

UNIVERSIDADE FEDERAL FLUMINENSE

AUGUSTO CÉSAR FADEL

ABORDAGEM DE OTIMIZAÇÃO PARA
ALOCÇÃO DE *PRIVACY BUDGET* NO
PROBLEMA DE AGRUPAMENTO
DIFERENCIALMENTE PRIVADO

NITERÓI

2026

UNIVERSIDADE FEDERAL FLUMINENSE

AUGUSTO CÉSAR FADEL

ABORDAGEM DE OTIMIZAÇÃO PARA
ALOCÇÃO DE *PRIVACY BUDGET* NO
PROBLEMA DE AGRUPAMENTO
DIFERENCIALMENTE PRIVADO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Doutor em Computação. Área de concentração: Ciência da Computação.

Orientador:

LUIZ SATORU OCHI

Co-orientador:

JOSÉ ANDRÉ DE MOURA BRITO

NITERÓI

2026

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

F144a Fadel, Augusto César
Abordagem de Otimização para Alocação de Privacy Budget
no Problema de Agrupamento Diferencialmente Privado / Augusto
César Fadel. - 2026.
241 f.: il.

Orientador: Luiz Satoru Ochi.
Coorientador: José André De Moura Brito.
Tese (doutorado)-Universidade Federal Fluminense, Instituto
de Computação, Niterói, 2026.

1. Proteção de dados. 2. Análise de agrupamento. 3.
Metaheurística. 4. Otimização (Computação). 5. Produção
intelectual. I. Ochi, Luiz Satoru, orientador. II. Brito,
José André De Moura, coorientador. III. Universidade Federal
Fluminense. Instituto de Computação. IV. Título.

CDD - XXX

AUGUSTO CÉSAR FADEL

ABORDAGEM DE OTIMIZAÇÃO PARA ALOCAÇÃO DE *PRIVACY BUDGET* NO
PROBLEMA DE AGRUPAMENTO DIFERENCIALMENTE PRIVADO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Doutor em Computação. Área de concentração: Ciência da Computação.

Aprovada em agosto de 2026.

BANCA EXAMINADORA

Prof. Luiz Satoru Ochi - Orientador, UFF

Prof. José André de Moura Brito - Co-orientador, ENCE/IBGE

Prof. Igor Machado Coelho, UFF

Prof. Fábio Protti, UFF

Prof. Gustavo Silva Semaan, UFRRJ

Prof. Gustavo da Silva Ferreira, ENCE/IBGE

Prof. Leonardo Silva de Lima, UFPR

Prof. Mauricio Guilherme de Carvalho Resende, UNIFESP

Niterói - 2026

Agradecimentos

Aos Professores José André de Moura Brito e Luiz Satoru Ochi, pelo acompanhamento e pela orientação desde a graduação e o mestrado, por toda a confiança, pela autonomia que sempre me concederam e por apontarem o caminho nas muitas vezes em que me perdi entre as ideias.

Ao Professor Gustavo Silva Semaan, pela referência que representa para mim na pesquisa acadêmica e pelo estímulo constante nessa área desde que nos conhecemos, quando eu ainda desenvolvia meu trabalho de conclusão de curso e ele, sua tese de doutorado.

Ao Professor Gustavo da Silva Ferreira, pela didática, pela paciência e pela capacidade de tornar simples e acessíveis os temas mais complexos, sem abrir mão do rigor científico.

A Andrea Borges Paim, Ingrid Christyne Luquett de Oliveira e Marcus Vinicius Moraes Fernandes, que apoiaram esta empreitada cada um à sua maneira e no seu tempo.

A Carlos Leonardo Henrique Zucarello Freire Feijó Braga e Diego de Oliveira Miro, pelas conversas às vezes profundas, outras nem tanto, mas sempre prazerosas.

A Edjalmo Borges Garcia, Antônio Carlos Pimentel Ferreira e Daniel Leite Lima, pela amizade, pelo companheirismo e por compartilharem comigo tanto conhecimento quanto vivência.

A Margareth Favoreto Fernandes, sem a qual dificilmente teria concluído este trabalho.

A José Roberto de Souza Almuíña e Diego Vargas Freire, amigos de todas as horas. A Saulo da Cruz Campos e Maria Del Carmen Ogando Rivas, amigos queridos de sempre, que compreenderam tantas ausências, mas com quem sei que sempre poderei contar.

A Camila Ceciliano Pereira, pelo amor incondicional, pelo companheirismo e pela enorme compreensão. Por saber o quanto custou chegar até aqui e por permanecer ao meu lado em todos os momentos.

Muito obrigado por fazerem parte da minha trajetória acadêmica, profissional e de vida. Serei eternamente grato.

Resumo

Nas últimas décadas, a crescente digitalização intensificou os riscos de reidentificação e exposição de dados sensíveis, tornando a privacidade um desafio central na análise de dados, ao mesmo tempo em que o compartilhamento de dados é cada vez mais estratégico. Nesse contexto, abordagens tradicionais de proteção são frequentemente custosas ou insuficientes, o que tem impulsionado o uso de Tecnologias de Aprimoramento da Privacidade (*Privacy Enhancing Technologies* - PETs), com destaque para a Privacidade Diferencial (PD), reconhecida por oferecer garantias formais e matematicamente rigorosas de proteção à privacidade.

Apesar de sua robustez teórica, a aplicação de PD ainda impõe desafios relevantes, sobretudo na definição e alocação adequada de *privacy budget*, parâmetro que afeta diretamente o equilíbrio entre privacidade e utilidade. Esse desafio torna-se particularmente relevante no contexto da análise de agrupamentos, em que os algoritmos podem realizar sucessivas operações sobre os dados durante o processo de formação dos grupos. Nesse cenário, o DPk-means, variante diferencialmente privada do k-means, destaca-se como uma das abordagens mais estudadas devido à sua simplicidade e escalabilidade. Entretanto, sua natureza iterativa torna crítica a alocação de *privacy budget*, uma vez que a perda de privacidade se acumula ao longo das atualizações dos centroides. Alocações inadequadas podem introduzir ruído excessivo e degradar significativamente a qualidade das soluções de agrupamento produzidas.

Embora diversos trabalhos proponham a incorporação de PD em algoritmos de agrupamento, a alocação de *privacy budget* ainda é majoritariamente abordada por meio de estratégias heurísticas, com limitada exploração de técnicas de otimização capazes de tratar sistematicamente o compromisso entre utilidade e privacidade. Nesse sentido, esta tese investigou o uso de metaheurísticas para o problema de alocação de *privacy budget* no DPk-means, partindo da hipótese de que abordagens baseadas em otimização podem aumentar a utilidade das soluções sem comprometer as garantias formais de privacidade.

Para isso, foi proposta uma representação baseada em chaves aleatórias para modelar a alocação de *privacy budget* ao longo das iterações do DPk-means, buscando melhorar a qualidade dos centroides e das partições geradas. Essa representação foi combinada com as metaheurísticas BRKGA, ILS e VNS, dando origem a três algoritmos implementados por meio do *framework* Random-Key Optimizer (RKO), que exploram o espaço de alocações em busca de configurações que maximizem a utilidade da solução de agrupamento sob restrições de privacidade.

Os experimentos computacionais realizados com 30 instâncias demonstraram superioridade consistente das abordagens propostas em relação às principais estratégias da literatura. Em particular, o algoritmo baseado no BRKGA apresentou o melhor desempenho global, em termos dos índices utilizados para avaliar a utilidade das soluções de agrupamento obtidas. Considerando o mecanismo de melhor desempenho, apresentou me-

diana do gap percentual em relação à melhor solução de aproximadamente 6%, enquanto a melhor estratégia da literatura obteve cerca de 110% sob o mesmo cenário. Esses resultados evidenciam a magnitude do ganho obtido e reforçam o potencial de abordagens baseadas em otimização para a alocação de *privacy budget*, contribuindo para o avanço do estado da arte no agrupamento diferencialmente privado.

Palavras-chave: Privacidade Diferencial, Análise de agrupamentos, Metaheurísticas, Otimização Contínua.

Abstract

In recent decades, increasing digitalization has intensified the risks of re-identification and exposure of sensitive data, making privacy a central challenge in data analysis, while data sharing is becoming increasingly strategic. In this context, traditional protection approaches are often costly or insufficient, which has driven the adoption of Privacy Enhancing Technologies (PETs), particularly Differential Privacy (DP), recognized for providing formal and mathematically rigorous privacy guarantees.

Despite its theoretical robustness, the application of DP still poses significant challenges, particularly regarding the definition and allocation of the privacy budget, a parameter that directly affects the trade-off between privacy and utility. This challenge becomes especially relevant in the context of clustering analysis, where algorithms may perform successive operations on the data throughout the cluster formation process. In this setting, DPk-means, the differentially private variant of k-means, stands out as one of the most widely studied approaches due to its simplicity and scalability. However, its iterative nature makes privacy budget allocation critical, since privacy loss accumulates over successive centroid updates. Inadequate allocations may introduce excessive noise and significantly degrade the quality of the resulting clustering solutions.

Although several studies propose incorporating DP into clustering algorithms, privacy budget allocation is still predominantly addressed through heuristic strategies, with limited exploration of optimization techniques capable of systematically handling the privacy-utility trade-off. In this context, this thesis investigated the use of metaheuristics for the privacy budget allocation problem in DPk-means, based on the hypothesis that optimization-based approaches can increase solution utility without compromising formal privacy guarantees.

To this end, a random-key-based representation was proposed to model privacy budget allocation across DPk-means iterations, aiming to improve the quality of the resulting centroids and partitions. This representation was combined with the BRKGA, ILS, and VNS metaheuristics, giving rise to three algorithms implemented within the Random-Key Optimizer (RKO) framework, which explore the allocation space in search of configurations that maximize the utility of clustering solutions under privacy constraints.

Computational experiments conducted on 30 instances demonstrated the consistent superiority of the proposed approaches over the leading strategies reported in the literature. In particular, the BRKGA algorithm achieved the best overall performance according to clustering quality metrics. Under the best-performing privacy mechanism, it achieved a median percentage gap of approximately 6% relative to the best-known solution, whereas the best strategy from the literature obtained a gap of about 110% under the same conditions. These results highlight substantial performance gains and reinforce the potential of optimization-based approaches for privacy budget allocation, contributing to the advancement of the state of the art in differentially private clustering.

Keywords: Differential Privacy, Clustering Analysis, Metaheuristics, Continuous Optimization.

Lista de Figuras

2.1	Exemplo de execução do k-means [94, p.453].	12
2.2	Analogia entre os processos de transmissão de uma mensagem criptografada e de consulta a bancos de dados estatísticos – adaptado de [189, 46].	16
2.3	O <i>trade-off</i> entre utilidade e risco de revelação no contexto de PD.	17
2.4	Mecanismo A aplicado a dois conjuntos de microdados adjacentes, D_1 e D_2 , produzindo duas distribuições resultantes, sendo S um subconjunto arbitrário de possíveis saídas de $A(D)$. – adaptado de [9].	22
2.5	Ilustração dos modelos de PD – adaptado de [46].	30
2.6	Exemplos de soluções do k-means com e sem aplicação de PD para perturbação dos centroides.	40
2.7	Exemplo da primeira iteração do k-means com perturbação aplicada através do mecanismo de Laplace com $b = 0,1$ e $b = 0,5$	41
2.8	Diagrama de execução da iteração t do DPk-means no modelo de PD global iterativo.	42
4.1	Fluxo de execução do algoritmo BRKGA proposto.	75
5.1	Distribuição do gap percentual por estratégias e cenários selecionados.	92
5.2	Distribuição do gap percentual por valores de <i>privacy budget</i> , independentemente do mecanismo adotado.	93
5.3	Resultados do teste pareado de Wilcoxon.	96
5.4	Distribuição do tempo de execução até a melhor solução, considerando, para cada cenário e algoritmo, os 300 tempos registrados na execução de 10 repetições em cada um dos 30 conjuntos de dados.	99
5.5	Diagramas de dispersão do resultado da aplicação do k-Means e do DPk-Means nos conjuntos de dados D4, iris e n6000k6.	103

A.1	Diferenças entre os problemas de <i>input</i> e <i>output privacy</i> – adaptado de [164].	136
B.1	Comparação entre as garantias de privacidade oferecidas pelos mecanismos de Laplace e Normal.	164
G.1	Evolução do <i>fitness</i> ao longo das gerações de uma execução do algoritmo BRKGAPD (Cenário: Laplace, $\epsilon = 10$; Conjunto de dados: D1).	193
G.2	Diagrama da recompensa média acumulada - BRKGAPD - Laplace, $\epsilon = 10$	195
G.3	AUC da recompensa média acumulada - BRKGAPD	196
G.4	AUC da recompensa média acumulada - ILSPD	196
G.5	AUC da recompensa média acumulada - VNSPD	197
G.6	Diagramas de dispersão do resultado da aplicação do k-Means ao conjunto de dados iris.	198

Lista de Tabelas

2.1	Número de Stirling de segundo tipo.	10
2.2	Representação de dois conjuntos de dados hipotéticos, D_1 e D_2 , considerando $D_2 \subset D_1$	13
2.3	Exemplo de ataque de inferência em conjunto de microdados sob k-anonimização e l-diversidade, com parâmetros $k = 3$ e $l = 2$	14
3.1	Comparação das abordagens identificadas na revisão da literatura	61
3.2	Consolidado dos trabalhos que abordam a alocação de <i>privacy budget</i> no DPk-means.	63
5.1	Conjuntos de dados utilizados no experimento	85
5.2	Espaço de parâmetros do algoritmo BRKGAPD	87
5.3	Espaço de parâmetros do algoritmo ILSPD	87
5.4	Espaço de parâmetros do algoritmo VNSPD	87
5.5	Mediana do gap percentual em relação à melhor estratégia por cenário, incluindo APAS.	90
5.6	Mediana do gap percentual em relação à melhor estratégia por cenário, excluindo APAS.	91
5.7	Gap percentual - estatísticas resumo.	94
5.8	Algoritmos propostos e estratégias da literatura ordenados segundo rank médio.	95
5.9	Resultados do experimento de estabilidade do BRKGAPD.	101
5.10	Comparativo dos mecanismos em termos da mediana do NICV para cada valor de ϵ e algoritmo, considerando, em cada caso, as 300 execuções referentes a 10 repetições para cada conjunto de dados.	104
5.11	Ranks por cenário dos algoritmos propostos.	105

5.12 P-valores dos testes de hipótese de significância entre diferenças de medianas de NICV para os mecanismos considerados.	106
A.1 Comparação das principais abordagens de <i>output privacy</i>	143
A.2 Exemplo de supressão (local) - Microdados.	143
A.3 Exemplo de recodificação - Microdados.	144
A.4 Exemplo k-anonimização.	146
C.1 Número de trabalhos identificados na RSL.	172
G.1 Combinações vencedoras - NICV	193
G.2 Parâmetros BRKGA	199
G.3 Parâmetros ILS	200
G.4 Parâmetros VNS	201
G.5 Teste de duas proporções - BRKGA	202
G.6 Teste de duas proporções - VNS	203
G.7 Frequência de conjuntos de dados com 10 repetições concluídas em cada cenário.	203
G.8 Conjuntos de dados em que a estratégia APAS foi executada com sucesso.	204
G.9 Mediana do NICV obtido para cada conjunto de dados e cenário.	205
G.10 Resultados do teste pareado de Wilcoxon - Mecanismo de Laplace.	220
G.11 Resultados do teste pareado de Wilcoxon - Mecanismo de Laplace*.	220
G.12 Resultados do teste pareado de Wilcoxon - Mecanismo Normal.	221

