



**USO DE DIFERENTES ALGORITMOS DE OTIMIZAÇÃO PARA
DEFINIÇÃO DE ÁREAS DE SERVIÇO DE ESQUADRAS DE
POLÍCIA PARA PORTUGAL**

André Filipe Mendes Duarte

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre em Ciência e Sistemas
de Informação Geográfica

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**USO DE DIFERENTES ALGORITMOS DE OTIMIZAÇÃO PARA
DEFINIÇÃO DE ÁREAS DE SERVIÇO DE ESQUADRAS DE
POLÍCIA PARA PORTUGAL**

Dissertação orientada por:

Orientador: Professor Doutor Roberto Henriques

Coorientadora: Professora Doutora Sara Ribeiro

Lisboa, 27 de Fevereiro de 2019

DECLARAÇÃO DE ORIGINALIDADE

Declaro que o trabalho contido neste documento é da minha autoria e não de outra pessoa. Toda a assistência recebida de outras pessoas está devidamente assinalada e é efetuada referência a todas as fontes utilizadas (publicadas ou não).

O trabalho não foi anteriormente submetido ou avaliado na NOVA *Information Management School* ou em qualquer outra instituição.

Lisboa, 27 de Fevereiro de 2019

André Filipe Mendes Duarte

AGRADECIMENTOS

Chegou ao fim mais uma etapa que dependeu direta ou indiretamente de um conjunto de pessoas, às quais quero agradecer por fazerem parte.

Agradeço ao Professor Roberto Henriques e à Professora Sara Ribeiro pela orientação e paciência.

A todos os docentes e colegas de mestrado que tive a oportunidade de conhecer com quem muito aprendi.

À administração e colegas do RAIZ (*part of The Navigator Company*) pela compreensão e pelos diversos incentivos.

À comunidade *Python* Portugal, onde existe sempre alguém do outro lado com quem partilhar e discutir ideias. Obrigados a todos da comunidade.

Ao Professor Hugo Martins pelo incentivo da utilização das bibliotecas da linguagem *Python* e ao Luís Calisto pelas dicas na otimização do código.

Quero deixar uma palavra de agradecimento à Professora Beatriz Fidalgo e ao Professor Raúl Salas pela amizade e pelos sábios conselhos ao longo destes anos todos.

Cindy, à mulher da minha vida que esteve sempre comigo neste percurso que também foi o dela. Por não desistir e acreditar apesar das grandes adversidades e dificuldades. Muito obrigado pelo amor e dedicação.

Por fim, a todos familiares e amigos que de uma forma ou de outra contribuíram nesta fase.

À Cindy e aos meus pais

“Anything's possible if you've got enough nerve.”

J.K. Rowling

Uso de diferentes algoritmos de otimização para definição de áreas de serviço de esquadras de polícia para Portugal

RESUMO

A segurança é considerada um direito fundamental nos estados democráticos. É uma condição que exige cada vez mais ferramentas avançadas de análise espacial, para apoiar a adequação de recursos e a disposição espacial das forças de segurança. A reorganização das forças de segurança no contexto atual depende da distribuição da população e do seu dinamismo. O processo de agrupar pequenas áreas geográficas para formar as áreas de serviço é designado por *districting*.

Esta dissertação tem como objetivo a otimização espacial das forças de segurança tendo em consideração a distribuição espacial da população residente do distrito administrativo de Setúbal.

A análise de dados da população residente foi efetuada com a análise *hot spot* para estudar a sua distribuição ao nível da freguesia. Implementou-se um algoritmo genético e efetuaram-se testes experimentais com os dados da população residente e dos grupos vulneráveis, para criar áreas de serviço das esquadras de polícia. Compararam-se os resultados obtidos com o *Automatic Zoning Procedure – Simulated Annealing (AZP-SA)* utilizando os mesmos dados.

A população do distrito concentra-se sobretudo na península de Setúbal, existindo uma grande assimetria na sua distribuição geográfica. Os testes experimentais com o algoritmo genético demonstram que as áreas de serviço criadas, apresentam uma soma das diferenças da população com uma grande variação. O AZP-SA obteve um desempenho ligeiramente superior ao algoritmo genético implementado.

A implementação do algoritmo permitiu obter soluções de áreas de serviço, no entanto, o desempenho do AZP-SA foi ligeiramente superior. A grande assimetria da população do distrito administrativo de Setúbal dificultou a criação de áreas de serviço mais equitativas.

Use of different optimization algorithms to define service areas of police stations in Portugal

ABSTRACT

Security is to consider being a fundamental right in democratic societies. It is a condition that requires advanced spatial analysis tools to support security forces in the spatial disposition and adequacy of resources. Nowadays security forces reorganization, depends on the population distribution and dynamic. Districting is the process of grouping small geographic areas in service areas.

The main objective of this thesis is the spatial optimization of security forces considering the spatial disposition of the population in Setubal administrative district.

Data analysis was done with the hot spot analysis to study the population and their distribution at *freguesia* level. A genetic algorithm was implemented to create service areas and experimental tests were performed with the population data and vulnerable groups. We compared the results with the Automatic Zoning Procedure - Simulated Annealing (AZP-SA).

The population concentrates in Setubal peninsula denoting a great asymmetry on their geographical distribution. Experimental tests with the genetic algorithm show a large variation of a sum of the population differences in service areas. AZP-SA performed better than the genetic algorithm.

The solutions for the service areas were obtained with the genetic algorithm. However, the performance of AZP-SA is slightly higher. The difficult to obtain equitable areas is due the great asymmetry of the population.

PALAVRAS-CHAVE

Esquadras de polícia

Otimização

Algoritmos genéticos

Áreas de serviço

Districting

AZP-SA

KEYWORDS

Police Station

Optimization

Genetic Algorithms

Service Areas

Districting

AZP-SA

ACRÓNIMOS

ADN	Ácido Desoxirribonucleico
AG	Algoritmo genético
AS	Áreas de serviço
AZP	<i>Automatic Zoning Procedure</i>
AZP-SA	Automatic Zoning Procedure - Simulated annealing
C	Linguagem de programação C
CAOP	Carta Administrativa Oficial de Portugal
CBPSB	<i>Constraint-Based Polygonal Spatial Clustering</i>
CE	Computação evolutiva
DGPJ	Direção Geral da Política de Justiça
DGT	Direção Geral do Território
ESRI	<i>Environmental Systems Research Institute</i>
EUA	Estados Unidos da América
EUROSTAT	<i>Statistical Office of the European Communities</i>
FCT	Fundação para a Ciência e Tecnologia
GeoSOM	<i>Geospatial Self Organizing Maps</i>
GNR	Guarda Nacional Republicana
GP-TS	<i>Graph-Partition-and-Tabu-Search</i>
hab./km ²	Habitante por quilómetro quadrado
IA	Inteligência Artificial
IDE	<i>Integrated development Environment</i>
INE	Instituto Nacional de Estatística
km	Quilómetros
km ²	Quilómetros quadrados
n.º hab.	número de habitantes
NUT	Nomenclatura das Unidades Territoriais
PCAM	<i>Patrol Car Allocation Model</i>
PDP	<i>Police Districting Problem</i>
PJ	Polícia Judiciária
PORDATA	Base de dados de Portugal Contemporâneo
PPAC	<i>Police Patrol Area Covering</i>

PSP	Polícia de Segurança Pública
PySAL	<i>Python Spatial Analysis Library</i>
r-districts	<i>reporting-districts</i>
RP-TS	<i>Random-Greedy-and-Tabu-Search</i>
SA	Simulated annealing
SIEJ	Estatísticas da Justiça
SIG	Sistemas de Informação Geográfica
SIM4SECURITY	Modelo Prospetivo e de Análise Espacial aplicado à Segurança Pública
SNPDP	<i>Street-Network Police Districting</i>
SOM	<i>Self Organizing Maps</i>
TS	<i>Tabu Search</i>
ZDES	<i>Zone-Design System</i>

INDICE DO TEXTO

AGRADECIMENTOS.....	iii
RESUMO.....	vi
ABSTRACT	vii
PALAVRAS-CHAVE	viii
KEYWORDS.....	viii
ACRÓNIMOS.....	ix
INDICE DE TABELAS.....	xiv
INDICE DE FIGURAS.....	xvi
1. Introdução.....	1
1.1. Contexto	1
1.2. Objetivos	2
1.3. Organização da dissertação	2
2. Descrição do Problema.....	4
2.1. Contextualização	4
2.1.1. A evolução da População Portuguesa	5
2.1.2. A distribuição das forças de segurança em Portugal Continental.....	8
2.1.3. Dados sobre criminalidade	11
2.2. Caso de estudo	13
3. Revisão da Literatura.....	14
3.1. Introdução	14
3.2. Visão geral dos problemas de <i>Districting</i>	14
3.3. <i>Police Districting Problem (PDP)</i>	20
3.3.1. Trabalhos de referência.....	21
3.4. Otimização	25
3.4.1. O processo de otimização	26

3.4.2.	Modelos de otimização espacial	27
3.5.	Meta-heurísticas	30
3.5.1.	Algoritmos Genéticos (AG)	30
3.5.2.	Terminologia e conceitos básicos	31
3.5.3.	Funcionamento básico de um AG	32
3.5.3.1.	Codificação.....	33
3.5.3.2.	População	34
3.5.3.3.	Função de Avaliação (<i>Fitness</i>)	35
3.5.3.4.	Métodos de seleção.....	35
3.5.3.5.	Operadores genéticos	36
3.5.3.6.	Estratégias evolucionárias	38
3.5.3.7.	Critérios de paragem	39
3.5.4.	Vantagens e limitações.....	39
3.5.5.	Principais áreas de aplicação	40
4.	Metodologia	41
4.1.	Área de estudo	41
4.2.	Dados e software.....	42
4.3.	<i>Hot Spot Analysis</i>	43
4.4.	Modelo de otimização para <i>police re(districting)</i>	43
4.4.1.	Proposta de resolução.....	44
4.5.	Implementação do AG	45
4.5.1.	Testes com diferentes parametrizações	47
4.5.2.	Definição de áreas de serviço.....	48
4.6.	Comparação com o AZP-SA.....	48
5.	Resultados e Discussão	49
5.1.	Análise exploratória de dados	49
5.1.1.	Estatística descritiva	49
5.1.2.	Análise da população por faixa etária de interesse	51

5.2. Resultados experimentais com o AG.....	53
5.2.1. Testes com as diferentes parametrizações	54
5.2.2. Definição de áreas de serviço com o AG	54
5.2.2.1. Comparação com o AZP-SA.....	58
5.2.3. Definição de áreas de serviço para a faixa etária maior que 65 anos.....	60
5.2.3.1. Comparação com o AZP-SA.....	62
6. Considerações finais	64
7. Referências bibliográficas	67
8. Anexos	74
8.1. Código da estatística descritiva	74
8.2. Código do algoritmo genético	75
8.3. Código do AZP-SA	80
8.4. Soluções obtidas com diferentes parametrizações com o AG	81
8.5. Identificação por freguesia das áreas de serviço determinadas.....	82
8.6. Média e soma da população residente por área de serviço	83

INDICE DE TABELAS

Tabela 1 - População residente em Portugal entre 1940 e 2011 (Fonte: INE, (2014)) ..5	
Tabela 2 - Tipos de crimes cometidos em 2011 e 2016 (PORDATA, 2018 e INE, 2017).	12
Tabela 3 - Comparação entre a terminologia biológica e computacional (Fonte: Sivanandam & Deepa, 2008).	32
Tabela 4 - Tema, formato e fonte dos dados.	42
Tabela 5 - Estatística descritiva da população residente e área no distrito de Setúbal.	50
Tabela 6 - Teste realizados com a área de estudo com diferentes parametrizações das populações.	54
Tabela 7 - Estatística descritiva da população afetada por área de serviço com a equação IX.	56
Tabela 8 - Estatística descritiva da população afetada por área de serviço com a equação XI.	57
Tabela 9 - Estatística descritiva da população afetada por área de serviço com o AZP-SA.....	59
Tabela 10 - Comparação entre os resultados obtidos entre o AG e o AZP-SA para as mesmas condições.	60
Tabela 11 - Estatística descritiva da população afetada por área de serviço com a equação XI.	61
Tabela 12 - Estatística descritiva da população afetada por área de serviço com o AZP-SA.....	63
Tabela 13 - Comparação entre os resultados obtidos entre o AG e o AZP-SA para as mesmas condições.	63
Tabela 14 – Todas as soluções obtidas com diferentes parametrizações.....	81
Tabela 15 - Identificação por freguesia das áreas de serviço criadas.	82
Tabela 16 - Média e soma da população residente por área de serviço com a equação IX.	83
Tabela 17 - Média e soma da população residente por área de serviço com a equação XI.	84
Tabela 18 - Média e soma da população residente por área de serviço com o AZP-SA.	85

Tabela 19 - Média e soma da população residente por área de serviço com a equação XI >65 anos.	86
Tabela 20 - Média e soma da população com mais de 65 anos por área de serviço com o AZP-SA.....	87

INDICE DE FIGURAS

Figura 1 - Esquema das tarefas do projeto SIM4SECURITY (Fonte: http://sim4security.novaims.unl.pt/).....	4
Figura 2 – População residente por município (a) e densidade populacional no ano 2011(b).....	6
Figura 3 - Variação da População por Município entre 2011 e 2016.....	7
Figura 4 - Distribuição dos postos territoriais e esquadras das forças de segurança em Portugal Continental.....	8
Figura 5 - Número de pessoas ao serviço da GNR e PSP (Fonte: INE,2017).....	9
Figura 6 - Número de pessoas ao serviço por cada 100 mil habitantes (Fonte: INE, 2017).	10
Figura 7 - Número de Polícias por cada 100.000 habitantes no ano de 2016 (Fonte: https://ec.europa.eu/eurostat/statistics-explained/)	10
Figura 8 - Números de crimes registados pelas policias em 2011 (a) e em 2016 (b). .	12
Figura 9 - Cartoon original do círculo eleitoral com a forma de salamandra (Levit & Wood, 2010).	16
Figura 10 - Diferentes tipos de grafos apresentados por Kalsics (2015).	19
Figura 11 - Dois tipos de partições com diferentes valores de compacidade (Kalsics et al., 2005).....	20
Figura 12 – Distritos de comando (direita) e os sectores de patrulha da.....	21
Figura 13 - Processo clássico de otimização (Adaptado de Chinneck, 2000).	27
Figura 14 - Algoritmo genético clássico (Simões, 1999).....	33
Figura 15 - Exemplo de codificação binária de 2 cromossomas com 10 genes (Fonte: Elaboração própria).	33
Figura 16 - Exemplo de codificação real de um cromossoma com 3 genes (Fonte: Elaboração própria).	34
Figura 17 - Exemplo de codificação por permutação (Sivanandam & Deepa, 2008)...	34
Figura 18 - Exemplo de cruzamento de um ponto de corte (Fonte: Elaboração própria).	36
Figura 19- Exemplo de cruzamento com dois pontos de corte (Fonte: Elaboração própria)	37
Figura 20 - Exemplo de cruzamento uniforme (Fonte: Sivanandam & Deepa, 2008)..	37
Figura 21 - Mutação por troca de 1 bit (Fonte: Elaboração própria).	38
Figura 22 - Mutação por troca de posição (Fonte: Elaboração própria).....	38

Figura 23 - Mutaç�o por adiç�o de um valor (Fonte: Elaboraç�o pr�pria).....	38
Figura 24 - Localizaç�o geogr�fica da �rea de estudo (Fonte: CAOP, 2017).	41
Figura 25 - Representaç�o dos centr�ides das freguesias e dos distritos (Fonte: Elaboraç�o pr�pria).	46
Figura 26 - Exemplo de uma matriz de contiguidade (Fonte: Elaboraç�o pr�pria).	46
Figura 27 - Populaç�o residente por freguesia representada pelo m�todo dos quantis em 2011.....	49
Figura 28 - Histograma da distribuiç�o da populaç�o residente por freguesia em 2011 (Fonte: INE, 2014b).	51
Figura 29 - An�lise hotspot realizada para a populaç�o residente (a) e para a faixa et�ria maior que 65 anos (b) do ano 2011.	52
Figura 30 - An�lise hotspot realizada para a faixa et�ria entre 20 e 24 anos (a) e para a faixa et�ria entre 24 e 64 anos (b).	53
Figura 31 - Desenho das �reas de serviço sem o fator dist�ncia tendo em conta a populaç�o total parcial por freguesia para 2011.....	55
Figura 32 - Desenho das �reas de serviço com o fator dist�ncia tendo em conta a populaç�o total parcial por freguesia para 2011.....	57
Figura 33 - �reas de serviço com o AZP-SA para a populaç�o residente.	59
Figura 34 - Desenho das �reas de serviço com o fator dist�ncia tendo em conta a populaç�o residente com mais de 65 anos, parcial por freguesia para 2011.	61
Figura 35 - �reas de serviço com o AZP-SA para a populaç�o com mais de 65 anos.	62

1. Introdução

1.1. Contexto

A segurança é tida como um direito fundamental para qualquer cidadão. No contexto de globalização em que vivemos, a segurança exige cada vez mais, novas abordagens e estratégias, visto que o crime é cada vez mais diversificado e organizado.

Aliado aos fenómenos de globalização deparamo-nos com a grande dinâmica da população o que obriga uma adequação das forças de segurança, quer ao nível de recursos quer ao nível da distribuição espacial (Painho & Ribeiro, 2018). A delimitação das áreas de serviço é uma das maiores preocupações das forças de segurança pelo que existe a necessidade de definir uma estratégia de distribuição do dispositivo territorial e da gestão das operações de forma a minimizar os custos operacionais (Henriques et al. 2018).

A delimitação de áreas de atuação ou de circunscrição para um determinado objetivo, é designada por *Districting*, e tem sido aplicada nas áreas da política, nos serviços de emergência, para as áreas de influência das escolas, regiões de vendas, entre outros (Kalsics, 2015). Segundo Mitchell (1972), as delimitações de áreas de ação são definidas e desenhadas pelos comandos das forças de segurança, o que de facto, se traduz numa tarefa complexa decorrente de diversas condicionantes e problemas associados.

Em Portugal, existem vários organismos públicos de forças e de serviços de segurança que garantem a segurança interna. De acordo com o artigo 25º, alínea 2 da Lei n.º53/2008 de 29 de Agosto, os principais órgãos com funções de segurança interna são a Guarda Nacional Republicana (GNR), a Polícia de Segurança Pública (PSP), Polícia Judiciária (PJ), entre outros.

Hoje em dia, temos à disposição um conjunto de ferramentas de apoio à decisão apoiadas na investigação operacional e na computação evolutiva (CE), como por exemplo a otimização combinatória. Importa aqui referir que a otimização é a arte e a ciência de alocar recursos escassos numa determinada tarefa.

1. Introdução

As técnicas de otimização, matematicamente visam determinar os extremos de funções (máximos e mínimos) em inúmeros domínios (Chinneck, 2000). Na prática é utilizada por cientistas, engenheiros e gestores para o apoio à decisão. Os métodos de otimização são utilizados na indústria, na engenharia e nos governos centrais, visto que a necessidade de minimizar custos e aumentar as receitas pode significar por vezes a falência técnica de um governo ou de uma indústria. Esta área de conhecimento é de igual modo, de extrema importância e aplicabilidade quando falamos de segurança dos cidadãos.

Esta dissertação surge no âmbito do projeto SIM4SECURITY – Modelo prospetivo e de Análise Espacial aplicado à Segurança Pública, financiado pela Fundação para a Ciência e Tecnologia (FCT) através do projeto PTDC/ATPDEM/1538/2014. Este projeto tem como objetivo criar uma solução tecnológica de apoio à decisão que permita, através dos Sistemas de Informação Geográfica (SIG) e de cenários demográficos, melhorar a eficácia da atividade operacional das forças de segurança.

1.2. Objetivos

Esta dissertação tem como objetivo propor novas metodologias de otimização da distribuição das forças de segurança, tendo em consideração a distribuição espacial e as características socioeconómicas da população em Portugal Continental. Os objetivos específicos são:

- Criar regiões espacialmente contíguas, através de ferramentas avançadas de análise espacial, como os algoritmos genéticos e com o *Automatic Zoning Procedure - Simulated Annealing (AZP-SA)*;
- Aplicação prática no distrito de administrativo de Setúbal tendo em conta a população residente por freguesia e os grupos etários mais vulneráveis, como a população residente com mais de 65 anos;
- Fornecer uma ferramenta de apoio à decisão que contribua para a melhoria do desempenho do dispositivo das forças de segurança.

1.3. Organização da dissertação

Esta dissertação está organizada da seguinte forma:

1. Introdução

- No segundo capítulo efetua-se a descrição e a contextualização do problema em estudo destacando os enfoques atuais;
- No terceiro capítulo efetua-se a revisão da literatura sobre os problemas de *districting* e *police districting*. Apresentam-se os métodos de resolução clássicos e da computação evolutiva.
- No quarto capítulo é apresentada a metodologia onde se efetua a descrição da análise exploratória de dados a aplicar, posteriormente aplicou-se um Algoritmo Genético (AG) e o AZP-SA para determinar potenciais áreas de serviço das forças de segurança.
- No quinto capítulo apresentam-se e discutem-se os resultados da análise exploratória e das experiências realizadas com diferentes funções de avaliação e algoritmos.
- No sexto capítulo apresentam-se as considerações finais, as conclusões obtidas, as contribuições, as recomendações e potenciais desenvolvimentos futuros.

2. Descrição do Problema

2.1. Contextualização

O problema em estudo resulta do projeto de investigação SIM4SECURITY (Modelo Prospetivo e de Análise Espacial aplicado à Segurança Pública). Os dois principais problemas que foram considerados no âmbito deste projeto prendem-se com a otimização da distribuição de recursos humanos, as infraestruturas das forças de segurança e a consciência do carácter dinâmico da reorganização territorial do sistema policial (Painho & Ribeiro, 2018). Esta dissertação procurou contribuir para a tarefa número 6 (Figura 1), que diz respeito à modelação da distribuição espacial das forças de segurança.

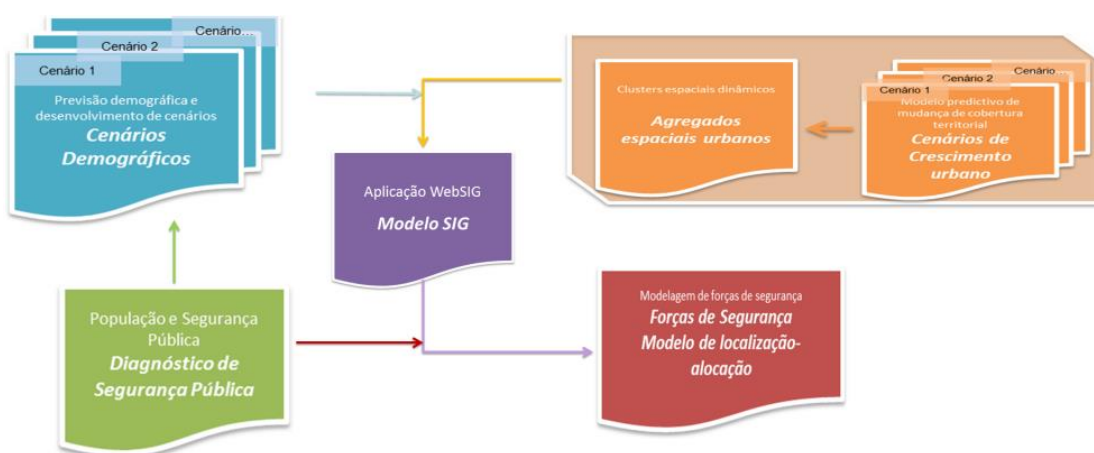


Figura 1 - Esquema das tarefas do projeto SIM4SECURITY (Fonte: <http://sim4security.novaims.unl.pt/>).

A tarefa número 6 tem como objetivos criar um modelo de localização-alocação, para as áreas de atuação das forças de segurança procurando uma distribuição simétrica e otimizada dos recursos em todo o território, tendo em consideração a distribuição da população, otimizando o rácio polícia/cidadão através de duas abordagens, os *Self Organizing Maps* (SOM) e Algoritmos genéticos (AG) (Painho & Ribeiro, 2018).

2. Descrição do Problema

Nesta dissertação implementou-se um algoritmo genético para criar regiões contíguas tendo em consideração a população residente. Por outro lado, no projeto (SIM4SECURITY), foram utilizados os *Self Organizing maps* (SOM) para agrupar os dados de entrada em regiões espacialmente consistentes de atributos socioeconómicos semelhantes (Painho & Ribeiro, 2018).

Este capítulo tem como objetivos descrever e enquadrar o problema, através de dados fornecidos pelo projeto e da pesquisa realizada. Numa primeira fase, observaram-se os grandes números da população portuguesa e a sua evolução, visto que as mudanças demográficas têm influência no rácio polícia/cidadão. Na segunda fase, realizou-se uma análise aos dados disponíveis sobre as forças de segurança em Portugal Continental e atual distribuição que de igual modo, são imprescindíveis para o planeamento securitário. Na terceira fase, procurou-se analisar os dados sobre os crimes em Portugal Continental. Por último, foi feita uma contextualização de um caso de estudo utilizando dados da população residente.

2.1.1. A evolução da População Portuguesa

A população portuguesa tem aumentado desde que há registo, com exceção do período entre 1960 e 1970, onde se verificou um decréscimo de 2,5% (Tabela 1). O maior crescimento teve lugar entre 1970 e 1981 com uma variação percentual de 13,5%. De acordo com Bandeira et al. (2014), este aumento relaciona-se com o regresso de cidadãos das ex-colónias e pelo retorno de emigrantes.

Ano do Recenseamento	População Total	Varição Percentual (%)
1940	7755423	
1950	8510240	9,7
1960	8889392	4,5
1970	8663252	-2,5
1981	9833014	13,5
1991	9867147	0,3
2001	10356117	5,0
2011	10562178	2,0

Tabela 1 - População residente em Portugal entre 1940 e 2011 (Fonte: INE, (2014))

Não obstante, os dados recolhidos pelas campanhas censitárias demonstram um aumento da população, verifica-se que pelas estimativas da população residente efetuadas pelo INE em 2018, a população tem diminuído desde 2011. Em 2017, a

2. Descrição do Problema

população total residente em Portugal era de 10291027 indivíduos, menos 18 546 face ao ano de 2016 (INE, 2017).

De acordo com Rodrigues et al. (2008), as profundas mudanças sociais, económicas e políticas que têm ocorrido desde meados do século XX, promoveram uma grande assimetria no território devido aos comportamentos migratórios que por consequência, potenciaram diferenças acentuadas da população residente entre o litoral e o interior do país. Os progressos científicos e técnicos refletiram-se de igual modo na população, nomeadamente na esperança média de vida e por outro lado, no decréscimo da fecundidade (Rodrigues & Ferreira, 2014). A diminuição da taxa de fecundidade e a intensificação dos processos migratórios tiveram impacto no aumento do envelhecimento da população promovendo o decréscimo do desenvolvimento económico (Moreira & Rodrigues, 2004). Se observarmos os dados censitários de 2011 (Figura 2a e 2b), os municípios com maior número de residentes e densidade populacional situam-se na faixa litoral, com particular destaque para a região da grande Lisboa e Porto, onde de acordo com Moreira & Rodrigues (2004) e Bravo et al.(2018), se situa a população da faixa etária mais jovem.

