

Regionalização via Skater

Eduardo C. G. Camargo¹ Antonio M. V. Monteiro¹

¹Instituto Nacional de Pesquisas Espaciais
Divisão de Processamento de Imagens

2010

- **Regionalização:** pode ser visto como um procedimento de classificação aplicado a *geo-objetos* com representação poligonal.
- Exige contigüidade entre *geo-objetos* de uma mesma classe.
"geo-objetos membros de uma mesma classe devem formar uma região única, homogênea e espacialmente contígua".

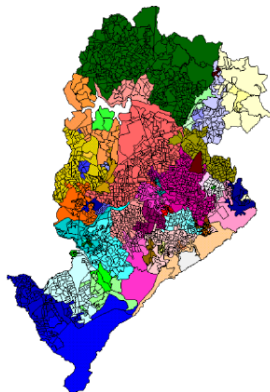
- **Regionalização:** pode ser visto como um procedimento de classificação aplicado a *geo-objetos* com representação poligonal.
- Exige contigüidade entre *geo-objetos* de uma mesma classe.

"geo-objetos membros de uma mesma classe devem formar uma região única, homogênea e espacialmente contígua".

- **Regionalização:** pode ser visto como um procedimento de classificação aplicado a *geo-objetos* com representação poligonal.
- Exige contigüidade entre *geo-objetos* de uma mesma classe.
"geo-objetos membros de uma mesma classe devem formar uma região única, homogênea e espacialmente contígua".

Introdução: Exemplo

Regionalização aplicada sobre os setores censitários de Belo Horizonte, considerando aspectos sócio-econômicos e condições dos domicílios.



Fonte: Neves, M. C. INPE (2003).

Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo:{nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo:{nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo: {nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo:{nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo:{nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

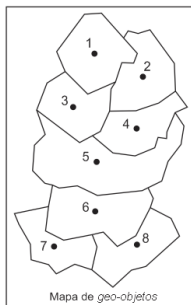
Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo: {nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

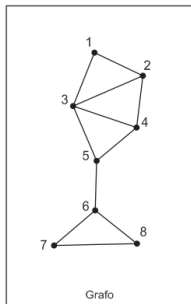
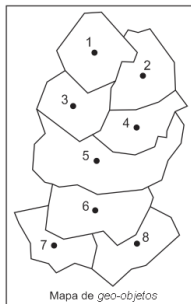
Skater: Características

- Uma ferramenta que executa a Regionalização;
- Considera a localização espacial dos *geo-objetos* (centróides);
- Se baseia na estrutura de vizinhança entre *geo-objetos* (grafo:{nós, arestas});
- Executa a regionalização via o método Árvore Geradora Mínima (AGM)
- Construção da AGM baseada em medidas de similaridade entre *geo-objetos*;
- Agrupa *geo-objetos* com características semelhantes;
- As características são estabelecidas pelo conjunto de variáveis de interesse.

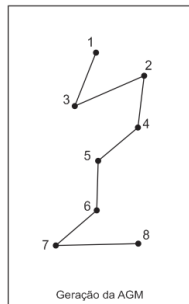
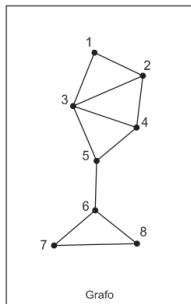
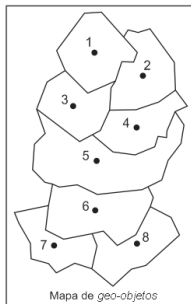
Passo 1: mapa de geo-objetos



Passo 2: construção do grafo

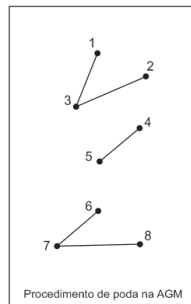
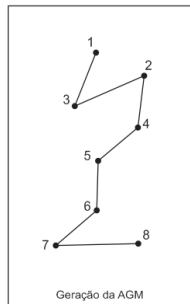
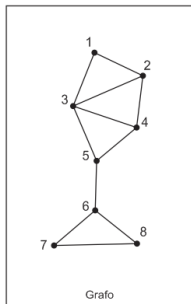
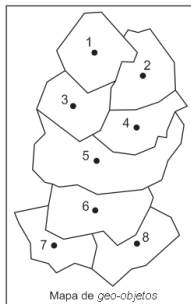


