

Gerson José de Mattos Freire

Utilização de Técnicas de
Georreferenciamento e Classificação
Não-Supervisionada na Coleta Seletiva
de Recicláveis em Belo Horizonte

IX Curso de Especialização em Geoprocessamento
2006



UFMG
Instituto de Geociências
Departamento de Cartografia
Av. Antônio Carlos, 6627 – Pampulha
Belo Horizonte
cartografia@igc.ufmg.br

Freire, Gerson José de Mattos

Utilização de Técnicas de Georreferenciamento e Classificação
Não-Supervisionada na Coleta Seletiva de Recicláveis em Belo
Horizonte / Gerson José de Mattos Freire. – Belo Horizonte, 2006.
iv, 40f.: il.

Monografia (Especialização) – Universidade Federal de Minas
Gerais. Instituto de Geociências. Departamento de Cartografia, 2006.
Orientador: Clodoveu Augusto Davis Jr.

1. Georreferenciamento. 2. Classificação não-supervisionada. 3.
Coleta Seletiva. 4. Recicláveis. 5. *Clusterização*. I. Título.

Utilização de Técnicas de Georreferenciamento e Classificação Não-Supervisionada na Coleta Seletiva de Recicláveis em Belo Horizonte

Gerson José de Mattos Freire

Orientador: **Clodoveu Augusto Davis Júnior**

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
Departamento de Cartografia – IGC
ESPECIALIZAÇÃO EM GEOPROCESSAMENTO 2006**

BELO HORIZONTE – MG

DEZEMBRO/2006

Ao meu filho João Victor, que dia a dia me ensina que o amor incondicional pode
também ser infinito.

AGRADECIMENTOS

À ASMARE - Associação de Catadores de Papel, Papelão e Material Reaproveitável de Belo Horizonte, que através do empenho e dedicação de todos os seus membros tornou possível este trabalho, especialmente ao Engenheiro Ambiental Diogo Tunes;

À equipe técnica da ALUVIAL ENGENHARIA E MEIO AMBIENTE LTDA., pelo apoio em todos os momentos, especialmente aos Engenheiros América Maria Eleuthério Soares e Marcílio Felício Pereira;

Aos colegas monitores Ana Maria, Sheila, Renata, Guilherme, Vladimir, pela paciência e auxílio, especialmente ao Diego Rodrigues Macedo;

A todos os meus amigos, o meu agradecimento de coração.

SUMÁRIO

Capítulo 1 – Introdução	3
1.1 – Justificativa.....	3
1.2 – Objetivos.....	4
1.2.1 – Objetivo Geral	4
1.2.2 - Objetivos específicos	4
Capítulo 2 – Trabalhos anteriores relacionados ao tema.....	5
2.1 - Geocodificação de endereços.....	5
2.2 - Programas de Coleta seletiva	5
2.3 - Classificação não-supervisionada / clustering espacial.....	6
2.3.1 - Descoberta de Conhecimento em Bancos de Dados	6
2.3.2 - Etapas do Processo de Descoberta de Conhecimento em Bancos de Dados	6
2.3.3 - Mineração de Dados em SIG	8
2.3.4 - O algoritmo K-Médias	9
2.3.5 - <i>Software</i> utilizado.....	10
Capítulo 3 – Metodologia e Preparação de Bancos de Dados.....	11
3.1 - Metodologia	11
3.1.1 - Origem de dados	11
3.1.2 - Classificação dos doadores por tipo de atividade.....	13
3.1.3 - Levantamento de quantitativos de material coletado	13
3.1.4 - Base de dados de vias e endereços da Prodabel.....	14
Capítulo 4 – Georreferenciamento, visualização, classificação	15
4.1 - Pré-processamento	15
4.2 - Geocodificação do banco de dados de doadores.....	16
4.2.1 - Visualização e quantificação do banco de dados e sua distribuição espacial	17
4.2.2 - Mapa de tipos de doadores.....	19
4.2.3 - Mapa temático de quantidades doadas.....	21
4.3 - Implementação de classificação não-supervisionada sobre a base de dados de doadores.....	23
4.3.1 - <i>Clusterização</i> puramente geográfica.....	23
4.3.2 - <i>Clusterização</i> ponderada.....	26
4.3.3 - Pós-processamento	27
4.4 - Comparação de resultados.....	31
4.4.1 - Distância percorrida em cada <i>cluster</i>	31
4.4.2 - Quantidade de material coletado e média de doadores em cada <i>cluster</i>	34
Capítulo 5 - Conclusão	35
Referências Bibliográficas.....	38
Anexos.....	40

ÍNDICE DE TABELAS

Tabela 1 – Dados conforme são coletados pela ASMARE	12
Tabela 2 – <i>Clusterização</i> local – total coletado e número de doadores em cada <i>cluster</i>	24
Tabela 3 – <i>Clusterização</i> ponderada – total coletado e número de doadores em cada <i>cluster</i>	27
Tabela 4 – <i>Clusters</i> com elementos e quantidade de material coletado acima dos parâmetros	28
Tabela 5 – <i>Clusterização</i> ponderada – total coletado e número de doadores em cada <i>cluster</i> corrigidos	29
Tabela 6 – <i>Clusterização</i> local – Distância total percorrida em cada <i>cluster</i>	32
Tabela 7 – <i>Clusterização</i> Ponderada – Distância total percorrida em cada <i>cluster</i>	33

ÍNDICE DE FIGURAS

Figura 1 – Etapas do processo de descoberta de conhecimento em bancos de dados.....	7
Figura 2 – Distribuição geográfica de doadores	18
Figura 3 – Percentual de doadores por regional do município de Belo Horizonte.....	19
Figura 4 – Tipos de doadores	20
Figura 5 – Número de doadores por regional do município de Belo Horizonte	21
Figura 6 – volume de doações	22
Figura 7 – Volume de doações por regional em quilogramas	23
Figura 8 – <i>Clusterização</i> local.....	25
Figura 9 – <i>Clusterização</i> Ponderada.....	30
Figura 10 – Distância percorrida em cada <i>cluster</i>	31

Capítulo 1 – Introdução

A ASMARE – Associação de Catadores de Papel, Papelão e Material Reaproveitável de Belo Horizonte é uma sociedade civil que se destina a gerir a coleta de materiais recicláveis na Região Urbana de Belo Horizonte. É apoiada pela Prefeitura Municipal de Belo Horizonte (PBH), que provê seus locais de funcionamento, bem como por outros elementos da sociedade civil, como a Pastoral de Rua da Igreja Católica e diversas fundações e entidades internacionais. Para estas atividades, a ASMARE dispõe de galpão situado na Avenida do Contorno, no bairro do Barro Preto, destinado ao recebimento e triagem de recicláveis oriundos de seus doadores regulares e parceiros de diversas naturezas, estabelecidos em todo o município. Atualmente é responsável pela triagem e reciclagem de cerca de 240 t/mês de papel, papelão e plásticos diversos, destinados à comercialização por atacado, gerando trabalho, renda e inclusão social para seus atuais 689 associados.

A ASMARE possui um banco de dados e um programa de recebimento de chamadas para coleta, que registra nome e endereço do doador, além de datas e horários disponíveis para coleta. Além disso, o programa armazena dados de contato e a natureza da atividade do doador. A administração da ASMARE esforça-se em manter atualizada a programação de coleta, com uma frequência mínima semanal para a coleta de recicláveis de todos os doadores. Para o recebimento de doações são utilizados quatro caminhões, que recebem sua programação de coleta diariamente, acrescida das coletas que não puderam ser efetuadas no dia anterior.

1.1 – Justificativa

As deficiências do sistema implantado dão fruto a diversas situações de atrasos em coletas programadas, com acúmulo de serviços, perda de produtividade e geração de gastos desnecessários com combustível e outros insumos. Além disso, a insatisfação gerada pela não ocorrência de coleta em datas e horários programados impede a fidelização do doador e é prejudicial à implantação e manutenção de programas de coleta seletiva nas unidades doadoras.

1.2 – Objetivos

1.2.1 – Objetivo Geral

O objetivo geral deste trabalho envolve a espacialização do banco de dados de doadores, efetuando um procedimento de classificação espacial no qual os doadores possam ser agrupados em categorias que levem a melhoramentos na distribuição e cumprimento das rotas de coleta.

1.2.2 - Objetivos específicos

Os seguintes objetivos específicos foram identificados para este trabalho:

- Criar mapa de doadores
- Criar mapas temáticos de quantidades coletadas em cada ponto
- Sugerir plano e mecanismos de divisão em classes de doadores
- Sugerir metodologia para inclusão de novos doadores.

O restante deste trabalho está organizado como se segue. O Capítulo 2 procura sintetizar os trabalhos relacionados com o tema proposto e o desenvolvimento teórico aplicado. O Capítulo 3 identifica as origens dos dados trabalhados, a metodologia desenvolvida, a forma utilizada para a consolidação dos dados e classificação dos doadores, bem como a sistemática aplicada no acompanhamento dos roteiros e levantamento de quantitativos. O Capítulo 4 discute o georreferenciamento do banco de dados, apresenta sua visualização e as formas de classificação não-supervisionadas adotadas. Finalmente, o Capítulo 5 apresenta as conclusões encontradas e sugestões para a continuidade dos estudos afeitos ao tema.

Capítulo 2 – Trabalhos anteriores relacionados ao tema

2.1 - Geocodificação de endereços

Diversas aplicações já foram desenvolvidas em Sistemas de Informação Geográfica - SIG para a Geocodificação de endereços, tais como os de MOURA et al. [7] desenvolvidos para a cidade de Viçosa – MG. Suas potencialidades já foram exploradas em campos tão diversos quanto a segurança, a saúde pública e a educação. A busca de sistemas flexíveis de geocodificação e sua utilização em endereçamento já foi abordada, entre outros pesquisadores, por DAVIS [4]. Atualmente, as pesquisas avançam em sistemas de localização urbana que permitam a localização de pontos de interesse a partir de endereços ou de referências a lugares urbanos, através da materialização de conceitos de raciocínio espacial qualitativo, tais como o LOCUS, desenvolvido por SOUZA et al. [14].

2.2 - Programas de Coleta seletiva

A utilização de SIG para o aprimoramento de Programas de Coleta Seletiva ainda tem pequena utilização, com maiores aplicações nos campos de roteirização, como visto em MELO [6] e outros autores; ROVIRIEGO [12] fez uma interessante abordagem, envolvendo comparação entre diversos sistemas de coleta seletiva, sua roteirização e a análise dos custos envolvidos em cada um destes sistemas; já DAGNINO [3] abordou aspectos sociais da coleta seletiva, da valorização dos catadores à sua inserção no mercado de trabalho formal, buscando uma abordagem geográfica setorial.

