



Universidade de Coimbra
Faculdade de Ciências e Tecnologia
Departamento de Matemática

Dissertação de Mestrado em Engenharia Geográfica
2012/2013

**Construção de um SIG para Apoio à Decisão numa
Empresa de Formação**

Building a GIS for Decision Support in a Training Company

Luís Miguel Conceição Silva dos Reis

Orientador: Cidália Fonte

Coimbra, 2013

Universidade de Coimbra
Faculdade de Ciências e Tecnologia
Departamento de Matemática

**Construção de um SIG para Apoio à Decisão numa
Empresa de Formação**

Dissertação para obtenção do grau de mestre em Engenharia Geográfica
no curso de mestrado em Engenharia Geográfica da Faculdade de Ciências
e Tecnologia da Universidade de Coimbra

Coimbra, 2013

"... A Geografia e a cartografia, em particular, são matérias que envolvem um conhecimento estratégico, o qual permite às pessoas que desconhecem seu espaço e sua representação, passarem a organizar e dominar esse espaço".

(Yves Lacoste)

Agradecimentos

Um agradecimento especial à minha esposa e à minha filha pela paciência demonstrada nos momentos que não pude estar presente.

Ao meu orientador, Professora Cidália Fonte, pelas sugestões, disponibilidade e partilha de conhecimentos. O seu espírito crítico contribuiu significativamente para melhorar a qualidade deste trabalho.

A todos os professores do mestrado em Engenharia Geográfica, pelos conhecimentos transmitidos e estímulo ao interesse pelo conhecimento científico.

Às colaboradoras da empresa ATEP, diretora pedagógica Eunice Lopes e coordenadora Catarina Silva, pelo interesse em implementar um projeto desta natureza e disponibilidade em fornecer todos os dados necessários.

Nº do aluno: 2003012972

Nome: Luís Miguel Conceição Silva dos Reis

Título da dissertação:

"Construção de um SIG para Apoio à Decisão numa Empresa de Formação"

Palavras-Chave:

- SIG
- sistema de apoio à decisão
- código aberto
- geocodificação
- análise espacial
- cartografia

Resumo

A crescente utilização dos sistemas de informação geográfica (SIG) tem originado a sua aplicação numa gama de setores da atividade económica cada vez mais abrangente. Muitos problemas podem ser resolvidos através da aplicação de análises SIG, nomeadamente, quando possuem uma componente geográfica. Este tipo de problemas dificilmente pode ser resolvido por outro tipo de sistema de informação, além disso, os SIG possuem características peculiares na visualização e manipulação da informação geográfica.

No presente trabalho reúnem-se meios e definem-se metodologias para a implementação de um projeto SIG numa empresa de formação, tais como, escolha, tratamento e organização de dados, escolha do *software*, desenvolvimento de rotinas e construção do projeto. Este procura dar resposta a vários problemas que surgem aquando o planeamento das atividades formativas na NUT III Médio Tejo, principalmente, na localização de locais para desenvolver ações de formação. A otimização da sua localização é feita sobre a rede de estradas em função da localização dos formandos inscritos na empresa. Numa segunda instância o SIG produz análises espaciais para uma organização eficiente da atividade formativa nas instalações da empresa, a seleção de formandos para ações de formação é feita de acordo com o custo de deslocação avaliado em função do tempo de viagem até à empresa.

Do ponto de vista estratégico o SIG ajudará o decisor a definir uma linha orientadora para a empresa. Permite a construção de cartografia temática baseada num indicador estatístico, calculado com informação estatística disponível no Instituto Nacional de Estatística, que permite identificar áreas com maior necessidade de formação (escolar). Para objetivos mais específicos, o utilizador pode criar mapas temáticos assentes em variáveis estatísticas dos Censos 2011.

Todo o projeto é desenvolvido sem custos adicionais para a empresa, fazendo uso de *software* de código aberto e dados gratuitos. As aplicações de *software* são personalizadas para que os colaboradores da empresa (não especialistas em SIG) possam realizar as análises SIG de forma autónoma. O SIG pode ser implementado e utilizado corretamente através da consulta de um manual de utilização.

Abstract

The growing use of geographic information systems (GIS) has led to its application in a range of sectors of economic activity increasingly comprehensive. Many problems can be solved through the application of GIS analysis, especially when they have a geographic component. This type of problem can hardly be solved by another type of information system, in addition, GIS have peculiar characteristics in visualization and manipulation of geographic information.

In this paper we define means and methodologies for the implementation of a GIS project in a training company, such as, choice, processing and organizing data, choosing the software, developing routines and construction of the project. This seeks to give the answer to several problems that arise during the planning of training activities in the NUT III Médio Tejo, mainly in locating sites to develop training activities. The optimizing of its location is made on the road network in function to the location of students enrolled in the company. In a second instance the GIS produces spatial analysis for an efficient organization of the training activity on company premises, the selection of students for training activities is done according to the cost of travel evaluated on travel time to the company.

From the strategic point of view the GIS will help the decision maker to set a guideline for the company. Allows the construction of thematic mapping based on a statistical indicator, calculated with statistical information available on the National Statistics Institute, which allows us to identify areas most in need of training (academic). For more specific objectives, the user can create thematic maps based on statistical variables of the 2011 Census.

The entire project is developed without additional costs for the company, making use of open source software and free data. Software applications that are customized to the company's employees (not GIS specialists) can perform GIS analysis independently. GIS can be implemented and used properly by consulting a user manual.

Índice

Agradecimentos.....	ii
Resumo.....	iv
Abstract.....	v
Índice.....	vi
Índice de Figuras.....	viii
Índice de Tabelas.....	x
Simbologia.....	xi
Siglas.....	xii
1. INTRODUÇÃO.....	1
1.1. Enquadramento e motivação.....	1
1.2. Objetivos.....	2
1.2.1. Objetivos gerais.....	2
1.2.2. Objetivos específicos.....	2
1.3. Metodologia geral.....	3
1.4. Organização da dissertação.....	5
1.5. Contribuições previstas.....	6
2. OS SIG NO APOIO À DECISÃO EMPRESARIAL.....	7
2.1. Introdução aos SIG.....	7
2.2. Bases de dados geográficas e realização de consultas.....	11
2.3. Mapas temáticos e Estatística.....	15
2.3.1. Determinação de um indicador estatístico de privação múltipla.....	18
2.4. Geocodificação.....	21
2.5. Otimização da localização em redes.....	25
2.5.1. Problemas de localização-alocação.....	27
2.5.2. Áreas de serviço.....	28
2.6. Os SIG no apoio à decisão.....	30
2.6.1. Processos de decisão e informação geográfica em contexto empresarial.....	33
2.7. Fases de implementação de um projeto SIG.....	34
3. IMPLEMENTAÇÃO DO PROJETO SIG-ATEP.....	37
3.1. Enquadramento empresarial.....	37
3.1.1. Atividade da empresa.....	37

3.1.2 Recursos humanos envolvidos	38
3.2. Identificação dos requisitos	39
3.3. Escolha do <i>software</i>	40
3.4. Dados.....	45
3.4.1. Fontes de dados.....	45
3.4.2. Tratamento dos dados	48
3.4.3. Esquema da Base de Dados.....	55
3.5. Desenho do SIG-ATEP	57
3.6. Geocodificação de formandos	60
3.7. Identificação dos locais para realizar formação	62
3.8. Definição de rede iso com custo em tempo de viagem.....	65
3.9. Cálculo de indicador estatístico de necessidade de formação (INF)	65
3.10. Construção de mapas temáticos	70
4. DISCUSSÃO DOS RESULTADOS OBTIDOS	72
4.1. Avaliação dos benefícios no planeamento da atividade empresarial	72
4.2. Avaliação dos benefícios na tomada de decisões estratégicas	75
5. CONCLUSÕES	78
6. BIBLIOGRAFIA	80
Anexo 1 – Dicionário de dados	88
Anexo 2 – Listagem de código usado em macros <i>VBA Microsoft Office Excel</i>	96
a) Macro “ <i>velocidade_max</i> ” usada na atribuição de velocidade máxima permitida à rede de estradas <i>OpenStreetMap</i>	96
b) Macro “ <i>geocoding</i> ” usada na organização dos dados para geocodificação	96
c) Macro “ <i>alocacao_formandos</i> ” usada na otimização da localização dos locais de formação.	96
d) Macro “ <i>categorias_INF</i> ” usada na atribuição de categorias ao Indicador de Necessidade de Formação (INF)	97
Anexo 3 – Manual de utilização.....	98

Índice de Figuras

Figura 1 – Área de trabalho do <i>Quantum GIS 1.8.0 “Lisboa”</i> com Carta Administrativa Oficial de Portugal (Dados: IGP, 2012).....	8
Figura 2 – Modelos de dados vetorial e matricial (Adaptado de ESRI - ArcGIS 10.0 Help, 2013) ..	9
Figura 3 – Excerto do mapa de Dr. John Snow (Adaptado de Jsi, 2012)	10
Figura 4 – Ambiente simplificado de um sistema de base de dados (Adaptado de Rigaux <i>et al.</i> , 2002)	12
Figura 5 – DER de base de dados geográfica (Adaptado de Worboys e Duckham, 2004)	13
Figura 6 – Análise espacial <i>buffer</i> (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)	14
Figura 7 – Análise espacial intersecção (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)	14
Figura 8 - Análise espacial união (Adaptado de ESRI - ArcGIS 10.0 Help, 2013).....	15
Figura 9 – Mapas no processo de análise (Adaptado de O’Sullivan e Unwin, 2003).....	15
Figura 10 – Área por concelho em Km ² para Portugal continental. Classificação por intervalos iguais, quebras naturais e desvios padrão (Dados: IGP, 2012)	17
Figura 11 – Diagramas de dispersão correspondentes a valores do coeficiente de correlação de <i>Pearson (r)</i> (Burt <i>et al.</i> , 2009).....	19
Figura 12 – Geocodificação baseada na rede de estradas, identificação do número de polícia (Golberg <i>et al.</i> , 2007).....	23
Figura 13 – Parâmetros adicionais do algoritmo de geocodificação (Golberg <i>et al.</i> , 2007).....	23
Figura 14 – Realidade do parcelamento ao longo da estrada (Golberg <i>et al.</i> , 2007).....	24
Figura 15 – Rede topológica com custo (distância) associado aos arcos.....	26
Figura 16 – Localização-alocação por métodos heurísticos (Fonte: ESRI - ArcGIS 10.0 Help, 2013)	28
Figura 17 – Definição de áreas de serviço para ambulâncias, com custo distância (Smith, 2013). ..	29
Figura 18 – Rede iso para ponto de oferta (a verde) com bandas de custo: 1000, 2000, 3000, >3000 metros, representadas pelas cores: cinzento, vermelho, castanho e verde, respetivamente. Resultados obtidos pela ferramenta <i>v.net.iso</i> do <i>software GRASS GIS</i> (Castells, 2013)	30
Figura 19 - Modelo concetual de um SIG usado para suporte à decisão, proposto por Mennecke (2000)	32
Figura 20 - Processo de implementação de um SIG (Adaptado de Somers, 2001)	36
Figura 21 – NUTIII Médio Tejo (Fonte: Comunidade Intermunicipal Médio Tejo, 2013)	38
Figura 22 – Ferramenta para geocodificação múltipla (<i>plugin mmqgis</i>).....	41
Figura 23 –Ferramenta <i>Sextante</i> localização-alocação implementado no <i>Software SIG</i> de código aberto <i>Kosmo</i>	42
Figura 24 – Alocação de áreas de serviço a pontos de oferta. Ferramenta <i>v.net.alloc</i> do <i>GRASS</i> : pontos de oferta representados por triângulos verdes e áreas de serviço (vermelho, verde e amarelo), (GRASS MANUAL, 2013)	42
Figura 25 – Janela de personalização do QGIS.....	44
Figura 26 – Tabela de atributos da rede de estradas <i>OpenStreetMap</i> (OpenStreetMap, 2013)	46
Figura 27 – Ficheiro <i>TXT</i> com dados sobre a localização dos locais de formação	46
Figura 28 – Tabela de atributos da <i>shapefile</i> Censos 2011 (INE, 2013)	47
Figura 29 – Tabela de atributos da <i>shapefile</i> CAOP 2012 (DGT, 2012)	47
Figura 30 – Tabela <i>CSV</i> com dados estatísticos dos Censos 2011 (INE, 2013).....	48
Figura 31 – Rede de estradas usada no projeto (Dados: OpenStreetMaps, 2013; CAOP, 2012) ..	49

Figura 32 – Quebra manual (traço a vermelho) das intersecções com vias rápidas (Dados: OpenStreetMaps, 2013)	50
Figura 33 – Calculadora de campo QGIS	52
Figura 34 – Tabela de atributos da <i>shapefile</i> com locais de formação	53
Figura 35 – Esquema da base de dados SIG-ATEP.....	56
Figura 36 – DER do dos ficheiros da pasta “ATEP”	57
Figura 37 – Estrutura de funcionamento do SIG-ATEP	58
Figura 38 – Barra de ferramentas ATEP construída através do <i>plugin</i> QGIS <i>Custom Launcher</i>	59
Figura 39 – Interface gráfica do QGIS personalizada para o SIG-ATEP.....	59
Figura 40 – Interface gráfica da macro <i>Excel</i> “geocoding” para apoio à geocodificação de formandos.....	60
Figura 41 – Tabela de atributos da <i>shapefile</i> resultante da geocodificação.....	61
Figura 42 – Estrutura de processos para otimizar a localização dos locais de formação.....	63
Figura 43 - Interface gráfica da macro <i>Excel</i> “alocacao_formandos” para apoio à otimização da localização dos locais de formação.....	64
Figura 44 – Mapa temático com necessidade de formação e rede iso. Dados: CAOP 2012 com freguesias da NUT III Médio Tejo e rede de estradas <i>OpenStreetMap</i>	70
Figura 45 - Resultado do <i>HTML Image Map Plugin</i> com visualização de um mapa e atributo da camada formandos (pontos a vermelho). A rede de estradas encontra-se dividida em duas áreas de serviço: linhas a azul e a amarelo. Dados: <i>OpenStreetMap</i> , <i>Google Maps</i>	71
Figura 46 – Análise de redes para localização ótima de locais de formação. Dados: <i>OpenStreetMap</i>	73
Figura 47 - Análise de redes para definição de rede iso com sobreposição de formandos. Dados: <i>OpenStreetMap</i>	74
Figura 48 – Indicador estatístico de necessidade de formação (INF) por freguesia. Dados: CAOP 2012, Censos 2011	75
Figura 49 – Número de desempregados por subsecção estatística e rede iso. Dados: CAOP 2012, Censos 2011, <i>OpenStreetMap</i>	76
Figura 50 - Número de desempregados por subsecção estatística e <i>Google Streets</i> . Dados: CAOP 2012, Censos 2011, <i>Google Maps</i>	77

Índice de Tabelas

Tabela 1 – Escala de avaliação qualitativa para coeficiente de correlação de Pearson (Adaptado de Hopkins, 2002)	20
Tabela 2 – Distribuição dos <i>Carstairs Scores</i> de privação por sectores de código postal escoceses, baseados nos Censos 2001 (adaptado de McLoone, 2004)	21
Tabela 3 – Caminhos possíveis para os pontos de oferta representados pelos nós 3 e 6. Custo de deslocação avaliado em quilómetros	26
Tabela 4 – Organização do ficheiro <i>TXT</i> com informação sobre códigos postais (Dados: CTT, 2013)	45
Tabela 5 – Velocidades máximas permitidas de acordo com o tipo de via (Dados de velocidade: PRP, 2013 e Wikipédia, 2013)	51
Tabela 6 – Dados estatísticos inseridos nas <i>shapefiles</i> (concelhos, freguesias, subsecção estatística). (Dados: INE, 2013)	54
Tabela 7 – Correlações entre variáveis para estatísticas dos concelhos	67
Tabela 8 – Correlações entre variáveis para estatísticas das freguesias	67
Tabela 9 – Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para os concelhos da NUT III Médio Tejo (resultados baseados nos Censos 2011)	68
Tabela 10 - Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para algumas freguesias da NUT III Médio Tejo (resultados baseados nos Censos 2011)	68
Tabela 11 - Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para algumas subsecções estatísticas da NUT III Médio Tejo (resultados baseados nos Censos 2011)	69
Tabela 12 – Tabela com classes de valores INF e escalas qualitativas para o INF, para: concelhos, freguesias e subsecção estatística	69

Simbologia

r – coeficiente de correlação de *Pearson*

Z-scores – método para calcular indicador de privação múltipla

z – grandeza atribuída a cada valor de uma variável. Indica a posição do valor relativamente à média da variável

z_i – z para cada valor da variável escolhida na construção do indicador de privação múltipla

z_i - valor da variável

X - média da variável

s – desvio-padrão da variável

carstairs score – valor do indicador de privação múltipla para cada caso numa série de dados

q - *offset* adicionado no início e fim da rua no processo de geocodificação

d - distância medida do eixo da via até à entidade geocodificada

θ – ângulo entre o segmento de reta correspondente a d e o eixo da via

G - rede topológica

N - nós numa rede topológica

A - arcos numa rede topológica

Δs - distância percorrida

Δt - tempo de viagem

v – velocidade

R25_64 - percentagem de indivíduos residentes com idades entre 25 a 64 anos

R3 - percentagem de indivíduos residentes com 3º ciclo do ensino básico

RF1D - percentagem de famílias clássicas com um desempregado

z_{R25_64} - valor de z para a variável **R25_64**

z_{R3} - valor de z para a variável **R3**

z_{F1D} - valor de z para a variável **F1D**

Siglas

ANQEP – Agência Nacional para a Qualificação e Ensino Profissional

API – *application programming interface*

ATEP – Associação Torrejana de Ensino Profissional

CAOP – Carta Administrativa Oficial de Portugal

CNO – Centro de Novas Oportunidades

DER – diagrama entidade relação

DGT – Direção Geral do Território

DMS – *degrees, minutes, seconds*

EPSG – *European Petroleum Survey Group*

EPTN – Escola Profissional de Torres Novas

ESRI – *Environmental Systems Research Institute*

GADS – *Geodata Analysis and Display System*

INE – Instituto Nacional de Estatística

INF – indicador de necessidade de formação

NUT III - Nomenclatura de Unidade Territorial de nível três

SEAD – sistemas espaciais de apoio à decisão

SGBD – sistema de gestão de bases de dados

SIG – sistema de informação geográfica

SIG-ATEP – sistema de informação geográfica da empresa ATEP

SIGO – sistema de informação e gestão da oferta educativa e formativa

SSD – sistema de suporte à decisão

VBA - *Visual Basic for Applications*

1. INTRODUÇÃO

1.1. Enquadramento e motivação

A utilização de sistemas de informação geográfica (SIG) em meio empresarial apresenta-se uma mais-valia. O decisor/gestor tem a vantagem de visualizar o seu território de atuação e de resolver vários problemas de índole espacial, proporcionando uma vantagem competitiva à sua empresa.

A maior parte das soluções existentes no mercado é dispendiosa para pequenas empresas e possui uma estrutura de implementação complexa, na medida em que é necessário conhecimento especializado. Perante esta realidade surgiu a oportunidade de implementar um projeto SIG na empresa de formação, Associação Torrejana de Ensino Profissional (ATEP), onde questões relacionadas com a localização das suas atividades têm uma componente espacial. Desta forma, a utilização de ferramentas SIG pode ser uma mais-valia. Como se trata de uma empresa de pequena dimensão, onde não há lugar para custos adicionais, optou-se por usar uma solução baseada em *software* e dados de acesso livre.

As formações promovidas pela ATEP são aprovadas pelo Ministério da Educação, sendo que esta entidade define o tipo de ações de formação que devem ser desenvolvidas pela empresa. A gestão destas formações é feita pela ATEP, nomeadamente, a abertura de grupos de formação, horário de funcionamento e localização das ações de formação.

Cabe à ATEP orientar os interessados para formação adequada ao seu perfil (académico e profissional), definir o número mínimo de inscritos para abertura de grupos de formação, procurar localizações favoráveis para ministrar formação, de acordo com a localização dos formandos inscritos, e promover as atividades da empresa em zonas com maior necessidade de formação, como por exemplo, freguesias onde a maioria dos indivíduos possui baixas qualificações académicas.

Visto que a empresa tem sobre a sua responsabilidade um vasto território (NUT III Médio Tejo), quando é necessário abrir novos locais de formação surge o problema da sua localização: será que é a melhor tendo em consideração a proveniência de todos os interessados? Quantos locais de

formação devem ser abertos? Que pessoas devem ser contactadas para frequentar determinado local de formação? Estas questões só podem ser respondidas convenientemente através da execução de operações de análise em SIG, pois a sua resolução envolve a execução de operações com uma componente espacial.

Por outro lado, a empresa precisa de ter orientações sobre o modo de atuação no mercado, nomeadamente, no tipo de formações a promover e para quem. Neste aspeto, os SIG também podem ajudar, através do uso de cartografia temática que identifique zonas preferenciais ao desenvolvimento de atividades de formação. Estas zonas podem ser identificadas através da utilização de mapas temáticos baseados variáveis estatísticas demográficas e sociais. Por exemplo, caso se pretenda promover formação direcionada a desempregados, podem-se identificar zonas prioritárias usando um mapa temático com o número de desempregados por freguesia.

1.2. Objetivos

1.2.1. Objetivos gerais

Face à pequena dimensão da empresa, pretende-se que a implementação do projeto assente em *software* de código aberto e fontes de dados gratuitas.

Este servirá para apoiar a decisão em várias situações: localização dos locais para ministrar formação, minimizando os custos de deslocação dos formandos e formadores e garantindo a cobertura de toda a NUT III Médio Tejo, assim como, escolha de locais para promover formação no sentido de abrir novos grupos de formação

O projeto será materializado por uma aplicação SIG *desktop*, personalizada de forma que possa ser utilizada por não-especialistas em SIG na visualização e manipulação de dados geográficos. A sua utilização será apoiada por um manual de utilização (Anexo 3).

1.2.2. Objetivos específicos

A construção do SIG com os objetivos acima indicados requer:

- Georreferenciação dos formandos registados na empresa através do processo de geocodificação. Este processo permitirá transformar os endereços postais, correspondentes aos formandos, em pontos num mapa;

- Obtenção de cartografia com a localização otimizada dos locais de formação, assim como, as suas áreas de influência. Esta cartografia irá basear-se na rede de estradas considerando o custo de deslocação em tempo de viagem;
- Obtenção de cartografia com a rede de estradas dividida em bandas de custo (medido em tempo de viagem) até à empresa de formação (ex.: 15, 30, 45 e >45 minutos). Esta irá permitir selecionar os formandos georreferenciados de acordo com o tempo de viagem, por exemplo: formandos localizados a menos de 15 minutos de viagem da ATEP;
- Criação de um indicador estatístico que permita avaliar a necessidade de formação (no âmbito da qualificação académica) ao nível do concelho, freguesia e subsecção estatística. Com este indicador irão ser construídos mapas temáticos que permitem identificar zonas de acordo com a necessidade de formação;
- Obtenção de cartografia temática baseada em variáveis estatísticas dos Censos 2011. Estas variáveis estão relacionadas com temas sociais/demográficos e serão disponibilizadas por concelho, freguesia e subsecção estatística;
- Através da análise da cartografia o utilizador deve conseguir identificar zonas prioritárias para a oferta de formação e relacioná-las com a proximidade à empresa. Ex.: “Quais as áreas com maior necessidade de formação a menos de 30 minutos da empresa?”. Como os formadores partem sempre da ATEP é preferível desenvolver formações próximo da mesma;
- Personalização da aplicação SIG de forma a torná-la intuitiva e de fácil utilização para funcionários da empresa sem formação na área dos SIG;
- Construção de um manual de utilização que permita aos utilizadores trabalharem autonomamente com o SIG;

1.3. Metodologia geral

Foi efetuada pesquisa bibliográfica sobre a utilização dos SIG em meio empresarial, no sentido de identificar e compreender os métodos de análise necessários ao cumprimento dos objetivos. Também foi realizada revisão bibliográfica sobre fundamentos básicos dos SIG, essenciais à compreensão dos processos de análise envolvidos no presente trabalho.

Para conseguir implementar o projeto SIG (denominado de SIG-ATEP) utilizando *software* de código aberto, fez-se pesquisa de forma a encontrar um *software* que possuísse as ferramentas necessárias para o cumprimento dos objetivos. As capacidades do *software* mais importantes para o projeto estão relacionadas com a georreferenciação dos formandos e análise de redes.

A georreferenciação de formandos é realizada através de ferramentas de geocodificação que transformam os dados dos formandos, particularmente endereços postais, num ficheiro contendo entidades geográficas, neste caso pontos. Os endereços postais provêm de registos existentes na empresa, sobre formandos que se encontram inscritos em ações de formação ou que realizaram formação na empresa.

A análise de redes está relacionada com a otimização da localização dos locais de formação usando a rede de estradas. Neste âmbito são usadas ferramentas SIG que permitem otimizar a localização de pontos na rede tendo em conta o custo de deslocação na rede, para este projeto o custo é determinado pelo tempo de viagem. Com o método adotado a localização dos locais de formação é efetuada em função da localização dos formandos, procurando minimizar o tempo de viagem dos formandos.

Usando ainda ferramentas de análise de redes é possível selecionar os formandos de acordo com o seu tempo de viagem à ATEP, através da divisão da rede de estradas em bandas de custo. A rede resultante desta operação é dividida em bandas de 15, 30, 45 e >45 minutos de viagem até à ATEP. O planeamento das atividades formativas, a desenvolver nas instalações da empresa, motivou a adoção deste tipo de análise.

A escolha dos dados para o projeto foi efetuada procurando o acesso livre e sua qualidade. Sempre que possível opta-se por dados fornecidos por entidades oficiais como a Direção Geral do Território (DGT) ou o Instituto Nacional de Estatística (INE). Quando os dados são provenientes de projetos colaborativos, onde os dados são inseridos e editados por voluntários, é efetuado tratamento e reorganização dos dados de modo a garantir a qualidade necessária.

Os dados estatísticos provenientes dos Censos 2011 servem de base para a construção de cartografia temática, sendo que as variáveis estatísticas a usar foram escolhidas em conjunto com os responsáveis pelo projeto educativo da ATEP. A construção do indicador estatístico de necessidade de formação baseia-se na utilização simultânea de algumas variáveis sociodemográficas. Este representa o nível de privação em formação (académica e profissional) de determinada população.

O manual de utilização foi construído no sentido de proporcionar ao utilizador do SIG autonomia na instalação do *software*, ferramentas SIG, integração dos dados no projeto e realização das análises SIG.

1.4. Organização da dissertação

A dissertação está estruturada em três secções: revisão bibliográfica sobre os SIG em contexto empresarial; descrição do trabalho desenvolvido; análise dos resultados obtidos.

Na **primeira secção** é feita uma revisão bibliográfica onde se abordam os temas relacionados com os SIG e suas capacidades, relevantes para o projeto, nomeadamente: principais características dos SIG, base de dados geográficas, realização de consultas de dados espaciais, mapas temáticos e representação de dados estatísticos, geocodificação, otimização da localização em redes, os SIG no apoio à decisão e seu enquadramento empresarial, fases de implementação de um projeto SIG.

Numa **segunda secção** desenvolve-se a implementação do SIG-ATEP, onde é feita uma análise do funcionamento e necessidades da ATEP e sua área de atuação, identificação dos requisitos para o SIG-ATEP, critérios usados para a escolha do *software*, dados utilizados (tratamento e organização), desenho do projeto e descrição de todos os procedimentos necessários para implementação do projeto. Neste último ponto são descritas metodologias para a geocodificação dos formandos, identificação de locais para realizar formação, definição da rede dividida por bandas de custo, cálculo de indicador estatístico de necessidade de formação e construção de mapas temáticos.

Na **terceira secção** discutem-se os resultados obtidos através da avaliação dos benefícios para a atividade da empresa, na perspetiva operacional e estratégica. São apresentadas as conclusões ao trabalho desenvolvido e sugestões para desenvolvimentos futuros.

Em anexo à dissertação encontram-se vários documentos: dicionário de dados onde consta descrição das tabelas e dos seus atributos; listagem do código utilizado no desenvolvimento do projeto; manual de utilização.

1.5. Contribuições previstas

O presente trabalho procura contribuir para uma maior eficácia do planeamento da oferta formativa da empresa e, ao mesmo tempo, reforçar a sua posição no mercado. Serve também para mostrar as potencialidades dos SIG nos negócios, principalmente, na forma como podem ser determinantes em atividades de gestão que tenham por base problemas espaciais.

Outro aspeto prende-se com a utilização de *software* de código aberto, espera-se demonstrar que empresas com menor dimensão podem também usufruir de todas as potencialidades dos SIG, sem custos adicionais com *software*. Ainda neste âmbito, procura-se desmistificar a dificuldade do uso de *software* de código aberto, principalmente por parte do utilizador/decisor.

Será possível um utilizador não-especialista realizar análises SIG complexas de forma intuitiva, sem necessidade de adquirir conhecimentos em análise de redes, sistemas de referência ou bases de dados. Neste aspeto, trata-se de uma “democratização” do uso dos SIG.

2. OS SIG NO APOIO À DECISÃO EMPRESARIAL

2.1. Introdução aos SIG

Um SIG possui um conjunto de ferramentas computacionais que permitem capturar, modelar, manipular, analisar e apresentar dados geograficamente referenciados (Worboys e Duckham, 2004; Choudhury *et al.*, 2009). Mais do que uma tecnologia que trabalha com informação geográfica é um sistema composto por, pessoas, *software*, *hardware*, dados e aplicações (Harmon e Anderson, 2003; Worboys e Duckham, 2004; Longley *et al.*, 2011).

A sua definição depende do contexto em que é utilizado, pode ser entendido como um sistema de apoio à decisão, uma ferramenta computadorizada para resolver problemas geográficos, uma ferramenta para automatizar operações sobre dados geográficos ou um simples repositório de mapas digitais (Longley *et al.*, 2011).

Os SIG estão associados a diferentes atividades utilizadoras de informação geográfica, é usado por organizações governamentais no planeamento territorial, empresas de logística no planeamento de rotas, empresas florestais na gestão das suas plantações, consultores de geomarketing na resolução de problemas de localização de novas lojas, turistas que consultam sites com serviços *WebMap* (serviço de mapas na Internet) no sentido de planearem visitas e caminhadas, etc (Choudhury *et al.*, 2009; Longley *et al.*, 2011).

Podemos considerar um SIG como um sistema digital capaz de representar dados geográficos (ex.: mapa da rede de estradas) e de lhe associar informação alfanumérica (atributos) através de uma tabela (tabela de atributos). Estes atributos podem ser qualitativos (ex.: classificação do uso do solo) ou quantitativos (ex.: densidade populacional). A informação geográfica, também referenciada como informação espacial, tem uma componente de localização, representada pela latitude e longitude. O SIG fornece um mapa digital dinâmico com toda a informação relevante para um projeto (Choudhury *et al.*, 2009). A Figura 1 apresenta a interface gráfica do *software* SIG *Quantum GIS 1.8.0 "Lisboa"*, com a Carta Administrativa Oficial de Portugal (CAOP), (por freguesias) e sua tabela de atributos. Na mesma visualização podemos ter vários níveis de informação (ou camadas) representado vários mapas digitais. Se a informação geográfica for

referente à mesma área pode existir sobreposição (ex.: Carta Administrativa Oficial de Portugal e estradas de Portugal).

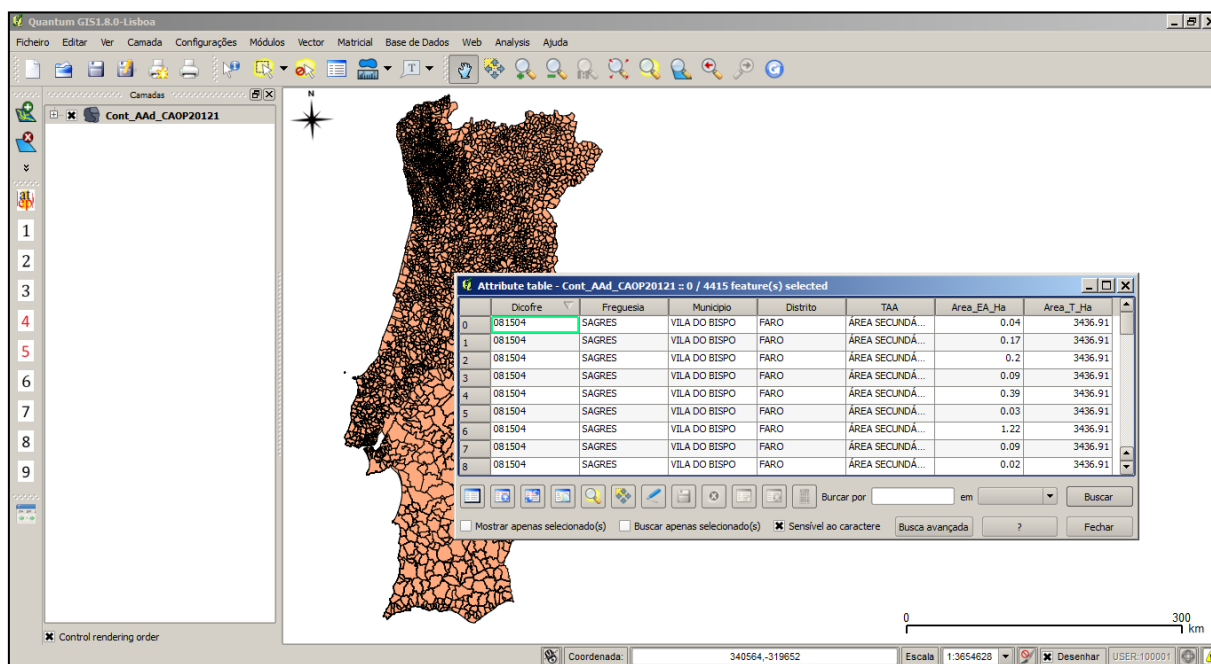


Figura 1 – Área de trabalho do *Quantum GIS 1.8.0 “Lisboa”* com Carta Administrativa Oficial de Portugal (Dados: IGP, 2012)

Existem dois métodos para representar os fenómenos geográficos num SIG, esses métodos denominam-se: vetorial e matricial (Burrough e McDonnel, 2000, O’Sullivan e Unwin, 2003; Longley *et al.*, 2011).

Vetorial: normalmente a representação dos fenómenos geográficos é efetuada através de objetos (entidades geográficas) pelo registo das suas coordenadas. As entidades podem ser pontos, linhas ou áreas (normalmente representadas através de polígonos). Tanto as linhas como os polígonos são representadas através de pontos ligados por segmentos de reta, o detalhe é maior quanto maior a densidade de pontos, pois estes representam os vértices destes dois tipos de geometria. Como o polígono representa uma área, as coordenadas do último vértice tem de coincidir com as do primeiro de maneira a construir uma figura fechada. Usando este método existem espaços vazios por entre as entidades geográficas (Figura 2a).

Matricial: representa o espaço geográfico através de uma grelha de células, designadas por píxeis. Cada píxel contém um valor representando a o valor do fenómeno geográfico no local ocupado pelo píxel (ex.: tipo de ocupação do solo). Com este método todo o espaço geográfico é representado através de píxeis de tamanho igual, normalmente retangulares ou quadrados. Como todos os píxeis assumem um valor a memória ocupada no computador é superior ao método vetorial (Figura 2b). A forma de dados matricial pode provir de satélites de deteção remota, como

Landsat Thematic Mapper. Neste formato a resolução (nível de detalhe) é maior quanto menor for o tamanho do píxel, tal como numa fotografia digital.