Lista de Abreviaturas e Siglas

CP	: <i>Computing Party</i> – parte que executa o processamento;
FL	: Aprendizado federado (do inglês, <i>Federated Learning</i>);
FHE	: Criptografia completamente homomórfica (do inglês, <i>Fully Homomorphic Encryption</i>);
INE	: Instituto Nacional de Estatística;
IP	: <i>Input Party</i> – parte proprietária do dado;
MIA	: Ataque de inferência de pertencimento (do inglês, <i>Membership Inference Attack</i>);
PD	: Privacidade Diferencial;
PET	: Tecnologias de aprimoramento da privacidade (do inglês, <i>Privacy Enhancing Technologies</i>);
PRAM	: Método de pós-aleatorização (do inglês, <i>Post-RAndomization Method</i>);
RP	: <i>Resulting Party</i> – parte que tem acesso ao resultado do processamento;
SDC	: Controle Estatístico de Confidencialidade (do inglês, <i>Statistical Disclosure Control</i>);
sMPC	: Computação segura entre múltiplas partes (do inglês, <i>Secure Multi-Party Computation</i>);
SSS	: Esquema de compartilhamento de segredo (do inglês, <i>Secret Sharing Scheme</i>);
TEE	: Ambiente de execução confiável (do inglês, <i>Trusted Execution Environment</i>);

Notação

A	: Mecanismo (algoritmo) aleatorizado (estocástico/não determinístico);
b	: Parâmetro de escala da distribuição de Laplace;
C_j	: Grupo j , $j = 1, \dots, k$;
c_j	: Centroide do grupo j , $j = 1, \dots, k$;
D	: Conjunto de dados $D \in \mathcal{D}$;
$\mathcal{D}$	: Conjunto de todas as bases de dados possíveis sobre um domínio X ;
Δf	: Sensibilidade global da função f ;
δ	: Probabilidade de falha de um mecanismo A ;
$d_M(x_1, x_2)$	: Distância entre os elementos x_1 e x_2 , em termos da métrica M ;
E	: Conjunto de soluções elite;
$\bar{E}$	: Conjunto de soluções não elite;
ϵ	: <i>Privacy budget</i> total;
ϵ_t	: Parcela de <i>privacy budget</i> alocada à iteração t do DPk-means;
ϵ_t^s	: Parcela de <i>privacy budget</i> alocada à iteração t do DPk-means referente à operação de soma dos atributos dos objetos alocados em cada grupo;
ϵ_t^c	: Parcela de <i>privacy budget</i> alocada à iteração t do DPk-means referente à operação de contagem (frequência) dos objetos alocados em cada grupo;
f_{decod}	: Decodificador;
f_{obj}	: Função objetivo;
h	: Número de atributos em um conjunto de dados D ;
$\text{Im}(A)$	: Conjunto de todas as saídas possíveis de A ;

k	:	Número de grupos (<i>clusters</i>);
$L(A(D_1), A(D_2), y)$	:	Perda de privacidade associada à divulgação da estatística y perturbada pelo mecanismo A ;
n	:	Número de registros (objetos) em um conjunto de dados D ;
ψ	:	Característica de interesse;
q	:	Consulta destinada à extração de informações agregadas de um conjunto de dados D ;
Q	:	Conjunto de q_t consultas sucessivas realizadas ao conjunto de dados D ;
$\mathcal{R}$	:	Contradomínio de A , tipicamente um subconjunto de $\mathbb{R}$ (ou $\mathbb{R}^h$, no caso de consultas vetoriais);
S	:	Suporte de $A(D)$;
S_j	:	Soma dos atributos dos objetos do grupo C_j , $j = 1, \dots, k$;
(s_q, a_q)	:	Par estado-ação do Q-Learning;
σ	:	Parâmetro de escala da distribuição Normal/Gaussiana;
t, T	:	Iterações do DPK-means ($t = 1, \dots, T$);
X	:	Domínio dos registros (universo) parcialmente representados em um conjunto de dados $D \in \mathcal{D}$;
v	:	Unidade de privacidade;
x_i	:	Registro (objeto) i do conjunto de dados D , $i = 1, \dots, n$;
y	:	Resultado da consulta q , aplicada ao conjunto de dados D , sob o mecanismo A , i.e., $A(q, D) = y$;

Sumário

1	Introdução	1
1.1	Objetivos	5
2	Definição do Problema	8
2.1	O problema de agrupamento	8
2.2	Métodos de agrupamento	9
2.3	Risco de violação da privacidade	12
2.4	Privacidade Diferencial	14
2.4.1	Formalização	21
2.4.2	Modelos	29
2.4.3	Mecanismos	32
2.5	O problema de agrupamento diferencialmente privado	34
3	Revisão da Literatura	43
3.1	Abordagens de PD para o problema de agrupamento	43
3.1.1	Métodos baseados em protótipos	44
3.1.2	Métodos baseados em densidade	55
3.1.3	Métodos baseados em grade	58
3.1.4	Métodos baseados em modelos	58
3.1.5	Métodos hierárquicos	59
3.2	Considerações finais relativas à RSL	60
4	Abordagem Proposta	64

4.1	Formulação proposta	65
4.2	<i>Framework</i> Random-Key Optimizer (RKO)	69
4.2.1	Perturbação	70
4.2.2	Recombinação	71
4.2.3	Busca local	71
4.3	Decodificador proposto	72
4.4	Função de custo (avaliação)	73
4.5	Algoritmos propostos	74
4.5.1	BRKGAPD	74
4.5.2	ILSPD	78
4.5.3	VNSPD	79
4.6	Calibração de parâmetros	81
5	Experimentos Computacionais	83
5.1	Conjuntos de dados	84
5.2	Parâmetros	86
5.3	Comparativo entre as estratégias da literatura e os algoritmos propostos . .	87
5.4	Tempos de execução	97
5.5	Avaliação de estabilidade do BRKGAPD	100
5.6	Mecanismos	104
6	Conclusões	107
6.1	Trabalhos futuros	111
	Referências	115
	Glossário	133
	Apêndice A – Tecnologias de Aprimoramento da Privacidade	135

A.1	<i>Input Privacy</i>	136
A.1.1	Computação segura entre múltiplas partes	137
A.1.2	Criptografia completamente homomórfica	138
A.1.3	Ambientes de execução confiáveis	138
A.1.4	Aprendizagem distribuída	139
A.2	<i>Output Privacy</i>	140
A.2.1	Supressão	143
A.2.2	Recodificação	144
A.2.3	k-Anonimização	145
A.2.4	l-Diversidade	146
A.2.5	PRAM	147
A.2.6	Microagregação	147
A.2.7	Geração de microdados sintéticos	147
A.2.8	Privacidade diferencial	148
Apêndice B – Mecanismos de PD		151
B.1	Mecanismo de resposta aleatorizada	151
B.2	Mecanismo exponencial	153
B.3	Definições complementares para o mecanismo de Laplace	154
B.3.1	Demonstração do mecanismo de Laplace	154
B.3.2	Demonstração alternativa do mecanismo de Laplace	155
B.3.3	Generalização (ϵ, δ) -PD	156
B.3.4	Acurácia (α, β)	158
B.3.5	Demonstração do mecanismo de Laplace vetorial	159
B.3.6	Cálculo do <i>privacy budget</i> mínimo (ϵ_m) por iteração do DPk-means	160
B.4	Definições complementares para o mecanismo Normal	162
B.5	Sensibilidade	164

B.6	Mecanismos aplicados ao DPk-means	165
Apêndice C – Metodologia Empregada na Revisão da Literatura		168
C.1	Planejamento	168
C.1.1	Protocolo	169
C.1.2	Crterios de avaliao de qualidade	170
C.2	Conduo	171
Apêndice D – Formulao alternativa para o modelo interativo		174
Apêndice E – Abordagem proposta para o modelo no interativo		178
E.1	<i>Coreset</i>	178
E.1.1	(k, z) -clustering	179
E.1.2	<i>Coreset</i> aplicado ao problema de agrupamento	179
E.1.3	<i>Coreset</i> diferencialmente privado	180
E.1.4	<i>Coreset</i> diferencialmente privado para o problema de agrupamento .	181
E.2	Formulao do problema	182
Apêndice F – Anlises de complexidade dos algoritmos		186
F.1	BRKGAPD	186
F.2	ILSPD	188
F.3	VNSPD	189
Apêndice G – Resultados e anlises complementares		192
G.1	Anlise das combinaes de parmetros dos algoritmos	192
G.2	Q-Learning - Anlise da recompensa	195
G.3	Visualizao dos resultados do agrupamento	197
G.4	Tabelas completas de resultados	199

Capítulo 1

Introdução

Durante grande parte da história humana, os indivíduos eram conhecidos em seus grupos locais e totalmente desconhecidos fora dele, criando um ambiente de pouca privacidade local, mas com baixo risco de roubo de dados pessoais.

Todavia, a partir do momento em que as informações se tornavam de conhecimento público – dentro de um círculo de convivência –, elas deixavam de oferecer vantagens significativas, uma vez que alguém tentando se passar por outro indivíduo seria rapidamente reconhecido [3, p.1]. Em suma, a familiaridade dentro do grupo proporcionava segurança.

Não obstante, com o avanço das tecnologias digitais nas últimas décadas, criou-se um verdadeiro “campo minado de privacidade” [3, p.2], ou seja, um cenário dinâmico, complexo e imprevisível, no qual dados anonimizados podem ser reidentificados e reutilizados de maneiras que comprometem severamente a privacidade, expondo informações pessoais de forma não intencional. Neste contexto, práticas aparentemente inofensivas, como aceitar termos de uso sem leitura atenta, autorizando a coleta e o uso de seus dados, podem representar o acionamento de “armadilhas” que comprometem profundamente a privacidade dos indivíduos. Em outras palavras, a privacidade, que outrora era garantida estruturalmente pelas limitações dos ambientes sociais, tornou-se uma responsabilidade individual, sujeita a riscos crescentes e dificilmente garantida em sua plenitude.

Apesar de regulações rigorosas desempenharem papel relevante na mitigação desses riscos, elas, isoladamente, não são suficientes para garantir a privacidade. Paralelamente, é inegável que, feito de forma responsável, o compartilhamento e o reuso de dados – incluindo dados sensíveis¹ – geram valor significativo para indivíduos, organizações e para

¹Dados sensíveis correspondem a informações cuja divulgação, identificação ou inferência pode causar algum prejuízo ou dano à pessoa ou organização à qual se referem. Tipicamente incluem, por exemplo, dados de saúde, financeiros, biométricos, comportamentais ou estratégicos.

a sociedade como um todo.

No âmbito dos produtores de estatísticas públicas, por exemplo, há um compromisso histórico e institucional não apenas com a produção e disseminação de informações de alta qualidade, mas também com a salvaguarda da privacidade dos respondentes, reconhecendo que a credibilidade das estatísticas depende, em grande medida, do uso responsável da informação [76].

Contudo, vivemos uma realidade paradoxal: ao mesmo tempo em que cresce a preocupação com privacidade, levando ao represamento de dados sensíveis, muitas vezes indisponíveis para análise ou uso como insumo em modelos estatísticos devido aos riscos associados à sua disponibilização, os próprios avanços tecnológicos fomentam uma demanda incessante por dados. Isso ocorre porque modelos analíticos e preditivos, cada vez mais sofisticados, como os utilizados em mineração de dados [13, 116, 39, 82], aprendizado de máquina [188, 190, 192] e *deep learning* [159, 185, 194, 158, 61], demandam grandes volumes de dados para atingir níveis satisfatórios de desempenho.

Além dos riscos aos indivíduos, há riscos institucionais relevantes. Dados abertos e pseudoanonimizados – disponibilizados por governos – podem ser reidentificados por meio de técnicas como *record linkage* [156] ou integração de múltiplas fontes de dados [125], resultando na exposição de informações sensíveis. Este risco é tanto uma preocupação para instituições públicas, que dependem desses dados para a formulação de políticas, quanto para a comunidade científica, que depende desses dados para pesquisa. Casos emblemáticos ilustram a gravidade da questão. Nos setores público e privado, respectivamente, destacam-se a recente análise de vulnerabilidades no contexto do US Census Bureau [2] e o caso do Netflix Prize [149, 172], que permanecem como exemplos clássicos dos riscos associados à reidentificação em dados anonimizados.

Frente a esta problemática, no contexto das soluções tecnológicas, as Tecnologias de Aprimoramento da Privacidade (do inglês, *Privacy Enhancing Technologies* - PET) [150] têm se consolidado como uma resposta robusta e necessária. Estas tecnologias vêm sendo incorporadas de maneira crescente em diferentes domínios e metodologias, com o objetivo de viabilizar análises úteis sem comprometer a privacidade dos indivíduos. Dentre as PET, destacam-se as técnicas de *input privacy*, que geralmente requerem infraestruturas especializadas, como *secure multiparty computation* e *trusted execution environments*, e as de *output privacy*, que oferecem maior flexibilidade operacional. Neste segundo grupo, a Privacidade Diferencial (PD) [57, 59] surge como uma das abordagens mais promissoras, por oferecer garantias formais de privacidade, sendo amplamente adotada em ambientes

acadêmicos, corporativos e governamentais [148].

No contexto dos Institutos Nacionais de Estatística (INEs), após a realização de uma operação estatística – relacionada a uma pesquisa amostral ou censo –, os dados coletados passam por etapas de crítica e tratamento, originando um cadastro ou conjunto de microdados representativo da população de interesse, que dá suporte aos produtos estatísticos disponibilizados ao público. No entanto, por conta de restrições de confidencialidade, tais produtos eventualmente não atendem a demandas mais específicas, como as de pesquisas acadêmicas. Tradicionalmente, esse problema é contornado por meio do uso de salas de acesso seguro ou da celebração de acordos de confidencialidade, sendo ambas alternativas burocráticas e onerosas.

Os acordos são definidos caso a caso entre o INE e instituições parceiras, enquanto o uso das salas exige a submissão e aprovação de um projeto, seguida da construção de um plano de consultas que, após auditoria, são liberadas ao pesquisador ou analista. Esse processo consome tempo e recursos, além de, por vezes, também limitar a utilidade dos resultados, uma vez que sua liberação pode ser condicionada à aplicação de técnicas tradicionais de controle estatístico de confidencialidade aos resultados produzidos [32, p.3]. Nesse sentido, a incorporação de métodos e algoritmos baseados em Privacidade Diferencial surge como uma alternativa robusta, capaz de melhorar a experiência do analista ao oferecer maior utilidade, garantir maior proteção à privacidade e reduzir o custo operacional para a instituição².

Apesar de sua robustez teórica, a aplicação prática de PD ainda impõe desafios significativos [45, p.33][117, 22], dentre outros aspectos, no que se refere à definição adequada do *privacy budget*, uma vez que, independentemente do algoritmo utilizado, a escolha desse parâmetro impacta diretamente o equilíbrio entre proteção da privacidade e utilidade, no que diz respeito à precisão das estatísticas produzidas. Com isso, um esforço recente na literatura tem se concentrado no desenvolvimento de algoritmos capazes de integrar a privacidade diferencial aos paradigmas atuais de aprendizado de máquina, incluindo *deep learning* e modelos baseados em IA. Variantes de métodos consagrados como SVM [192] e SGD [61], associados a mecanismos dinâmicos de controle de *privacy budget* e ao uso de inteligência artificial [11], têm sido propostas para equilibrar o *trade-off* entre privacidade e utilidade.