2.3 - Classificação não-supervisionada / *clustering* espacial

2.3.1 - Descoberta de Conhecimento em Bancos de Dados

Conforme NEVES [8], a necessidade de recolher e armazenar dados de diversos tipos, formatos e origem, superou a capacidade humana de analisar, sintetizar e extrair conhecimento a partir desses dados. Os bancos de dados fornecem as ferramentas necessárias ao armazenamento e a utilização de grandes quantidades de dados; porém, a compreensão e a análise desses dados requerem a utilização de ferramentas apropriadas, que automatizem as atividades. Tais ferramentas usam recursos de áreas tais como a inteligência artificial, a aprendizagem automática, o reconhecimento de padrões e a estatística, e encontram-se em constante evolução. As aplicações integram teorias, métodos e algoritmos provenientes dessas áreas, tendo como objetivo a extração de conhecimento a partir de grandes bancos de dados.

Os algoritmos utilizados para procurar padrões nos dados são denominados algoritmos de *mineração de dados* (ou *Data Mining*, DM). O processo global de descoberta de conhecimento inclui a gestão dos algoritmos de DM e a interpretação dos padrões encontrados pelos mesmos.

2.3.2 - Etapas do Processo de Descoberta de Conhecimento em Bancos de Dados

O processo de Descoberta de Conhecimento em Bancos de Dados (ou *Knowledge Discovery in Databases*, KDD) tem quatro etapas sucessivas (conforme figura 1): limpeza e integração dos dados; seleção e transformação dos dados; mineração e avaliação; e apresentação do conhecimento [HAN, 9]. Essas etapas podem ser generalizadas em três grandes fases: pré-processamento, mineração e pós-processamento de dados.

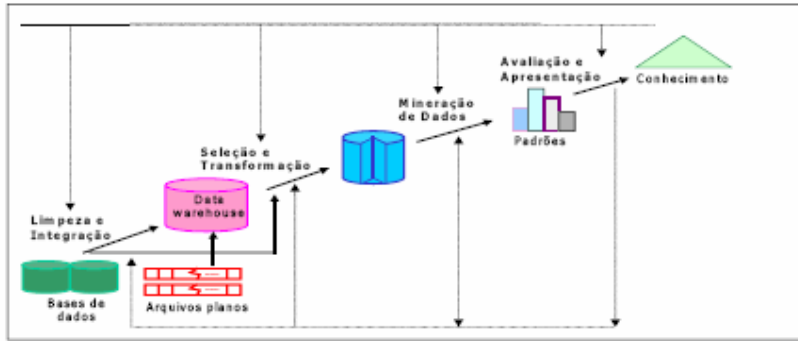


Figura 1 – Etapas do processo de descoberta de conhecimento em bancos de dados (Fonte: BOGORNY [1])

O *pré-processamento* é uma das fases mais demoradas do processo. Nessa fase são realizadas as seguintes tarefas:

- determinação dos objetivos da descoberta: define-se claramente o problema;
- limpeza dos dados: eliminação de ruído e inconsistência dos dados;
- integração dos dados: podem ser combinados dados de múltiplas fontes;
- seleção de dados: os dados relevantes para a mineração de dados são identificados e agrupados, gerando uma amostra do banco de dados;
- transformação de dados: conversão dos dados para um formato interpretável pelas ferramentas de mineração de dados.

A *mineração de dados* é a etapa em que se definem os padrões a serem usados para a extração do conhecimento e selecionam-se os algoritmos de mineração, que serão aplicados sobre os dados selecionados. Consiste na aplicação de algoritmos específicos, que tenham uma limitação aceitável de eficiência computacional e que sejam capazes de produzir uma enumeração particular de padrões. Esses algoritmos utilizam técnicas de aprendizado indutivo sobre bancos de dados e são capazes de extrair conhecimento através de exemplos, aplicando métodos iterativos por repetidas vezes.

O *pós-processamento* é realizado através das etapas de avaliação e apresentação dos padrões, que são responsáveis pela identificação e análise dos padrões interessantes que representam conhecimento, bem como, a forma como o conhecimento extraído será apresentado ao usuário.

2.3.3 - Mineração de Dados em SIG

Para o uso em Sistemas de Informação Geográficos, as principais técnicas de mineração de dados podem ser classificadas em: generalização, associação espacial, aproximação e agregação, classificação e *clusterização*.

Na definição de HAN [9], o *agrupamento* ou *clusterização* consiste em identificar coleções de objetos semelhantes. O agrupamento dos dados pode ser baseado em funções de distância, que podem ser especificadas para diferentes contextos da aplicação. Pode-se, por exemplo, agrupar casas de uma área, de acordo com sua categoria, área construída e localização geográfica.

A tarefa básica da *clusterização* é agrupar um conjunto de objetos em subconjuntos, de acordo com os critérios apropriados [NEVES 8]. Esses subconjuntos agrupam elementos que têm um alto grau de semelhança ou similaridade, enquanto elementos pertencentes a grupos distintos têm pouca semelhança entre si.

Uma característica que torna o *clustering* uma das técnicas mais utilizadas é a sua habilidade de identificar estruturas diretamente dos dados, sem que haja um conhecimento prévio dos mesmos. Essa técnica tem sido bastante utilizada em análise exploratória de dados espaciais e em procedimentos de regionalização [OHATA, 10].

Ainda segundo NEVES (op.cit.) os critérios mais comuns adotados na técnica de agrupamento são a homogeneidade e a separação. A primeira refere-se a objetos pertencentes a um mesmo *cluster*, onde os objetos são o mais similares possível. Na segunda, os objetos de diferentes *clusters* devem ser os mais distintos possíveis.

A qualidade dos *clusters* gerados depende de uma série de definições estabelecidas pelo usuário como, por exemplo, escolha dos atributos, medidas de dissimilaridade, critérios de agrupamento, escolha do algoritmo e definição do número de *clusters*. A dissimilaridade normalmente é utilizada por essa técnica para avaliar o grau de semelhança entre dois objetos durante o processo de agrupamento. Muitas vezes, essa medida é apresentada como sendo a distância entre dois objetos [HAN, op.cit.]

Os métodos de particionamento buscam encontrar a melhor partição dos n objetos em k grupos. Normalmente os k *clusters* encontrados são de melhor qualidade do que os k *clusters* produzidos pelos métodos hierárquicos. Os métodos de particionamento mais utilizados são baseados em um ponto central (média dos atributos dos objetos – *K-média*) ou em um objeto representativo para o *cluster* (*k-medoids*) [BOGORNY, op.cit.].

2.3.4 - O algoritmo K-Médias

O K-Médias é um algoritmo para partição de um conjunto de elementos em determinado número de agrupamentos (*clustering*) K, utilizando o método dos mínimos quadrados e o ajustamento das médias (média móvel) no decorrer dos processos de iterações. O uso de um *software* para realizar esta operação se estrutura apenas na simples implementação de um algoritmo que se baseia nos mínimos quadrados. Este algoritmo, de forma geral, possui apenas duas etapas:

- A determinação dos centróides, que é a posição “média” de todas as amostras de um agrupamento ou classe, sendo o número de classes determinado previamente pelo usuário, em redor do qual se originarão os agrupamentos (classes), funcionando os mesmos como “sementes”, e a utilização destas para a formação de novos agrupamentos.
- A associação de cada elemento do conjunto à semente considerada “mais próxima”.

Um dos principais problemas para a aplicação desse método consiste em determinar maneiras de estabelecer a posição média entre objetos e a distância entre eles. Quando os dados não são georreferenciados, é necessário recorrer a todo tipo de artifício, formulando equações que sejam capazes de quantificar, por exemplo, a diferença entre atributos categóricos de objetos (por exemplo, renda “alta”, “média” ou “baixa”). No caso de dados espaciais, essa posição média e a distância podem ser obtidos como um subproduto do georreferenciamento, de forma bastante natural.

Existem mais de 20 tipos de algoritmos descritos como algoritmos de partições K-média. Eles são conhecidos não apenas por sua velocidade, como também por possuírem boas propriedades de convergência (cf. THEILER et GISLER [16]). No decorrer das iterações, o programa tenta minimizar a soma dos resíduos do processo dos mínimos quadrados, sobre todos os grupos e dentro de cada grupo isoladamente em relação à distância de cada objeto para o respectivo centróide de seu grupo. A convergência é alcançada quando o resíduo da soma dos quadrados não pode mais ser reduzido. Os agrupamentos obtidos são geometricamente os mais compactos possíveis,

em redor de seus respectivos centróides. Assim, o objetivo do k-média é dividir conjuntos de elementos em K classes, da forma mais simples possível.

Já a definição de centróide é dada por BRANSON [2] e descrita a seguir.

Seja q_1, q_2, \dots, q_n uma distribuição em n partes distintas de uma grandeza (tal como área, comprimento, massa ou volume), cada uma localizada, respectivamente, em $(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)$. As coordenadas cartesianas do Centróide (x_c, y_c, z_c) de um corpo rígido serão, então, expressas por:

$$X_c = \frac{\sum_{i=1}^n x_i q_i}{\sum_{i=1}^n q_i}; Y_c = \frac{\sum_{i=1}^n y_i q_i}{\sum_{i=1}^n q_i} \text{ e } Z_c = \frac{\sum_{i=1}^n z_i q_i}{\sum_{i=1}^n q_i}$$

2.3.5 - Software utilizado

Para o procedimento de clusterização, utilizou-se o *Software* TANAGRA, desenvolvido por pesquisadores da Universidade de Lyon [15]. O TANAGRA é um *software* livre para mineração de dados, para finalidades acadêmicas e de pesquisa. Propõe diversos métodos para mineração de dados, análise exploratória, classificação estatística, de computação e de pesquisa em bases de dados. TANAGRA conta com processos de classificação supervisionada e não-supervisionada, tais como clusterização, análise fatorial, estatísticas parametrizadas e não parametrizadas, regras de associação, seleção de características e algoritmos da construção. TANAGRA é um "*open source project*" (programa de código fonte aberto), no qual cada pesquisador pode acessar seu código de fonte e adicionar seus próprios algoritmos. Sua finalidade principal é fornecer a pesquisadores e estudantes um *software* de mineração de dados com facilidade de uso, adaptando-se às normas atuais do desenvolvimento do *software* no domínio da estatística, particularmente em sua interface homem-máquina e permitindo a análise de dados reais ou sintéticos. Outra finalidade é propor aos pesquisadores uma arquitetura que permita a adição de seus próprios métodos de mineração de dados, para comparar seus resultados.