Passo 3: geração da AGM



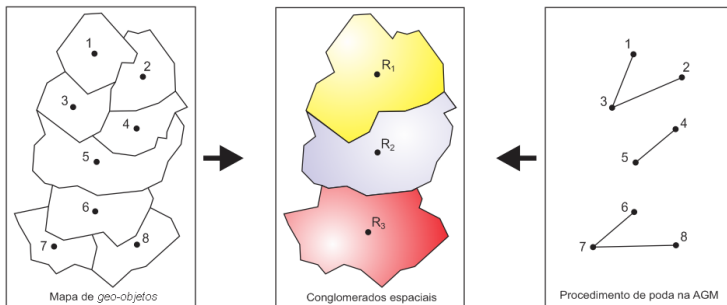
Skater: Visão geral

Passo 4: poda da AGM



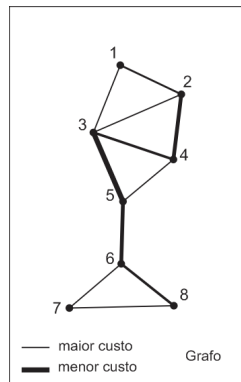
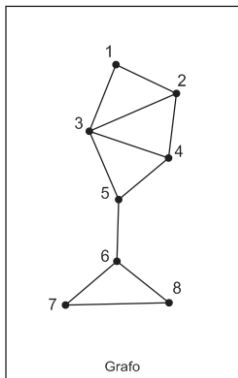
Skater: Visão geral

Passo 5: mapa de conglomerados espaciais



AGM: Construção da Árvore Geradora Mínima

Questão: Como são atribuídos os "custos" nas arestas (entre geo-objetos)?

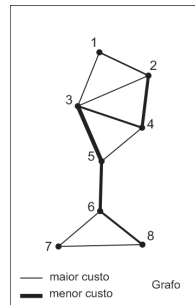


AGM: Cálculo dos custos

- Inicialmente os custos são calculados através de uma métrica que avalia a semelhança entre dois geo-objetos;
- Métrica -> coeficiente de similaridade, denotado por S ;
- Os coeficientes de similaridade entre todos geo-objetos podem ser condensados em uma matriz $S_{n \times n}$.

$$S = \begin{bmatrix} 0 & & & & & \\ S_{21} & 0 & & & & \\ S_{31} & S_{32} & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ S_{n1} & S_{n2} & \dots & S_{n,n-1} & 0 \end{bmatrix}$$

S : matriz diagonal de similaridade
 n : número de nós (geo-objetos)



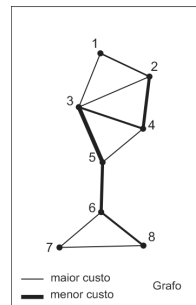
AGM: Cálculo dos custos

- Inicialmente os custos são calculados através de uma métrica que avalia a semelhança entre dois geo-objetos;
- Métrica -> coeficiente de similaridade, denotado por S ;
- Os coeficientes de similaridade entre todos geo-objetos podem ser condensados em uma matriz $S_{n \times n}$.

$$S = \begin{bmatrix} 0 & & & & & \\ S_{21} & 0 & & & & \\ S_{31} & S_{32} & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ S_{n1} & S_{n2} & \dots & S_{n,n-1} & 0 \end{bmatrix}$$

S : matriz diagonal de similaridade

n : número de nós (geo-objetos)



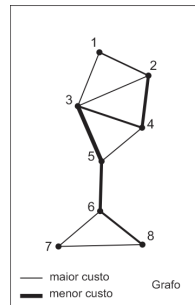
AGM: Cálculo dos custos

- Inicialmente os custos são calculados através de uma métrica que avalia a semelhança entre dois geo-objetos;
- Métrica -> coeficiente de similaridade, denotado por S ;
- Os coeficientes de similaridade entre todos geo-objetos podem ser condensados em uma matriz $S_{n \times n}$.

$$S = \begin{bmatrix} 0 & & & & & \\ S_{21} & 0 & & & & \\ S_{31} & S_{32} & 0 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ S_{n1} & S_{n2} & \dots & S_{n,n-1} & 0 \end{bmatrix}$$

S : matriz diagonal de similaridade

n : número de nós (*geo-objetos*)



AGM: Cálculo dos custos

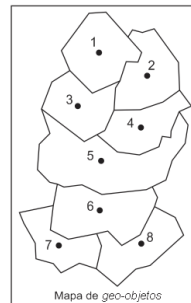
- Da mesma forma, os p atributos ou variáveis associados a cada um dos n geo-objetos também podem ser representados por uma matriz $\mathbf{X}_{n \times p}$.

