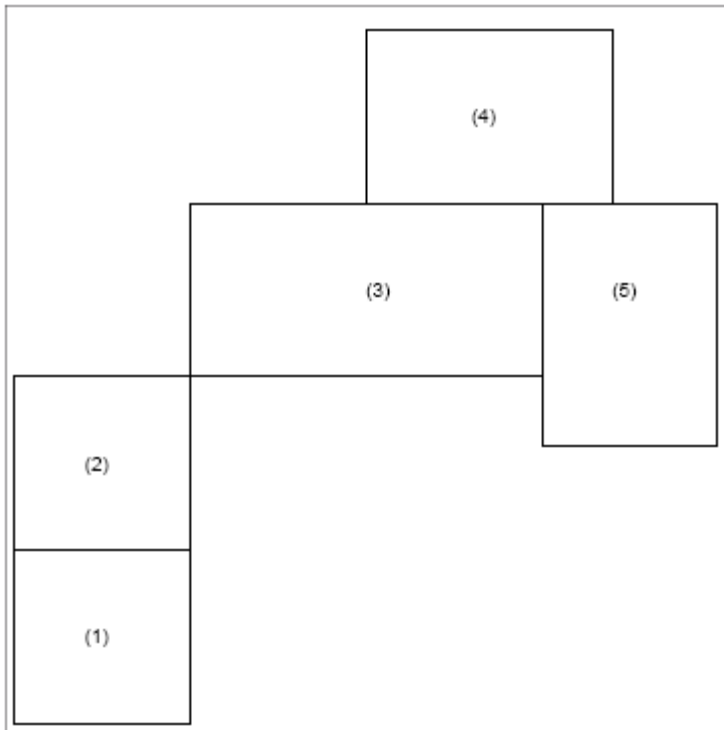
Opérateur décalage spatial (2)

► Décalage spatial

- Soit y un vecteur de dimension $N \times 1$ et W une matrice de poids de dimension $N \times N$, la **variable spatialement décalée** est le vecteur Wy de dimension $N \times 1$ et de terme général :

$$\sum_{j \neq i} w_{ij} y_j \text{ pour chaque } i = 1, \dots, N$$

- Si W est une matrice standardisée en ligne, le $i^{\text{ème}}$ élément du vecteur Wy représente la moyenne spatialement pondérée des valeurs prises par la variable y dans les unités spatiales j voisines de i
- le décalage spatial ne contient pas y_i



Matrice de contiguité - tour (frontière commune)

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

Matrice de contiguité standardisée

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

$$y = [y_1 \quad y_2 \quad y_3 \quad y_4 \quad y_5]'$$

$$y^* = [y_1^* \quad y_2^* \quad y_3^* \quad y_4^* \quad y_5^*]'$$

$$y^* = Cy$$

$$\begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \\ y_5^* \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