Capítulo 3 – Metodologia e Preparação de Bancos de Dados

3.1 - Metodologia

As atividades desenvolvidas podem ser sintetizadas como se segue.

- Pré-processamento: (1) Origem dos dados. (2) Consolidação do banco de dados de doadores na região urbana de Belo Horizonte. (3) Classificação dos doadores por tipo de atividade. (4) Levantamento de dados quantitativos de material coletado. (5) Compatibilização do banco de dados de doadores obtida com as disponibilizadas pela PRODABEL – endereços BH
- Criação de pontos georreferenciados para o banco de dados de doadores
- Visualização e quantificação do banco de dados e sua distribuição espacial
- Classificação geográfica destes doadores através da Análise de *Clusters*
- Análise das classes geográficas obtidas.

3.1.1 - Origem de dados

Banco de dados de doadores cadastrados na RMBH

O banco de dados de doadores foi coligido por pessoal próprio da ASMARE, que utilizou critérios subjetivos e pessoais para a entrada dos dados, resultando em um banco de dados de baixa coerência e praticamente incompatível com as bases de dados de vias e endereços georreferenciados, como nos exemplos que se seguem.

Tabela 1 – Dados conforme são coletados pela ASMARE

Nome	Logradouro	Bairro	Peso Médio (kg)	Obs.
INSTITUTO SAGRADA FAMILIA	AV. CARLOS LUZ, 535	CAIÇARA	200	quinta
Rede Globo	Av. Américo Vespúcio, 2.045	Caçara	150	quinta
Posto REM – PETROBRÁS	Av. Amazonas, 4.594	Nova Suiça	50	sexta
MG MASTER LTDA	RUA ABAETE, 211	BONFIM	30	quinta
Refil Gráfico	Contorno, 6545 4º Andar	Savassi	30	
Correio	R. Boa Ventura, 401 portaria Sul	Jaraguá	100	
GRAFICA DA PREFEITURA	AV. PRESIDENTE CARLOS LUZ, 1275	CAIÇARA	500	quinta
EDITORA FORUM LTDA	RUA PIAUI, 1627 LOJA 01	FUNCIONARIOS	50	
ERICA	AV. BERNARDO GUIMARAES, 1151 APTO 402	FUNCIONARIOS	30	
Clínica Oculare	Rua Maranhão, 655	Santa Efigênia	80	
ED. ROSA MARIA	RUA NIQUELINA, 58	SANTA EFIGENIA	40	
PRESTAR RECURSOS HUMANOS	RUA NIQUELINA, 72	SANTA EFIGENIA	150	
HEMOMINAS	RUA GRAO PARA, 882 - 504	SANTA EFIGENIA	80	
Edifício Ravel	Rua Ceará, 1.695	Funcionários	30	
ED. IVONE COLAUD	RUA SERGIPE, 1348	SAVASSI	120	
INSTITUTO INOVACAO	RUA CLAUDIO MANOEL, 237	FUNCIONARIOS		
ED. JUNIA MARTINS	RUA DOS OTONI, 66 APTO 301	FUNCIONARIOS		
DAISY MARIA	RUA DOS TIMBIRAS 261 APT 1502	FUNCIONARIOS		
ED. PRES JUCELINO	RUA SANTA RITA DURÃO, 206 APTO 10	FUNCIONARIOS		
ED. DONATO DA FONSECA	RUA TIMBIRAS, 1484	FUNCIONARIOS		
Condominio Maria Alexandrina	Rua Aimorés, 612	Funcionários		
Edifício Henrique Gomes	Rua Maranhão 1673 apto 701	Funcionários		
Condomínio Montauban	Rua Maranhão, 1567	Funcionários		
Grupo Krypton	Rua Paraíba, 1051	Funcionários		
Maria Costa	Rua Santa Rita Durão, 206 - apart: 10	Funcionários		
Solar Alvorada	Rua Sergipe, 457	Funcionários		
YEDA	RUA AFONSO RIBEIRO, 190	PALMEIRAS		
Set Point	Av. Contorno, 3257	Santa Efigenia		
PATOLOGIA SAO PAULO	RUA DOS OTONI, 826	SANTA EFIGENIA		
SANTA BARBARA ENGENHARIA	RUA PADRE MARINHO, 37 2 ANDAR	SANTA EFIGENIA		
Ação Mineira para Educação	Rua Álvares Maciel, 59 - 9º andar	Santa Efigênia		
Clínica São Paulo (Laboratório)	Rua dos Otoni, 826	Santa Efigênia		
Escola Santa Luzia	Rua Itabirito, 333	Santa Efigênia		
Algodão Doce	Rua Maranhão, 181	Santa Efigênia		
BANCO REAL - GETULIO VARGAS	AV. GETULIO VARGAS,1245 4 ANDAR	SAVASSI		
Vila Franca	Rua Alagoas, 1453 apto 1303	Savassi		
Condominio Vivaldi	Rua Cláudio Manoel, 521	Savassi		
NEXTEL	RUA INCONFIDENTES, 1180	SAVASSI		
VISANET	Rua Paraíba, 1174 - Pilotis	Savassi		
Edifício Ivone	Rua Sergipe, 1348 - (Perto do Café Cancun)	Savassi		
MARIA DE FATIMA	RUA SERGIPE1014	SAVASSI		
PDV	Rua Tome de Souza, 273 - 9o andar	Savassi		

O banco de dados existente compreende as seguintes informações:

Doador, endereço, tipo de doador, frequência de coleta, data programada atualmente para coleta, veículo adotado atualmente para coleta.

Este banco de dados é atualizado mensalmente, sendo acrescido de novos doadores e expurgados os doadores inativos há mais de seis meses.

3.1.2 - Classificação dos doadores por tipo de atividade

A classificação adotada pela Associação tem uma lógica própria, derivada da experiência e necessidades do usuário, e é detalhada como se segue.

- Banco – são grandes doadores de papel, com volumes de grande grau de constância e homogeneidade.
- Comercial – Doam principalmente papelão em caixas, em pequenos volumes; dependendo da atividade, podem ser grandes doadores de latas de alumínio.
- Empresa - Principalmente papelão, em volumes maiores; pequena ocorrência de plásticos;
- Escola – Principalmente papel, em pequenas quantidades, com sazonalidade ligada à ocorrência de atividades escolares;
- Hospital – Grandes doadores de plásticos e papelões;
- Órgão Público - São grandes doadores de papel;
- Residencial – Pequenas ocorrências de papel; grandes volumes de plásticos e latas de alumínio.

Para os efeitos do presente trabalho, esta classificação tornou-se puramente indicativa, uma vez que a análise levou em consideração o posicionamento geográfico do doador e a quantidade de material doado.

3.1.3 - Levantamento de quantitativos de material coletado

A Aluvial Engenharia e Meio Ambiente Ltda. é empresa colaboradora da ASMARE, atuando na gestão de informações e quantificação de resíduos coletados, bem como no apoio técnico-operacional em geral. Dados foram levantados no período de Outubro de 2005 a Junho de 2006, compreendendo o acompanhamento de roteiros cumpridos na coleta e levantamento de quantitativos gerados em cada doador.

Para aferição de quantitativos de material doado foi empregada a estratégia de acompanhamento diário de rotas de coleta, durante um período de nove meses, nos quais foram acompanhados todos os veículos envolvidos, cada um por um período de

pelo menos duas semanas não consecutivas. Este acompanhamento permitiu o levantamento de algumas premissas, úteis para a aplicação da classificação, como se verá. As principais são as seguintes:

- A média de material coletado é de cerca de 1.500 kg de material por dia, por veículo;
- Os veículos de coleta têm a capacidade de efetuar de 20 a 35 coletas por dia. A grande variação se deve a fatores de difícil quantificação, tais como desconhecimento das rotas por parte dos motoristas, volume de tráfego nas vias e atrasos eventuais em pontos de coleta (dificuldade em localizar responsável pela guarda do material, demora na resposta ao chamado da campanha, etc.).
- Ocorrem atrasos de grande monta (de uma a duas horas) no início da operação de coleta, por falta de materiais de coleta (*big bags*) adequados.
- Ocorre um número considerável de coletas vazias, ou seja, coletas que falham, pois o doador visitado não tinha efetuado a coleta seletiva necessária até a data estipulada.

Devido às dificuldades operacionais (ausência de balança nos veículos, inexperiência e falta de preparo da mão-de-obra envolvida, etc.) foram efetuadas extrapolações para a quantidade de material coletada para cada categoria de doadores, utilizando-se o critério de proporcionalidade dos dados aferidos em campo.

3.1.4 - Base de dados de vias e endereços da Prodabel

A Prodabel, órgão de gerenciamento de dados da Prefeitura Municipal de Belo Horizonte dispõe de dados georreferenciados da tessitura urbana de Belo Horizonte. Estes dados foram recebidos em formato *.TAB, compatível com MapInfo 8.0, consistindo das seguintes informações: Tabelas de arruamentos, regiões de BH e bairros metropolitanos. Estes dados foram utilizados como referência de georreferenciamento e como auxiliares na tomada de decisões.

Capítulo 4 – Georreferenciamento, visualização, classificação

4.1 - Pré-processamento

Apesar de aparentemente simples e do número relativamente baixo de registros, o banco de dados de doadores apresentou grandes dificultadores em sua conversão e georreferenciamento, exigindo operações cansativas e repetitivas, entre as quais citamos os principais.

Ajustes Coluna LOGRADOURO

Necessitou ser transformada e subdividida em três colunas, conforme o banco de dados de endereços da PRODABEL:

TPLOG	NOLOG	NUIMOV
-------	-------	--------

No banco de dados PRODABEL, a coluna TPLOG apresenta um número limitado de definições, o que exigiu operações de compatibilização como as que se seguem.

Filtrar registros para LOGRADOURO contendo a expressão “av.”

Para todos os registros filtrados, TPLOG = “AVE”

Na coluna LOGRADOURO, substituir em todos os registros a expressão “av.” por “”

Filtrar registros para LOGRADOURO contendo a expressão “R.”