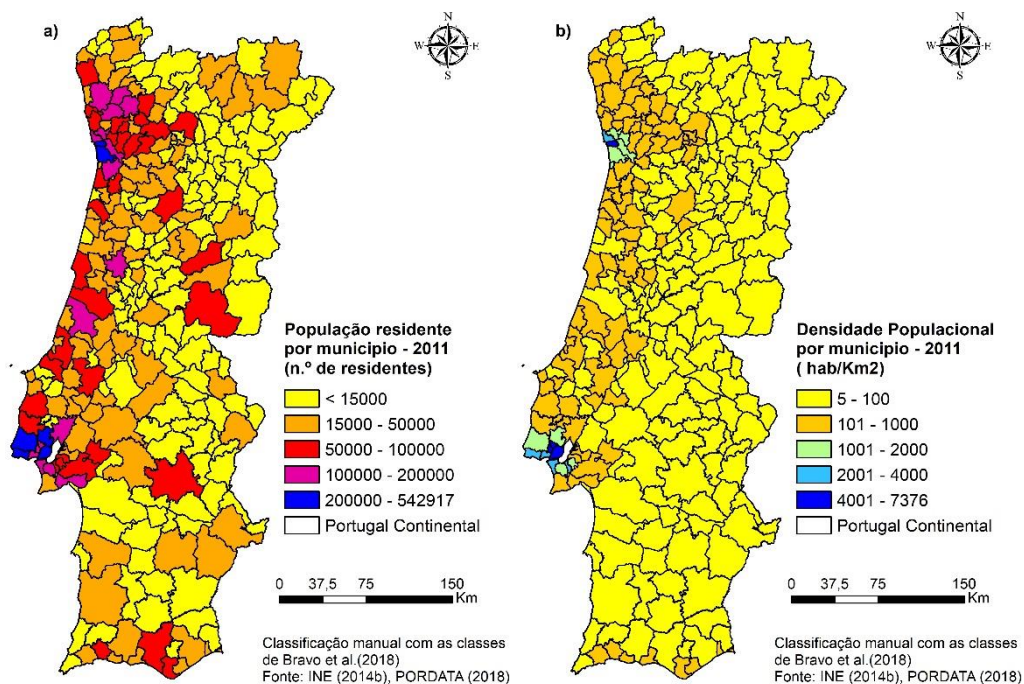


Figura 2 – População residente por município (a) e densidade populacional no ano 2011(b).

A variação da população mostra em certa medida, a deslocação da população, sendo por isso um indicador importante para estudar a localização e a distribuição das forças de segurança. No que diz respeito à variação da população entre 2011 e as

2. Descrição do Problema

projeções de 2016 (Figura 3), o município de Alcoutim teve uma perda de 17,6% da população enquanto que Arruda dos Vinhos aumentou 7,13%. Na região da grande Lisboa, à exceção do próprio município de Lisboa os saldos são positivos. Os concelhos mais interiores dos distritos de Castelo Branco e de Portalegre apresentam saldo negativo, ao contrário dos municípios do litoral.

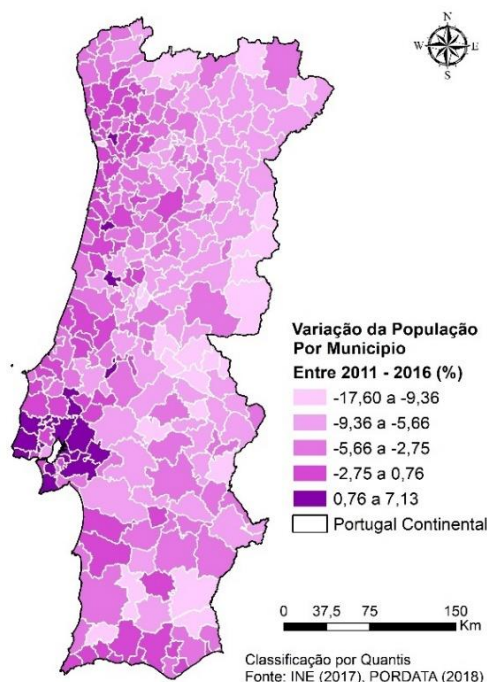


Figura 3 - Variação da População por Município entre 2011 e 2016.

As projeções da população residente entre 2011 e 2040 foram estudadas no âmbito do projeto por Bravo et al. (2018), e apresentaram indicadores importantes para a estratégia de segurança pública para Portugal. Dos diferentes cenários construídos foi possível inferir que entre 2011 e 2040 haverá um decréscimo da população em Portugal continental entre 7% e 18%, considerando o cenário 3 (otimista) e o cenário 4 (cenário tendencial natural pesada, assumindo um saldo migratório nulo), respetivamente. O grupo etário com mais de 65 anos, aumentará em 50%, sendo um valor preocupante sob o ponto de vista de aplicação de políticas públicas adequadas (Bravo et al., 2018). Quanto à distribuição da população haverá alterações caso se mantenha a tendência, existirão progressivamente mais idosos a residir em áreas rurais menos habitadas (Bravo et al., 2018).

Apesar da população mais jovem ter tendência a concentrar-se no litoral, vão existir residentes de todas as faixas etárias no interior do país, ainda que em muito menor número. O planeamento das forças de segurança tem que forçosamente

2. Descrição do Problema

acompanhar as tendências demográficas. Para além do combate à criminalidade o policiamento deverá ser de proximidade para existir um sentimento de segurança nas faixas etárias mais vulneráveis (Bravo et al., 2018).

2.1.2. A distribuição das forças de segurança em Portugal Continental

Conhecer o dispositivo atual das forças de segurança é um ponto de partida para projetar cenários estratégicos. As áreas de atuação da PSP estão concentradas nos principais centros urbanos, enquanto que a GNR atua fora dos centros urbanos (Cabral et al., 2018). O dispositivo territorial e a gestão das operações são definidas por freguesias ou bairros. A Figura 4 retrata a localização geográfica das esquadras da PSP e apenas os postos territoriais da GNR. Esta localização geográfica pode ser obtida no Mapa do Cidadão¹ e nos endereços eletrónicos da PSP e GNR.

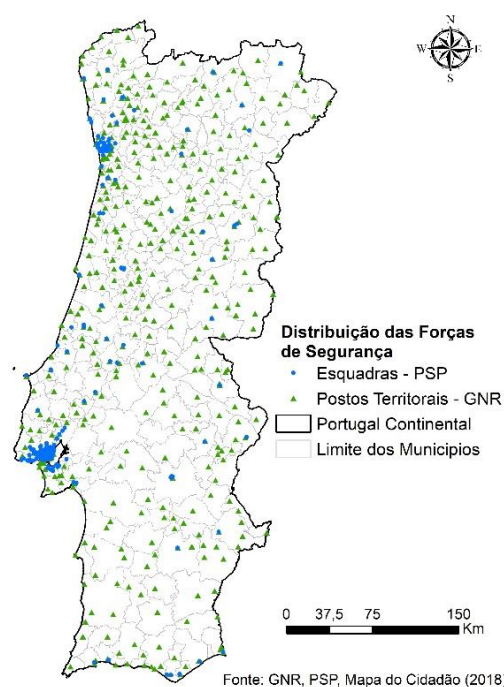


Figura 4 - Distribuição dos postos territoriais e esquadras das forças de segurança em Portugal Continental.

No que concerne às áreas de atuação de cada força de segurança, os registos disponíveis apresentam algumas lacunas, tal como foi referido por Pereira (2017). A falta de dados desagregados é a maior limitação para o presente estudo (Cabral et al., 2018). O conhecimento das áreas geográficas de atuação de cada força ao nível da

¹ <https://www.mapadocidadao.pt/>

2. Descrição do Problema

freguesia, isto é, com maior desagregação, permitiria avaliar as atuais áreas de influência, verificar quais seriam as freguesias que estão afetadas a uma determinada esquadra, que grupos etários cobre, quantos crimes anuais são cometidos, a tipologia de crimes, entre outros. Os dados sobre o crime estão apenas disponíveis por concelho o que para esta análise é insuficiente. No que concerne aos locais onde ocorreu o crime também não existe informação georreferenciada disponível publicamente. Um histórico geográfico dos locais do crime permitiria planificar com maior rigor e critério as patrulhas, o número de efetivos, de viaturas necessárias e quantificar as cargas de trabalho associadas.

Para poder planificar uma estratégia, o número de pessoas ao serviço é um fator essencial. De acordo com os dados do INE, verifica-se que na série temporal a que tivemos acesso, a PSP apresenta números mais constantes do que a GNR. Analisando a Figura 5, o ano em que o número de pessoas ao serviço mais reduzido da PSP foi em 1995 com 18204 pessoas, enquanto que o mais elevado foi no ano 2010 com 21558 pessoas. Quanto à GNR verifica-se que em 2016 apresenta 22591 pessoas e no ano 2000 teve 25904 pessoas ao serviço, sendo este o maior valor registado nas séries representadas na Figura 5.

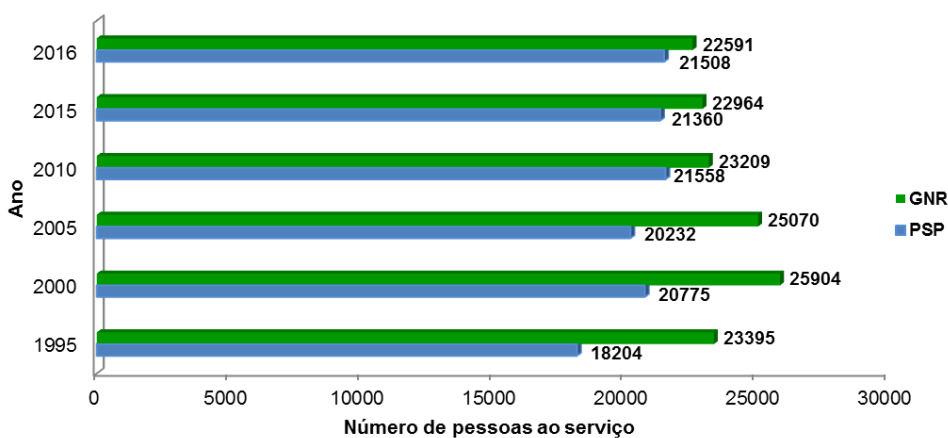


Figura 5 - Número de pessoas ao serviço da GNR e PSP (Fonte: INE,2017).

Uma das métricas acessíveis sobre os rácios polícia/cidadão, é o número de pessoas que estão ao serviço por cada 100 mil habitantes. Conhecer os números do estado atual é um ponto de partida para compreender os rácios polícia por habitante praticados. Em 1995 a PSP apresentava 181 pessoas ao serviço, enquanto que em 2016 estavam 209 (Figura 6).

A GNR tinha 251 pessoas ao serviço no ano 2000 enquanto que em 2016 apresentava o menor valor das séries com 219.

2. Descrição do Problema

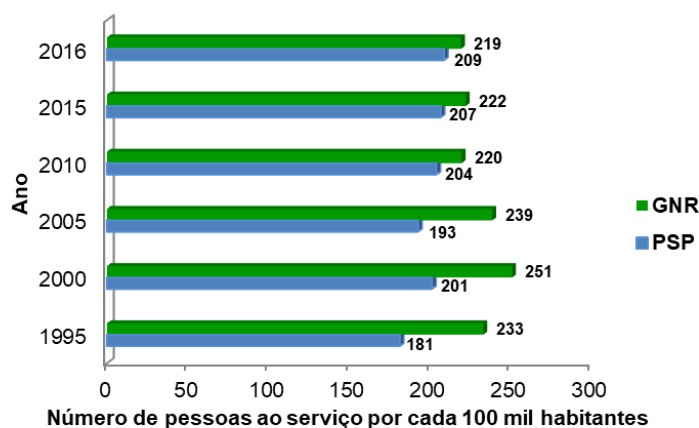


Figura 6 - Número de pessoas ao serviço por cada 100 mil habitantes (Fonte: INE, 2017).

No contexto europeu, de acordo com os dados disponibilizados pelo EUROSTAT, Portugal é sétimo país com o maior número de pessoas ao serviço por cada 100 mil habitantes (Figura 7). A Finlândia é o país com menor número de polícias por cada 100 mil habitantes enquanto que Chipre apresenta cerca de 573 efetivos (Figura 7).

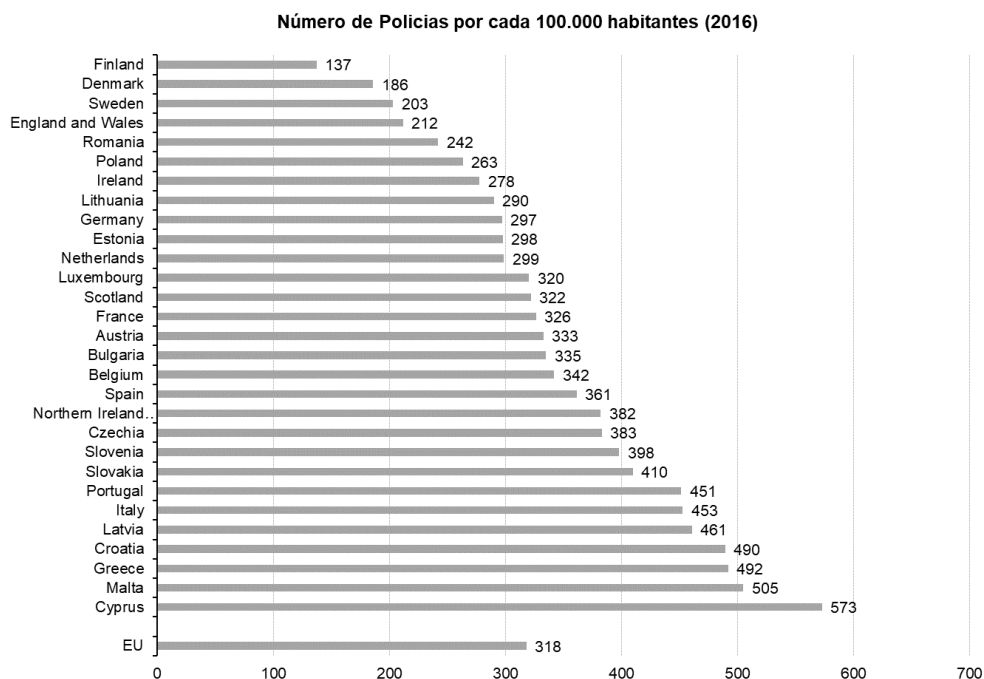


Figura 7 - Número de Polícias por cada 100.000 habitantes no ano de 2016 (Fonte: <https://ec.europa.eu/eurostat/statistics-explained/>)

Para completar esta exposição dos dados sobre a distribuição das forças de segurança não poderíamos deixar de partilhar as conclusões do estudo realizado no âmbito do projeto (SIM4SECURITY) sobre esta temática. Os modelos de análise

2. Descrição do Problema

espacial avançada desenvolvidos por Ribeiro et al. (2017), Pereira (2017) e Cabral et al. (2018), permitiram conhecer com maior detalhe a distribuição territorial das forças de segurança. Através da combinação do modelo da alteração do uso do solo, dos mapas dasimétricos e o cálculo das distâncias ao dispositivo, verificaram que 55% da população portuguesa habita áreas que distam menos de 5km do dispositivo, por rede viária, enquanto que 20% da população vive a mais de 10km (Cabral et al., 2018). Quanto à faixa etária maior que 65 anos (em 2011) cerca de 20% da população residia em locais a mais de 10km do dispositivo. Nas projeções para 2030 e 2040, concluíram que a manter-se a tendência, haverá um decréscimo para 18% e 15%, respetivamente (Cabral et al., 2018). Perante estes cenários de envelhecimento da população residente nestes locais, esta alteração poderá implicar um ajuste da distribuição do dispositivo das forças de segurança (Cabral et al., 2018).

2.1.3. Dados sobre criminalidade

Quanto aos crimes registados pelas polícias, podemos verificar que a maioria foi cometida em centros urbanos com maior número de residentes. Em 2011, o município com menor número de crimes foi Barrancos com 23, por outro lado em Lisboa contabilizaram-se 42501 crimes (Figura 8a e 8b). Em 2016 houve um decréscimo do número de crimes nos grandes centros urbanos, em relação ao número total de crimes cometidos em 2011. O município onde se registaram menos crimes foi Barrancos com 30, enquanto que Lisboa foram cometidos 36782 crimes (Figura 8b).

2. Descrição do Problema

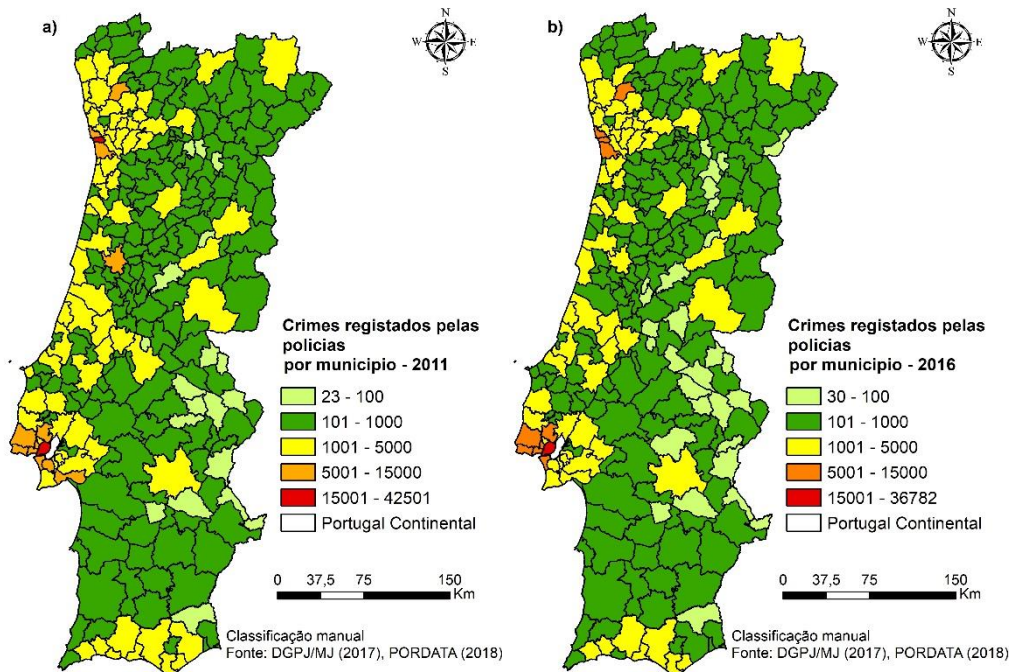


Figura 8 - Números de crimes registados pelas policias em 2011 (a) e em 2016 (b).

Nas duas séries analisadas, os tipos de crimes com maior expressão em Portugal são cometidos contra o património (Tabela 2). Em 2011, foram cometidos cerca de 220379 crimes e em 2016 baixou para 164650. Os crimes menos cometidos foram contra a identidade cultural, integridade pessoal, registando-se 4 crimes em 2011 e 33 em 2016. Os crimes contra animais de companhia passaram a ser tipificados em 2015, pelo que não existem registos em 2011, no entanto, em 2016 registaram-se 1513 (Tabela 2).

Tipo de Crime	Anos	
	2011	2016
Contra as pessoas	84783	74631
Contra o património	220379	164650
Contra a vida social	44150	40667
Contra o estado	5992	5459
Contra a identidade cultural, integridade pessoal	4	33
Contra animais companhia	0	1513
Legislação Avulsa e Outros	39644	25875
Total	396963	314844

Tabela 2 - Tipos de crimes cometidos em 2011 e 2016 (PORDATA, 2018 e INE, 2017).

Ribeiro & Henriques (2018), com a aplicação do *hot spot analysis* e os SOM nos tipos de crime por município, definiram diferentes níveis de vulnerabilidade face à

2. Descrição do Problema

criminalidade. Concluíram que as taxas de criminalidade mais elevadas se situam nas zonas urbanas de Lisboa e do Algarve. No norte do País identificaram também, zonas com elevada incidência de criminalidade. Na combinação entre a criminalidade e as variáveis socioeconómicas obtiveram resultados importantes para a definição de zonas homogéneas.

2.2. Caso de estudo

Este caso de estudo visa contribuir em certa medida para o projeto SIM4SECURITY e ao mesmo tempo para a maior das preocupações das forças de segurança que é a otimização do dispositivo territorial.

Para a persecução dos objetivos acima referidos, procura-se a delimitação adequada das áreas de ação das forças de segurança para que possibilitem tempos de resposta mais baixos, melhorar e dividir de forma equitativa a distribuição das cargas de trabalho perante os recursos disponíveis, aumentar a visibilidade dos agentes no terreno, otimizar os custos com as patrulhas. Por outro lado, existe a preocupação de apanhar os suspeitos, localizar e identificar testemunhas, a recolha de provas, o socorro à vítima e melhorar a reputação para que os cidadãos se sintam mais seguros (Henriques et al., 2018).

A satisfação de todos estes objetivos é uma tarefa que requer antes de mais dados desagregados sobre a população, sobre o crime, o número de viaturas ao serviço, atuais circuitos de patrulhas, entre outros. Contudo, a dificuldade de obter dados desta natureza, devido à importância que apresentam para a segurança nacional, não nos permite abordar o problema tendo em conta todos os objetivos referidos. Nesse sentido, foi necessário adequar o problema ao tipo de dados disponíveis.

O que nos propusemos a realizar foi simular um conjunto de áreas de serviço num distrito tendo em conta a população residente e por faixas etárias de grupos vulneráveis (exemplo: faixa etária maior do que 65 anos). Selecionou-se o distrito de Setúbal por ser um dos mais populosos e por apresentar duas situações antagónicas no que concerne à distribuição geográfica da população. Na península de Setúbal concentra-se a maioria da população residente (cerca de 82%), enquanto que no restante território consagra cerca de 18%. Pretende-se criar áreas de serviço, para cada esquadra de polícia, que sejam espacialmente contíguas tendo em conta os limites de freguesia, a população residente e grupos vulneráveis identificados na descrição do problema.

3. Revisão da Literatura

3.1. Introdução

Este capítulo tem como objetivos rever os conceitos e métodos de localização-alocação aplicados em contexto real. Como se trata de encontrar soluções para delimitar as áreas de serviço das forças de segurança, numa primeira instância efetuou-se uma visão geral dos problemas de *districting*, mencionando a sua tipologia e a miríade de aplicações no campo do planeamento e organização do território. Na seção 3.3. são descritos os conceitos de base do *police districting* e efetuou-se uma revisão dos trabalhos anteriores. Seguidamente, abordaram-se os conceitos fundamentais da otimização espacial culminando com os métodos e os modelos de localização-alocação. Por fim, efetuou-se uma revisão da terminologia e do modo de funcionamento de um algoritmo genético visto que foi implementado para a resolução deste problema.

3.2. Visão geral dos problemas de *Districting*

Os problemas de partição em zonas, conhecidos na terminologia anglo-saxónica por *Districting Problems* ou *Zone Design*, consistem na divisão de regiões ou áreas geográficas em áreas mais pequenas sendo estas designadas por zonas. Esta partição pode ser implementada através de agrupamentos de unidades territoriais contíguas mais pequenas até formar uma zona (Pereira, 2006). O objetivo é facilitar determinadas operações sobre o território e otimizar o planeamento mediante determinados critérios e restrições (Kalcsics, et al. 2005; Zhang, 2014; Bucarey et al., 2015).

Os problemas de partição em zonas ou do território são utilizados para a definição do desenho dos círculos eleitorais (Hojati, 1996; Bação et al., 2005; Goderbauer & Winandy, 2017) , distribuição de esquadras da polícia e zonas de patrulhas (D'Amico et al., 2002; Kistler, 2009; Curtin et al., 2010; Bucarey et al., 2015), distribuição de escolas (Ferland & Guénette, 1990; Lee, 2015), serviços de emergência

3. Revisão da Literatura

(Blais et al., 2003), rede elétrica (Bergey et al., 2003; Assis et al., 2014), recolha de resíduos sólidos (Hanafi et al., 1999; Lin & Kao, 2008). De acordo com Pereira (2006), o que diferencia o tipo de partição são os critérios e as restrições que permitem avaliar uma dada área. De igual modo, as áreas podem ser divididas consoante o contexto e motivação do problema a resolver.

Kalcsics (2015), dividiu em quatro grandes áreas de aplicação: *political districting*, *sales territory design*, *distribution districting* e *service districting*, onde descreve sumariamente inúmeros exemplos e critérios a ter em consideração na resolução deste tipo de problemas. Seguidamente efetua-se uma breve abordagem à definição dos círculos eleitorais por ser um problema clássico, e posteriormente são abordados os problemas de zonas de vendedores e de serviços.

1. The political districting

O desenho dos círculos eleitorais é o mais recorrente na literatura porque é particularmente essencial nas democracias onde cada círculo eleitoral elege apenas um único representante para a assembleia parlamentar. A Alemanha, os Estados Unidos da América, o Canadá e a Nova Zelândia são os exemplos mais conhecidos (Kalcsics et al., 2005; Kalcsics, 2015).

De acordo com Pereira (2006) e Goderbauer & Winandy (2017), a definição dos círculos eleitorais consiste na criação de zonas do território de um país ou de outra unidade política (como um estado ou região), num conjunto de círculos eleitorais por forma a otimizar um ou mais critérios. Este processo é realizado consoante os períodos eleitorais nos países que têm este sistema eleitoral e de acordo com a variação da população (Kalcsics et al., 2005; Bucarey et al., 2015; Kalcsics, 2015). As flutuações da população, neste caso eleitores, promovem a necessidade de ajustar as delimitações das zonas para que o total dos círculos eleitorais coincida com o total dos mandatos e total dos eleitores (Pereira, 2006).

A resolução deste tipo de problema deve ser realizada com muita clareza para evitar suspeitas de manipulação da delimitação dos círculos eleitorais com o objetivo de beneficiar algum candidato ou força política. Esta motivação parte de um fenómeno ocorrido no século XIX em *Massachusetts*, designado por *gerrymandering*. O governador *Elbridge Gerry*, propôs dividir o estado em formato de salamandra (Figura 9), para beneficiar os candidatos republicanos (Levit & Wood, 2010).

3. Revisão da Literatura



Figura 9 - Cartoon original do círculo eleitoral com a forma de salamandra (Levit & Wood, 2010).

De acordo com alguns autores, particularmente, para evitar este fenómeno são utilizados critérios/restrições como a compacidade que visa a definição de círculos eleitorais com formas geográficas compactas, a contiguidade para que todas as zonas estejam ligadas geograficamente, a igualdade populacional em cada zona para respeitar o princípio de “um homem um voto” (Bação et al., 2005; Kalcsics et al., 2005; Pereira 2006). A soma das diferenças das populações da zona devem ser semelhantes tanto quanto possível entre todos os grupos ou estar situada num intervalo pré-determinado, tal como podemos observar nos trabalhos de Bozkaya et al. (2003), e em Vanneschi et al. (2017).