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \dots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

$\mathbf{X}$: matriz de atributos

n : número de nós (*geo-objetos*)

p : número de variáveis



Métrica de Minkowski:

$$S_{ij}^{(\lambda)} = \left[\sum_{l=1}^p |x_{il} - x_{jl}|^\lambda \right]^{1/\lambda} \quad \lambda > 0$$

em que:

i e j : indexadores dos geo-objetos;

l : indexador da variável (atributo);

p : número de variáveis;

x_{il} e x_{jl} : valor da l -ésima variável associada ao i -ésimo e j -ésimo geo-objeto, respectivamente;

λ : é um parâmetro; maiores valores de $\lambda \Rightarrow$ enfatizar a variável com maior diferença entre x_{il} e x_{jl} .

Métrica de Minkowski:

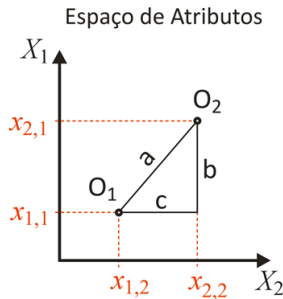
$$S_{ij}^{(\lambda)} = \left[\sum_{l=1}^p |x_{il} - x_{jl}|^\lambda \right]^{1/\lambda} \quad \lambda > 0$$

$$\text{Para } \lambda = 2 \Rightarrow S_{ij} = \sqrt{\sum_{l=1}^p |x_{il} - x_{jl}|^2}$$

Neste caso, o coeficiente de similaridade entre dois geo-objetos é obtido através da **distância euclidiana** calculada sobre o espaço de atributos.

AGM: Coeficiente de Similaridade

Exemplo do cálculo do coeficiente de similaridade entre dois geo-objetos, O_1 e O_2 , com duas variáveis associadas X_1 e X_2 .



$$S_{ij} = \sqrt{\sum_{l=1}^p |x_{il} - x_{jl}|^2}$$

$$S_{12} = \sqrt{(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2}$$

$$S_{12} = a = \sqrt{b^2 + c^2}$$

Lembrando que:

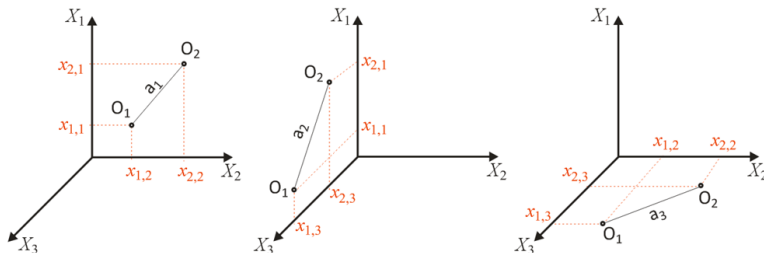
p : é o número de atributos (ou variáveis);

l : é o indexador dos atributos, varia de 1 a p ;

i e j : são indexadores dos geo-objetos.

AGM: Coeficiente de Similaridade

Outro exemplo: dois geo-objetos com 3 variáveis.

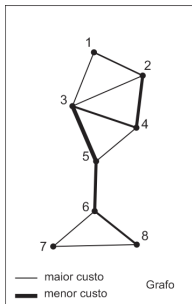


$$S_{ij} = \sqrt{\sum_{l=1}^p |x_{il} - x_{jl}|^2}$$

$$S_{12} = \sqrt{\underbrace{[(x_{11} - x_{21})^2 + (x_{12} - x_{22})^2]}_{a_1} + \underbrace{[(x_{11} - x_{21})^2 + (x_{13} - x_{23})^2]}_{a_2} + \underbrace{[(x_{12} - x_{22})^2 + (x_{13} - x_{23})^2]}_{a_3}}$$

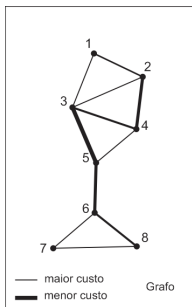
AGM: Resumo

- Inicialmente associa-se "custos" às arestas;
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".