Para todos os registros filtrados, TPLOG = “RUA”

Na coluna LOGRADOURO, substituir em todos os registros a expressão “R.” por “”

Filtrar registros para LOGRADOURO contendo a expressão “RUA”

Para todos os registros filtrados, TPLOG = “RUA”

Na coluna LOGRADOURO, substituir em todos os registros a expressão “RUA” por “”

Filtrar registros para LOGRADOURO contendo a expressão “PRAÇA”

Para todos os registros filtrados, TPLOG = “PCA”

Na coluna LOGRADOURO, substituir em todos os registros a expressão “PRAÇA” por “”.

Filtrar registros para LOGRADOURO contendo a expressão “PÇA”

Para todos os registros filtrados, TPLOG = “PCA”

Na coluna LOGRADOURO, substituir em todos os registros a expressão “PÇA” por “”

Também foram efetuadas as seguintes operações acessórias:

- Converter todos os registros em maiúsculas
- Remover todas as formas de acentuação
- Substituir todos os “Ç” por “C”
- Separar os campos NOLOG e NUIMOV
- Separar em coluna específica atributos especiais de endereçamento – andar, número de apartamento, indicações de proximidade ou outras referências – nome de portaria de acesso, etc.

Além disso, foram corrigidas e verificadas uma a uma as seguintes informações:

- COLAB – Razão social ou outro identificador do colaborador. Trata-se da referência pela qual o colaborador está registrado no banco de dados da ASMARE, refletindo na maior parte das vezes o relacionamento entre o doador e a Associação.
- TIPCOLAB – Tipologia da atividade exercida pelo doador cadastrado.

4.2 - Geocodificação do banco de dados de doadores

Após os trabalhos de pré-processamento, o banco de dados foi submetida à operação de busca de endereços para sua geocodificação, através do *software* MapInfo, utilizando-se como referência o já citado banco de endereços da PRODABEL.

Utilizou-se a função SQL Select do MapInfo, cruzando-se as bases

- *Base_INIC2* (base de doadores ASMARE)
- *Enderecotransp* (base de endereços de Belo Horizonte – Prodabel)

A condição de pesquisa foi implementada como:

Base_INIC2 . NOLOG = enderecotransp . NOLOG and Base_INIC2 . NUIMOV = enderecotransp . NUIMOV

Ordenando-se os resultados por

Base_INIC2 . IDEND

Dos 559 registros iniciais do banco de dados, a seleção SQL conseguiu localizar 480 endereços na primeira tentativa; outros 83 registros puderam ser geocodificados após correções diversas no banco de dados (correções estas principalmente em nomes de ruas e numeração de endereços). Após a eliminação de registros repetidos, chegou-se ao banco de dados com 436 registros, que foi o destinado à análise através do processo de *clusterização*, como se verá a seguir.

4.2.1 - Visualização e quantificação do banco de dados e sua distribuição espacial

Após o georreferenciamento, foi criado um objeto pontual para cada doador; O resultado foi mapeado através do *software* MapInfo, resultando na visualização dos doadores sobre a malha urbana de Belo Horizonte, como se vê na figura 2.

Doadores de Material Reciclável em Belo Horizonte

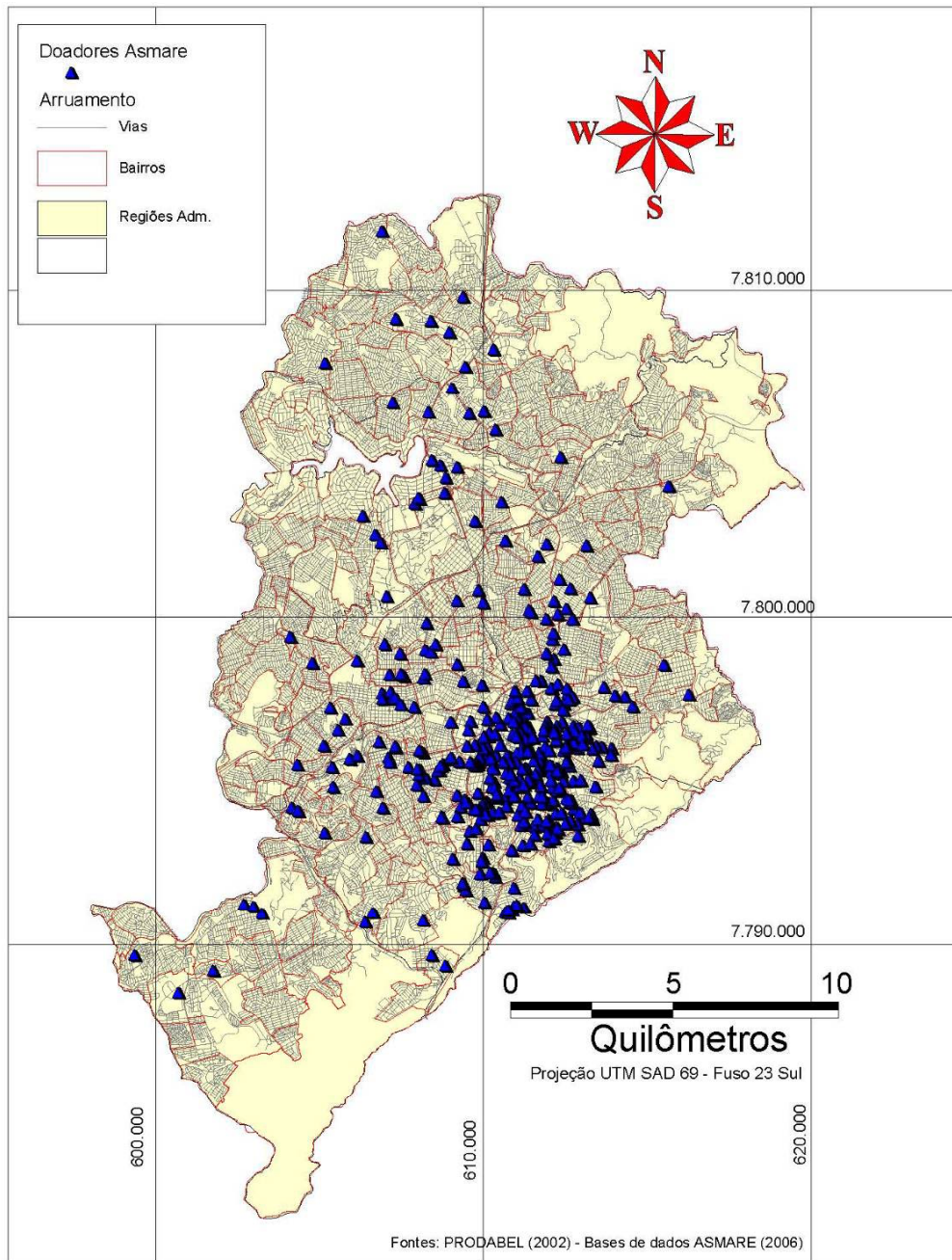


Figura 2 – Distribuição geográfica de doadores

Percebe-se que os pontos estão concentrados na regional Centro-Sul, mas existem grupos significativos de outros pontos nas regiões leste, oeste e noroeste. A figura 3 indica a divisão percentual dos doadores nas regionais do município.

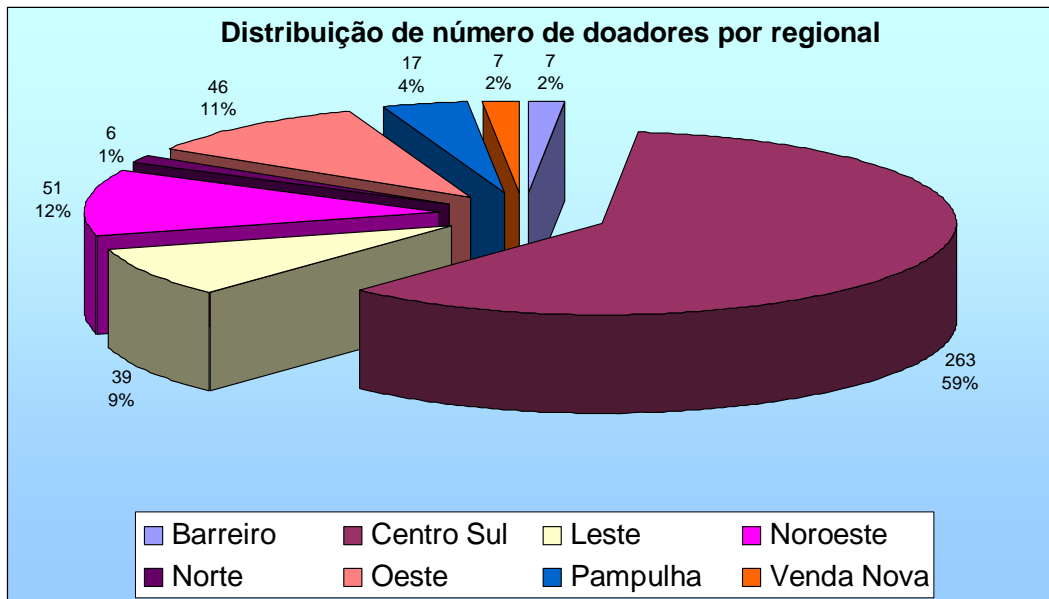


Figura 3 – Percentual de doadores por regional do município de Belo Horizonte

4.2.2 - Mapa de tipos de doadores

Para demonstrar outras potencialidades da análise, a distribuição espacial por tipo de doador é apresentada na figura 4. Através dela e também da figura 5, verifica-se a grande participação de doações de particulares, além da expressiva contribuição do setor bancário.