²Atualmente, uma parcela significativa de recursos – sobretudo de pessoal – é dedicada ao atendimento de solicitações de pesquisadores, jornalistas e cidadãos em geral que demandam estatísticas em níveis de desagregação mais detalhados do que aqueles disponibilizados nos canais convencionais dos INE. A aplicação de Privacidade Diferencial pode contribuir para simplificar ou até mesmo automatizar esse processo, reduzindo de forma substancial os custos associados.

Diante desse cenário, destaca-se a importância da pesquisa no contexto de algoritmos de agrupamento diferencialmente privados, que permitem identificar padrões e estruturas em dados sensíveis sem comprometer a privacidade dos indivíduos. Algoritmos de agrupamento, por sua natureza, são amplamente utilizados em diversas aplicações, desde análises exploratórias até modelos de suporte à decisão [66]. No entanto, observa-se que a maioria dos trabalhos que incorporam PD a esses algoritmos não faz uso de técnicas de otimização para equilibrar o *trade-off* entre utilidade e privacidade – um desafio crítico –, sobretudo para instituições que lidam com a disseminação de dados estatísticos para múltiplos públicos e finalidades.

No contexto do agrupamento diferencialmente privado, a alocação do *privacy budget* caracteriza-se como um problema de otimização – combinatória ou contínua, a depender da formulação adotada – de elevada complexidade³, com múltiplas restrições e funções-objetivo não lineares, o que inviabiliza a aplicação de abordagens exatas na maioria das aplicações reais. Esta característica naturalmente direciona para a adoção de metaheurísticas como uma estratégia eficaz e viável, no que concerne à busca de soluções de boa qualidade em tempo computacional aceitável. Assim, pelo que foi apresentado e com base no que está disponível na literatura, a presente pesquisa foi desenvolvida no sentido de incorporar tais técnicas para propor nova metodologia de alocação de *privacy budget* em problemas de agrupamento que, além de satisfazer PD, otimize o *trade-off* entre utilidade e privacidade, contribuindo para o avanço do estado da arte na proteção da privacidade em mineração de dados e oferecendo subsídios metodológicos relevantes para práticas de controle de confidencialidade e privacidade adotadas por instituições produtoras ou detentoras de dados.

Portanto, o objetivo geral deste trabalho de doutorado foi propor, desenvolver e avaliar uma nova abordagem para a alocação de *privacy budget* em problemas de agrupamento sob privacidade diferencial, visando maximizar a utilidade dos dados para um nível de privacidade previamente definido. Para isso, foram propostos algoritmos baseados em metaheurísticas, integrados a uma metodologia adaptativa de calibração de parâmetros, capaz de otimizar o equilíbrio entre privacidade e utilidade.

³Em [129, p.9], a alocação de *privacy budget* na execução do k-means diferencialmente privado foi formulada como um problema de otimização combinatória, classificado pelos autores como da classe NP-difícil.

1.1 Objetivos

A privacidade diferencial tem sido adotada em muitas áreas, permitindo que análises de dados sensíveis sejam conduzidas, enquanto a privacidade dos indivíduos neles representados seja preservada. No entanto, embora possibilite estimar a perda cumulativa de privacidade, sua implementação adequada é crucial para garantir o equilíbrio entre utilidade e privacidade [55].

Muitas abordagens disponíveis na literatura propõem formas de incorporar PD a algoritmos de agrupamento [214, 129, 203, 80, 29]. No entanto, as abordagens atualmente disponíveis, em sua maioria, não adotam técnicas de otimização para equilibrar o *trade-off* entre utilidade e privacidade – questão importante para os produtores e detentores de dados, considerando as informações que são divulgadas diariamente e os seus diferentes níveis de granularidade e públicos-alvo.

Assim, pelo que foi apresentado e o que é proposto na literatura, foi identificada lacuna referente à aplicação de tais técnicas para propor abordagens que, além de satisfazer PD, otimizem esse *trade-off*. Mais especificamente, destaca-se a relevância do uso de metaheurísticas, uma vez que o problema de alocação de *privacy budget* pode ser abordado como um problema de otimização e que tem funções-objetivo não lineares, como apresentado na Seção 4.1.

Partindo dessa hipótese, o objetivo geral da pesquisa foi desenvolver e validar uma nova metodologia para alocação de *privacy budget* no contexto do problema de agrupamento, orientada a maximizar a utilidade das soluções de agrupamento, superando limitações das principais abordagens disponíveis na literatura.

Como desdobramento do objetivo geral, foram definidos objetivos específicos voltados à construção e avaliação de uma abordagem de otimização aplicada ao problema de alocação de *privacy budget* no contexto do DPK-means. Nesse sentido, os seguintes objetivos específicos foram desenvolvidos:

- realizar uma revisão sistemática da literatura para caracterizar o estado da arte e identificar oportunidades concretas de contribuição científica.
- propor uma formulação baseada em chaves aleatórias para o problema de alocação de *privacy budget* e, a partir de sua implementação no *framework* RKO [37], implementar três algoritmos de otimização baseados nas metaheurísticas BRKGA, ILS e VNS;

- utilizar o método *Q-Learning* para realizar a calibração adaptativa dos hiperparâmetros associados às metaheurísticas, contribuindo para maior robustez e desempenho das soluções obtidas;
- investigar, por meio de experimentos e testes de hipótese, o uso dos mecanismos de Laplace e Normal no contexto do DPK-means, considerando também duas abordagens distintas para o cálculo da sensibilidade e diferentes valores de *privacy budget*, avaliando 18 cenários resultantes dessas combinações para analisar, de forma abrangente, o *trade-off* entre utilidade e privacidade e validar o potencial da abordagem proposta como alternativa eficaz para a alocação de *privacy budget* no Problema de Agrupamento Diferencialmente Privado.

Essa tese conta com cinco capítulos, além da introdução. No Capítulo 2 são apresentados o Problema de Agrupamento Clássico, os riscos associados à violação da privacidade na aplicação desses métodos a dados sensíveis, a técnica de Privacidade Diferencial e a descrição e formalização do problema de alocação de *privacy budget* no contexto do agrupamento diferencialmente privado. O Capítulo 3 traz uma revisão sistemática da literatura, com os principais métodos propostos para abordar esse problema. O Capítulo 4 apresenta uma formulação do problema no contexto de otimização e propõe uma abordagem baseada em um conjunto de metaheurísticas que estão implementadas no otimizador de chaves aleatórias denominado RKO [37]. O otimizador oferece uma estrutura modular, e a aplicação das metaheurísticas exige, basicamente, implementar os procedimentos de decodificação e cálculo de função objetivo. O Capítulo 5 apresenta os resultados do experimento computacional realizado em um conjunto de 30 instâncias. Cinco estratégias de alocação de *privacy budget* identificadas como mais relevantes na revisão da literatura foram comparadas às soluções produzidas pelos algoritmos propostos.

De maneira geral, a partir da análise dos resultados, incluindo a aplicação de testes de hipótese e avaliação de gaps do NICV, foi observada uma superioridade consistente das abordagens baseadas em otimização desenvolvidas nesta tese, evidenciando o potencial das metaheurísticas para explorar de maneira eficaz o espaço de soluções do problema de alocação de *privacy budget*. Em particular, o algoritmo baseado no BRKGA (BRKGAPD) apresentou o melhor desempenho global em termos do NICV das soluções de agrupamento produzidas, obtendo medianas de *gap* percentual inferiores às observadas em todas as demais abordagens avaliadas, incluindo os outros dois algoritmos propostos. Os resultados obtidos indicam ganhos relevantes em termos de utilidade das soluções de agrupamento, reflexo da menor degradação causada pela alocação otimizada do *privacy budget* e melhor

equilíbrio no *trade-off* entre privacidade e qualidade do agrupamento.

Além disso, os experimentos reforçam a adequação da modelagem do problema como um problema de otimização contínua, bem como a eficácia da representação baseada em chaves aleatórias e do uso de metaheurísticas para abordar o problema de alocação de *privacy budget* no DPk-means. Por fim, o Capítulo 6 traz as conclusões acerca da abordagem proposta, destacando a contribuição na redução segura de barreiras de acesso a dados sensíveis, favorecendo o desenvolvimento de metodologias analíticas e estudos com potencial de impacto científico, institucional e social. Esse capítulo traz ainda algumas propostas de trabalhos futuros identificados ao longo da pesquisa.

Capítulo 2

Definição do Problema

Este capítulo estabelece o contexto em que a pesquisa realizada nesta tese se insere. Inicialmente, é apresentado o problema de agrupamento clássico e os principais métodos de agrupamento utilizados em análise de dados, com destaque para o k-means. Em seguida, discute-se o risco de violação da privacidade quando tais métodos são aplicados a dados sensíveis, enfatizando como a extração de informação pode, inadvertidamente, expor características individuais ou permitir inferências indesejadas. À luz dessas questões, a técnica de Privacidade Diferencial é introduzida, bem como o arcabouço teórico que fundamenta a proteção formal da privacidade, permitindo a formulação do problema de agrupamento diferencialmente privado, no qual se busca conciliar, de maneira rigorosa, a utilidade dos agrupamentos com garantias matemáticas de privacidade.

2.1 O problema de agrupamento

Resolver um problema de agrupamento consiste, basicamente, em particionar um conjunto de objetos representados em uma base de dados - por meio de seus registros -, de maneira que objetos alocados a um mesmo **grupo** tenham grande similaridade entre si em relação ao seu conjunto de atributos¹. Além disso, os objetos alocados em grupos distintos devem ser o mais dissimilares possível, com base nos mesmos critérios. Em outras palavras, deseja-se maximizar a similaridade interna dos grupos (similaridade dos objetos alocados ao mesmo grupo) e a separação entre os grupos.

Formalmente, o problema de agrupamento clássico consiste em, dado um conjunto de

¹A similaridade entre os objetos e, de maneira análoga, a dissimilaridade, é comumente medida através de métricas de distância. Para conjuntos de dados com atributos numéricos é frequente o uso de distâncias baseadas na métrica de Minkowski, como as distâncias Euclidiana e de Manhattan.

dados D com n objetos (relacionados a registros) definidos pelos seus h atributos, de forma que $D = \{x_1, \dots, x_i, \dots, x_n\}$, $x_i \in \mathbb{R}^h$, $\forall i \in \{1, \dots, n\}$, construir k grupos (*clusters*) homogêneos² ($C_1, \dots, C_j, \dots, C_k$). Mais especificamente, no agrupamento completo e exclusivo [184, p.492-493], as restrições 2.1 e 2.2 devem ser respeitadas, garantindo que todos os objetos sejam alocados a um e somente um grupo. A restrição 2.3 garante que todos os grupos tenham, ao menos, um objeto alocado.

$$\bigcup_{j=1}^k C_j = D \quad (2.1)$$

$$C_j \cap C_l = \emptyset, \forall j, l \in \{1, \dots, k\}, j \neq l \quad (2.2)$$

$$|C_j| > 0, \forall j \in \{1, \dots, k\} \quad (2.3)$$

O problema de agrupamento é um problema de classificação não-supervisionado [184, p.491], i.e., os métodos de agrupamento realizam a partição dos dados de maneira autônoma e sem necessidade de um conhecimento prévio sobre classes ou valores associados a eles, ainda que seja importante a validação das soluções obtidas, seja por meio de índices de validação [186] ou mesmo pela avaliação de especialistas no domínio de estudo.

No processo de agrupamento, os grupos são formados com base em um princípio indutivo, também chamado de função objetivo ou critério de agrupamento [147]. Essa função é utilizada para identificar um modelo ou estrutura que melhor represente os grupos em um determinado conjunto de dados. Após definir a função objetivo, o desafio do agrupamento passa a ser otimizar essa função para o conjunto de dados analisado.

2.2 Métodos de agrupamento

Os métodos de agrupamento são abordagens que se propõem a produzir soluções de boa qualidade (não necessariamente ótimos globais) para o problema de agrupamento. Para isso, geralmente adotam estratégias baseadas em heurísticas, i.e., abordagens que dispensam enumeração e avaliação exaustiva de todas as possíveis soluções visando garantir o ótimo global em relação à função objetivo considerada, o que é impraticável na maioria dos casos. Como os problemas de agrupamento são, em geral, NP-hard [10, 124], é comum

²Um grupo é homogêneo se os objetos alocados a ele são suficientemente similares entre si, segundo medida de similaridade previamente definida (e.g., distância euclidiana).

a adoção de métodos heurísticos ou metaheurísticas, dado que a enumeração exaustiva das soluções é inviável. No Problema de Agrupamento Clássico [94, p.444-445], o número total de soluções possíveis é definido pelo número de Stirling de segundo tipo [92, p.257], que permite calcular o número de maneiras de particionar um conjunto de n objetos em exatamente k grupos não vazios. A Tabela 2.1 apresenta o número de Stirling para alguns valores de n e k , sendo possível ver que, mesmo para valores relativamente pequenos de ambos os parâmetros, é inviável enumerar todas as possíveis soluções (partições).

Tabela 2.1: Número de Stirling de segundo tipo.

n	$k = 2$	$k = 3$	$k = 5$
10	5,110e+02	9,330e+03	4,252e+04
20	5,243e+05	5,806e+08	7,492e+11
50	5,629e+14	1,196e+23	7,401e+32
100	6,338e+29	8,590e+46	6,574e+67

Tabela elaborada pelo autor.

Os métodos de agrupamento podem ser classificados segundo sua exclusividade [147, p.22]. Métodos exclusivos, também chamados de *crisp* ou *hard*, são aqueles nos quais os objetos são alocados a um e somente um grupo. Métodos não exclusivos, também chamados de *fuzzy* ou *soft*, são aqueles em que cada um dos objetos recebe uma probabilidade de pertencimento a cada um dos grupos. Na presente pesquisa foram considerados apenas os métodos exclusivos.

É possível dividir os métodos de agrupamento também em relação à estratégia de particionamento adotada [94, p.448][184, p.493]. Os métodos hierárquicos são aqueles em que as soluções são obtidas através de sucessivas divisões ou junções (de objetos e/ou grupos já formados), gerando uma sequência de partições aninhadas. No caso dos métodos não hierárquicos, como aqueles baseados em protótipos, o número de grupos deve ser definido a priori, uma vez que cada grupo é representado por seu protótipo, sendo a alocação dos objetos realizada de forma iterativa em função da similaridade com esses protótipos. Em outras palavras, um grupo é um conjunto de objetos, sendo cada objeto mais similar (ou mais próximo) ao protótipo que define o grupo do que ao protótipo de qualquer outro grupo [184, p.494]. Os métodos baseados em densidade formam grupos com base na distribuição dos objetos no espaço de atributos considerado, identificando regiões de alta densidade de objetos como pertencentes a um mesmo agrupamento e separando-as por regiões de baixa densidade, interpretadas como fronteiras naturais entre diferentes grupos. Existem ainda os métodos baseados em grade [68, p.269], que particionam o espaço de atributos associados aos objetos em uma grade de células e realizam o agrupamento com

base na distribuição dos objetos dentro dessas células. Em [66], os autores apresentam uma revisão recente e abrangente sobre métodos de agrupamento.

Dentre os métodos de agrupamento mencionados, apenas o k-means foi explorado no contexto do problema de agrupamento diferencialmente privado e, por isso, será descrito a seguir. O k-means tem por objetivo particionar um conjunto de n objetos em k grupos [94, p.451] completos e exclusivos, i.e., todo objeto é alocado a somente um grupo. É baseado em protótipos definidos em termos de centroides, sendo o centroide de um grupo correspondente à média de cada um dos atributos dos objetos nele contidos. Em geral, adota como medida de dissimilaridade entre objetos (ou objetos e centroides) a distância euclidiana ao quadrado. Apesar das diferentes propostas de algoritmos para implementação [77, 137, 99], o método se destaca por sua eficiência computacional, sendo uma alternativa particularmente interessante para lidar com grandes conjuntos de dados [161, p.6].