Para além dos critérios referidos Bozkaya et al. (2003), propuseram ter em consideração: os limites geográficos naturais, as subdivisões político administrativas, a homogeneidade socioeconómica, a semelhança com o plano atual para que o candidato possa concorrer da mesma maneira num distrito semelhante, a consideração pela integridade das comunidades e por fim a igualdade da probabilidade de representação para a garantia de que os grupos minoritários tenham uma partilha justa de representantes.

II. Sales territory design

A definição de zonas de vendedores é outro tipo de problema de *districting* que tem sido estudado devido à importância na competitividade do ramo empresarial. As empresas procuram aumentar a eficácia das suas vendas através de estratégias de subdivisão de áreas de mercado, designadas também por regiões de responsabilidade

3. Revisão da Literatura

(Kalcsics, 2015). O princípio das zonas de vendedores não é muito diferente dos círculos eleitorais, visto que o objetivo é agrupar pequenas áreas geográficas em zonas de maior dimensão tendo em conta determinados critérios e restrições. A grande diferença reside na inclusão da carga de trabalho.

Neste tipo de problemas o número de clientes e o volume de vendas entre as regiões devem ser equilibrados para que a carga de trabalho e os tempos de viagem do vendedor estejam uniformemente distribuídos. Outro critério indissociável é a maximização do lucro ou potencial de crescimento, tendo em conta as regiões de vendas (Salazar-Aguilar et al., 2011; Ríos-Mercado & López-Pérez, 2012).

De forma sucinta de Zoltners & Sinha (1983), apontou como critérios mais utilizados para o desenho de zonas de vendas a carga de trabalho, volume de vendas, potencial de crescimento e medidas de distância.

III. Service districting

A definição de zonas de serviços pode ser utilizada em vários contextos e apresenta inúmeras aplicações independentemente se as infraestruturas são fixas ou não (por exemplo: hospitais, escolas, esquadras de polícia, serviços de emergência, recolha de resíduos sólidos).

Na maioria dos casos o desenho das zonas, neste contexto tem como finalidade definir as condições de oferta que melhor respondem às condições de procura para o objeto dos serviços. O que está em causa é a contribuição que um determinado equipamento oferece para servir o cidadão. O grande objetivo é determinar áreas contíguas que tenham uma boa acessibilidade para que a utilização do equipamento sirva o cidadão de forma eficiente (Kalcsics, 2015).

Como exemplos clássicos temos tal como já foi referido, o desenho de zonas de escolas cujos critérios são a capacidade de utilização equitativa, a distância, a boa acessibilidade e o balanço racial mais usualmente utilizado nos Estados Unidos (Ferland & Guénette, 1990 e Kalcsics, 2015). Os designados equipamentos sociais, como os hospitais e outras infraestruturas de utilidade pública são determinadas em função do número habitantes residentes, tendo em conta as limitações de capacidade máxima que o equipamento pode oferecer e a contiguidade espacial entre as mesmas (Kalcsics, 2015).

A recolha de resíduos sólidos urbanos é outros dos problemas muito referenciados na literatura pela importância que apresentam ao nível do planeamento

3. Revisão da Literatura

municipal e regional. Os critérios utilizados são o balanço no que respeita aos tempos de viagem, compacidade e a contiguidade (Hanafi et al. 1999; Lin & Kao, 2008).

No que concerne aos serviços de emergência os exemplos mais citados na literatura é o trabalho de D'Amico et al. (2002) sobre o designado problemas de *police districting*, e o caso de estudo de Baker, et al. (1989) onde o foco é a desenho de zonas para a primeira intervenção de ambulâncias (*Emergency Service Districting*). Para além dos critérios utilizados como o equilíbrio da carga de trabalho, a compacidade e a contiguidade, estes problemas destacam-se pelos tempos de resposta a chamadas de serviços, que devem ser minimizados e terem valores limite de resposta (Kalcsics et al., 2005).

Existem ainda outros tipos de aplicações como o caso da definição de zonas de redes elétricas e os clássicos problemas de transportes coletivos que para esta investigação não são aprofundados por serem serviços de outra tipologia.

3.2.1. Componentes e critérios - definições

Um problema de *districting*, pode ser formalmente descrito como a partição de todas as unidades básicas J num número de p distritos ou setores que satisfaz os critérios de equitabilidade ou balanço, compacidade e contiguidade, se requerido o centro de cada um dos distritos².

Os problemas de *districting*, são constituídos por um conjunto de unidades básicas, $J=\{1, \dots, n\}$, designadas por unidades de cobertura ou unidades geográficas representadas comumente no plano, através das primitivas geográficas. Um endereço pode ser representado por um ponto, uma avenida ou rua através de linhas e os limites de freguesias, concelhos ou locais através de polígonos (Kalsics et al., 2005).

A distância entre duas ou mais unidades básicas é dada por $d_{ij}=d(i,j)$ onde i é a unidade básica, e $j \in J$. No caso de linhas e polígonos as distâncias são calculadas através de pontos que representam os centróides dos polígonos, as linhas podem ser pontos médios ou podemos calcular o tempo de viagem para a visita de um determinado local (Kalsics, 2015).

Relativamente às relações de vizinhança num grafo, dado por $G = (V,E)$ traduzindo a vizinhança ou contiguidade onde $V_j \in V$ corresponde a $j \in J$ e $\{V_i, V_j\}$ é d_{ij} . Quanto à adjacência $N(j) \subseteq V$, denota o conjunto das unidades básicas adjacentes de V_j

² O termo distritos neste contexto, não tem o mesmo significado que distrito administrativo.

3. Revisão da Literatura

$\mathcal{C} \cup V$. Um distrito D_k , $1 \leq k \leq p$, é um subconjunto de unidades básicas, onde p é o número total de distritos (Kalsics, 2015).

O termo balanço significa a igualdade de tamanho, equilíbrio, no que respeita a um determinado critério (ex: número de residentes, número de chamadas, números de crimes, carga de trabalho). Considera-se uma solução equilibrada quando o número de residentes é servido pelo mesmo número de esquadras de polícia ou como no caso do *political districting* um homem um voto (Kalsics, 2015).

A contiguidade é outro critério fundamental, visto que permite reduzir os tempos de viagem no caso da polícia e previne o *gerrymandering* no caso dos círculos eleitorais. De acordo com Garfinkel & Nemhauser (1970), um setor ou distrito é contíguo se for possível percorrer um caminho entre duas localizações sem ter que cruzar outra que não seja vizinho. No caso dos grafos, um distrito é contíguo se e só se caracterizar um grafo conexo (Garfinkel & Nemhauser, 1970). Se cada equipamento for um ponto, por exemplo, a localização de esquadras de polícia, não é fácil e nem é claro conhecer, quais são os vizinhos, mas se forem linhas ou polígonos é mais óbvio o estabelecimento das relações de vizinhança.

Com pontos temos ao nosso dispor vários tipos de grafos de contiguidade. Geralmente, utilizam-se grafos baseados nos diagramas de *voronoi* e um dos seus derivados os grafos de Gabriel (Figura 10).

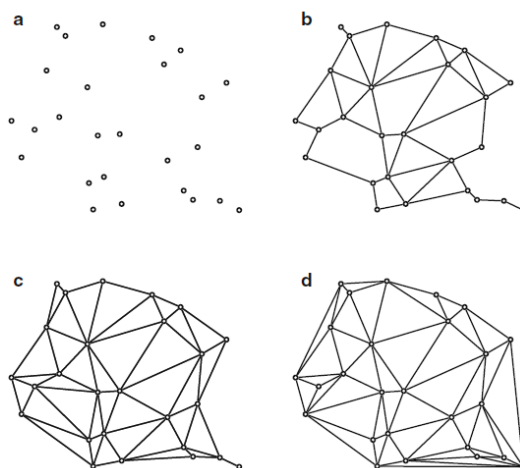


Figura 10 - Diferentes tipos de grafos apresentados por Kalsics (2015).

No que respeita às medidas geométricas, se não existir informação da vizinhança a melhor alternativa é determinar a sobreposição entre cada distrito através

3. Revisão da Literatura

do *Convex Hull*. A grande vantagem desta abordagem é que os distritos convexos geralmente impedem o cruzamento de rotas de diferentes distritos.

Não existe uma definição rigorosa para a compactidade. Wenwen et al. (2013), define compactidade de uma forma, como uma medida de quantificação para representar o quanto a mesma é compacta. Um distrito ou setor é geograficamente compacto se tiver formas circular ou retangular e sem distorção. As formas compactas implicam uma máxima acessibilidade para todas as partes, podendo partilhar atributos e propriedades comuns, referido por Tobler (Wenwen et al., 2013). Este critério tem objetivos idênticos à contiguidade. Se as unidades básicas forem polígonos, as medidas geométricas a utilizar são baseadas na relação perímetro/área.

No que concerne a medidas baseadas na distância comumente utilizada é a distância euclidiana. A Figura 11, mostra dois tipos de partições consoante os diferentes valores de compactidade.

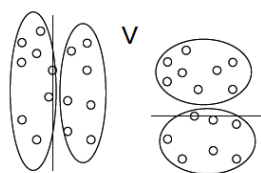


Figura 11 - Dois tipos de partições com diferentes valores de compactidade (Kalsics et al., 2005).

3.3. *Police Districting Problem (PDP)*

The police districting problem é uma designação originária dos Estados Unidos da América para delinear e caracterizar as partições dos departamentos da polícia de acordo com a sua jurisdição e estrutura hierárquica (Camacho-Collados, et al., 2015). D'Amico et al. (2002) e Camacho-Collados et al. (2015) referem que a estrutura hierárquica, genericamente é composta por distritos de comandos (*precincts*), sectores de patrulha (*beats*) e por *reporting districts* (*r-districts*). Os distritos de comando estão subdivididos em sectores de patrulha e nos *r-districts* que constituem o elemento geográfico mais pequeno onde são processadas as estatísticas do crime. Segundo Sarac et al. (1999), os *r-districts* podem coincidir com os pequenos grupos censitários (*census block groups*) e são delimitados de acordo com elementos físicos como avenidas, ruas, autoestradas, rios.

3. Revisão da Literatura

Existem inúmeros exemplos sobre a representação do PDP e por isso partilhamos o caso da Polícia Metropolitana de Washington D.C.³, onde podemos encontrar mapas interativos e relatórios (Figura 12).

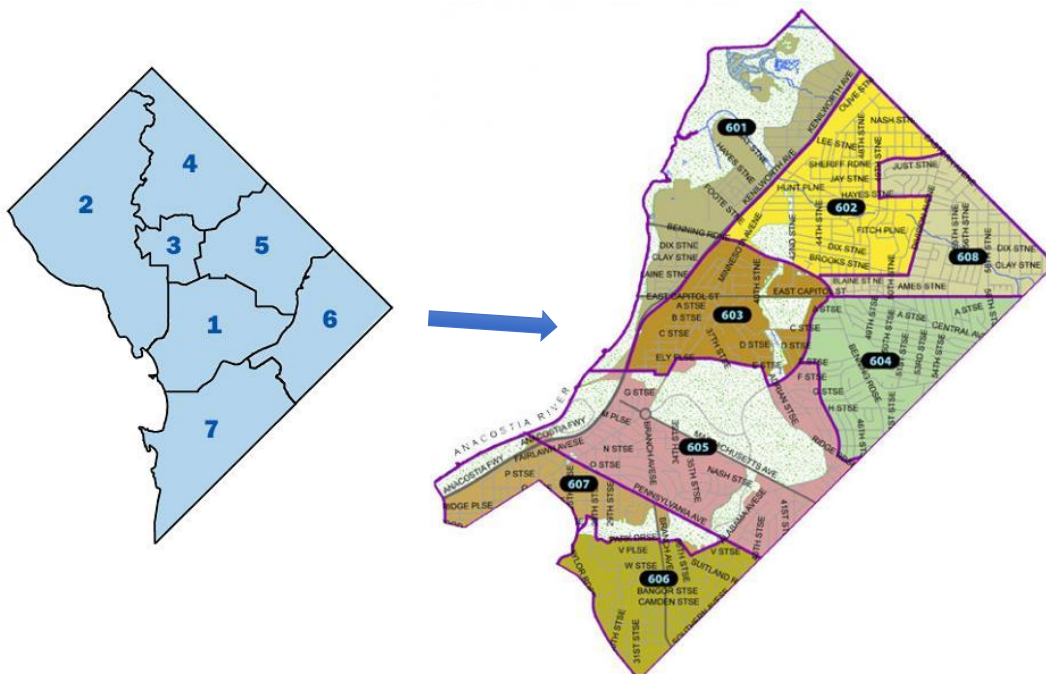


Figura 12 – Distritos de comando (direita) e os sectores de patrulha da Polícia de Washington D.C³ (esquerda).

Na página onde estão estes recursos o cidadão comum tem acesso ao número de efetivos de toda a estrutura hierárquica, dados sobre todos os tipos de crimes e a respetiva localização geográfica.

Na Europa as estruturas territoriais dos departamentos da polícia dependem do país e da região, no entanto a estrutura adotada, no geral é semelhante. Os limites de cada comando não estão disponíveis para consulta.

Algumas abordagens à resolução deste tipo de problemas têm sido propostas na literatura. A seguir faremos uma breve análise dos métodos e critérios que são utilizados.

3.3.1. Trabalhos de referência

O primeiro trabalho de referência na literatura sobre o PDP foi de Mitchell (1972), onde propõe um novo desenho para rotas de patrulhas baseada no peso da distância ponderada total esperada, definida pela soma do tempo do serviço e o tempo de viagem,

³ <https://mpdc.dc.gov/page/police-districts-and-police-service-areas>

3. Revisão da Literatura

para a sua conclusão, utilizando a heurística *min-max clustering*. Os dados utilizados foram referentes a um ano de incidentes na cidade de Anaheim (Califórnia).

Bodily (1978), utilizou um método de procura local (*local search algorithm*), incorporando a eficiência e a igualdade de serviço avaliadas de acordo com grupos de interesse. Em 1985 o primeiro autor a incluir a igualdade da carga de trabalho (*workload equalization*) no processo de otimização foi Benveniste. Para resolver o problema de otimização utilizou um modelo estocástico não linear através de um método de aproximação (cit por Camacho-Collados et al., 2015).

Sarac et al. (1999), estudou a reconfiguração dos distritos da polícia da cidade de *Buffalo*, em Nova York. Neste trabalho, discute a formulação de um problema de partição multiobjectivo e apresenta uma aplicação prática baseada na classificação das entidades geográficas censitárias. Devido à performance pouco satisfatória do algoritmo de otimização acabaram por abandonar a ideia e levaram a cabo a utilização das seções estatísticas (*Census Block Groups*). No entanto, importa reter que os critérios utilizados no problema de otimização, foram a homogeneidade tanto quanto possível da população, da área e do volume de chamadas. Cada distrito deveria ser compacto e contíguo. A nova configuração deveria ser facilmente adaptada a outras agências. Por fim, os limites dos distritos deveriam respeitar os 5 limites existentes na cidade de *Buffalo*. Os resultados obtidos foram mais satisfatórios com a utilização dos *Census Block Groups* do que com o método da partição multiobjectivo.

O trabalho de D'Amico et al. (2002) visou propor novos limites de comando na cidade de Nova York, departamento da policia de *Buffalo*, através da *meta-heurística, simulated annealing*. Utilizaram para modelar o problema, grafos das partições que envolviam a área de comando. Os grafos estavam sujeitos a restrições de contiguidade, compacidade, convexidade, tamanho e tempo de resposta iguais. Após cada iteração do algoritmo aplicaram o modelo do software *Patrol Car Allocation Model (PCAM)* para encontrar a alocação ótima dos carros de patrulha. Os resultados obtidos mostram que conseguiram melhorar a realocação dos carros de patrulha o que reduziu consideravelmente a variação da carga de trabalho. Conseguiram também que o tempo de resposta não fosse aumentado, reduzindo-o na maioria dos casos. Relativamente às possíveis melhorias do algoritmo implementado apontam a inicialização do mesmo com partições que minimizem o limite médio de cada comando geográfico. Concluíram ainda que, a introdução de sistema de informação geográfica poderia melhorar os cálculos dos tempos de viagem comparativamente com o PCAM.

3. Revisão da Literatura

Curtin et al. (2005), utilizaram um modelo para determinar uma distribuição espacial mais eficiente de patrulhas na cidade de Dallas no Texas, designado por *Police Patrol Area Covering* (PPAC). Este modelo é baseado nos designados problemas da máxima cobertura em que o objetivo é a maximização do número de eventos que estão mais perto do centro do equipamento. Os autores tinham como objetivos aumentar o nível de serviço e encontrar alocações espaciais mais eficientes. Os modelos aplicados incluem variações na prioridade nas chamadas de serviços com base no tipo de crime que foi cometido. No trabalho subsequente realizado pelos mesmos autores, nomeadamente em 2010, aplicaram o método tradicional da máxima cobertura com *backup* para fornecer soluções ótimas e para melhorar o uso dos métodos quantitativos utilizando um SIG. Os modelos foram testados com os limites geográficos dos diferentes tipos de área de intervenção da polícia de Dallas, Texas. Compararam os atuais com a sua proposta e demonstraram que o delineamento de áreas de patrulhas e de áreas administrativas podem ser melhoradas através dos métodos da máxima cobertura. Por outro lado, a integração do SIG e da programação linear oferece um método muito prático para construção de alternativas.

Zhang & Brown (2013) e Zhang (2014), apresentam um método para resolver o problema *police patrol districting*, em três passos, baseando na simulação e avaliação usando um modelo de agentes e em SIG. O caso de estudo foi realizado com os dados do departamento de polícia de *Charlottesville, VA*, Estados Unidos. No primeiro passo, aplicaram um algoritmo semelhante ao *Constraint-Based Polygonal Spatial Clustering* (CBPSB), para gerar automaticamente os distritos usando pequenas unidades geográficas atômicas como os *building blocks*. O segundo passo, foi analisar e reduzir o número de planos promissores. No último passo, avaliaram o desempenho através de um modelo baseado em agentes implementados em *Java* em ambiente SIG. Avaliaram a relação entre os parâmetros dos distritos efetuados no primeiro passo com as variáveis resposta e utilizam as regras de *pareto*. Concluíram que, as soluções geradas apresentavam uma melhoria nos tempos de resposta e na variação da carga de trabalho.

Bucarey et al. (2015), no seu capítulo dedicado ao *Police Districting*, efetuam uma breve revisão sobre este tema e propõem um modelo baseado nas *p*-medianas com restrições de igualdade de recursos entre cada esquadra da polícia chilena. Este modelo tem como objetivo a minimização da soma das distâncias entre cada bloco ao centro da instalação, as restrições de igualdade de recursos com limites superior e inferior também foram aplicadas a cada quadrante. Como medida de compacidade

3. Revisão da Literatura

utilizaram a minimização da distância aos centros e o perímetro de cada distrito. O caso de estudo foi realizado para Santiago do Chile, Ñuñua. A heurística utilizada permitiu obter apenas 5% de variação entre cada distrito ao contrário do método manual que apresentava variações de 77%.

Cammacho-Collados et al. (2015), definiram um método multicritério com o Corpo Nacional da Polícia Espanhola para o *PDP*. No primeiro modelo consideraram os atributos de área, risco, compacidade e apoio mútuo. A análise de decisão foi suportada pela preferência dos decisores no que respeita aos atributos, balanço na carga de trabalho e eficiência. A área de estudo foi no distrito central de Madrid e utilizaram os algoritmos gulosos aleatórios e um algoritmo de procura local para resolver o problema de otimização. As soluções encontradas foram comparadas com as atuais configurações para as patrulhas e avaliaram a sua qualidade. Empiricamente o algoritmo gera rapidamente soluções melhores que as existentes. No artigo de revisão subsequente desenvolvido pelos mesmos autores, Liberatori & Cammacho-Collados (2016), efetuam a comparação de métodos de procura local multicritério para o problema em grafos.

Chen et al. (2018), propuseram uma metodologia em que usam as ruas como unidades básicas. O problema foi designado por *Street-Network Police Districting* (SNPDP). O estudo foi aplicado no nordeste da cidade de Londres, designada por *Candem*. A carga de trabalho foi definida tendo em conta a combinação do peso de 3 atributos. O risco de crime, área e diâmetro do distrito. Efetuaram a comparação com o PPAC de Curtin et al. (2010). O modelo PPAC procura a máxima cobertura dos eventos, enquanto que o SNPDP, procura baixar as médias da carga de trabalho utilizando as ruas como unidades básicas. Os métodos computacionais utilizados para comparação foram, um método exato de programação linear mista com o GUROBI e as heurísticas *Random-Greedy-and-Tabu-Search* (RG-TS), *Graph-Partition-and-Tabu-Search* (GP-TS). Os resultados obtidos mostram que o GP-TS apresenta soluções muito mais rápidas e de alta qualidade comparativamente com as obtidas com o GUROBI. O GP-TS também apresentou uma maior performance e rapidez comparativamente com o RG-TS.

3.4. Otimização

A otimização também designada por programação matemática, é a arte ou ciência de alocar recursos escassos para o melhor efeito possível (Chinneck, 2000). A otimização permite tornar as decisões o mais funcionais e efetivas possíveis, pelo que a sua utilização se centra na procura de soluções que proporcionem o máximo desempenho, minimização de custos, maximização do lucro, na gestão de recursos materiais, de pessoas, de dinheiro, do território, entre outros (Chinneck, 2000).

Segundo Chinneck (2000) e Talbi (2009), a otimização é aplicada amplamente no ramo dos negócios, indústria, governos centrais, na engenharia, na ciência da computação e na geografia para resolver problemas reais e complexos.

O conceito básico de otimização é encontrar a melhor solução ou alternativa possível para um dado problema. Na sua génese, o problema de otimização é uma classe de problemas que procuram maximizar ou minimizar uma função matemática com um determinado número de variáveis sujeitas a determinadas restrições (Sarker & Newton, 2008). A função matemática a otimizar é designada por função objetivo que geralmente contém várias variáveis. Se o problema em causa tiver mais do que uma função objetivo é designado por multiobjectivo (Sarker & Newton, 2008).

Um problema de otimização é constituído por três elementos principais: a) função objetivo; b) restrições; c) variáveis de decisão e parâmetros:

- a) A função objetivo designada muitas vezes por função de custo ou de energia, é uma expressão matemática que combina as variáveis para determinar o valor alvo (*target*) ou a qualidade da solução em função das variáveis e dos parâmetros;
- b) As restrições são expressões de igualdade ou desigualdade (inequações) que combinam os limites das variáveis no conjunto de soluções possíveis;
- c) As variáveis de decisão podem ser designadas apenas por variáveis e são valores desconhecidos que são determinados através do modelo. Estas variáveis podem ser contínuas, discretas e ou binárias.

Na prática, imaginemos que pretendemos implementar num determinado concelho com 10 freguesias, duas esquadras de polícia. As restrições podem ser representadas por uma área máxima ou mínima de atuação. As variáveis de decisão correspondem às freguesias onde serão instaladas as esquadras e a respetiva área de serviço.

3. Revisão da Literatura

A função objetivo e as restrições podem ser descritas pelas seguintes expressões segundo Reeves (1997) :

Minimização $g(x)$ Equação I

Sujeita a:

$f_i(x) \leq b_j \forall i$ Equação II

condições x

Onde x são as variáveis de decisão desconhecidas, $x = [x_1, x_2, x_3, \dots, x_n]$, $g(x)$ e $f_i()$ são funções de x e b_j é o coeficiente que limita o valor de cada função $f_i()$. A condição x consiste na combinação de valores reais, inteiros ou binários, não negativos, isto é, as restrições (Tong & Murray, 2012).

Uma função objetivo pode ser expressa através de relações lineares entre as variáveis do modelo e não lineares desde que envolvam pelo menos uma relação não linear. As relações não lineares aumentam significativamente a complexidade do problema. As restrições limitam os valores das variáveis de decisão e o espaço viável das soluções. As relações não lineares nas restrições aumentam a complexidade do problema. As variáveis de decisão são os valores desconhecidos cujo objetivo é encontrar essas variáveis que satisfaçam as restrições e otimizem o valor da função objetivo.

A função objetivo, as variáveis de decisão e as restrições são combinadas para resolver o problema de otimização através de métodos exatos ou por heurísticas (algoritmos de aproximação).

3.4.1. O processo de otimização

Um processo de otimização para ser bem-sucedido deverão os intervenientes considerar todas as fases típicas de um projeto desta natureza. Muitas vezes, pode ser necessário modificar o modelo para que as soluções procuradas sejam alcançadas. A Figura 13, permite-nos ter uma visão geral de um ciclo de otimização.

3. Revisão da Literatura

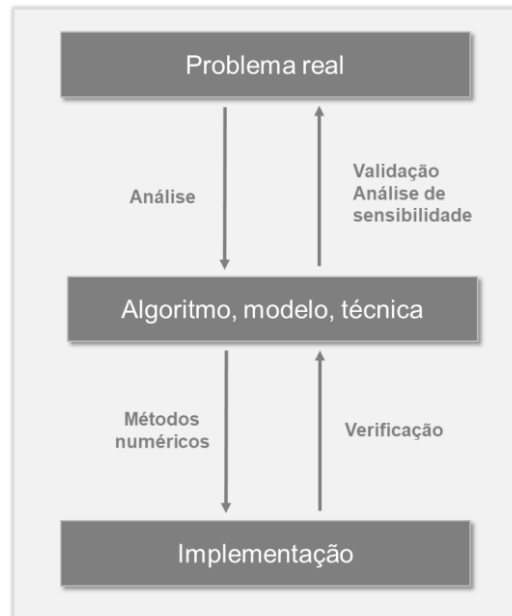


Figura 13 - Processo clássico de otimização (Adaptado de Chinneck, 2000).

Numa primeira instância, o processo surge para resolver um problema real, onde se efetua uma análise prévia recolhendo toda a informação relevante e necessária para formular o problema. Após este trabalho de abstração passamos à fase de escolher um algoritmo ou desenvolver um modelo ou técnica para implementar num programa comercial ou então construir um. Depois da sua implementação efetua-se a verificação do algoritmo ou do modelo utilizado efetuando testes à implementação. Por fim, para a garantia do seu funcionamento e aderência à realidade efetuam-se testes de validação e a análise de sensibilidade (Chinneck, 2000).

3.4.2. Modelos de otimização espacial

Os problemas de otimização espacial à semelhança dos não espaciais utilizados na matemática e nas engenharias têm como elementos principais o objetivo a ser otimizado, as restrições e as variáveis de decisão (Tong & Murray, 2012).

Anselin et al. (2013), referem que a otimização espacial é aplicada num contexto geográfico em que os objetivos e as restrições são definidas pelo espaço e pelas relações espaciais. As principais diferenças da otimização espacial da não espacial são as relações topológicas e a interdependência espacial como por exemplo a distância, adjacência, conectividade, intersecção, forma, padrões e distritos (Tong & Murray, 2012).