Doadores ASMARE - Tipos de Doadores

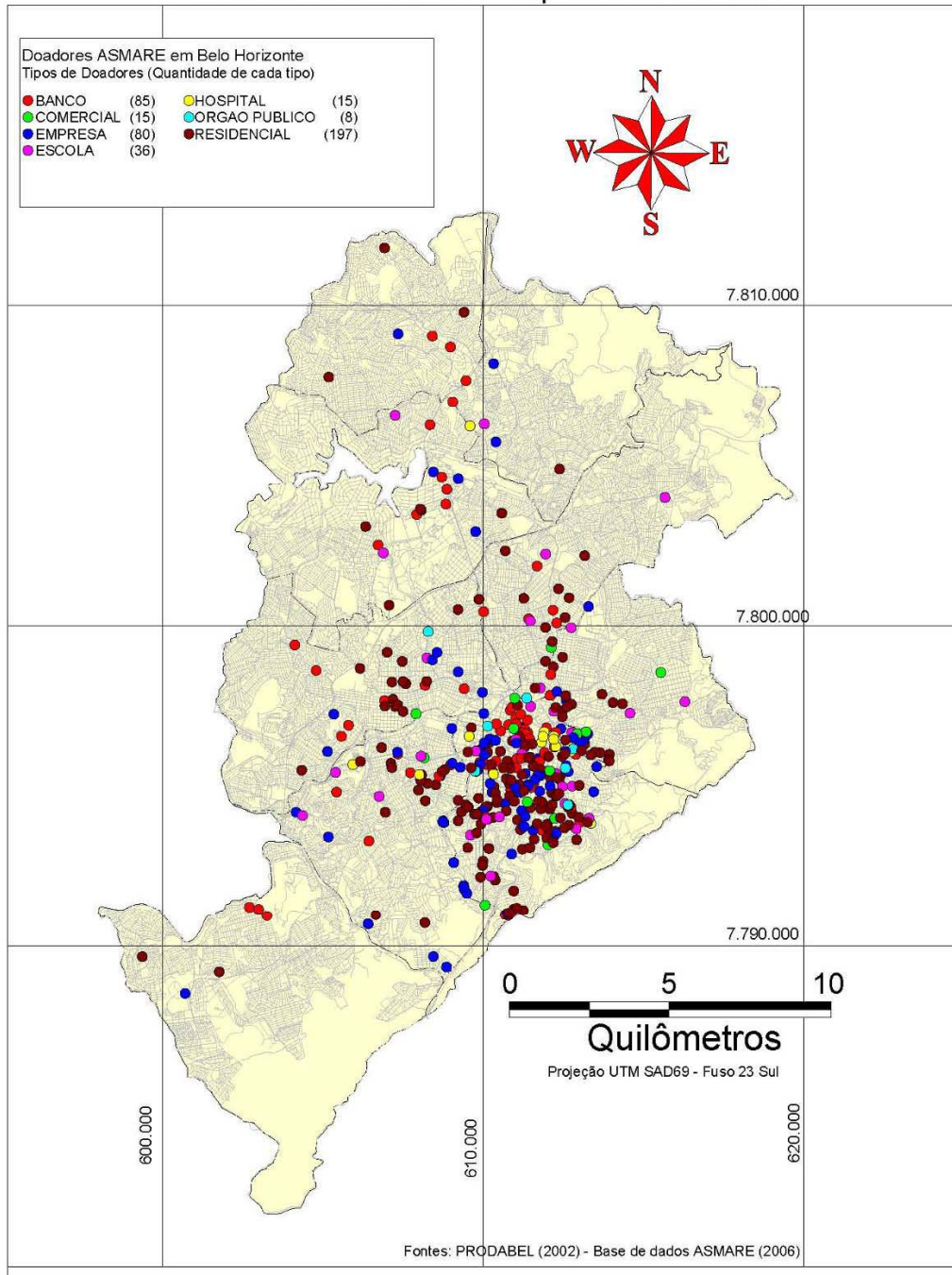


Figura 4 – Tipos de doadores

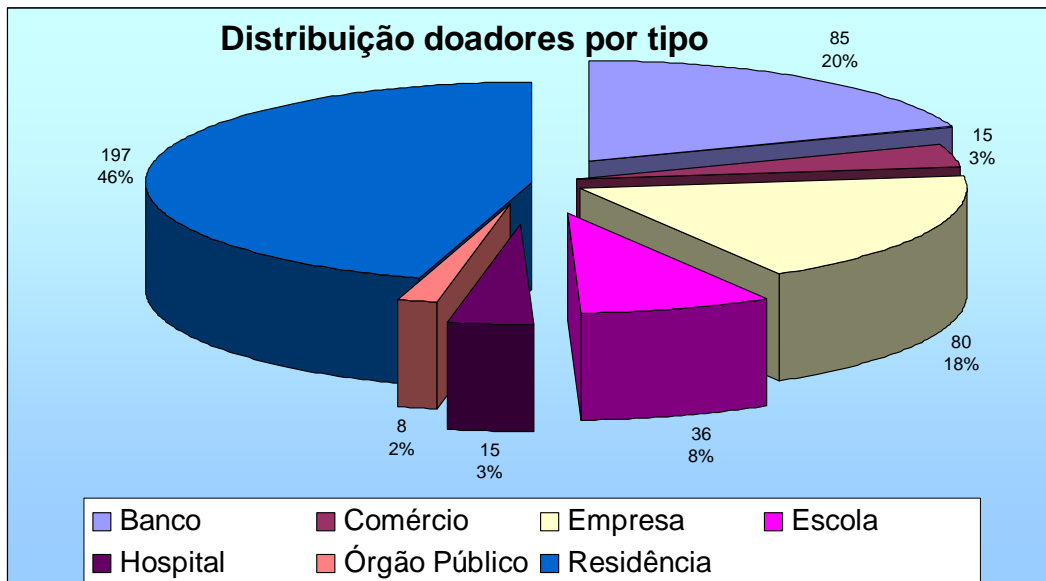


Figura 5 – Número de doadores por regional do município de Belo Horizonte

4.2.3 - Mapa temático de quantidades doadas

Elaborou-se também mapa temático indicando a análise quantitativa de material recolhido. Mais uma vez verificou-se a grande concentração das regiões Centro-sul e Noroeste do município, conforme figuras 6 e 7.

Doadores ASMARE BH - Quantidade doada em cada local

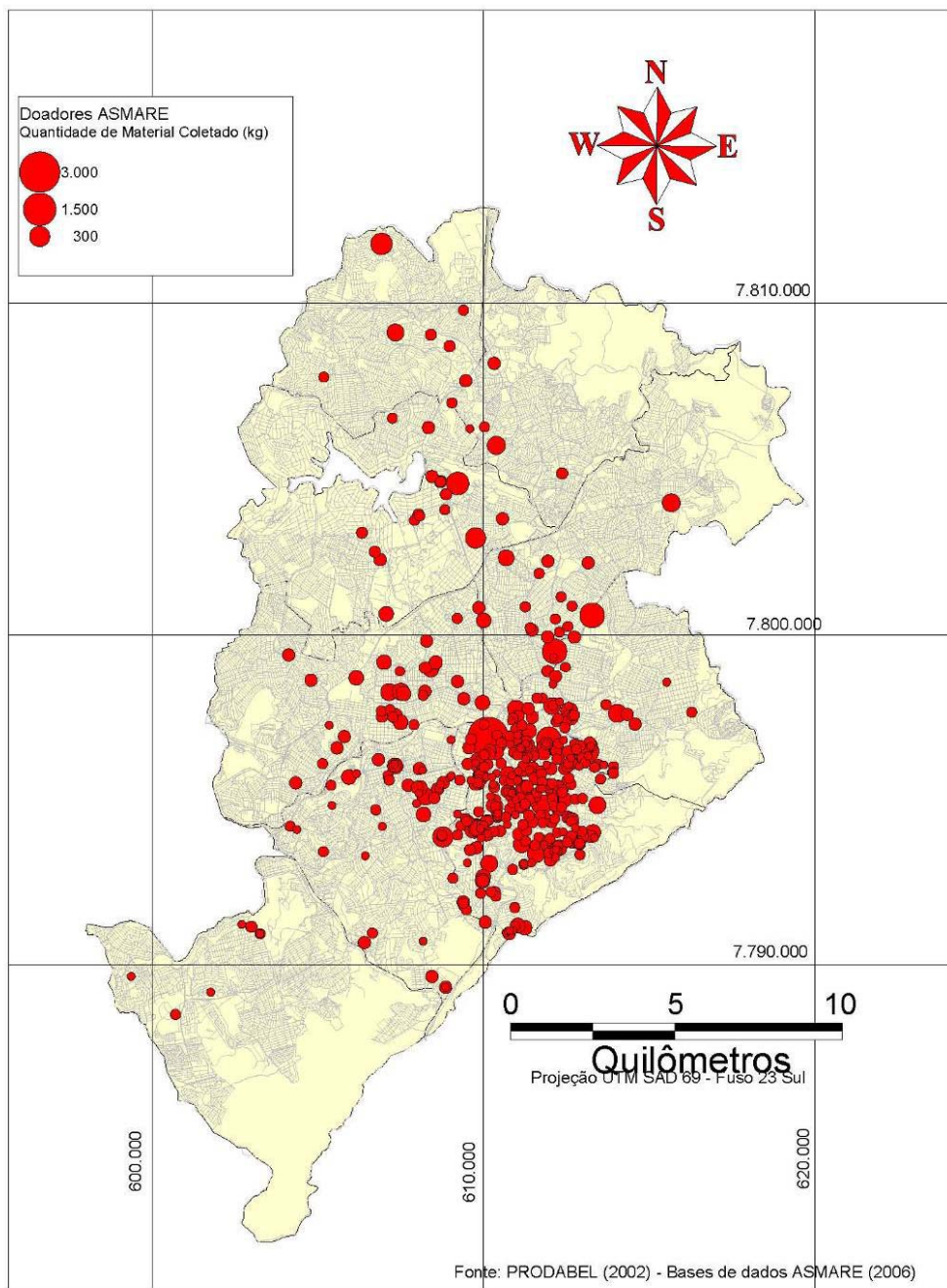


Figura 6 – volume de doações

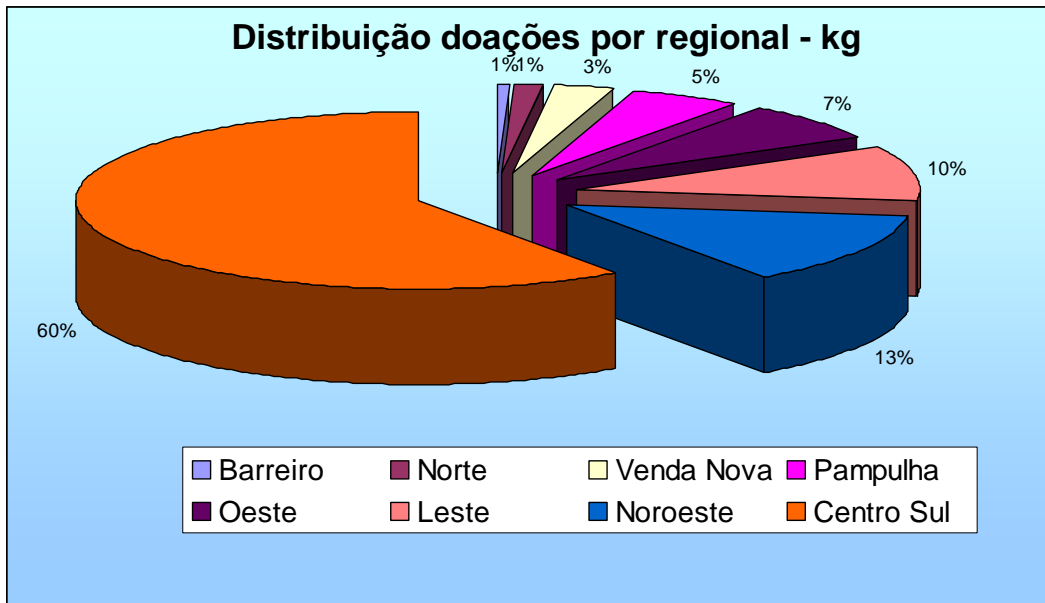


Figura 7 – Volume de doações por regional em quilogramas

4.3 - Implementação de classificação não-supervisionada sobre a base de dados de doadores

Efetuiu-se a divisão do banco de dados obtido, de 436 doadores, em 32 classes. Este número foi adotado para que cada classe representasse um veículo/rota/dia, para um total de 04 (quatro) veículos de coleta, considerando duas rotas por veículo por dia, durante quatro dias por semana, perfazendo, portanto, um total de 32 roteiros. Foi reservado o equivalente a um dia da semana para a coleta corretiva, no qual são realizadas as coletas que, por quaisquer motivos, não possam ser efetuadas no dia apropriado.