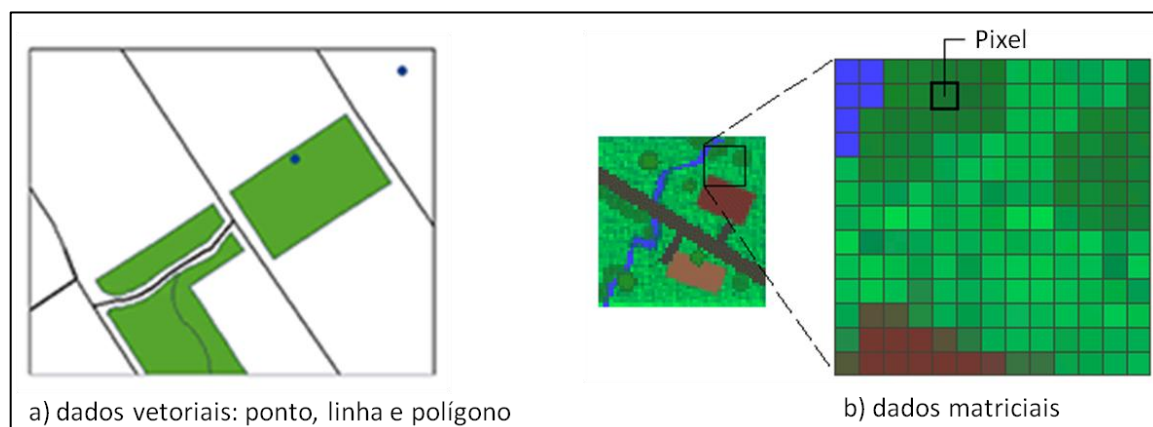


Figura 2 – Modelos de dados vetorial e matricial (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)

Os SIG permitem representar o mundo real de uma forma simplificada. A percepção do mundo real depende dos seus observadores, como tal, os SIG representam o mundo de acordo com as necessidades dos utilizadores. A juntar a esta característica existe a enorme complexidade de representação do mundo real, pois pode ser descrito ao nível de pequenas partículas até ao nível dos oceanos e continentes (Choudhury *et al.*, 2009).

A modelação do mundo real através dos SIG assume-se como uma relação entre o detalhe com que pretendemos representar os fenómenos (escala geográfica), o tipo de entidades geográficas e seus atributos. Um exemplo apresentado por O'Sullivan e Unwin (2003) mostra como esta relação é importante. Uma estação de comboios pode ser representada a determinada escala com um ponto, a uma escala maior, como um polígono e a uma escala muito maior como um conjunto de linhas, representando as ferrovias.

A representação das entidades geográficas deve ser coerente com os objetivos do projeto. Estas devem ser, identificáveis, relevantes e facilmente descritas através dos seus atributos (O'Sullivan e Unwin, 2003).

A escala usada nos valores dos atributos pode ser de vários tipos: **nominal**, que serve para identificar ou distinguir uma entidade de outra (ex.: nomes de lugares), é um atributo qualitativo; **ordinal**, para valores que possuem determinada ordem (ex.: médio, alto, muito alto), podem ser expressos em valores numéricos, códigos ou nomes; **intervalar**, usa valores numéricos podendo-se quantificar as diferenças entre valores, no entanto a escala não tem um zero absoluto (ex.: a escala de temperatura, 0 graus celsius não representam ausência de temperatura); **razão**, pode-se

acontece o fenómeno, comparar diferentes propriedades do mesmo lugar, descobrir relações, estabelecer correlações e procurar explicações. A segunda baseia-se no conceito que a separação das entidades geográficas pode ter um significado útil, isto é, pode dizer algo sobre os mecanismos responsáveis pela sua localização ou propriedades (Longley *et al.*, 2011).

2.2. Bases de dados geográficas e realização de consultas

Uma base de dados geográfica pode ser considerada como um conjunto de dados integrados sobre um assunto em particular, tendo os dados uma componente espacial (Longley *et al.*, 2011).

Uma base de dados geográfica é parte crítica de um SIG devido ao custo da sua criação, manutenção e ao impacto que tem sobre os resultados das operações de análise e consequentemente a tomada de decisões baseadas nesses resultados (Longley *et al.*, 2011). A sua criação torna-se numa tarefa complexa porque envolve recolha de dados, bem como a sua verificação e estruturação. A informação pode estar disponível em diferentes formatos, digitais ou analógicos, tais como: fotografias aéreas, imagens de satélite ou tabelas (Burrough e McDonnel, 2000, Rigaux *et al.*, 2002). O armazenamento digital dos dados geográficos caracteriza-se pelo seu grande volume (Rigaux *et al.*, 2002; Longley *et al.*, 2011), por exemplo, uma imagem global do satélite *Landsat*, com resolução de 15 metros, ocupa 6,5 Tera bytes (Longley *et al.*, 2011).

Para manipular estas bases de dados é necessário um sistema que permita: inserção de dados e verificação; armazenamento e manipulação dos dados; exportação e apresentação de dados; transformação de dados e interação com os utilizadores, isto é, um sistema de gestão de bases de dados (SGBD), (Rigaux *et al.*, 2002). Estas características são úteis para os SIG, daí que as maiores bases de dados geográficas sejam geridas por um SGBD. A maior parte dos *softwares* SIG possuem tecnologia SGBD ou possuem interfaces para a integrar (Longley *et al.*, 2011).

Genericamente um SGBD facilita os seguintes processos: definição da base de dados (tipos de dados, estruturas e restrições a ter em conta); construção da base de dados (armazenamento dos dados); manipulação da base de dados; realização de consultas de forma a obter dados específicos; atualização da base de dados (Rigaux *et al.*, 2002; Longley *et al.*, 2011). As operações de consulta são entendidas como uma inquirições à base de dados sobre determinados critérios, por exemplo, numa base de dados com informação demográfica sobre cidades pode-se procurar a tabela com informação do número de habitantes e questionar: “Quais as cidades com mais de 10 000 habitantes?”.

A Figura 4 representa o ambiente simplificado de um sistema de base de dados. Pode-se ver que o SGBD funciona como mediador entre os utilizadores ou aplicações de programas e os

dispositivos de armazenamento de dados. Dentro do SGBD o *software* deve processar consultas e o acesso aos dados. A par dos dados armazenados existem os metadados, que permitem compreender os dados e a estrutura da base de dados, pela descrição dos dados (ex.: significado, origem, qualidade, história, etc.) e suas relações (Longley *et al.*, 2011).

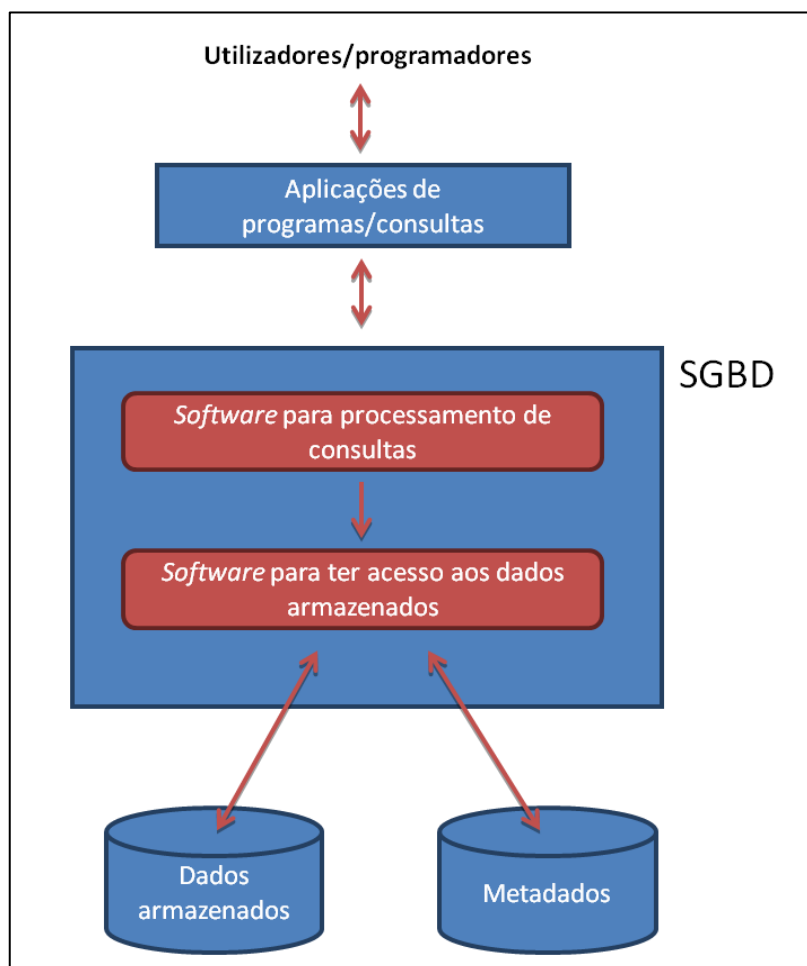


Figura 4 – Ambiente simplificado de um sistema de base de dados (Adaptado de Rigaux *et al.*, 2002)

Num SGBD a informação geográfica é armazenada segundo classes de entidades que representam determinado tema (ex.: rios, cidades, distritos, etc.). As classes são armazenadas em tabelas com várias linhas e colunas, sendo as linhas usadas para armazenar as várias entidades (ex.: várias cidades) e as colunas os atributos da entidade (ex.: nome, densidade populacional, área, etc.), os dados que compõem uma linha são referidos como valores (Worboys e Duckham, 2004; Longley *et al.*, 2011). As tabelas compostas por informação geográfica distinguem-se das não-geográficas pela presença de colunas com a geometria das entidades (Longley *et al.*, 2011).

As tabelas relacionam-se entre si através de atributos em comum que permitem a junção entre tabelas. Este modelo relacional permite a junção de tabelas para criar novas visualizações da base

de dados, sobre a tabela resultante da junção podem-se executar operações como consultas ou apresentações de dados (Longley *et al.*, 2011).

A propriedade que permite classificar as relações entre tabelas denomina-se de cardinalidade. As possibilidades de relações são: *um para um* (1:1), *um para muitos* (1:N), *muitos para um* (N:1) e *muitos para muitos* (N:M), (Harmon e Anderson, 2003). Na Figura 5 representa um diagrama entidade relação (DER) que esquematiza as relações entre tabelas de uma base de dados geográfica. Existe uma relação de *muitos para muitos*, entre a tabela ESTRADA e a tabela CIDADE, isto é, uma cidade pode estar ligada a várias estradas e uma estrada pode estar ligada a várias cidades. Os círculos representam a possibilidade de a conexão ser nula (ex.: uma estrada pode não estar conectada com uma cidade) e o traço significa a obrigatoriedade da conexão (ex.: uma cidade tem de estar conectada pelo menos a uma estrada). A relação *um para muitos* (1:N) indica que uma cidade pode ter vários cinemas, mas um cinema só pode estar localizado numa cidade, a simbologia aparece ao contrário, N:1 em vez de 1:N porque indica o sentido da relação. Entre as tabelas CINEMA e GERENTE estabelece-se uma relação de *um para um* (1:1), pois um cinema só pode ter um gerente e este só pode gerir um cinema. A chave-primária das tabelas é assinalada com sublinhado.

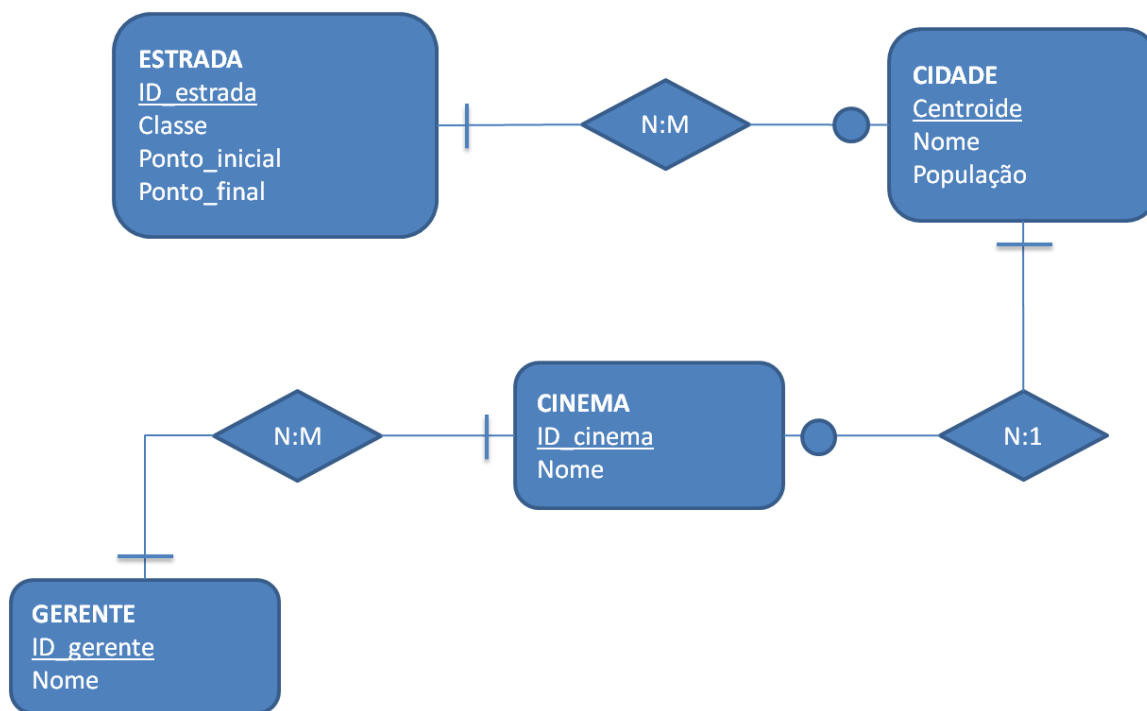


Figura 5 – DER de base de dados geográfica (Adaptado de Worboys e Duckham, 2004)

O DER da base de dados deve ser acompanhado de um dicionário de dados que descreve as tabelas: nome dos atributos, tipo (ex.: numérico), tamanho (número de caracteres) e descrição (Harmon e Anderson, 2003).

Sobre a base de dados podem-se realizar consultas de modo a encontrar dados segundo critérios definidos pelo utilizador. Estas podem incidir sobre uma tabela ou sobre várias tabelas (Longley *et al.*, 2011). Para a base de dados da Figura 5, caso se queira encontrar as cidades com mais de x população só é necessário a tabela CIDADE, no entanto, caso se pretenda saber qual ou quais gerentes trabalham na cidade x , precisa-se de usar três tabelas: CIDADE, GERENTE e CINEMA. Os resultados visualizados através de tabelas de atributos e como se está a trabalhar com entidades geográficas podem também ser visualizados num mapa.

A natureza geométrica dos dados geográficos permite estabelecer relações espaciais entre as entidades e suportar análises espaciais. Estas relações podem ser testadas através de condições sobre a relação entre figuras geométricas, como por exemplo: **igualdade** – se as figuras são iguais; **disjunção** – se as figuras partilham um ponto em comum; **intersecção** – se as figuras intersectam-se; **sobreposição** – se as figuras sobrepõem-se (devem ser da mesma dimensão); **inclusão** – se está uma figura contida em outra; etc (Longley *et al.*, 2011).

Podem também suportar análises espaciais, como os exemplos seguintes: **buffer** - devolve uma figura geométrica que representa todos os pontos cuja distância à figura geométrica de entrada é igual ou menor à definida pelo utilizador (Figura 6); **intersecção** - devolve uma figura geométrica que contém pontos comuns às duas figuras geométricas de entrada (Figura 7); **união** - devolve uma figura geométrica que contém todos os pontos das duas figuras geométricas de entrada (Figura 8), (Longley *et al.*, 2011).

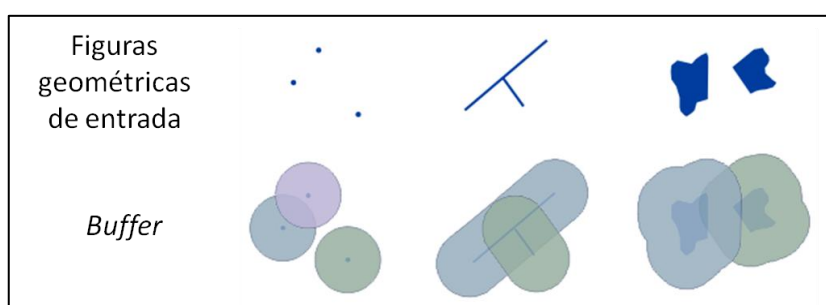


Figura 6 – Análise espacial *buffer* (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)

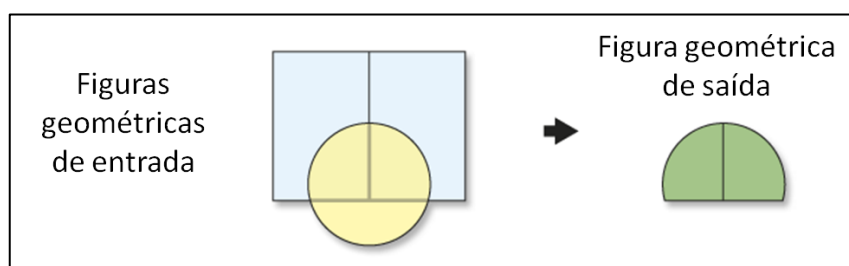


Figura 7 – Análise espacial intersecção (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)

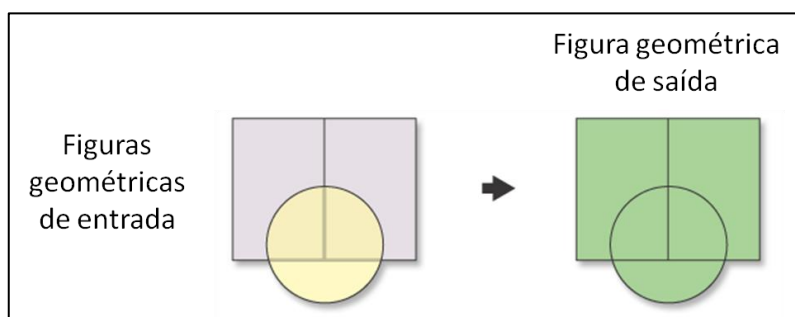


Figura 8 - Análise espacial união (Adaptado de ESRI - ArcGIS 10.0 Help, 2013)

Este tipo de análises permite obter informação baseada nas relações espaciais entre as entidades geográficas. Por exemplo: na base de dados da Figura 5, ao selecionar as cidades que intersectam a estrada x e fazer um *buffer* de x quilómetros a partir do centróide¹ de cada cidade, pode-se selecionar os cinemas contidos nos *buffer* criados e obter os cinemas localizados até x quilómetros de cada cidade.

2.3. Mapas temáticos e Estatística

A apresentação de mapas representa uma das funções mais básicas de um SIG (O'Sullivan e Unwin, 2003). Os mapas temáticos distinguem-se dos demais por serem construídos para transmitir informações sobre um único assunto ou tema (ex.: mapa com densidade populacional), (Sommer, 2006). São uma forma eficiente de sumarizar e comunicar os resultados do SIG e representam importantes ferramentas de apoio à decisão. A Figura 9, apresenta o processo de análise desenvolvido pelo decisor em torno da utilização dos mapas temáticos. Este pode construir mapas em função das questões que formulou e dos dados que possui, sendo que a análise dos mapas pode suscitar novas questões. O processo deverá permitir chegar a importantes conclusões (O'Sullivan e Unwin, 2003).

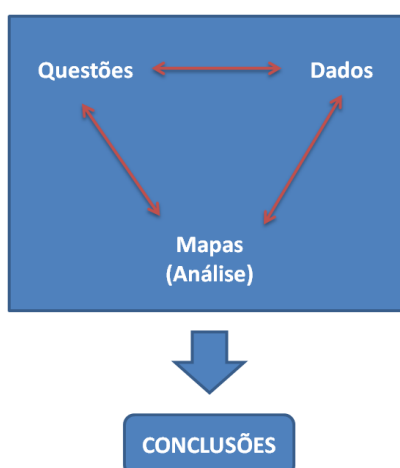


Figura 9 – Mapas no processo de análise (Adaptado de O'Sullivan e Unwin, 2003)

¹ Ponto que representa o centro massa da entidade geográfica (Sommer, 2006)

Os mapas podem apresentar: resultados de análises (ex.: localização ótima de uma nova loja); comunicar relações espaciais entre diferentes fenómenos, no mesmo mapa, em diferentes mapas ou para diferentes áreas (ex.: sobreposição de um mapa contendo focos de incêndios florestais com um mapa da ocupação do solo); permitem identificar padrões de distribuição espacial; entre outros. Para que um mapa funcione como suporte à decisão é necessário que a mensagem seja interpretada pelo decisor, não se limitando a transmitir informação conhecida acerca do mundo (Longley *et al.*, 2011).

Os dados estatísticos são um recurso para a construção de mapas temáticos, através desta relação é possível representar geograficamente diversas realidades, como por exemplo: preferências dos consumidores; intenções de voto eleitoral; indicadores de desenvolvimento social; etc (Longley *et al.*, 2011). Este tipo de dados é armazenado na tabela de atributos e a sua apresentação no mapa pode ser feita por vários métodos, de acordo com o tipo de elemento geométrico usado (ex.: linha, ponto, área). Esta apresentação deve respeitar regras de simbologia (Dente *et al.*, 2009).

O mapa coropleto é um tipo de mapa muito utilizado para representar dados estatísticos censitários (ex.: variáveis estatísticas dos Censos 2011). É composto por polígonos que podem representar áreas administrativas (ex.: freguesias), secções estatísticas, polígonos representativos de determinado fenómeno, etc. Cada área é preenchida com cor ou textura de forma a simbolizar o valor da variável. Para indicar a intensidade dos valores é definida uma escala de cores (ex.: castanho claro a castanho escuro), com os valores separados em classes (Dente *et al.*, 2009; Longley *et al.*, 2011).

A classificação dos dados deverá ser coerente com o fenómeno a representar. Assim, o esquema escolhido deverá respeitar o nível de generalização pretendido e a natureza dos dados. Pode-se estabelecer o número de classes e optar por vários métodos de classificação, produzindo amplitudes de classe diferentes (Dente *et al.*, 2009).

De entre vários métodos de classificação destacam-se: **intervalos iguais**; **quebras naturais** e **desvios-padrão** (Dente *et al.*, 2009; Longley *et al.*, 2011).

- **Intervalos iguais:** os valores são divididos em classes de igual amplitude, independentemente da distribuição dos dados;
- **Quebras naturais:** as classes são definidas de acordo com grupos naturais, isto é, determinados por quebras na série de dados. O método de cálculo procura reduzir a variância dentro das classes e maximizar a variância entre classes;

- **Desvios-padrão:** classificação de acordo com o desvio relativamente à média. São geradas classes de valores representando o número de desvios-padrão acima ou abaixo da média.

Se os valores tiverem uma ordem intrínseca, como a escala ordinal, basta atribuir uma escala de intensidade de cor a cada categoria, ou variação de cores (Longley *et al.*, 2011). Por exemplo, para a escala: Baixo (Verde), Médio (Amarelo), Alto (Vermelho).

Os *softwares* SIG possuem várias ferramentas para a construção deste tipo de mapas, nomeadamente: determinação automática de classes (usando vários métodos), atribuição de paletes de cores (ex.: intensidade da cor e variação de cores), personalização do número de classes, etc.

A Figura 10 mostra três mapas coropletos com os concelhos de Portugal continental classificados segundo sua área total. Constata-se que os diferentes métodos de classificação originam diferentes distribuições espaciais. Fazendo uma análise visual simples sobre os padrões de distribuição verificam-se diferenças significativas entre os métodos intervalos iguais e quebras naturais.

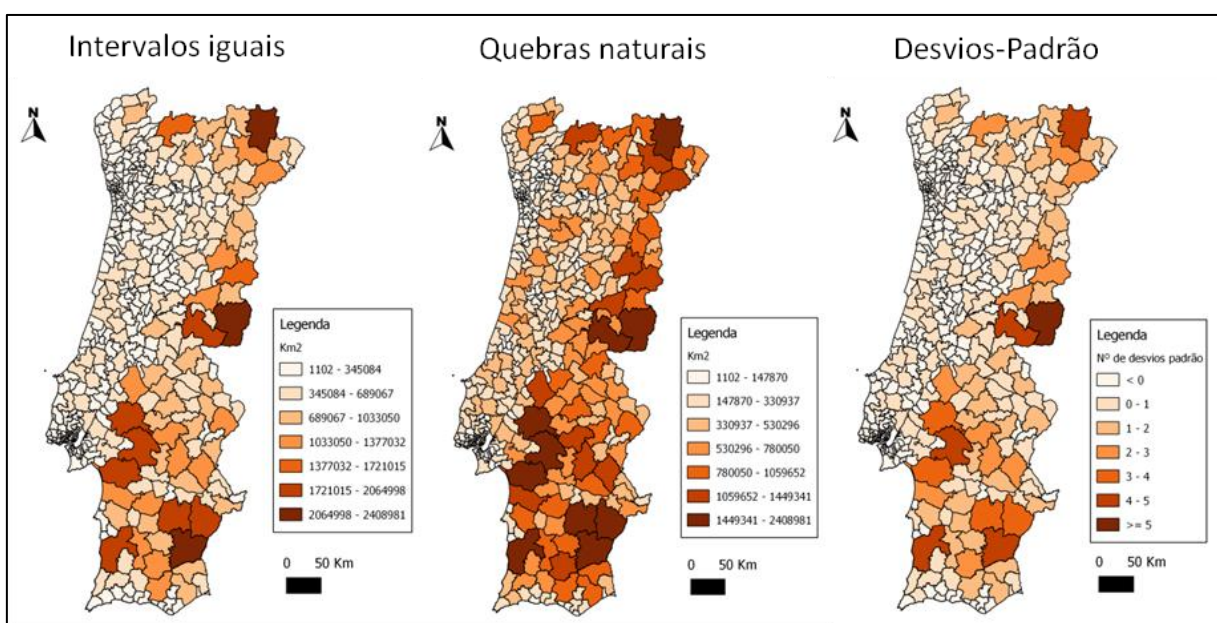


Figura 10 – Área por concelho em Km² para Portugal continental. Classificação por intervalos iguais, quebras naturais e desvios padrão (Dados: IGP, 2012)

2.3.1. Determinação de um indicador estatístico de privação múltipla

Vários estudos têm-se debruçado na aquisição de informação sobre privação múltipla (nível privação de uma população segundo vários indicadores: social, material, educacional, etc) e sua distribuição geográfica. São estudos que procuram avaliar a pobreza, exclusão social, saúde ou educação em determinada região ou país. Esta tarefa é feita através da construção de um indicador que reúne várias variáveis: económicas, sociais, materiais e demográficas (Social Disadvantage Research Centre, 2003).

Mcloone (2004) define um indicador de privação sociomaterial através de variáveis derivadas dos Censos 2001, no sentido de identificar zonas geográficas em privação material e social na Escócia, nomeadamente, variáveis reveladoras de baixo crescimento económico, desemprego masculino, classes sociais baixas e pessoas sem carro próprio (Mcloone, 2004).

Para poder usar as variáveis no indicador o autor teve que verificar o seu grau de associação, este foi quantificado através do coeficiente de correlação de *Pearson (r)*. Trata-se de uma medida estatística que mede o grau de relação linear entre um par de variáveis quantitativas. O seu valor varia entre -1 e 1, assumindo valores negativos para correlações negativas e positivos para correlações positivas, sendo que o valor absoluto mede a intensidade da associação linear. No extremo, o coeficiente assume o valor de 1 para correlações perfeitas positivas, -1 para correlações perfeitas negativas e 0 quando não existe correlação (Murteira *et al.*, 2007).

Este tipo de relações é facilmente visualizado através de diagramas de dispersão (Figura 11), onde os valores das variáveis são representados por pontos. Na correlação positiva os valores das variáveis evoluem no mesmo sentido (Figura 11a) e na correlação negativa os valores evoluem em sentido inverso (Figura 11b), neste último os valores de uma variável aumentam quando os valores da outra diminuem. Quando não existe correlação os valores não evoluem linearmente, em toda a série de valores (Figura 11c, d). Dentro dos limites referidos podem existir várias intensidades de associação (Figura 11e, f).

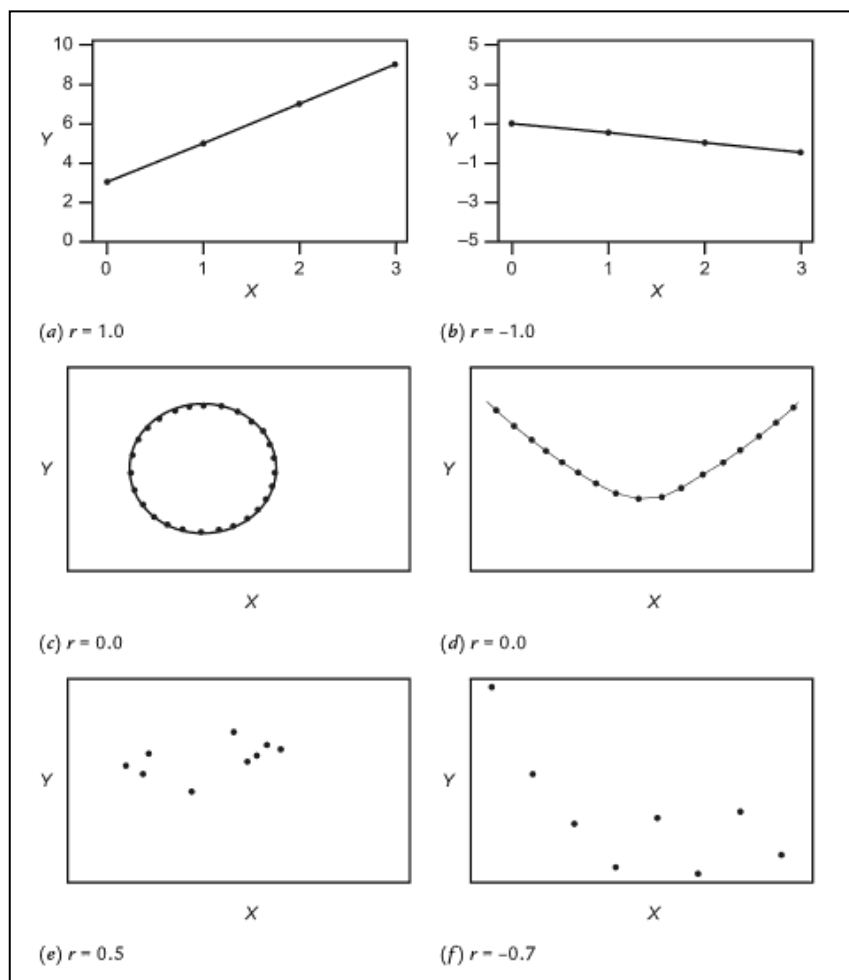


Figura 11 – Diagramas de dispersão correspondentes a valores do coeficiente de correlação de *Pearson* (r) (Burt *et al.*, 2009)

Este coeficiente não faz distinção entre variáveis dependentes e independentes, dificilmente se pode afirmar que uma variável varia em função de outra. Caso o valor de r seja significativo podemos dizer apenas que existem semelhanças entre a distribuição dos valores das variáveis (Filho e Júnior, 2009).

A interpretação do valor de r pode ser facilitada pela implementação de uma classificação qualitativa dos valores. Para ciências sociais pode-se adotar a escala descrita na Tabela 1 (Hopkins, 2002).

Tabela 1 – Escala de avaliação qualitativa para coeficiente de correlação de Pearson (Adaptado de Hopkins, 2002)

Coeficiente de correlação de <i>Pearson</i> (<i>r</i>)	Classificação qualitativa
0	Nula
< 0.1	Insignificante
0.1 – 0.3	Fraca
0.3 – 0.5	Moderada
0.5 – 0.7	Forte
0.7 – 0.9	Muito Forte
> 0.9	Quase Perfeita
1	Perfeita

A verificação de associação entre as variáveis permite determinar a sua validade para entrar na construção do indicador estatístico. Consideram-se resultados aceitáveis a partir de correlações moderadas.

McLoone (2004) utiliza o método dos *Z-Scores* para calcular o indicador de privação. Este método permite determinar uma grandeza (*z*) que indica a posição de um valor relativamente à média do conjunto de dados. É expresso em desvios-padrão considerando-se que para $z = 0$ o valor é equivalente à média $z > 0$ é superior à média e $z < 0$ é inferior à média (O’Sullivan e Unwin, 2003).

De acordo com a fórmula de McLoone (2004) e O’Sullivan e Unwin (2003), *z* é calculado para cada valor de cada variável escolhida para o indicador, representado por *zi* na fórmula (2.1), *xi* representa o valor da variável, *X* a média da variável, *s* o seu desvio-padrão.

$$z_i = \frac{x_i - X}{s} \quad (2.1)$$

Aplicando a fórmula (2.1) ficamos com um valor *z* para cada variável (ex.: densidade populacional) e para cada caso (ex.: concelho). O valor do indicador é obtido fazendo a soma de todos os valores *z* em cada caso. Este valor é referenciado por McLoone (2004) como *Carstairs Score* e representa o número de desvios-padrão, do conjunto das variáveis, em relação à média.

Na Tabela 2 apresenta-se os resultados obtidos pelo autor (McLoone) para as unidades territoriais correspondentes a sectores de código postal, na Escócia. Na coluna da esquerda o autor definiu uma escala qualitativa, classificando as zonas como: áreas muito ricas (valores abaixo da média, a partir de -5); média escocesa (valores próximos da média) e áreas muito carentes (valores acima

da média, a partir de 6). Na coluna do meio encontram-se as classes de valores (*Carstairs Score*) e na coluna da direita o número de sectores de código postal por cada classe.

Tabela 2 – Distribuição dos *Carstairs Scores* de privação por sectores de código postal escoceses, baseados nos Censos 2001 (adaptado de McLoone, 2004)

	<i>Carstairs Score</i>	Sectores de código postal
<i>Classes qualitativas</i>	<i>Classes</i>	<i>Número</i>
<i>Áreas ricas</i>	-8 a -7	1
	-7 a -6	22
	-6 a -5	39
	-5 a -4	87
	-4 a -3	118
	-3 a -2	131
	-2 a -1	124
<i>Média escocesa</i>	-1 a 0	121
<i>Média escocesa</i>	0 a 1	88
<i>Áreas pobres</i>	1 a 2	73
	2 a 3	57
	3 a 4	52
	4 a 5	16
	5 a 6	25
	6 a 7	12
	7 a 8	9
	8 a 9	11
	9 a 10	6
	10 a 11	4
	11 a 12	6
	12 a 13	4
	13 a 14	3
14 a 15	0	
15 a 16	1	

O método dos *Z-scores* é útil quando se pretende resumir várias variáveis num valor de modo a evitar a sua análise em separado. Permite captar as múltiplas facetas da privação dos territórios através de uma única medida (Nogueira e Remoaldo, 2010).

2.4. Geocodificação

Geocodificação é um método de georreferenciação que permite a transformação da descrição de um lugar para uma referência geográfica absoluta. Um exemplo clássico é a conversão de endereços postais em pontos num mapa, atribuindo-lhes um par de coordenadas geográficas (latitude e longitude), (Bakshi *et al.*, 2004; Goldberg *et al.*, 2007; Zandberger, 2007; Ameluxen, 2010; Longley *et al.*, 2011).

Este processo associa coordenadas a dados a que pode ser associada uma localização geográfica através algoritmos que fazem a correspondência entre os dados inseridos por um utilizador e dados de referência. Os primeiros referem-se, por exemplo, a nomes e números usados pelo utilizador para descrever a localização da entidade (ex.: nome de rua, instituição, empresa, número de polícia, etc.), os segundos dizem respeito ao conjunto de dados existentes numa base de dados com elementos necessários à identificação unívoca dos locais, para uma região, país ou mundo (ex.: nomes de ruas, números de polícia, localidades, etc.), (Goldberg *et al.*, 2007; Zandberger, 2007).

O serviço de geocodificação é assegurado por grandes empresas (ex.: *Google*, *Yahoo*, *Microsoft*, etc.) que desenvolvem os algoritmos e mantêm os dados de referência atualizados. Normalmente é um serviço pago, mas dentro de determinado limite pode-se subscrever um serviço gratuito. A título de exemplo, a *Google* permite usar a sua *API Geocoding*², gratuitamente, até 2500 pedidos por dia (Google Developers, 2013). Existem programas SIG (ex.: *ArcGIS*, *Quantum GIS*, *gvSIG*, etc.) que incorporam ferramentas de geocodificação, permitindo ao utilizador converter listas de endereços para um ficheiro de pontos, porém, é nos serviços *WebMap* que a tecnologia tem maior expressão, tais como: *Google Maps*, *Bings Maps*, *Yahoo Maps*, entre outros.