Os modelos de otimização espacial têm sido aplicados em problemas de localização, planeamento do uso do solo e desenho de redes e em *(re)districting*

3. Revisão da Literatura

(Anselin et al., 2013). A quantidade de aplicações e de modelos são variados. Na literatura encontram-se muitos trabalhos de investigação nesta área, como Hosage & Goodchild (1986) em que utilizam os algoritmos genéticos para resolver um problema de localização-alocação, Openshaw & Rao (1995) para problemas de regionalização, entre muitos outros.

A primeira abordagem matemática para o problema de *(re)districting* foi proposta por Hess et al. (1965), formularam um problema de localização de equipamentos capacitado designado por *p*-medianas⁴. É um problema de atribuição com restrições adicionais, em que é atribuída a cada unidade básica um distrito. A função objetivo (Equação III), minimiza a distância euclidiana entre cada unidade, no entanto, não existe a garantia de contiguidade. O custo da atribuição c_{ij} ao centro do distrito i é a distância euclidiana entre cada centro do distrito j . A distância d_{ij} é definida entre cada centro das unidades i e j .

$$\sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij} \quad \text{Equação III}$$

$$\sum_{j \in J} x_{ij} = 1 (i \in I) \quad \text{Equação IV}$$

$$\sum_{j \in J} x_{ij} = m \quad \text{Equação V}$$

$$a \leq \sum_{i \in j} p_i x_{ij} \leq b (j \in J) \quad \text{Equação VI}$$

$$x_{ij} \leq x_{jj} (i \in I, j \in J) \quad \text{Equação VII}$$

$$x_{ij} = 0 \text{ ou } 1 (i \in I, j \in J) \quad \text{Equação VIII}$$

A restrição 1 (Equação IV), garante que cada unidade da população i é atribuída exatamente a um distrito. A restrição 2 (Equação V), garante que os distritos M sejam escolhidos. A restrição 3 (Equação VI), garante o equilíbrio da população entre os distritos. Este modelo não garante a convergência, todas as soluções não contiguas têm de ser rejeitadas. Permite que o desvio da entre a população mínima e máxima não tenha controlo.

⁴ *p*-medianas: modelo de localização-alocação, que permite definir a localização de equipamentos, perante um conjunto de clientes que devem ser servidos de forma a otimizar um determinado critério (Alp et al., 2003).

3. Revisão da Literatura

Garfinkel & Nemhauser (1970), apresentaram um algoritmo de procura em árvore em que numa primeira fase começa numa unidade arbitrária que se junta a unidades contíguas até que a população que é combinada se torna viável. Se o distrito é compacto então tem viabilidade, caso não tenha o algoritmo retrocede e volta à fase inicial.

Hosage & Goodchild (1986), apresentam uma das primeiras investigações na área da otimização espacial utilizando um AG, para resolver problemas de localização-alocação. Implementaram um AG para resolver o problema das p -medianas em que os resultados não foram os esperados, visto que as soluções “caíram” maioritariamente em ótimos locais. Contudo, demonstraram que os AG's apresentam a vantagem de lidar com tarefas muito específicas e que são ajustáveis a problemas de otimização de difícil resolução.

Bozkaya et al. (2003), utilizaram os critérios descritos na seção 3.2. Propuseram um modelo multicritério baseado na atribuição de pesos nos critérios e efetuaram uma análise de sensibilidade. Desenvolveram um sistema interativo para visualizar as soluções com o intuito de comparação.

Bação et al. (2005), propuseram um algoritmo genético para resolver um problema de *districting* eleitoral. Os critérios para o desenho dos distritos são o equilíbrio da população, contiguidade espacial e a compacidade geográfica. A avaliação da contiguidade espacial foi realizada através do método da soma das diferenças da população. A medida de compacidade espacial utilizada foi a soma da divisão do perímetro ao quadrado de cada zona pela área.

Castelli et al. (2015), aplicaram a programação genética com operadores semânticos a um problema de *districting* eleitoral. Os critérios para o problema são muito semelhantes aos aplicados por Bação et al. (2005). Todos os distritos devem ser iguais em termos de população, cada distrito deve ser um território único e contíguo e devem ser compactos. A igualdade populacional também foi avaliada com a soma das diferenças absolutas das populações dos distritos com a média da população do distrito.

Existem ainda outros trabalhos que merecem destaque como o de Ricca & Simeone (2008), em que utilizaram o *Tabu Search* (TS) para um problema de *political districting*.

3.5. Meta-heurísticas

Na secção seguinte apresentamos uma abordagem aos conceitos de meta-heurísticas dando especial realce aos algoritmos genéticos. As meta-heurísticas são utilizadas geralmente quando não se conhece a priori, um algoritmo eficiente para resolver um determinado problema. As meta-heurísticas são métodos abstratos e não fornecem etapas específicas para encontrar uma solução ótima. A sua natureza estocástica permite adicionar aleatoriedade às soluções. Ao contrário dos métodos exatos que garantem a solução ótima, as meta-heurísticas não garantem a mesma. Os métodos exatos conhecidos demoram muito tempo para encontrar uma solução ótima ao contrário das meta-heurísticas que o tempo de procura é reduzido podendo ser utilizadas em problemas de grandes dimensões (Malczewski & Rinner, 2015).

Os algoritmos evolucionários são exemplos de meta-heurísticas baseadas em processos naturais de evolução. As meta-heurísticas mais utilizadas em problemas de districting são o *simulated annealing* (SA), *Tabu Search* (TS), GRASP e os Algoritmos genéticos (AG).

3.5.1. Algoritmos Genéticos (AG)

Os algoritmos genéticos (AG) foram introduzidos por John Holland e pelo seu grupo entre os anos 60 e 70, ideia que se materializou em 1975 com o livro “*Adaptation in Natural and Artificial Systems*” (Goldberg, 1989; Mitchell, 1998). Em 1989, Goldberg aluno de *Holland*, edita um dos livros mais completos e importantes sobre os algoritmos genéticos designado por *Genetic algorithms for search, optimization, and machine learning*. Após este período, a utilização dos AG tem-se vulgarizado na resolução de inúmeros problemas complexos.

Os AG são um dos ramos da computação evolutiva (CE) para a resolução de problemas baseados na genética e na teoria da evolução com grande desenvolvimento na área da inteligência artificial (IA). São métodos de procura adaptativos utilizados na pesquisa de soluções, na resolução de problemas de otimização e de *machine learning*⁵, inspirados na teoria da seleção natural e da evolução das espécies de *Charles Darwin* (Goldberg, 1989; Beasley et al., 1993; Costa & Simões, 2008).

De acordo com Goldberg (1989), os AG são diferentes dos métodos tradicionais de procura e otimização, porque trabalham com a codificação das potenciais soluções

⁵ O significado de *Machine learning* é aprendizagem máquina.

3. Revisão da Literatura

e não com as próprias soluções, efetuam a procura num conjunto de pontos permitindo realizar buscas simultâneas em vários conjuntos, usam a informação de troca da função objetivo, e usam regras de transição probabilísticas em vez de regras determinísticas.

3.5.2. Terminologia e conceitos básicos

A compreensão deste algoritmo carece da descrição dos processos biológicos implícitos, pois a inspiração do AG foi baseada nos mesmos. Os AG utilizam a analogia entre os processos naturais da evolução e a resolução dos problemas, no entanto, os processos da biologia são mais complexos. Genericamente, de acordo com Mitchell (1998) e Costa & Simões (2008), todos os organismos vivos são constituídos por células que contém o mesmo conjunto de um ou mais cromossomas. O cromossoma é uma longa sequência do Ácido Desoxirribonucleico (ADN), constituído por vários genes que codificam as características dos indivíduos. A sua posição na cadeia de ADN, é designada por locus, e os valores de diferentes de cada gene são designado por alelos. Todo o conjunto do material genético é designado por genoma e o conjunto de genes que constitui o mesmo designa-se por genótipo. O conjunto das características visíveis, codificado pelo genótipo designa-se por fenótipo. Feita esta revisão dos conceitos da biologia agora passamos à reprodução e mutação.

Segundo Mitchell (1998) e Costa & Simões (2008), os indivíduos cujos cromossomas estão dispostos aos pares são chamados de diplóides e utilizam a reprodução sexuada, ao contrário dos haplóides que a reprodução é assexuada. Durante a reprodução sexuada ocorre a recombinação, ou crossover, em que cada gene de cada um dos pais são trocados. Cada par de cromossomas, neste caso gametas, quando se fundem cria um par de cromossomas diplóides. Nos organismos haplóides a reprodução dá-se através da fissão binária de uma célula-mãe que dá origem a 2 células filhas que contém material genético da progenitora.

A descendência em ambos os casos está sujeita a mutação que provocam alterações do material genético. A aptidão de um organismo é tipicamente definida como a probabilidade de o organismo viver e de se reproduzir.

No que concerne ao AG, o termo cromossoma refere-se às soluções candidatas ao problema, que é codificado por exemplo, com uma cadeia de *bits*. Os genes são os *bits*, ou a cadeia de *bits* adjacentes que codificam um parâmetro. Um alelo na cadeia de *bits* tem o valor 0 ou 1, pelo que é possível ter mais alelos em cada locus. O crossover consiste em trocar material genético entre os 2 pais, selecionados de acordo com a função de avaliação (*fitness*). Os novos indivíduos podem ser sujeitos a mutação através

3. Revisão da Literatura

da troca de valores de um ou mais genes (Mitchell, 1998). A Tabela 3, faz uma sinopse da comparação entre os termos biológicos e os computacionais.

Termos biológicos	Termos computacionais
Cromossoma, individuo	Vetor
Gene	Caractere
Alelo	Valor do caractere
Lócus	Posição do caractere
Genótipo	Representação do individuo
Fenótipo	Interpretação do vetor de caracteres

Tabela 3 - Comparação entre a terminologia biológica e computacional (Fonte: Sivanandam & Deepa, 2008).

Segundo Mitchell (1998), a maioria das aplicações dos AG são em individuos haplóides, com um único cromossoma. O genótipo de um individuo de um AG consiste na sequência de genes do seu cromossoma.

3.5.3. Funcionamento básico de um AG

Os AG permitem-nos resolver problemas complexos e muitas vezes intratáveis por métodos clássicos. A chave para que o AG resolva um determinado problema específico é a codificação correta do problema a resolver. Por outro lado, a definição do tipo de representação dos individuos da população, o método de seleção, os operadores genéticos a aplicar, a função de avaliação e as características da população são aspetos primordiais para que o AG tenha sucesso.

Genericamente, a etapa inicial é a geração da população⁶ inicial que normalmente é feita de modo aleatório ou utilizando heurísticas. Posteriormente os melhores individuos da população são selecionados perante a sua qualidade, através da função de avaliação. Os melhores individuos, neste caso os progenitores, iram-se reproduzir e trocar o seu material genético para dar origem à descendência (*crossover*). Seguidamente, os novos individuos podem ser alterados pela mutação. As soluções obtidas a partir do *crossover* e da mutação constituem a geração seguinte. Todos estes passos repetem-se sucessivamente até encontrar a solução ótima ou próxima da ótima (Tutorialspoint, 2016). O funcionamento de um AG básico pode ser observado na Figura 14.

⁶ População é diferente de população residente. Num AG a população é um conjunto de cromossomas que irão ser testados.

3. Revisão da Literatura

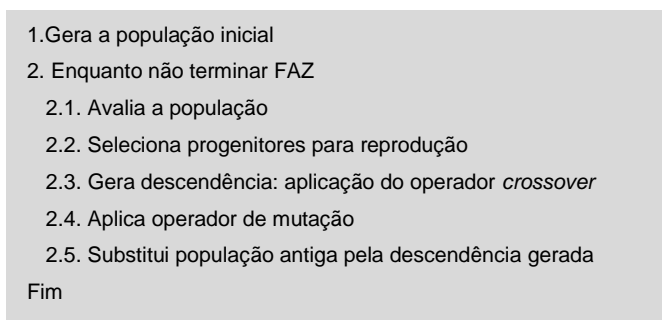


Figura 14 - Algoritmo genético clássico (Simões, 1999).

Nas subseções seguintes iremos analisar cada um dos elementos básicos do AG, destacando alguns tipos de codificação.

3.5.3.1. Codificação

A codificação é o processo de representação dos genes dos indivíduos. Podem ser representados por uma cadeia de *bits*, números, árvores, listas, matrizes ou qualquer outro tipo de objeto (Sivanandam & Deepa, 2008). A representação de um cromossoma é designado como o genótipo do indivíduo (Beasley et al., 1993).

A codificação dos cromossomas é dos passos mais importantes para que o AG funcione e não comprometa os operadores genéticos. A codificação depende muito do problema que pretendemos resolver. Apresentamos aqui alguns dos tipos de codificação que consideramos importantes, a binária, números reais e por permutação. A codificação binária dos cromossomas foi a primeira a ser apresentada por *Holland* em 1962 sendo esta uma das mais adotadas. Neste caso, cada cromossoma é constituído por uma sequência de *bits* e geralmente têm tamanho fixo (Tutorialspoint, 2016). Este tipo de representação permite-nos representar um levado número de cromossomas utilizando apenas 2 alelos, o 0 e o 1. A figura 15 representa um exemplo de uma codificação binária.

Cromossoma 1	1	0	0	0	1	1	0	0	1	1
Cromossoma 2	0	1	1	0	0	1	1	0	0	1

Figura 15 - Exemplo de codificação binária de 2 cromossomas com 10 genes (Fonte: Elaboração própria).

A codificação com números reais, em que queremos definir os genes com variáveis contínuas num problema, em vez de discretas, podemos usar números. A

3. Revisão da Literatura

precisão desses números dos valores reais ou ponto flutuante é, no entanto, limitada à capacidade de processamento do computador (Figura 16).

Cromossoma 1	0.234	4.435	3.769
Cromossoma 2	0.126	2.760	4.567

Figura 16 - Exemplo de codificação real de um cromossoma com 3 genes (Fonte: Elaboração própria).

Neste caso os operadores de cruzamento e de mutação serão distintos aos da codificação binária.

A codificação por permutação, utilizada em muitos problemas de otimização combinatória, cada cromossoma é uma cadeia de números inteiros ou reais que representa um número numa sequência (Figura 17).

Cromossoma 1	1	5	3	2	6	4	7	9	8
Cromossoma 2	8	5	6	7	2	3	1	4	9

Figura 17 - Exemplo de codificação por permutação (Sivanandam & Deepa, 2008).

Este tipo de codificação é útil para problemas de ordenação (Sivanandam & Deepa, 2008), como por exemplo o problema do caixeiro viajante.

Existem outros tipos de representação como por exemplo, em árvore, que é muito utilizada na programação genética.

3.5.3.2. População

A população é o conjunto de indivíduos que serão testados os seus parâmetros fenotípicos. Existem dois aspetos fundamentais sobre a população, a geração inicial da população e o tamanho (Sivanandam & Deepa, 2008). De acordo com Goldberg (1989), estes aspetos, devem ser estabelecidos empiricamente e de acordo com a disponibilidade dos recursos computacionais. Nas recomendações de Sivanandam & Deepa (2008), nos casos em que que a população é inicializada aleatoriamente, a primeira população deve ser a maior possível para podermos explorar todo o espaço de pesquisa. No caso da utilização de uma heurística para inicializar uma população, a aptidão média da mesma é maior, pelo que pode ajudar o AG a encontrar melhores soluções. Se a população não tem diversidade o algoritmo explora apenas uma pequena

3. Revisão da Literatura

parte do espaço de procura e pode nunca encontrar soluções ótimas globais (Sivanandam & Deepa, 2008).

3.5.3.3. Função de Avaliação (*Fitness*)

A avaliação de um individuo é o valor da função objetivo para o seu fenótipo (Sivanandam & Deepa, 2008). O cálculo da função de avaliação indica-nos a qualidade da solução, no entanto, também nos indica qual poderá ser o cromossoma ideal (Tutorialspoint, 2016).

A escolha da função de avaliação depende do tipo de problema a resolver. A correta definição desta função juntamente com a codificação é o aspeto crucial de um AG (Beasley et al., 1993).

3.5.3.4. Métodos de seleção

A seleção é o processo de escolher os progenitores para o cruzamento. O propósito da seleção é encontrar os melhores cromossomas individuais da população para reprodução que irão proporcionar filhos (*offspring*), com maior mérito. Se a seleção favorecer os melhores cromossomas irá resultar uma convergência elevada promovendo a estagnação num máximo local, no entanto, este facto também pode acontecer de forma prematura, pelo que se deve tentar preservar sempre a diversidade da população. O método de seleção deverá ser conjugado com outros operadores genéticos, para que o sucesso do AG não seja comprometido (Costa & Simões, 2008).

Existem vários métodos de seleção propostos na literatura, pelo que aqui vamos apresentar os mais tradicionais. Uma das técnicas tradicionais mais utilizadas é o método da roleta (*Roulette Wheel*), também conhecido pela seleção proporcional ao mérito. Consiste na utilização dos valores atribuídos a cada individuo pela função de avaliação e na qualidade média da população. Perante a qualidade de cada individuo, atribui-se uma porção de um círculo (roleta) a cada um deles. A roleta é girada o número de vezes tantas quantas o número de indivíduos da população, parando numa porção correspondente a um dos indivíduos (Tutorialspoint, 2016). Após este procedimento ser efetuado n vezes estarão selecionados os progenitores que irão passar os seus genes para a próxima geração.

A seleção por ranking (*Rank selection*) de acordo com Mitchell (1998), procura evitar a rápida convergência do AG com o máximo local. Os indivíduos da população são ordenados de acordo com o seu mérito e a possibilidade de escolha é atribuída de acordo com a posição que ocupa.

3. Revisão da Literatura

A seleção por torneio (*Tournament selection*) consiste em escolher sucessivamente grupos de soluções com maior mérito e depois dentro de cada um desses grupos são selecionados os melhores. De acordo com Sivanandam & Deepa (2008), é um método bastante eficiente e que leva à solução ideal.

Existem ainda mais alguns métodos de seleção como a seleção aleatória, a amostragem estocástica universal, a seleção por truncatura e a seleção de *Boltzmann* (Sivanandam & Deepa, 2008).

3.5.3.5. Operadores genéticos

Outro dos aspetos essenciais na implementação de um AG são os operadores genéticos a utilizar. Aqui tudo depende da estratégia que foi utilizada na codificação (Mitchell, 1998). Apresentamos aqui algumas operações de cruzamento (*crossover*) e de mutação.

1. *Crossover*

O cruzamento (*crossover*), é um processo pelo qual os progenitores selecionados para a reprodução trocam o seu material genético e vão gerar os filhos. As 3 formas de cruzamento mãos comuns são o *crossover* de 1 ponto de corte, com dois ou mais pontos de corte e o uniforme. O cruzamento de um ponto consiste em dividir os cromossomas selecionados num ponto aleatório. Posteriormente, troca-se o material genético dos 2 progenitores tendo como referência o ponto de corte selecionado (Figura 18).

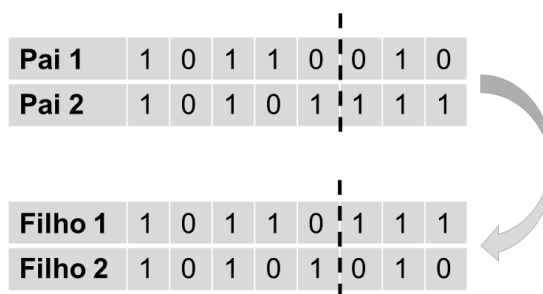


Figura 18 - Exemplo de cruzamento de um ponto de corte (Fonte: Elaboração própria).

O cruzamento de 2 pontos de corte funcionam de forma análoga ao de um ponto de corte (Figura 19).

3. Revisão da Literatura

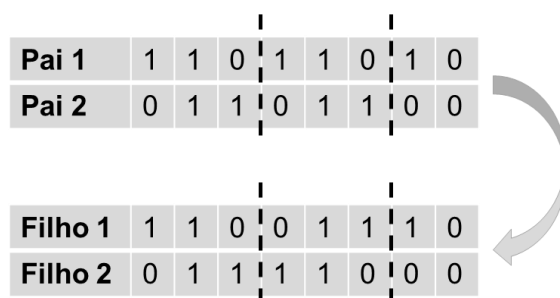


Figura 19- Exemplo de cruzamento com dois pontos de corte (Fonte: Elaboração própria) .

O *crossover* uniforme, sugerido por Syswerda em 1989, consiste na utilização de uma máscara, coincidente com o tamanho do cromossoma, gerada aleatoriamente, que fornece a posição dos genes que deverão ser trocados. Por exemplo, um *bit* 0 na máscara indica que o gene será do progenitor 1, um *bit* 1 indica que o gene pertence ao outro progenitor. Para o segundo filho o processo é inverso (Figura 20).



Figura 20 - Exemplo de cruzamento uniforme (Fonte: Sivanandam & Deepa, 2008).

Existem outros tipos de *crossover* como por exemplo, o cruzamento parcialmente semelhante, o cruzamento ordenado muito utilizados na codificação por permutação, que podem ser consultados nas obras de Mitchell (1998) e de Sivanandam & Deepa (2008).

Outro parâmetro básico no *crossover* é a probabilidade de cruzamento em que descreve a frequência que se vai realizar o mesmo. Se a probabilidade é 0% então a geração seguinte será a cópia dos progenitores, se for 100% todos os descendentes provêm do cruzamento. Este operador é aplicado para que os novos cromossomas contenham as melhores partes dos progenitores.

II. Mutação

O operador genético mutação, possui um papel fundamental no AG, porque garante a diversidade, introduz aleatoriedade, evita a convergência prematura e previne que o algoritmo caia nos mínimos locais. Este operador consiste na alteração do valor

3. Revisão da Literatura

de 1 ou mais genes de um cromossoma. Os tipos de mutação mais utilizados são a troca de 1 bit, troca de posição de 2 bits e a soma ou subtração de um pequeno número em codificações reais (Figuras 21, 22 e 23).

Cromossoma 1	1	0	1	1	0	0	1	0
Mutação	1	1	1	1	0	0	1	0

Figura 21 - Mutação por troca de 1 bit (Fonte: Elaboração própria).

Cromossoma 2	1	3	2	4	5	8	9	7
Mutação	1	3	2	9	5	8	4	7

Figura 22 - Mutação por troca de posição (Fonte: Elaboração própria).

Cromossoma 3	0.234	4.435	3.769
Mutação	0.234	4.516	3.769

Figura 23 - Mutação por adição de um valor (Fonte: Elaboração própria).

A probabilidade de mutação determina quantas vezes uma determinada parte do cromossoma é mutada. Geralmente a probabilidade é bastante baixa, por exemplo 0,001.

3.5.3.6. Estratégias evolucionárias

Após aplicar os processos de cruzamento e mutação irão resultar cromossomas que farão parte da nova geração. A nova geração torna-se a nova população de origem (McCall, 2005). Nesta fase podemos optar por uma estratégia evolucionária para efetuar a escolha das soluções que irão formar as próximas gerações. Estas opções permitem aumentar a diversidade e ainda aproveitar partes da população que de outra forma se perderiam. Para este efeito, existem algumas estratégias de reposição.

A estratégia geracional em que substitui todos os cromossomas de uma população pelos seus descendentes (Tutorialspoint, 2016). Neste caso, todos os cromossomas são gerados através do processo de seleção e de recombinação, onde existe a probabilidade de se perder diversidade na população (McCall, 2005).

A reposição por elitismo, introduzida por *Kenneth de Jong* (em 1975), em que os melhores cromossomas individuais são preservados e todos os outros são substituídos (Mitchell, 1998 e McCall, 2005).

3. Revisão da Literatura

A reposição em estado estacionário (*steady-state*) em que na nova população é gerado um novo cromossoma que irá substituir o pior da população de origem (McCall, 2005 e Tutorialspoint, 2016).

3.5.3.7. Critérios de paragem

Existem vários critérios de paragem que deveremos aplicar (Sivanandam & Deepa, 2008). O AG para quando um determinado número de gerações foi atingido, por acabar o tempo pré-determinado ou se não existirem alterações na *fitness*.

3.5.4. Vantagens e limitações

Haupt & Haupt (2004) e Sivanandam & Deepa (2008), apontam algumas das vantagens e limitações dos AG. As vantagens são:

- espaço de solução é mais amplo;
- otimiza um número elevado de variáveis;
- lidam bem com funções objetivo complexas;
- resistência aos ótimos locais e encontram facilmente os ótimos globais;
- facilidade de modificação para diferentes problemas;
- facilidade de combinação com outras heurísticas;
- facilidade de implementação;
- entre outros.

No que respeita às limitações os AG são normalmente lentos, pois quando ainda estão a avaliar a população inicial os métodos clássicos já encontraram a solução. Ficam aqui mais algumas das limitações apontadas por Haupt & Haupt (2004) e Sivanandam & Deepa (2008):

- definição da representação para o problema;
- o problema de escolher os vários parâmetros como o tamanho da população, taxa de mutação, taxa de cruzamento, o método de seleção;
- impossibilidade de usar gradientes;
- não é bom na identificação de ótimos locais;
- necessidade de ser acoplado a uma técnica de busca local;
- dificuldade em encontrar o ótimo global exato;
- exigir um grande número de avaliações da função *fitness*.

3.5.5. Principais áreas de aplicação

Como já foi referido no início da secção 3, os AG são utilizados para resolver problemas difíceis, como o NP-difíceis, com múltiplos máximos e mínimos, para os quais não existe nenhum algoritmo de otimização. Segundo Beasley et al. (1993) e Sivanandam & Deepa (2008), algumas das aplicações são as seguintes:

- planeamento de trajetórias de robôs;
- planeamento estratégico;
- otimização combinatória como os problemas do caixeiro viajante, máxima cobertura, partição de grafos, p-medianas, *(re)districting problems*;
- controlo de gasodutos, evasão de mísseis, perseguição;
- design de aviões, computadores, teclados e redes de comunicação, distribuição de energia elétrica;
- alocação de recursos;
- *machine learning* – desenho de redes neuronais, sistemas de classificação; otimização de redes neuronais artificiais;
- processamento digital de imagem;
- na música.

4. Metodologia

4.1. Área de estudo

A área de estudo foi o distrito de Setúbal que se situa a sul da capital Lisboa. De acordo com a nomenclatura das unidades territoriais (NUTS II), o sector norte do distrito (península de Setúbal) pertence á Área Metropolitana de Lisboa, enquanto que o sector sul pertence ao Alentejo Litoral. Apresenta uma área total de aproximadamente 5214 Km² sendo constituído por 13 municípios e 55 freguesias de acordo com a nova reorganização administrativa. A Figura 24 mostra a localização geográfica da área de estudo e os respetivos limites das freguesias.