4.3.1 - Clusterização puramente geográfica

Inicialmente, a divisão em classes utilizou critérios puramente geográficos. O algoritmo k-média foi aplicado sobre o banco de dados no estado em que esta se encontrava, definindo-se como variáveis de entrada os valores de coordenadas x e y, representando as coordenadas geográficas de cada doador.

A tabela resultante encontra-se no anexo I.

Os dados foram resumidos na tabela 2 e mapeados na figura 8.

Tabela 2 – *Clusterização* local – total coletado e número de doadores em cada *cluster*

Número do <i>cluster</i>	Total de material coletado	Contagem de doadores
1	405	6
2	945	12
3	120	2
4	735	19
5	1605	13
6	1050	20
7	1050	24
8	210	8
9	240	4
10	1650	30
11	165	7
12	1200	26
13	240	6
14	465	11
15	315	5
16	4605	27
17	630	9
18	870	13
19	1320	24
20	465	7
21	210	1
22	1110	16
23	195	7
24	1140	11
25	1425	17
26	720	7
27	420	9
28	750	14
29	630	17
30	1065	20
31	915	18
32	1065	26
Total	27930	436

Clusterização Local

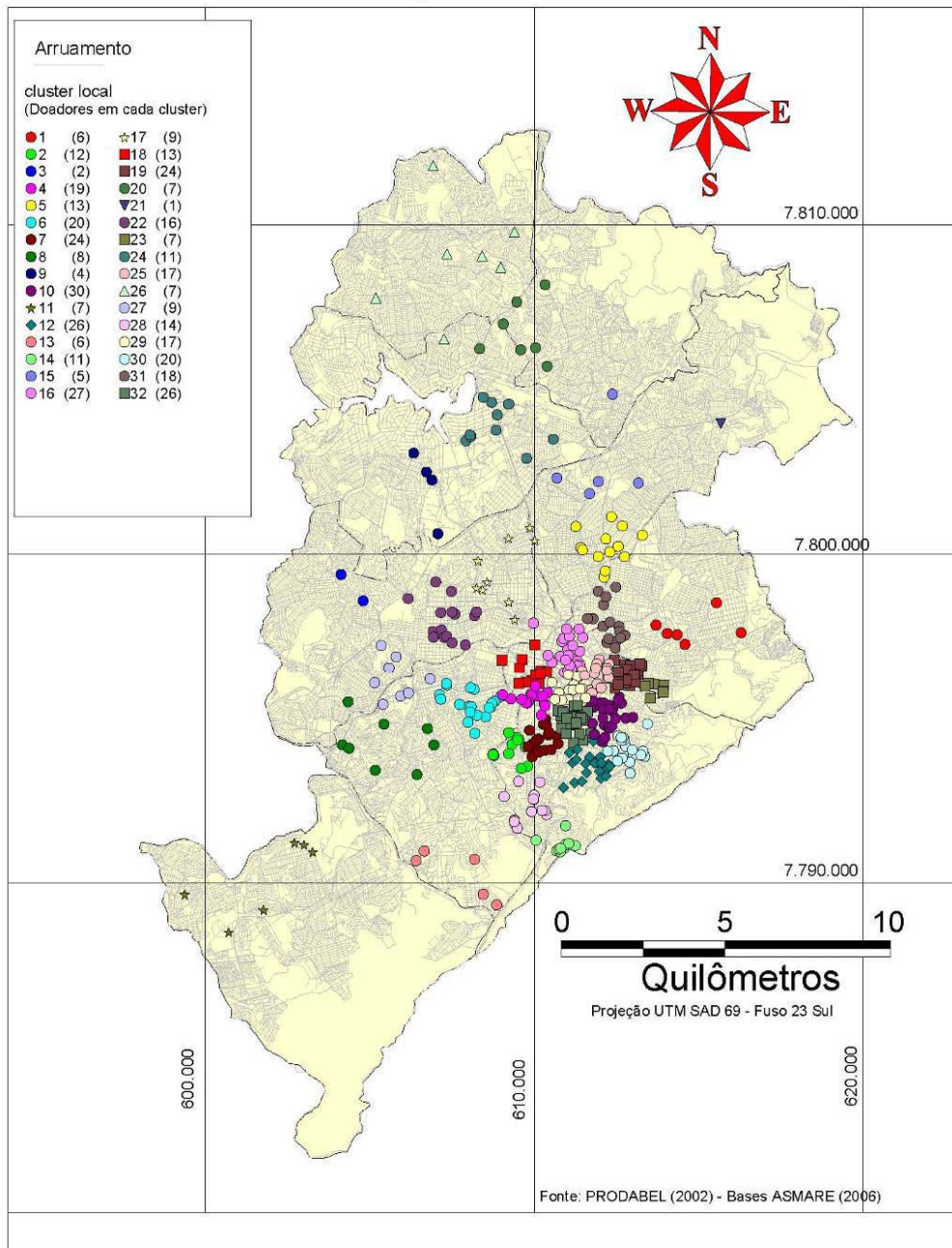


Figura 8 – Clusterização local

4.3.2 - *Clusterização ponderada*

Alternativamente à classificação por critérios puramente geográficos, buscou-se uma classificação que levasse em conta também a quantidade de material coletado em cada doador. Devido às limitações do algoritmo implementado no *software* utilizado, foi necessário construir artificialmente um peso através de um processo de normalização.

Inicialmente, determinou-se que uma quantidade de material coletado igual a 15 kg corresponderia a um peso 1; Todos os demais valores de material coletado foram definidos como múltiplos inteiros deste valor, e replicados conforme este múltiplo, gerando uma tabela artificialmente maior, com 1.865 registros; a tabela resultante se encontra no anexo II. O algoritmo foi então aplicado sobre o banco de dados resultante, definindo-se como variáveis de entrada os valores de coordenadas x e y, representando as coordenadas geográficas de cada doador, repetidas *n* vezes nos casos em que o material coletado neste doador era um múltiplo da quantidade básica.

A tabela resultante encontra-se resumida no anexo III.

Os dados foram resumidos na tabela 3.

Tabela 3 – *Clusterização* ponderada – total coletado e número de doadores em cada cluster

Número do <i>cluster</i>	Total de material coletado	Contagem de doadores
1	1.140,00	11
2	3.000,00	1
3	3.420,00	50
4	210,00	1
5	360,00	6
6	1.455,00	21
7	465,00	11
8	690,00	14
9	0	0
10	0	0
11	210,00	3
12	1.440,00	29
13	1.200,00	31
14	375,00	6
15	1.290,00	28
16	360,00	1
17	1.935,00	35
18	0	0
19	1.710,00	32
20	240,00	6
21	510,00	13
22	1.020,00	23
23	1.260,00	24
24	240,00	4
25	165,00	7
26	1.215,00	22
27	750,00	14
28	495,00	11
29	405,00	6
30	840,00	7
31	1.020,00	11
32	510,00	8
Total	27930	436

Observe-se que a numeração identificadora de *clusters* não guarda qualquer razão de semelhança ou posicionamento, tendo sido gerada aleatoriamente pelo próprio algoritmo utilizado.

Observe-se ainda que a *clusterização* ponderada chegou a três *clusters* sem quaisquer elementos (contagem de doadores igual a zero).

4.3.3 - Pós-processamento

Das premissas encontradas anteriormente, na fase de levantamento de dados de coleta, verificou-se que operacionalmente eram verificadas as seguintes condições:

- Média de material coletado de cerca de 1.500 kg de material por dia, por veículo;
- Capacidade de efetuar de 20 a 35 coletas por dia, por veículo.

Adotadas estas premissas, observa-se que os *clusters* de números 3 e 17 extrapolaram estes valores, seja no número de doadores ou no volume de material coletado.

Tabela 4 – *Clusters* com elementos e quantidade de material coletado acima dos parâmetros

Número do cluster	Total de material coletado	Contagem de doadores
3	3.420,00	50
17	1.935,00	35

Considerando-se este fato e o de que foram encontrados 3 clusters sem nenhum elemento, propõe-se a divisão dos elementos encontrados em cada um destes *clusters* em dois outros. Para tanto, foi aplicado novamente o algoritmo K-média. Os resultados se encontram listados na tabela 5.

Tabela 5 – *Clusterização* ponderada – total coletado e número de doadores em cada *cluster* corrigidos

Número do <i>cluster</i>	Total de material coletado	Contagem de doadores
1	1140	11
2	3000	1
3	1815	23
4	210	1
5	360	6
6	1455	21
7	465	11
8	690	14
9	1605	27
10	885	11
11	210	3
12	1440	29
13	1200	31
14	375	6
15	1290	28
16	360	1
17	1050	24
19	1710	32
20	240	6
21	510	13
22	1020	23
23	1260	24
24	240	4
25	165	7
26	1215	22
27	750	14
28	495	11
29	405	6
30	840	7
31	1020	11
32	510	8
TOTAL	27930	436

Verificou-se que ainda restou um *cluster* com número de doadores igual a zero. Sugere-se que o veículo e dia designados sejam reservados também para a efetivação de coleta corretiva.