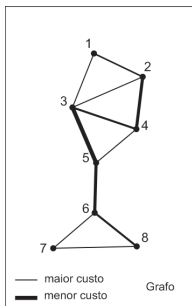


1

Geração da AGM

AGM: Resumo

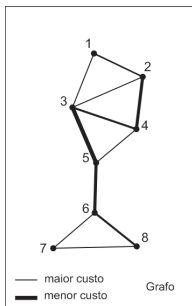
- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



Geração da AGM

AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".

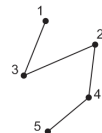
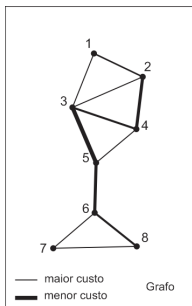






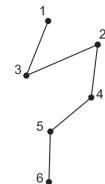
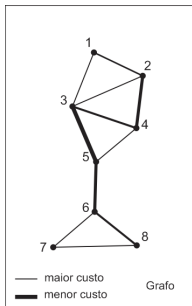
AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



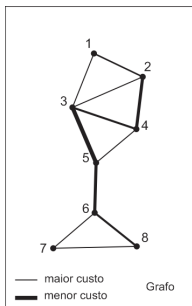
AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



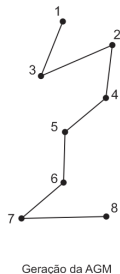
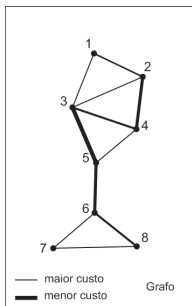
AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



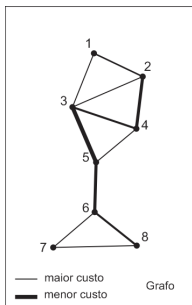
AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



AGM: Resumo

- Inicialmente as arestas recebem "custos";
- Os "custos" são calculados em função da similaridade entre os geo-objetos;
- Depois são eliminadas as arestas de menor "custo".



No final do procedimento tem-se:

n nós (no exemplo = 8) e

$n-1$ arestas (no exemplo = 7)

Poda da AGM: Idéia

- Nesta etapa do procedimento a forma de atribuir custos às arestas é modificada, de modo a obter melhores resultados:
 - regiões mais homogêneas, e
 - mais equilibradas em termos de números de geo-objetos por região.
- Depois remove-se as arestas de menores custos.

Poda da AGM: Idéia

- Nesta etapa do procedimento a forma de atribuir custos às arestas é modificada, de modo a obter melhores resultados:
 - regiões mais homogêneas, e
 - mais equilibradas em termos de números de geo-objetos por região.
- Depois remove-se as arestas de menores custos.

Poda da AGM: Idéia

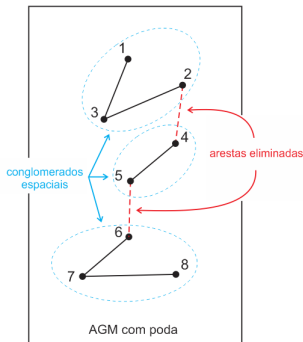
- Nesta etapa do procedimento a forma de atribuir custos às arestas é modificada, de modo a obter melhores resultados:
 - regiões mais homogêneas, e
 - mais equilibradas em termos de números de geo-objetos por região.
- Depois remove-se as arestas de menores custos.

Poda da AGM: Idéia

- Nesta etapa do procedimento a forma de atribuir custos às arestas é modificada, de modo a obter melhores resultados:
 - regiões mais homogêneas, e
 - mais equilibradas em termos de números de geo-objetos por região.
- Depois remove-se as arestas de menores custos.

Poda da AGM: Idéia

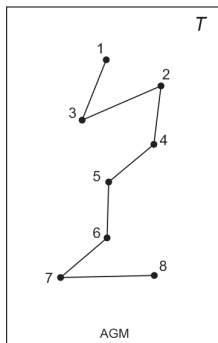
- Nesta etapa do procedimento a forma de atribuir custos às arestas é modificada, de modo a obter melhores resultados:
 - regiões mais homogêneas, e
 - mais equilibradas em termos de números de geo-objetos por região.
- Depois remove-se as arestas de menores custos.



Poda da AGM: Custos das arestas

- O "custo" de remover uma aresta l da árvore T é dada por:

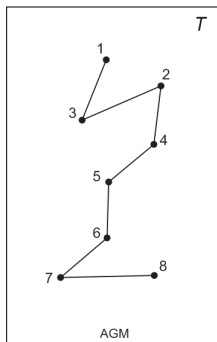
$$l = SQD_T - SQD_l$$



- 1) SQD_T é a soma dos quadrados dos desvios da árvore T ;
- 2) SQD_l refere-se a soma de duas parcelas obtidas dos quadrados dos desvios.