A geocodificação divide-se em quatro componentes principais: dados de entrada; dados de saída; algoritmo de processamento e dados de referência. Os dados de entrada referem-se à descrição da localização que o utilizador pretende georreferenciar, estes devem conter atributos que possam ser identificados pelo algoritmo para efetuar a análise de correspondência com os dados de referência, normalmente é usado o endereço postal (nome da rua, número polícia e localidade) porque é a forma mais divulgada de referenciar um local (Bakshi *et al.*, 2004; Zandberger, 2007).

O processo de geocodificação também pode ser realizado de forma inversa, isto é, pesquisando os locais através das coordenadas geográficas obtém-se a descrição da localização. No *Google Maps* estas são inseridas no formato *DMS* (*degrees, minutes e seconds*) ou graus decimais, em sistema de referência *WGS84* (Google Developers, 2013).

O método de geocodificação baseado na rede de estradas é o mais difundido. Este método considera a rede de estradas como um conjunto de segmentos com os nomes das ruas e intervalo de números de polícia, estes últimos permitem identificar a posição do fenómeno relativamente ao segmento, incluindo o lado em que se encontra (ex.: ímpares do lado direito e pares do lado esquerdo), (Bakshi *et al.*, 2004; Zandberger, 2007).

² Interface de programação de aplicativos para geocodificação

A identificação do ponto é feita por interpolação linear, de acordo com o intervalo de números de polícia. Na Figura 12, para números entre 601 e 649 (ímpares), o número 631 estará posicionada a uma distância do início do segmento correspondente a 64% (fazendo a contagem no sentido 601 para 649) do seu comprimento e estará ao lado esquerdo da rua (Golberg *et al.*, 2007).

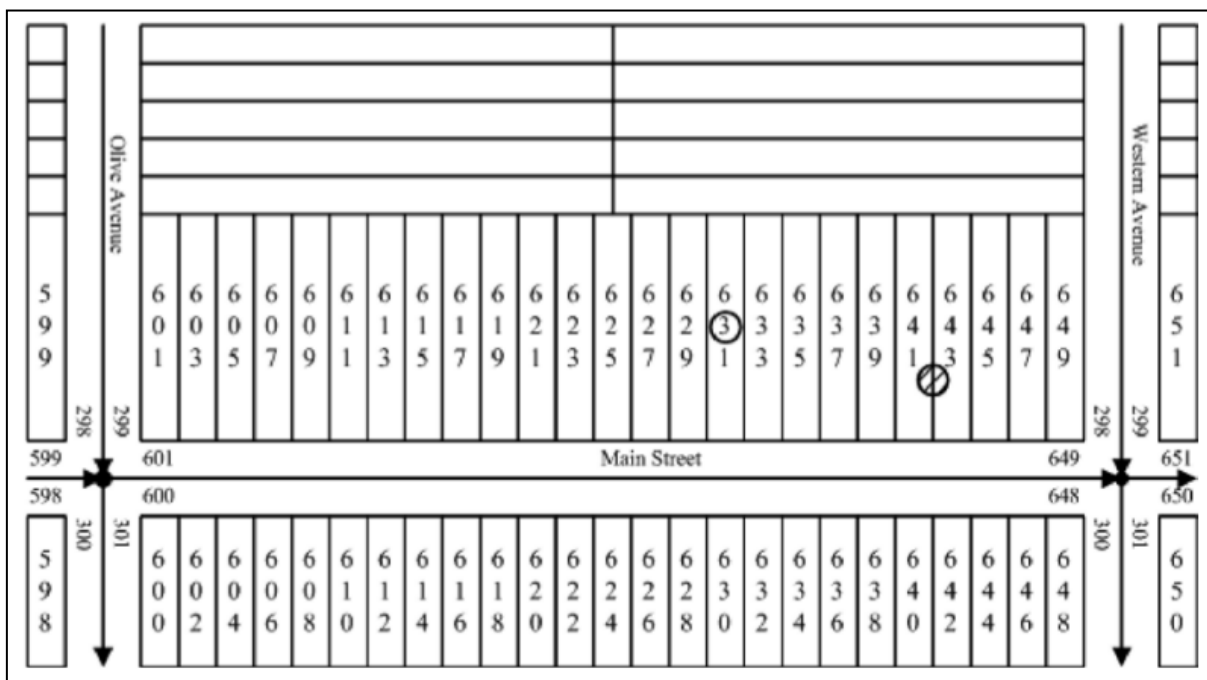


Figura 12 – Geocodificação baseada na rede de estradas, identificação do número de polícia (Golberg *et al.*, 2007)

Podem existir parâmetros adicionais permitindo um aumento de exatidão posicional. Na Figura 13 é considerado um *offset* no início e fim da rua (distância q); uma distância a partir do eixo da via (d) e uma direção dada pelo ângulo θ , este último define a orientação relativamente ao eixo da via, de modo a que o ponto seja colocado no centro da parcela (Golberg *et al.*, 2007; Zandbergen, 2007).

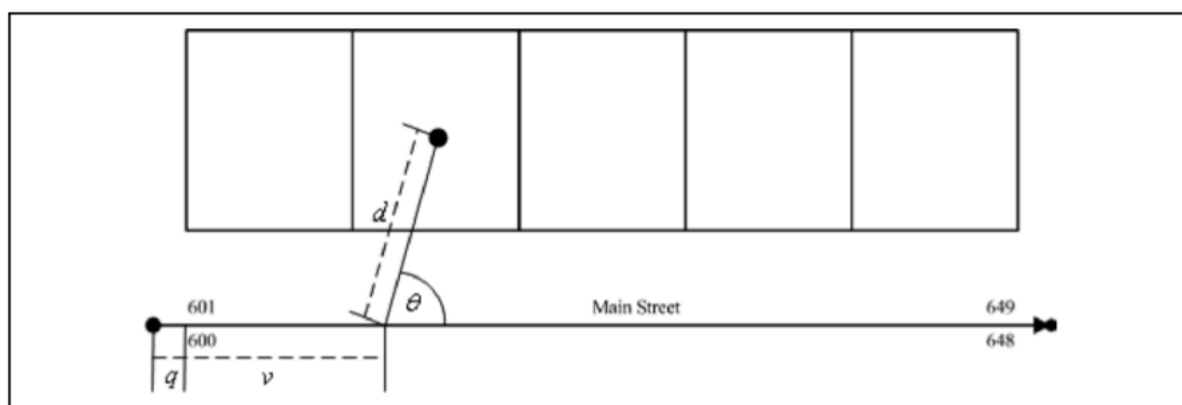


Figura 13 – Parâmetros adicionais do algoritmo de geocodificação (Golberg *et al.*, 2007)

É um método que apresenta algumas limitações de exatidão, nomeadamente por considerar o parcelamento homogéneo (Figura 12). Como esta situação nem sempre se verifica, surgem erros de posicionamento que degradam a exatidão posicional (Bakshi *et al.*, 2004; Golberg *et al.*, 2007). Pode-se ver pela Figura 14 que o ponto correspondente ao número de polícia 631 foi mal georeferenciado, na realidade as parcelas possuem uma configuração muito diferente. A sua verdadeira posição é representada por um círculo preenchido a traços na Figura 12.

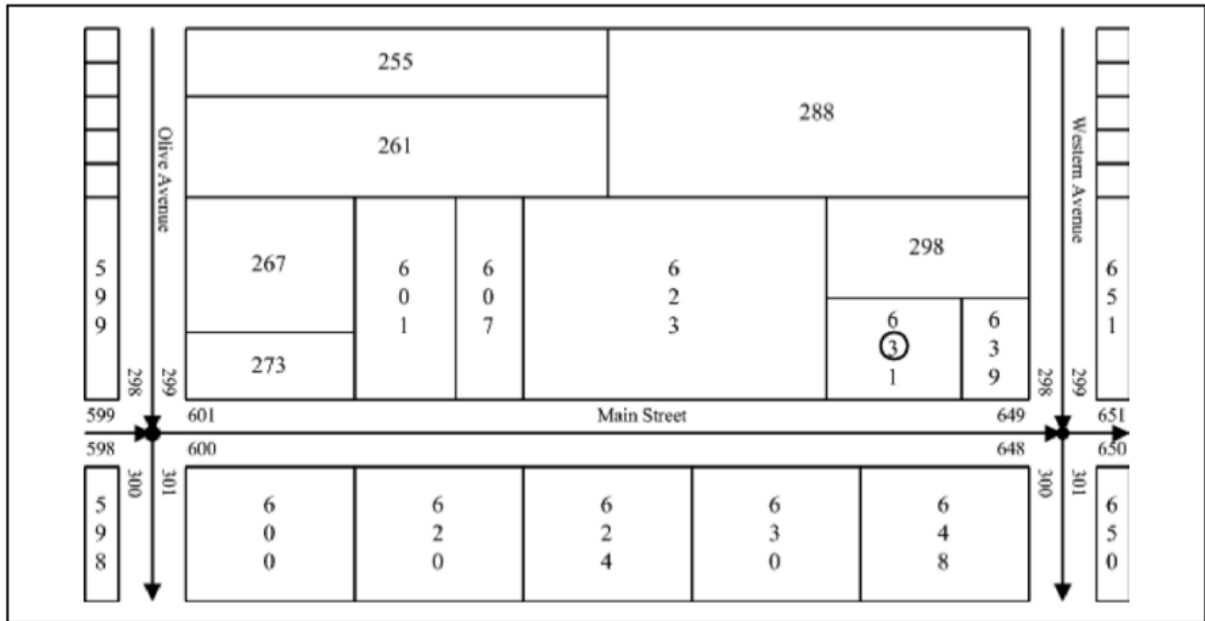


Figura 14 – Realidade do parcelamento ao longo da estrada (Golberg *et al.*, 2007)

Existem outros métodos de geocodificação que permitem obter melhores resultados na exatidão posicional. Como por exemplo, o método baseado nos limites de propriedade, este possui informação sobre os limites de propriedade e uma descrição mais completa dos locais (ex.: informação do que existe na propriedade, podendo haver mais que um endereço). Estes detalhes permitem obter uma maior exatidão posicional, pois a georreferenciação passa a ser baseada no centróide da parcela. Caso exista mais que um endereço na parcela a sua georreferenciação também é baseada no centróide, situação que representa uma limitação do método (Zandbergen, 2007; Golberg *et al.*, 2007).

O método baseado na rede de estradas está disponível para a maior parte dos países (Longley *et al.*, 2011), inclusive para Portugal. Como referido anteriormente pode ser executado através de serviços *WebMap* como *Google Maps*, *Bing Maps*, *Sapo Mapas*, etc.

2.5. Otimização da localização em redes

Através da análise de redes pode-se implementar algoritmos para otimizar a localização de infraestruturas, tais como, lojas, filiais, estabelecimentos de produção, centros de distribuição, etc., designadas daqui em diante por pontos de oferta. A otimização destes pontos de oferta é feita em função da distribuição da procura (clientes), (Ghosh e Rushton, 1987). O objetivo será maximizar um benefício ou minimizar um custo associado à distribuição espacial da oferta de um bem ou de um serviço (Lorena et al., 2001; Gonçalves e Matos, 2005). A abordagem através da maximização de um benefício é normalmente adotada pelo setor público pois pretende a satisfação dos clientes em detrimento dos custos necessários para o alcance do objetivo (ex.: localização de hospitais, escolas, postos de polícia, etc.). No setor privado considera-se que cada cliente tem necessidades que podem ser satisfeitas pela oferta, assim, os pontos de oferta devem ser localizados de forma a servir o máximo número de clientes possível minimizando os custos de deslocação na rede (Lorena *et al.*, 2001).

Este tipo de análise utiliza redes topológicas baseadas na teoria matemática dos grafos, nestas redes são descritas ligações e relações entre os elementos, sem usar coordenadas (Smith *et al.*, 2013). Uma rede $G = (N, A)$ possui um conjunto de nós (N) (pontos) e um conjunto de arcos (A), sendo um arco uma linha que une dois nós.

Numa aplicação real os pontos de oferta e clientes podem ser representados pelos nós da rede, e as estradas pelos arcos. Para encontrar o ponto de oferta que melhor sirva todos os clientes, pode-se encontrar a combinação que minimize o custo de deslocação. Esta pode ser determinada pela soma dos custos de deslocação (ex.: distância, tempo de viagem, etc.) de todos os caminhos possíveis, partindo dos clientes para os pontos de oferta. O ponto de oferta escolhido deve ser o que apresente um custo de deslocação menor.

Na Figura 15, apresenta-se um exemplo de rede topológica onde os valores associados aos arcos representam distância em quilómetros e os números dentro das caixas o número do nó. Os nós 3 e 6 são considerados pontos de oferta e os restantes os clientes (1, 2, 4, 5, 7). A Tabela 3 apresenta os custos de deslocação em todos os caminhos possíveis (colunas 2 a 4), dos clientes para os pontos de oferta 3 e 6. Verifica-se que o valor total de custo de deslocação é menor no ponto de oferta 6 (1025 Km), e por esta razão, é o ponto de oferta que melhor serve os clientes.

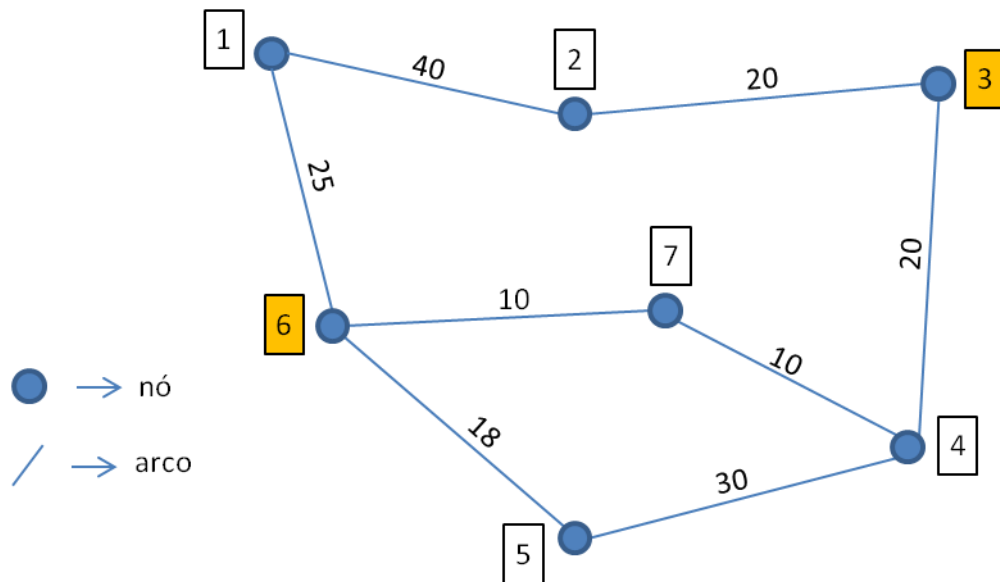


Figura 15 – Rede topológica com custo (distância) associado aos arcos

Tabela 3 – Caminhos possíveis para os pontos de oferta representados pelos nós 3 e 6. Custo de deslocação avaliado em quilómetros

Caminhos possíveis				
Cientes	Ponto de oferta (3)			Subtotais
1	60	65	93	218
2	20	105	133	258
4	20	105	133	258
5	50	135	103	288
7	30	95	143	268
Total (Km)				1290

Cientes	Ponto de oferta (6)			Subtotais
1	25	100	128	253
2	65	60	88	213
4	105	20	48	173
5	18	50	135	203
7	10	115	58	183
Total (Km)				1025

2.5.1. Problemas de localização-alocação

Os problemas localização-alocação permitem otimizar a localização de pontos de oferta e, simultaneamente, alocar os clientes a esses pontos (Ghosh e Rushton, 1987). Pode-se considerar o exemplo de uma empresa de venda a retalho. A otimização da localização de novas lojas é feita de forma que a solução contenha as lojas que ficam mais perto dos clientes e que cada cliente é vinculado (alocado) à loja que lhe fica mais perto.

Como a localização ótima não é igual para todas as atividades estes problemas podem ser abordados segundo várias formas: **cobertura máxima**, **p-centro** e **p-mediana**, (Mirchandani e Francis, 1990; Eiselt e Sandblom, 2004; Gonçalves e Matos, 2005).

Nos problemas de **cobertura máxima** o número de pontos de oferta é racionalizado no sentido de obter máxima cobertura da procura. Pode-se definir o número de pontos de oferta e permitir o aumento da distância percorrida na rede de forma a aumentar a cobertura, ou aumentar o número de pontos de oferta de forma a diminuir o número de clientes sem cobertura. Para aplicar este modelo é necessário fornecer o número de pontos de oferta e valores (ex.: distância, tempo de viagem, etc.) para a cobertura máxima admitida (Eiselt e Sandblom, 2004).

O problema **p-centro** é direcionado para os pontos de oferta, procura-se minimizar a distância mais longa entre um cliente e o seu ponto de oferta mais próximo (Eiselt e Sandblom, 2004). Este tipo de problema também é usado em serviços de emergência, por exemplo, se pretendermos localizar meios de combate a incêndios de forma a minimizar o tempo máximo de resposta (Mirchandani e Francis, 1990; Longley *et al.*, 2011).

Nos problemas de localização-alocação o **p-mediana** é o mais utilizado. Este procura a localização dos pontos de oferta de forma a minimizar a soma das distâncias entre cada cliente e o ponto de oferta que lhe está associado (Ghosh e Rushton, 1987; Mirchandani e Francis, 1990; Lorena *et al.*, 2001; Eiselt e Sandblom, 2004; Longley *et al.*, 2011). É muito utilizado em empresas que pretendem a localização ótima das suas lojas/filiais, sem ter de definir critérios para cobertura máxima admitida (como no problema de cobertura máxima), ainda que, muitas vezes são atribuídos pesos aos pontos de oferta de forma a conseguir uma solução mais enquadrada no mundo real (Mirchandani e Francis, 1990). Por exemplo, se o decisor atribuir o volume de negócios por loja (seus pontos de oferta) como peso. As lojas com maior volume de negócios podem ter maior possibilidade de fazer parte da solução.

Devido à complexidade do problema a sua solução computacional é classificada como *NP-Hard* (*Non-deterministic Polynomial-time hard*), isto é, encontrar a solução ótima é teoricamente possível por

pesquisa exaustiva, enumeração e avaliação de todas as possíveis soluções (combinações), porém, é inviável seguir essa estratégia uma vez que o número de combinações possíveis aumenta exponencialmente com o tamanho do problema. Considera-se que a solução ótima não é realizável em tempo computacional útil (Dias, 2005). Para conseguir soluções em tempo computacional útil são usados algoritmos baseados em métodos aproximados (heurísticos). Segundo Smith (2013), existem métodos heurísticos que produzem soluções muito próximas da solução ótima.

A Figura 16 mostra a solução encontrada pelo *software* comercial *ArcGIS 10.0* da *ESRI* para resolução do problema p -mediana. Pode-se visualizar os pontos de oferta considerados na solução e seus clientes alocados, estes últimos são nós na rede ligados simbolicamente (pelas linhas rosa) ao seu ponto de oferta. Neste caso, o decisor definiu que a solução devia ter 7 pontos de oferta dos 16 possíveis.

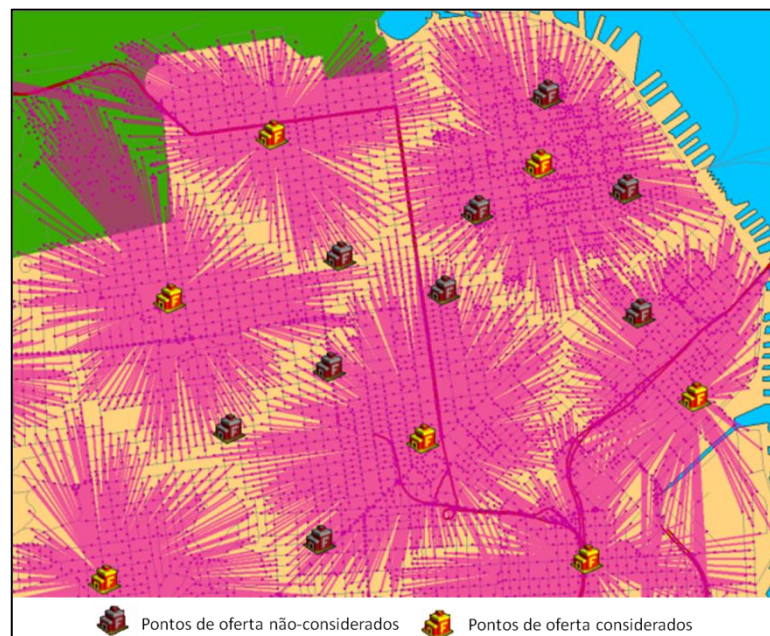


Figura 16 – Localização-alocação por métodos heurísticos (Fonte: ESRI - ArcGIS 10.0 Help, 2013)

2.5.2. Áreas de serviço

Entende-se por áreas de serviço a definição das áreas mais próximas dos pontos de oferta em termos de custo (ex.: distância, tempo de viagem, etc.), na prática são alocadas sub-redes sem considerar os clientes e mantendo os pontos de oferta fixos (Smith, 2013). Trata-se de um método mais simples que os anteriores (ponto 2.5.1), pois não existe variação no número de pontos de oferta. Na Figura 17 encontram-se definidas áreas de serviço para ambulâncias, como se trata de um serviço público em que o objetivo é o máximo benefício, a rede é alocada à quantidade de recursos disponíveis (ambulâncias) de forma a cobrir toda a procura.



Figura 17 – Definição de áreas de serviço para ambulâncias, com custo distância (Smith, 2013)

Também se podem definir bandas de custo através do corte da rede, usando linhas de igual custo (isolinhas). Esta operação pode resultar numa rede composta por várias bandas de custo (ex.: 15, 20, 25 minutos de viagem) ou em polígonos, cada um representando uma banda de custo (Smith, 2013), estas bandas de custo podem partir de um ponto de oferta ou de vários. A partir deste momento a rede resultante desta operação designa-se de rede iso. A Figura 18 mostra uma rede iso correspondente a um ponto de oferta (a verde), contendo bandas de custo em metros: 1000, 2000, 3000 e >3000. Estando estas simbolizadas pelas cores: cinzento, vermelho, castanho e verde, respetivamente.

Com esta abordagem podem-se realizar análises interessantes, por exemplo, quantos clientes estão a menos de 2000 metros do ponto de oferta?

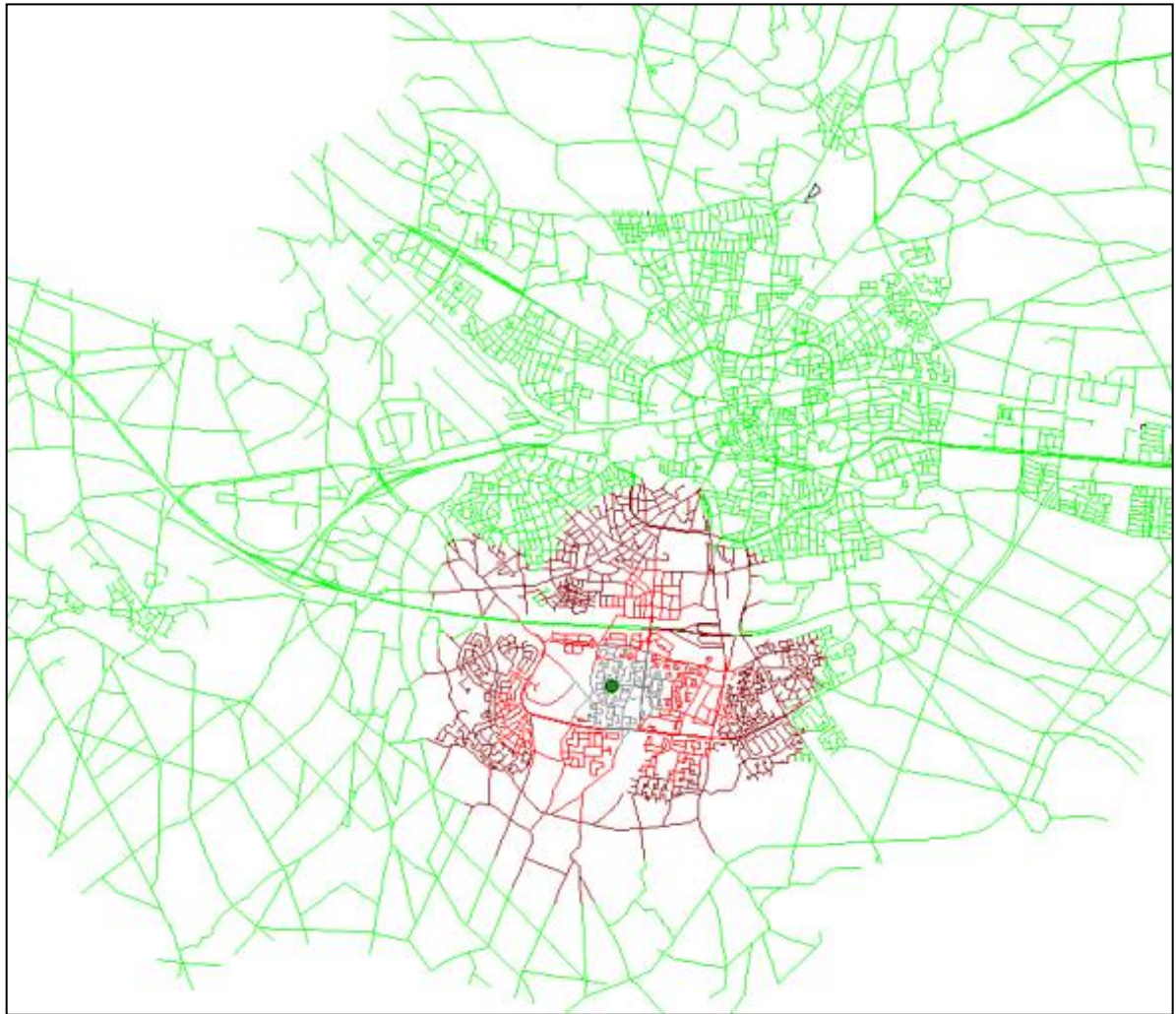


Figura 18 – Rede iso para ponto de oferta (a verde) com bandas de custo: 1000, 2000, 3000, >3000 metros, representadas pelas cores: cinzento, vermelho, castanho e verde, respetivamente. Resultados obtidos pela ferramenta *v.net.iso* do software *GRASS GIS* (Castells, 2013)

2.6. Os SIG no apoio à decisão

O conceito de sistema de suporte à decisão (SSD) começou a ser desenvolvido nos anos 70 do século XX e rapidamente foi reconhecido como parte dos sistemas de informação (SI), (Nag e Sengupta, 2008). É um sistema que possui ferramentas para apresentação e modelação de dados de forma a ajudar os decisores a compreender problemas e encontrar soluções. Neste contexto entende-se modelação como um conjunto de regras e procedimentos usado para representar um fenómeno ou fazer previsões (Somers, 2006).

Os SSD evoluíram dentro do mundo dos negócios tratando apenas de informação financeira e outros dados inerentes à atividade empresarial. O volume de dados era relativamente pequeno, quando comparado com os dados geográficos, o que constituía um entrave à integração de informação geográfica. Com a grande evolução dos sistemas computacionais foi possível

incorporar alguma informação espacial, ainda com um ambiente gráfico básico. Um bom exemplo usado em análise de dados geográficos e sua apresentação gráfica foi o GADS (*Geodata Analysis and Display System*) desenvolvido na década de 70 do século XX. Era usado para resolver problemas relacionados com planejamento de rotas, mas possuía gráficos muito limitados e capacidade de processamento inadequada (Burstein e Holsapple, 2008; Nag e Sengupta, 2008)

A par do desenvolvimento dos SSD houve uma expansão dos SIG, a capacidade de processamento de dados melhorou e desenvolveram-se novas aplicações. Porém, a integração de informação geográfica nos SSD revelou-se ser um processo lento, principalmente devido à complexidade computacional no tratamento deste tipo de informação (Nag e Sengupta, 2008).

O crescente desenvolvimento tecnológico e o interesse pela integração de informação geográfica nos processos de decisão levaram ao aparecimento dos sistemas espaciais de apoio à decisão (SEAD). Estes sistemas podem ser vistos como uma integração entre *software* e *hardware* num contexto espacial, usando as funcionalidades dos SIG sobre a gestão da base de dados, gestão de conhecimento, análise de critérios, modelação e apresentação de dados (geográficos e não-geográficos), (Li *et al.*, 2004).

Para Longley *et al.* (2011) os SEAD são SIG adaptados para apoiar processos de decisão. Devem fornecer resposta imediata sobre as várias implicações que advém das propostas dos decisores e permitir cenários alternativos. Normalmente possuem interfaces especiais que apresentam apenas as funcionalidades relevantes para o projeto.

A Figura 19 mostra o modelo concetual de um SIG para suporte à decisão proposto por Mennecke (2000). Este modelo atribui maior relevância à informação geográfica face ao modelo proposto por Densham (1991), procura integrar os SSD com os SIG e contribuir para a definição dos SEAD, é por isso o modelo base adotado para a esquematização dos SEAD (Keenan, 2005; Nag e Sengupta, 2008; Sugumaran e DeGroot, 2011; Azaz, 2011). Verifica-se que a informação geográfica entra em todos os processos do modelo, contribuindo para a gestão dos dados, modelação e apresentação de resultados.

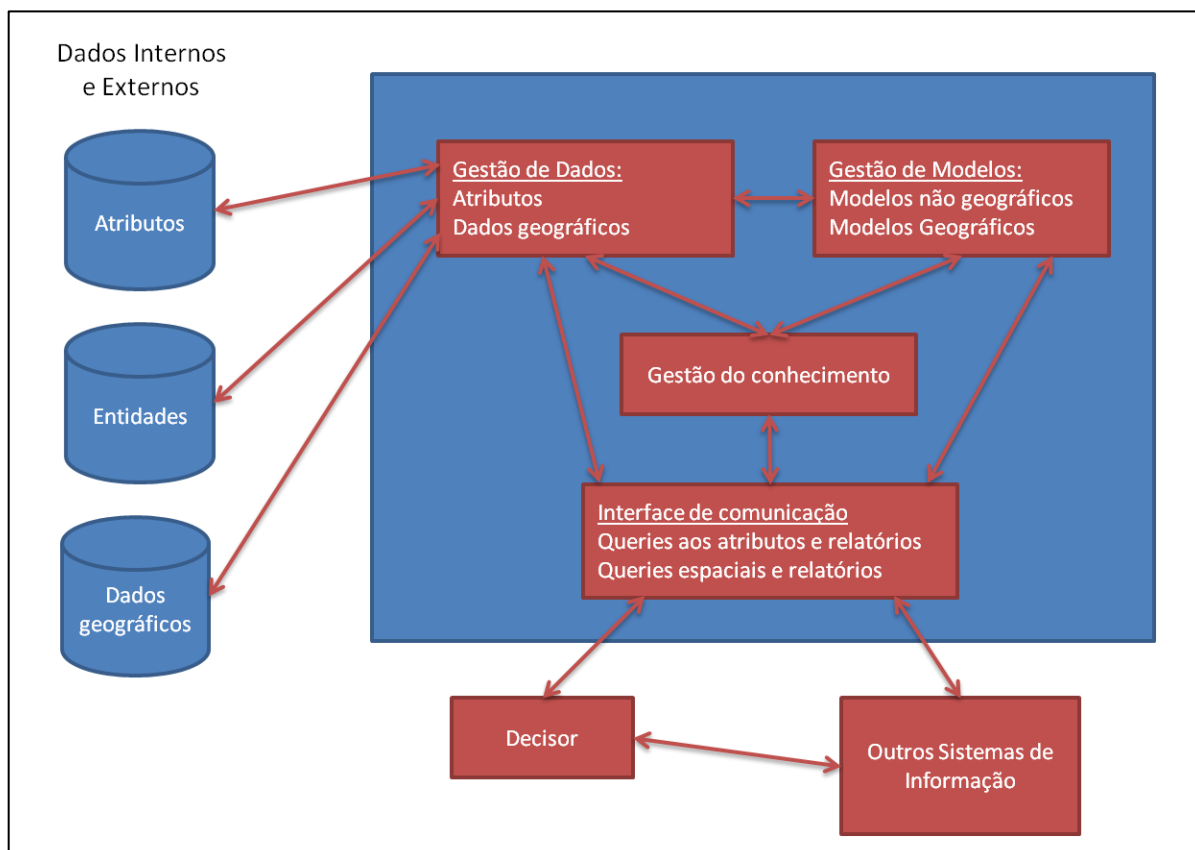


Figura 19 - Modelo conceitual de um SIG usado para suporte à decisão, proposto por Mennecke (2000)

O interesse pela incorporação da informação e análise geográfica nos processos de decisão deve-se ao facto de 80% da informação necessária para negócios e governo possuir base geográfica. As potencialidades económicas da informação geográfica levaram a que muitas organizações passassem a incorporar dados geográficos nas suas atividades (Sugumaran e DeGroot, 2011).

Não obstante das vantagens dos SEAD as características inerentes aos problemas de decisão espacial/geográfica incrementam complexidade na análise dos dados e consequentemente no processo de tomada de decisão. Os autores Nag e Sengupta (2008) referem particularidades inerentes a este tipo de problemas:

- Existe grande número de decisões alternativas;
- Os resultados ou consequências das decisões alternativas variam espacialmente;
- Cada alternativa é avaliada tendo como base múltiplos critérios;
- Tipicamente existe mais do que um decisor envolvido no processo de decisão;
- Os decisores têm diferentes preferências relativamente à importância dos critérios e consequências das decisões;
- As decisões estão frequentemente rodeadas de incertezas.

2.6.1. Processos de decisão e informação geográfica em contexto empresarial

Nos últimos 20 anos tem-se verificado um aumento do uso dos SIG em meio empresarial: banca, serviços financeiros, logística e transportes, imobiliárias e análise de mercados, no entanto, estes já eram usados pelo sector público, em atividades relacionadas com a produção de cartografia, educação, administração do território, entre outras (Pick, 2007; Longley *et al.*, 2011). O meio empresarial levou mais tempo a adotar os SIG devido ao seu elevado custo e indefinição dos seus benefícios. Em termos computacionais as base de dados geográficas são mais pesadas que as não-geográficas e a análise de dados geográficos requer maior capacidade de processamento. Ao nível dos recursos humanos, no meio empresarial as pessoas possuem menos conhecimentos de SIG e existem maiores dificuldades em conseguir formação em SIG (Pick, 2007).

Estes sistemas tornam-se cruciais para responder a questões do tipo “Onde?”, como por exemplo, qual o melhor local para implementar a empresa e encontrar potenciais clientes (Grismshaw, 2000). Os SIG podem ajudar uma cadeia de lojas a retalho a localizar as suas lojas de forma a obter uma rede bem estruturada que vai de encontro às necessidades dos clientes e ajudar os seus especialistas de Marketing a encontrar os melhores locais para colocar/distribuir publicidade. Estes sistemas permitem perceber, questionar, interpretar e visualizar os dados de forma diferente da análise das linhas e colunas de uma tabela. É possível colocar questões do tipo: “Onde?”, “Porquê?” e “Como?” (Grismshaw, 2000, Azaz, 2011).

Os SIG podem ser aplicados em várias áreas de negócio, mostrando versatilidade e transversalidade. A seguir dão-se alguns exemplos de possíveis aplicações (Mennecke, 2000; Thomas e Ospina, 2004; Keenan, 2005; Pick, 2005; Posthumus, 2008; Azaz, 2011; Longley *et al.*, 2011).