O mapeamento correspondente se encontra na figura 9.

Clusterização Ponderada

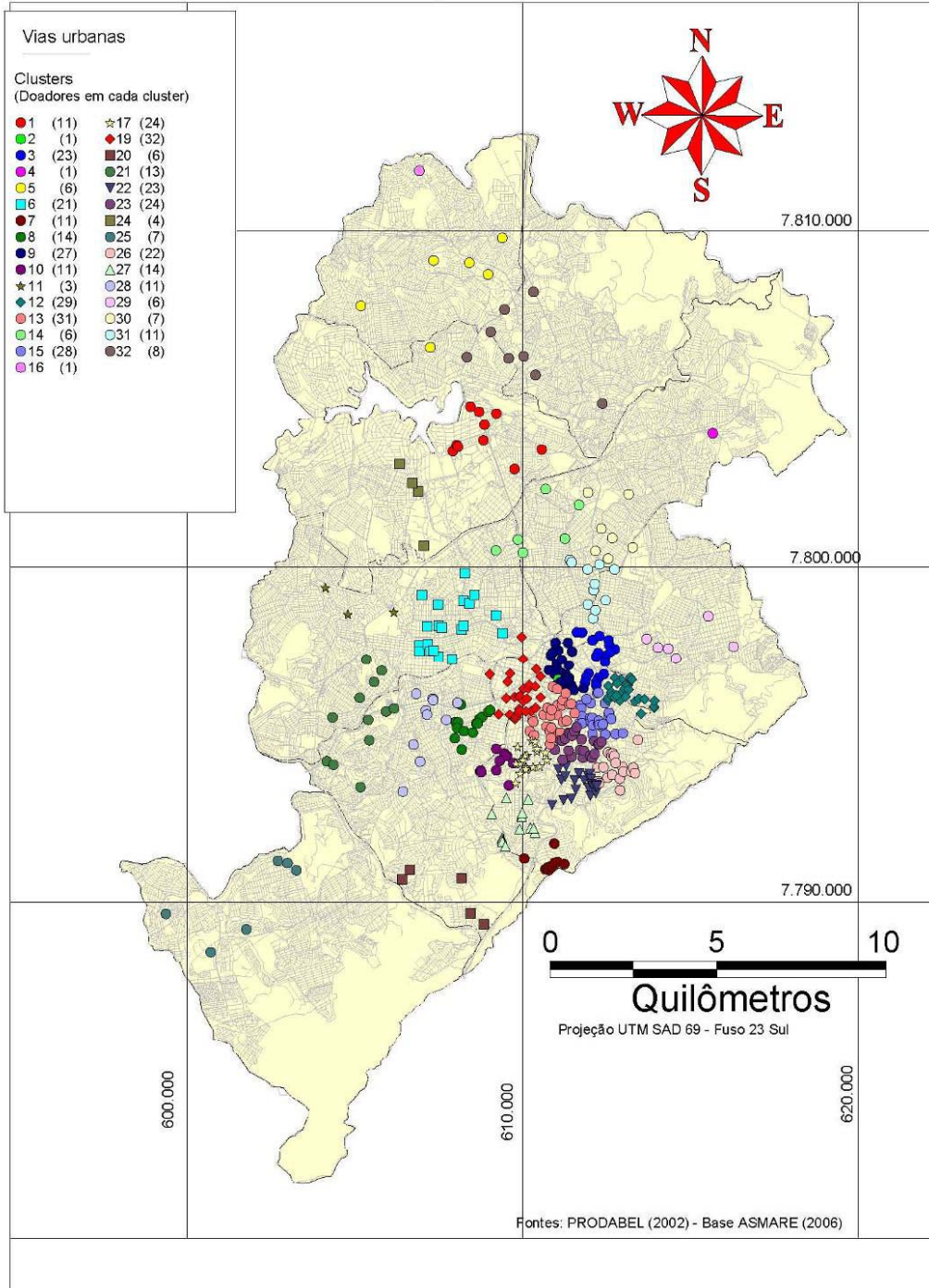


Figura 9 – Clusterização Ponderada

4.4 - Comparação de resultados

4.4.1 - Distância percorrida em cada *cluster*

Para comparar os resultados encontrados nos dois métodos utilizados para *clusterização*, buscou-se comparar índices que melhor representassem a distribuição dos elementos em cada *cluster* encontrado. Um dos índices definido foi a distância percorrida para se atingir todos os elementos pertencentes a um mesmo *cluster*.

Um cálculo exato desta distância envolveria a elaboração de um roteiro físico para cada um dos *clusters*, com a identificação de vias, suas mãos direcionais e um correto sistema de primazia entre os elementos; trabalho de grande vulto, e que fugiria ao escopo desta apresentação. Para efeitos de simplificação, adotou-se o conceito de distância euclidiana em cada *cluster*, como se segue.

Num Sistema de Coordenadas Cartesianas, a função que usualmente se emprega para medir a distância entre dois pontos P_1 e P_2 de coordenadas (x_1, y_1) e (x_2, y_2) , respectivamente, é dada pela fórmula

$$d(P_1, P_2) = d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

que é uma decorrência do Teorema de Pitágoras da Geometria Euclidiana Plana e, por isso, é chamada de distância Euclidiana.

Dentro de cada *cluster*, ordenaram-se os elementos da menor coordenada x para a maior, conforme representação genérica na figura 10; calculou-se então a distância percorrida através da fórmula acima.

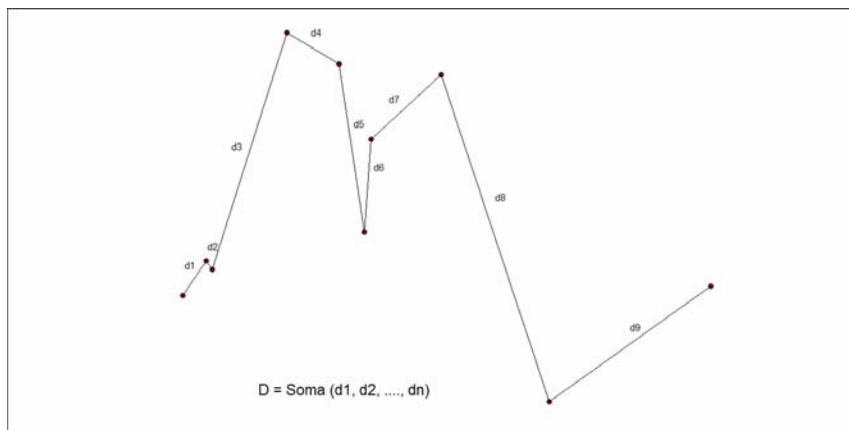


Figura 10 – Distância percorrida em cada *cluster*

Os valores encontrados estão apresentados nas Tabelas 6 e 7.

Tabela 6 – *Clusterização local – Distância total percorrida em cada cluster*

Identificação do <i>cluster</i>	Número de elementos no <i>cluster</i>	Distância total percorrida em cada <i>cluster</i> (m)
1	6	5.413,22
2	12	8.451,11
3	2	1.040,38
4	19	8.243,47
5	13	14.056,43
6	20	18.230,12
7	24	13.462,15
8	8	15.983,33
9	4	3.829,87
10	30	21.249,52
11	7	7.899,55
12	26	15.315,90
13	6	6.642,24
14	11	3.843,09
15	5	8.390,73
16	27	16.830,29
17	9	8.657,21
18	13	7.060,50
19	24	12.603,84
20	7	7.992,45
21	1	-
22	16	16.372,85
23	7	2.085,82
24	11	11.479,74
25	17	8.203,51
26	7	14.657,13
27	9	6.878,56
28	14	7.220,94
29	17	8.190,78
30	20	10.401,47
31	18	11.822,29
32	26	13.789,69
TOTAL	436	316.298,18

Tabela 7 – *Clusterização Ponderada – Distância total percorrida em cada cluster*

Identificação do <i>cluster</i>	Número de elementos no <i>cluster</i>	Distância total percorrida em cada <i>cluster</i> (m)
1	11	6.341,98
2	1	-
3	23	11.640,78
4	1	-
5	6	7.890,05
6	21	13.817,76
7	11	2.637,29
8	14	4.409,39
9	27	14.644,78
10	11	3.475,99
11	3	2.409,64
12	29	7.738,17
13	31	14.785,83
14	6	5.822,37
15	28	11.277,34
16	1	-
17	24	7.647,30
19	32	19.105,91
20	6	3.546,76
21	13	17.390,41
22	23	8.429,84
23	24	7.565,59
24	4	2.627,81
25	7	5.902,13
26	22	6.941,29
27	14	5.507,74
28	11	8.500,26
29	6	3.878,27
30	7	6.969,35
31	11	7.302,05
32	8	10.298,81
TOTAL	436	228.504,89

Desprezou-se o cálculo de distâncias naqueles *clusters* com apenas um elemento. Ainda assim, obteve-se uma distância média percorrida em cada cluster 20,02 % menor na *clusterização* ponderada (média de 10.203,17 m na *clusterização* local versus 8.160,89 m na *clusterização* ponderada).

Além disso, calculado um fator de dispersão F_1 dado por

$$F_1 = D_n / N_n \quad \text{onde:}$$

D_n é a distância percorrida em cada *cluster*

N_n é o número de elementos pertencentes a cada *cluster*, obteve-se:

Para *clusterização* local:

$$F_{1, \text{médio}} = 833,42 \text{ m/elemento}$$

Para a *clusterização* ponderada:

$$F_{1, \text{médio}} = 624,65 \text{ m/elemento}$$

Correspondendo a uma dispersão 25,05 % menor para a *clusterização* ponderada.

4.4.2 - Quantidade de material coletado e média de doadores em cada *cluster*

Dos dados das tabelas 2 e 5, obteve-se 872,81 kg /*cluster* na *clusterização* local versus 900,67 kg/*cluster* na *clusterização* ponderada. Ainda dos dados das tabelas 2 e 5, obteve-se médias de 13,625 doadores /*cluster* na *clusterização* local versus 14,065 doadores/*cluster* na *clusterização* ponderada; Um resultado 3,22 % maior no segundo caso. O resultado encontrado, decorrente do número de *clusters* com zero elementos no segundo caso, aponta para uma maior rentabilidade no processo de coleta na *clusterização* ponderada.