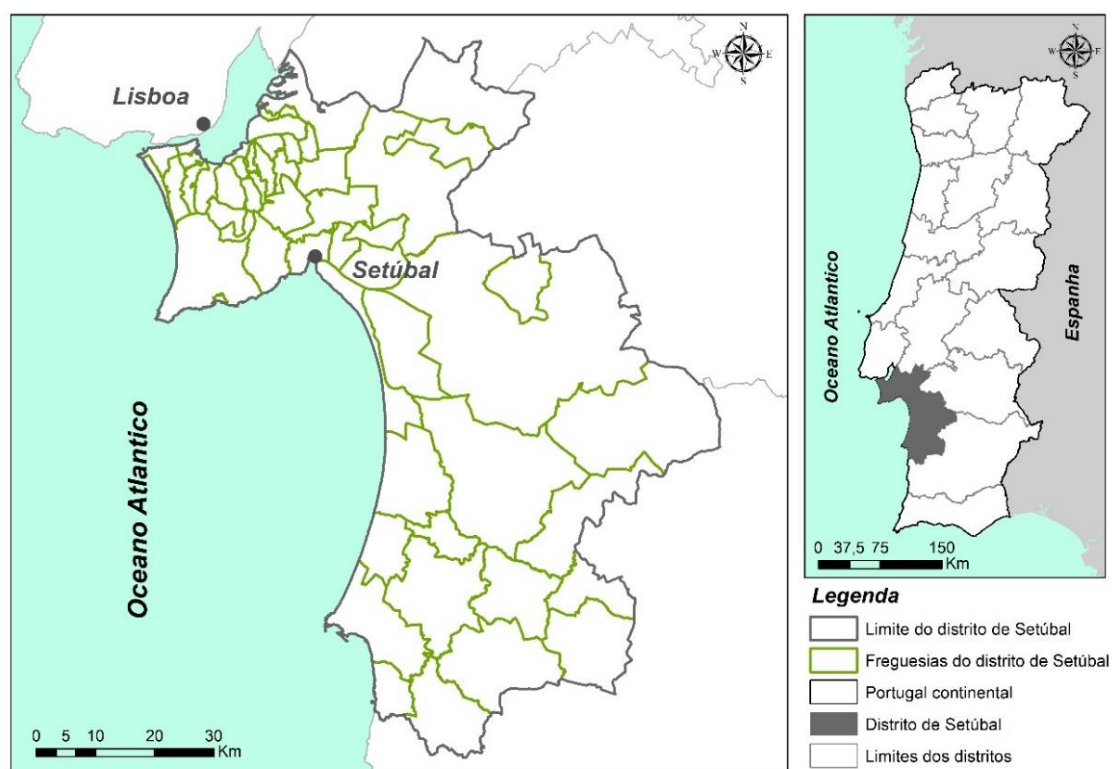


Figura 24 - Localização geográfica da Área de estudo (Fonte: CAOP, 2017).

A capital de distrito é Setúbal e em 2017 tinha uma população total de 116.330 habitantes de acordo com as projeções do INE (2018) e PORDATA (2018).

4. Metodologia

O distrito apresenta uma população total de 851 258 habitantes de acordo com os censos de 2011 (INE, 2014b).

4.2. Dados e software

Os dados utilizados dizem respeito à população residente dos censos de 2011 e as projeções de 2016. A sua recolha foi realizada no sítio do Instituto Nacional de Estatística (INE). Para este estudo efetuamos a agregação dos dados de acordo com a nova reorganização administrativa fornecida pela Carta Administrativa Oficial de Portugal (CAOP).

A divisão administrativa utilizada neste estudo foi a freguesia. Na Tabela 4 podem ser observados os dados utilizados tendo em conta o tema, formato e fonte.

Tema	Formato	Fonte
Limites administrativos	Vetorial	CAOP (2013 e 2017)
Censos de 2011	Alfanumérico	INE (2014b)

Tabela 4 - Tema, formato e fonte dos dados.

Os dados da população da área de estudo foram previamente preparados no *ArcMAP 10.5*, onde se efetuou a junção dos dados populacionais às freguesias da área de estudo. Calcularam as suas coordenadas dos eixos X e Y no sistema de referência de coordenadas Lisboa *Hayford-Gauss* militares. As soluções obtidas no AG foram acopladas na camada de freguesias do distrito de Setúbal.

O desenvolvimento prático desta dissertação foi realizado recorrendo à linguagem *Python*⁷, que foi criada por *Guido Van Rossum* no final dos anos 80 (Lutz, 2003). É uma linguagem interpretada, de alto nível e que suporta vários paradigmas da programação como o procedimental, funcional e orientada aos objetos. De acordo com Vasconcelos & Barão (2017), a linguagem *Python* oferece-nos um conjunto de mecanismos de programação avançada, para a implementação de algoritmos e resolução de problemas. A grande vantagem da sua utilização é podermos misturar com C ou *Java* (Costa, 2015).

A linguagem *Python* apresenta vários ambientes de desenvolvimento integrado (*Integrated Development Environment*), o IDE, para facilitar o desenvolvimento, teste e execução dos programas. Por outro lado, oferece-nos uma diversidade de bibliotecas

⁷ <https://www.python.org/>

4. Metodologia

em inúmeras áreas de conhecimento como a ciência dos dados (*scipy*, *numpy*), análise de dados (*Pandas*), análise gráfica (*matplotlib* e *Seaborn*), análise espacial (*PySAL – Python Spatial Analysis Library*), análise de clusters (*Clusterpy*).

O interpretador utilizado para o desenvolvimento dos algoritmos utilizados nesta dissertação foi o *Spyder* que faz parte do *software Anaconda* versão 2⁸ com a versão 2.7 do *Python*. Este *software* é muito mais do que um interpretador, visto que nos permite gerir e adicionar vários elementos ao nosso ambiente de desenvolvimento.

4.3. *Hot Spot Analysis*

A análise exploratória dos dados da população total residente, por freguesia e por faixas etárias foi efetuada com a *hot spot analysis*, que nos permite determinar a estatística *Getis-Ord Gi** para um dado conjunto de dados. Os valores do *z-score* e do *p-value* indicam-nos onde se situam os *clusters* com os valores mais baixos e mais altos. Um atributo com um valor mais alto pode não ser significativo. Para ser um *hot spot* estatisticamente significativo os atributos (freguesias) deverão ter vizinhos cujos valores sejam igualmente elevados (ESRI, 2009).

A soma local de um atributo, com os seus vizinhos é comparada proporcionalmente com o somatório de todos os atributos. Um *z-score* com significância estatística ocorre quando a soma local é muito diferente da esperada, e se a diferença não é um resultado aleatório. Os *cold spot*, ocorrem quando o *z-score* é menor (ESRI, 2009).

4.4. Modelo de otimização para *police re(districting)*

Os problemas de *districting* são considerados problemas de otimização combinatória, sendo por isso designados de NP-difíceis. Nos problemas de *districting* para a definição de áreas de serviço de esquadras de polícia, como já afloramos, são usados como critérios, o equilíbrio da carga de trabalho, a compacidade, contiguidade e os tempos de resposta a chamadas de serviços. Perante a ausência deste tipo de dados, foi necessário adequar o problema ao tipo de informação a que neste momento tínhamos acesso.

As propostas de Bação et al. (2005), com algoritmos genéticos, Castelli et al. (2015), através da programação genética, Kim et al. (2018) com SA e TS, oferecem

⁸ <https://www.anaconda.com/download/>

4. Metodologia

boas possibilidades ao nível de desempenho e eficiência. Não obstante, as aplicações terem sido aplicadas a problemas de (re)districting eleitoral consideramos que é um ponto de partida para definir as áreas de serviço com base em dados da população residente e por faixas etárias.

A nossa proposta de resolução foi no sentido da implementação de um AG baseado nas funções objetivo e restrições que os autores referenciados acima aplicaram.

4.4.1. Proposta de resolução

Segundo Bação et al. (2005) e Castelli et al. (2015) os referidos distritos, áreas de serviço designadas, devem ter as seguintes características:

- Equilíbrio da População, em que todos os distritos devem ser iguais em termos da população;
- Contiguidade, cada distrito deve ser um território único e contíguo;
- Compacidade geográfica, os distritos devem ser compactos;

A igualdade populacional foi avaliada pela soma das diferenças entre a população do distrito e a média da população de todas os distritos. Bação et al. (2005), partilham a função objetivo seguinte:

$$\min \sum_j |P_j - \mu| \quad \text{Equação IX}$$

Em que P_j representa a população da j ...ésima zona e μ a média da população por zona. Como medida de compacidade utilizou-se a soma das distâncias euclidianas entre os centróides, dada por:

$$\text{Compacidade radial} = \sum_j \sum_{i \in Z_j} d_{ij} \quad \text{Equação X}$$

Em que d_{ij} representa a distância euclidiana entre o i ...ésimo unidade de área e o j ...ésimo do centro da zona. Para incluir uma medida de compacidade na função objetivo pode ser através da soma ou produto:

4. Metodologia

$$\min \sum_j \left(|P_j - \mu| + \sum_{i \in Z_j} d_{ij} \right) \quad \text{Equação XI}$$

Outra medida de compacidade proposta por Bação et al. (2005) é a divisão entre a soma dos perímetros de cada distrito ao quadrado pela área do distrito. Aplicou-se esta medida para avaliar a compacidade das soluções obtidas no AG e pelo AZP-SA como descrevemos no ponto 4.6.

$$\sum_i \frac{Pr_i^2}{a_i} \quad \text{Equação XII}$$

A verificação da restrição da contiguidade foi também realizada através da construção de uma matriz binária, com o *PySal* com relações de vizinhança do tipo *Queen*.

No que concerne à complexidade do problema aqui resolvido é importante referir que a unidade administrativa básica foi a freguesia. Serão utilizadas 55 freguesias para 28 áreas de influência, isto é, implementação potencial para 28 esquadras. Para o problema em concreto temos cerca de 2441460 possibilidades. A dimensão do problema combinatório é dada pela expressão seguinte:

$$S = \frac{(n-1)!}{(k-1)!(n-k)} \quad \text{Equação XIII}$$

Em que:

n - Número de freguesias

k – Número de distritos

4.5. Implementação do AG

Os passos fundamentais para a implementação de um AG são a escolha da codificação e dos operadores genéticos a utilizar. Existem inúmeras formas para codificar um problema de *districting*. A codificação que foi adotada é muito semelhante à utilizada para resolver problemas das p -medianas. Os exemplos mais conhecidos são os de Correa et al., (2001) e Alp et al. (2003).

A unidade elementar escolhida para os centróides foi a freguesia. O centróide da freguesia é um cliente e ao mesmo tempo uma potencial instalação. O centro do distrito é representado por um centróide que irá representar todos os outros, princípio

4. Metodologia

semelhante ao das p -medianas em que cada centróide é alocado a uma mediana (Figura 25).

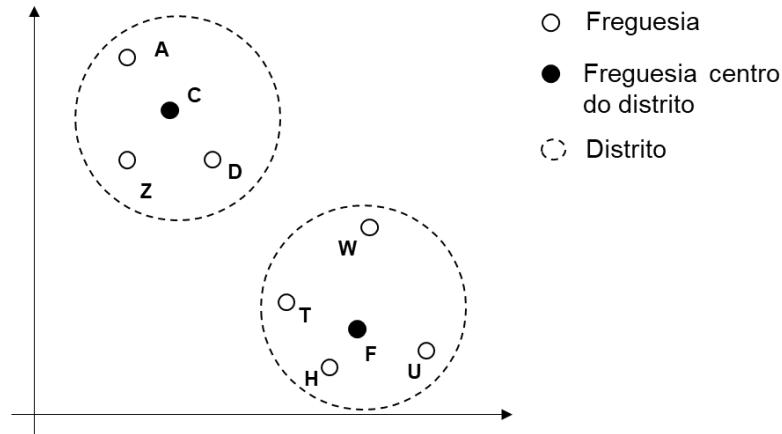


Figura 25 - Representação dos centróides das freguesias e dos distritos (Fonte: Elaboração própria).

Para representar os centróides candidatos codificaram-se os mesmos com letras do abecedário. Esta operação foi realizada na tabela de atributos da camada de polígonos. O exemplo representado na Figura 25 será dado por um dicionário de listas como : {'C': ['A', 'Z', 'D'], 'F': ['W', 'H', 'U', 'T']}.

O comprimento do cromossoma é dado pelo número de zonas a serem criadas, em que é codificado com uma *string* de comprimento k (k é o número de zonas) onde o n ésimo elemento da *string* indica a posição do centro da zona (Bação et al., 2005). Os genes assumem aqui letras que correspondem ao centróide de cada freguesia. No caso do distrito administrativo de setúbal estabelecemos que deveria ter 28 esquadras o que perfaz áreas de serviço com a capacidade de servir 30000 residentes. Os distritos formam as áreas de serviço das forças de segurança.

Para forçar a contiguidade das soluções determinaram-se os vizinhos de primeira ordem de cada freguesia através biblioteca *PySal*. Esta biblioteca permite determinar uma matriz de contiguidade através de uma camada de polígonos. A leitura da matriz de contiguidade é muito simples, por exemplo, o centróide A tem como vizinho de primeira ordem o B, de segunda ordem o D (Figura 26).

	A	B	C	D
A	0	1	0	1
B	0	0	1	0
C	0	1	0	0
D	0	1	1	0

Figura 26 - Exemplo de uma matriz de contiguidade (Fonte: Elaboração própria).

4. Metodologia

Aplicou-se um mecanismo de procura, através de ciclos, na matriz para posteriormente ser utilizado na produção das soluções iniciais.

A estratégia de inicialização tem em conta os vizinhos de primeira ordem fornecidos pela matriz de contiguidade. Só é construída a solução inicial se os centróides forem contíguos, caso contrário, o algoritmo corrige e elimina os cromossomas inválidos. Aqui há garantias de que as soluções iniciais são contíguas.

A avaliação das soluções foi implementada para a equação IX e XI. Com a equação IX apenas utilizamos os valores da população residente enquanto que com a XI adicionamos o fator distância.

O método de seleção dos pais foi através da ordenação descendente, isto é, do cromossoma com maior valor na função de avaliação para o menor valor. O *crossover* nos progenitores parametrizado foi de um ponto de corte aleatório. Posteriormente, após o *crossover* aplicou-se a condição de que os novos cromossomas devem ser corrigidos caso tenham algum gene repetido. A taxa de *crossover* é de 100%, isto é, ocorre sempre.

Outro operador genético utilizado foi a mutação para introduzir diversidade nas soluções através da troca de posição de genes. Aqui é realizada a troca de posição de forma aleatória da chave ou valor. A taxa de mutação é de 1%, isto é, 1 em cada 100 cruzamentos ocorre a mutação. Se a divisão entre o valor da população com a taxa de mutação for resto 0 (zero), então a mutação ocorre, caso contrário não ocorre. Posteriormente à aplicação dos operadores genéticos a melhor população passa à geração seguinte. Por fim, o critério de paragem do AG utilizado foi a geracional. Quando o algoritmo atinge o número de iterações definidas este para e gera a solução final.

Todas as soluções finais foram analisadas ao nível de contiguidade para atestar a sua validade. O código do *script* utilizado está disponível no Anexo 8.2.

4.5.1. Testes com diferentes parametrizações

A verificação da performance e consistência do algoritmo foi realizada tendo em conta as medidas estatísticas de tendência central, máximos e mínimos obtidos. Realizaram-se 25 corridas em 7 repetições diferentes. As 6 primeiras repetições foram implementadas com o número de residente total por freguesia e na sétima utilizou-se o número de residentes com mais de 65 anos. Utilizaram-se diferentes tamanhos da população inicial (10, 50, 100), com as equações XI e XIII.

4.5.2. Definição de áreas de serviço

As áreas de serviço foram definidas de acordo com os menores valores das funções objetivo determinadas nas experiências com maior consistência e desempenho no ponto 4.5.3. As soluções foram acopladas à informação geográfica para obter uma visualização espacial.

Para comparar as soluções obtidas foram analisados os valores da soma das diferenças da população dada pela equação IX e a compacidade dada pela equação XII.

4.6. Comparação com o AZP-SA

Com o intuito de verificar e confrontar o desempenho do AG utilizou-se a biblioteca *clusterpy* construída por Duque et al. (2011), que nos fornece um conjunto de algoritmos de *clustering* e de regionalização. Os algoritmos disponíveis são o AZP com as variantes do SA e o TS, os SOM, Geospatial SOM, entre outros.

O AZP-SA é um algoritmo que agrega N zonas (áreas) em regiões, que segundo Openshaw & Rao (1995), as regiões resultantes devem ser formadas por zonas contíguas e conectadas internamente. A função objetivo é uma variante do AZP mas com a inclusão da meta-heurística SA (Kirkpatrick et al., 1983), para efetuar a procura das soluções. O SA começa a procura de forma aleatória até encontrar uma melhoria local na função objetivo. Quando não encontra uma melhoria então é calculada a probabilidade através da equação de *Boltzman* (Openshaw & Rao, 1995).

A grande vantagem desta biblioteca é que basta parametrizar o diretório e o campo de uma *shapefile* com os dados a analisar e posteriormente devolve, um *output* no mesmo formato com a solução. Aplicou-se o AZP-SA na população total residente por freguesia, e por faixa etária maior que 65 anos para comparar nas mesmas condições que o AG. O código do *script* utilizado está disponível no Anexo 8.3.

5. Resultados e Discussão

5.1. Análise exploratória de dados

5.1.1. Estatística descritiva

Na Figura 27 apresenta-se a distribuição do número de residentes representada pelo método de 5 quantis do distrito administrativo de Setúbal. A sul e a este do distrito o número de habitantes residentes é menor enquanto que na península de Setúbal é onde se concentra cerca de 82% da população residente, como referido anteriormente.

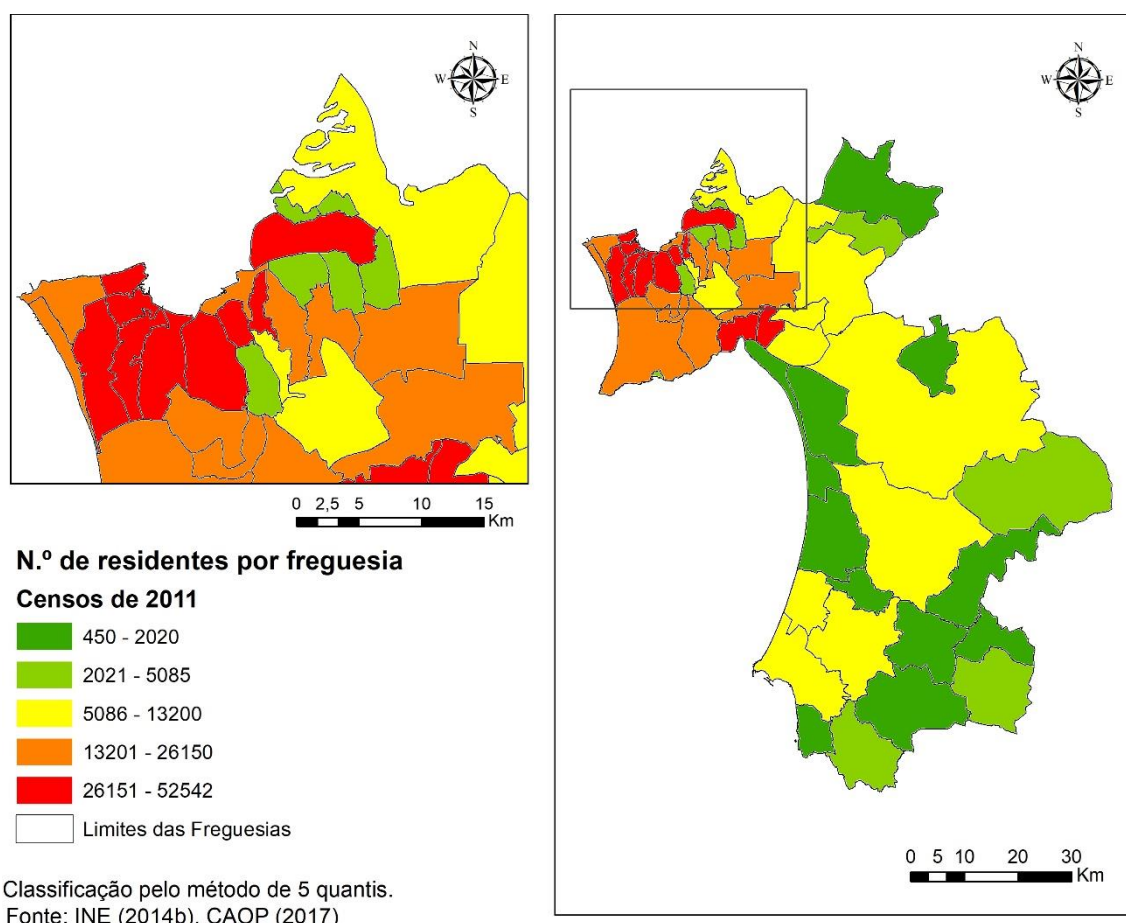


Figura 27 - População residente por freguesia representada pelo método dos quantis em 2011.

5. Resultados e discussão

O distrito de Setúbal como a maioria dos distritos administrativos em Portugal continental, apresenta uma grande assimetria no que concerne à distribuição da população. Este acontecimento dificulta um planeamento equitativo quer das forças de segurança quer dos demais serviços públicos.

As medidas de tendência central e de dispersão são reflexo da discrepância entre os valores mínimos e máximos de residentes que habitam em cada freguesia (Tabela 5). A freguesia com menor número de residentes é São Martinho do município de Alcácer do Sal com 450 habitantes residentes. A freguesia com maior número de residentes é Setúbal (São Sebastião) com 52542 habitantes residentes.

	População Residente (n.º hab.)	Área (km²)	Densidade Populacional(hab./Km²)
Média	15477,42	94,80	987,93
Mediana	10647,00	50,72	232,11
Desvio-padrão	15865,98	142,47	1636,36
Mínimo	450,00	2,02	4,08
Máximo	52542,00	888,37	8068,40

Tabela 5 - Estatística descritiva da população residente e área no distrito de Setúbal.

No que concerne à área, a freguesia com menor expressão é Sesimbra (Santiago) com 2.01 Km², a maior é a União de freguesias de Santa Maria do Castelo, Santiago e Santa Susana, com 888,36Km².

Observando o histograma do número de residentes por freguesia (Figura 28) confirma-se o que se tinha constatado pelas medidas de tendência central, a assimetria é negativa, enviesada à esquerda, em que a mediana do número de residentes é inferior à média. Quer dizer que temos mais freguesias com menor número de residentes comparativamente com as freguesias mais populosas. Na estratificação de acordo com o histograma (Figura 28), 24 freguesias apresentam entre 450 e 7891 residentes e a classe com menor número de freguesias é entre 30216 e 37658 residentes com 2.

5. Resultados e discussão

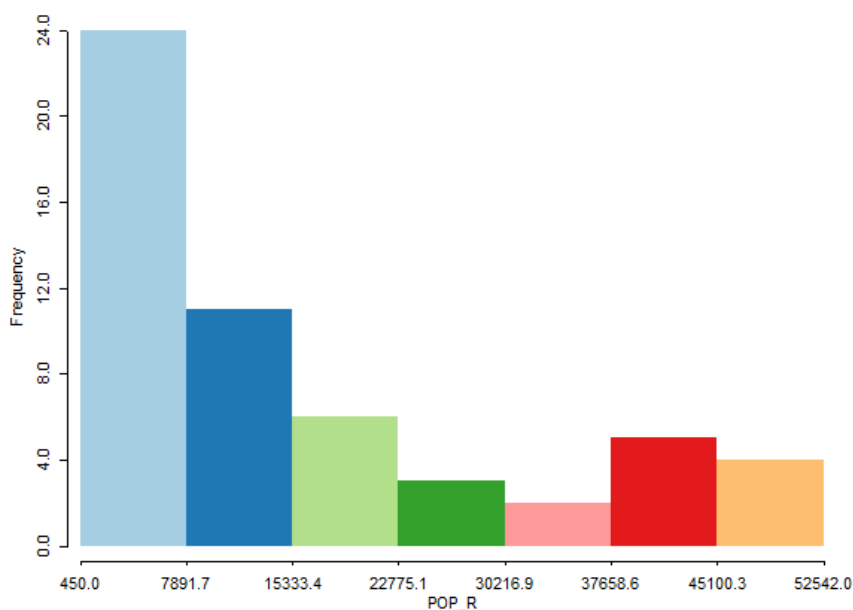


Figura 28 - Histograma da distribuição da população residente por freguesia em 2011 (Fonte: INE, 2014b).

5.1.2. Análise da população por faixa etária de interesse

A análise da população foi efetuada apenas nas faixas etárias cujos residentes se encontrem no ativo e na faixa etária maior que 65 anos, isto é, grupos vulneráveis. A península de Setúbal apresenta a maior concentração de população residente (Figura 29a), sendo por isso um *hot spot* estatisticamente positivo, enquanto que na zona sul do distrito situam-se os *cold spot*, isto é, menor concentração de residentes. A restante área não apresenta significância perante o valor estatístico Getis-Ord G_i^* . As freguesias de Abela e a União de Freguesias de São Domingos e Vale de Água apresentam um *cold spot* de 95%. Por sua vez, Ermidas do Sado, Alvalade, Azinheira dos Barros e a União de Freguesias de Santiago do Cacém, S. Cruz e S. Bartolomeu da Serra apresentam um *cold spot* de 90%.

Os *hot spot* com um maior nível de significância da faixa etária superior aos 65 anos situam-se na península de Setúbal, à exceção da freguesia de Sesimbra (Santiago), enquanto que os *cold spot* estão localizados no sul do distrito (Figura 29b). Os *hot spots* com um nível de significância de 95% são as freguesias de Azeitão, União de freguesias do Montijo e Afonsoeiro, Quinta do Anjo e Samouco. A freguesia de São Francisco apresenta o valor de 90%. Relativamente aos *cold spot*, a freguesia de Abela apresenta 95% enquanto que Alvalade, a União de freguesias de São Domingos e Vale de Água e a União de Freguesias de Santiago do Cacém, S. Cruz e S. Bartolomeu da

5. Resultados e discussão

Serra apresentam *cold spot* de 90%. Todas as restantes freguesias não apresentam valores significativos.