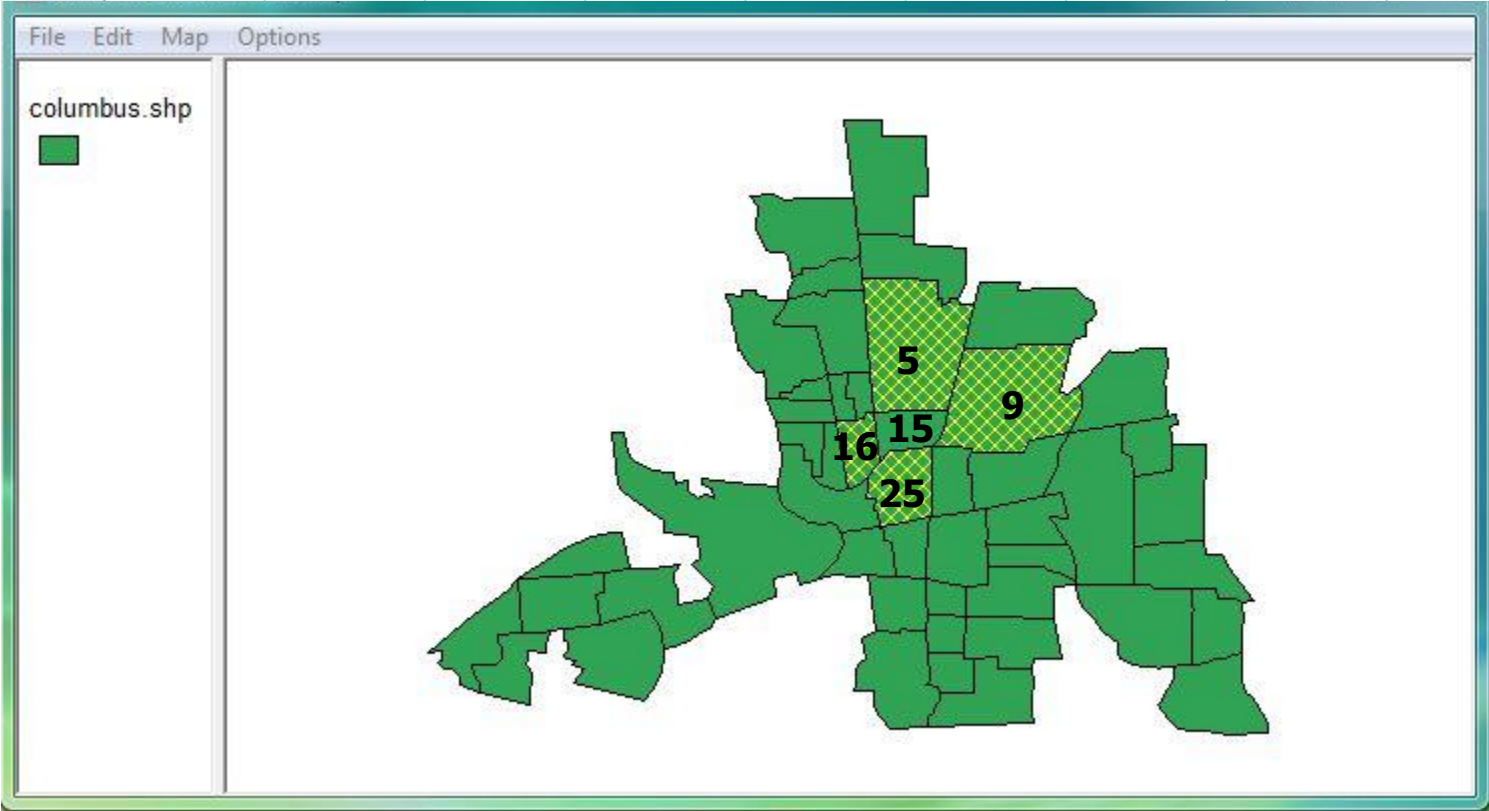
$$\begin{pmatrix} y_1^* \\ y_2^* \\ y_3^* \\ y_4^* \\ y_5^* \end{pmatrix} = \begin{pmatrix} y_2 \\ y_1 \\ 1/2y_4 + 1/2y_5 \\ 1/2y_3 + 1/2y_5 \\ 1/2y_3 + 1/2y_4 \end{pmatrix}$$

Interprétation du décalage spatial

- ▶ Association linéaire = Autocorrélation spatiale
 - comparaison de la valeur de y localisée en i à la moyenne pondérée des valeurs dans les localisations voisines
 - ⇒ y_i et $(Wy)_i$ similaires = autocorrélation spatiale positive (élevé-élevé, faible-faible)
 - ⇒ y_i et $(Wy)_i$ dissimilaires = autocorrélation spatiale négative (faible-élevé, élevé-faible)

Exemple : Criminalité dans la ville de Columbus

	AREA	PERIMETER	COLUMBUS_	COLUMBUS_I	POLYID	NEIG	HOVAL	INC	CRIME
16	0.093154	1.340060	17	36	16	36	18.800000	7.625000	54.838700
25	0.171680	1.666490	26	32	25	32	17.900000	8.461000	61.299200
5	0.488888	2.997130	6	7	5	7	23.225000	11.252000	50.731500
9	0.500755	3.169710	10	18	9	18	52.600000	17.586000	30.515900
15	0.106653	1.437610	16	9	15	9	18.000000	9.873000	48.585500
47	0.245249	2.079990	48	15	47	15	42.500000	18.950000	27.822900
14	0.060884	1.128420	15	40	14	40	42.900000	9.963000	57.066100



Variable spatialement décalée dans GeoDa

WCRIME(15) =

$1/4 * \text{CRIME}(16) + 1/4 * \text{CRIME}(25) + 1/4 * \text{CRIME}(5) + 1/4 * \text{CRIME}(9) =$

$1/4 * 54,8387 + 1/4 * 61,2992 + 1/4 * 50,7315 + 1/4 * 30,5159 = 49,3463$

Cambriolages et vols
de véhicule pour 1000
ménages en 1980

Variable
spatialement
décalée

POLYID	NEIG	HOVAL	INC	CRIME	WCRIME
16	36	18.800000	7.625000	54.838700	52.5989000
25	32	17.900000	8.461000	61.299200	48.6884000
15	9	18.000000	9.873000	48.585500	49.3463000
5	7	23.225000	11.252000	50.731500	38.4120000
9	18	52.600000	17.586000	30.515900	39.0099000
47	15	42.500000	18.950000	27.822900	17.5732000
46	48	76.100000	18.324000	16.530500	16.7033000