Ambos os termos são apresentados em detalhes a seguir.

$$\blacksquare I = SQD_T - SQD_I$$



1) SQD_T é dado por:

$$SQD_T = \sum_{j=1}^m \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$

em que:

m : é o número de variáveis consideradas na análise;

j : é o indexador das variáveis, varia de 1 a m ;

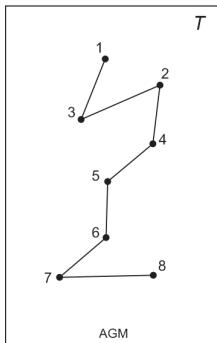
n : é o número total de nós (*geo-objetos*) em T ;

i : é o indexador de *geo-objetos*, varia de 1 a n ;

x_{ij} : representa a j -ésima variável associada ao i -ésimo *geo-objeto*;

$\bar{x}_j$: é o valor médio da j -ésima variável.

Poda da AGM: O termo SQDT

■ Ainda sobre o termo SQD_T 

$$SQD_T = \sum_{j=1}^m \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$



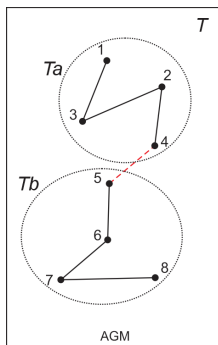
ID	Centróide X	Centróide Y	X1	X2	X3
1	607674.545	807362.205	0.6689	-0.2865	-0.3743
2	609474.035	799640.440	0.4351	-1.3384	-0.4970
3	609964.900	794709.440	-1.7143	0.2067	1.5887
4	609434.419	807784.450	0.7433	-4.6903	-0.3841
5	609249.130	808024.620	0.7433	0.0854	-0.0701
6	610067.635	806984.075	0.7340	1.1062	0.2489
7	611928.180	795947.495	-1.8663	-0.4164	1.0636
8	618578.180	795334.422	-1.5320	0.6789	0.0946

m=3

n=8

arquivo de dados

$$I = SQD_T - SQD_I$$



2) SQD_I é a soma das duas parcelas obtidas da soma dos quadrados dos desvios das duas sub-árvores, Ta e Tb , geradas pela retirada da aresta I da árvore T :

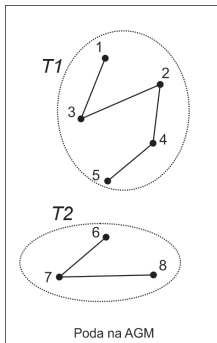
$$SQD_I = SQD_{Ta} - SQD_{Tb}$$

No cálculo de SQD_{Ta} e SQD_{Tb} , considera-se apenas os atributos associados aos geo-objetos pertencentes a cada sub-árvore SQD_{Ta} e SQD_{Tb} .

No exemplo da figura calcula-se SQD_I sete vezes, que é o número de arestas da árvore T .

Poda da AGM: Etapa final

$$I = SQD_T - SQD_I$$



Depois de calculado todos os custos das arestas (SQD_I), remove-se aquela de menor custo;

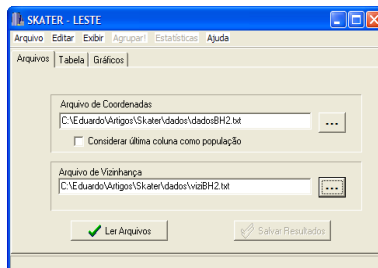
Ao remover a aresta de menor custo duas sub-árvores são criadas $T1$ e $T2$;

Repete-se o procedimento em $T1$ e $T2$, e assim sucessivamente até um critério de parada;

O critério de parada é definido pelo analista como, por exemplo, o número de classes desejadas.