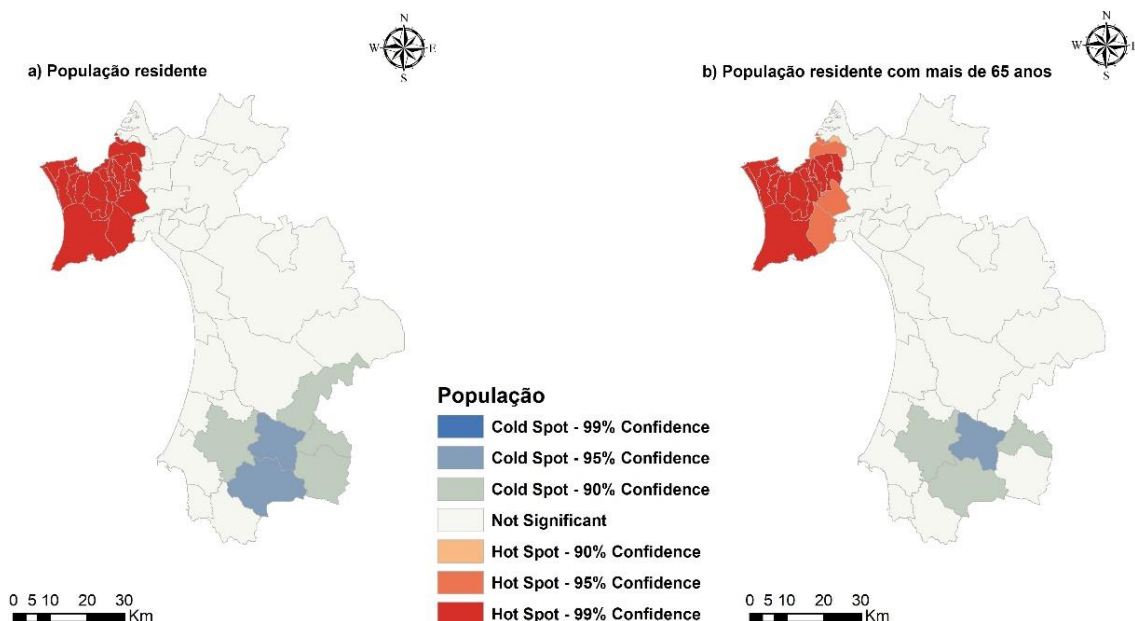


Figura 29 - Análise hotspot realizada para a população residente (a) e para a faixa etária maior que 65 anos (b) do ano 2011.

No que respeita à faixa etária entre os 20 e 24 anos é nas freguesias da península de Setúbal que se concentram os *hot spots* com 99% de confiança à exceção da freguesia do Samouco com 95% (Figura 30a). No que concerne aos *cold spot*, à semelhança do que acontece com a população total residente, a freguesia de Abela apresenta 95% enquanto que Alvalade, a união de freguesias de São Domingos e Vale de Água e a União de Freguesias de Santiago do Cacém, S. Cruz e S. Bartolomeu da Serra apresentam *cold spot* de 90%.

A faixa etária dos 24 aos 64 anos apresenta os mesmos níveis de significância no *hot spot* e no *cold spot* que a população residente (Figura 30b). Em todos os testes Getis-Ord G_i^* a freguesia de Sesimbra (Santiago) não é significativa.

Esta análise dos *hot spot* e *cold spot* permitiu averiguar onde se concentra o maior e menor número de residentes por faixas etárias de interesse. De facto, é na península de Setúbal que se concentram o maior número de residentes e por faixa etária, o que significa que do ponto de vista de planeamento existe a necessidade de ter um maior dispositivo quer seja em número de esquadras, quer seja número de efetivos e de viaturas. Por outro lado, as freguesias cujo número de residentes é menor situam-se no sul e centro do distrito administrativo confirmando a grande variação da população residente. Sob o ponto de vista da otimização consideramos que o maior

5. Resultados e discussão

desafio prende-se com uma agregação adequada de freguesias mediante a distribuição assimétrica da população.

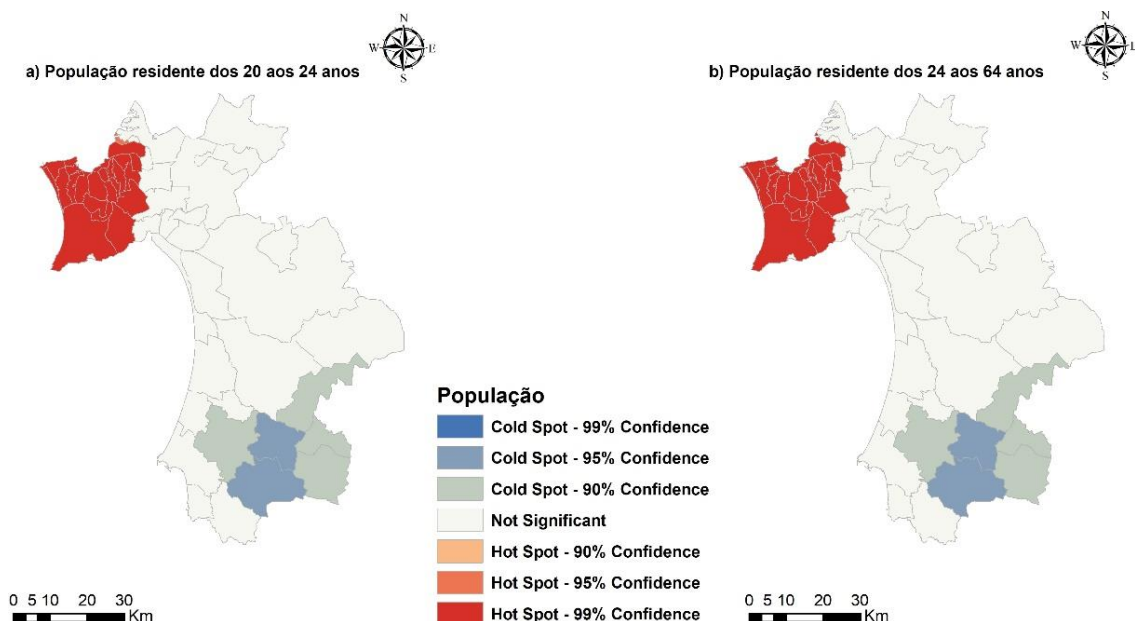


Figura 30 - Análise hotspot realizada para a faixa etária entre 20 e 24 anos (a) e para a faixa etária entre 24 e 64 anos (b).

Para além da análise da distribuição da população residente por faixa etária com *hot spot analysis* seria proveitoso efetuar este exercício com o número de crimes e sua tipologia por freguesia. Chow et al. (2015), efetuaram previamente uma análise da distribuição dos crimes relacionando com a localização das esquadras de polícia para posteriormente aplicarem um algoritmo adequado. No nosso entendimento explorando as taxas de crime com uma unidade de área desagregada e com a localização atual das esquadras de polícia facilita a análise a montante. Ribeiro & Henriques (2018), aplicaram esta análise por municípios com os dados da criminalidade porque não existem dados desagregados ao nível da freguesia, disponíveis e publicados pelas forças de segurança.

5.2. Resultados experimentais com o AG

Os resultados experimentais dos testes iniciais foram apresentados sob forma de tabela para comparar com as diferentes parametrizações. Posteriormente, espacializaram-se as soluções com a função de avaliação mais baixa para representar as áreas de serviço de cada potencial localização de esquadra de polícia. Aqui não houve uma diferenciação do tipo de corpo de polícia.

5. Resultados e discussão

5.2.1. Testes com as diferentes parametrizações

A Tabela 6 resume os 150 testes e 6 repetições realizadas com diferentes populações para um total de 28 áreas de serviço.

A média da função de avaliação sem o fator distancia, mais baixa obtida nesta amostra de testes foi de 625603,1, com uma população inicial de 100 indivíduos e uma taxa de mutação de 1%. A média mais elevada foi obtida com 10 indivíduos no valor de 805821,3. Todas as outras médias das medidas de tendência central foram igualmente mais baixas com 100 indivíduos, à exceção do desvio padrão obtido com 50 populações cujo valor foi de 41965,3.

No que respeita à média da função de avaliação com o fator distância, a mais baixa foi obtida com 100 indivíduos e a mais elevada com 10 indivíduos, cujos valores são 855679 e 1047218,5, respetivamente.

Todas as outras médias das medidas de tendência central foram igualmente mais baixas com 100 indivíduos. Aumentando o tamanho da população independentemente da função de avaliação utilizada, os valores obtidos foram sempre mais baixos.

N.º de Populações	Equação IX			Equação XI		
	10	50	100	10	50	100
Média	805821,3	692329,4	625603,1	1047218,5	927521	855679
Mediana	742112,0	694997,0	635745,0	1021655,2	905623,9	845633
Desvio-padrão	146582,2	41965,3	55688,9	120543,4	87664,35	44492,8
Mínimo	635392,0	602039,0	521625,0	803367,0	799243,9	789905
Máximo	1230799,0	776293,0	751331,0	1319550,6	1099561	957646

Tabela 6 - Teste realizados com a área de estudo com diferentes parametrizações das populações.

Pela análise da Tabela 6, o número de indivíduos das populações teve muito impacto nos resultados. Neste caso, quanto maior o tamanho da população quer seja com o fator distancia ou não, menor valor da função de avaliação. Estes resultados indiciam que com populações com menos indivíduos poderá ter ocorrido convergência prematura, isto é, convergem para um ótimo local (Malczewski & Rinner, 2015). A posição privilegiada de alguns genes pode ter dominado rapidamente a população o que a força a cair em ótimos locais (Malik & Wadhwa, 2014).

5.2.2. Definição de áreas de serviço com o AG

De acordo com a parametrização realizada sem o fator distância (Equação IX) para a população residente por freguesia, 100 populações e uma taxa de mutação de 1%, podemos observar espacialmente pela Figura 31, as potenciais localizações e

5. Resultados e discussão

respetivas áreas de serviço das esquadras de polícia simuladas com o centróide de cada polígono das freguesias.

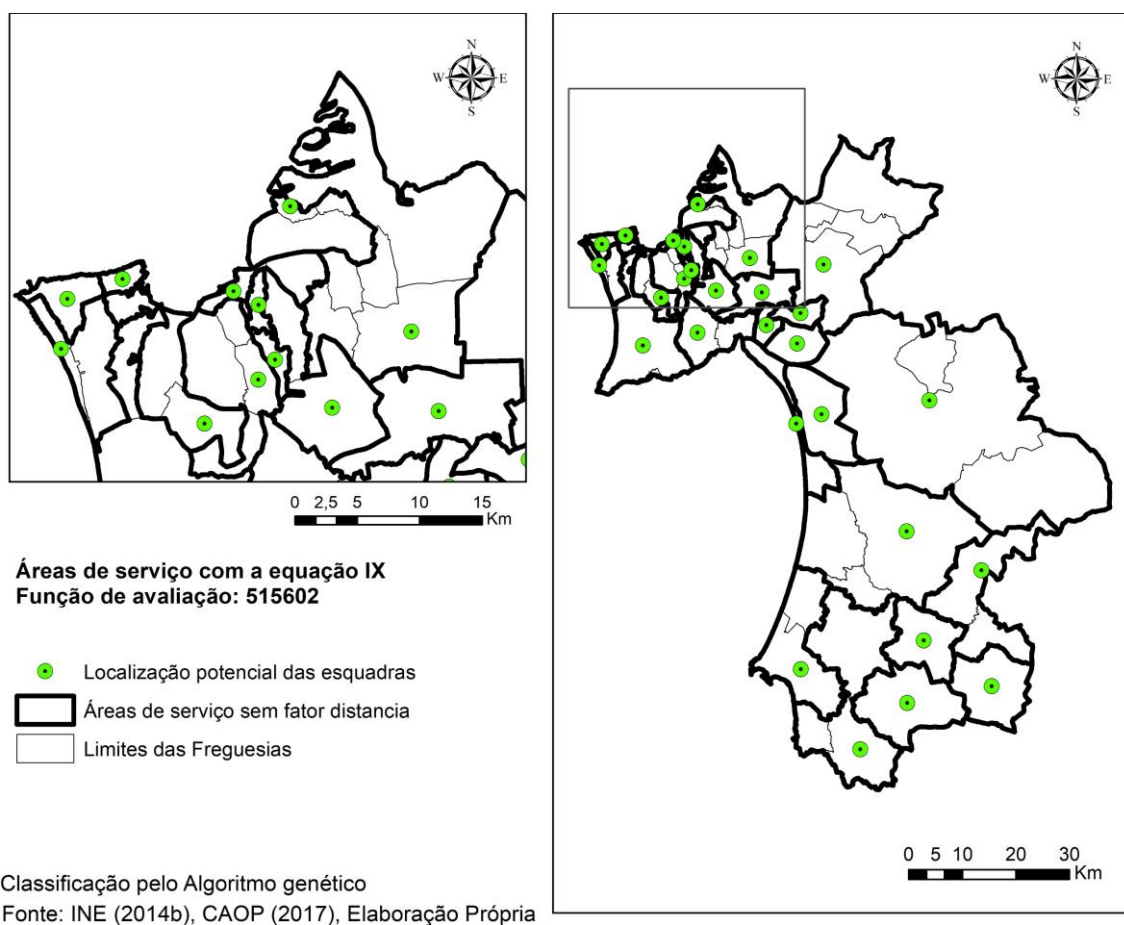


Figura 31 - Desenho das áreas de serviço sem o fator distância tendo em conta a população total parcial por freguesia para 2011.

Verifica-se que a maioria das potenciais localizações das esquadras de polícia localizam-se península de Setúbal, devido à maior concentração da população comprovada na análise exploratória efetuada dos pontos quentes. No que concerne à soma da população residente afetada a cada potencial esquadra, verifica-se uma grande variação (Tabela 7), que de facto se deve em grande parte à distribuição assimétrica da população residente que dificulta o algoritmo na procura das combinações entre freguesias. Como podemos observar pela Tabela 7, a área de serviço com menor população afetada é de 890 residentes que corresponde ao centróide da freguesia de Abela. A área de serviço com maior soma da população afetada diz respeito à agregação das freguesias da Quinta do Conde, Palhais e Coina, união de freguesias do Seixal, Arrentela e Aldeia do Paio Pina e por fim, com a união de freguesias do Alto do Seixalinho, Santo André e Veredena com 115877 habitantes.

5. Resultados e discussão

Neste caso específico, a média da população nesta área de serviço é de 28969 residentes.

	População afetada (n.º hab.)	Área (km²)
Média	30402,1	186,8
Mediana	20664	79,9
Desvio-padrão	29239,8	278,6
Mínimo	890,0	6,2
Máximo	115877,0	1353,5

Tabela 7 - Estatística descritiva da população afetada por área de serviço com a Equação IX.

No que respeita à área ocupada por Km², a área de serviço menor tem 6,2 Km² e a maior apresenta 1353,5km².

Sob o ponto de vista de otimização a soma total da população por área de serviço deveria se situar em média na ordem de 30.000 residentes. No entanto, existem várias freguesias deste distrito que apresentam valores de população residente superiores ao estipulado para cada esquadra de polícia o que dificulta a distribuição das áreas de serviço. Se analisarmos os valores médios por cada área de serviço a discrepância diminui, mas ainda assim o dispositivo policial nestes casos careceria de um número elevado de polícias ao serviço bem como uma carga de trabalho muito exigente.

A situação contrária passa-se na zona sul do distrito, as somas da população afetada por cada potencial esquadra são mais baixas devido ao número de residentes que também é muito menor, tal como verificamos na análise *hot spot*.

Com a inclusão do fator distância na função de avaliação as médias melhoraram significativamente, no entanto, à semelhança do que sucedeu com a simulação sem o fator distância, os valores da soma da população ultrapassam o valor estipulado inicialmente. A Figura 32 apresenta as potenciais localizações e respetivas áreas de influência das esquadras de polícia simuladas com o centróide de cada polígono das freguesias adicionando à equação o fator distância.

5. Resultados e discussão

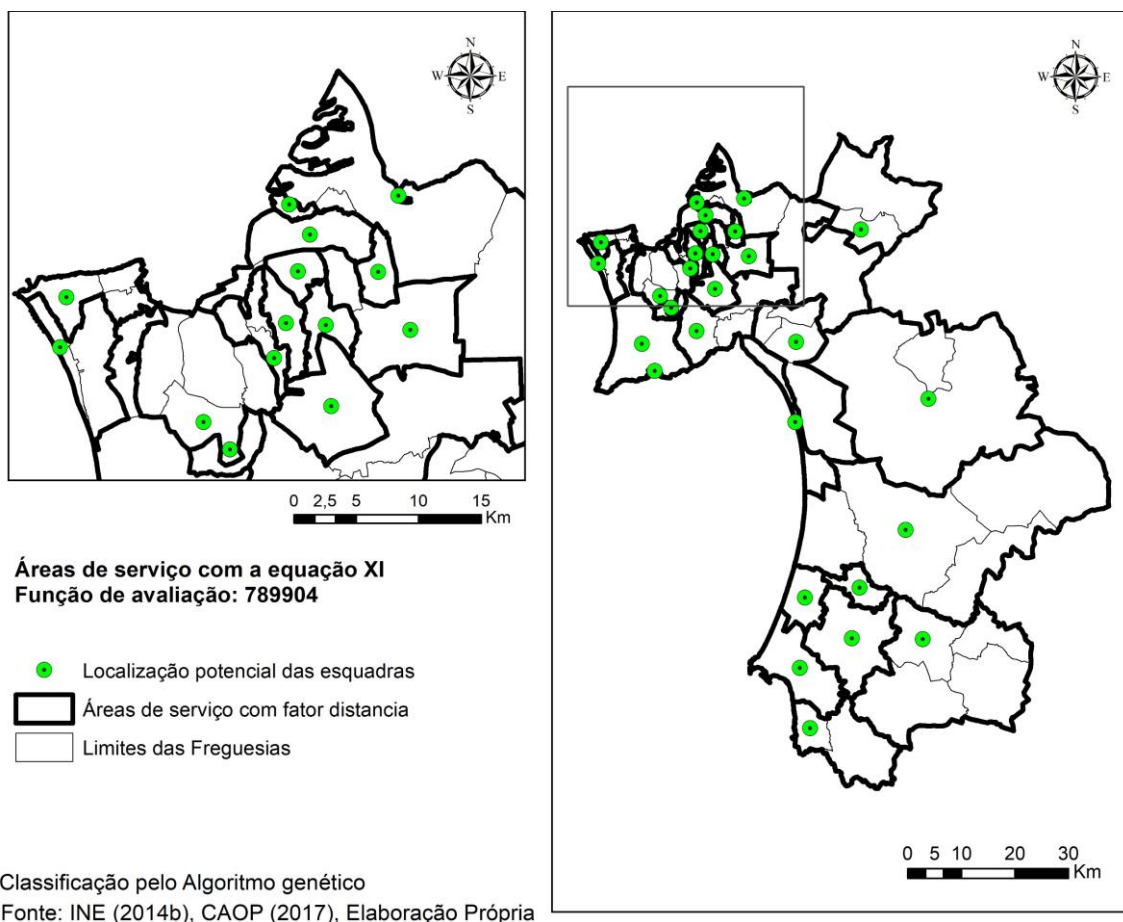


Figura 32 - Desenho das áreas de serviço com o fator distância tendo em conta a população total parcial por freguesia para 2011.

A esquadra com menor população afetada é de 809 residentes que corresponde ao centróide da freguesia de São Francisco da Serra. A área de serviço com maior soma da população afetada é constituída pelas uniões de Freguesias de Almada, Cova da Piedade, Pragal e Cacilhas, União das freguesias da Caparica e Trafaria, e União de freguesias do Laranjeiro e Feijão. Este valor máximo não foi caso único, visto que obtivemos outra área de serviço com uma soma da população afetada de 114199 cuja média é de 28550 residentes (Tabela 8). Neste caso, é uma solução inaceitável sob o ponto de vista de planeamento e organização das forças de segurança porque a soma total desta área de serviço é de 115683 residentes.

	População afetada (n.º hab.)	Área (km ²)
Média	30402,1	186,8
Mediana	14125	70,9
Desvio-padrão	34600,5	280,5
Mínimo	890,0	2,02
Máximo	115683,0	1119,2

Tabela 8 - Estatística descritiva da população afetada por área de serviço com a Equação XI.

5. Resultados e discussão

Relativamente à área ocupada por Km², a área de serviço menor tem 2,02Km² e a maior apresenta 1119,2km² (Anexo 8.6).

No nosso entendimento estes resultados devem-se às fortes restrições (*hard constraints*) de vizinhança implementadas logo na inicialização do algoritmo, que limita o espaço de procura. Como já foi referido, só foram considerados os vizinhos de primeira ordem o que não facilita a combinação entre freguesias.

Se observamos os resultados obtidos por Bação et al. (2005), a soma da população por zona é muito semelhante entre si, o que não aconteceu na nossa experiência. Castelli et al. (2015), embora tenham aplicado um algoritmo genético através da programação genética e os números de zonas sejam diferentes, podemos verificar que a variação da população por zona criada é muito menor do que a obtida por nós.

A grande desvantagem de simular áreas de serviço com os limites de freguesias e respetivo centróide, em grandes centros urbanos, não permite que exista uma divisão equitativa da população. Nestes casos, a opção de simular por bairros permitiria melhorar a análise. Por outro lado, o tipo de codificação com os pares de coordenadas em X e Y aplicada por Bação et al. (2005), permitiria aumentar o espaço de procura não forçando a que os centros fossem apenas os centróides das freguesias. Nos Anexos 8.5 e 8.6 podemos averiguar quais são as freguesias afetadas a cada área de serviço por cada algoritmo.

5.2.2.1. Comparação com o AZP-SA

A Figura 33 apresenta o resultado da aplicação do algoritmo AZP-SA para o mesmo número de áreas de serviço e para a população residente. Com este algoritmo de regionalização, como já foi afluado anteriormente, apenas obtemos os limites das áreas de serviço com as respetivas métricas da soma da população total e médias. Nos Anexos 8.5 e 8.6 podemos averiguar quais são as freguesias afetadas a cada área de serviço por cada algoritmo.

5. Resultados e discussão

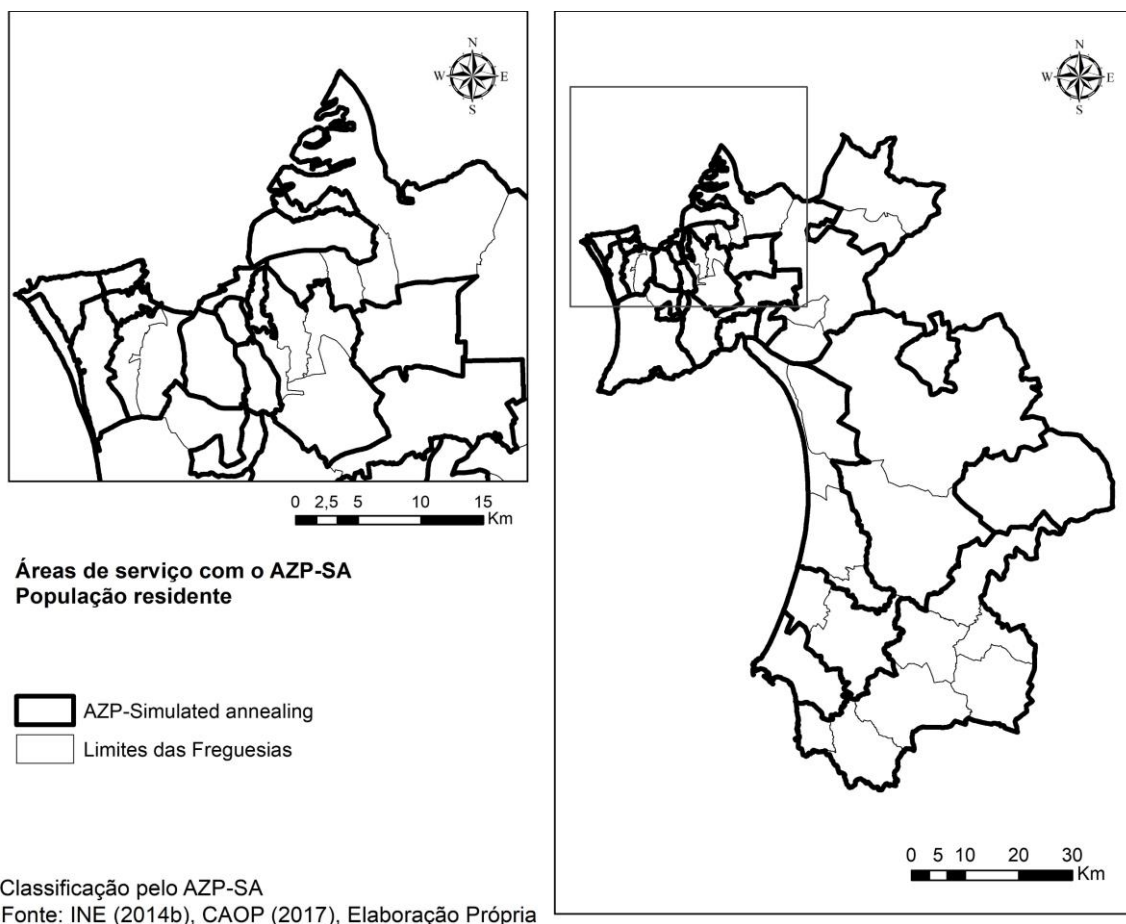


Figura 33 - Áreas de serviço com o AZP-SA para a população residente.

Pela análise da Figura 33, podemos observar que a maior concentração de áreas de serviço se situa na península de Setúbal, devido às razões já referenciadas. A área de serviço com menor valor é constituída pela freguesia de São Martinho com 450 habitantes residentes. A área de serviço com maior soma da população residente é constituída pelas freguesias de Corroios e Amora com um total de 92290 (Tabela 9).

	População afetada (n.º hab.)	Área (km ²)
Média	29980,6	186,8
Mediana	33653,5	34,2
Desvio-padrão	22138,5	349,3
Mínimo	450,0	2,02
Máximo	96290,0	1592,6

Tabela 9 - Estatística descritiva da população afetada por área de serviço com o AZP-SA.

Relativamente à área ocupada por Km², a área de serviço menor tem 2,02Km² e a maior apresenta 1592,6km².

5. Resultados e discussão

Ao nível da soma da população por área de serviço os resultados do AZP-SA são ligeiramente melhores do que foram obtidos com o AG que foi aplicado. Não obstante, as áreas de serviço continuam a apresentar uma variação entre máximos e mínimos com uma grande diferença. Prova destas constatações são as somas das diferenças da população que foram obtidas. Com a equação XI obtivemos a maior soma das diferenças da população comparativamente com a equação XIII e o AZP-SA (Tabela 10).

	Compacidade	Soma das diferenças da população
Equação IX	1351,4	678070,4
Equação XI	1332,7	789786,9
AZP-SA	1332,9	501320,7

Tabela 10 - Comparação entre os resultados obtidos entre o AG e o AZP-SA para as mesmas condições.

Ao contrário dos resultados obtidos por Bação et al. (2005), em que obtiveram somas da população com menores, com o seu AG, com as mesmas equações utilizadas por nós, do que com o ZDES cujo algoritmo é o AZP-SA. As medidas de compacidade são muito semelhantes, mas na soma das diferenças da população o AZP-SA teve melhor desempenho. No entanto, a medida de compacidade utilizada acaba por ser muito discutível, visto que este índice favorece áreas com maiores dimensões, que no nosso entendimento são de difícil gestão e distribuição da carga de trabalho, para as forças de segurança. Nos Anexos 8.5 e 8.6 podemos averiguar quais são as freguesias afetadas a cada área de serviço por cada algoritmo.

5.2.3. Definição de áreas de serviço para a faixa etária maior que 65 anos

De acordo com a parametrização realizada com o fator distância para a população residente com mais de 65 anos por freguesia, podemos observar espacialmente pela Figura 33, as potenciais localizações e respetivas áreas de serviço das esquadras de polícia simuladas com o centróide de cada polígono das freguesias.