Sem abordar as especificidades dos algoritmos, o método pode ser descrito através dos seguintes passos gerais, quais sejam: (i) A partir do parâmetro de entrada k , referente ao número de grupos a serem gerados, k centroides - correspondentes a objetos da base - são inicialmente selecionados, cada um associado a um grupo; (ii) cada um dos $(n - k)$ objetos restantes tem então sua dissimilaridade avaliada em relação aos k centroides e a alocação é feita ao grupo cujo centroide apresentar menor dissimilaridade; (iii) os centroides são então atualizados a partir dos grupos formados³. Após a atualização dos centroides, se um objeto estiver mais próximo de um centroide diferente daquele ao qual foi anteriormente associado, ele é realocado para o grupo correspondente ao centroide mais próximo. Esse procedimento é repetido até que seja satisfeito algum critério de parada, por exemplo, as alterações nas coordenadas dos centroides c_j serem pouco significativas (e.g., inferior a ξ) entre duas iterações consecutivas, i e $i + 1$, de tal forma que $\|c_j^i - c_j^{i+1}\| \leq \xi, \forall j \in \{1, \dots, k\}$. A Figura 2.1 ilustra o funcionamento do k-means, mais especificamente, (a) inicialização e agrupamento inicial – Figura 2.1a; (b) iterações intermediárias – Figura 2.1b; e (c) solução de agrupamento final – Figura 2.1c.

³Diferentemente dos medoids, os centroides não representam objetos do conjunto de dados. Portanto, mesmo que a inicialização se dê pela seleção de k objetos, a partir da primeira atualização, os centroides necessariamente passam a representar pontos no espaço de atributos.

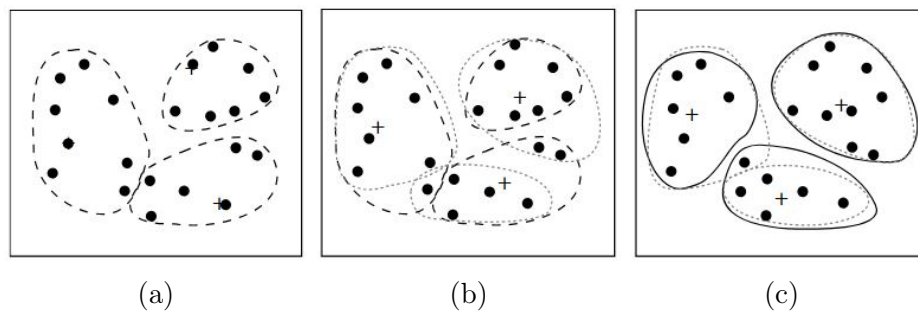


Figura 2.1: Exemplo de execução do k-means [94, p.453].

Este método tem como principais limitações ser pouco efetivo na identificação de grupos de formas arbitrárias, i.e., não esféricas, e estar sujeito a influência de valores extremos (*outliers*), que podem ter grande influência na definição dos centroides e, consequentemente, na alocação dos objetos; o que pode provocar distorções na solução produzida [94, p.454].

2.3 Risco de violação da privacidade

No contexto do problema de agrupamento, considere que uma solução foi produzida aplicando-se o k-means em um conjunto de dados D_1 , com n registros e h atributos, por um agente com acesso legítimo aos dados. Esse agente (e.g., pesquisador acadêmico), ao publicar os resultados do seu estudo ou análise, divulga os centroides finais, ou seja, a informação agregada, por atributo, quanto aos objetos (registros da base original) que compõem cada grupo. Um analista mal-intencionado (i.e., intruso ou adversário), dotado de conhecimento prévio (e.g., fontes de dados auxiliares), pode fazer uso dessa informação para violar a privacidade dos indivíduos representados no estudo [155, p.207]. Mais especificamente, considere que o analista tem acesso a um conjunto de dados D_2 , que é subconjunto de D_1 . É possível combinar essa informação com as coordenadas dos centroides divulgados na pesquisa para reconstruir D_1 , oferecendo ao intruso acesso ilegítimo a dados de indivíduos que ele não detinha anteriormente.

No pior caso, considere que D_2 difere de D_1 em apenas 1 registro, i.e., D_2 contém $n - 1$ dos n registros de D_1 e os mesmos h atributos⁴, como representado na Tabela 2.2. Com isso, o intruso conhece as coordenadas de $n - 1$ registros de D_1 . Além disso, através dos resultados publicados, conhece também as coordenadas dos k centroides obtidos a partir de D_1 . Uma vez que o cálculo para obtenção dos centroides é conhecido, é possível

⁴No contexto de confidencialidade ou privacidade de dados, esse cenário representa o pior caso, pois é aquele em que o conhecimento prévio do adversário é máximo.

identificar a qual grupo pertence o registro faltante e reconstruir suas coordenadas ou atributos a fim de alcançar o mesmo centroide [135, p.1541].

Tabela 2.2: Representação de dois conjuntos de dados hipotéticos, D_1 e D_2 , considerando $D_2 \subset D_1$.

Idade	Altura	Peso
28	1,79	93
29	1,86	90
21	1,85	97
23	1,68	98
50	1,79	58
55	1,82	68
47	1,79	63
49	1,74	67
31	1,66	66
37	1,51	61
35	1,61	66
36	1,73	77

Tabela elaborada pelo autor.

(a) Conjunto D_1

Idade	Altura	Peso
28	1,79	93
29	1,86	90
21	1,85	97
23	1,68	98
50	1,79	58
55	1,82	68
47	1,79	63
49	1,74	67
31	1,66	66
37	1,51	61
35	1,61	66

Tabela elaborada pelo autor.

(b) Conjunto D_2

Como exemplo, considere estudo hipotético realizado em determinada população a fim de identificar prevalência de doenças graves a partir de atributos socioeconômicos. Nesse caso, os pesquisadores foram bem sucedidos em isolar um grupo de indivíduos com alta probabilidade de desenvolver uma doença específica. O perfil desses indivíduos, dado pelos atributos utilizados no estudo, pode ser levado em conta na elaboração de políticas públicas voltadas à prevenção e tratamento de tal doença. Entretanto, uma operadora de plano de saúde pode fazer uso do mesmo resultado para identificar indivíduos propensos a essa condição e rejeitar ou rescindir seus contratos. Esse tipo de conduta importaria dificuldades aos participantes do estudo, o que, além das consequências potencialmente severas para eles, provavelmente desestimularia sua participação em estudos futuros.

Riscos como esse podem ser mitigados através de técnicas como a Privacidade Diferencial (PD), apresentada na Seção 2.4. A PD faz parte do arcabouço das Tecnologias de Aprimoramento da Privacidade (do inglês, *Privacy Enhancing Technologies* - PET) [150, 71], apresentadas no Apêndice A (p.135).

2.4 Privacidade Diferencial

As Tecnologias de Aprimoramento da Privacidade (do inglês, *Privacy Enhancing Technologies* - PET) [150, 71] compreendem um amplo conjunto de abordagens voltadas para a proteção da privacidade de dados pessoais. Dentre elas, a Privacidade Diferencial (PD) se destaca por ter fundamentação matemática sólida e pela capacidade de efetivamente garantir a privacidade de dados pessoais, quando corretamente aplicada. Abordagens como k-anonimização [182] (Seção A.2.3, p.145) e l-diversidade [136] (Seção A.2.4, p.146) são conhecidamente vulneráveis a ataques de inferência⁵ [135, p.1542], por exemplo. A Tabela 2.3 apresenta um conjunto de microdados hipotético no qual estão presentes informações sensíveis referentes a condições de saúde de pacientes de um hospital. A fim de reduzir o risco de revelação, foram aplicados métodos de k-anonimização e l-diversidade, com parâmetros $k = 3$ e $l = 2$. No entanto, uma operadora de plano de saúde que tenha em seu cadastro um paciente desse hospital do sexo masculino, com idade entre 40 e 45 anos, residente na região representada pelo CEP 148xx, pode inferir que ele tem probabilidade de 0,66 de ser portador de HIV e optar por cancelar sua cobertura. A PD pode ser usada para inibir esse tipo de ataque de forma efetiva. O restante desta seção busca apresentar de forma intuitiva os conceitos de PD e formalizá-los em seguida.

Tabela 2.3: Exemplo de ataque de inferência em conjunto de microdados sob k-anonimização e l-diversidade, com parâmetros $k = 3$ e $l = 2$

CEP	Idade	Sexo	Condição
130xx	30-35	F	Câncer
130xx	30-35	F	Câncer
130xx	30-35	F	Gripe
148xx	40-45	M	HIV
148xx	40-45	M	HIV
148xx	40-45	M	Asma

Tabela elaborada pelo autor.

No contexto de criptografia, sem entrar em pormenores, diz-se que a segurança semântica (*semantic security*) é garantida se um intruso⁶ não é capaz de aprender nada novo sobre a mensagem original não criptografada (*plaintext*) a partir da mensagem criptografada (*ciphertext*). Esse entendimento é frequentemente utilizado para tentar definir

⁵Ataques de inferência de pertencimento/adesão/associação (do inglês, *Membership Inference Attack* - MIA), como o ataque de rastreamento [60, p.70-77][44, p.266-267], incorrem em uma violação de privacidade sutil, em que um invasor usa estatísticas ruidosas e dados de um indivíduo alvo para determinar se esse indivíduo está representado em um conjunto de microdados específico [60, p.71], revelando informações sensíveis.

⁶O intruso representa uma parte ilegítima na comunicação, i.e., uma parte que tenta acessar indevidamente a mensagem ou informação trocada entre remetente e destinatário.

privacidade no contexto de análise de dados [59, p.13]. Porém, nesse contexto, o destinatário da mensagem é um analista, i.e., um profissional ao qual é concedido o direito de extrair informações ou estatísticas agregadas de um [conjunto de microdados](#) construído, por exemplo, a partir de uma pesquisa amostral realizada em uma população⁷ alvo. É, entretanto, necessário garantir que esse analista não seja capaz de aprender nada sobre nenhum indivíduo específico dessa população.

A analogia é, portanto, falha [59, p.13], pois, na transmissão de mensagens ou pacotes criptografados, há duas partes legítimas – remetente e destinatário – e uma parte ilegítima – intruso (adversário), como ilustrado na Figura 2.2. No contexto de análise de dados com controle do [risco de revelação](#), existem apenas duas partes e ambas são legítimas⁸, como ilustrado na Figura 2.2. O proprietário ou [curador](#) (*curator*) do dado corresponde à parte emissora (remetente), e o [analista](#) (*querier*) corresponde, simultaneamente, às partes receptora e intrusa (destinatário e adversário). Ele buscará extrair o máximo de informação para atender ao seu propósito e, mesmo que não tente deliberadamente extrair informações sensíveis sobre indivíduos específicos, o controlador deve garantir que não seja possível fazê-lo⁹. Ou seja, ao contrário dos sistemas criptografados, na análise de dados com preservação de privacidade há que se ter parcimônia, pois não é possível negar nem oferecer toda informação ao analista.

Essa noção pode ser representada pelo conceito de [utilidade](#). Se a utilidade é máxima, toda informação é disponibilizada ao analista (i.e., acesso total e irrestrito ao [conjunto de microdados](#)), ao passo que, se a utilidade é zero, nenhuma informação pode ser obtida por ele (i.e., acesso bloqueado ao [conjunto de microdados](#)). A utilidade zero é indesejada por todas as partes envolvidas, já a utilidade máxima representa um problema para o controlador do conjunto de microdados e para os indivíduos nele representados.

O analista deseja sempre maximizar a utilidade, ao passo que o curador deseja permitir que o máximo de informações sejam extraídas, porém sem que nenhuma informação individual seja exposta ao analista. Essa noção introduz o conceito de [risco de revelação](#), que faz referência à probabilidade de o analista aprender algo sobre indivíduos específicos

⁷Uma população é o conjunto completo de todas as unidades de análise (e.g., indivíduos, eleitores, empresas, produtos etc.) que possuem uma característica em comum que se deseja estudar ou sobre a qual se deseja obter informações.

⁸Evidentemente é possível existir uma parte ilegítima, que tente acessar o conjunto de microdados e extrair informações sem permissão, porém tal situação é tratada no contexto de segurança da informação, ou seja, está fora do escopo da presente pesquisa.

⁹Para alcançar essa proteção, o curador pode, por exemplo, restringir o nível de desagregação que o analista tem acesso ou adicionar ruído (perturbar) ao resultado da consulta realizada por ele, como ilustrado na Figura 2.2.

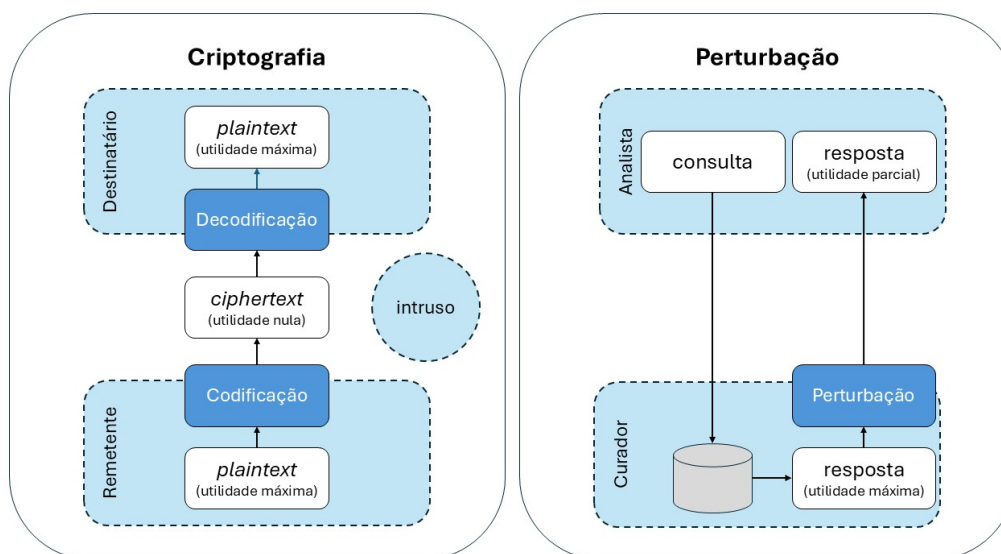


Figura 2.2: Analogia entre os processos de transmissão de uma mensagem criptografada e de consulta a bancos de dados estatísticos – adaptado de [189, 46].

a partir das informações extraídas. A redução da utilidade é um recurso utilizado pelo curador para mitigar o risco de revelação de informações sensíveis.

A Privacidade Diferencial (PD) [57] é uma abordagem que adiciona um ruído aleatório, gerado segundo uma distribuição de probabilidade conhecida, a toda informação extraída pelo analista, reduzindo a utilidade a fim de controlar o risco de revelação¹⁰. Evidentemente que a intensidade do ruído é aspecto fundamental para aplicações bem sucedidas dessa abordagem. A robustez da fundamentação matemática está associada ao fato de PD oferecer garantia quantitativa sobre o quão distinguíveis as informações de um indivíduo podem ser para um potencial intruso [44, p.8], em meio às informações dos demais indivíduos, proporcionando critério para especificação dos parâmetros da distribuição do ruído, descrita adiante.