Prática: Disponibilidade do Skater

- Disponível em: <http://www.est.ufmg.br/leste/skater.htm>
- Interface do Skater



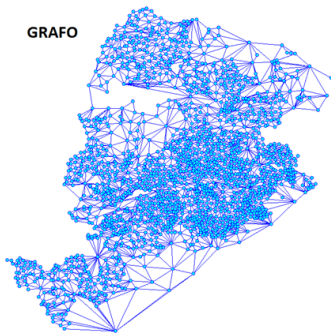
Prática: Dados de entrada

■ Exemplo

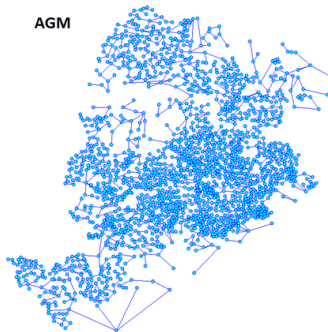


Prática: Visualização do Grafo e da AGM

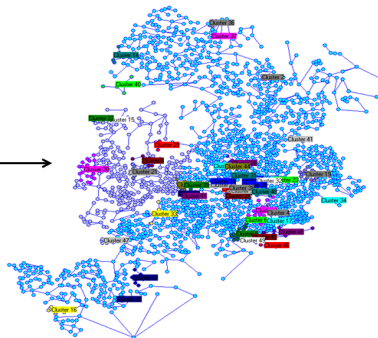
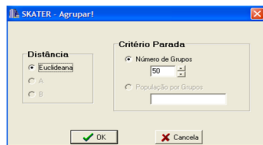
GRAFO



AGM



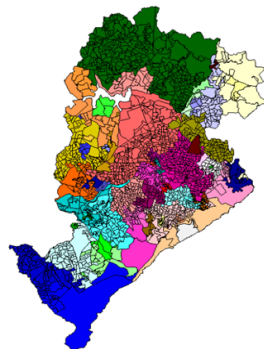
Prática: Geração e Visualização da AGM com poda



Prática: Resultado Final - Conglomerados Espaciais

Arquivo da AGM com poda

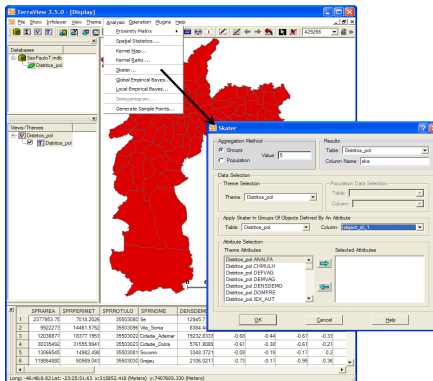
Classe	ID
8	1786
8	190
8	1326
9	1192
9	1186
9	1185
9	1588
9	1191
9	1193
9	1587
9	1590
10	1144
11	815
12	1111
12	1115
12	1118
13	1093
14	1101
14	821
14	1083
14	1103
14	813
14	812
15	1939
15	1938
15	1937
15	1932



Fonte: Neves, M. C. INPE (2003).

Prática: Outra possibilidade para executar o Skater

- Disponível em: <http://www.dpi.inpe.br/terraview>
- O programa TerraView



Referências Bibliográficas

- 1 Lage J.P, Assunção R.M., Reis E.A. (2001). A Minimal Spanning Tree Algorithm Applied to Spatial Cluster Analysis. In: Electronic Notes in Discrete Mathematics, Jayme Szwarcfiter and Siang Song (editores), Elsevier Science Publishers, Vol. 7, online publication.
- 2 Neves C.M., Câmara G., Assunção R.M. e Freitas C.C. (2002) Procedimentos Automáticos e Semi-automáticos de Regionalização por Árvore Geradora Mínima. In: Simpósio Brasileiro de Geoinformática, GeoInfo 2002 (4 : 2002 dez 5-6 : Caxambú – MG), pp. 109-116. Anais Editado por Davis Jr C. A. e Borges K.A.V. Belo Horizonte (MG) : SBC.
- 3 Neves, M. C. Procedimentos Eficientes para Regionalização de Unidades Socioeconômicas em Bancos de Dados Geográficos. Tese de Doutorado do Curso de Pós-Graduação em Sensoriamento Remoto, INPE, 2003.
- 4 Assunção R.M., Lage J.P. e Reis E.A. (2003). Análise de Conglomerados Espaciais via Árvore Geradora Mínima. Revista Brasileira de Estatística, v. 63, n. 220, p. 7-24, 2004.
- 5 Castro, M. S. M.; Silva, B. F. A.; Assunção, R. M.; Beato Filho, C. C. Regionalização como estratégia para a definição de políticas públicas de controle de homicídios. Cad. Saúde Pública. 2004, vol.20, n.5, pp. 1269-1280.
- 6 Laboratório de Estatística Espacial (LESTE) da UFMG, <http://www.est.ufmg.br/leste/skater.htm>