Gestão de instalações, informação sobre localização, condições dos edifícios, área geográfica que servem, distribuição geográfica dos edifícios, planta interna, capacidade de armazenamento, monitorização dos edifícios em tempo real, etc. As empresas de distribuição de gás, água e eletricidade (*utilities*) incorporam SIG no seu *Software* de gestão;

Marketing, talvez a área que maior partido tirou desta tecnologia, o denominado geomarketing tem provas dadas em situações relacionadas com, localização do mercado, características dos consumidores e sua localização, análise da concorrência, locais preferenciais para colocação de painéis de publicidade, distribuição de publicidade, etc. A referir que a cadeia de restaurantes MacDonald’s lidera o uso de SIG para estudo do mercado.

Logística e Transporte, área tradicional no uso dos SIG, embora sejam usados principalmente na análise de redes (ex.: caminho mais curto, problema do caixeiro-viajante), sistemas de inventariação e despacho de encomendas. Empresas de transportes lideram na utilização destes sistemas.

Decisão Estratégica e Planeamento, parte da informação usada por gestores provem de um número variado de fontes, pouco estruturada e num horizonte temporal grande, isto é, carece de consistência. Assim, surge a necessidade de implementar modelos analíticos que consigam construir vários cenários (ex.: análise multicritério) e possam ser usados por não especialistas em SIG, como gestores. É neste contexto que emerge o modelo SEAD, flexível e com interface amigável, mas que mesmo assim implica alterações na cultura empresarial, pois trata-se de fornecer aos gestores uma visão geográfica do negócio. Existem algumas empresas que fazem este tipo de aproximação ao usarem os SIG, como, o já referido *MacDonald's* ou a cadeia de lojas *Marks and Spencer*. Conseguem orientar a sua oferta (ex.: abertura de novos estabelecimentos) através do geoprocessamento de dados demográficos e económicos.

Num ambiente empresarial tomar a decisão certa no momento certo implica usar um formato de informação intuitivo e de leitura rápida. Um SIG permite reunir informação proveniente de várias fontes num só mapa. Esta capacidade possibilita a comparação entre as múltiplas variáveis de um problema, tornando o processo de tomada de decisão mais célere (Posthumus, 2008).

2.7. Fases de implementação de um projeto SIG

O processo de implementação de um SIG é complexo, dependendo do tamanho da organização e da quantidade de informação geográfica que é necessária. Também requer boas capacidades de gestão de projetos (Harmon e Anderson, 2003; Longley *et al.*, 2011).

Ainda que existam inúmeros tipos de SIG, com características e objetivos diferentes, o esquema usado para implementação do SIG é comum a todos. Para o setor privado os SIG tendem a ser mais específicos, com funcionalidades muito próprias e bases de dados distintas, não podendo ser usadas em outros projetos. Quando se inicia a implementação os objetivos do SIG devem estar bem definidos, ainda que possam ser moldados face a condicionantes futuras (Somers, 2001).

A Figura 20 apresenta o esquema de implementação de um SIG. Este é composto pelas seguintes fases (Somers, 2001; Harmon e Anderson, 2003; Longley *et al.*, 2011):

Planificação geral, neste ponto identifica-se dois tipos de planificação: plano estratégico, onde se expõe o papel do SIG na empresa e os objetivos a atingir; plano operacional onde se define a

natureza do SIG, de que tipo é, por quem vai ser usado, que departamentos da empresa o vão utilizar, se os dados espaciais e tecnologia vão ser integrados em outros SI, que impactos vai ter na forma como a empresa conduz os negócios, etc. Numa análise mais aprofundada devem ser bem identificados os participantes do projeto, assim como, os recursos afetados. Também se deve definir formação adequada para os utilizadores da tecnologia SIG e feita uma estimativa dos benefícios;

Análise dos requisitos e resultados a esperar, realiza-se uma análise exaustiva dos requisitos que o SIG deve ter, nomeadamente, os processos de trabalho, tipos de dados, fontes de dados, ligação com outros SI, análise dos processos de gestão que vão interferir com o SIG, entrevistas com os recursos humanos e avaliação dos conhecimentos de SIG dos futuros utilizadores. Por fim deve ser feita uma análise cuidada dos resultados a esperar, sendo possível identificar, custos e benefícios, constrangimentos, oportunidades ou problemas previsíveis;

Design do Sistema, definição das ferramentas/capacidades do SIG para cumprir os requisitos, anteriormente determinados. Em primeiro lugar deve-se desenhar a base de dados, pois é um importante componente do SIG, casos de estudo referem que pode representar até 80% do orçamento. De seguida, procede-se à escolha do *software* e aplicações, definindo o tipo de funções para aplicações que irão suportar as ferramentas SIG e para a gestão da base de dados. Também se define de que forma vão ser acedidos os dados e que requisitos o sistema deve ter, no que respeita à integração dos dados. É determinado o arranjo global do SIG, em aspetos relacionados com interfaces, integração, planeamento das tarefas de análise e gestão de todos os componentes que compõem o SIG. Por fim, são avaliados os resultados esperados com o design que foi concebido;

Aquisição e Desenvolvimento, aquisição de *software*, *hardware*, dados e sua integração no sistema. Nesta fase pode existir necessidade de adquirir dados a outras organizações ou através de trabalho de campo. Por fim é feita a integração de todas as ferramentas de maneira que o SIG fique operacional para os utilizadores;

Operação e Manutenção, este último item trata do uso do SIG e sua manutenção. É um processo crítico e normalmente lento, pois os utilizadores têm de se familiarizar com as ferramentas e estar atentos a possíveis problemas de implementação. Normalmente são feitos testes piloto no sentido de avaliar a integridade dos dados e as capacidades do *software*. Nesta fase é importante reconhecer que o SIG não está acabado, mas antes, em melhoria contínua.

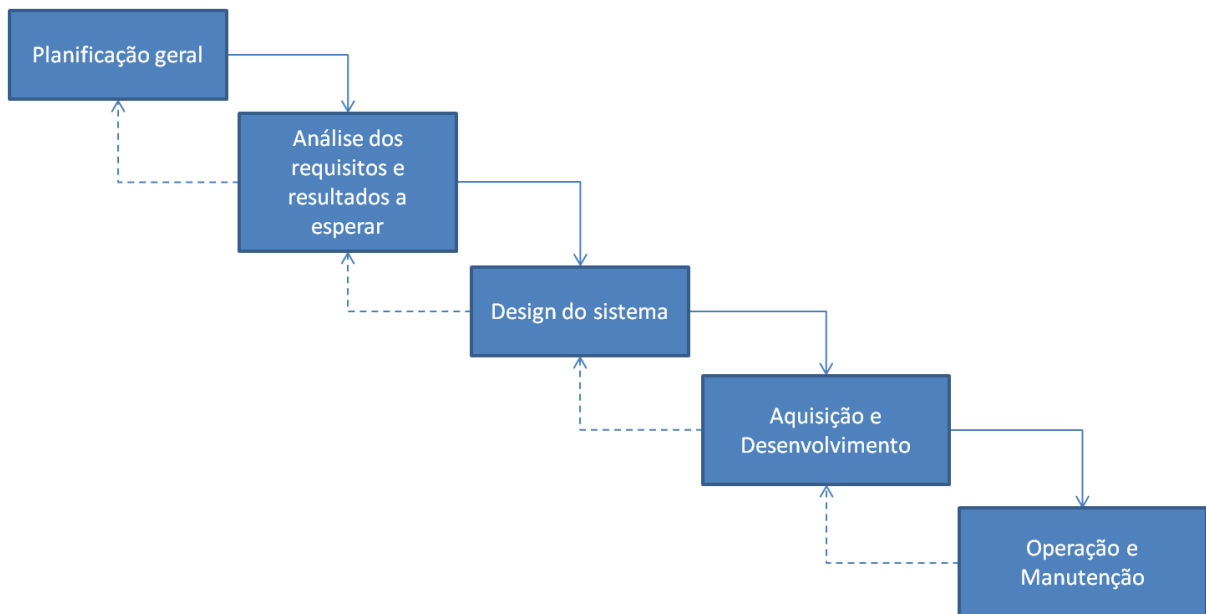


Figura 20 - Processo de implementação de um SIG (Adaptado de Somers, 2001)

Como se pode constatar este processo deve estar aberto à possibilidade de fazer reformulações ou alterações pontuais, com vista a resolver problemas identificados quando do seu uso. É por isso que o processo de implementação deve ser entendido também como processo de gestão do SIG.

3. IMPLEMENTAÇÃO DO PROJETO SIG-ATEP

3.1. Enquadramento empresarial

3.1.1 Atividade da empresa

A empresa ATEP desenvolve a sua atividade na área de formação de jovens e adultos. Tem sobre a sua responsabilidade uma escola profissional, Escola Profissional de Torres Novas (EPTN), um Centro de Novas Oportunidades (CNO) e vários projetos de formação de curta duração (formações modulares). O projeto SIG vai apoiar o CNO e as formações modulares, porque é na educação de adultos que existe maior exigência no planeamento de horários e abertura de turmas. A maior parte da atividade formativa é realizada em horário pós-laboral e em itinerâncias, ou seja, em locais de formação criados para o efeito, fora das instalações da ATEP (ex.: juntas de freguesia, associações culturais, etc.).

No exercício das atividades formativas surgiu a necessidade de otimizar a distribuição dos locais de formação, no sentido de conseguir uma gestão mais eficiente das deslocações, procurando minimizar os tempos de viagem para formadores e formandos.

Esta necessidade deriva da vastidão do território onde a empresa tem de assegurar formação, NUT III Médio Tejo composta por 11 concelhos (Figura 21). Ainda que existam mais empresas a promover formação neste território, todas têm a obrigação de satisfazer as necessidades formativas na totalidade da NUT III.

A empresa utiliza uma plataforma *on-line*, Sistema de Informação e Gestão da Oferta Educativa e Formativa (SIGO), para gestão dos formandos: inscrições; inserção de dados sobre ações de formação; certificações obtidas; transferência de/para outras instituições de formação; inserção/atualização de dados pessoais; etc. O SIGO é mantido pela Agência Nacional para a Qualificação e Ensino Profissional (ANQEP) cabendo à ATEP a utilização e atualização de dados, referentes aos seus formandos.

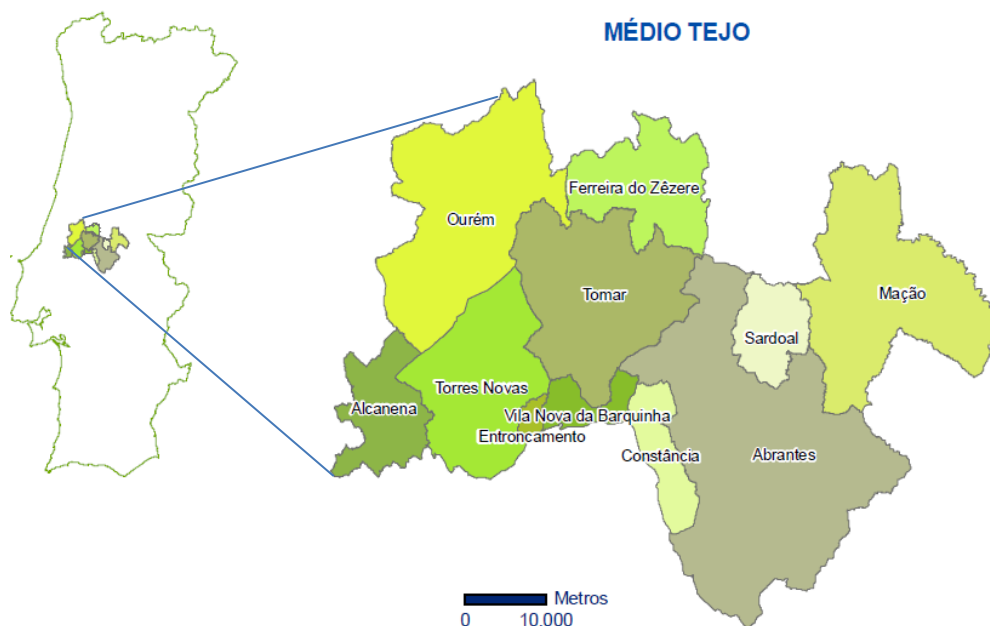


Figura 21 – NUTIII Médio Tejo (Fonte: Comunidade Intermunicipal Médio Tejo, 2013)

Aquando o planeamento da abertura de novas turmas surge o interesse em encontrar as zonas (ex.: freguesias) com maior necessidade de formação, porque além das inscrições feitas por iniciativa das pessoas é útil realizar ações de divulgação no sentido de captar mais interessados. Neste aspeto será importante utilizar indicadores estatísticos que forneçam informação fidedigna sobre a realidade sociodemográfica da região.

3.1.2 Recursos humanos envolvidos

Como a empresa não possui funcionários com conhecimentos de SIG ou cartografia é pertinente fazer uma reflexão sobre o seu enquadramento no projeto, pois o SIG deve ser usado por estes, de forma autónoma.

As pessoas que irão usar o SIG são de ciências sociais e jurídicas possuindo responsabilidades na gestão e planeamento das atividades formativas. Realizam análises financeiras aos projetos; fazem planeamento ao nível de alocação de recursos humanos e calendarização das ações de formação; divulgam as atividades da empresa em toda a NUT III e coordenam as equipas de trabalho.

Devido a esta realidade será necessário promover formação de curta duração e construir um manual de utilização, direcionado para o projeto em questão. Não será útil fazer uma abordagem muito abrangente aos SIG e suas ferramentas, pois a especificidade do projeto não implica essa necessidade.

3.2. Identificação dos requisitos

O SIG-ATEP deve-se basear em *software* de código aberto permitindo a sua utilização e atualização sem custos financeiros (GNU, 2013). A aplicação SIG que irá ser utilizada pelos colaboradores da ATEP deve ter uma interface amigável e intuitiva permitindo aceder às ferramentas de análise através de botões personalizados para o SIG-ATEP.

As ferramentas utilizadas devem permitir a importação de dados provenientes do SIGO, existentes em formato *XLSX*³, e manipulação do seu conteúdo. Os dados dos formandos (ex.: endereço postal, nome, etc.) devem ser georreferenciados, de forma a serem usados nas operações de análise espacial. Este processo terá que ser concretizado através de geocodificação utilizando um serviço gratuito.

O SIG-ATEP deve possuir ferramentas para análise de redes, nomeadamente na resolução de problemas de localização-alocação, definição de áreas de serviço e rede iso. Os dados resultantes das análises deverão ser exportados para uma localização pré-definida. Esta localização deverá ser numa pasta que contenha os dados necessários ao funcionamento do SIG, dados a importar para o SIG e exportados por este.

A aplicação SIG deve executar análises espaciais como: interseção, sobreposição e união, e deve permitir a realização de consultas. Na apresentação de resultados, deve possuir ferramentas para visualizar informação proveniente de serviços *WebMap* (ex.: *Google Maps*) e manipulação dos níveis de informação (camadas), no que respeita à alteração de simbologia dos mapas, escolha do sistema de referência, edição da tabela de atributos, edição das entidades geográficas, etc.

Os dados utilizados devem ser gratuitos e estar num formato que possa ser usado pelo SIG (ex.: *shapefile*⁴). Também devem ter qualidade mínima (geométrica e temática) para integrar no SIG-ATEP. Serão necessários dados sobre: rede de estradas, com informação que permita calcular o tempo de viagem; locais onde se pode desenvolver formação; limites administrativos ao nível do concelho, freguesia, subsecção estatística; estatísticas sociodemográficas dos Censos 2011. Os dados dos Censos 2011 devem servir de base ao cálculo de um indicador estatístico de privação múltipla, que seja indicador do nível de necessidade de formação da população. Este será designado como indicador de necessidade de formação (INF) a partir daqui.

³ Extensão de ficheiros Microsoft Office Excel

⁴ Arquivo usado para armazenar dados geoespaciais do tipo vetorial, desenvolvido e regulamentado pela *ESRI*

3.3. Escolha do *software*

Foi feita uma pesquisa exaustiva de forma a selecionar uma aplicação SIG de código aberto e uma folha de cálculo para tratamento de dados provenientes do SIGO e de fontes estatísticas (ex.: Censos 2011).

A escolha da aplicação SIG de código aberto foi bastante difícil, visto que a informação existente sobre este tipo de *software* encontra-se disseminada por várias fontes e com pouco suporte oficial. De acordo com os requisitos definidos no ponto 3.2. apresentam-se os critérios de escolha:

- ✓ Ser de código aberto;
- ✓ Ter interface amigável;
- ✓ Ser facilmente personalizável;
- ✓ Permitir trabalhar com os formatos mais comuns de dados (ex.: *shapefile*);
- ✓ Permitir executar geocodificação de vários registos a partir de um ficheiro;
- ✓ Possuir algoritmos para análise de redes (localização-alocação, áreas de serviço e definição de rede iso);
- ✓ Permitir carregar *WebMaps* (exs: *Google Maps*; *Bing Maps*; etc.);
- ✓ Permitir a exportação de dados para suportes/formatos universais (ex.: *JPG*, *HTML*).

Existem muitos *softwares* SIG de código aberto com diferentes funcionalidades e capacidades de análise. A maioria cumpre as funcionalidades básicas para gestão de um SIG, nomeadamente as referenciadas nos critérios de escolha. Para necessidades específicas as opções diminuem bastante, chegando mesmo a não existir opção.

Dentro das possibilidades verificou-se que poucos *softwares* dispõem de ferramentas de geocodificação, a maioria dos *softwares* testados só permite entrada de registos individuais, a título de exemplo, *MapWindow* e *gvSIG*. Para este critério também se deve ter em conta que é necessário possuir uma *API Google* para geocodificar. A subscrição deste serviço implica custos para uma utilização superior a 2500 pedidos diários, no entanto, em utilizações inferiores existe a possibilidade de a subscrever de forma gratuita (Google Developers, 2013). É através desta *API* que conseguimos aceder ao serviço de geocodificação da *Google*, a forma como é feita a entrada e saída dos registos depende do *software* que estamos a usar, depende da interface desse *software*.

Para conseguir a geocodificação de vários registos (designada a partir daqui como geocodificação múltipla) seria importante que a aplicação permitisse o carregamento dos dados a partir de um

ficheiro e geocodificasse os registos (moradas, códigos postal, etc.) sem pedir informação adicional, como a *Google API key* (chave que permite a utilização da *API*).

O *software Quantum GIS 1.8.0 “Lisboa”* (Quantum GIS, 2013) possui um *plugin (mmqgis)*, (Minn, 2013) que permite geocodificação múltipla, com *API* da *Google*, a partir de um ficheiro *CSV*⁵. Esta ferramenta é bastante fácil de utilizar, precisando apenas de especificar os campos do *CSV* a que corresponde: endereço, cidade, estado e país (Figura 22). A informação também pode ser manipulada no ficheiro, por exemplo, no campo cidade podemos ter a localidade e no estado, o concelho/distrito. Após a geocodificação são criados dois ficheiros: uma *shapefile* de pontos com os atributos referentes à informação contida nas linhas do *CSV* (em sistema de referência *WGS84*) e um *CSV* com os registos que não foram encontrados. Caso seja inserida informação adicional no *CSV* (exs.: nomes, telefones, etc.), esta não é eliminada, ficando igualmente registada na tabela de atributos.

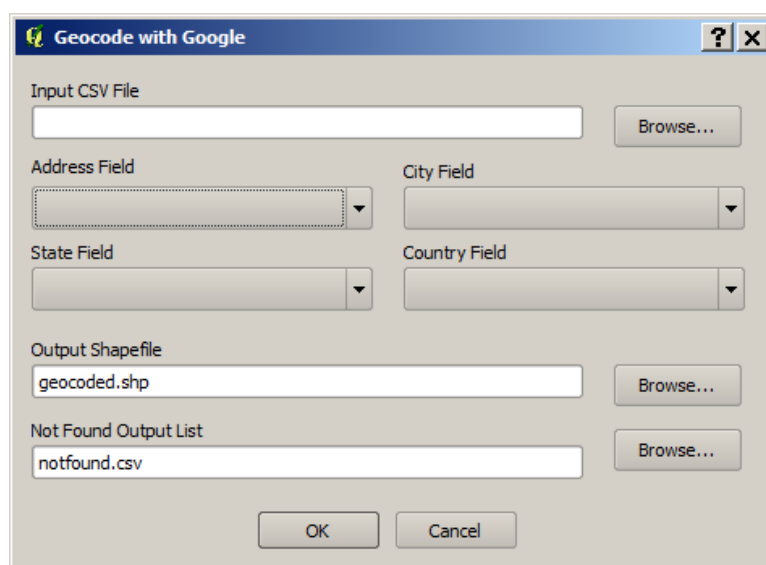


Figura 22 – Ferramenta para geocodificação múltipla (*plugin mmqgis*)

No critério de escolha para análise de redes o número de aplicações existente é mais reduzido, sendo as capacidades de análise bastante variáveis. De todos os *software* de código aberto analisados não foi encontrado nenhum com análise localização-alocação em redes, a solução mais próxima do objetivo encontrava-se nos *Softwares gvSIG* (Comunidad gvSIG, 2013) e *Kosmo* (Kosmo, 2013) usando a extensão *Sextante* (Olaya, 2013). Esta possui uma ferramenta para análise localização-alocação que implementa os métodos mais comuns (Figura 23), no entanto, a solução é desenvolvida usando distâncias euclidianas, não permitindo o uso de redes.

⁵ Formato que guarda os dados em cadeias de caracteres separados, normalmente, por vírgulas. É um formato suportado por muitas aplicações de *software*.

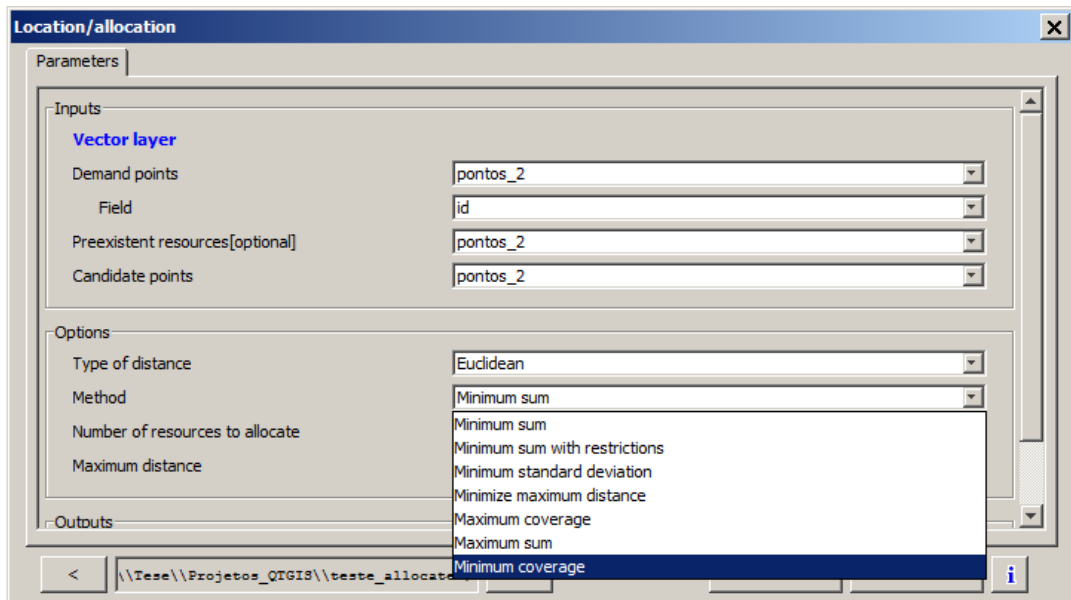


Figura 23 – Ferramenta *Sextante* localização-alocação implementado no *Software SIG* de código aberto *Kosmo*

O *Software GRASS* também possui alguns algoritmos para análise de redes, ainda que não tenha algoritmos para análise localização-alocação. De acordo com os objetivos foram testados e analisados as ferramentas: *v.net.alloc* e *v.net.iso*. A ferramenta *v.net.alloc* permite alocar áreas de serviço a pontos oferta (ver Figura 24 e legenda), a que podem ser atribuídos pesos (custos). As sub-redes obtidas correspondem às zonas da rede mais próximas de cada ponto de oferta. O custo de deslocação na rede pode-se basear na distância dos arcos da rede (custo pré-definido pela ferramenta) ou atribuir custos como tempo de viagem, consumo de combustível, etc. (GRASS MANUAL, 2013).

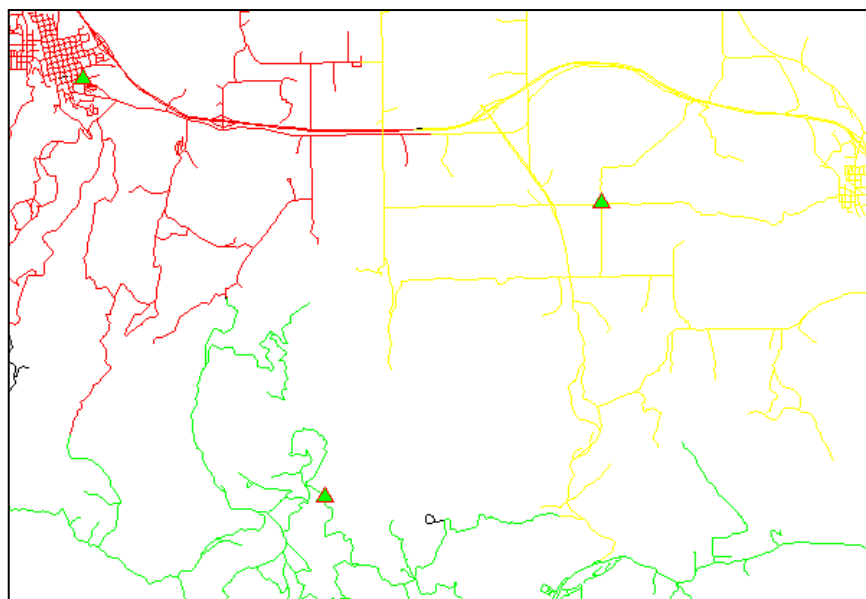


Figura 24 – Alocação de áreas de serviço a pontos de oferta. Ferramenta *v.net.alloc* do *GRASS*: pontos de oferta representados por triângulos verdes e áreas de serviço (vermelho, verde e amarelo), (GRASS MANUAL, 2013)

A ferramenta *v.net.iso* também define áreas de serviço para os pontos de oferta, só que aplica critérios diferentes. Corta as áreas de serviço em bandas delimitadas por isolinhas de custo, a partir do centro do ponto de oferta (ver Figura 18). Tal como no anterior o custo associado aos arcos pode ser personalizado (GRASS MANUAL, 2013).

De salientar que várias ferramentas *GRASS* estão integradas no *software (Quantum GIS) QGIS*, incluindo as duas apresentadas. Este *software* permite a extensão das suas funcionalidades pela instalação de *plugin* adicionados através de um repositório na Internet. Este repositório é dinamizado por uma comunidade de desenvolvedores, que criam novos *plugin* e mantêm os existentes atualizados.

Numa pesquisa mais abrangente foram identificados *software* gratuitos, nomeadamente, *Flowmap* (Universiteit Utrecht, 2013) e *S-Distance* (DPRD, 2006). Estes possuem algoritmos de análise localização-alocação, no entanto, as suas licenças de utilização são apenas para fins académicos impedindo utilização em contexto profissional. Além disso, trabalham com formatos próprios, com interface pouco amigável e não permitem uma utilização generalista, como a maior parte das soluções SIG, assim, não se optou por esta solução.

Relativamente aos restantes critérios pode-se constatar que o QGIS é a solução mais equilibrada, pois possui *plugin* para carregar serviços *WebMap (OpenLayers Plugin)*, com vários tipos de mapas: estradas, satélite e híbrido.

A exportação de dados (mapas) é possível para vários formatos, por exemplo: *JPG/TIFF/PDF*, através do compositor de impressão (ferramenta que prepara o mapa de saída) e *HTML*, usando o *plugin Html Image Map* (Duivenvoorde, 2013). Este último é muito importante, pois permite a interação com o utilizador, como por exemplo, a visualização dos atributos das entidades geográficas clicando com o rato sobre estas.

Para tornar a interface gráfica da aplicação mais amigável e com ferramentas que permitam automatizar tarefas inerentes ao projeto, o QGIS possui vários *plugin*: *Custom Launcher*, para lançar tarefas específicas ou mesmo programas externos ao QGIS; *Script Runner*, para correr algoritmos escritos em *python*; *Plugin Builder*, construção de novos *plugin*, entre outros.

Relativamente ao *design* do *software*, o QGIS permite personalizar a sua interface através da manipulação dos menus e botões visíveis (Figura 25). Esta característica é muito importante para este tipo de aplicação, pois permite dar-lhe um aspeto minimal condicionando as opções do utilizador e direcionando-o para as ações essenciais/funcionais ao projeto SIG, tendo como linha de orientação a construção de uma interface intuitiva e amigável.

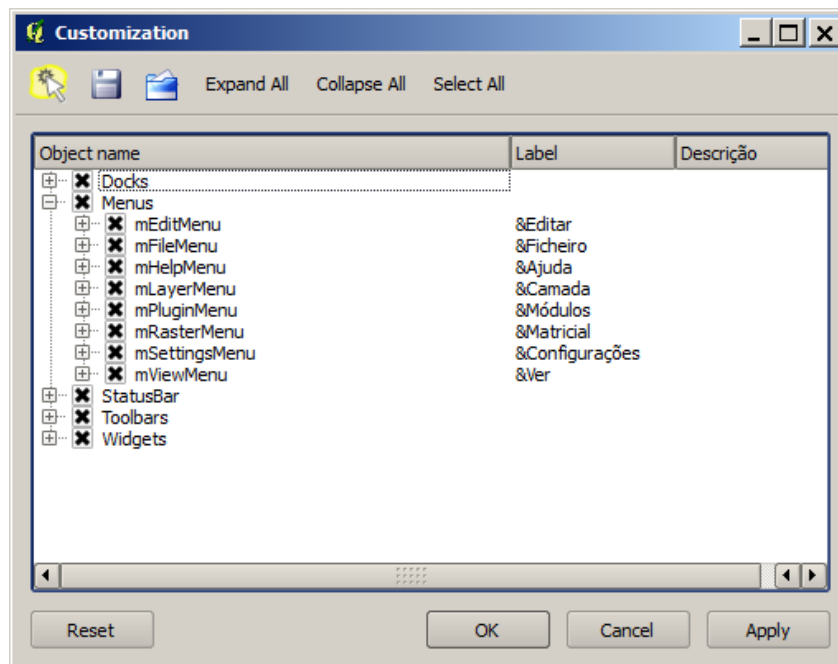


Figura 25 – Janela de personalização do QGIS

Pelos motivos expostos concluiu-se que o *software* QGIS (com ferramentas GRASS) é o que mais se adequa aos critérios de escolha definidos. A falta de algoritmos para a análise localização-alocação representa uma condicionante aos propósitos do projeto, desta forma, assumiu-se um compromisso entre esta lacuna e as restantes capacidades do *software*. Considerou-se que a otimização da localização dos locais de formação pode ser desenvolvida através da ferramenta GRASS *v.net.alloc*.

Como a empresa usa o *Microsoft Office Excel Professional* diariamente decidiu-se adotar este *software* para o tratamento estatístico dos dados na construção de indicador estatístico de necessidade de formação, a organização dos dados provenientes da Plataforma SIGO e a manipulação de dados em tabelas DBF⁶. A edição de tabelas DBF é possível após a instalação do *plugin* gratuito *ExceIToDBF* (Bonfiglio, 2009). A par destas capacidades possui características muito importantes para o projeto, pois permite a extensão de funcionalidades usando macros programadas em linguagem VBA (*Visual Basic for Applications*) e trabalha com dados alfanuméricos num número variado de formatos.

Todo o *software* vai correr sobre o sistema operativo *Windows 7 Professional*, pois é o sistema usado na empresa.

⁶ Formato de dados usado em tabelas de bases de dados. É também o formato usado para tabelas de atributos das *shapefile*

3.4. Dados

3.4.1. Fontes de dados

Uma das fontes de dados usadas é a base de dados do SIGO, que inclui a informação sobre os formandos. O SIGO permite a exportação dos dados para um ficheiro *Excel* com informação útil para o processo de geocodificação, nomeadamente: morada, código postal e localidade. Contudo os dados estão organizados em 35 colunas contendo muita informação não necessária para a geocodificação, a informação será reorganizada através da macro *Excel* “geocoding” (Anexo 2b) e guardada num ficheiro *CSV*, para que os seus registos sejam geocodificados pelo *plugin mmqgis* (Figura 22).

Para apoio ao processo de geocodificação é usada informação sobre os códigos postais da NUT III Médio Tejo com extensão de 3 dígitos. Esta encontra-se disponível para descarregar no site dos CTT em formato *TXT*, possuindo todos os códigos postais existentes no país. Os dados estão organizados em 17 campos, separados por ponto e vírgula (Tabela 4). Para o projeto interessam os campos 15 a 17.

Tabela 4 – Organização do ficheiro *TXT* com informação sobre códigos postais (Dados: CTT, 2013)

Nº Campo	Designação	Conteúdo	Tipo
1	DD	Código do Distrito	Alfa-numérico, sempre preenchido
2	CC	Código do Concelho	Alfa-numérico, sempre preenchido
3	LLLL	Código da localidade	Alfa-numérico, sempre preenchido
4	LOCALIDADE	Nome da localidade	Alfa-numérico, sempre preenchido
5	ART_COD	Código da Artéria	Alfa-numérico
6	ART_TIPO	Artéria - Tipo (Rua, Praça, etc)	Alfa-numérico
7	PRI_PREP	Primeira preposição	Alfa-numérico
8	ART_TITULO	Artéria - Título (Doutor, Eng.º, Professor, etc)	Alfa-numérico
9	SEG_PREP	Segunda preposição	Alfa-numérico
10	ART_DESIG	Artéria - Designação	Alfa-numérico
11	ART_LOCAL	Artéria - Informação do Local/Zona	Alfa-numérico
12	TROÇO	Descrição do troço	Alfa-numérico
13	PORTA	Número da porta do cliente	Alfa-numérico, vazio para códigos postais geográficos(1)
14	CLIENTE	Nome do cliente	Alfa-numérico, vazio para códigos postais geográficos (1)
15	CP4	N.º do código postal	Alfa-numérico, sempre preenchido
16	CP3	Extensão do n.º do código postal	Alfa-numérico, sempre preenchido
17	CPALF	Designação Postal	Alfa-numérico, sempre preenchido

Para identificar a localização ótima dos locais de formação é necessário estar definida uma rede, através da qual se fazem as deslocações de formandos e formadores. Como os formandos normalmente se deslocam em transporte próprio considerou-se que a rede poderia considerar apenas vias para uso de transporte particular. Sendo um dos objetivos o uso de dados gratuitos, optou-se pela rede de estradas disponível no *OpenStreetMap*, e que pode ser descarregada para formato *ESRI shapefile* (OpenStreetMap, 2013). A *shapefile* encontra-se no sistema de referência *WGS 84* e com sistema de coordenadas geográficas, não tendo projeção cartográfica definida. Trata-se de uma *shapefile* de linhas com os atributos: *osm_id* (código da linha), *name* (nome da via), *ref* (designação da via), *type* (tipo de via), *oneway* (sentido único ou dois sentidos: 1 e 0 respetivamente), *bridge* (ponte: 1=sim; 0=não), *tunnel* (túnel: 1=sim; 0=não) e *maxspeed* (velocidade máxima permitida), (Figura 26).