A Figura 2.3 apresenta uma intuição do *trade-off* entre utilidade e risco de revelação (ou perda de privacidade) no contexto de PD. Em 2.3a é apresentada a informação original, i.e., a imagem não perturbada. Nela é possível ver cinco veículos, suas respectivas cores e modelos e até mesmo as placas de alguns. É possível ver ainda uma placa com os dizeres “*No Parking Allowed At Any Time*”, que, em tradução livre, significa “não é permitido estacionar em nenhum momento”. Assim, a divulgação desta imagem sujeita os proprietários dos veículos à aplicação de eventuais sanções pecuniárias, como multa por estacionamento proibido. Considere, no entanto, que esta imagem se destina a permitir

¹⁰No contexto de PD, o risco de revelação é quantificado através da perda de privacidade (do inglês, *privacy loss*), formalizada na Definição 9 (p.25)

contabilizar ou auditar o número de veículos infratores e não efetivamente a aplicação de sanções. Em 2.3b é apresentada uma versão embaçada, i.e., perturbada, da imagem original. Considerando que o critério de violação seja a exposição da placa, três dos quatro veículos ficaram protegidos. Em 2.3c é aplicada uma perturbação mais intensa, a fim de ocultar as placas de todos os veículos. Pelo critério de exposição da placa, todos os veículos estão agora protegidos. Mas uma avaliação mais rigorosa poderia considerar tal critério insuficiente, uma vez que o segundo veículo se destaca substancialmente dos demais. Digamos, por hipótese, que seja um modelo raro, do qual existem poucas unidades. A imagem 2.3c poderia oferecer informações suficientes para permitir sua identificação por parte de um especialista. Devido a suas características (atributos) diferenciadas, esse veículo exige que uma perturbação mais intensa seja aplicada, e a imagem resultante é apresentada em 2.3d. Neste caso, mesmo sob critério mais conservador, não há risco de revelação de nenhum dos veículos, porém, poderia-se argumentar que o propósito original, de contagem dos veículos, ficou comprometido, em função da drástica redução da utilidade. Portanto, quanto mais heterogênea é a população considerada, mais desafiador é encontrar o equilíbrio entre utilidade e risco de revelação.



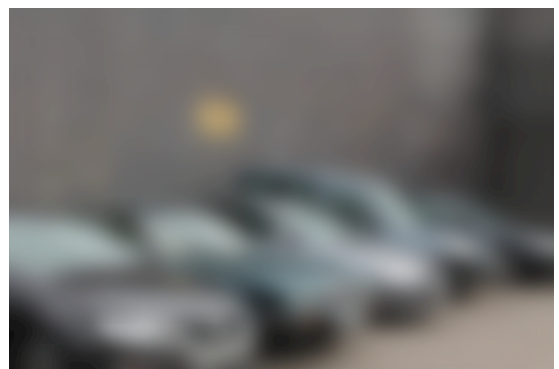
(a)



(b)



(c)



(d)

Figura 2.3: O *trade-off* entre utilidade e risco de revelação no contexto de PD.

No que se refere à produção do dado, a PD pode ser aplicada essencialmente através dos modelos global ou local¹¹. A resposta aleatorizada [195, 41] é o primeiro exemplo formalmente documentado de abordagem que busca controlar o risco de revelação através de um processo de adição de ruído aleatório e é especialmente apropriada ao modelo de PD local, no qual o ruído é adicionado no momento da geração (ou coleta) do dado.

Essa abordagem foi idealizada com o objetivo de reduzir a não resposta durante a coleta de quesitos dicotômicos, cujas respostas possam ser consideradas sensíveis (e.g., vítimas de violência, usuários de drogas etc.). O método consiste em o respondente realizar o lançamento de uma moeda honesta¹² e, se obtiver coroa, responder honestamente. Caso contrário, ele deverá realizar um novo lançamento e responder “sim” se obtiver cara e “não” se obtiver coroa, independentemente de qual seja a resposta verdadeira, o que corresponde à adição de ruído. Uma vez que os resultados dos lançamentos da moeda não são armazenados, a informação de quem respondeu honestamente ou aleatoriamente não pode ser recuperada.

Considere um quesito que investigue, em uma amostra de tamanho n , a presença de uma característica de interesse ψ , com duas possíveis respostas: sim ou não; cuja proporção real, desconhecida, é dada por p_ψ . Sob PD Local, a proporção p_ψ pode ser estimada por $\hat{p}_\psi = 2(\frac{1}{n} \sum_{i=1}^n \psi_i) - \frac{1}{2}$, onde $\psi_i = 1$ se o indivíduo i possui a característica de interesse ψ e $\psi_i = 0$ caso contrário. Assintoticamente, a proporção observada na amostra, dada pelo termo $\frac{1}{n} \sum_{i=1}^n \psi_i$, é acrescida de $\frac{1}{4}$ de casos positivos ($\psi_i = 1$) advindos do processo de aleatorização, uma vez que metade das respostas simplesmente registra o segundo lançamento da moeda, cuja proporção esperada é $\frac{1}{2}$. O fator 2 recompõe a proporção da característica de interesse, mas também duplica os casos positivos introduzidos pelo processo aleatório. A subtração do termo $\frac{1}{2}$ desconta esses casos, corrigindo parcialmente o ruído adicionado.

No modelo de PD global, o dado é coletado e armazenado em sua forma original, sendo o ruído introduzido apenas no resultado ou saída produzido a partir de cada [consulta](#) realizada. A essência por trás da ideia de PD é impedir que um analista seja capaz de identificar a presença ou ausência de indivíduos em um conjunto de dados e, consequentemente, revelar informações sobre indivíduos específicos.

É possível ilustrar esse conceito através do exemplo¹³ em que um professor que, após

¹¹Os modelos de PD são detalhadamente descritos na Seção 2.4.2.

¹²Em uma moeda honesta, a probabilidade de se obter o resultado “cara” deve ser a mesma de se obter o resultado “coroa”, i.e., $P(cara) = P(coroa) = 0,5$.

¹³Adaptado de [44, p.8-10].

corrigir as provas de uma de suas turmas, com 10 alunos, escreve no quadro a nota média: 8,5. Entretanto, antes de entregar as provas, recebe a notícia de que um dos alunos desistiu do curso. O professor então atualiza a nota desconsiderando o resultado desse aluno: 8,7. Embora possa não ser evidente a princípio, é trivial recuperar a nota do aluno desistente.

Seja n o número de alunos ativos no curso, x_i a nota do aluno i e $\bar{x}_1$ a nota média observada. Assim, a média inicial foi obtida por

$$\bar{x}_1 = \frac{1}{10} \sum_{i=1}^{10} x_i = 8,5$$

Após a desistência de um dos alunos, a média atualizada foi obtida por

$$\bar{x}_2 = \frac{1}{9} \sum_{i=1}^9 x_i = 8,7$$

Denotando a nota do aluno desistente por x_* , temos então que a diferença entre as médias é dada por

$$\bar{x}_1 - \bar{x}_2 = \frac{1}{10} \sum_{i=1}^{10} x_i - \frac{1}{9} \sum_{i=1}^9 x_i = \left(\frac{1}{10} - \frac{1}{9} \right) \sum_{i=1}^9 x_i + \frac{1}{10} x_* = -0,2$$

Podemos então isolar o valor de x_* , recuperando a nota do aluno desistente:

$$x_* = 10 \cdot \left(-0,2 + \frac{1}{90} \sum_{i=1}^9 x_i \right) = 10 \cdot \left(-0,2 + \frac{1}{10} \bar{x}_2 \right) = -2 + \bar{x}_2 = -2 + 8,7 = 6,7$$

Generalizando o problema, considere dois conjuntos de microdados, denotados por D_1 e D_2 , respectivamente de tamanhos n e $n - 1$, que diferem entre si por apenas um registro¹⁴. Em outras palavras, todos os indivíduos representados em D_1 estão também representados em D_2 , com exceção de um único indivíduo, representado apenas em D_1 ¹⁵,

¹⁴Esse critério emula a condição na qual o intruso tem o máximo conhecimento prévio possível. A PD garante que, mesmo quando um intruso obtém todas as informações, exceto as do alvo, ele não consegue determinar se as informações do alvo estão nos dados atacados, ou seja, a privacidade diferencial pode resistir a ataques na condição de máximo conhecimento prévio por parte do adversário [127, p.1].

¹⁵Neste contexto, assume-se dados representados no formato análogo ao de bases de dados relacionais, em que cada registro é armazenado em uma linha e, após as devidas integrações de tabelas auxiliares e tratamento de duplicidade de registros, cada registro da base final corresponde a um indivíduo. Em outras formas de representação, como grafos [26, 183], por exemplo, é necessário avaliar se a granularidade

e D_1 e D_2 são ditos conjuntos de dados adjacentes¹⁶. Uma [consulta](#) q , formulada com o objetivo de extrair estatísticas sobre uma característica de interesse ψ dos indivíduos, terá resultados diferentes quando aplicada em ambos os conjuntos de microdados, uma vez que, em D_2 , haverá um indivíduo a menos com a informação referente a essa característica representada.

O exemplo da nota média da turma pode ser facilmente adaptado a essa notação. Os conjuntos de dados D_1 e D_2 correspondem, respectivamente, à turma inicial, de 10 alunos, e à final, excluído o aluno desistente. O atributo ou característica de interesse ψ é a nota obtida pelos estudantes. A [consulta](#) q representa a função ou algoritmo a ser aplicado aos conjuntos de microdados com o objetivo de extrair informações, ou seja, a média das notas dos estudantes.

No caso mais geral, o intruso não está interessado, especificamente, em recuperar a nota do aluno desistente, pois sequer tem conhecimento desse tipo de evento. Todavia, pode estar interessado em explorar alguma vulnerabilidade por ventura existente. Nesse cenário, a estratégia do intruso poderia ser executar até n vezes a [consulta](#) q em D_1 , removendo um dos registros de D_1 em cada execução, ou seja, executar $q(D_1 \setminus x_i^1)$ para todo $i \in \{1, \dots, n\}$, onde $D_1 = \bigcup_{i=1}^n x_i^1$, até obter o mesmo resultado de $q(D_2)$, i.e., até que, para algum i , $q(D_1 \setminus x_i^1) = q(D_2)$, isolando assim o registro de D_1 ausente em D_2 . A PD adiciona um ruído aleatório ao resultado de todas as execuções da [consulta](#) q , de forma a limitar a diferença entre $q(D_1 \setminus x_i^1)$ e $q(D_2)$, a fim de inviabilizar essa estratégia.

Mesmo quando o indivíduo ausente é conhecido, como no exemplo da nota média da turma, é possível adotar essa abordagem baseada na adição de ruído, com a diferença de que apenas duas consultas precisam ser consideradas. Os alunos receberiam a informação de que as estatísticas divulgadas possuem um ruído adicionado sem, é claro, que o valor do ruído seja revelado [44, p.12]. Com isso, a estratégia utilizada anteriormente para recuperar a nota do aluno desistente precisaria incorporar duas novas variáveis, r_1 e r_2 , correspondentes ao ruído adicionado às médias das notas dos 10 e 9 alunos, respectivamente.

$$\bar{x} - \bar{x}' = \left(\frac{1}{10} \sum_{i=1}^{10} x_i + r_1 \right) - \left(\frac{1}{9} \sum_{i=1}^9 x_i + r_2 \right) = \left(\frac{1}{10} - \frac{1}{9} \right) \sum_{i=1}^9 x_i + \frac{1}{10} x_{10} + r_1 - r_2$$

Consequentemente, a equação não pode mais ser resolvida analiticamente, uma vez mínima a ser considerada está associada aos vértices (representando indivíduos, por exemplo) ou arestas (representando relacionamentos entre os vértices) do grafo [138, 210, 221, 176].

¹⁶O conceito de conjunto de dados adjacentes é formalizado na Definição 3 (p.23).

que, além da nota do aluno desistente, x_* , há outras duas variáveis desconhecidas, r_1 e r_2 . Esse exemplo também evidencia a importância de especificar adequadamente a distribuição a partir da qual o valor do ruído será selecionado. Considerando a escala do valor original da estatística, se o ruído for muito elevado, a divulgação da informação perturbada, i.e., adicionada do ruído, terá pouco relevância, ou seja, pouca utilidade (e.g., Figura 2.3d). Por outro lado, um ruído de pouca magnitude permitirá que os termos r_1 e r_2 sejam desconsiderados na equação anterior, às custas de uma pequena imprecisão no resultado final ao recuperar o valor de x_* .

Ademais, o ruído adicionado deve ser suficiente para garantir proteção para todos os alunos, na hipótese de qualquer um deles abandonar o curso. Essa noção está associada ao conceito de [sensibilidade](#), cuja intuição é dada a partir de pares de conjuntos de dados adjacentes que maximizem a diferença entre as saídas de uma função [44, p.15]. No exemplo das médias, o pior caso se daria se todos os alunos tirassem zero, exceto um, que tirou 10 ou, inversamente, de todos os alunos tirassem 10, exceto um, que tirou zero. Ambos os casos resultariam em uma diferença absoluta de 1 ponto entre as médias inicial, com 10 alunos, e final, com 9. Complementando o exemplo numérico¹⁷, sem risco de revelar a nota de qualquer aluno, o professor poderia divulgar os seguintes valores de média [44, p.19]: $\bar{x}_1 = 8,365904769923894$ e $\bar{x}_2 = 8,955108151796736$. Cabe destacar que a privacidade garantida pela PD não é inerente à informação divulgada, mas à função utilizada para produzi-la [44, p.26], como apresentado na Definição 6 (p.24).

2.4.1 Formalização

Formalmente, a Privacidade Diferencial (PD) requer que os resultados produzidos por uma consulta, quando aplicada em pares de conjuntos de microdados adjacentes ou vizinhos, pertençam à mesma distribuição [181, p.258] ou, mais especificamente, estabelece que o resultado de uma consulta é diferencialmente privado se não for possível observar alterações significativas devido à influência (presença) de nenhum indivíduo (registro) específico [44, p.26].

Por exemplo, considere o caso hipotético ilustrado na Figura 2.4. Dois conjuntos de dados, D_1 e D_2 , contêm informações do tempo de tela do mesmo grupo de usuários, exceto por John, presente apenas no conjunto D_1 . Uma função com componente aleatória (e.g., um mecanismo de PD A , no qual o resultado é perturbado com a adição de ruído

¹⁷Neste exemplo, a privacidade diferencial foi aplicada utilizando o mecanismo de Laplace, descrito na Seção 2.4.3, adotando os parâmetros $\Delta f = 1$ e $b = 1$.

aleatório), fixados os parâmetros, foi aplicada múltiplas vezes em ambos os conjuntos de dados, de maneira a produzir uma amostra de resultados para cada um deles. A função aplicada será considerada privada se a diferença entre as distribuições observadas nas duas amostras for suficientemente pequena. Isso dependerá essencialmente de quanto o tempo de tela de John destoa dos demais indivíduos, presentes em ambos os conjuntos de dados. Mais especificamente, a magnitude do ruído deverá ser suficiente para mascarar a contribuição do valor associado a John. A PD estabelece critérios para quantificar e calibrar o ruído de forma a atingir esse objetivo. As definições a seguir buscam formalizar os conceitos e a definição de PD.