Capítulo 5 - Conclusão

Os resultados encontrados servirão à ASMARE como ponto de partida para a seleção de áreas preferenciais para campanhas de implantação e aprimoramento de coleta seletiva. Os *clusters* de número 7, 8, 12, 13, 15, 17, 19, 20, 21, 22, 23, 25 e 27 apresentam média de coleta abaixo da do conjunto, descontados aqueles *clusters* de somente um doador com valor de coleta expressivo. A estatística é indicativa de capacidade ociosa em cada um destes agrupamentos.

Avaliações econômicas detalhadas devem ser efetuadas para os *clusters* de número 5, 21 e 32, que apresentam o maior coeficiente de distância percorrida por kg de material coletado – respectivamente 21,92, 34,10 e 35,77 m/ kg; Vide figura 11.

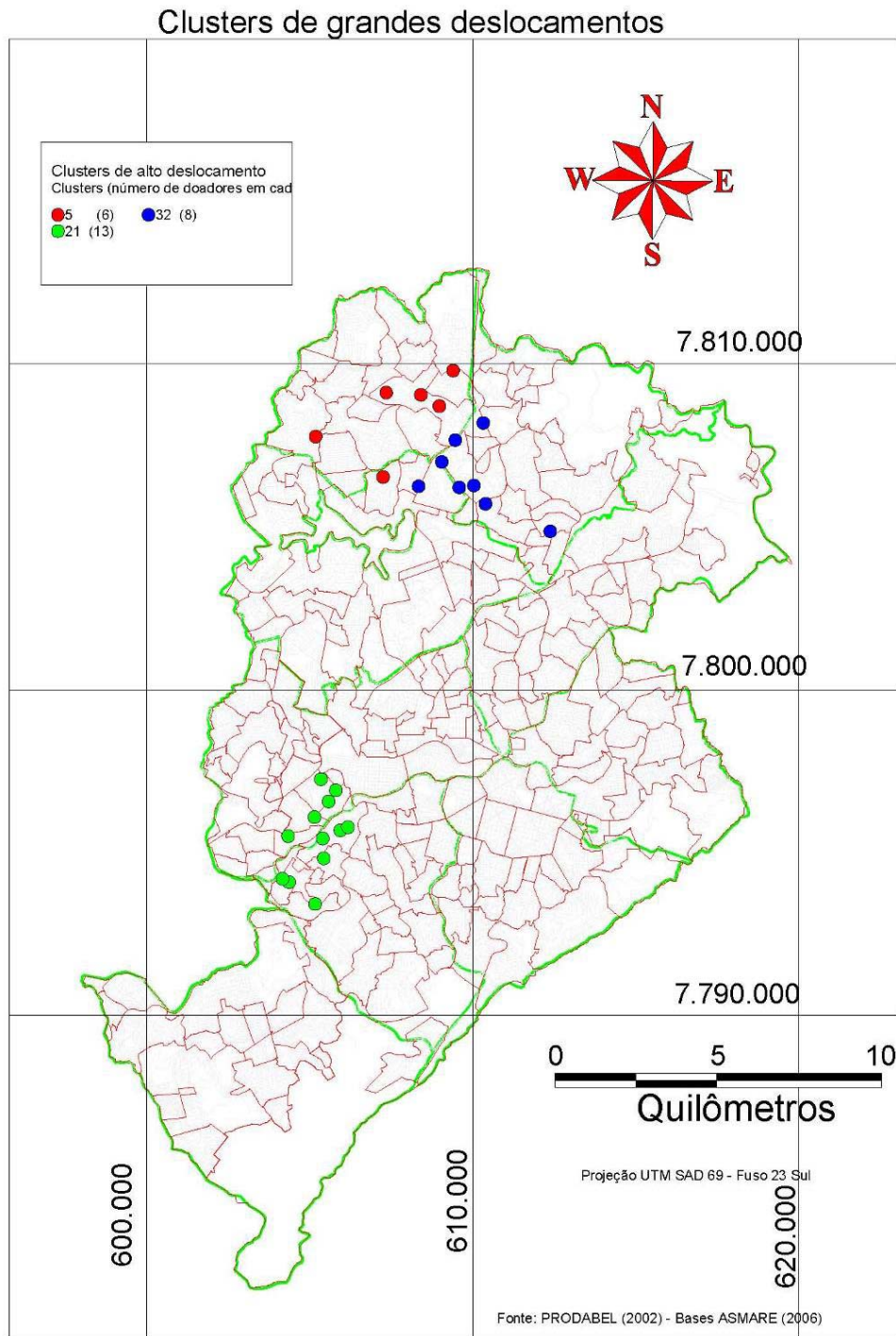


Figura 11 – Clusters de grandes deslocamentos

Não pouco a propósito, estes *clusters* se encontram também mais distantes dos pontos de maior concentração de doadores. A análise deve levar em consideração ou um aumento de intervalo de coleta ou a busca de outros possíveis doadores nestas regiões, que justifiquem economicamente os deslocamentos de coleta.

A *clusterização* demonstrou ser uma ferramenta poderosa de classificação não supervisionada de conjuntos; sua aplicação resultou em subconjuntos (*clusters*) homogêneos, de grande utilidade prática. Entretanto, a *clusterização* ponderada pela quantidade de material coletado demonstrou um inequívoco ganho qualitativo nos *clusters* formados, que resultará em economia de rotas e energia na coleta. O resultado evidencia a oportunidade de desenvolvimento de nova ferramenta, um algoritmo k-média ponderado, que leve em conta um fator de atribuição de valor ou quantidade a cada um dos elementos geográficos individuais. Suas aplicações seriam muitas, tais como a implantação de centros de distribuição de medicamentos e insumos; combate a pandemias através de vacinação; implantação de centrais de segurança e postos policiais; e muitas outras que fogem ao escopo deste trabalho.

Referências Bibliográficas

1. BOGORNY, Vânia. **Algoritmos e Ferramentas de Descoberta de Conhecimento em Bancos de Dados Geográficos**. Trabalho Orientado pelo Prof. Dr. Luís Otávio Campos Álvares – Porto Alegre: PPGC da UFRGS, 2003.43.: il. – (TI III - 1120).
2. BRANSON, L. K. **Mecânica, Estática e Dinâmica**, Livros Técnicos e Científicos Editora S. A., Rio de Janeiro, 1974. 430 pp.
3. DAGNINO, R. S. - **Um olhar geográfico sobre a questão dos materiais recicláveis em Porto Alegre: sistemas de fluxos e a (in)formalidade, da coleta à comercialização**. / Trabalho de Conclusão do Curso de Geografia. - Universidade Federal do Rio Grande do Sul. Instituto de Geociências. Porto Alegre, RS - BR, 2004.
4. DAVIS JR., C. A.; FONSECA F. T., BORGES, K. A. V., 2003, **A Flexible Addressing System for Approximate Geocoding**. V *Simpósio Brasileiro de Geoinformática - GeoInfo 2003, Campos do Jordão (SP)*
5. DAVIS JÚNIOR, Clodoveu Augusto. **Introdução aos sistemas de informação geográficos**. Belo Horizonte: PRODABEL, 2001. 275f.
6. MELO, A. C. S.; GIANARELLI, P. C.; GOMES, E. G.; FERREIRA FILHO, V. J. M. **Sistemas de roteamento de veículos e gestão da cadeia de suprimentos: uma abordagem analítica**. In XXXI SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL - SBPO, 1999, Juiz de Fora. Anais... Juiz de Fora: SOBRAPO, 1999, p. 690-704. (CD-ROM).
7. MOURA, M.C., RIBEIRO, C.A.A.S, TIBIRIÇÁ, A.C.G. - **Criação da rede e Base De Dados para o Sistema Viário de Viçosa – MG, Utilizando SIG** - Universidade Federal de Viçosa (UFV) – in Revista Brasileira de Cartografia Nº 54 – Viçosa – MG – Brasil
8. NEVES, M. C.; FREITAS, C. C.; CAMARA, G. **Mineração de Dados em Grandes Bancos de Dados Geográficos**. INPE, 2001. Relatório Técnico.
9. NG, R. T.; HAN, J. **Efficient and Effective Clustering Methods for Spatial Data Mining**. In: Twentieth International Conference on Very Large Data Base, Santiago, 1994.

10. OHATA, Arlete Tiekó; QUINTANILHA, José Alberto - **O uso de algoritmos de clustering na mensuração da expansão urbana e detecção de alterações na Região Metropolitana de São Paulo (1991-2002)** – in Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, 16-21 abril 2005, INPE, p. 647-655. 1PTR/EPUSP - Universidade de São Paulo
11. ROCHA, César Henrique Barra & FILHO, Luiz Fernandes de Brito. **Locais Para Aterro Sanitário de Mangaratiba / Rj: Seleção por Geoprocessamento.** Monografia apresentada a Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2000.
12. ROVIRIEGO, L.F.V – Proposta de uma metodologia para a avaliação de sistemas de coleta seletiva de resíduos sólidos domiciliares - /dissertação apresentada à Escola de Engenharia de São Carlos – USP – Orientador: Prof. Dr. Edson M. Aguiar, São Paulo, 2005
13. SILVA, Jorge Xavier da. **Geoprocessamento para análise ambiental.** Ed. Rio de Janeiro – RJ - 2001.
14. SOUZA, L.A.; LAENDER, A.H. F.; DAVIS JR., C.A., BORGES, K.A.V. e outros - **LOCUS: um localizador espacial urbano** - UFMG – Universidade Federal de Minas Gerais –Departamento de Ciência da Computação – Av. Presidente Antônio Carlos, 6627, 31270-901, Belo Horizonte, MG, Brasil
15. TANAGRA: Ricco Rakotomalala, **TANAGRA: Um logiciel de data mining gratuit pour l'enseignement et la recherche.** em Actes de EGC'2005, RNTI-E-3, Vol. 2, pp.697-702, 2005. (tradução) Ricco RAKOTOMALALA, “TANAGRA: um software livre para a pesquisa e finalidades académicos”, nas continuações de EGC'2005, RNTI-E-3, Vol. 2, pp.697-702, 2005. (em francês) Ricco RAKOTOMALALA Lyon - FRANCE Janeiro 2004. Última modificação: Maio 18o, 2006. Software livre disponível para download em <http://eric.univ-lyon2.fr/~ricco/tanagra/>
16. THEILER, J. et GISLER, Galen - **A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation** – Astrophysics and Radiation Measurements Group, NIS-2 – Nonproliferation and International Security Division, MS-D436 – Los Alamos National Laboratory, Los Alamos, NM 87545 USA

Anexos