Attribute table - roads :: 0 / 320667 feature(s) selected								
	osm_id	name	ref	type	oneway	bridge	tunnel	maxspeed
0	2954258	Rua da Lapa	NULL	residential	0	0	0	NULL
1	3979151	Rua dos Pardais	NULL	residential	0	0	0	NULL
2	3979152	NULL	N 247	secondary	0	0	0	NULL
3	3979153	Travessa da Escola	NULL	residential	0	0	0	NULL
4	3979154	Rua de Baixo	NULL	residential	0	0	0	NULL
5	3979155	Rua da Quinta	NULL	steps	0	0	0	NULL

Figura 26 – Tabela de atributos da rede de estradas *OpenStreetMap* (OpenStreetMap, 2013)

Os locais de formação representam instituições onde é possível ministrar ações de formação (ex.: juntas de freguesia). Estes foram coletados para um ficheiro *TXT*, através da pesquisa no *GoogleMaps* pelas juntas de freguesia existentes na NUT III Médio Tejo. Neste ficheiro a informação foi organizada em três campos separados por ponto e vírgula: Latitude, Longitude e Nome (Figura 27), num total de 105 registos. Na freguesia onde se localiza a empresa, a junta de freguesia foi substituída pelas instalações da ATEP.

locais_formacao_inicial.txt - Bloco de notas		
Ficheiro	Editar	Formatar
Ver	Ajuda	
Latitude,	Longitude,	Nome
39.55423,	-7.99635,	Junta_Freg_Macao
39.58579,	-8.05634,	Junta_Freg_Aboboreira
39.65910,	-8.06636,	Junta_Freg_Amendoa
39.54262,	-8.03835,	Junta_Freg_Penhascoso
39.66653,	-7.99079,	Junta_Freg_Cardigos
39.62256,	-7.92237,	Junta_Freg_Carvoeiro
39.56853,	-7.87470,	Junta_Freg_Envendos
39.48432,	-8.02013,	Junta_Freg_Ortiga
39.54256,	-8.27152,	Junta_Freg_Aldeia_do_Mato

Figura 27 – Ficheiro *TXT* com dados sobre a localização dos locais de formação

No SIG-ATEP a construção de mapas temáticos é feita a partir de variáveis estatísticas dos Censos 2011 e cartografia com limites administrativos de, concelhos, freguesias e subsecções estatísticas. Para este fim utilizou-se a *shapefile* disponibilizada pelo INE com os limites das subsecções estatísticas (INE, 2013) e pela *shapefile* com a CAOP 2012.1 (Carta Administrativa Oficial de Portugal versão 2012.1) disponibilizada pela Direção Geral do Território (DGT) para os limites dos concelhos e freguesias (DGT 2012). Estas *shapefile* encontram-se no sistema de referência *ETRS 89*, com projeção *Transverse Mercator (TM06 Portugal)*.

As *shapefile* referidas no parágrafo anterior são compostas por polígonos com vários atributos, representando sobretudo códigos das unidades administrativas. A Figura 28 mostra os atributos existentes na tabela da *shapefile* proveniente do INE, como os limites das subsecções estatísticas: DTMN11 (identificador único de concelho); FR11 (identificador de freguesia no concelho); SEC11 (identificador de secção estatística na freguesia); SS11 (identificação da subsecção estatística na secção estatística); BGRI11 (Base Geográfica de Referência da Informação 2011 que identifica de forma unívoca a subsecção estatística); LUG11 (identificador de lugar INE); LUG11DESIG (nome de lugar INE), (INE, 2013).

	DTMN11	FR11	SEC11	SS11	BGRI11	LUG11	LUG11DESIG
0	1401	04	004	12	14010400412	999999	Residual
1	1401	16	002	02	14011600202	023290	Vale das Mós
2	1401	04	004	01	14010400401	023254	Baralho
3	1401	16	001	03	14011600103	023290	Vale das Mós
4	1401	16	001	02	14011600102	023290	Vale das Mós
5	1401	16	002	01	14011600201	023290	Vale das Mós

Figura 28 – Tabela de atributos da *shapefile* Censos 2011 (INE, 2013)

A CAOP 2012 também possui vários atributos para cada entidade geográfica (Figura 29): DICOFRE (identificador único de freguesia); FREGUESIA (nome da freguesia); MUNICÍPIO (nome do Concelho); GEO_COD (identificador único de concelho); DISTRITO (nome do distrito); TAA (identificação do tipo de área administrativa); AREA_EA_HA (valor área da área administrativa); AREA_T_HA (valor total da área da freguesia), (DGT, 2012).

	DICOFRE	FREGUESIA	MUNICIPIO	GEO_COD	DISTRITO	TAA	AREA_EA_HA	AREA_T_HA
39	141001	SÃO JOÃO BAPT...	ENTRONCAMENTO	1410	SANTARÉM	ÁREA PRINCIPAL	456.05	456.05
40	141002	NOSSA SENHOR...	ENTRONCAMENTO	1410	SANTARÉM	ÁREA PRINCIPAL	916.76	916.76
41	140801	CONSTÂNCIA	CONSTÂNCIA	1408	SANTARÉM	ÁREA PRINCIPAL	883.15	883.15
42	141916	ZIBREIRA	TORRES NOVAS	1419	SANTARÉM	ÁREA PRINCIPAL	1050.52	1050.52
43	140802	MONTALVO	CONSTÂNCIA	1408	SANTARÉM	ÁREA PRINCIPAL	1281.44	1281.44
44	140208	MONSANTO	ALCANENA	1402	SANTARÉM	ÁREA PRINCIPAL	1846.34	1846.34

Figura 29 – Tabela de atributos da *shapefile* CAOP 2012 (DGT, 2012)

Os atributos mais importantes para o projeto são os que identificam univocamente as entidades geográficas, nomeadamente: BGRI11 para identificar as subsecções estatísticas na *shapefile* do INE; DICOFRE e MUNICÍPIO, para identificar freguesia e concelho respetivamente, na CAOP.

Juntamente com a *shapefile* do INE foi descarregado um ficheiro *CSV* com os dados estatísticos dos Censos 2011 (resultados definitivos), distribuídas por 122 variáveis. Estas baseiam-se em temas sociais, económicos e de educação, estando disponíveis para os níveis do concelho, freguesia e subsecção estatística. A Figura 30 apresenta um excerto da referida tabela onde podemos ver algumas colunas, inclusive, o código identificador da entidade geográfica (GEO_COD) que permite transferir os dados estatísticos das tabelas *CSV* para as tabelas de atributos das *shapefile*. Para uma compreensão dos nomes das colunas existe o ficheiro *CSV* “C2011_FSINTESE_VARIAVEIS” (disponível na pasta “dados_estatisticos” no CD-ROM fornecido) onde podemos encontrar a descrição de cada um.

	A	B	C	D	E
1	ANO	GEO_COD	GEO_COD_DSG	NIVEL	NIVEL_DSG
2	2011	'PT	PT		1 Total Nacional
3	2011	'1	Continente		2 NUT1
4	2011	'16	Centro		3 NUT2
5	2011	'16C	Médio Tejo		4 NUT3
6	2011	'1421	Ourém		5 Municipio
7	2011	'142101	Alburitel		6 Freguesia
8	2011	'142102	Atouguia		6 Freguesia
9	2011	'142103	Casal dos Bernard		6 Freguesia

Figura 30 – Tabela *CSV* com dados estatísticos dos Censos 2011 (INE, 2013)

Irão também ser utilizados serviços *WebMap* que permitem a visualização de cartografia no QGIS (ex.: *Google Maps*, *OpenStreetMap*, etc.). Estes estão disponíveis através do *plugin* QGIS *OpenLayers* e encontram-se no Sistema de Referência *WGS 84*, com projeção *Pseudo Mercator (EPSG:3857)*, (Alastair Aitchison, 2011).

3.4.2. Tratamento dos dados

No sentido de poder integrar os dados no SIG-ATEP foi necessário proceder ao tratamento de alguns dos dados usados.

Para apoio ao processo de geocodificação foi necessário preparar o ficheiro *TXT* com os códigos postais. Visto que interessavam apenas os códigos postais da NUT III Médio Tejo procedeu-se à sua seleção no ficheiro *TXT*. Estes foram copiados para um novo ficheiro *TXT* segundo o

código de concelho (campo 2 da Tabela 4), tarefa realizada para os 11 Concelhos que compõem a NUT III. De seguida, os dados foram reorganizados de forma a manter apenas os campos 15 a 17 (Tabela 4). Com esta operação os dados estavam disponíveis para serem usados na macro *Excel* “geocoding” (Anexo 2b) apresentada no ponto 3.6.

A rede de estradas proveniente do *OpenStreetMaps* foi restringida à área de estudo selecionando as estradas correspondentes e exportando a seleção para uma nova *shapefile*. A rede abrangida passou um pouco os limites da NUT III Médio Tejo de maneira a garantir a sua conectividade junto dos limites (Figura 31).

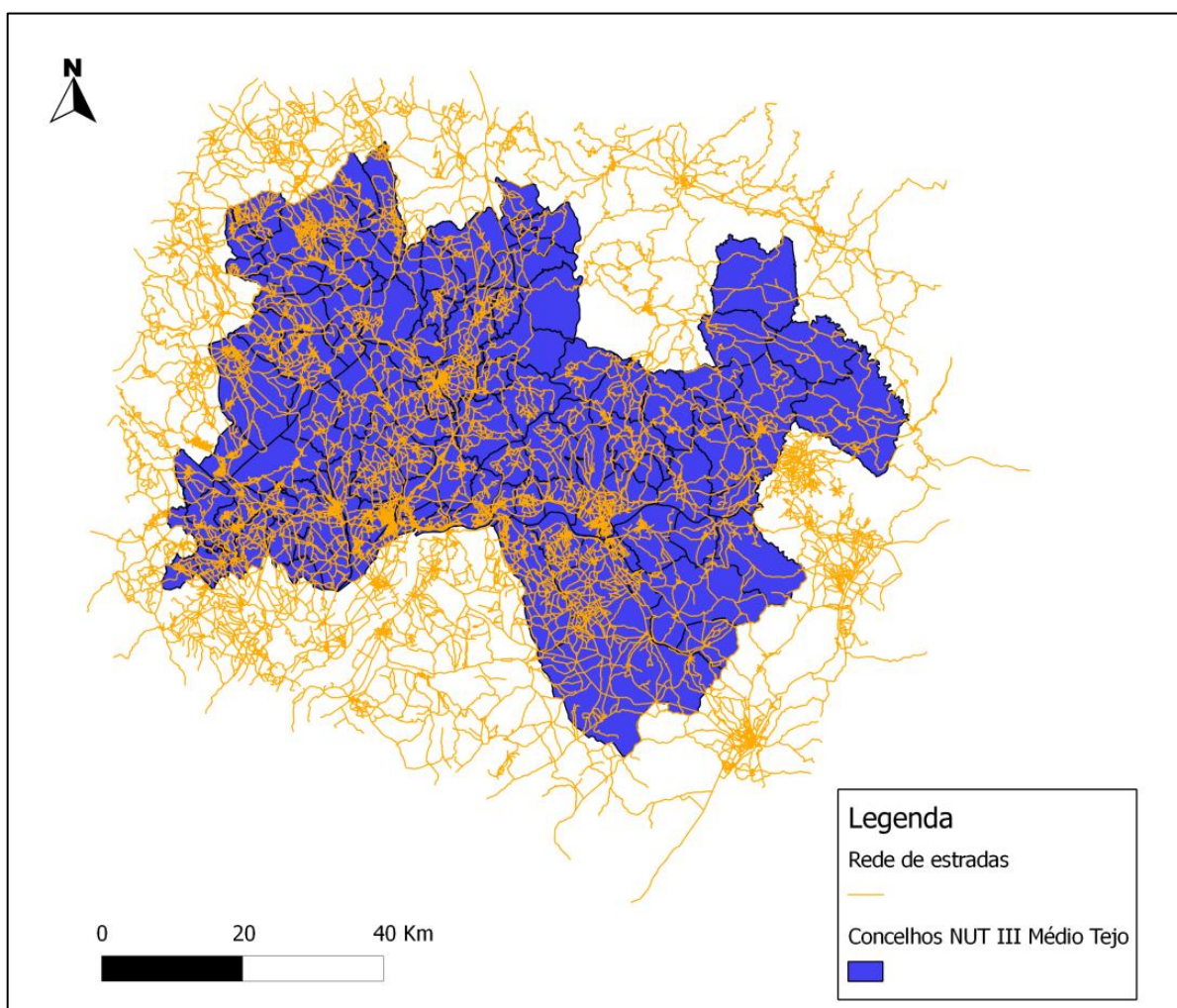


Figura 31 – Rede de estradas usada no projeto (Dados: OpenStreetMaps, 2013; CAOP, 2012)

Analisou-se de forma cuidada a qualidade geométrica da rede e verificou-se a existência de muitas linhas sem conectividade e linhas sem quebras nas interseções. Para garantir a conexão entre linhas, nas interseções, foi necessário executar a ferramenta *GRASS v.clean.break* que quebra as linhas em cada interseção. Como não se pretendia quebrar vias rápidas como autoestradas (ex.: A1) ou itinerários complementares (ex.: IC9), visto que estas têm ligação com as restantes vias

apenas em nós (saídas) construídos para o efeito, foram realizadas consultas ao atributo *type* (Figura 26) e criada *shapefile* só com essas vias. Depois, foram unidas as duas redes quebrando-se manualmente as vias onde intersectavam os nós das vias rápidas, conforme exemplo mostrado na Figura 32.

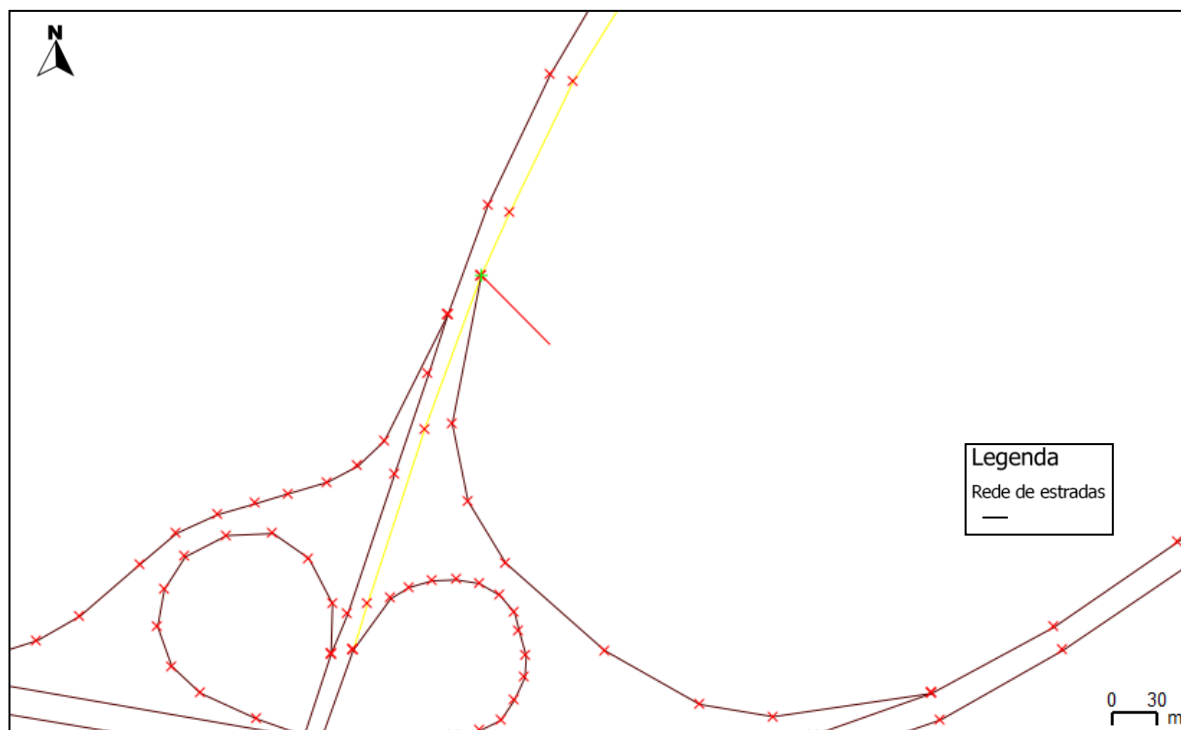


Figura 32 – Quebra manual (traço a vermelho) das intersecções com vias rápidas (Dados: OpenStreetMaps, 2013)

Para verificar se existiam linhas desconectadas da rede utilizou-se a ferramenta de alocação de redes *v.net.alloc*. Para poder usá-la criaram-se alguns pontos “teste” (simulação de locais de formação) que foram conectados à rede, através da ferramenta *GRASS v.net* (ferramenta que permite conectar pontos à rede). Pela análise do resultado verificou-se que partes da rede tinham sido ignoradas pela ferramenta *v.net.alloc*, devido a dois tipos de erro: linhas desconectadas e linhas que não quebraram pela aplicação da ferramenta *v.clean.break*. A correção destes erros foi feita manualmente, quebrando linhas e conectando linhas mal digitalizadas.

Com a rede consistente do ponto de vista geométrico foi avaliada a qualidade temática. Os atributos *name*, *ref* e *maxspeed*, possuíam muitos registos sem valor (*NULL*), o que poderia originar problemas. Desta forma, os atributos *name* e *ref* foram eliminados na tabela e no atributo *maxspeed* foram eliminados os seus registos. Manteve-se este último atributo porque será necessário para o cálculo do tempo de viagem. O atributo *type* não possuía valores nulos e encontrava-se coerente com a geometria. Através da inspeção de algumas linhas verificou-se que a classificação do tipo de via (*type*) estava correta, desta forma mantiveram-se os nomes do tipo de via originais (inglês).

De acordo com os objetivos o custo da rede será determinado pelo tempo de viagem, deste modo, é necessário preencher todos os registos no atributo *maxspeed*. Foram identificadas as velocidades máximas permitidas e associadas a cada tipo de via, segundo o atributo *type* (Tabela 5). As velocidades estão conforme a PRP (Prevenção Rodoviária Portuguesa), para as rodovias, nos caminhos pedestres e ciclovias foram adotadas as velocidades médias, 6 e 20 respetivamente (Wikipédia, 2013).

Tabela 5 – Velocidades máximas permitidas de acordo com o tipo de via (Dados de velocidade: PRP, 2013 e Wikipédia, 2013)

Tipo de via	Velocidade máxima (Km/h)
<i>tertiary</i>	70
<i>tertiary_link</i>	50
<i>residential</i>	40
<i>unclassified</i>	50
<i>trunk</i>	90
<i>trunk_link</i>	50
<i>primary</i>	80
<i>motorway</i>	120
<i>motorway_link</i>	50
<i>unclassified</i>	50
<i>primary_link</i>	50
<i>cycleway</i>	20
<i>pedestrian</i>	6
<i>rest_area</i>	20
<i>construction</i>	50
<i>road</i>	50
<i>living_street</i>	40
<i>secondary_link</i>	50
<i>secondary</i>	70
<i>service</i>	40
<i>track</i>	50

Para a atribuição da velocidade máxima foi programada a macro *Excel* “velocidade_max” (Anexo 2a) que permite a atribuição da velocidade máxima a cada entidade geográfica de acordo com o atributo *type*, esta tarefa é realizada através da edição da tabela *DBF* da *shapefile*.

Para calcular o tempo de viagem é necessário determinar o comprimento das vias. Através da “calculadora de campo” do *QGIS* disponível na tabela de atributos, foi calculado o comprimento das linhas com criação de um novo atributo (*COMP*). Como mostra a Figura 33 o comprimento está disponível na função *length* do grupo da geometria. Ao aplicar esta função são adicionados valores ao novo atributo, correspondentes ao comprimento de cada entidade geográfica.

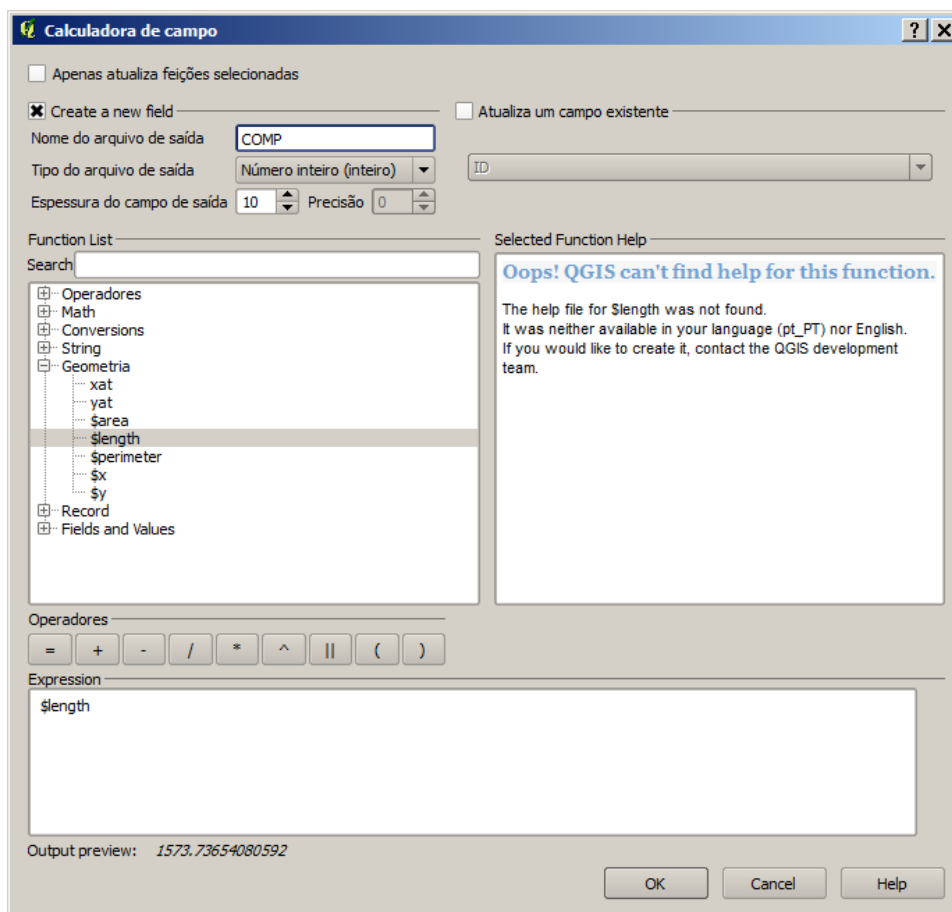


Figura 33 – Calculadora de campo QGIS

A fórmula o tempo de viagem (3.2) foi deduzida a partir da fórmula da velocidade (3.1). A implementação da fórmula foi idêntica ao caso anterior, criou-se um novo campo (*TEMPO_VIAGEM*) e em “Expression” (Figura 33) escreveu-se a fórmula (3.2). Para Δs entraram os valores do atributo *COMP* e para v os valores do atributo *maxspeed*. Como os tempos de viagem normalmente não ultrapassam 1hora o resultado da fórmula foi multiplicado por 60, de forma a ter o tempo de viagem em minutos.

$$v = \frac{\Delta s}{\Delta t} \quad (3.1)$$

$$\Delta t = \frac{\Delta s}{v} \quad (3.2)$$

v (velocidade máxima); Δs (Comprimento da linha); Δt (Tempo de Viagem)

Os atributos: *osm_id*, *oneway*, *bridge* e *tunnel* não serão usados na análise, no entanto, mantêm-se na *shapefile* devido à coerência de valores (sem valores nulos), apenas se alterou o seu nome para maiúsculas. O atributo *osm_id*, foi renomeado para ID.

O ficheiro *TXT* com dados relativos aos locais de formação foi convertido para *shapefile* de pontos através da ferramenta QGIS *Add Delimited Text Layer*. Na Figura 34 apresentam-se os atributos da *shapefile*: *LATITUDE*, *LONGITUDE*, *NOME* e *ESTADO*, este último atributo é editado aquando a análise SIG, nomeadamente pela ferramenta *GRASS v.net.alloc*. Nesta ferramenta, quando o local de formação é considerado na análise deve ter valor de custo “0” e quando não é considerado deve ter valor “-1”.

	LATITUDE	LONGITUDE	NOME	ESTADO
0	39.55422999...	-7.99635	Junta Freg Macao	0
1	39.58579	-8.05634	Junta Freg Aboboreira	0
2	39.6591	-8.06636	Junta Freg Amendoa	0
3	39.54261999...	-8.0383499...	Junta Freg Penhascoso	0
4	39.66653	-7.99079	Junta Freg Cardigos	0
5	39.62256	-7.92237	Junta Freg Carvoeiro	0
6	39.56853	-7.8747	Junta Freg Envendos	0
7	39.48431999...	-8.02013	Junta Freg Ortiga	0
8	39.54256	-8.27152	Junta Freg Aldeia do Mato	0

Figura 34 – Tabela de atributos da *shapefile* com locais de formação

As *shapefile* CAOP 2012 e Censos 2011 foram editadas de maneira a constituir novas *shapefile* com concelhos, freguesias e subsecções estatísticas da NUT III Médio Tejo. Na CAOP 2012 procedeu-se à seleção dos 11 Concelhos que compõem a NUT III e guardou-se a seleção para uma nova *shapefile*. Na *shapefile* Censos 2011 o procedimento teve de ser diferente, porque cada *shapefile* que se descarrega através do site do INE corresponde a um concelho. Assim, foi efetuada a união de 11 *shapefile* através da ferramenta QGIS: *Mescla arquivos shape em um*.

As tabelas *CSV* com os dados estatísticos também são descarregadas por concelho, deste modo, tiveram que ser copiadas para um ficheiro único. Em conjunto com os responsáveis da empresa selecionaram-se 16 variáveis, consideradas úteis para os utilizadores do SIG-ATEP. Foram criados 16 atributos em ambas as *shapefile* e atribuído um código a cada variável, pois o QGIS não permite nomear um atributo com mais de 10 caracteres. O código e descrição das variáveis apresentam-se na Tabela 6.

Como os valores quantitativos e qualitativos do indicador estatístico (determinado no ponto 3.9.) também serão transferidos para estas *shapefile*, foram criados mais dois atributos, *Z_SCORE* e *SCORE_CLAS*, para armazenar valores quantitativos e qualitativos, respetivamente.

Tabela 6 – Dados estatísticos inseridos nas *shapefiles* (concelhos, freguesias, subsecção estatística). (Dados: INE, 2013)

Código	Descrição da variável estatística (INE)
PRESENTES	<i>Total de indivíduos presentes</i>
RESIDENTES	<i>Total de indivíduos residentes</i>
RESIDENTES_H	<i>Total de homens residentes</i>
RESIDENTES_M	<i>Total de mulheres residentes</i>
RESI_25_64	<i>Indivíduos residentes com idade entre 25 e 64 anos</i>
RESID_64	<i>Indivíduos residentes com idade superior a 64 anos</i>
N_ANALF	<i>Indivíduos residentes sem saber ler nem escrever</i>
RESID_1BAS	<i>Indivíduos residentes com o 1º ciclo do ensino básico completo</i>
RESID_2BAS	<i>Indivíduos residentes com o 2º ciclo do ensino básico completo</i>
RESID_3BAS	<i>Indivíduos residentes com o 3º ciclo do ensino básico completo</i>
RESID_SEC	<i>Indivíduos residentes com o ensino secundário completo</i>
RESID_PSEC	<i>Indivíduos residentes com o ensino pós-secundário</i>
RESID_SUP	<i>Indivíduos residentes com um curso superior completo</i>
RESI_DES1	<i>Indivíduos residentes desempregados à procura do 1º emprego</i>
RESID_DESN	<i>Indivíduos residentes desempregados à procura de novo emprego</i>
RESID_EMP	<i>Indivíduos residentes empregados</i>
RESID_REF	<i>Indivíduos residentes pensionistas ou reformados</i>
RESID_S_AC	<i>Indivíduos residentes sem atividade económica</i>
PRIMARIO	<i>Indivíduos residentes empregados no sector primário</i>
SECUNDAR	<i>Indivíduos residentes empregados no sector secundário</i>
TERCIARIO	<i>Indivíduos residentes empregados no sector terciário</i>

A transferência dos dados foi efetuada através do *Excel* usando colunas comuns nos ficheiros *DBF* e no ficheiro *CSV* que contém as estatísticas. Tendo em conta que as *shapefile* possuem uma ordenação diferente do *CSV* foi criada uma coluna temporária nas tabelas *DBF*, composta por uma série de números (1,2,3,..., n). Esta operação permite salvaguardar a ordenação original da *shapefile*. Para a transferência de dados procedeu-se à ordenação (do mais pequeno para o maior) de acordo com os valores das colunas comuns: GEO_COD (concelhos); DICOFRE (freguesias) e BGRI11 (Subsecção estatística). Os dados estatísticos foram copiados da tabela *CSV* para as tabelas *DBF*, respeitando o código das entidades geográficas e por fim procedeu-se à sua reordenação, de acordo com a coluna de dados temporária.

Como irão ser utilizados serviços *WebMap* é pertinente que toda a informação esteja Sistema de Referência usado por estes serviços (ver ponto 3.4.1), pois é o único que não se consegue alterar. Deste modo, todas as *shapefile* do projeto SIG-ATEP serão guardadas no sistema de referência usado por este tipo de serviços: *WGS 84*, com projeção *Pseudo Mercator (EPSG:3857)*, (Alastair Aitchison, 2011). A conversão de sistemas de referência foi realizada fazendo “guardar como” a cada *shapefile* e definindo o sistema pretendido na nova *shapefile*. Optou-se por este método porque a ferramenta *QGIS Definir projeção atual* não produzia efeito sobre as *shapefile*.

3.4.3. Esquema da Base de Dados

O desenho do esquema para a base de dados foi feito em consonância com as características do *software* QGIS e GRASS. Como o GRASS não lê diretamente *shapefile* foi criado um *mapset* que é uma pasta usada pelo GRASS para armazenamento de ficheiros GRASS. O *mapset* permite armazenar dados tipo matricial e vetorial, pode-se dizer que a sua função é análoga à *geodatabase* usada pela ESRI.

Tal como as *shapefile* os ficheiros GRASS são compostos por uma tabela DBF e ficheiros com a geometria das entidades geográficas. Num *mapset* as tabelas DBF encontram-se ligadas entre si estabelecendo relações de cardinalidade.

No QGIS a manipulação destes ficheiros é feita através de um navegador GRASS que permite eliminar ficheiros, adicionar ficheiros ao gestor de camadas do QGIS, renomear ficheiros e copiar ficheiros. Para realizar outro tipo de operações o utilizador tem de fazer uso das ferramentas GRASS. Por exemplo, para importar *shapefile* para o *mapset* o utilizador tem de usar a ferramenta GRASS *v.in.ogr*.

Devido ao exposto nos parágrafos anteriores os dados do SIG-ATEP foram armazenados em duas pastas, representando a base de dados SIG-ATEP (ver Figura 35):

- A pasta **“Dados”**, que contém as *shapefile* descritas no ponto 3.4.2., a *shapefile* com o resultado da geocodificação e várias *shapefile* com o resultado das ferramentas GRASS;
- A pasta **“ATEP”**, *mapset* que contém ficheiros GRASS resultantes da importação de *shapefile* e da análise SIG: locais de formação, estradas (rede de estradas *OpenStreetMap*), formandos (formandos geocodificados), áreas de serviço (resultado da ferramenta *v.net.alloc*), rede iso (resultado da ferramenta *v.net.iso*).

No SIG-ATEP utilizador não tem acesso à pasta “ATEP”, pois os seus dados são para uso exclusivo das ferramentas GRASS que materializam a análise SIG.

A pasta “Dados” é uma pasta que o utilizador pode editar, serve como pasta de armazenamento de *shapefile* necessárias à análise SIG e pasta de destino dos resultados produzidos pela análise SIG. Ao contrário da pasta “ATEP” na pasta “Dados” os ficheiros não possuem ligações entre si, pois esta funciona como um repositório de dados.

No Anexo 1 (dicionário de dados) encontra-se uma descrição pormenorizada dos ficheiros (tabelas e atributos) envolvidas no projeto SIG-ATEP.

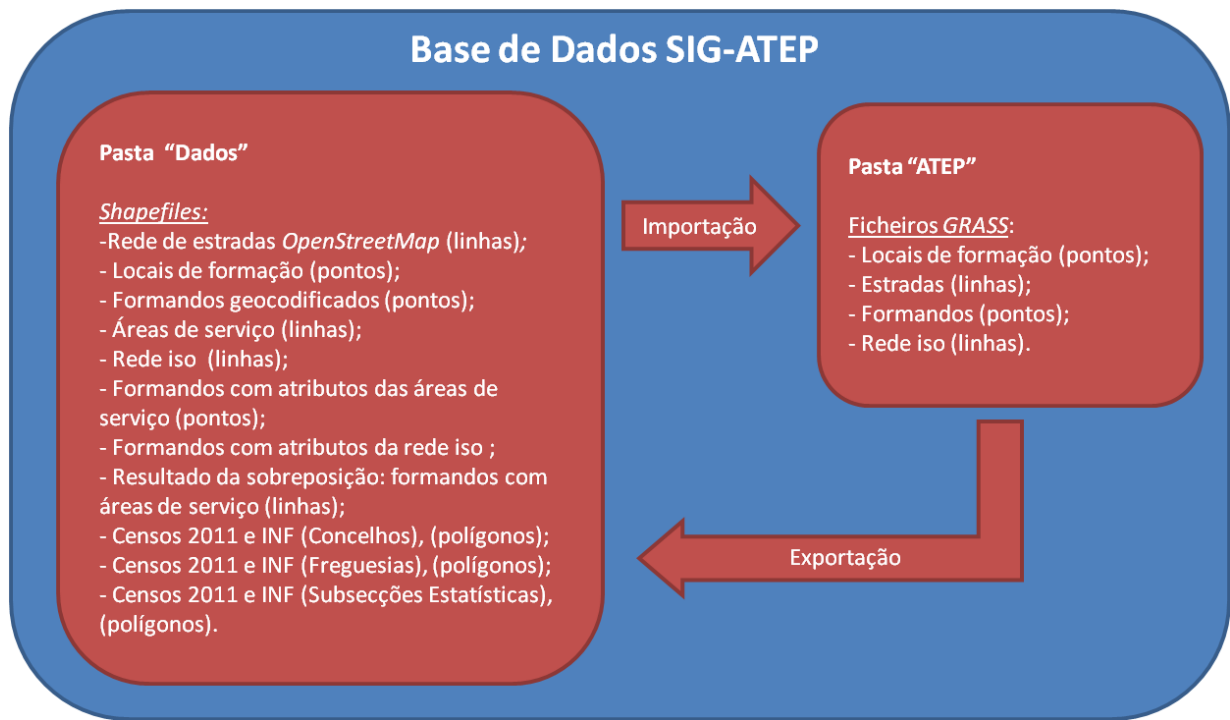


Figura 35 – Esquema da base de dados SIG-ATEP

Na Figura 36 apresenta-se o DER com as tabelas que compõem a pasta "ATEP", mostrando relações estabelecidas e atributos de cada uma. A relação entre as tabelas locais de formação e áreas de serviço tem, cardinalidade de *um para um (1:1)*, isto é, um local de formação só pode pertencer a uma área de serviço e uma área de serviço só pode estar alocada a um local de formação. As restantes relações têm cardinalidade de *um para muitos (1:N)*, por exemplo, uma estrada pode estar ligada a muitos formandos, mas um formando só pode estar ligado a uma estrada. Pode existir casos em que a relação não exista, por exemplo, uma estrada pode não ter formandos conectados. Nestes casos a relação é simbolizada com um círculo, do lado em que a relação possa ser nula. As relações obrigatórias são simbolizadas por um traço, por exemplo, um formando tem de estar sempre conectado com uma estrada.

A chave primária é representada pelo atributo *cat* porque é a terminologia adotada pelo GRASS sempre que se importa uma *shapefile* para o *mapset*. A descrição dos atributos pode ser consultada no dicionário de dados (Anexo 1).