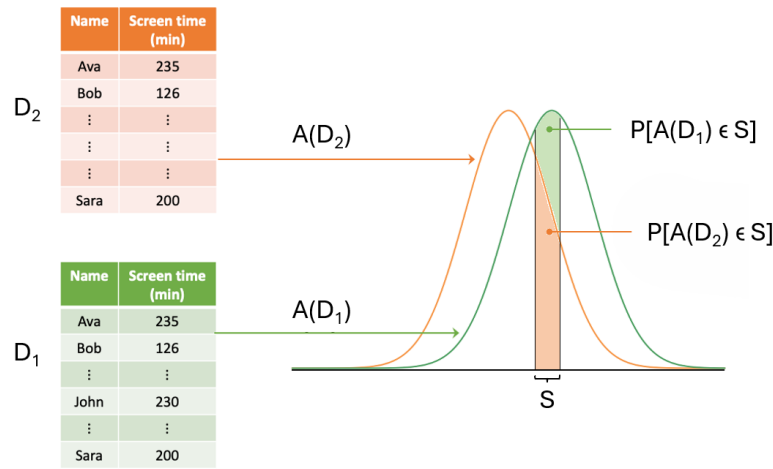


Figura 2.4: Mecanismo A aplicado a dois conjuntos de microdados adjacentes, D_1 e D_2 , produzindo duas distribuições resultantes, sendo S um subconjunto arbitrário de possíveis saídas de $A(D)$. – adaptado de [9].

Definição 1. Curador. O curador [44, p.18] é a entidade com acesso completo ao [conjunto de microdados](#) sensíveis e que tem interesse em permitir a extração de informações a partir dele, zelando, porém, pela confidencialidade ou privacidade.

Definição 2. Métricas de distância. A distância, em termos da métrica M , entre os elementos x_1 e x_2 , pertencentes ao domínio $\mathcal{X}$, é dada por $d_M(x_1, x_2)$ [44, p.55].

Formalmente, uma métrica de distância $d(x_1, x_2)$ é uma função $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ tal que, para quaisquer elementos $x_1, x_2, x_3 \in \mathcal{X}$, as seguintes propriedades se verificam [44, p.303]:

- **não negatividade** $d(x_1, x_2) \geq 0$ e $d(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2$
- **simetria** $d(x_1, x_2) = d(x_2, x_1)$
- **desigualdade triangular** $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$

A métrica de distância mais elementar é a distância absoluta entre dois números reais x_1 e x_2 , que corresponde à diferença absoluta entre eles, dada pela Eq. 2.4 [44, p.55]. Esse conceito pode ser generalizado para pares de registros (objetos) definidos por h atributos através da soma das diferenças absolutas em cada dimensão. Essa métrica é denominada distância de Manhattan e é dada pela Eq. 2.5. Outra métrica de destaque quando se trata da distância entre pares de objetos multidimensionais é a distância euclidiana, dada pela Eq. 2.6.

$$d_{\text{abs}}(x_1, x_2) = |x_1 - x_2| \quad (2.4)$$

$$d_{L_1}(x_1, x_2) = \sum_{l=1}^h |x_1^l - x_2^l| = \|x_1 - x_2\|_1 \quad (2.5)$$

$$d_{L_2}(x_1, x_2) = \sqrt{\sum_{l=1}^h (x_1^l - x_2^l)^2} = \|x_1 - x_2\|_2 \quad (2.6)$$

No contexto de PD, $\mathcal{D}$ define o espaço de todos os conjuntos de microdados possíveis e constitui um caso particular do domínio genérico $\mathcal{X}$, sobre o qual duas métricas de distância são especialmente relevantes para a definição de adjacência entre conjuntos de microdados (Definição 3). Nesse caso, a distância, em termos da métrica M , entre os conjuntos de microdados D_1 e D_2 , pertencentes ao domínio $\mathcal{D}$, é dada por $d_M(D_1, D_2)$. A **distância simétrica** entre dois conjuntos¹⁸ D_1 e D_2 , dada pela Eq. 2.7 [44, p.55], indica o número de objetos que aparecem em apenas um dos conjuntos de microdados, i.e., que não faz parte da interseção entre eles. A **distância de Hamming** entre dois conjuntos de microdados D_1 e D_2 ordenados e de mesmo tamanho n (número de registros ou objetos), dada pela Eq. 2.8 [44, p.57], indica o número de registros que diferem entre eles, sendo $\mathbb{I}[\cdot]$ a função indicadora, que assume valor 1 se a condição entre colchetes for verdadeira e 0 caso contrário.

$$d_{\text{sim}}(D_1, D_2) = |D_1 \triangle D_2| \quad (2.7)$$

$$d_{\text{ham}}(D_1, D_2) = \sum_{i=1}^n \mathbb{I}[D_{1i} \neq D_{2i}] \quad (2.8)$$

¹⁸A distância simétrica considera os conjuntos de microdados como *multisets*, que são conjuntos de microdados que podem ter múltiplas instâncias de cada registro.

Definição 3. Conjuntos de microdados adjacentes. Conjuntos de microdados vizinhos ou adjacentes (e.g., D_1 e D_2) são aqueles que diferem entre si de forma mínima, segundo uma métrica de distância predefinida. Nesse caso, utiliza-se a notação $D_1 \sim D_2$ para indicar que D_1 e D_2 são adjacentes.

Mais especificamente, dois conjuntos de microdados D_1 e D_2 são considerados p -adjacentes em termos da métrica M se $d_M(D_1, D_2) \leq p$ [44, p.57]. A definição de vizinhança imediata pode ser então formalizada adotando a distância simétrica¹⁹ (Eq. 2.7) e $p = 1$, e, portanto, D_1 e D_2 são 1-adjacentes se $d_{\text{sim}}(D_1, D_2) \leq 1$. Isso indica que D_1 e D_2 diferem entre si por apenas um registro²⁰.

Quando o tamanho (número de registros ou objetos) é conhecido²¹ para todos os possíveis conjuntos de microdados, essa definição pode ser feita com base na distância de Hamming (Eq. 2.8) com $p = 1$, e, portanto, D_1 e D_2 são 1-adjacentes se $d_{\text{ham}}(D_1, D_2) \leq 1$.

Definição 4. Característica de interesse. A característica de interesse ψ faz referência a uma informação disponível para os registros do [conjunto de microdados](#) (e.g., sexo, cor ou raça, renda mensal etc.). Corresponde a um atributo do [conjunto de microdados](#).

Definição 5. Consulta. Uma consulta q , composta de uma ou mais funções f , é um algoritmo determinístico²² [59, p.11, 25] que tem como saída uma estatística sobre características de interesse da população representada no [conjunto de microdados](#).

Definição 6. Sensibilidade. Intuitivamente, a sensibilidade mede o quanto a saída de uma função f pode variar quando atributos de um único indivíduo no conjunto de dados são modificados ou removidos. Essa medida é essencial para calibrar a quantidade de ruído necessária para proteger a contribuição individual.

De forma geral, seja $f : \mathcal{D} \rightarrow \mathcal{R}$ e D_1 e D_2 ($D_1, D_2 \in \mathcal{D}$) conjuntos de dados adjacentes, a sensibilidade global de f é definida como:

$$\Delta f = \max |f(D_1) - f(D_2)| \quad (2.9)$$

¹⁹A distância simétrica é substituída pela distância de Hamming quando a diferença entre os conjuntos de microdados se dá não na presença dos registros (e.g., indivíduos), mas na alteração/atualização das informações ou atributos armazenados sobre eles.

²⁰Adotando a distância simétrica, a vizinhança imediata de um [conjunto de microdados](#) é formada pelos conjuntos de microdados que diferem dele pela presença ou ausência de exatamente um registro.

²¹O contexto no qual o tamanho dos conjuntos de microdados é conhecido, é denominado PD limitada (do inglês, *bounded DP*). No caso de PD ilimitada (do inglês, *unbounded DP*), a definição de adjacência baseada na distância simétrica é mais apropriada [44, p.58].

²²Neste contexto, algoritmo determinístico corresponde a um sequência de instruções que não considere componentes aleatórias, i.e., que, dada uma mesma entrada e configuração de parâmetros, produzirá sempre uma mesma saída ou resultado.

Essa definição contempla o pior caso de variação da função entre quaisquer dois conjuntos de dados adjacentes. A sensibilidade global é particularmente importante porque fornece um limite superior para todos os possíveis conjuntos de dados, permitindo a construção de mecanismos com garantias formais independentes de registros específicos. No entanto, essa abordagem pode ser conservadora, uma vez que considera o pior cenário possível.

No exemplo da nota média dos alunos, $f(D_1) = \frac{1}{|D_1|} \sum_{i \in D_1} x_i$ e $f(D_2) = \frac{1}{|D_2|} \sum_{i \in D_2} x_i$, onde x_i representa a nota obtida pelo aluno i . Uma vez que as notas variam no intervalo $[0, 10]$, o pior caso se daria se apenas um aluno tirasse 0, enquanto todos os demais tirassem 10, e justamente o aluno que não pontuou²³ representasse o registro faltante em D_2 . Neste caso, $f(D_1) = \frac{90}{10} = 9$ e $f(D_2) = \frac{90}{9} = 10$ e, portanto, $\Delta f = 1$.

Definição 7. Mecanismo. Um mecanismo A é um algoritmo aleatorizado (ou estocástico, não determinístico) que, a partir de uma consulta a um conjunto de dados, produz como saída o resultado da consulta adicionado de ruído gerado aleatoriamente [59, p.16].

Definição 8. Unidade de privacidade. A unidade de privacidade (do inglês, *privacy unit*) v_i corresponde à entidade à qual se deseja garantir privacidade²⁴ (e.g., indivíduo, organização, domicílio etc.) [44, p.27]. Geralmente está associada à unidade de análise representada nos registros do conjunto de dados.

Definição 9. Perda de privacidade. Seja y a saída de uma consulta q feita ao conjunto de dados D e perturbada pelo mecanismo A , i.e., $A(q, D) = y$. Um intruso pode tentar violar a privacidade de uma unidade v_i investigando a probabilidade de y ser observado na presença ou não de v_i nos conjuntos de dados adjacentes D_1 e D_2 . Em outras palavras, se a saída y foi produzida a partir da aplicação de A em D_1 ($P[A(D_1) = y]$), que inclui o registro de v_i , ou em D_2 ($P[A(D_2) = y]$), que difere do anterior apenas por não incluir o registro de v_i ($D_1 \sim D_2$). Assim, a divergência entre essas duas probabilidades (e.g., Figura 2.4) oferece uma medida do poder de distinção entre os dois casos e, portanto, de o quão confiante o intruso pode ser na escolha que fizer, i.e., se o resultado y , produzido pela consulta q e perturbado pelo mecanismo A , inclui ou não o registro referente a v_i . A **perda de privacidade** (do inglês, *privacy loss*) pode então ser quantificada pela razão entre as probabilidades dos dois eventos [44, p.28], dada por

²³Analogamente, o pior caso também se verificaria se todos os alunos tirassem 0, exceto um, que obteve nota 10 e, justamente esse, estivesse ausente em D_2 .

²⁴Naturalmente, quanto maior a influência de uma unidade de privacidade no conjunto de dados a qual pertence, i.e., quanto mais ela se destacar dos demais registros no que diz respeito à característica de interesse, maior será a amplitude do ruído a ser adicionado para mascarar sua presença ou ausência na saída das consultas realizadas.

$$L(A(D_1), A(D_2), y) = \ln \left(\frac{P[A(D_1) = y]}{P[A(D_2) = y]} \right) \quad (2.10)$$

Se $L(A(D_1), A(D_2), y) = 0 \implies P[A(D_1) = y] = P[A(D_2) = y]$ e o intruso não obtém nenhuma vantagem. Se $L(A(D_1), A(D_2), y) > 0 \implies P[A(D_1) = y] > P[A(D_2) = y]$, sugerindo ao intruso que a consulta foi aplicada em D_1 ²⁵, i.e., que o resultado contempla informação sobre v_i . Finalmente, se $L(A(D_1), A(D_2), y) < 0 \implies P[A(D_2) = y] > P[A(D_1) = y]$, induzindo o intruso a acreditar que a consulta foi aplicada em D_2 , no qual v_i está ausente.

Definição 10. Privacidade Diferencial. Formalmente [59, p.17], a PD diz respeito à aplicação de um mecanismo $A : \mathcal{D} \rightarrow \mathcal{R}$ que satisfaz (ϵ, δ) -PD se, para quaisquer conjuntos de dados adjacentes $D_1 \sim D_2$ e para todo $S \subseteq \mathcal{R}$:

$$P[A(D_1) \in S] \leq e^\epsilon P[A(D_2) \in S] + \delta \quad (2.11)$$

Sendo:

- A é um mecanismo²⁶ (algoritmo) aleatorizado.
- $\mathcal{D}$ o conjunto de todas as bases de dados possíveis sobre um domínio X , $\mathcal{D} = \bigcup_{n \geq 0} X^n$, i.e., corresponde ao conjunto de todos os *multisets* finitos de elementos de X .
- $\mathcal{R}$ o contradomínio de A ²⁷.
- S um subconjunto arbitrário de todas as possíveis saídas (imagem) de $A(D)$.
- ϵ o parâmetro que quantifica o nível de privacidade²⁸ garantido pelo mecanismo.
- δ a probabilidade de falha²⁹ do mecanismo [54, p.12].

²⁵A perda de privacidade tende a ser sempre positiva quando $A(D_1) = y$, caso contrário não haveria utilidade [44, p.29]. Uma violação de privacidade só ocorre se a perda de privacidade for muito grande em magnitude, mais especificamente se $L(A(D_1), A(D_2), y) > \epsilon$, como formalizado na Definição 10.

²⁶Um mecanismo de PD pode ser formalmente entendido como uma variável aleatória que associa um conjunto de dados de entrada a uma distribuição de probabilidades sobre possíveis saídas, de modo que execuções distintas sobre a mesma entrada possam produzir resultados diferentes em razão da aleatoriedade introduzida através do ruído.

²⁷Em função das consultas e mecanismos adotados, tipicamente $\mathcal{R} \subset \mathbb{R}$.

²⁸Controla o quanto as distribuições de saída podem divergir quando o conjunto de dados é modificado em um único registro. Valores menores implicam maior perturbação (menor utilidade) e garantias mais fortes de privacidade.

²⁹Probabilidade de que A produza saídas com diferenças mais significativas a partir de conjuntos de dados adjacentes, i.e., que a condição de privacidade diferencial seja descumprida.

O parâmetro ϵ , também denominado *privacy budget*, é limitado inferiormente em zero, configuração em que a utilidade tende a ser nula³⁰ e o analista não aprende nada a partir do resultado da consulta. Na analogia com a segurança semântica, seria o equivalente a tratar o analista como a parte ilegítima. Não há limite superior teórico para o parâmetro ϵ , porém, quanto maior o valor de ϵ , menor é o ruído adicionado e, consequentemente, maior é o risco de revelação. Novamente, na analogia com o contexto de criptografia, a definição de um valor de ϵ muito elevado³¹ equivaleria a tratar o analista como parte receptora, i.e., com acesso a toda a informação. O conceito de *privacy budget* representa um limite superior para a perda de privacidade cumulativa e oferece previsibilidade quanto à degradação da garantia de privacidade em cenários em que um mesmo conjunto de microdados é objeto de múltiplas consultas [119, p.765].

Quanto ao parâmetro δ , tipicamente assume valores menores que o inverso de qualquer polinômio no tamanho do conjunto de dados³². Se $\delta = 0$, diz-se que A é $(\epsilon, 0)$ -diferencialmente privado, ou satisfaz ϵ -PD, o que corresponde à chamada privacidade diferencial pura.