À semelhança dos testes que foram efetuados com diferentes populações realizados com a população residente, efetuaram-se 25 corridas e apenas uma repetição com 100 populações e 500 gerações. Os resultados obtidos foram idênticos aos testes com diferentes parametrizações e podem ser observados no anexo 8.4.

A área de serviço com menor população afetada apresenta 161 residentes com idade superior a 65 anos, que corresponde ao centróide da freguesia de São Martinho. A área de serviço com maior soma da população afetada é constituída por duas freguesias, a União de Freguesias de Almada, Cova da Piedade, Pragal e Cacilhas e

5. Resultados e discussão

pela União das freguesias da Caparica e Trafaria, com 18493 habitantes cuja média é 9247 (Figura 34).

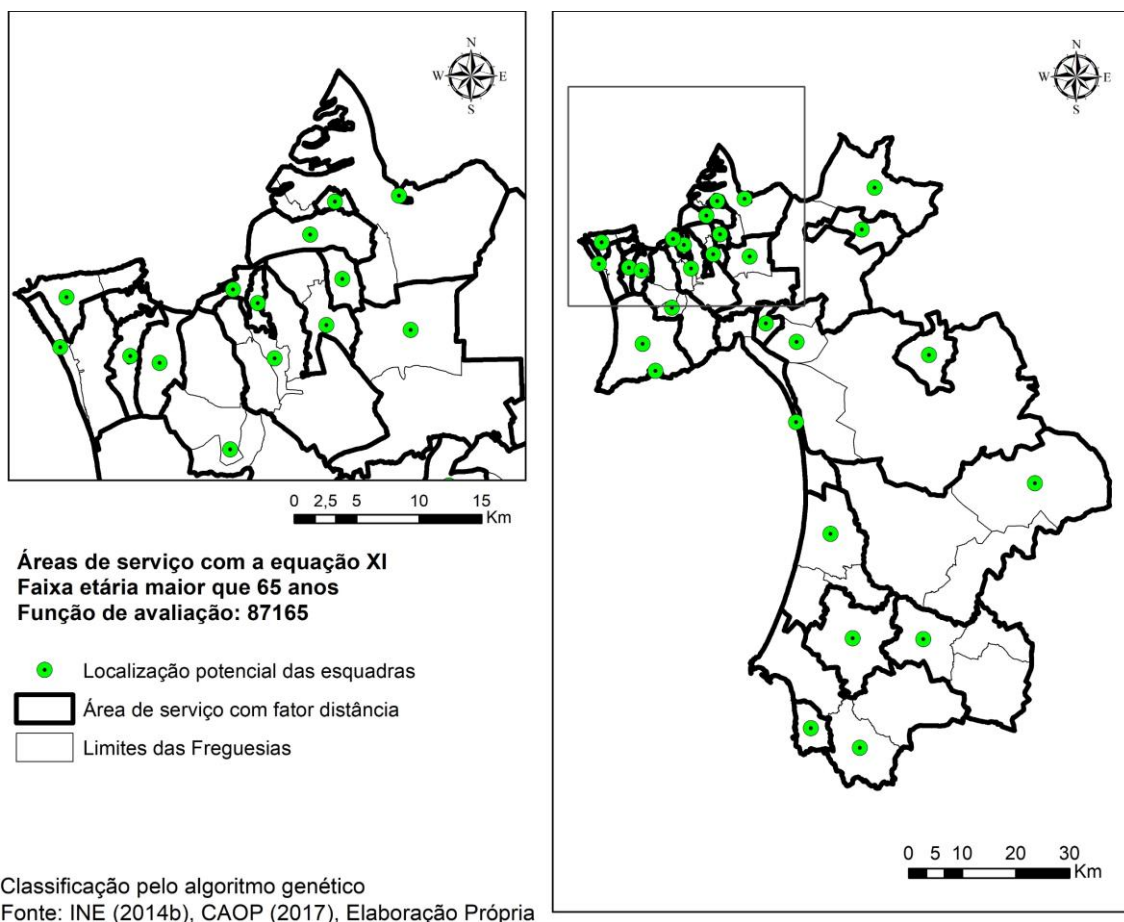


Figura 34 - Desenho das áreas de serviço com o fator distância tendo em conta a população residente com mais de 65 anos, parcial por freguesia para 2011.

À semelhança do AG aplicado à população total existe uma grande discrepância entre a população residente com mais de 65 anos por área de serviço (Tabela 11). Pela análise do *hot spot* tínhamos verificado que a assimetria nesta faixa etária era bastante evidente, no entanto, numa dimensão um pouco diferente em relação à população residente como pode ser constatado pela análise *hot spot*.

	População afetada (n.º hab.)	Área (km ²)
Média	5599,8	186,8
Mediana	3696,5	84,4
Desvio-padrão	5278,6	282,3
Mínimo	161,0	2,02
Máximo	18493,0	1140,8

Tabela 11 - Estatística descritiva da população afetada por área de serviço com a Equação XI.

5. Resultados e discussão

Relativamente à área ocupada por Km², a área de serviço menor tem 2,02Km² e a maior apresenta 1140,8km² (Anexo 8.6).

A ideia de criar áreas de serviço para a faixa etária maior do que 65 anos pode permitir promover o sentimento de segurança e o aumento de confiança nas forças de segurança, visto que de acordo com Bravo et al. (2018), perspectiva-se o aumento desta faixa etária.

5.2.3.1. Comparação com o AZP-SA

A Figura 35 apresenta o resultado da aplicação do algoritmo AZP- SA para o mesmo número de áreas de serviço e para a faixa etária maior do que 65 anos.

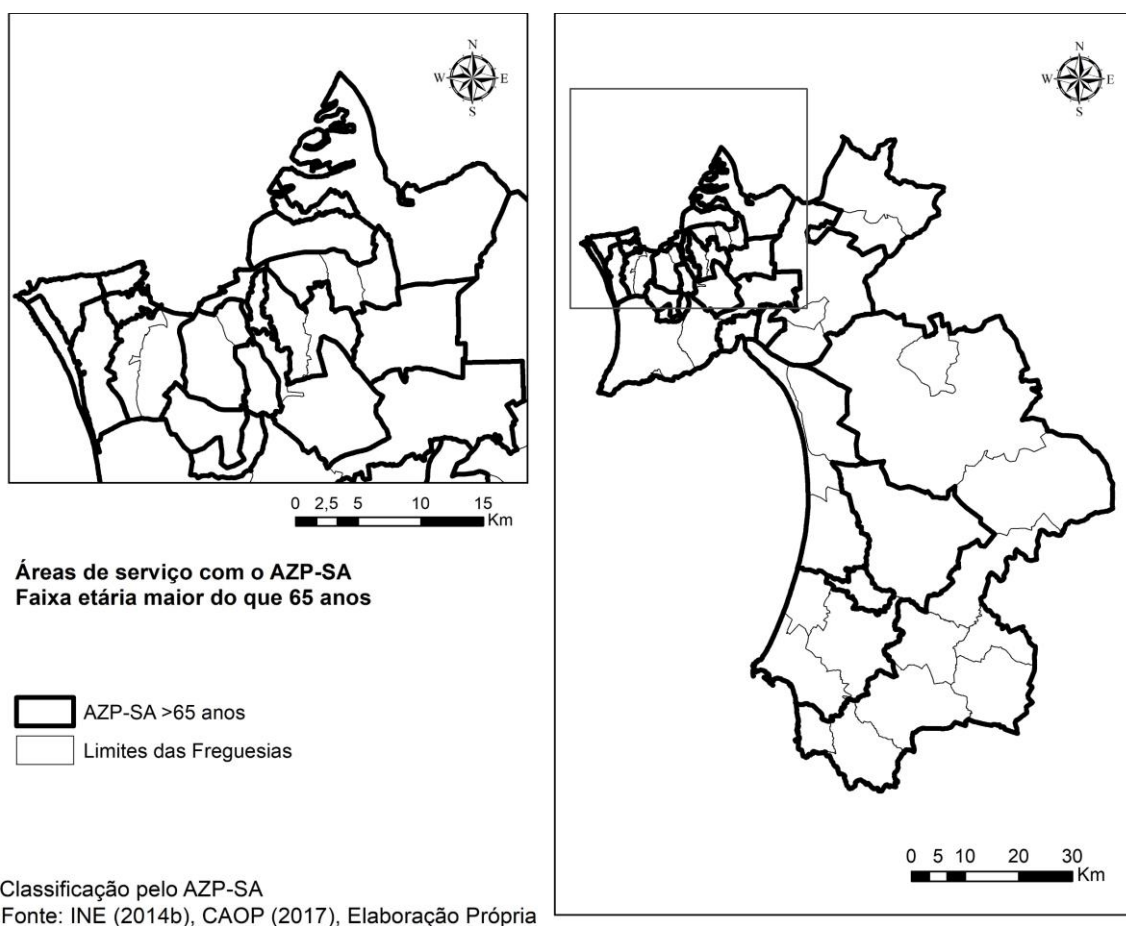


Figura 35 - Áreas de serviço com o AZP-SA para a população com mais de 65 anos.

Pela análise da figura 35, podemos observar que a maior concentração de áreas de serviço se situa na península de Setúbal, tal como aconteceu nas outras simulações. A agregação de freguesias também é mais evidente no sul e no centro do distrito administrativo pelas razões que já explorámos.

5. Resultados e discussão

A área de serviço com maior valor da soma da população apresenta 16154 habitantes, sendo constituída pela União de freguesias do Seixal, Arrentela e Aldeia Paio Pires e a união de freguesias do Alto do Seixalinho, Santo André e Verderena (Tabela 12). A área de serviço com menor valor da soma da população é São Francisco da Serra com 296 residentes.

	População afetada (n.º hab.)	Área (km²)
Média	5599,8	186,8
Mediana	4580,0	36,4
Desvio-padrão	4345,0	455,1
Mínimo	296,0	6,2
Máximo	16154,0	2305,1

Tabela 12 - Estatística descritiva da população afetada por área de serviço com o AZP-SA.

Obtiveram-se valores de compacidade próximos entre os dois algoritmos, no entanto, a soma das diferenças da população é muito menor no AZP-SA com 75691,5 enquanto que o AG obteve 114030,5 (Tabela 13).

	Compacidade	Soma das diferenças da população
Equação IX	1387,3	114030,5
AZP-SA	1372,3	75691,5

Tabela 13 - Comparação entre os resultados obtidos entre o AG e o AZP-SA para as mesmas condições.

Nesta experiência o resultado obtido é muito semelhante com a população residente total por freguesia. Confrontando com os resultados de Bação et al. (2005), chegamos às mesmas conclusões da experiência anterior.

Não obstante, os resultados do AG apresentarem menor desempenho comparativamente com o AZP-SA, a diferença entre as soluções é pouco significativa. As áreas de serviço em média apresentam valores semelhantes o que já não acontece com os valores máximos e mínimos. No caso do AZP-SA o valor máximo das áreas de serviço é sempre mais elevado o que no nosso entendimento não é benéfico para as forças de segurança. Porém as soluções do AG acabam por ser competitivas em relação ao AZP-SA. O incremento das relações de vizinhança de segunda ou mais ordens pode aumentar a eficiência do AG. Por outro lado, a inicialização do algoritmo também necessita de uma melhoria porque está sujeita à combinação com os vizinhos de primeira ordem reduzindo assim a diversidade de soluções iniciais. A inclusão de uma heurística para a inicialização também pode melhorar significativamente o desempenho do AG.

6. Considerações finais

A dinâmica da população e o atual contexto da nossa sociedade forçam a que exista uma necessidade de repensar estrategicamente a forma de atuação quer das forças de segurança, quer dos serviços públicos.

Como podemos observar, as novas linhas de pensamento ligadas à computação evolutiva fornecem-nos ferramentas importantes para planejar em diferentes cenários de alterações na sociedade. Cada vez mais temos acesso a ferramentas mais sofisticadas que nos ajudam a explorar os dados e a obter conclusões mais precisas. A computação evolutiva em consonância com os SIG permitem-nos hoje explorar contextos de difícil análise e resolução como a redistribuição das áreas de serviço das forças de segurança ou outros tipos de problemas de localização-alocação.

Na descrição do problema identificamos os constrangimentos e oportunidades desta linha de investigação que grande parte já tinha sido explorada com o projeto SIM4SECURITY. De facto, as constantes alterações populacionais, nomeadamente o envelhecimento da população e a desertificação do interior trazem-nos novos desafios quer ao nível do planeamento quer ao nível da atuação. A reorganização das forças de segurança é um ponto essencial para a garantia da segurança interna, pois promove uma sociedade mais segura e com maior equilíbrio.

Constatamos com a revisão bibliográfica que grande parte dos estudos sobre este tema debruçam-se sobre a otimização dos circuitos das patrulhas, nas configurações dessas mesmas unidades e poucos são os que se dedicam à localização das instalações e respetivas áreas de serviço. A maioria dos estudos foram realizados nos Estados Unidos à exceção de alguns que decorreram nomeadamente em Espanha e Itália.

Devido à falta de dados desagregados ao nível da freguesia sobre criminalidade optámos por utilizar apenas os dados disponíveis da população residente para propor novas áreas de serviço.

6. Considerações finais

Na análise exploratória dos dados dos censos de 2011 para o distrito administrativo de Setúbal averiguamos a existência de uma grande assimetria no que concerne à distribuição da população residente. A península de Setúbal alberga cerca de 82% da população do distrito administrativo. Aprofundamos mais este facto quando aplicamos a análise *hot spot* à população residente total e por faixas etárias. Os *hot spots*, no geral, situam-se sobretudo na península de Setúbal enquanto que os *cold spot* no sul e centro do distrito administrativo. Esta análise permitiu-nos conhecer em detalhe a distribuição da população residente e verificar que o problema é de difícil resolução com estas configurações.

De modo a responder aos desafios desta dissertação implementou-se um algoritmo genético para propor uma nova localização das esquadras de polícia e respetivas áreas de serviço.

Nos testes efetuados com diferentes números de populações verificamos que o aumento de indivíduos da população tem um grande impacto nos valores das funções de avaliação. Os valores elevados obtidos com um menor valor de indivíduos provavelmente devem-se à convergência para ótimos locais.

Na definição de áreas de serviço com a equação IX foi possível estabelecer 28 áreas de serviço e 28 potenciais localizações baseadas nos centróides das freguesias. Porém, a soma da população residente por área de serviço apresenta uma grande amplitude entre si. O mesmo aconteceu com a aplicação da equação XI onde obtivemos resultados semelhantes, mas com uma ligeira diferença na compacidade e na soma das diferenças da população. O Algoritmo AZP-SA aplicado nas mesmas condições apresentou um desempenho mais eficiente que o AG, visto que os valores de compacidade e da soma das diferenças da população foram mais baixos. No entanto, no que respeita às áreas ocupadas por cada área de serviço, foi o AZP-SA a atingir os maiores valores.

No que concerne à aplicação do AG e do AZP-SA na população com uma faixa etária superior a 65 anos os resultados foram semelhantes aos alcançados com a população residente. O AZP-SA permitiu-nos efetuar comparações de áreas de serviço para o distrito administrativo de Setúbal com o AG.

O AG construído apresenta algumas limitações que se devem às restrições estabelecidas logo na estratégia de codificação das soluções iniciais. Na matriz de vizinhança construída recorrendo ao *PySal*, apenas definimos as relações de vizinhança de primeira ordem o que limita fortemente o espaço de procura do algoritmo. Esta opção de limitar o espaço de procura garante que as áreas de serviço sejam contíguas, por

6. Considerações finais

outro lado, limita a variabilidade das soluções. Neste caso estas restrições tão limitativas promovem que as soluções caiam facilmente em ótimos locais. Outro fato que deve ser sublinhado foi a utilização de apenas uma da taxa de mutação (1%) e a não utilização de outros operadores genéticos para comparação. As experiências com diferentes percentagens de taxas de mutação permitir-nos-ia obter uma maior diversidade, assim como a utilização de outros operadores genéticos.

A ideia de criar áreas de serviço através de *districting* permitiu-nos definir as áreas de serviço. Após esta fase, adicionando dados desagregados sobre o crime, o número de efetivos por esquadra, o número de viaturas disponíveis para cada área de serviço, poderemos aplicar outros métodos de otimização para quantificar e melhorar as atuais áreas.

Espera-se que com esta abordagem, que já tinha sido aplicada por Bação et al. (2005) e Castelli et al. (2015), para problemas de *districting* eleitoral, seja a porta de embarque para potenciar o interesse das forças de segurança na aplicação deste tipo de algoritmos no teatro real das operações.

Como trabalho futuro pretende-se melhorar a atual estratégia de codificação e explorar outras. Testar várias taxas de mutação e operadores genéticos para compará-los entre si. A utilização de outras equações de *districting* e outros algoritmos é um objetivo futuro visto que nos permitirá comparar diferentes soluções. A programação genética é um dos caminhos pelos quais temos interesse em explorar dado os excelentes resultados que Castelli et al. (2015), já obtiveram.

7. Referências bibliográficas

- Alp, O., Erkut, E., & Drezner, Z. V. I. (2003). An Efficient Genetic Algorithm for the p -Median Problem, *Annals of Operations Research*, 122, 21–42 doi: 10.1023/A:1026130003508.
- Anselin, L., Murray, T., & Rey, S. (2013). Spatial Analysis. In *The Oxford Handbook of Quantitative Methods*. Oxford University Press.
- Assis, L. S. De, Franca, P. M., & Usberti, F. L. (2014). A redistricting problem applied to meter reading in power distribution networks, *Computers & Operations Research*, 41(1), 65–75. doi: 10.1016/j.cor.2013.08.002.
- Baçaõ, F., Lobo, V., & Painho, M. (2005). Applying genetic algorithms to zone design. *Soft Computing*, 9(5), 341–348, doi: 10.1007/s00500-004-0413-4.
- Baker, J. R., Clayton, E. R., & Moore, L. J. (1989). Redesign of primary response areas for county ambulance services. *European Journal of Operational Research*, 41(1), 23–32, doi:10.1016/0377-2217(89)90035-0.
- Bandeira, M. L., Azevedo, A. B., Gomes, C. S., Mendes, M. F., Baptista, M. I., & Cabral, M. V. (2014). *Dinâmicas demográficas e envelhecimento da população portuguesa - 1950-2011 Evolução e Perspectivas*. Lisboa: Fundação Francisco Manuel dos Santos.
- Beasley, D., Bull, D. R., & Martin, R. R. (1993). An overview of genetic algorithms : Part 1, fundamentals. *University Computing*, 2(15), 1–16, doi: 10.1017/CBO9781107415324.004.
- Bergey, P. K., Ragsdale, C. T., & Hoskote, M. (2003). A Simulated Annealing Genetic Algorithm for the Electrical Power Districting Problem. *Annals of Operations Research*, 121(1), 33–55, doi:10.1023/A:1023347000978.
- Blais, M., Lapierre, S. D., & Laporte, G. (2003). Solving a home-care districting problem in an urban setting. *Journal of the Operational Research Society*, 54(11), 1141–1147, doi:10.1057/palgrave.jors.2601625.
- Bodily, S. (1978). Police Sector Design Incorporating Preferences of Interest Groups for Equality and Efficiency. *Management Science*, 24(12), 1301–1313.

7. Referências bibliográficas

- Bozkaya, B., Erkut, E., & Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research*, 144(1), 12–26, doi:10.1016/S0377-2217(01)00380-0.
- Bravo, J., Rodrigues, T., Ribeiro, S., & Inácio, A. (2018). Projeções da População Residente 2011-2040. In *Modelos Preditivos & Segurança Pública*, Porto:Fronteira do Caos, pp. 169–208.
- Bucarey, V., Ordóñez, F., & Bassaletti, E. (2015). Shape and Balance in Police Districting. In *Applications of Location Analysis*, Santiago, Chile: Springer International Publishing Switzerland, 329-347 doi:10.1007/978-3-319-20282-2_14.
- Cabral, P., Ribeiro, S., Pereira, J., & Painho, M. (2018). Análise espacial avançada no contexto da segurança interna. In *Modelos Preditivos & Segurança Pública*, Porto: Fronteira do Caos, pp. 231–251.
- Camacho-Collados, M., Liberatore, F., & Angulo, J. M. (2015). A multi-criteria Police Districting Problem for the efficient and effective design of patrol sector. *European Journal of Operational Research*, 246(2), 674–684. doi:10.1016/j.ejor.2015.05.023.
- CAOP (2013). Carta Administrativa Oficial de Portugal Continental. Direção-Geral do Território, (URL: http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop_/caop__download_/, consultado em outubro de 2018).
- CAOP (2017). Carta Administrativa Oficial de Portugal Continental. Direção-Geral do Território, (URL: http://www.dgterritorio.pt/cartografia_e_geodesia/cartografia/carta_administrativa_oficial_de_portugal_caop_/caop__download_/, consultado em outubro de 2018).
- Castelli, M., Vanneschi, L., Henriques, R., (2015). A geometric semantic genetic programming system for the electoral redistricting problem. *Neurocomputing.*, 154, pp.200-207.
- Chinneck, J. (2000). *Practical optimization: A gentle introduction*. Ottawa: Carleton University.
- Chen, H., Tao Cheng & Xinyue Ye (2018): Designing efficient and balanced police patrol districts on an urban street network, *International Journal of Geographical Information Science*, doi: 10.1080/13658816.2018.1525493.
- Chow, A., Cheung, Y., Yoon, T. (2015). Optimization of Police Facility Locationing. *Transportation Research Record: Journal of the Transportation Research Board*, 2528,60-68, doi: 10.3141/2528-07.

7. Referências bibliográficas

- Costa, E. (2015). *Programação em Python - Fundamentos e Resolução de problemas*. Lisboa: FCA- Editora de Informática.
- Costa, & Simões. (2008). *Inteligência Artificial - Fundamentos e Aplicações* (2.^a edição). Lisboa: FCA - Editora de Informática.
- Curtin, K. M., Hayslett-McCall, K., & Qiu, F. (2010). Determining optimal police patrol areas with maximal covering and backup covering location models. *Networks and Spatial Economics*, 10(1), 125–145. doi:10.1007/s11067-007-9035-6.
- Curtin, K. M., Qui, F., Hayslett-McCall, K., & Bray, T. M. (2005). Integrating GIS and Maximal Covering Models to Determine Optimal Police Patrol Areas. In F. Wang (Ed.), *Geographic Information Systems and Crime Analysis*, Hershey, PA: IGI Global. pp. 214-235, doi:10.4018/978-1-59140-453-8.ch013.
- DGPJ/MJ (2017). Crimes registados pelas autoridades policiais - 2016. (URL:http://www.dgpj.mj.pt/sections/noticias/crimes-registados-pelas_2, consultado em outubro de 2018).
- D'Amico, S., Wang, S.-J., Batta, R., & Rump, C. (2002). A Simulated Annealing Approach to police district design. *Computers & Operations Research*, 29 (1999), 55–65.
- Duque, J. C., Dev, B., Betancourt, A., & Franco, J. L. (2011). ClusterPy: Library of spatially constrained clustering algorithms, Versão 0.9.9., Colombia. (URL: <http://www.rise-group.org> consultado em outubro de 2018).
- ESRI. (2009). How Hot Spot Analysis: Getis-Ord Gi* (Spatial Statistics) works. (URL: [http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How Hot Spot Analysis: Getis-Ord Gi* \(Spatial Statistics\) works](http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How Hot Spot Analysis: Getis-Ord Gi* (Spatial Statistics) works), consultado em outubro de 2018).
- Ferland, J. A., & Guénette, G. (1990). Decision Support System for the School Districting Problem. *Operations Research*, 38(1), 15–21.
- Garfinkel, R. S., & Nemhauser, G. L. (1970). Optimal Political Districting by Implicit Enumeration Techniques. *Management Science*, 16(8), B495--B508.
- Goderbauer, S., & Winandy, J. (2017). Political Districting Problem: Literature Review and Discussion with regard to Federal Elections in Germany. (URL:http://www.or.rwth-aachen.de/research/publications/LitSurvey_PoliticalDistricting__Goderbauer_Winandy_20171123.pdf, consultado em outubro de 2018).
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning* (First). Addison-Wesley Publishing Company.
- Haupt,R. & Haupt, S. (2004). *Practical Genetic Algorithms* (2^a edição). Wiley-Interscience.

7. Referências bibliográficas

- A Jonh wiley & Sons, Inc.
- Hanafi, S., Freville, A., & Vaca, P. (1999). Municipal Solid Waste Collection: An Effective Data Structure For Solving The Sectorization Problem With Local Search Methods. *INFOR: Information Systems and Operational Research*, 37(3), 236–254. doi:10.1080/03155986.1999.11732383.
- Henriques, R., Ribeiro, S., & Castelli, M. (2018). Modelo de otimização. In *Modelos Preditivos & Segurança Pública*, Porto: Fronteira do Caos, pp. 291–302.
- Hess, S. W., Weaver, J. B., Siegfeldt, H. J., Whelan, J. N., & Zitlau, P. A. (1965). Nonpartisan Political Redistricting by Computer. *Oper. Res.*, 13(6), 998–1006. doi:10.1287/opre.13.6.998.
- Hojati, M. (1996). Optimal Political Districting. *Computers Ops Res.*, 23(12), 1147–1161.
- Hojage, C. & Goodchild, M. (1986). Discrete space location-allocation solutions from genetic algoritms. *Annals of Operations Research*, 2, 35-46.
- INE (2014). Censos em Portugal de 1864 a 2011. (URL: http://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_historia_pt, consultado em outubro de 2018).
- INE (2014b). Censos 2011 - População residente por freguesia, CAOP 2013. (URL: https://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos_quadros, consultado em outubro de 2018).
- INE (2017). Estimativas da População Residente em Portugal-2016. (URL: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaquas&DESTAQUESdest_boui=315156710&DESTAQUESmodo=2, consultado em outubro de 2018).
- INE (2018). Estimativas da População Residente em Portugal-2017. (URL: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_destaquas&DESTAQUESdest_boui=315156710&DESTAQUESmodo=2, consultado em outubro de 2018).
- Kalcsics, J. (2015). Districting Problems. In *Location Science*, Karlsruhe: Springer International Publishing Switzerland. doi:10.1007/978-3-319-13111-5_23. pp. 595–622.
- Kalcsics, J., Nickel, S., & Schröder, M. (2005). Towards a unified territorial design approach - applications, algorithms and GIS integration. *Top*, 13(1), 1–56, doi:10.1007/bf02578982.
- Kim, K., Chun, Y., & Kim, H. (2018). A Robust Heuristic Approach for Regionalization Problems BT - GeoComputational Analysis and Modeling of Regional Systems. In J.-C. Thill & S. Dragicevic (Eds.), Cham: Springer International Publishing, 305–324, doi: 10.1007/978-3-319-59511-5_16.