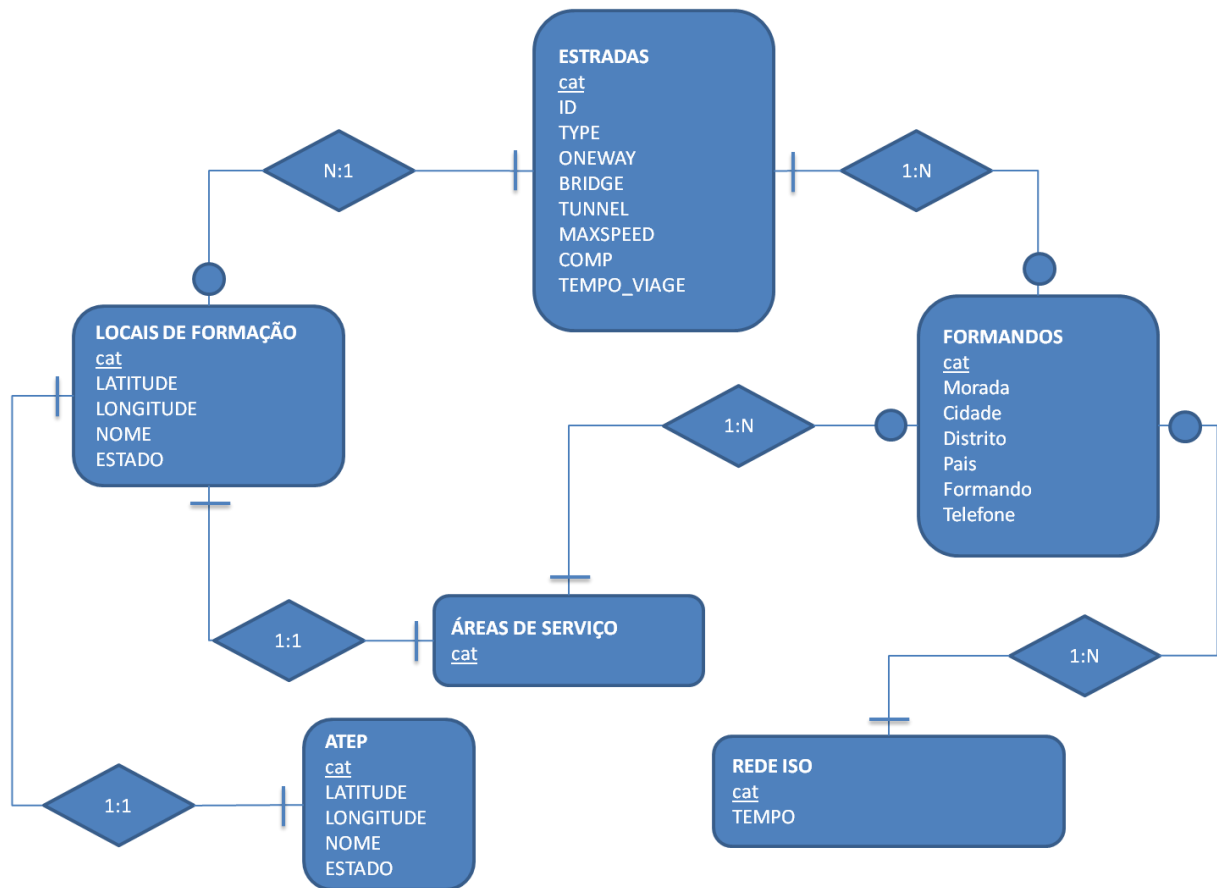


Figura 36 – DER do dos ficheiros da pasta “ATEP”

3.5. Desenho do SIG-ATEP

O desenho da estrutura de funcionamento do projeto pode ser considerada como parte do *design* do Sistema, pois explicita as ferramentas e dados a utilizar. Na Figura 37 são apresentados os dados e ferramentas necessárias à implementação do projeto. As caixas verdes representam os dados de entrada: dados sobre os formandos, rede de estradas, locais de formação, limites administrativos, dados dos Censos 2011 e INF. As caixas vermelhas representam os dados de saída: cartografia temática; locais de formação (locais de formação seleccionados pela análise); formandos geocodificados; áreas de serviço; rede iso e ficheiros com resultados da união entre *shapefiles*. As caixas azuis representam as ferramentas necessárias à análise SIG. Todas as tarefas necessárias ao funcionamento do SIG-ATEP estão descritas no manual de utilização.

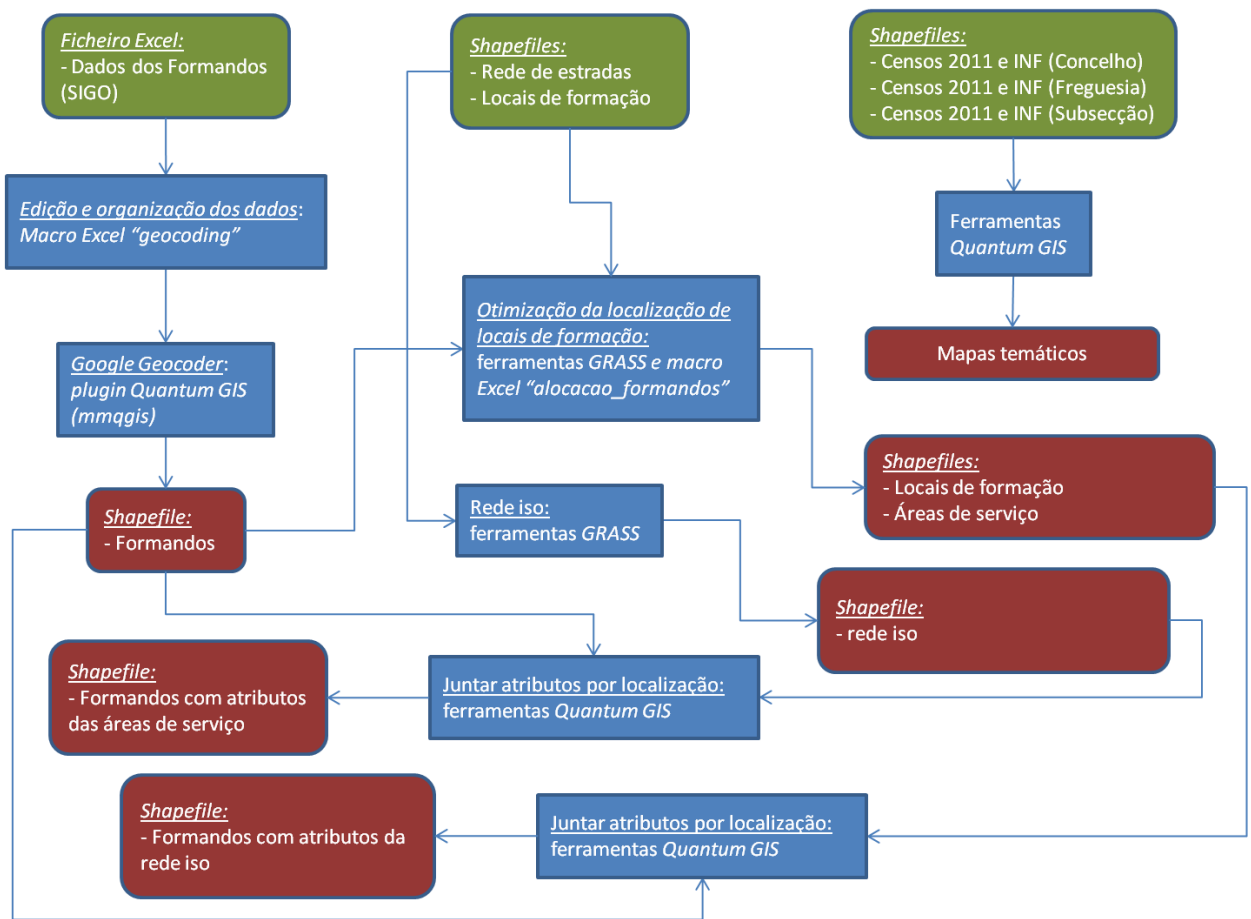


Figura 37 – Estrutura de funcionamento do SIG-ATEP

Para auxiliar o utilizador foi implementada uma barra de ferramentas através do *plugin* QGIS: *Custom Launcher* (Denis, 2012). Este *plugin* permite adicionar botões a uma barra de ferramentas criada na interface gráfica do QGIS. Os botões adicionados podem lançar várias ações, tais como: programas, explorador do *Windows*, algoritmos *GRASS*, ficheiros, etc.

A barra de ferramentas construída para o SIG-ATEP possui uma sequência de processos que materializam toda a análise SIG do projeto (Figura 38). Apenas a cartografia temática não é construída através desta barra de ferramentas.

No primeiro botão, com logotipo da empresa, é aberto um ficheiro com ajuda, isto é, abre uma imagem com instruções para o uso da barra de ferramentas, os botões seguintes, enumerados de 1 a 9, representam uma sequência lógica de processos que nos conduzem através de toda a análise SIG, desde a importação de ficheiro com os formandos (1), instruções para geocodificação (2), localização otimizada de locais de formação (3, 4 e 5), exportação de dados (6), instruções para união de *shapefiles* (7 e 9), construção da rede iso (8). O seu uso deverá ser acompanhado pela consulta do manual de utilização, pois existem especificidades da análise que não são descritas nas instruções lançadas pelos botões desta barra.

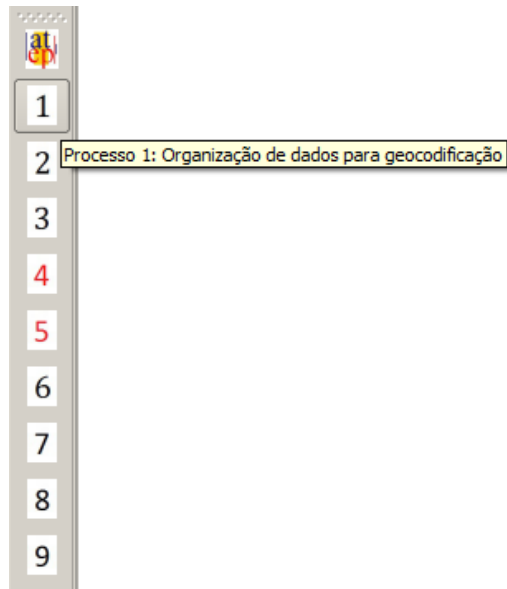


Figura 38 – Barra de ferramentas ATEP construída através do *plugin QGIS Custom Launcher*

No sentido de melhorar a usabilidade do QGIS procedeu-se à personalização da sua interface gráfica (ver Figura 39). Ficaram apenas visíveis as ferramentas relacionadas com a manipulação do mapa no ambiente de trabalho (*zoom, pan, zoom to selection*, etc), o gestor de camadas, ficheiro, impressão, medição de distâncias, anotações, seleção, identificação de elementos, importação de camadas vetoriais, exportação de mapas para *HTML* e a barra de ferramentas ATEP. Os menus mantiveram-se todos porque o utilizador pode necessitar de utilizar ferramentas adicionais.

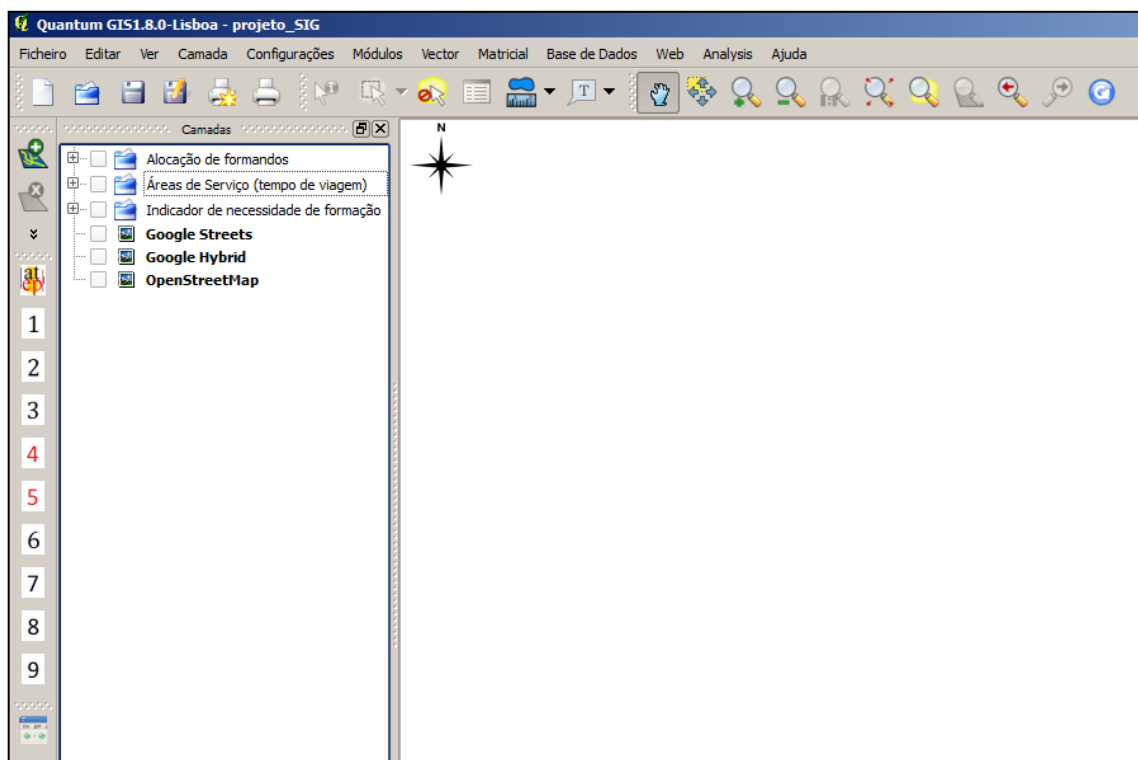


Figura 39 – Interface gráfica do QGIS personalizada para o SIG-ATEP

3.6. Geocodificação de formandos

Para operacionalizar a geocodificação o utilizador terá que usar dois botões da barra de ferramentas ATEP: **botão 1** para lançar a macro *Excel* “geocoding” e **botão 2** para visualizar instruções de uso do *plugin* de geocodificação *mmqgis*. De seguida, descreve-se o modo de funcionamento da macro “geocoding” e *plugin mmqgis*.

Como os dados provenientes do SIGO tem de ser reorganizados, de modo a serem usados pelo *plugin mmqgis*, foi construída a macro “geocoding” (Anexo 2b). Esta organiza e exporta os dados para o formato *CSV*.

Ao lançar a macro “geocoding” surge uma interface amigável construída através de *Windows Forms* (Figura 40). Esta possui vários e 3 botões: “Importar Dados”, permite a importação do ficheiro *XLSX* (proveniente do SIGO); “Organizar Dados”, organização dos dados para geocodificação; “Exportar Dados”, exportação dos dados para *CSV*.

Na organização dos dados a macro executa várias tarefas automáticas, nomeadamente: redução do número de colunas, de 35 para 6 (Morada, Cidade, Distrito, País, Nome; Telefone); substituição do código postal pelo nome da localidade a que pertence; remoção de elementos não desejáveis: acentos, pontos, “ç”, etc. Esta preparação é necessária devido à organização/formatação imposta pelo *plugin mmqgis* (Figura 22).

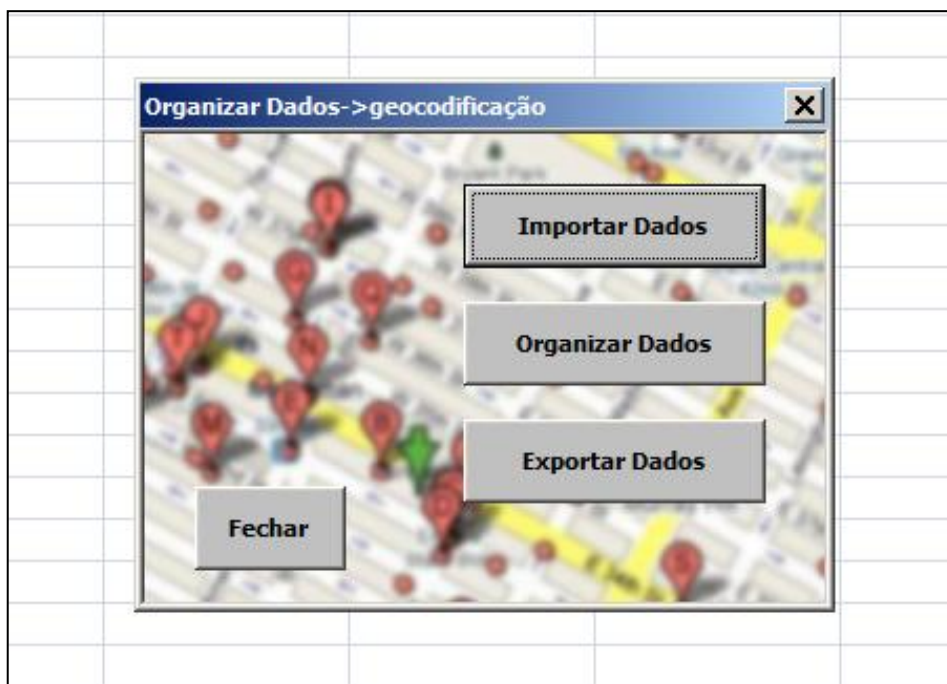


Figura 40 – Interface gráfica da macro *Excel* “geocoding” para apoio à geocodificação de formandos

As instruções lançadas pelo botão 2 permitem ao utilizador executar o *plugin mmqgis* que materializa a geocodificação. Desta operação resulta uma *shapefile* de pontos com os atributos definidos pela macro “geocoding”. Pode-se ver na Figura 41 os atributos desta *shapefile*: Morada (rua e número polícia do formando); Cidade (localidade); Distrito; País; Formando (nome do formando) e Telefone (contato existente, telemóvel ou telefone).

	Morada	Cidade	Distrito	Pais	Formando	Telefone
0		TORRES NOVAS	Santarem	Portugal		
1		RIO DE MOINHOS	Santarem	Portugal		
2		VILA NOVA DA B...	Santarem	Portugal		
3		TORRES NOVAS	Santarem	Portugal		
4		ASSENTIZ	Santarem	Portugal		
5		TORRES NOVAS	Santarem	Portugal		

Figura 41 – Tabela de atributos da *shapefile* resultante da geocodificação

Juntamente com a *shapefile* referida é criado um ficheiro *CSV* contendo os registos não encontrados pelo *plugin mmqgis*. Desta forma o utilizador tem possibilidade de alterar o ficheiro *XLSX* no sentido de corrigir erros ou falta de informação e realizar nova geocodificação.

Os ficheiros criados pela macro “geocoding” e pelo *plugin mmqgis* são guardados automaticamente na pasta “Dados” (Figura 35). Assim, o utilizador tem a informação necessária ao SIG numa só localização, sem ter de escolher ou criar pastas.

Para que o resultado da geocodificação fosse guardado na pasta “Dados” com o sistema de referência adotado no SIG-ATEP, o código do *plugin mmqgis* teve de ser alterado. Esta modificação é vantajosa para o utilizador não-especialista em SIG, pois automatiza algumas tarefas, uma vez que não é necessário alterar o sistema de referência da *shapefile*, escolher pasta de destino, dar nome ao ficheiro ou carregá-lo para o gestor de camadas do QGIS.

A alteração deste *plugin mmqgis* também apresenta vantagens para a análise da rede. Dado que alguns procedimentos da análise serão automatizados através de botões da barra de ferramentas ATEP, exige-se que os ficheiros de entrada estejam numa localização exata e com determinado nome. Por exemplo, quando se executa a ferramenta *GRASS* para importação da *shapefile* (*v.in.ogr*) com o resultado da geocodificação, este vai sempre buscar o ficheiro correto na pasta “Dados”.

3.7. Identificação dos locais para realizar formação

Face às limitações para conseguir implementar a análise localização-alocação, tal como foi definida no ponto 2.5.1, optou-se por uma solução que utiliza a ferramenta *GRASS* para definição de áreas de serviço (*v.net.alloc*) e construiu-se a macro *Excel* “alocacao_formandos” (Anexo 2c). Esta solução baseia-se num processo iterativo que procura minimizar o número de locais de formação a abrir e garantir que todos os formandos estão servidos pelo local de formação mais próximo.

A ferramenta *GRASS* referida é aplicada sobre a rede de estradas definindo as áreas de serviço para todos os locais de formação, isto é, com base na posição dos locais de formação identificam-se as zonas da rede mais próximas de cada um desses pontos.

Após a análise referida no parágrafo anterior faz-se a interseção dos formandos com as áreas de serviço e fica-se a saber o número de formandos alocados a cada local de formação. Com estes dados surge um problema de cobertura a partir do qual se pode minimizar o número de locais de formação a considerar, ou seja, define-se um número mínimo de formandos (ex.: 10) por local de formação e retiram-se da análise os locais de formação que não possuam esse número mínimo de formandos na sua área de serviço.

A gestão do número de locais de formação é efetuada através da macro “alocacao_formandos” que procede à edição da tabela *DBF* contendo os locais de formação. Para retirar um local de formação da análise a macro atribui o valor “-1” ao atributo ESTADO (ver Figura 34).

Depois de redefinir o número de locais de formação realiza-se uma primeira iteração usando novamente a ferramenta *GRASS v.net.alloc*. Nesta segunda análise obtêm-se áreas de serviço em menor número, mas com dimensões maiores. Este procedimento assegura sempre a cobertura dos formandos pelos locais de formação, sendo que o número de formandos por área de serviço aumenta.

Caso o utilizador pretenda fazer nova iteração pode usar novamente a macro “alocacao_formandos” para redefinir o número de locais de formação e correr a ferramenta *v.net.alloc*. Tal como na primeira iteração o número de locais de formação irá diminuir e a dimensão das áreas de serviço irá aumentar. No limite, as iterações levaram a um resultado composto por um local de formação e uma área de serviço.

Pelos resultados obtidos pode-se dizer que a solução adotada representa um problema de máxima cobertura, onde não se define a cobertura máxima admitida (ver ponto 2.5.1).

A operacionalização da análise descrita nos parágrafos anteriores é feita a partir dos botões 3 a 6 da barra de ferramentas ATEP. A Figura 42 apresenta de modo esquemático o que cada botão faz:

- **Botão 3**, prepara os dados para a análise: rede de estradas e locais de formação. Importa a *shapefile* com os locais de formação para a pasta “ATEP” (Figura 35) e reconstrói a topologia da rede. Este procedimento é necessário sempre que se inicia uma análise nova, pois a macro “alocacao_formandos” altera a tabela da *shapefile* com os locais de formação.
- **Botão 4**, executa a ferramenta *GRASS v.net.alloc* para definição de áreas de serviço e prepara os dados a editar pela macro “alocacao_formandos”. Começa por definir as áreas de serviço para os locais de formação, importa a *shapefile* com os formandos geocodificados, conecta-os à rede de estradas e reconstrói a topologia da rede. Depois, intersecta os formandos geocodificados com as áreas de serviço e exporta o resultado através de uma *shapefile* para a pasta “Dados”.
- **Botão 5**, abre pasta com ficheiro *Excel* com a macro “alocacao_formandos”.
- **Botão 6**, exporta os resultados da análise, em formato *shapefile*, para a pasta “Dados”: rede de estradas dividida por áreas de serviço e locais de formação considerados.

Como se pode verificar na Figura 42 a utilização dos botões não é sequencial. A análise deve ser iniciada com o botão 3, seguindo do botão 4 e 5. Após seleccionar o número de locais de formação na macro “alocacao_formandos”, deve-se voltar ao botão 4 para definir novas áreas de serviço, isto é, primeira iteração. Para visualizar os resultados da iteração o utilizador deve carregar no botão 6 de forma a exportar os resultados.

Se o utilizador não estiver satisfeito com os resultados da primeira iteração pode voltar ao botão 5 e proceder à iteração seguinte (botão 4).

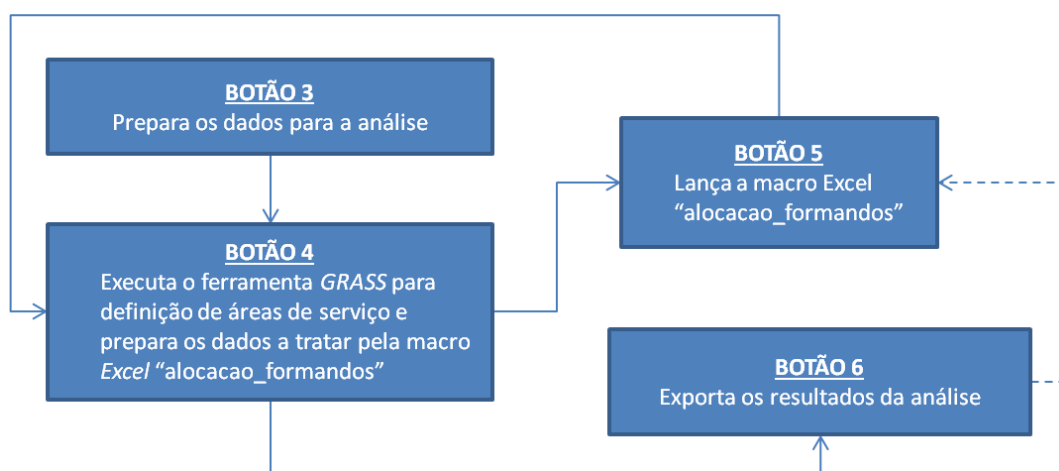


Figura 42 – Estrutura de processos para otimizar a localização dos locais de formação

A visualização dos resultados é efetuada usando três *shapefile* diferentes: uma com os formandos geocodificados, outra com os locais de formação considerados e outra com as áreas de serviço. Para que o utilizador possa visualizar a informação usando apenas uma *shapefile*, o **botão 7** apresenta instruções para unir *shapefile* por localização e criar uma nova *shapefile*. É feita a união da *shapefile* contendo as áreas de serviço com a *shapefile* dos formandos, usando a intersecção como condição para a união. Esta operação permite ao utilizador realizar consultas sobre a tabela de atributos da nova *shapefile*, como por exemplo, “Que formandos foram alocados ao local de formação x?”.

Para que o utilizador possa ter uma antevisão do número de locais de formação a considerar na análise, a macro “alocacao_formandos” possui uma interface *Windows Forms* (ver Figura 43) que informa do número de locais a ser considerado pela ferramenta *GRASS v.net.alloc*. O utilizador seleciona o número mínimo de formandos por local de formação, através da lista pendente, e carrega em “Pré-visualizar escolha”. Logo de seguida aparece uma caixa de diálogo com o número de locais de formação a considerar pela análise da ferramenta *GRASS* referida. De maneira a preparar os dados para a análise basta carregar em “Finalizar”.

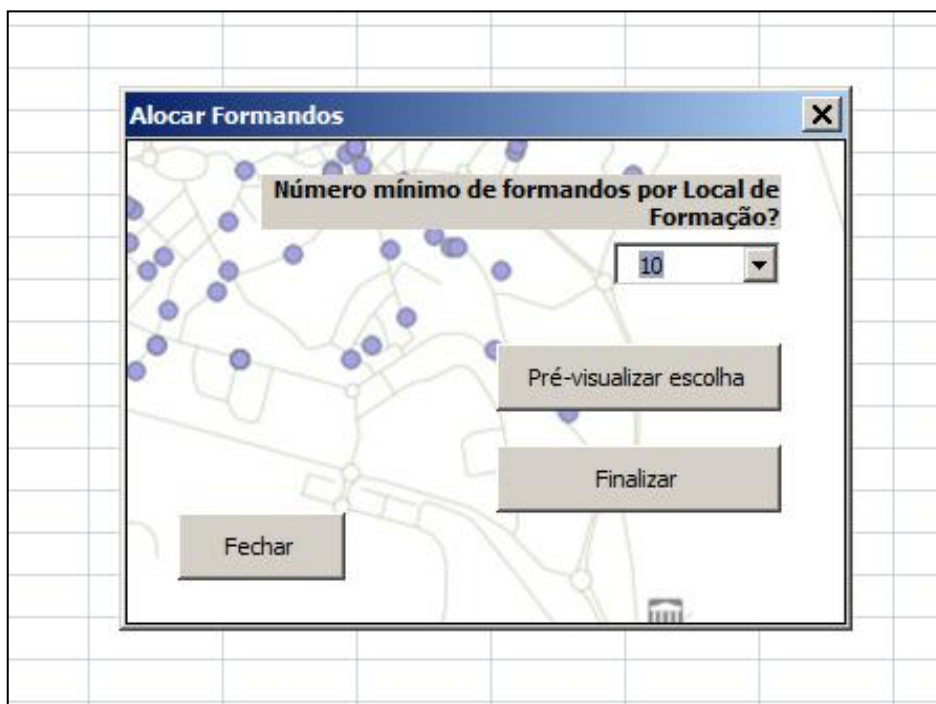


Figura 43 - Interface gráfica da macro *Excel* “alocacao_formandos” para apoio à otimização da localização dos locais de formação

3.8. Definição de rede iso com custo em tempo de viagem

A rede iso fornece ao utilizador informação sobre o custo de deslocação na rede de estradas, de acordo com vários níveis de custo. É constituída por várias isolinhas de custo em tempo de viagem, tendo-se definido as classes de valores: 15, 30, 45 e >45 minutos. Esta disposição foi definida em conjunto com os responsáveis pelo projeto educativo da empresa.

A construção da rede iso é efetuada através da ferramenta *GRASS v.net.iso* que corta a rede de estradas através de isolinhas de custo. A sua operacionalização é efetuada pelo **botão 8** que realiza a análise e exporta os resultados para uma *shapefile*, localizada na pasta “Dados”.

A informação contida nesta *shapefile* é mais útil quando cruzada com a informação existente na *shapefile* que contém os formandos geocodificados. Tal como no ponto anterior o utilizador pode usar o **botão 9** para consultar instruções de forma a unir as duas *shapefile*. Após a união o utilizador pode realizar consultas sobre a tabela de atributos da nova *shapefile*, como por exemplo, “Quais os formandos que se encontram a 30 minutos da ATEP?”.

3.9. Cálculo de indicador estatístico de necessidade de formação (INF)

No sentido de fornecer ao utilizador uma ferramenta que permitisse identificar áreas de acordo com a necessidade de formação, foi construído um indicador estatístico de necessidade de formação (INF) baseado em variáveis estatísticas dos Censos 2011. Trata-se de um indicador resumo que irá permitir identificar zonas com maior necessidade de formação, ao nível de concelho, freguesia e subsecção estatística. A construção deste indicador baseia-se no método adotado por Mcloone (2004), explicitado no ponto 2.3.1.

Procura-se construir um indicador que revele privação em educação e empregabilidade, daí a designação: indicador de necessidade de formação. As variáveis que compõem o indicador foram selecionadas em conjunto com os responsáveis pelo projeto educativo da empresa. O método de escolha foi empírico, sendo a tomada de decisão baseada em anos de experiência acumulada. Assim, foram escolhidas as seguintes variáveis dos Censos 2011:

- 1) *Número de indivíduos residentes com idades 25 a 64 anos;*
- 2) *Número de indivíduos residentes com 3º ciclo do ensino básico;*
- 3) *Número de famílias clássicas com 1 desempregado.*

A variável 1 foi escolhida por representar o intervalo de idades elegíveis para formação financiada, no âmbito de projetos educativos desenvolvidos pela empresa, quanto mais população nesta faixa etária, melhor para ATEP. A variável 2 foi identificada como sendo reveladora de populações com lacunas em formação escolar, quanto maior o número de pessoas com este nível de ensino, maior a necessidade de escolarização. A variável 3 apresenta indício de agregados familiares com problemas de empregabilidade, é preferível à taxa de desemprego porque as famílias com desempregados apresentam mais determinação para ingressar em formação, seja escolar ou profissional. Como a taxa de desemprego é um indicador genérico não seria possível identificar essa especificidade.

De acordo com a metodologia adotada as variáveis dos Censos 2011 devem ser transformadas para valores relativos, deste modo, procedeu-se ao cálculo das variáveis: R25_64 (percentagem de indivíduos residentes com idades 25 a 64 anos), calculada através da fórmula (3.3); R3 (percentagem de indivíduos residentes com 3º ciclo do ensino básico), fórmula (3.4); F1D (percentagem de famílias clássicas com um desempregado), fórmula (3.5).

Os denominadores das fórmulas são variáveis dos Censos 2011, que representam o total de indivíduos residentes e de famílias clássicas, de acordo com o tema das variáveis escolhidas (numeradores).

$$R25_64 (\%) = \frac{\text{Número de indivíduos residentes com idades 25 a 64 anos}}{\text{Número de indivíduos residentes}} \times 100 \quad (3.3)$$

$$R3 (\%) = \frac{\text{Número de indivíduos residentes com 3º ciclo do ensino básico}}{\text{Número de indivíduos residentes}} \times 100 \quad (3.4)$$

$$F1D (\%) = \frac{\text{Número de famílias clássicas com 1 desempregado}}{\text{Número de Famílias clássicas}} \times 100 \quad (3.5)$$

O grau de associação entre as variáveis foi testado através do coeficiente de correlação de *Pearson*, determinado para todos os pares de variáveis. Para o efeito foi usado o *Excel* que possui uma função para calcular este coeficiente.

Nas Tabelas 7 e 8 apresentam-se os resultados obtidos para os concelhos e freguesias, respetivamente. Para os concelhos os resultados variam entre correlação moderada (0,47) a forte (0,62 e 0,71), nas freguesias as correlações são moderadas (0,40 a 0,48). Como seria de esperar o sentido das correlações é positivo, para áreas com maior número de indivíduos entre 25 e 64 é

existe maior número de indivíduos com 3º ciclo do ensino básico e maior número de famílias com um membro desempregado.

Tabela 7 – Correlações entre variáveis para estatísticas dos concelhos

		R25_64	R3	F1D
R25_64	<i>Coeficiente de Pearson</i>	-	0,47	0,71
R3	<i>Coeficiente de Pearson</i>	0,47	-	0,62
F1D	<i>Coeficiente de Pearson</i>	0,71	0,62	-
	<i>Nº Concelhos</i>	11	11	11

Tabela 8 – Correlações entre variáveis para estatísticas das freguesias

		R25_64	R3	F1D
R25_64	<i>Coeficiente de Pearson</i>	-	0,48	0,45
R3	<i>Coeficiente de Pearson</i>	0,48	-	0,40
F1D	<i>Coeficiente de Pearson</i>	0,45	0,40	-
	<i>Nº de Freguesias</i>	112	112	112

Para as subsecções estatísticas não foram calculados os coeficientes de correlação. A presença de elevado número de subsecções com valor zero inviabilizou a obtenção de resultados aceitáveis, pois nestas condições correlação é insignificante ou fraca. Se as variáveis registassem para todos os casos valores zero a correlação seria nula.

Mesmo com esta situação pode-se proceder ao cálculo do INF porque os valores apresentados para concelhos e freguesias são suficientes para explicar a existência de associação entre as variáveis.

O cálculo do INF partiu da aplicação da fórmula (2.1), a qual permitiu obter o valor de z para todas variáveis e todos os casos. O valor do INF é o valor de *Carstairs Score* definido no ponto 2.3.1, ou seja, a soma dos z para cada caso. A Tabela 9 apresenta os valores das variáveis que entram no cálculo do indicador, seus valores z (Z_{R25_64} , Z_{R3} , Z_{F1D}) e valores INF obtidos para os concelhos da NUTIII Médio Tejo.

Analisando a Tabela 9 verifica-se que existem concelhos com valores de INF favoráveis à empresa, como por exemplo, Entroncamento, o seu valor indica que está posicionado 4,26 desvios-padrão relativamente à média. Para este valor de INF é esperado que os valores das variáveis sejam dos mais elevados da série. Analisando novamente a tabela verifica-se que os valores deste concelho são os mais elevados para as variáveis R25_64 e R3 e acima da média para F1D. Pode-se identificar também concelhos com menos necessidade de formação: Ferreira do Zêzere, Mação e Ourém, assim como, Concelhos próximos da média: Alcanena, Tomar, Vila Nova da Barquinha e Tomar.