A ϵ -PD garante que a probabilidade de observar qualquer saída de um dado mecanismo A é aproximadamente a mesma³³ para qualquer par de conjuntos de microdados adjacentes [44, p.29]. Quando $\delta = 0$, a definição dada pela Eq. 2.11, segundo leitura proposta por [20, p.2], pode ser interpretada intuitivamente como a afirmação de que qualquer analista que observar o resultado da execução de A em D_1 não poderá inferir a presença de um indivíduo em D_1 com mais de $100 \times \left| 1 - \frac{1}{\exp(\epsilon)} \right|$ % de confiança ou, de forma mais rigorosa, que a perda de privacidade é limitada superiormente por ϵ , i.e., $L(A(D_1), A(D_2), y) \leq \epsilon$ [44, p.29]. Essa garantia é robusta e independe de aspectos como fontes auxiliares às quais o intruso tenha ou possa vir a ter acesso. Entretanto, a depender da distribuição de D , satisfazer ϵ -PD pode implicar perda de utilidade expressiva.

³⁰Nesta configuração, a depender do mecanismo de geração do ruído, o ruído terá a mesma ordem de grandeza do resultado da consulta, anulando sua utilidade.

³¹A título de referência, para $\epsilon = 5$, a razão de chances dada pela Eq. 2.10 é superior a 140, indicando que o adversário pode se sentir extremamente confiante ao tentar distinguir entre dois conjuntos de dados vizinhos [44, p.33]. Embora a definição do parâmetro ϵ seja dependente do contexto, de maneira geral $0,01 \leq \epsilon \leq 1$ é considerado um intervalo adequado para dados sensíveis ou quando se deseja priorizar a proteção da privacidade, ao passo que $1 < \epsilon \leq 10$ é adotado em contextos menos sensíveis, quando se deseja priorizar a utilidade.

³²Geralmente δ é escolhido como um valor desprezível em função do tamanho do conjunto de dados, e.g., $\delta = 1/n^2$. Valores de δ da ordem de $\frac{1}{n}$ são perigosos, pois garantem a preservação da privacidade da enorme maioria dos registros, mas implicam divulgação dos registros completos de um pequeno número de indivíduos do conjunto de microdados [59, p.18], a exemplo da abordagem “*just a few*” [59, p.9].

³³Mais especificamente, garante que, entre bases adjacentes, as saídas diferem por no máximo um fator multiplicativo e^ϵ .

O parâmetro δ representa uma relaxação dessa garantia, sendo tipicamente adotado com $\delta > 0$ quando ϵ -PD implica perda de utilidade considerada inaceitável. Em outras palavras, o parâmetro δ representa a probabilidade de que possa haver uma diferença significativa nas probabilidades para alguns pares de conjuntos de microdados adjacentes em algumas saídas do mecanismo. Por isso, a privacidade diferencial (ϵ, δ) , também chamada de privacidade diferencial aproximada, é considerada uma relaxação, permitindo uma pequena probabilidade δ de falha.

Em ϵ -PD, as distribuições de probabilidade são controladas simultaneamente em todos os conjuntos de microdados adjacentes, fornecendo uma garantia de privacidade muito forte. Já em (ϵ, δ) -PD, a garantia é condicional. Para a maioria dos pares de microdados adjacentes e para a maioria dos resultados, as probabilidades são controladas, mas há uma pequena probabilidade δ de que, para alguns pares e resultados específicos, a garantia possa não ser tão rigorosa. Isso implica que, após observar uma saída, há uma pequena chance³⁴ de que se possa encontrar um [conjunto de microdados](#) adjacente onde essa saída seja muito mais provável. Chance essa que é nula sob ϵ -PD.

Por exemplo, para $\epsilon = 0,1$ e $\delta = 0$, temos:

$$P[A(D_1) \in S] \leq \exp(0,1)P[A(D_2) \in S] + 0$$

$$P[A(D_1) \in S] \leq 1,105P[A(D_2) \in S]$$

Ao passo que, para $\epsilon = 0,1$ e $\delta = 0,01$, temos:

$$P[A(D_1) \in S] \leq \exp(0,1)P[A(D_2) \in S] + 0,01$$

$$P[A(D_1) \in S] \leq 1,105P[A(D_2) \in S] + 0,01$$

Isso significa que, para os resultados obtidos para pares de conjuntos de microdados adjacentes, as probabilidades estarão dentro do fator $\exp(0,1)$, mas há uma pequena chance ($\delta = 0,01$) de que isso não seja estritamente válido.

No contexto do problema de agrupamento diferencialmente privado (Seção 2.5), ao aplicar um algoritmo sobre dois conjuntos adjacentes D_1 e D_2 , espera-se que a alocação dos objetos e as coordenadas dos centroides sejam muito semelhantes em ambos, de modo que um adversário não consiga inferir, com elevada confiança, se um determinado indivíduo está representado ou não nas soluções de agrupamento. Quando $\delta = 0$, essa limitação é estritamente válida para todas as possíveis saídas do mecanismo. No exemplo

³⁴Essa propriedade é vista como uma “garantia posterior ao fato” (*ex post facto*) [59, p.18].

apresentado para $\epsilon = 0,1$, isso significa que qualquer conjunto de centroides produzidos a partir de D_1 não poderá ser mais que aproximadamente 1,105 vezes mais provável do que a saída produzida a partir de D_2 . Por outro lado, em (ϵ, δ) -PD, admite-se uma pequena probabilidade δ de ocorrência de saídas excepcionais para as quais essa limitação pode não ser estritamente respeitada. Isso pode corresponder, por exemplo, a situações raras em que o ruído adicionado pelo mecanismo resulta em centroides onde a influência de um indivíduo específico ainda pode ser isolada.

2.4.2 Modelos

No que se refere ao estágio de geração, produção ou coleta do dado no qual a PD é aplicada, os dois modelos tipicamente adotados são local e global, representados na Figura 2.5. No modelo local, a camada de privacidade é introduzida na origem (e.g., no momento da coleta ou geração do dado), de modo que o dado é armazenado e processado já perturbado, i.e., modificado pela adição de ruído. Esse modelo é adotado quando os dados são coletados, gerados e/ou processados de forma distribuída e não há confiança na entidade responsável por centralizá-los ou quando essa entidade não está disposta a gerenciar o risco intrínseco a administrar dados sensíveis. Requer, porém, significativamente mais dados para satisfazer a mesma garantia de privacidade do modelo global e, por isso, tende a ser usado em contextos em que há uma quantidade muito grande de dados [64, 139], o que pode compensar a perda de utilidade [44, p.34].

No modelo global, os dados são coletados (ou consolidados) e armazenados por um curador confiável e, portanto, armazenados sem perturbação, resultando em utilidade plena para o curador. A camada de privacidade é aplicada em momento posterior, como previamente à divulgação do resultado de uma consulta, por exemplo.

No que diz respeito à forma de interação do analista com o conjunto de dados, o que corresponde à etapa de disseminação, a implementação de uma solução baseada em PD pode se dar essencialmente através dos modelos não interativo (ou *offline*) ou interativo (ou *online*).

Seja D um conjunto de dados contendo n registros individualizados e possivelmente identificados, descritos por h atributos (não perturbados). No modelo não interativo ou *offline* [59, p.11], o curador produz, a partir de D , um objeto D^* tal como um **conjunto de microdados** sintético³⁵, uma coleção de estatísticas resumo (e.g., tabelas, estatísticas

³⁵Ao contrário dos métodos tradicionais de SDC para geração de microdados sintéticos, neste contexto o conjunto gerado não necessariamente reproduz o *layout* de D , podendo consolidar, agregar ou até

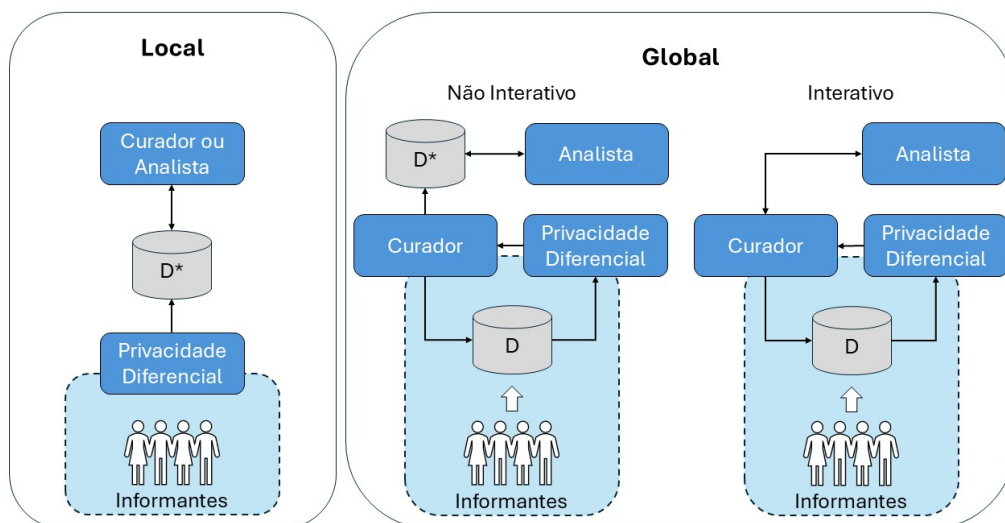


Figura 2.5: Ilustração dos modelos de PD – adaptado de [46].

descritivas etc.) ou um **conjunto de microdados** protegido (do inglês, *sanitized*). Uma vez divulgado D^* , não tem mais participação no processo, i.e., não há intermediação do acesso do analista a D^* , e o analista em nenhum momento interage com D . Alguns exemplos de abordagens baseadas nesse modelo são [58, 16, 24].

O modelo interativo ou *online* [59, p.11], por sua vez, permite que um analista, por intermédio do curador, execute um conjunto de T consultas sucessivas $Q = \{q_1, \dots, q_t, \dots, q_T\}$ ao conjunto de dados D , de maneira que cada consulta q_t possa oferecer informações para compor a consulta seguinte q_{t+1} , permitindo uma abordagem investigativa personalizada. Nesse caso, o curador deve sempre intermediar o acesso do analista, que recebe as respostas às suas consultas perturbadas por um mecanismo de PD.

Quando o universo de todas as consultas possíveis de serem executadas em D é determinado e conhecido a priori³⁶, o modelo não interativo tende a oferecer resultados com menor perturbação [59, p.12]. Entretanto, quando esse não é o caso, o modelo não interativo é praticamente inviabilizado, uma vez que precisaria oferecer respostas para todas as possíveis consultas. Além disso, a utilidade (no que se refere ao ruído adicionado) é deteriorada progressivamente em função do número de consultas feitas [59, p.12].

Abordagens baseadas em conceitos similares aos do modelo não interativo, independentemente do uso de PD, são bem consolidadas entre diversas instituições produtoras de dados, especialmente entre Institutos Nacionais de Estatística (INE). Após a realização

mesmo duplicar informações.

³⁶Isso significa dizer que todas as informações que serão produzidas a partir do **conjunto de microdados** são previamente determinadas e, além disso, que as consultas correspondentes já foram elaboradas.

de uma operação estatística, os dados coletados são submetidos a diversas etapas de crítica e tratamentos estatísticos, resultando em um [conjunto de microdados](#) representativo da população de interesse, que, por sua vez, dá origem a produtos estatísticos, que são disponibilizados para analistas, pesquisadores e o público em geral. Esses produtos, no entanto, eventualmente não atendem a necessidades mais específicas, como é geralmente o caso de pesquisas acadêmicas, por exemplo.

Nesses casos, é necessário adotar abordagens baseadas em conceitos do modelo interativo. Nos INE, isso tradicionalmente é feito através do uso de salas de acesso seguro ou da assinatura de acordos de confidencialidade, ambas alternativas que demandam substancial carga burocrática. Os acordos são instrumentos específicos, discutidos caso a caso e geralmente celebrados entre o INE e instituições parceiras. O procedimento para uso das salas de acesso seguro geralmente inicia com a submissão de um projeto por parte do analista interessado. Uma vez aprovado o projeto, é concedido ao analista o acesso ao conjunto de dados através do ingresso à sala de acesso seguro. Ele ou ela então constrói e submete um conjunto de consultas. Porém, os resultados das consultas realizadas só ficam acessíveis ao analista após uma auditoria ser realizada por equipe específica, a fim de avaliar eventuais violações às políticas de confidencialidade da instituição. Esse processo é custoso tanto para o curador quanto para o analista, e pode produzir resultados de baixa utilidade, a depender dos métodos empregados.

A proposta de incorporar PD oferece garantia mais robusta, permitindo alcançar melhor equilíbrio entre privacidade e utilidade, embora essa afirmação não seja consenso entre pesquisadores da área [51, 22]. Além de aprimorar substancialmente a experiência do analista, o melhor equilíbrio nesse *trade-off* representa maior robustez e segurança no que se refere à privacidade e economia de recursos humanos e de tempo para a instituição detentora do dado.

A PD permite ainda viabilizar aplicações específicas e de grande relevância, tais como monitoramento de vírus e gestão de desastres. Aplicações como essas envolvem consultas contínuas de informações de localização, o que é desafiador do ponto de vista da garantia da privacidade, mas abordagens baseadas em PD oferecem solução satisfatória para o problema [87].

De maneira mais geral, a PD viabiliza o acesso a conjuntos de dados de maneira mais dinâmica e segura. Na era do *Big Data* e da inteligência artificial, onde a boa performance dos modelos é cada vez mais dependente de grandes volumes e múltiplas fontes de dados, tecnologias como essa são de fundamental importância. Em [159], os autores apresentam

uma revisão abrangente sobre *deep learning* diferencialmente privado, abordando tanto ataques à privacidade quanto mecanismos de proteção.

2.4.3 Mecanismos

Mecanismos de PD são algoritmos aleatorizados que atuam retornando a saída de uma função f adicionada de ruído gerado (amostrado) segundo uma distribuição de probabilidade conhecida, cujo parâmetro de posição central é o zero e o parâmetro de escala é calibrado de acordo com a sensibilidade de f . Os mecanismos alcançam privacidade por meio de perturbação do valor original da saída da consulta [44, p.37]. A ideia é baseada em garantir que as distribuições de probabilidade usadas para gerar essas saídas permaneçam suficientemente semelhantes quando o mecanismo é aplicado a qualquer par de conjuntos de microdados adjacentes³⁷. Essa condição limita a confiança com a qual um analista pode inferir sobre a presença ou ausência de qualquer indivíduo no [conjunto de microdados](#), pois a probabilidade de obter uma saída específica permanece aproximadamente a mesma, independentemente de os dados de um indivíduo específico estarem incluídos.

Neste contexto, há alguns mecanismos clássicos na literatura de PD, sendo o mecanismo de resposta aleatorizada, já mencionado, o mais elementar. Esse mecanismo é apropriado para variáveis dicotômicas e é detalhadamente descrito no Apêndice B.1 (p.151). O mecanismo exponencial [140][59, p.37-39] é utilizado para seleção em um conjunto limitado de candidatos e pode ser facilmente aplicado a dados categóricos, conforme exemplificado no Apêndice B.2 (p.153). O mecanismo geométrico [44, p.93-94] se destina a saídas numéricas inteiras e os mecanismos Normal (Gaussiano) [59, p.261-266][44, p.123-126] e de Laplace, descritos a seguir e adotados na abordagem proposta nesta tese, se destinam a saídas numéricas reais^{38 39}. Há ainda uma variedade de outros mecanismos disponíveis na literatura [146, 160, 144, 85, 84, 140].