7. Referências bibliográficas

- Kirkpatrick S, Gelatt C D, Vecchi M P, (1983). Optimisation by simulated annealing. *Science*, 220, 671-680.
- Kistler, A. (2009). Tucson Police Officers Redraw Division Boundaries to Balance Their Workload. *Geography & Public Safety*, 1(3), 1–2. URL: [em: https://ric-zai-inc.com/Publications/cops-w0550-pub.pdf](https://ric-zai-inc.com/Publications/cops-w0550-pub.pdf)<http://www.nij.gov/topics/technology/maps/documents/gps-bulletin-v1i3.pdf>, consultado em outubro de 2018.
- Lei n.º 53/2008.D.R. I Série, 167(2008-08-29), 6135 - 6141.
- Lee, Y. S. (2015). School districting and the origins of residential land price inequality. *Journal of Housing Economics*, 28, 1–17, doi:10.1016/j.jhe.2014.12.002.
- Levit, J., & Wood, E. (2010). A citizen's guide to redistricting (New York U) (URL<https://www.brennancenter.org/page/-/CGR%20Reprint%20Single%20Page.pdf>, consultado em outubro de 2018).
- Lin, H.-Y., & Kao, J.-J. (2008). Subregion Districting Analysis for Municipal Solid Waste Collection Privatization. *Journal of the Air & Waste Management Association*, 58(1), 104–111. doi:10.3155/1047-3289.58.1.104.
- Lutz, M. (2003). *Learning Python*, (2.ª edição), Sebastopol, CA, USA: O'Reilly, pp. 621.
- Malik, S., Wadhwa (2014). Preventing premature convergence in genetic algorithm using dgca and elitist technique, *International Journal of Advanced Research in Computer Science and Software Engineering*, 4, 410-418.
- Malczewski, J., & Rinner, C. (2015). *Multicriteria Decision Analysis in Geographic Information Science*. (Springer, Ed.), *Analysis methods*. New York. doi:10.1007/978-3-540-74757-4.
- McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, 184(1), 205–222. doi:10.1016/j.cam.2004.07.034.
- Mitchell, P. (1972). Optimal Selection of Police Patrol Beats. *The Journal of Criminal Law, Criminology and Police Science*, 63(4).
- Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge, Massachusetts: MIT Press Cambridge, ISBN: 0262631857.
- Moreira, M. J. G., & Rodrigues, T. F. (2004). As Regionalidades Demográficas do Portugal Contemporâneo, 1–38. (URL: <http://www.cepese.pt/portal/investigacao/working-papers/populacao-e-prospectiva/as-regionalidades-demograficas-do-portugal-contemporaneo/As-Regionalidades-Demograficas-do-Portugal.pdf>, consultado em outubro de 2018).

7. Referências bibliográficas

- Openshaw, S., & Rao, L. (1995). Algorithms for reengineering 1991 Census geography, 27(April 1994), 425–446.
- Painho, M., & Ribeiro, S. (2018). O projeto SIM4SECURITY. Um exemplo de aplicação de tecnologias diferenciadas em sede de políticas públicas. In *Modelos Preditivos & Segurança Pública*, Porto: Fronteira do Caos, pp. 157–167.
- Pereira. (2006). Partição Multicritério de um Território em Zonas: Modelos, Algoritmos e Aplicações, Tese de Doutoramento. Universidade de Coimbra. Faculdade de Economia. Coimbra.
- Pereira, J. (2017). *O Mapeamento Dasimétrico enquanto Ferramenta de Apoio à Decisão: Distribuição da População e Caracterização da Área de Intervenção da PSP e da GNR*. Tese de Mestrado. NOVAIMS. Lisboa.
- PORDATA (2018). Base de dados de Portugal contemporâneo. URL: <https://www.pordata.pt/>, consultado em outubro de 2018.
- Reeves, C. R. (1997). Feature Article — Genetic Algorithms for the Operations Researcher. *INFORMS Journal on Computing*.
- Ribeiro, S., Cabral, P., & Bravo, J. (2017). O impacto das tendências demográficas nas políticas de segurança pública. In *International Conference Risks, Security and Citizenship*. Setúbal. (URL: <http://sim4security.novaims.unl.pt/>, consultado em outubro de 2018).
- Ribeiro, S., & Henriques, R. (2018). Aplicação dos Self-organizing Maps na análise da criminalidade em Portugal, 2011,2016. In *Modelos Preditivos & Segurança Pública* Porto: Fronteira do Caos, pp. 253–280.
- Ricca, F., & Simeone, B. (2008). Local search algorithms for political districting, 189, 1409–1426, doi:10.1016/j.ejor.2006.08.065.
- Ríos-Mercado, R. Z., & López-Pérez, J. F. (2012). Commercial territory design planning with realignment and disjoint assignment requirements. *Omega (United Kingdom)*, 41(3), 525–535. doi:10.1016/j.omega.2012.08.002.
- Rodrigues. (2008). *História da População Portuguesa. Das longas permanências à conquista da modernidade*. Porto: Afrontamento, CEPESSE.
- Rodrigues, & Ferreira. (2014). Portugal e a globalização das migrações. Desafios de segurança. *População e Sociedade*, 22, 137–155.
- Salazar-Aguilar, M. A., Ríos-Mercado, R. Z., & González-Velarde, J. L. (2011). A bi-objective programming model for designing compact and balanced territories in commercial districting. *Transportation Research Part C: Emerging Technologies*, 19(5), 885–895, doi:10.1016/j.trc.2010.09.011.

7. Referências bibliográficas

- Sarac, A., & Batta, R. (1999). Reconfiguring Police Reporting Districts In The City of Buffalo. *O.R. Insight*, 12(3), 16–24.
- Sarker, R., & Newton, C. (2008). *Optimization Modelling- A practical Approach* (Taylor & F). New York: CRC Press.
- Simões, A. (1999). *Transposição: Estudo de um Novo Operador Genético Inspirado Biologicamente*. Tese de mestrado. Universidade de Coimbra. Coimbra.
- Sivanandam, S. N., & Deepa, S. N. (2008). *Introduction to Genetic Algorithms* (Springer S). Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-73189-4.
- Talbi, E. G. (2009). *Metaheuristics: From Design to Implementation*. (Wiley, Ed.), *Metaheuristics: From Design to Implementation*. New Jersey: A JOHN WILEY & SONS, INC, PUBLICATION. doi:10.1002/9780470496916.
- Tong, D., & Murray, A. T. (2012). Spatial Optimization in Geography. *Annals of the American Association of Geographers*, 102(6), 1290–1309. doi:10.1080/00045608.2012.685044.
- Tutorialspoint (2016). Genetic Algorithms. (URL: https://www.tutorialspoint.com/genetic_algorithms/index.htm, consultado em outubro de 2018).
- Vanneschi, L., Henriques, R., & Castelli, M. (2017). Multi-objective genetic algorithm with variable neighbourhood search for the electoral redistricting problem. *Swarm and Evolutionary Computation*, 36(February), 37–51. doi:10.1016/j.swevo.2017.04.003.
- Vasconcelos, & Barão. (2017). *Ciência dos dados nas organizações*, (1.^a edição), (Lisboa: Lidel, Ed).
- Wenwen Li , Michael F. Goodchild & Richard Church (2013): An efficient measure of compactness for two-dimensional shapes and its application in regionalization problems, *International Journal of Geographical Information Science*, doi:10.1080/13658816.2012.752093.
- Zhang, Y., Brown, D. (2013). *Police patrol districting method and simulation evaluation using agent-based model & GIS*. *Security Informatics*, 2(7),pp.1-13, doi:10.1186/2190-8532-2-7.
- Zhang, Y. (2014). *Stochastic Simulation Optimization for Districting Problems with Application to Police Patrol District Design*. Doctoral Thesis. Faculty of School Engineering and Applied Science. University of Virginia.
- Zoltners, A. A., & Sinha, P. (1983). Sales Territory Alignment: A Review and Model. *Management Science*, 29(11), 1237–1256. doi:10.1287/mnsc.29.11.1237.

8. Anexos

8.1. Código da estatística descritiva

```
# -*- coding: utf-8 -*-
"""
NOVA Information Management School
MCSIG - Tese de Mestrado

Estatística descritiva
Autor: Andre Duarte
Nome do ficheiro: estatistica_descritiva.py
"""
#importacao das bibliotecas com alias
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

##importacao e leitura do ficheiro csv com dados da populacao
pop_r = pd.read_csv('centroids.csv')

##estatistica descritiva
table = pop_r.describe()
table
##histogramas, caixa de bigodes, violin e swarm plot
#_sns.violinplot(x='POP_R', data = pop_r, inner=None)
#_sns.boxplot(x='POP_R', data = pop_r, whis=np.inf)
_ = sns.swarmplot(x='POP_R', data = pop_r, color="red", size=5)
_ = plt.hist(pop_r['POP_R'], bins=20)
_ = plt.xlabel('Numero de Residentes')
_ = plt.ylabel('Numero de Freguesias')
plt.show()
```

8.2. Código do algoritmo genético

```
# -*- coding: utf-8 -*-
"""
NOVA Information Management School
MCSIG - Tese de Mestrado

Algoritmo genetico
Autor: Andre Duarte
Nome do ficheiro: ga_final_entrega.py
"""

import pandas as pd
import numpy as np
import random
import math
import copy
import pysal as ps
import sys
import matplotlib as plt

####Variaveis de entrada####

fich_ori = pd.read_csv('centroides.csv')
#fich_ori

nodes = fich_ori["COD_GA"].tolist() #centroides dos poligonos
#nodes

pop_resi = dict(zip(fich_ori.COD_GA, fich_ori.POP_R)) #constroi
dicionario centroe (node) e populacao
#pop_resi

##Vizinhos de 1 e 2 ordem com o pysal
w = ps.queen_from_shapefile(('SETUBAL_TUDO.shp'), "COD_GA")# determina
os vizinhos do tipo queen
Wmatrix, idx = w.full()
#Wmatrix

W_df = pd.DataFrame(Wmatrix, columns = nodes, index = nodes)#constroi
matrix binaria 0 e 1 (u-matrix)
#W_df

d = {}
for k, subdict in W_df.to_dict().items(): #procura a vizinhanca na
matrix 0 e 1
    neighbours = []
    for k2,value in subdict.items():
        if value:
            neighbours.append(k2)
    d[k] = neighbours

'''
def find_second_order(d):
    d2 = {}
```

8. Anexos

```
    for k, neighbours in d.items():
        new_neighbours = set(neighbours) # Procura vizinhos e vizinhos
dos vizinhos
        for neighbour in neighbours:
            new_neighbours = (new_neighbours|set(d[neighbour]))-{k}
            d2[k]=list(new_neighbours)
        return d2
#print find_second_order(d)
'''
#Vizinhos = find_second_order(d)

Vizinhos = d

### Coordenadas em x e y dos centroides

X_Coord = dict(zip(fich_ori.COD_GA, fich_ori.X_Coord))
#X_Coord

Y_Coord = dict(zip(fich_ori.COD_GA, fich_ori.Y_Coord))
#Y_Coord

#### Variaveis ou parametros globais do Algoritmo genetico ####

gPopulation = 50
gNumberOfGenerations = 500
gNumberOfAreas = 6
gMutationRate= 100 #
gKeepParents = True

##Gerar cromossomas identico ao p-medianas
def generateRandomCromossome():

    heads={}
    while(len(heads)<gNumberOfAreas):
        index = random.randint(0,24)
        if( nodes[index] not in heads):
            heads[nodes[index]]=Vizinhos[nodes[index]][:]

    return correctCromossome(heads)

##Cromossoma feito com as chaves
def getCromossomeList(aCromossome):
    cromossomeValues = aCromossome.keys()
    for item in aCromossome:
        for v in aCromossome[item]:
            cromossomeValues.append(v)
    return cromossomeValues

##Corrige e elimina os cromossomas invalidos tendo em conta a chave
def correctCromossome(aCromossome):

    rCromossome = aCromossome.copy()
    cromossomeKeys = rCromossome.keys()
    for item in rCromossome:
        for v in rCromossome[item][:]:
            if(v in rCromossome):
                rCromossome[item].remove(v)
            else:
```

8. Anexos

```
        for key in cromossomeKeys:
            if(key != item and v in rCromossome[key]):
                rCromossome[item].remove(v)
                break
cromossomeValues = getCromossomeList(rCromossome)

for node in nodes:
    if( node not in cromossomeValues):
        for item in rCromossome:
            if( node in Vizinhos[item]):
                rCromossome[item].append(node)
return rCromossome

##Crossover de um ponto de corte
def crossoverCromossomes(aCromossome1, aCromossome2):

    c1A = dict(aCromossome1.items()[len(aCromossome1)/2:])
    c1B = dict(aCromossome1.items()[:len(aCromossome1)/2])

    c2A = dict(aCromossome2.items()[len(aCromossome2)/2:])
    c2B = dict(aCromossome2.items()[:len(aCromossome2)/2])

    #descendent1
    r1 = copy.deepcopy(c1A)
    r1.update(copy.deepcopy(c2B))
    #descendent2
    r2 = copy.deepcopy(c2A)
    r2.update(copy.deepcopy(c1B))
    r1x = correctCromossome(r1)
    r2x = correctCromossome(r2)
    return [r1x,r2x]

##Mutacao com troca de posicao de um elemento chave ou valor
def mutation(aCromossome):

    rCromossome = copy.deepcopy(aCromossome)
    allList = getCromossomeList(aCromossome)
    fromMutate = random.randint(0, len(nodes)-1)
    toMutate = random.randint(0, len(nodes)-1)
    if ( nodes[fromMutate] in rCromossome):
        #change the key
        rCromossome[nodes[toMutate]] =
rCromossome.pop(nodes[fromMutate])
    else:
        for key in rCromossome:
            if( nodes[fromMutate] in rCromossome[key]):

                rCromossome[key].append(nodes[toMutate])
    return correctCromossome(rCromossome)

##Funcao de avaliacao dos cromossomas (prof. Bacao)
def fitnessFunc(aCromossome):

    if( len(getCromossomeList(aCromossome)) != len(nodes)):
        return 50000000
    if( len(aCromossome) != gNumberOfAreas ):
        return 50000000
```

8. Anexos

```
zonesCount = {}
zonesDist = {}
for zone in aCromossome:

    zonesCount[zone] = pop_resi[zone]
    zonesDist[zone] = 0
    for area in aCromossome[zone]:
        zonesCount[zone] += pop_resi[zone]
        zonesDist[zone] += math.sqrt(((X_Coord[zone]-
X_Coord[area])**2) + (Y_Coord[zone]-Y_Coord[area])**2)
    #print zonesCount
    #print zonesDist
    #print'-----'
-----'

avg = sum(zonesCount.values())/len(zonesCount)
retVal = 0
for zone in zonesCount:
    retVal += math.fabs((zonesCount[zone]-avg) + zonesDist[zone])

return retVal

def printCromossome(aCromossome):
    cromossomeValues = getCromossomeList(aCromossome)
    return "CROMOSSOME (LEN="+str(len(cromossomeValues))+")" +
str(aCromossome) + " | Fitness = " + str(fitnessFunc(aCromossome))

def singleCromossomesTests():
    c1 = generateRandomCromossome()
    c2 = generateRandomCromossome()
    r = crossoverCromossomes(c1,c2)
    print(printCromossome(r[0]))
    rx = mutation(r[0])
    print(printCromossome(rx))

def takeSecond(elem):
    return elem[1]

def main(args):

    print("Initial population creation")
    population = []
    for i in range(gPopulation):
        population.append([generateRandomCromossome(),0])

    bestCromossome = []
    bestFitness = 5000000
    avgPopulation = 0

    for generation in range(gNumberOfGenerations):
        avgPopulation = 0 #inicializa a avg
        bestFitnessGeneration = 5000000
        for i in range(gPopulation):

            population[i][1] = fitnessFunc(population[i][0])
            avgPopulation += population[i][1]
            if( population[i][1] < bestFitnessGeneration):
```


8. Anexos

```
        bestFitnessGeneration = population[i][1]
        if( population[i][1] < bestFitness):
            print "Find better fitness: "+ str(population[i][1]) +
" than " + str(bestFitness)
            bestFitness = population[i][1]
            bestCromossome = copy.deepcopy(population[i][0])
            avgPopulation /= gPopulation

    population.sort(key=takeSecond)
    newPopulation = []

    for i in range(0,gPopulation/2,2):
        offSprings = crossoverCromossomes(population[i][0],
population[i+1][0])
        newPopulation.append([offSprings[0], 0])
        if( (i%gMutationRate) == 0):
            newPopulation.append([mutation(offSprings[1]), 0])
        else:
            newPopulation.append([offSprings[1], 0])

        newPopulation.append([copy.deepcopy(population[i][0]), 0])
        newPopulation.append([copy.deepcopy(population[i+1][0]),
0])

    population = newPopulation[:]
    print "Geracao:" + str(generation)+ " | Melhor individual: " +
printCromossome(bestCromossome)
    print ("Geracao:" + str(generation)+ " | Populacao
Avg:"+str(avgPopulation)+ " Melhor
geracao:"+str(bestFitnessGeneration)+" | Melhor ate agora:" +
str(bestFitness))
    print ("Fitness:"+str(fitnessFunc))
    print
(str(generation)+"", "+str(avgPopulation)+'', '+str(bestFitnessGeneration)
+', '+ str(bestCromossome))

    return 0

if __name__ == '__main__':
    import sys
    sys.exit(main(sys.argv))

f = open('pop_50_06_500_5.txt', 'w')
sys.stdout = f
#print >>f, "Geracao:" + str(generation)+", "+"Populacao
Avg:"+str(avgPopulation)+'', '+ "Melhor
geracao:"+str(bestFitnessGeneration)+", "+"Melhor individual:" +
str(bestCromossome)
f.close()
```

8.3. Código do AZP-SA

```
# -*- coding: utf-8 -*-
"""
NOVA Information Management School
MCSIG - Tese de Mestrado

Algoritmo AZP-SA
Autor: Andre Duarte
Nome do ficheiro: AZP_SA.py
"""

import clusterpy
calif =
clusterpy.importArcData("D:\\AD_SIM4SECURITY\\genetico\\azp_shape")
calif.fieldNames
dataOperations = {'D_N_65':['mean', 'sum']}
calif.exportArcData("D:\\AD_SIM4SECURITY\\genetico\\azpSA1_input")
calif.cluster('azpSA', ['POP_R'], 28, std= 1,dissolve= 1,
dataOperations=dataOperations)
calif.results[0].exportArcData("D:\\AD_SIM4SECURITY\\genetico\\azpSA_6
5")
```

8. Anexos

8.4. Soluções obtidas com diferentes parametrizações com o AG

Corrida/População	Equação IX			Equação XI			Equação XI- >65 anos
	10	50	100	10	50	100	100
1	735079	683107	521625	1081303	776409,7	831502,1	108023
2	735266	643904	596025	1115939	839766,6	858314,2	91797
3	1115103	602039	614509	995223,3	860078,2	818657,5	112678
4	797094	696250	619295	1113017	905623,9	883532,7	107037
5	921227	667863	630318	932326,7	878291,4	813939,8	112486
6	735968	674909	531957	1021655	811855,1	920737,4	116527
7	1021856	677952	603512	1150497	919315,9	790673,9	106038
8	669920	658041	679568	989850,1	1099561	904755,4	109916
9	745421	709990	680010	1013289	1075531	823979,8	98730
10	773574	629808	643633	859612,4	920525,3	926045,1	83610
11	728915	695971	640183	1141612	799243,9	845632,8	111970
12	721744	690029	663344	916123,6	887062,2	816015,9	105693
13	742538	684977	751331	1319551	835157,9	837242,7	93598
14	923789	720309	668100	1018214	1053171	840607,4	116270
15	842901	776293	527717	1208272	890258,5	874287,2	87165
16	1002109	636615	614909	1182645	962125,4	789904,8	94562
17	742112	705119	554233	965118,9	843090,1	876016,2	103562
18	898867	694997	642768	803367	836958,6	890202,9	97618
19	922471	753546	635745	930933,9	784686,9	842197	111480
20	865058	716878	644303	903214,8	1003644	863986	133179
21	908298	771510	670344	1105660	1055344	957645,9	106640
22	884746	732109	668733	1142299	1034307	791332,7	107864
23	804858	716679	685153	1139687	892854,8	905832,5	96062
24	749165	661535	570779	1020022	981719,7	836304,7	91458
25	748369	707804	581984	1111033	975451,7	852626,8	111582

Tabela 14 – Todas as soluções obtidas com diferentes parametrizações.

8. Anexos

8.5. Identificação por freguesia das áreas de serviço determinadas

Nome da Freguesia simplificado	Áreas de Serviço - Identificação por freguesia				
	Equação IX	Equação XI	Equação XI >65 anos	AZP-SA	AZP-SA >65 anos
Comporta	14	16	17	8	27
Carvalhal	13	15	15	8	27
Sado	15	16	17	13	12
São Martinho	16	17	16	18	20
Santa Maria do Castelo e Santiago e Santa Susana	16	17	17	15	20
S.Julião, N.S. da Anunciada e S.Maria da Graça	23	23	15	26	3
Setúbal (São Sebastião)	18	16	19	7	21
Alcochete	28	13	18	13	22
Gâmbia-Pontes-Alto da Guerra	19	16	17	13	12
Quinta do Conde	6	21	21	6	2
Fernão Ferro	21	14	21	19	4
Canha	11	4	14	24	14
Palmela	20	23	28	14	16
Palhais e Coia	6	14	23	5	1
São Domingos e Vale de Água	22	9	12	23	20
Quinta do Anjo	25	25	23	21	7
Santo António da Charneca	24	24	23	21	7
Seixal, Arrentela e Aldeia de Paio Pires	6	14	21	2	23
Corroios	26	22	24	10	17
Amora	21	14	25	10	17
Alto do Seixalinho, Santo André e Verderena	6	24	23	22	23
Charneca de Caparica e Sobreda	27	26	26	11	5
Costa da Caparica	27	26	26	19	9
Laranjeiro e Feijó	3	3	25	17	25
Moita	28	27	27	21	18
Pinhal Novo	28	28	28	28	11
Caparica e Trafaria	3	3	2	9	28
Alvalade	4	9	7	23	20
Baixa da Banheira e Vale da Amoreira	5	24	3	12	8
Alhos Vedros	25	1	23	21	18
Almada, Cova da Piedade, Pragal e Cacilhas	1	3	2	25	10
Barreiro e Lavradio	2	11	1	3	15
Gaio-Rosário e Sarilhos Pequenos	2	2	27	13	13
Sarilhos Grandes	28	28	6	13	13
Ermidas-Sado	9	9	7	23	20
Atalaia e Alto-Estanqueiro-Jardia	28	8	18	13	13
Abela	7	9	7	23	20
Pegões	11	4	4	24	14
Santiago do Cacém, S.Cruz e S.Bartolomeu da Serra	12	6	5	15	24
Santo André	17	5	9	15	24
São Francisco da Serra	8	7	9	8	6
Montijo e Afonsoeiro	10	11	10	4	26
Azinheira Barros e São Mamede do Sádio	9	10	11	23	20
Poceirão e Marateca	11	13	14	13	12
São Francisco	10	13	8	20	19
Melides	8	10	9	8	27
Grândola e Santa Margarida da Serra	8	10	11	15	24
Torrão	16	10	11	1	20
Cercal	12	20	12	23	20
Porto Covo	12	20	20	23	20
Sesimbra (Santiago)	26	19	13	16	16
Sines	17	18	12	27	24
Azeitão (São Lourenço e São Simão)	23	23	21	14	16
Sesimbra (Castelo)	26	22	22	19	16
Samouco	10	12	18	20	19

Tabela 15 - Identificação por freguesia das áreas de serviço criadas.

8. Anexos

8.6. Média e soma da população residente por área de serviço

Equação IX		
Área de Serviço	Media da população	Soma da população
1	49661	49661
2	12127	24254
3	33011	66022
4	2098	2098
5	30949	30949
6	28969	115877
7	890	890
8	4434	13301
9	1362	2724
10	14147	42441
11	4696	14087
12	4285	12854
13	1630	1630
14	1268	1268
15	5783	5783
16	3926	11778
17	11924	23847
18	52542	52542
19	5885	5885
20	17481	17481
21	32844	65688
22	1469	1469
23	28488	56975
24	11536	11536
25	13458	26915
26	23852	71555
27	29174	58347
28	12680	63401

Tabela 16 - Média e soma da população residente por área de serviço com a equação IX.

8. Anexos

Equação XI		
Área de Serviço	Média da População	Soma da população
1	15050	15050
2	2377	2377
3	38561	115683
4	2801	5602
5	10647	10647
6	8454	8454
7	809	809
8	5085	5085
9	1619	6477
10	3873	15491
11	29494	58988
12	3143	3143
13	7637	22911
14	28550	114199
15	1630	1630
16	16370	65478
17	4742	9483
18	13200	13200
19	4841	4841
20	2200	4400
21	25606	25606
22	33357	66714
23	24819	74456
24	28082	84245
25	11865	11865
26	29174	58347
27	17653	17653
28	14212	28424

Tabela 17 - Média e soma da população residente por área de serviço com a equação XI.

8. Anexos

AZP-SA – População Residente		
Área de Serviço	Média da População	Soma da População
1	2295	2295
2	44920	44920
3	21877	21877
4	37111	37111
5	3591	3591
6	25606	25606
7	52542	52542
8	1341	5365
9	26150	26150
10	48145	96290
11	44929	44929
12	30949	30949
13	6182	43278
14	18179	36358
15	9742	38968
16	4841	4841
17	39872	39872
18	450	450
19	16510	49530
20	2665	5330
21	14026	56104
22	41760	41760
23	1654	11581
24	2801	5602
25	49661	49661
26	38098	38098
27	13200	13200
28	13200	13200

Tabela 18 - Média e soma da população residente por área de serviço com o AZP-SA.

8. Anexos

Equação XI - >65 anos		
Área de Serviço	Média da População	Soma da População
1	4682	4682
2	9246	18493
3	5320	5320
4	813	813
5	2164	2164
6	757	757
7	489	1468
8	299	299
9	833	2501
10	5568	5568
11	1230	3691
12	1234	3702
13	1550	1550
14	1121	2243
15	4549	9098
16	161	161
17	1080	4321
18	1053	3159
19	8056	8056
20	255	255
21	3924	15696
22	3204	3204
23	3364	16820
24	6924	6924
25	7606	15213
26	4963	9926
27	1761	3522
28	3593	7187

Tabela 19 - Média e soma da população residente por área de serviço com a equação XI >65 anos.

8. Anexos

AZP-SA – População com mais de 65 anos		
Área de Serviço	Média da População	Soma da População
1	644	644
2	2997	2997
3	8857	8857
4	3227	3227
5	7307	7307
6	296	296
7	3958	1979
8	5320	5320
9	2619	2619
10	14015	14015
11	3686	3686
12	3628	1209
13	2143	714
14	1324	662
15	4682	4682
16	11352	2838
17	14831	7416
18	5495	2748
19	778	389
20	6399	640
21	8056	8056
22	1760	1760
23	16154	8077
24	8871	2218
25	7306	7306
26	5568	5568
27	1042	347
28	4478	4478

Tabela 20 - Média e soma da população com mais de 65 anos por área de serviço com o AZP-SA.



C& SIG



UNIGIS PT