Tabela 9 – Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para os concelhos da NUT III Médio Tejo (resultados baseados nos Censos 2011)

Concelho	R25_64 (%)	R3 (%)	F1D (%)	Z _{R25_64}	Z _{R3}	Z _{F1D}	INF
Abrantes	52,45	16,90	11,35	0,32	0,50	1,14	1,96
Alcanena	53,07	15,90	9,28	0,53	-0,50	-0,02	0,01
Constância	54,34	16,69	10,73	0,94	0,29	0,80	2,03
Entroncamento	55,57	18,64	10,51	1,35	2,24	0,67	4,26
Ferreira do Zêzere	48,40	14,25	6,31	-1,00	-2,15	-1,69	-4,85
Mação	43,65	16,38	6,13	-2,56	-0,02	-1,80	-4,38
Sardoal	50,62	16,91	11,16	-0,28	0,51	1,04	1,27
Tomar	51,26	16,32	11,00	-0,07	-0,08	0,95	0,80
Torres Novas	52,75	16,23	9,16	0,42	-0,17	-0,09	0,17
Vila Nova da Barquinha	52,21	16,46	8,42	0,25	0,06	-0,51	-0,20
Ourém	51,77	15,73	8,43	0,10	-0,68	-0,50	-1,08
<i>Média</i>	<i>51,46</i>	<i>16,40</i>	<i>9,32</i>				

Na Tabela 10 apresentam-se resultados para algumas freguesias da NUT III, o elevado número de freguesias impede que sejam apresentados valores para todas as freguesias⁷. Analisando os valores de INF verifica-se que existe um padrão de distribuição idêntico aos valores obtidos para os concelhos: os valores mais elevados para as variáveis correspondem a valores INF positivos e maiores (ex.: Concovada); os valores mais baixos para as variáveis correspondem a valores INF negativos e menores (ex.: Aldeia do Mato). Neste âmbito a diferença entre concelhos e freguesias está na amplitude dos valores de INF, sendo superior para as freguesias.

Tabela 10 - Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para algumas freguesias da NUT III Médio Tejo (resultados baseados nos Censos 2011)

Freguesia	R25_64 (%)	R3 (%)	F1D (%)	Z _{R25_64}	Z _{R3}	Z _{F1D}	INF
Aldeia do Mato	39,00	7,48	8,21	-2,73	-3,65	-0,23	-6,61
Alferrarede	53,32	17,87	11,25	0,79	1,21	0,86	2,86
Alvega	48,83	13,74	11,53	-0,32	-0,72	0,96	-0,07
Bemposta	52,03	13,59	9,99	0,47	-0,79	0,41	0,09
Martinchel	51,49	15,89	10,04	0,34	0,28	0,43	1,05
Mouriscas	41,38	16,92	7,54	-2,15	0,76	-0,47	-1,86
Pego	53,06	17,28	12,80	0,72	0,93	1,42	3,08
Rio de Moinhos	50,83	17,80	9,86	0,18	1,17	0,36	1,71
Rossio ao Sul do Tejo	50,94	17,15	13,19	0,20	0,87	1,57	2,64
São Facundo	46,39	13,05	11,53	-0,92	-1,04	0,97	-1,00
Concovada	52,68	17,30	19,05	0,63	0,94	3,68	5,25

⁷ Todos os resultados do INF encontram-se nos ficheiros XLSX: “Z_scores_concelho”; “Z_scores_freguesia” e “Z_scores_subsecao”, localizados na subpasta “dados_estatisticos”, da pasta “Ficheiros_SIG_ATEP” no CD-ROM fornecido.

A Tabela 11 mostra resultados para algumas subsecções estatísticas (por razões análogas aos resultados por freguesia). O padrão de distribuição de valores é idêntico ao dos concelhos e freguesias, sendo que para as subsecções os valores do INF apresentam a maior amplitude.

Tabela 11 - Valores das variáveis que compõem o INF, valores Z calculados para cada variável e valores obtidos para INF. Valores para algumas subsecções estatísticas da NUT III Médio Tejo (resultados baseados nos Censos 2011)

Subsecção estatística (GEO_COD)	R25_64 (%)	R3 (%)	F1D (%)	Z _{R25_64}	Z _{R3}	Z _{F1D}	INF
14010100102	57,14	0,00	33,33	0,54	-1,01	1,83	1,36
14010100103	41,67	0,00	0,00	-0,09	-1,01	-0,55	-1,66
14010100104	44,44	0,00	0,00	0,02	-1,01	-0,55	-1,54
14010600120	100,00	75,00	100,00	2,31	4,75	6,60	13,65
14010100106	31,25	6,25	16,67	-0,52	-0,53	0,64	-0,41
14010100107	18,18	0,00	16,67	-1,06	-1,01	0,64	-1,43
14010100108	23,08	7,69	0,00	-0,86	-0,42	-0,55	-1,83
14010100109	26,67	0,00	0,00	-0,71	-1,01	-0,55	-2,27
14010100110	23,08	7,69	0,00	-0,86	-0,42	-0,55	-1,83
14200200311	3,57	0,00	0,00	-1,66	-1,01	-0,55	-3,23
14010100212	50,00	25,00	0,00	0,25	0,91	-0,55	0,61

A análise dos valores de INF reveste-se de alguma complexidade porque a sua amplitude varia em função da desagregação territorial. Para os concelhos os valores das variáveis são menos heterogéneos que nas freguesias e subsecções estatísticas, logo a amplitude do INF é menor. No sentido de facilitar a análise deste indicador foi construída uma escala qualitativa (Tabela 12), onde os valores de INF são agrupados em classes de igual amplitude.

Tabela 12 – Tabela com classes de valores INF e escalas qualitativas para o INF, para: concelhos, freguesias e subsecção estatística

	Classes de valores INF	Escalas qualitativas INF
<i>Concelhos</i>	$\geq -4,85$ e $< -2,57$	Fraca
	$\geq -2,57$ e $< -0,30$	Potencial
	$\geq -0,30$ e $< 1,98$	Emergente
	$\geq 1,98$ e $< 4,26$	Forte
<i>Freguesias</i>	$\geq -6,61$ e $< -3,45$	Fraca
	$\geq -3,45$ e $< -0,68$	Potencial
	$\geq -0,68$ e $< 2,29$	Emergente
	$\geq 2,29$ e $< 5,25$	Forte
<i>Subsecções estatísticas</i>	$\geq -3,37$ e $< 1,36$	Fraca
	$\geq 1,36$ e $< 6,10$	Potencial
	$\geq 6,10$ e $< 10,84$	Emergente
	$\geq 10,84$ e $< 15,57$	Forte

Os valores INF foram classificados qualitativamente através da macro *Excel* “categorias_INF” (Anexo 2d). Esta tarefa não foi feita manualmente porque seria muito exaustiva, principalmente para os dados das subsecções estatísticas que representam cerca de 10 000 entidades geográficas. A macro foi adaptada para cada nível (concelho, freguesia e subsecção estatística) de modo a substituir os valores INF pelos valores da escala qualitativa apresentada na Tabela 12. Os valores obtidos foram carregados nas *shapefile*, Censos 2011 e CAOP 2012, de forma a permitir a produção de mapas temáticos de necessidade de formação.

3.10. Construção de mapas temáticos

Com os dados resultantes das análises o utilizador pode construir diferentes mapas temáticos, compará-los a diferentes níveis de desagregação territorial e partilhá-los por toda a organização fazendo uso do QGIS. No âmbito deste projeto é útil elaborar mapas temáticos com o INF ou outros dados estatísticos dos Censos 2011, de modo a identificar áreas prioritárias para promover atividades de formação. Também pode ser pertinente juntar a estes mapas a informação contida nas *shapefiles*: rede iso, áreas de serviço, locais de formação ou formandos geocodificados.

A Figura 44 apresenta um mapa construído através do compositor de impressão do QGIS, onde se sobrepõem a rede iso com os valores de INF por Freguesia. Este mapa permite responder questões do tipo: “Quais as áreas com maior necessidade de formação, localizadas a menos de 30 minutos da ATEP?”.

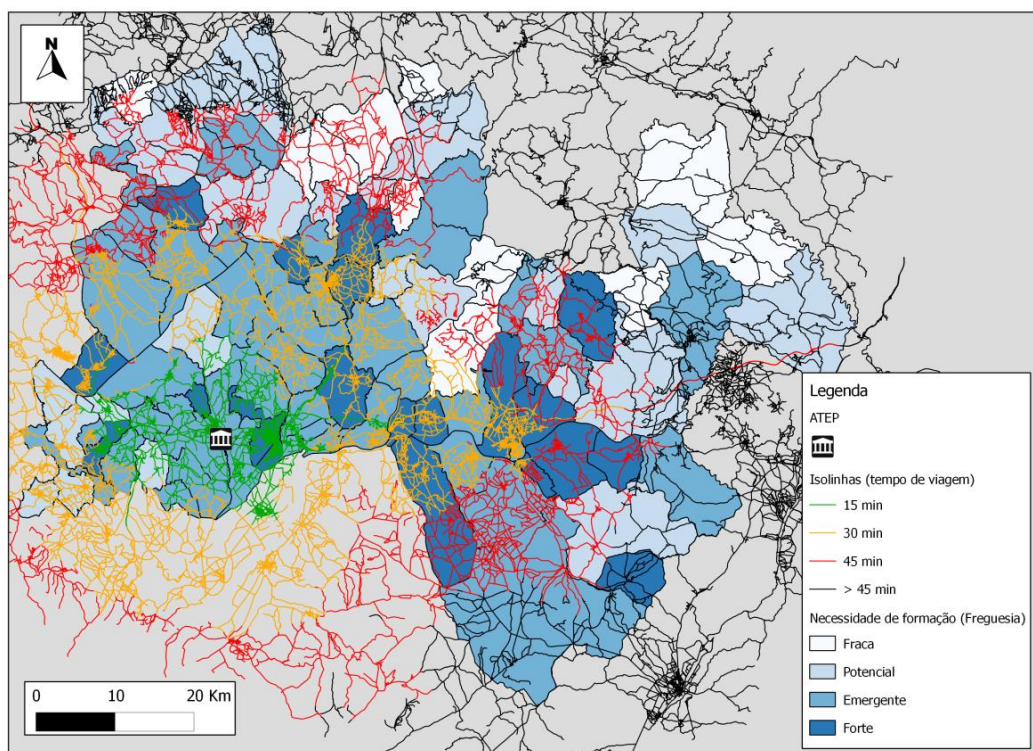


Figura 44 – Mapa temático com necessidade de formação e rede iso. Dados: CAOP 2012 com freguesias da NUT III Médio Tejo e rede de estradas *OpenStreetMap*

O *plugin HTML Image Map Plugin* permite a exportação rápida do mapa para formato *HTML*, dispensando o *QGIS* para visualizar o mapa e atributos das entidades geográficas, basta ter um *browser* de Internet (ex.: *Mozilla FireFox*). Apresenta-se um exemplo na Figura 45 onde é possível visualizar o *Google Streets*, duas áreas de serviço, formandos geocodificados e os locais de formação. Para visualizar o nome dos formandos basta clicar sobre os pontos.

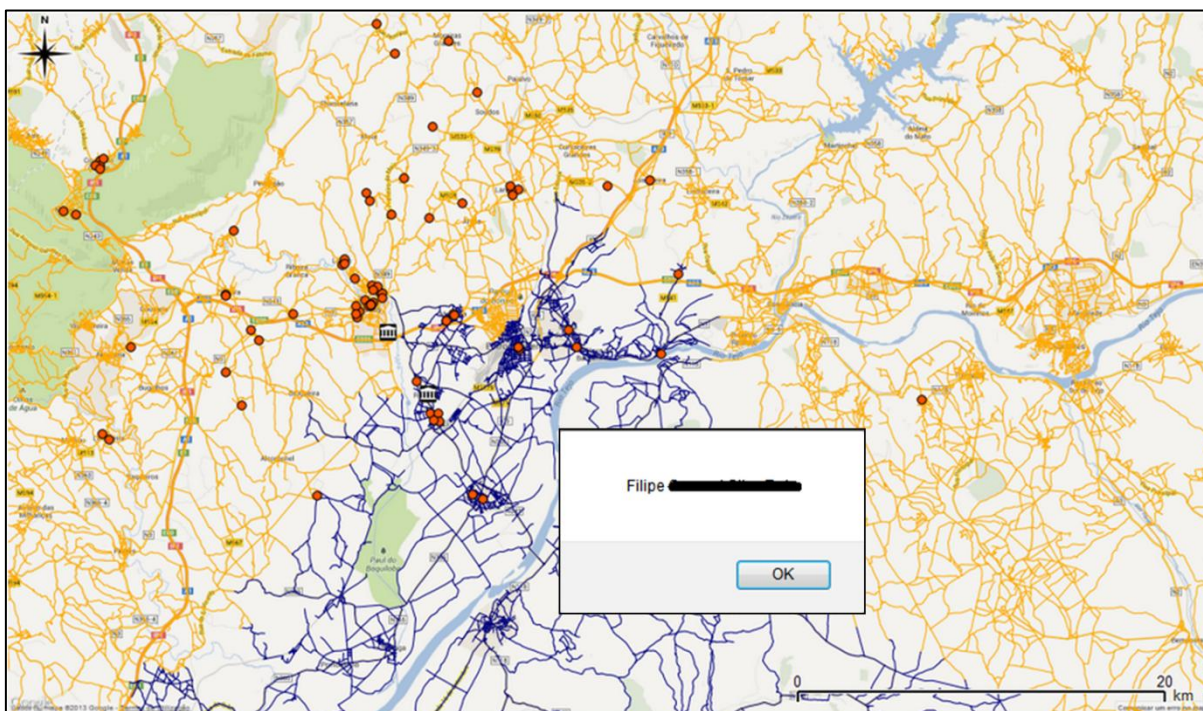


Figura 45 - Resultado do *HTML Image Map Plugin* com visualização de um mapa e atributo da camada formandos (pontos a vermelho). A rede de estradas encontra-se dividida em duas áreas de serviço: linhas a azul e a amarelo. Dados: *OpenStreetMap*, *Google Maps*

Para conseguir uma leitura completa do mapa é preferível utilizar o compositor de impressão e exportar o mapa como imagem ou *PDF*, em vez do *HTML Image Map Plugin*. Em mapas temáticos deve-se apresentar legenda com a descrição de todos os elementos presentes.

4. DISCUSSÃO DOS RESULTADOS OBTIDOS

Após implementação do projeto SIG foi feita uma avaliação dos resultados obtidos e de que forma estes podem ajudar a empresa. Foram realizados testes com grande número de formandos de maneira a explorar todas as potencialidades deste tipo de análise. Com os resultados obtidos pretende-se avaliar os benefícios para a empresa no que respeita ao planeamento/gestão de locais de formação e tomada de opções estratégicas.

4.1. Avaliação dos benefícios no planeamento da atividade empresarial

Numa primeira abordagem deve-se destacar a possibilidade dos responsáveis pelo projeto educativo poderem visualizar espacialmente o seu território formativo: distribuição dos locais de formação; formandos; estatísticas sobre educação; etc. Esta visão pode mudar a forma como as opções de planeamento são tomadas, por exemplo: ao analisar a cartografia com distribuição de formandos rapidamente se pode identificar zonas com maior densidade e, a partir daí, realizar ações para contato de formandos no sentido de abrir novos locais de formação.

Em termos operacionais a análise para otimização da localização de locais de formação mostra ser uma boa opção, principalmente, quando não se percebe bem quais os locais que devem ser abertos em detrimento de outros. O resultado da análise fornece elementos necessários para que os decisores possam fazer um planeamento correto e rápido da abertura de novos grupos, para o demonstrar, foi simulada uma situação realista com os dados provenientes da plataforma SIGO.

Na seleção dos dados optou-se pelos formandos em situação de diagnóstico e acolhimento, isto é, aguardam integração em turma para iniciar processo de formação. A aplicação deste filtro no *Excel* resultou em 134 formandos para geocodificação.

Após geocodificação definiram-se de áreas de serviço para todos os locais de formação, com o número mínimo de formandos por local de formação definido em 10. O resultado desta análise devolveu 3 locais elegíveis para a nova definição de áreas de serviço. Na figura 46 pode-se ver o resultado desta análise, estão definidas 3 áreas de serviço, formandos geocodificados e locais de formação considerados na solução.

É possível visualizar a influência das vias rápidas na dimensão e distribuição geográfica de determinadas áreas de serviço, como as áreas associadas à ATEP. Pode-se verificar também, que a existência de uma via rápida, na parte este da referida área de serviço, contribuiu para que esta tivesse uma maior abrangência geográfica. Se a rede de estradas não possuísse vias rápidas a área de serviço, junta de freguesia Riachos, seria muito maior.

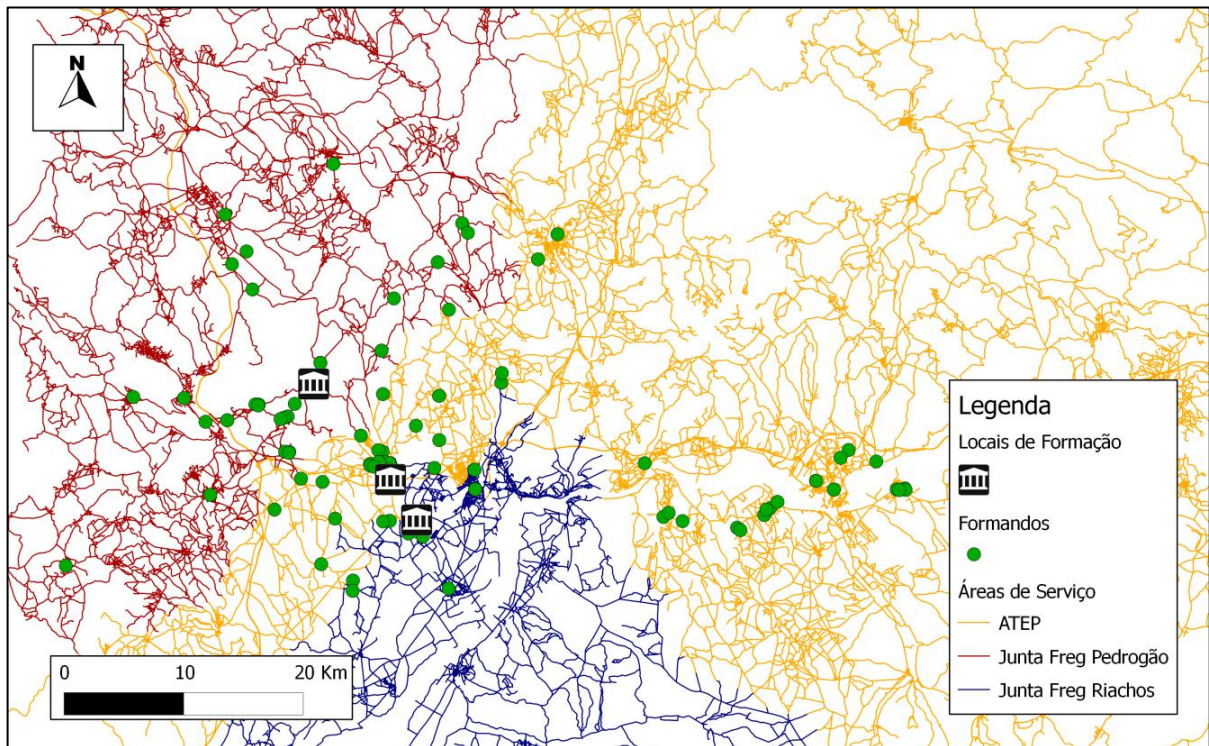


Figura 46 – Análise de redes para localização ótima de locais de formação. Dados: *OpenStreetMap*

Após a visualização geográfica o utilizador pode realizar consultas à tabela de atributos que resulta da união das tabelas: formandos e áreas de serviço. A informação contida na nova *shapefile* é útil, por exemplo, quando se pretende contactar os potenciais formandos associados a um dos locais de formação presentes na solução. Esta operação ajuda a agilizar a constituição de grupos de formação. Antes a seleção era realizada em tabelas *Excel* através das moradas, agora existe a possibilidade de fazer uma seleção mais rigorosa baseada no tempo de viagem que o formando supostamente demora a chegar ao local onde é ministrada a formação.

A construção de redes iso também tem grande importância na atividade da empresa, principalmente no planeamento de ações de formação de curta duração (formações modulares). Estas são quase todas dinamizadas na empresa e são bastante frequentadas por antigos formandos.

Para este objetivo foram selecionados formandos certificados com 3º ciclo do ensino básico em situação de desemprego (total de 114), de modo a planear a abertura de grupos para formação

modular de Inglês. O resultado da análise pode ser visualizado no mapa da Figura 47. Verifica-se uma grande concentração de formandos junto da ATEP, localizando-se a maior parte a menos de 15 minutos.

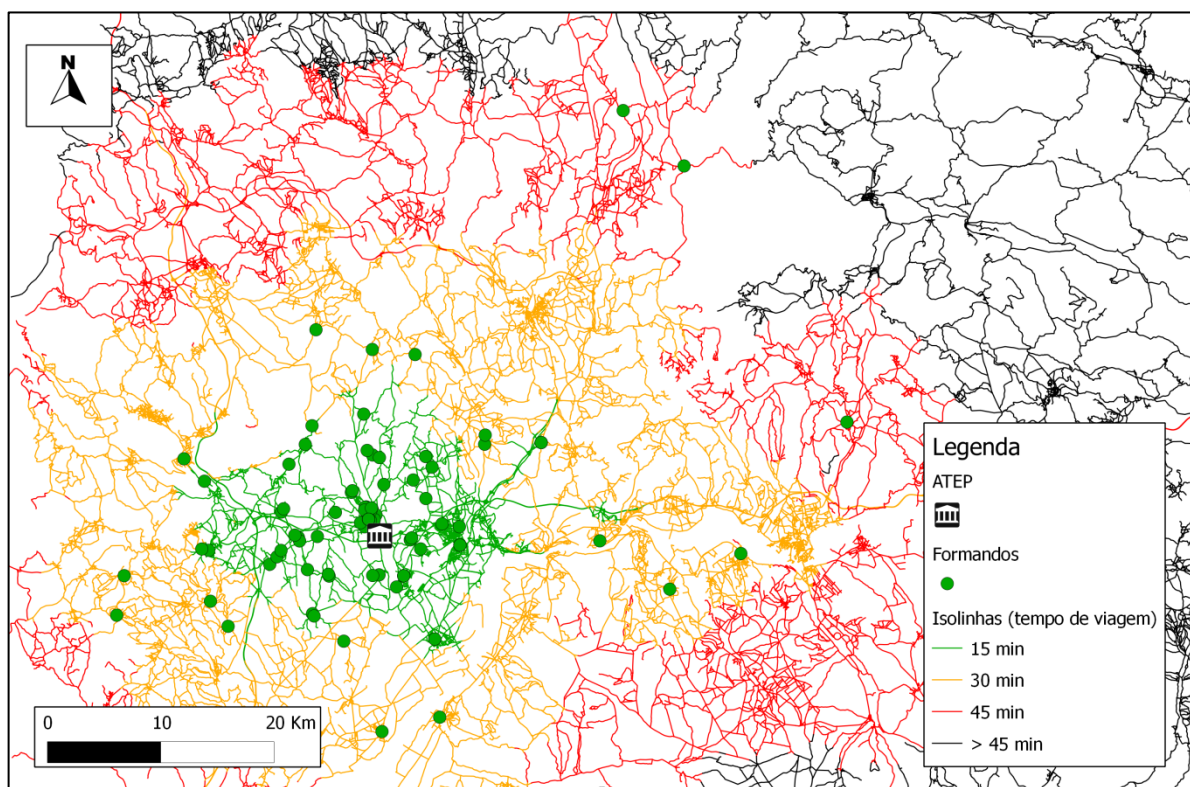


Figura 47 - Análise de redes para definição de rede iso com sobreposição de formandos. Dados: *OpenStreetMap*

Para planejar a formação poder-se-ia seleccionar os formandos a menos de 15 minutos de viagem da empresa e informá-los da oferta formativa. Procedeu-se à união das *shapefile* formandos e rede iso, seguida de uma consulta à tabela de atributos da nova *shapefile*, usando o argumento 15 para o atributo “TEMPO_VIAGEM”. O resultado devolveu uma lista com 80 formandos.

Este resultado aumenta a possibilidade de os formandos responderem positivamente ao convite para uma formação, pois o tempo de viagem até à empresa é bastante favorável. Sem o SIG-ATEP o processo de seleção dos formandos a contactar pode ser demorado e desprovido de rigor, pois o decisor utiliza a tabela *Excel* para seleccionar os formandos de acordo com a morada. No entanto, tem de se ter em conta que existem formandos dispostos a percorrer grandes distâncias para beneficiarem destas formações.

4.2. Avaliação dos benefícios na tomada de decisões estratégicas

A escolha de locais para realizar formação não se deve basear somente na distribuição geográfica dos formandos já registados no SIGO. É necessário que se procure mais mercado pela identificação de potenciais formandos. Esta tarefa pode ser concretizada usando indicadores estatísticos que permitam identificar áreas com maior necessidade de formação. Com a definição destas áreas os decisores podem definir a estratégia que querem para a empresa, relativamente à escolha de locais para divulgar os serviços de formação, seleção de locais para realizar as formações de acordo com o tipo de formação (ex.: áreas com maior número de desempregados devem ser orientadas para formações de curta duração), escolha de entidades para estabelecer parcerias; etc.

Como os dados estão disponíveis para vários níveis (concelho, freguesia e subsecção estatística) o utilizador pode construir mapas de acordo com o detalhe que pretende. Para uma análise mais geral pode optar pelas estatísticas do concelho ou freguesia. O mapa da Figura 48 permite fazer uma análise do INF por freguesia. Pode-se constatar que as freguesias com maior necessidade de formação (Emergente e Forte) localizam-se na região Sul da NUT III Médio Tejo, ainda que a sua distribuição seja algo heterogénea.

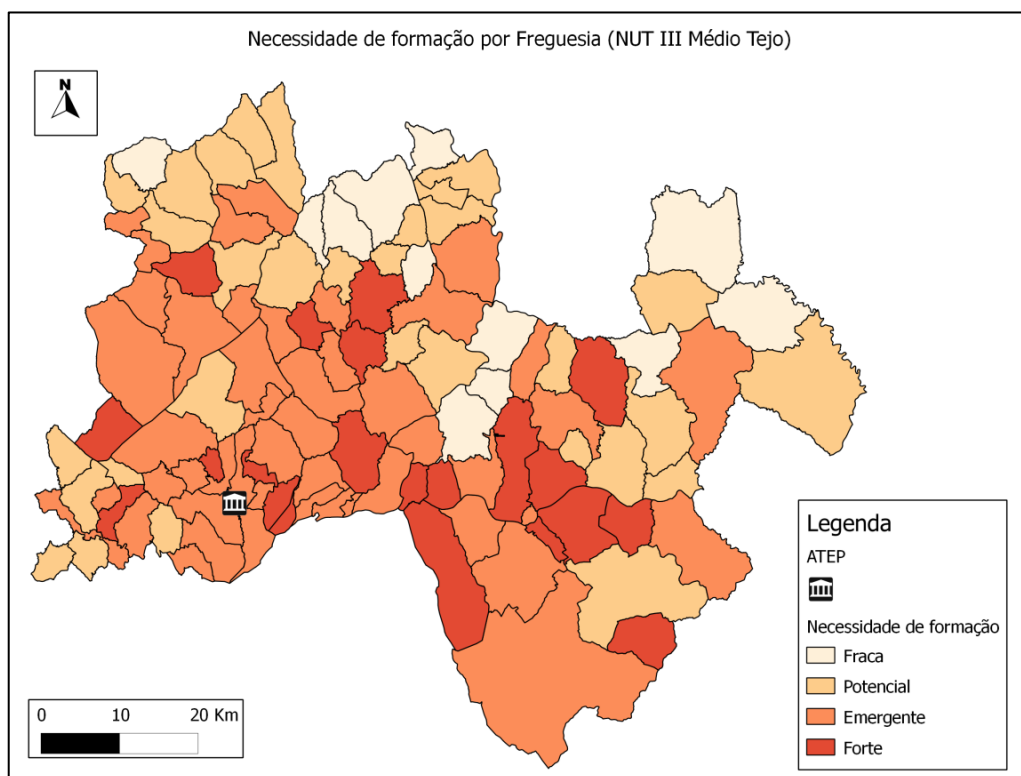


Figura 48 – Indicador estatístico de necessidade de formação (INF) por freguesia. Dados: CAOP 2012, Censos 2011

Para fazer uma análise dentro dos limites de uma freguesia pode-se utilizar os dados por subsecção estatística. Pode-se ainda cruzar esta informação com a rede iso de maneira a escolher áreas de interesse a determinado tempo de viagem da ATEP. O mapa da Figura 49 apresenta o número de desempregados (organizados em classes) por subsecção estatística, rede iso e localização da ATEP. Pela sua análise, podem-se identificar subsecções estatísticas preferenciais ao desenvolvimento de formação direcionada a desempregados. Basta identificar subsecções estatísticas com maior número de desempregados (7 - 30) a menos de 15 minutos da ATEP.

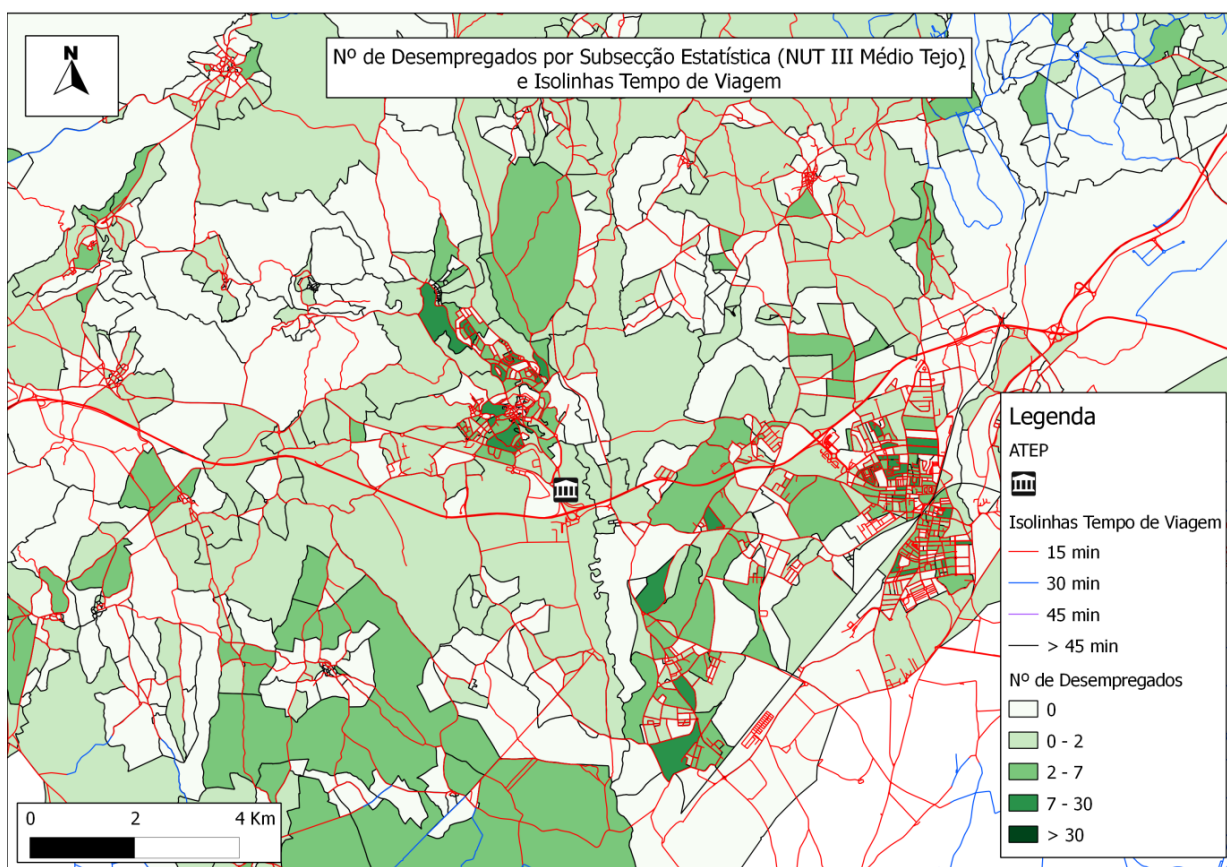


Figura 49 – Número de desempregados por subsecção estatística e rede iso. Dados: CAOP 2012, Censos 2011, *OpenStreetMap*

Para que o utilizador possa identificar os locais no mapa (ex.: Rua, Localidade, etc.) existe a possibilidade de carregar camadas do *Google Maps*, *OpenStreetMap* ou outro. Esta informação pode ser cruzada com os dados estatísticos de forma a agilizar a identificação dos locais para desenvolver formação. O mapa da Figura 50 permite associar as subsecções estatísticas ao nome das ruas, porque possui a rede de estradas do *Google Maps* (*Google Streets*) e o número de desempregados por subsecção estatística. Generalizando um pouco pode-se dizer que o mapa permite identificar as ruas de acordo com o número de desempregados.

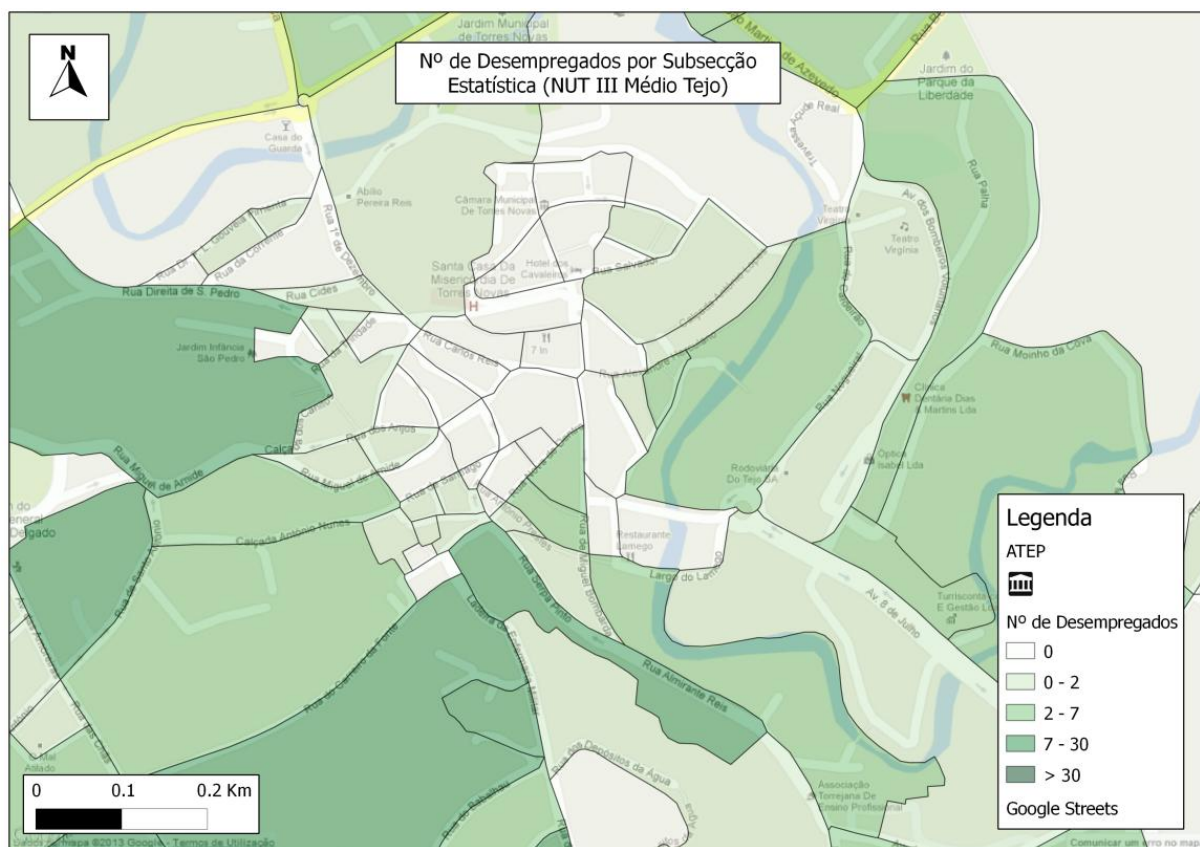


Figura 50 - Número de desempregados por subsecção estatística e *Google Streets*. Dados: CAOP 2012, Censos 2011, *Google Maps*

Este tipo de análises permite à empresa definir estratégias para melhorar a sua presença junto da população, por via de ações de divulgação/sensibilização mais eficazes. Esta ferramenta pode transformar-se numa vantagem competitiva essencial para reforçar a posição da empresa face à concorrência porque, apesar da NUT III Médio Tejo estar sobre a sua responsabilidade, existem outras empresas com a mesma missão.

5. CONCLUSÕES

A aplicação dos SIG em empresas que não trabalham com cartografia não é muito comum, principalmente em empresas de pequena dimensão. A incorporação de informação geográfica nos processos de decisão pode ser uma mais-valia para as empresas, pois parte da informação necessária à sua atividade tem componente geográfica. O presente trabalho mostra que é possível implementar um projeto SIG numa empresa de pequenas dimensões sem custos adicionais em termos de *software* e dados.

A utilização de dados gratuitos revelou-se uma aposta ganha, os problemas encontrados centraram-se apenas na *shapefile* proveniente do *OpenStreetMap*, nomeadamente ao nível da qualidade geométrica e temática. Estes foram ultrapassados com relativa facilidade não colocando em causa a sua integridade e usabilidade. No entanto, para aplicações que incidissem sobre uma região mais vasta (ex.: Portugal) a sua aplicação poderia estar comprometida, face à quantidade de alterações necessárias.

As dificuldades inerentes à utilização de *software* de código aberto levaram a um ajustamento do método usado na análise localização-alocação. O método adotado não automatiza o processo de otimização da localização dos locais de formação, mas permite ao utilizador decidir qual o cenário que mais lhe convém, tendo em conta o número de formandos que pretende alocar a cada local de formação. Para o contexto em que foi aplicado possui vantagens evidentes, o decisor pode controlar o número de locais de formação em função do número mínimo de formandos que pretende alocar. Em outros contextos deverá existir uma reanálise e devido enquadramento.

A definição da rede iso mostrou-se bastante pertinente para o planeamento das atividades de formação na ATEP, pois permite ao decisor selecionar formandos de acordo com o custo de deslocação contabilizado em termos de tempo de viagem. Através desta análise o decisor pode gerir as ações de formação de forma mais eficiente. Pode minimizar o custo de deslocação convocando apenas os formandos localizados a menor tempo de viagem, a partir da ATEP.

Os dados estatísticos usados representam uma mais-valia para o planeamento estratégico da empresa. São disponibilizados por uma fonte fidedigna (INE), apresentam grande nível de detalhe (subsecção estatística) e referem-se a um período recente (Censos 2011). Através do INF

o decisor pode identificar áreas prioritárias para promover e desenvolver formação, dentro do território de atuação da ATEP (NUT III Médio Tejo). Pode também fazer uso das várias variáveis estatísticas disponibilizadas, no sentido de obter uma caracterização sociodemográfica do território.

A maior dificuldade centrou-se na personalização da aplicação SIG de maneira que pudesse ser utilizada por não-especialistas. A inclusão de ferramentas *GRASS* no *QGIS* reveste-se de características muito próprias, dificultando a sua integração com ferramentas existentes no *QGIS* ou seus *plugin*. A construção de uma barra de ferramentas através do *plugin custom launcher* teve como objetivo a integração de várias ferramentas provenientes do *GRASS*, *Excel* e *QGIS*, procurando implementar todos os processos necessários à análise SIG, organizados num esquema sequencial.

A solução apresentada utiliza o *Excel*, um *software* proprietário, devido ao facto deste ser usado na ATEP. No entanto, a solução poderia utilizar *software* de código aberto, como a aplicação *Calc* do *Open Office* que tem uma interface semelhante e também permite a construção de macros.

O sistema pode ser adaptado a outros casos, no âmbito da formação escolar e profissional. Se a abrangência geográfica for diferente, basta alterar os dados referentes à rede de estradas, limites administrativos e variáveis estatísticas. Em outras realidades territoriais é possível que as variáveis que compõem o INF tenham de ser modificadas. Por exemplo, em territórios urbanos dinâmicos a integração de população com 3º ciclo pode não fazer sentido.

Poderia implementar-se a análise localização-alocação em plenitude através da programação de um novo *plugin* para o *QGIS*. Este pode ser implementado em linguagem de programação *python*, usando o *Plugin Builder*. A versão *python* para *Quantum GIS* possui uma livraria que possibilita a construção de grafos.

O custo de deslocação ao longo da rede foi definido em função do tempo de viagem, no entanto pode ser alterado para outro que satisfaça os objetivos do projeto. Pode ser usada a distância ou estimar um valor para o custo de viagem, como por exemplo: custo financeiro, consumo de combustível, etc.

6. BIBLIOGRAFIA

Referências a livros:

- BURROUGH, PETER e MCDONNELL, RACHAEL (2000) – *Principles of Geographical Information Systems*, Oxford
- DENT, B.; TORGUSON, J.; HODLER, T. (2009) – *Cartography, Thematic Map Design*, 6ª Edição, McGrawHill
- GHOSH, AVIJIT e RUSHTON, GERAD (1987) – *Spatial Analysis and Location-Allocation Models*, Van Nostrand Reinhold Company, New York
- GNU PSPP (2012) - *PSPP Users' Guide (Statistical Analysis Software)*, Release 0.7.9-gca93ab, Free Software Foundation, Inc
- GRISMSHAW, DAVID J. (2000) - *Binging Geographical Information Systems into Business*. 2ª Edição, John Wiley & Sons
- HARMON, JOHN e ANDERSON, STEVEN (2003) – *The design and implementation of Geographic Information Systems*, John Wiley & Sons
- MIRCHANDANI, PITU B. e FRANCIS, RICHARD L. (1990) – *Discrete Location Theory*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Advisory Editors: Ronald L. Graham, Jan Karel Lenstra and Robert E. Tarjan
- NAG, PRITHUISH e SENGUPTA, SMITA (2008) – *Introduction to Geographical Information System*, Ashok Kumar Mittal
- LONGLEY, P.; GOODCHILD, M.; MAGUIRE, D.; RHIND, D. (2011) – *Geographic Information Systems & Science*, 3ª Edição, John Wiley & Sons
- PICK, JAMES (2007) – *Geo-Business, GIS in The Digital Organization*, John Wiley & Sons
- QGIS DEVELOPMENT TEAM (2013) - *QGIS User Guide*, Release 1.8.0, QGIS Project

- SOCIAL DISADVANTAGE RESEARCH CENTRE (2003) - *Scottish Index of Multiple Deprivation*, Department of social policy and social work, University of Oxford. Disponível em <http://www.scotland.gov.uk/library5/social/siod-00.asp>
- SMITH, M.; GOODCHILD, M. e LONGLEY, P. (2013) - *Geospatial Analysis, A Comprehensive Guide to Principles, Techniques and Software Tools*, 4ª Edição. Disponível em <http://www.spatialanalysisonline.com/HTML/index.HTML>
- SOMMER, SHELLY e WADE, TASHA (2006) – *A to Z GIS: An Illustrated Dictionary of Geographic Information Systems*, ESRI Press (conteúdo disponível em: <http://support.esri.com/en/knowledgebase/Gisdictionary/browse>)
- SUGUMARAN, RAMANATHAN e DEGROOTE, JOHN (2011) – *Spatial Decision Support Systems, Principles and Practices*, CRC Press
- WORBOYS, M. e DUCKHAM, M. (2004) – *GIS. A computing perspective*, Second Edition, CRC Press

Referências a capítulos em livros:

- BURT, JAMES E.; BARBER, G. e RIGBY, D. (2009) – *Elementary Statistics for Geographers*, Third Edition, The Guilford Press, 156 – 194
- BURSTEIN, FRADA e HOLSAPPLE, CLYDE (2008) – *Hand on Decisions Support Systems*, Volume 2, Springer, 71 - 73
- CHOUDHURG, SUJIT; CHAKRABARTI, D.; CHOUDHURG, SUCHANDRA (2009) – *An introduction to geographic information technology*, I.K. Internacional, 75 - 85
- DENSHAM P. J. (1991) *Spatial decision support systems*, In: D. J. Maguire, M. S. Goodchild and D. W. Rhind (eds) *Geographical information systems: principles and applications*, London: Longman, 403 – 412
- DIAS, ANTÓNIO (2005) – *O problema da p-mediana aplicado ao problema da gestão óptima da diversidade*, Dissertação de Mestrado apresentada à Universidade de Aveiro para obtenção do grau de Mestre em Matemática Aplicada, 5 - 11
- EISELT, H. e SANDBLOM, C. (2004) – *Decision Analysis, Location Models and Scheduling Problems*, Springer, 171 - 189

- GONÇALVES, ALEXANDRE e MATOS, JOÃO (2005) – *Modelação Geográfica de Problemas de Localização de Equipamentos*. Actas da IV Conferência Nacional de Cartografia e Geodesia, LIDEL (Edições Técnicas), 71-80
- LI, Z.; ZHOU, Q.; KAINZ, W. (2004) – *Advances in Spatial Analysis and Decision Making*, Swets & Zeitlinger Publishers, 203 - 208
- LORENA, L.; SENNE, E.; PAIVA, J. e PEREIRA, M. (2001) – *Integração de Modelos de Localização a Sistemas de Informação Geográficas*, Gestão & Produção 8 n.º 2, 180 – 195
- MURTEIRA, B.; RIBEIRO, C; SILVA, J e PIMENTA, C. (2007) – *Introdução à Estatística*, McGraw-Hill, 2ª edição, 1 – 46
- NOGUEIRA, HELENA e REMOALDO, C. (2010) – *Olhares Geográficos sobre a Saúde*, Edições Colibri, 106 - 113
- O’SULLIVAN, DAVID e UNWIN, DAVID (2003) – *Geographic Information Analysis*, John Wiley & Sons, 384 – 411
- PICK, JAMES (2005) – *Geographical Information Systems In Business*, Idea Group Publishing, 1 - 20
- RIGAUX, P; SCHOLL, M.; NOISARD, A. (2002) – *Spatial Databases with application to GIS*, Elsevier Science, 3 – 26
- THOMAS, CRISTOPHER e OSPINA, MILTON (2004) – *Measuring UP, The Business case of GIS*, ESRI, 1 - 15

Referências a artigos em publicações periódicas:

- AMELUXEN, CHRISTOF (2010) – *An Approach to geocoding based on Volunteered Spatial Data*, Geoinformatik 2010. Die Welt im Netz, 2010, 7 - 12
- AZAZ, LOTFY (2011) – *The use of Geographic Information Systems (GIS) in Business*, International Conference on Humanities, Geography and Economics (ICHGE'2011) Pattaya, 299 - 303
- BAKSHI, R.; KNBLOCK, C.; THAKKAR, S. (2004) – *Exploiting online sources to accurately geocode addresses*, GIS 2004, Proceedings of the 12th annual ACM international workshop on Geographic information systems, 194 - 203

- FILHO, DALSON e JÚNIOR, JOSÉ (2009) – *Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r)*, Revista Política Hoje, Vol. 18, nº 1, 115 - 146
- GOLDBERG, D.; WILSON, J. e KNOBLOCK (2007) – *From Text to Geographic Coordinates: The Current State of Geocoding*, URISA Journal, 33 - 46
- KEENAN, PETER (2005) – *Concepts and Theories of GIS in Business*, Idea Group Publishing, 1 - 19
- MCLOONE, PHILIP (2004) - *Carstairs scores for Scottish postcode sectors from the 2001 Census*, MRC Social & Public Health Sciences Unit - University of Glasgow, 1 - 57
- MENNECKE, BRIAN E. (2000) – *Understanding the Role of Geographic Information Technologies in Business: Applications and Research Directions*. Journal of Geographic Information and Decision Analysis, vol. 1, n. ° 1, 44 – 68
- POSTHUMUS, RUDI (2008) – *Gis as a tool in Business Intelligence*. Application technical, Position IT (Maio/Junho, 2008), 71 - 74
- SOMERS, REBECCA (2001) – *Quick Guide to GIS Implementation and Management*, URISA, 1 - 34
- ZANDBERGEN, P. (2007) – *A comparison of address point, parcel and street geocoding techniques*. ELSEVIER – Computers, Environment and Urban Systems, 214 - 232

Referências a páginas na internet:

- ATCHISON, ALASTAIR (2011) – *Página sobre os Sistemas de Referência usados pelos serviços WebMap*, <http://alastaira.wordpress.com/2011/01/23/the-google-maps-bing-maps-spherical-mercator-projection/>, acessada a 02 de Maio de 2013
- BONFIGLIO, ANDREA (2009) – *Página oficial do projeto ExceIToDBF*, <http://ExceItodbf.sourceforge.net/>, acessada a 12 de Maio de 2013
- CATELLS, JORDI (2013) – *Página pessoal de Jordi Castells com tutorial para o uso de ferramentas GRASS*, <http://jcastellssala.wordpress.com/2012/05/07/basic-network-analysis-with-grass/>, acessada a 19 de Agosto de 2013
- COMUNIDAD GVSIG (2013) – *Página oficial do projeto gvSIG*, <http://www.gvsig.org/web/>, acessada a 02 de Maio de 2013

- COMUNIDADE INTERMUNICIPAL DO MÉDIO TEJO (2013) – *Página oficial da Comunidade Intermunicipal do Médio Tejo*, <http://www.mediatejodigital.pt/pt/conteudos/comunidade+intermunicipal+do+medio+tejo/>, acedida a 02 de Maio de 2013
- DENIS (2012) - *Página do autor do plugin Quantum GIS custom launcher*, <http://plugins.qgis.org/plugins/customlauncher/>, acedida a 09 de Maio de 2013
- DPRD (2006) – *Página do Departamento de Planeamento e Desenvolvimento Regional da Universidade de Thessaly (Grécia), responsável pelo projeto S-Distance*, http://www.prd.uth.gr/sites/spatial_analysis/Software/SdOverview_en.HTML#Anchor_new1, acedida a 02 de Maio de 2013
- DUIVENVOORDE, RICHARD (2013) - *Página do autor do plugin Quantum GIS HTML Image Map Plugin*, http://plugins.qgis.org/plugins/imagemap_plugin/, acedida a 02 de Maio de 2013
- ESRI - ArcGIS 10.0 Help (2013) – *Página de ajuda do ArcGIS 10.0*, <http://help.arcgis.com/en/arcgisdesktop/10.0/help/>, acedida a 18.02.2013
- GOOGLE DEVELOPERS (2013) – *Página da Google Geocoding API*, <https://developers.google.com/maps/documentation/geocoding/#ComponentFiltering>, acedida a 30 de Abril de 2013
- GNU (2013) – *Página oficial da licença GPL (General Public License)*, <http://www.gnu.org/licenses/gpl.HTML>, acedida a 15 de Maio de 2013
- GNU PSPP (2012) – *Página oficial do projeto PSPP*, <http://www.gnu.org/Software/pspp/>, acedida a 15 de Maio de 2013
- GRASS GIS (2013) – *Página oficial do projeto GRASS GIS*, <http://grass.osgeo.org/>, acedida a 30 de Abril de 2013
- GRASS MANUAL (2013) – *Página do Manual de referência GRASS*, <http://grass.osgeo.org/grass65/manuals/index.HTML>, acedida a 02 de Maio de 2013
- HOPKINS, W. (2002) – *Página com escala qualitativa usada no coeficiente de correlação de Pearson: A New View of Statistics*, acedida a 12 de Junho de 2013

- IBM (2013) – *Página oficial do Software IBM SPSS*, <http://www-01.ibm.com/Software/analytics/spss/>, acessada a 15 de Maio de 2013
- JSI (2012) – *Página oficial do projeto John Snow*, <http://www.jsi.com>, acessada a 24 de Junho de 2013
- KOSMO (2013) – *Página oficial do projeto Kosmo (Plataforma SIG livre)*, <http://www.opengis.es/index.php>, acessada a 02 de Maio de 2013
- MILLER, BILL (2013) – *Página do autor do Software OpenStat*, <http://www.statprograms4u.com/OpenStatMain.htm>, acessada a 15 de Maio de 2013.
- MINN, MICHAEL (2013) – *Página do autor do plugin Quantum GIS mmqgis*, <http://michaelminn.com/linux/mmqgis/>, acessada a 02 de Maio de 2013
- OLAYA, VICTOR (2013) – *Página oficial do autor da extensão Sextante*, <http://www.sextantegis.com/>, acessada a 02 de Maio de 2013
- QUANTUM GIS (2013) – *Página oficial do projeto Quantum GIS*, <http://www.qgis.org/en/about-qgis/qgis-releases/149-qgis-18-release.HTML>, acessada a 02 de Maio de 2013
- PRP (2013) – *Página oficial da Prevenção Rodoviária Portuguesa (valores para velocidades máximas permitidas)*, <http://www.velocidade.prp.pt/default.aspx?Page=4036>, acessada a 12 de Maio de 2013
- UNIVERSITEIT UTRECHT (2013) – *Página oficial do projeto Flowmap*, <http://flowmap.geog.uu.nl/whyuseflowmap.php>, acessada a 02 de Maio de 2013
- WIKIPÉDIA (2013) – *Página com velocidades médias de locomoção*, http://pt.wikipedia.org/wiki/Quil%C3%B4metro_por_hora, acessada a 12 de Maio de 2013
- WIKIPÉDIA (2013) – *Página com diversos Softwares de código aberto para estatística*, http://en.wikipedia.org/wiki/List_of_statistical_packages, acessada a 15 de Maio de 2013
- WU (2013) – *Página oficial do projeto R Project, Institute for Statistics and Mathematics of WU (Wirtschaftsuniversität Wien)*, <http://www.r-project.org/>, acessada a 15 de Maio de 2013

Referências a fontes de dados:

- DGT (2012) – *Página oficial da Direção Geral do Território (DGT) com Carta Administrativa Oficial de Portugal*, <http://www.igeo.pt/produtos/cadastro/caop/inicial.htm>, acedida a 02 de Maio de 2013
- INE (2013) - *Página dos Censos 2011 (INE)*, <http://mapas.ine.pt/download/index2011.pHTML>, acedida a 02 de Maio de 2013
- CTT (2013) – *Página dos CTT com ficheiros txt contendo Códigos Postais*, http://www.ctt.pt/feapl_2/app/open/tools.jspx?tool=1, acedida a 09 de Maio de 2013
- OPENSTREETMAP (2013) - *Página contendo shapefiles com dados do openstreetmap*, <http://download.geofabrik.de/openstreetmap/>, acedida a 02 de Maio de 2013

Anexos

Anexo 1 – Dicionário de dados

Tabelas das *shapefiles* na pasta “Dados”

Rede de estradas *OpenStreetMap* (tipo de geometria: linhas)

Rede de estradas extraída do OpenStreetMap modificada para o SIG-ATEP

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
ID	Número inteiro	20	Número único identificador da linha
TYPE	Texto	15	Tipo de estrada (ex.: secondary, residential)
ONEWAY	Número inteiro	1	Se tem um sentido (valor 1) ou não (valor 2)
BRIDGE	Número inteiro	1	Se a linha corresponde a uma ponte. Se sim tem valor 1, se não tem valor 0
TUNNEL	Número inteiro	1	Se a linha corresponde a um túnel. Se sim tem valor 1, se não tem valor 0
MAXSPEED	Número inteiro	6	Velocidade máxima permitida em Km/h
COMP	Número real	20	Comprimento da linha em metros
TEMP_VIAGE	Número real	20	Tempo de viagem em minutos

Locais de formação (tipo de geometria: pontos)

Locais de formação considerados no SIG-ATEP. Referem-se às sedes de junta de freguesia da NUT III Médio Tejo e instalações da ATEP

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
LATTITUDE	Número real	20	Latitude em sistema de referência WGS 84
LONGITUDE	Número real	20	Longitude em sistema de referência WGS 84
NOME	Texto	35	Nome do local de formação
ESTADO	Número inteiro	2	Estado do local de formação para considerar na análise de áreas de serviço. Valor 0 é considerado, valor -1 não é considerado.

Formandos geocodificados (tipo de geometria: pontos)

Resultado do processo de geocodificação com Google geocoder. Os registos referem-se a dados retirados do SIGO

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
Morada	Texto	80	Morada do formando (rua e número de polícia)
Cidade	Texto	80	Localidade da morada (cidade, vila, etc.)
Distrito	Texto	80	Distrito a que pertence a morada
País	Texto	80	País a que pertence a morada
Formando	Texto	80	Nome completo do formando
Telefone	Texto	80	Número de telefone ou telemóvel do formando

Áreas de serviço (tipo de geometria: linhas)*Áreas de serviço construídas com a ferramenta GRASS de análise de redes v.net.alloc*

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
LATTITUDE	Número real	20	Latitude em sistema de referência WGS 84
LONGITUDE	Número real	20	Longitude em sistema de referência WGS 84
NOME	Texto	35	Nome do local de formação
ESTADO	Número inteiro	2	Estado do local de formação para considerar na análise de áreas de serviço. Nesta tabela assume valor 0.

Rede iso (tipo de geometria: linhas)*Rede iso construída com ferramenta GRASS de análise de redes v.net.iso*

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
TEMPO	Número inteiro	20	Custo em tempo de viagem, assumindo a escala de valores: 15, 30, 45 e >45 minutos

Formandos geocodificados com atributos das áreas de serviço (tipo de geometria: pontos)*Resultado do processo de união por localização das shapefiles: formandos geocodificados e áreas de serviço*

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
Morada	Texto	80	Morada do formando (rua e número de polícia)
Cidade	Texto	80	Localidade da morada (cidade, vila, etc.)
Distrito	Texto	80	Distrito a que pertence a morada
País	Texto	80	País a que pertence a morada
Formando	Texto	80	Nome completo do formando
Telefone	Texto	80	Número de telefone ou telemóvel do formando
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
LATTITUDE	Número real	20	Latitude em sistema de referência WGS 84
LONGITUDE	Número real	20	Longitude em sistema de referência WGS 84
NOME	Texto	35	Nome do local de formação
ESTADO	Número inteiro	2	Estado do local de formação para considerar na análise de áreas de serviço. Nesta tabela assume valor 0

Formandos geocodificados com atributos da rede iso (tipo de geometria: pontos)

Resultado do processo de união por localização das shapefiles: formandos geocodificados e rede iso

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
Morada	Texto	80	Morada do formando (rua e número de polícia)
Cidade	Texto	80	Localidade da morada (cidade, vila, etc.)
Distrito	Texto	80	Distrito a que pertence a morada
País	Texto	80	País a que pertence a morada
Formando	Texto	80	Nome completo do formando
Telefone	Texto	80	Número de telefone ou telemóvel do formando
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
TEMPO	Número inteiro	20	Custo em tempo de viagem, assumindo a escala de valores: 15, 30, 45 e >45 minutos

Resultado da sobreposição entre formandos geocodificados e áreas de serviço (geometria: linhas)

Resultado do processo de sobreposição com ferramenta GRASS v.select.overlap

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS

Censos 2011 e indicador de necessidade de formação (INF) por concelho, (tipo de geometria: polígonos)*Dados da CAOP 2012, variáveis dos Censos 2011 e INF, para NUT III Médio Tejo*

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
DICOFRE	Texto	6	Identificador único de freguesia
FREGUESIA	Texto	31	Nome de freguesia
MUNICIPIO	Texto	22	Nome de concelho
GEO_COD	Texto	20	Identificador único de concelho
DISTRITO	Texto	8	Nome de distrito
TAA	Texto	15	Identificação do tipo de área administrativa
AREA_EA_HA	Número real	20	Valor área da área administrativa
AREA_T_HA	Número real	20	Valor total da área da Freguesia
Z_SCORE	Número real	20	Valor do INF
SCORE_CLAS	Texto	11	Categoria do INF na escala: Fraca, Potencial, Emergente e Forte
PRESENTES	Número inteiro	20	Total de indivíduos presentes
RESIDENTES	Número inteiro	20	Total de indivíduos residentes
RESIDENTES_H	Número inteiro	20	Total de homens residentes
RESIDENTES_M	Número inteiro	20	Total de mulheres residentes
RESI_25_64	Número inteiro	20	Indivíduos residentes com idade entre 25 e 64 anos
RESID_64	Número inteiro	20	Indivíduos residentes com idade superior a 64 anos
N_ANALF	Número inteiro	20	Indivíduos residentes sem saber ler nem escrever
RESID_1BAS	Número inteiro	20	Indivíduos residentes com o 1º ciclo do ensino básico completo
RESID_2BAS	Número inteiro	20	Indivíduos residentes com o 2º ciclo do ensino básico completo
RESID_3BAS	Número inteiro	20	Indivíduos residentes com o 3º ciclo do ensino básico completo
RESID_SEC	Número inteiro	20	Indivíduos residentes com o ensino secundário completo
RESID_PSEC	Número inteiro	20	Indivíduos residentes com o ensino pós-secundário
RESID_SUP	Número inteiro	20	Indivíduos residentes com um curso superior completo
RESI_DES1	Número inteiro	20	Indivíduos residentes desempregados à procura do 1º emprego
RESID_DESN	Número inteiro	20	Indivíduos residentes desempregados à procura de novo emprego
RESID_EMP	Número inteiro	20	Indivíduos residentes empregados
RESID_REF	Número inteiro	20	Indivíduos residentes pensionistas ou reformados
RESID_S_AC	Número inteiro	20	Indivíduos residentes sem atividade económica
PRIMARIO	Número inteiro	20	Indivíduos residentes empregados no sector primário
SECUNDAR	Número inteiro	20	Indivíduos residentes empregados no sector secundário
TERCIARIO	Número inteiro	20	Indivíduos residentes empregados no sector terciário

Censos 2011 e indicador de necessidade de formação (INF) por freguesia, (tipo de geometria: polígonos)*Dados da CAOP 2012, variáveis dos Censos 2011 e INF, para NUT III Médio Tejo*

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
DICOFRE	Texto	6	Identificador único de freguesia
FREGUESIA	Texto	31	Nome de freguesia
MUNICIPIO	Texto	22	Nome de concelho
GEO_COD	Texto	20	Identificador único de concelho
DISTRITO	Texto	8	Nome de distrito
TAA	Texto	15	Identificação do tipo de área administrativa
AREA_EA_HA	Número real	20	Valor área da área administrativa
AREA_T_HA	Número real	20	Valor total da área da Freguesia
Z_SCORE	Número real	20	Valor do INF
SCORE_CLAS	Texto	11	Categoria do INF na escala: Fraca, Potencial, Emergente e Forte
PRESENTES	Número inteiro	20	Total de indivíduos presentes
RESIDENTES	Número inteiro	20	Total de indivíduos residentes
RESIDENTES_H	Número inteiro	20	Total de homens residentes
RESIDENTES_M	Número inteiro	20	Total de mulheres residentes
RESI_25_64	Número inteiro	20	Indivíduos residentes com idade entre 25 e 64 anos
RESID_64	Número inteiro	20	Indivíduos residentes com idade superior a 64 anos
N_ANALF	Número inteiro	20	Indivíduos residentes sem saber ler nem escrever
RESID_1BAS	Número inteiro	20	Indivíduos residentes com o 1º ciclo do ensino básico completo
RESID_2BAS	Número inteiro	20	Indivíduos residentes com o 2º ciclo do ensino básico completo
RESID_3BAS	Número inteiro	20	Indivíduos residentes com o 3º ciclo do ensino básico completo
RESID_SEC	Número inteiro	20	Indivíduos residentes com o ensino secundário completo
RESID_PSEC	Número inteiro	20	Indivíduos residentes com o ensino pós-secundário
RESID_SUP	Número inteiro	20	Indivíduos residentes com um curso superior completo
RESI_DES1	Número inteiro	20	Indivíduos residentes desempregados à procura do 1º emprego
RESID_DESN	Número inteiro	20	Indivíduos residentes desempregados à procura de novo emprego
RESID_EMP	Número inteiro	20	Indivíduos residentes empregados
RESID_REF	Número inteiro	20	Indivíduos residentes pensionistas ou reformados
RESID_S_AC	Número inteiro	20	Indivíduos residentes sem atividade económica
PRIMARIO	Número inteiro	20	Indivíduos residentes empregados no sector primário
SECUNDAR	Número inteiro	20	Indivíduos residentes empregados no sector secundário
TERCIARIO	Número inteiro	20	Indivíduos residentes empregados no sector terciário

Censos 2011 e indicador de necessidade de formação (INF) por subsecção estatística, (tipo de geometria: polígonos)

Dados da CAOP 2012 disponibilizada pelo INE, variáveis dos Censos 2011 e INF, para NUT III Médio Tejo

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
DTMN11	Texto	4	Identificador único de concelho
FR11	Texto	2	Identificador de freguesia no concelho
SEC11	Texto	3	Identificador de secção estatística na freguesia
SS11	Texto	2	Identificação da subsecção estatística na secção estatística
BGRI11	Texto	11	Base Geográfica de Referenciação da Informação 2011 que identifica de forma unívoca a subsecção estatística
LUG11	Texto	6	Identificador de lugar INE
LUG11DESIG	Texto	27	Nome de lugar INE
Z_SCORE	Número real	20	Valor do INF
SCORE_CLAS	Texto	11	Categoria do INF na escala: Fraca, Potencial, Emergente e Forte
PRESENTES	Número inteiro	20	Total de indivíduos presentes
RESIDENTES	Número inteiro	20	Total de indivíduos residentes
RESIDENTES_H	Número inteiro	20	Total de homens residentes
RESIDENTES_M	Número inteiro	20	Total de mulheres residentes
RESI_25_64	Número inteiro	20	Indivíduos residentes com idade entre 25 e 64 anos
RESID_64	Número inteiro	20	Indivíduos residentes com idade superior a 64 anos
N_ANALF	Número inteiro	20	Indivíduos residentes sem saber ler nem escrever
RESID_1BAS	Número inteiro	20	Indivíduos residentes com o 1º ciclo do ensino básico completo
RESID_2BAS	Número inteiro	20	Indivíduos residentes com o 2º ciclo do ensino básico completo
RESID_3BAS	Número inteiro	20	Indivíduos residentes com o 3º ciclo do ensino básico completo
RESID_SEC	Número inteiro	20	Indivíduos residentes com o ensino secundário completo
RESID_PSEC	Número inteiro	20	Indivíduos residentes com o ensino pós-secundário
RESID_SUP	Número inteiro	20	Indivíduos residentes com um curso superior completo
RESI_DES1	Número inteiro	20	Indivíduos residentes desempregados à procura do 1º emprego
RESID_DESN	Número inteiro	20	Indivíduos residentes desempregados à procura de novo emprego
RESID_EMP	Número inteiro	20	Indivíduos residentes empregados
RESID_REF	Número inteiro	20	Indivíduos residentes pensionistas ou reformados
RESID_S_AC	Número inteiro	20	Indivíduos residentes sem atividade económica
PRIMARIO	Número inteiro	20	Indivíduos residentes empregados no sector primário
SECUNDAR	Número inteiro	20	Indivíduos residentes empregados no sector secundário
TERCIARIO	Número inteiro	20	Indivíduos residentes empregados no sector terciário

Tabelas na pasta “ATEP”

Locais de formação (tipo de geometria: pontos)

Locais de formação considerados no SIG-ATEP. Referem-se às sedes de junta de freguesia da NUT III Médio Tejo e instalações da ATEP

Atributo	Tipo	Tamanho (caracteres)	Descrição
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
LATTITUDE	Número real	20	Latitude em sistema de referência WGS 84
LONGITUDE	Número real	20	Longitude em sistema de referência WGS 84
NOME	Texto	35	Nome do local de formação
ESTADO	Número inteiro	2	Estado do local de formação para considerar na análise de áreas de serviço. Valor 0 é considerado, valor -1 não é considerado.

Estradas (tipo de geometria: linhas)

Rede de estradas extraída do OpenStreetMap modificada para o SIG-ATEP

Atributo	Tipo	Tamanho (caracteres)	Descrição
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
ID	Número inteiro	20	Número único identificador da linha
TYPE	Texto	15	Tipo de estrada (ex.: <i>secondary, residential</i>)
ONEWAY	Número inteiro	1	Se tem um sentido (valor 1) ou não (valor 2)
BRIDGE	Número inteiro	1	Se a linha corresponde a uma ponte. Se sim tem valor 1, se não tem valor 0
TUNNEL	Número inteiro	1	Se a linha corresponde a um túnel. Se sim tem valor 1, se não tem valor 0
MAXSPEED	Número inteiro	6	Velocidade máxima permitida em Km/h
COMP	Número real	20	Comprimento da linha em metros
TEMP_VIAGE	Número real	20	Tempo de viagem em minutos

Formandos (tipo de geometria: pontos)

Resultado do processo de geocodificação com Google geocoder. Os registos referem-se a dados retirados do SIGO

Atributo	Tipo	Tamanho (caracteres)	Descrição
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
Morada	Texto	80	Morada do formando (rua e número de polícia)
Cidade	Texto	80	Localidade da morada (cidade, vila, etc.)
Distrito	Texto	80	Distrito a que pertence a morada
País	Texto	80	País a que pertence a morada
Formando	Texto	80	Nome completo do formando
Telefone	Texto	80	Número de telefone ou telemóvel do formando
ESTADO	Número inteiro	2	Estado do local de formação para considerar na análise de áreas de serviço. Valor 0 é considerado, valor -1 não é considerado.

Rede iso (tipo de geometria: linhas)

Rede iso construída com ferramentas GRASS de análise de redes

<i>Atributo</i>	<i>Tipo</i>	<i>Tamanho (caracteres)</i>	<i>Descrição</i>
cat	Número inteiro	10	Número único identificador do ponto atribuído pelo GRASS
TEMPO	Número inteiro	20	Custo em tempo de viagem, assumindo a escala de valores: 15, 30, 45 e >45 minutos

Anexo 2 – Listagem de código usado em macros VBA Microsoft Office Excel

a) Macro “velocidade_max” usada na atribuição de velocidade máxima permitida à rede de estradas *OpenStreetMap*

```
Sub velocidade_max()
```

```
‘ Macro para atribuição automática da velocidade máxima às estradas, em função do tipo de estrada
```

```
Dim myList, myRange
```

```
Set myList = Sheets("Folha1").Range("A:A") ‘ Lista com valores das velocidades máximas
```

```
Set myRange = Sheets("roads_Medio_Tejo_WGS84").Range("H:H") ‘ Intervalo de dados para atribuição de velocidades
```

```
‘ Ciclo for que permite percorrer a lista com tipo de estrada e valores de velocidade
```

```
For Each cel In myList.Columns(1).Cells
```

```
‘ Atribuição da velocidade através da função localizar e substituir..  
myRange.Replace What:=cel.Value, Replacement:=cel.Offset(0, 1).Value,  
LookAt:=xlWhole
```

```
Next cel
```

```
End Sub
```

b) Macro “geocoding” usada na organização dos dados para geocodificação

Listagem do código em formato *TXT* e ficheiro *Excel* com a macro “geocoding” estão disponíveis na subpasta “geocoding_xls”, localizada na pasta “Anexos_Dissertacao” no CD-ROM fornecido.

c) Macro “alocacao_formandos” usada na otimização da localização dos locais de formação

Listagem do código em formato *TXT* e ficheiro *Excel* com a “alocacao_formandos” macro estão disponíveis na subpasta “alocacao_xls”, localizada na pasta “Anexos_Dissertacao” no CD-ROM fornecido.

d) Macro "categorias_INF" usada na atribuição de categorias ao Indicador de Necessidade de Formação (INF)

Sub categorias_INF()

' Macro para atribuição automática de categorias aos valores do INF. Como a amplitude das classes de valores varia entre os níveis: concelho, freguesia e subsecção estatística, as condições para atribuição de categorias variam ligeiramente. Na prática fazem-se variar os valores da "cell4" dentro do ciclo for. A restante sintaxe mantém-se igual para as três situações.

Range("Folha2!A:A").Select ' Definição de intervalo de dados a editar

Dim cell4 As Range ' Definição de célula para valores no intervalo

' Ciclo for que permite percorrer as linhas no intervalo de dados e substituir

For Each cell4 In Selection

' Imposição de condições para atribuição das categorias

If cell4.Value < -1 Then

cell4.Value = "1_Fraca"

ElseIf cell4.Value >= 2 Then

cell4.Value = "4_Forte"

ElseIf cell4.Value >= 1 Then

cell4.Value = "3_Emergente"

ElseIf cell4.Value >= -1 Then

cell4.Value = "2_Potencial"

End If

Next cell4

End Sub

Anexo 3 – Manual de utilização