Entre os mecanismos fundamentais destacam-se o mecanismo de Laplace e o mecanismo Normal, que diferem tanto na métrica de sensibilidade adotada quanto na garantia

³⁷Ao controlar a divergência entre as distribuições, é possível garantir que a informação de nenhum indivíduo seja exposta, preservando sua privacidade.

³⁸O mecanismo Normal tende a exigir muito menos ruído para o mesmo nível de privacidade ao responder a um grande número de consultas, em comparação ao mecanismo de Laplace com composição avançada [44, p.123]. Entretanto, o mecanismo Normal não satisfaz ϵ -DP [44, p.124] e, sob (ϵ, δ) -DP, os benefícios da composição são em grande parte desperdiçados [44, p.126].

³⁹Cabe destacar que as formulações clássicas de mecanismos como os de Laplace e Normal assumem aritmética real. Em implementações práticas, entretanto, a geração de ruído depende de algoritmos numéricos e da representação finita de números em ponto flutuante, o que pode introduzir vulnerabilidades. Em particular, ataques explorando a representação em ponto flutuante de algoritmos de amostragem podem comprometer as garantias formais de PD [44, p.125][117].

de privacidade fornecida.

O mecanismo de Laplace constitui uma das construções mais elementares para garantir privacidade diferencial pura, i.e., ϵ -PD.

Seja $f : \mathcal{D} \rightarrow \mathcal{R}^h$ e D_1 e D_2 conjuntos de dados adjacentes ($D_1, D_2 \in \mathcal{D}$), a sensibilidade global de f em norma ℓ_1 é definida como:

$$\Delta f = \max_{D_1, D_2} \|f(D_1) - f(D_2)\|_1$$

Essa quantidade mede a máxima variação na saída de f decorrente da modificação (ou remoção/inclusão) de um único registro no conjunto de dados.

O mecanismo de Laplace A_L associado a f , com parâmetro de privacidade ϵ , é definido por:

$$A_L(D) = f(D) + (Y_1, \dots, Y_h)$$

onde Y_i é uma variável aleatória independente extraída da distribuição de Laplace centrada em zero com parâmetro de escala $b = \frac{\Delta f}{\epsilon}$, cuja densidade de probabilidade é dada por:

$$\text{Lap}(y \mid b) = \frac{1}{2b} \exp\left(-\frac{|y|}{b}\right)$$

Assim, o mecanismo de Laplace adiciona ruído independente a cada coordenada da saída de f . A calibração do ruído pela razão $\frac{\Delta f}{\epsilon}$ garante que a influência de qualquer registro na saída do mecanismo seja limitada. Formalmente, o mecanismo de Laplace satisfaz ϵ -PD, i.e., para qualquer par de conjuntos de dados adjacentes D_1 e D_2 , e qualquer subconjunto arbitrário S de possíveis saídas $A_L(D)$ [59, p.32]:

$$P[A_L(D_1) \in S] \leq e^\epsilon P[A_L(D_2) \in S]$$

O mecanismo Normal, por sua vez, generaliza a ideia de adição de ruído ao contexto de privacidade diferencial aproximada, i.e., (ϵ, δ) -PD.

Seja $f : \mathcal{D} \rightarrow \mathcal{R}^h$ e D_1 e D_2 conjuntos de dados adjacentes ($D_1, D_2 \in \mathcal{D}$), a sensibilidade global⁴⁰ de f em norma ℓ_2 é definida como:

⁴⁰A escolha da norma na definição de sensibilidade está relacionada à distribuição de ruído utilizada,

$$\Delta_2 f = \max_{D_1, D_2} \|f(D_1) - f(D_2)\|_2$$

O mecanismo Normal A_N é definido como:

$$A_N(D) = f(D) + Z$$

onde $Z \sim \mathcal{N}(0, \sigma^2 I_h)$, i.e., um vetor aleatório com distribuição normal multivariada de média zero e matriz de covariância $\sigma^2 I_h$.

A escolha do parâmetro σ é fundamental para garantir privacidade. Em particular, dados os parâmetros $\epsilon \in (0, 10)$ e $\delta > 0$, o mecanismo Normal satisfaz (ϵ, δ) -PD sempre que [59, p.53]:

$$\sigma \geq c \frac{\Delta_2 f}{\epsilon}$$

sendo c uma constante tal que $c^2 > 2 \ln(1,25/\delta)$.

Assim, diferentemente do mecanismo de Laplace, o mecanismo Normal não garante privacidade pura, mas sim uma versão relaxada que permite uma pequena probabilidade δ de violação da garantia de privacidade.

Os conceitos, modelos e mecanismos apresentados fundamentam o desenvolvimento de métodos diferencialmente privados, nos quais a proteção da privacidade é obtida por meio da adição de ruído. Na Seção 2.5, esses fundamentos foram aplicados ao problema de agrupamento diferencialmente privado, com ênfase no DPk-means, considerando especificamente o emprego do mecanismo de Laplace para perturbação de consultas referentes à atualização dos centroides.

2.5 O problema de agrupamento diferencialmente privado

O uso de métodos de agrupamento em dados sensíveis, mesmo em aplicações legítimas, pode levar à exposição de informações confidenciais, como descrito na Seção 2.3 através do estudo hipotético que identificou um grupo com alta probabilidade de desenvolver uma doença grave com base em atributos socioeconômicos. Resultados como esse podem conforme descrito no Apêndice B.5, p.164.

orientar políticas públicas de prevenção e tratamento com grande impacto social, mas também podem ser usados por operadoras de planos de saúde para recusar ou rescindir contratos de indivíduos específicos, por exemplo.

Como apresentado na Seção 2.4, a PD consiste em perturbar os dados introduzindo um ruído controlado no momento da coleta ou geração do dado, no caso dos modelos local e global não interativo, ou a cada consulta realizada, no caso do modelo global interativo, a fim de impedir diferenciar ou isolar qualquer registro individual. Essa estratégia pode ser incorporada a qualquer abordagem de análise aplicada a um [conjunto de microdados](#), desde a extração de estatísticas e estimativas [56] até a aplicação de técnicas sofisticadas de *deep learning* [159].

Dada a relevância dos métodos de agrupamento, que viabilizam a extração de padrões em dados, aspecto crucial em aplicações analíticas e de apoio à decisão, torna-se evidente a importância da pesquisa em algoritmos de agrupamento que incorporem abordagem baseada em privacidade diferencial, a fim de permitir que tais análises sejam realizadas em dados sensíveis, porém com risco de revelação controlado. Nesse contexto, a incorporação de PD ao Problema de Agrupamento depende do método e algoritmo adotados, sendo necessário definir, além do modelo de PD, a calibração do ruído e a etapa do processo em que ele será introduzido.

Dentre as muitas abordagens propostas nas últimas décadas para o Problema de Agrupamento Clássico, i.e., sem PD, o k-means, apesar de limitações conhecidas⁴¹, permanece como uma das principais escolhas, devido à sua simplicidade, eficiência computacional e versatilidade em diversas aplicações. Essas características o tornam especialmente atraente em cenários com grandes volumes de dados, justificando o interesse contínuo em seu uso e aprimoramento [161, p.2].

O k-means consiste em um algoritmo que, ao ser aplicado, divide um conjunto D de n objetos x_i , $i \in \{1, \dots, n\}$, descritos por seus h atributos, em k grupos C_j , $j \in \{1, \dots, k\}$, de forma a minimizar a dissimilaridade entre objetos do mesmo grupo e maximizar a dissimilaridade entre objetos de grupos distintos. Preliminarmente à execução desse algoritmo, o conjunto de dados tem seus atributos normalizados a fim de evitar que atributos de escalas maiores dominem aqueles de escalas menores. A primeira etapa consiste então na inicialização dos centroides, i.e., definir os k centroides $\{c_1, \dots, c_j, \dots, c_k\}$ a serem considerados na primeira iteração⁴². Com isso se dá início a um processo iterativo que

⁴¹A sensibilidade a *outliers* e a dificuldade em capturar estruturas não esféricas são as principais limitações do k-means apontadas na literatura [161, p.2].

⁴²A inicialização dos centroides pode ser feita, por exemplo, selecionando aleatoriamente k objetos

consiste de duas etapas, quais sejam, (i) alocar cada objeto ao grupo no qual o centroide correspondente apresenta menor dissimilaridade; e (ii) atualizar os centroides a partir da nova composição dos grupos.

Para a etapa de alocação, uma métrica de dissimilaridade é adotada a fim de identificar, para cada um dos n objetos, o centroide de menor dissimilaridade. Assumindo a distância euclidiana quadrática⁴³ como função de dissimilaridade, a distância entre o centroide do grupo C_j , definido por c_j , e o objeto x_i é dada pela Eq. 2.12 [100, p.461].

$$d_{L_2}^2(c_j, x_i) = \sum_{l=1}^h (c_j^l - x_i^l)^2 = \|c_j - x_i\|^2 \quad (2.12)$$

No que se refere à atualização dos centroides, eles representam o vetor de médias dos atributos relativos aos objetos que compõem os grupos e, portanto, devem ser atualizados a cada modificação na composição do grupo correspondente. Para todo $j \in \{1, \dots, k\}$, as coordenadas do centroide $c_j = (c_j^1, c_j^2, \dots, c_j^l, \dots, c_j^h)$ são calculadas, na abordagem clássica (sem perturbação), através da razão entre a soma dos valores dos atributos dos objetos atribuídos ao grupo C_j (S_j) e a contagem (frequência) de objetos em C_j ($|C_j|$), conforme Eq. 2.13.

$$c_j^l = \frac{S_j^l}{|C_j|} = \frac{\sum_{x_i \in C_j} x_i^l}{|C_j|}, \quad \forall l \in \{1, \dots, h\} \quad (2.13)$$

Uma vez que se trata de um método iterativo, as etapas de alocação dos objetos e atualização dos centroides são repetidas até que um critério de parada seja satisfeito⁴⁴. A fim de direcionar o processo, é definida uma função objetivo, responsável por medir a qualidade de cada solução obtida, e que é calculada ao final de cada iteração⁴⁵.

Uma função objetivo deve refletir os aspectos de interesse na aplicação do algoritmo, traduzindo em termos quantitativos os critérios utilizados para orientar a busca por solu-

como centroides iniciais ou distribuindo aleatoriamente os n objetos que compõem o conjunto de dados em k grupos e calculando os centroides correspondentes.

⁴³A distância euclidiana quadrática não constitui uma métrica de distância em sentido formal, pois não satisfaz a propriedade de desigualdade triangular (ver Definição 2, p.22). No entanto, seu uso no k-means permanece adequado, sustentado pela propriedade de que a média minimiza a soma dos quadrados dos desvios [114, p.654], independentemente das propriedades formais de uma métrica. A escolha de $d_{L_2}^2$ (Eq. 2.12), em detrimento de d_{L_2} (Eq. 2.6, p.23), evita o custo computacional da raiz quadrada, vantagem relevante dado o caráter iterativo do k-means.

⁴⁴São exemplos de critérios de parada o número máximo de iterações e a estabilidade da solução (e.g., nenhuma alteração na alocação dos objetos em determinado número de iterações consecutivas).

⁴⁵Nesse caso, o critério de parada pode ser estabelecido com base em um valor alvo da função objetivo ou no incremento (ou decremento, no caso de minimização) alcançado em iterações sucessivas.

ções que melhor atendam aos objetivos do problema. Para o k-means, o principal aspecto é a homogeneidade interna, i.e., minimizar a dissimilaridade entre objetos alocados ao mesmo grupo⁴⁶.

A função objetivo do k-means, a ser minimizada, é dada pela Eq. 2.14 [100, p.462].

$$f_{\text{obj}_1} = \sum_{j=1}^k \sum_{x_i \in C_j} d_{L_2}^2(c_j, x_i) \quad (2.14)$$

Complementarmente, deve-se considerar restrições que garantam que todos os objetos foram alocados a exatamente um grupo (Eq. 2.15 e Eq. 2.16) e que todo grupo tenha ao menos um objeto (Eq. 2.17) [225, p.12].

$$\bigcup_{j=1}^k C_j = D \quad (2.15)$$

$$C_i \cap C_j = \emptyset, \quad \forall i \neq j \quad (2.16)$$

$$C_j \neq \emptyset, \quad j \in \{1, \dots, k\} \quad (2.17)$$

As etapas anteriores definem a versão clássica do k-means. Porém, quando aplicado a conjuntos de dados sensíveis, o processo de atualização dos centroides oferece uma potencial vulnerabilidade de privacidade [23, p.5], uma vez que os centroides e as estatísticas utilizadas no seu cálculo podem contribuir na revelação de informações sobre indivíduos específicos representados nos dados, conforme detalhado na Seção 2.3, justificando o uso de PD. Uma variação do método que satisfaz PD consiste em introduzir ruído ao final da etapa de atualização dos centroides, perturbando o cálculo de suas coordenadas. Dada a natureza iterativa do método, no qual a atualização contínua dos centroides e a realocação dos objetos dependem dos resultados das iterações anteriores, a escolha mais direta em termos do modelo de PD é o global iterativo⁴⁷, embora seja possível explorar variantes

⁴⁶Outro aspecto relevante para o k-means é maximizar a dissimilaridade entre objetos alocados em grupos distintos. Ambos os aspectos podem ser quantificados conjuntamente através do índice de Calinski-Harabasz [30][44, p.166], dado por $CH = \frac{BCSS \cdot (n-k)}{WCSS \cdot (k-1)} = \frac{\sum_{j=1}^k |C_j| \|c_j - c\|^2 \cdot (n-k)}{\sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2 \cdot (k-1)}$, onde c corresponde ao centroide global, i.e., às médias dos atributos de todos os objetos.

⁴⁷O modelo global iterativo responde a consultas de forma sequencial e adaptativa, sendo mais indicado para cenários em que as consultas são desconhecidas ou dependem de resultados anteriores [59, p.11], como em análises exploratórias ou iterativas.

do k-means tanto no modelo local, tipicamente na presença de dados distribuídos, quanto no global não interativo, como apresentado na Seção 3.1.1.

O modelo interativo permite incorporar mecanismos de privacidade diferencial de forma adaptativa ao longo das iterações, possibilitando o controle do ruído através da distribuição do *privacy budget* (ϵ), enquanto se preserva a dinâmica de refinamento da solução de agrupamento. Isso pode ser feito incorporando o mecanismo de PD na etapa de atualização dos centroides e o vetor de coordenadas do centroide j perturbado, denotado por c_j^* , é dado por [129, p.3]:

$$c_j^* = \frac{S_j^*}{|C_j^*|} = \frac{\sum_{x_i \in C_j} x_i + Y_S}{|C_j| + Y_C}$$

Sendo $Y_S \sim \text{Lap}(\Delta f^s / \epsilon^s)$ o ruído adicionado à soma dos atributos e $Y_C \sim \text{Lap}(\Delta f^c / \epsilon^c)$ o ruído adicionado à frequência dos objetos no grupo j . Dessa forma, o ruído adicionado à consulta de soma segue uma distribuição de Laplace com parâmetro de escala $b = \Delta f^s / \epsilon^s$, sendo Δf^s a sensibilidade da consulta de soma e ϵ^s a parcela de *privacy budget* alocada a essa consulta. A distribuição do ruído adicionado à consulta de frequência dos objetos é definida de maneira análoga.

Uma vez que essa etapa é repetida a cada iteração t , $t \in \{1, \dots, T\}$, a consulta referente à atualização dos centroides é executada T vezes e, portanto, o *privacy budget* (ϵ) deve ser distribuído entre as T iterações. Esse aspecto é formalizado e explorado de forma mais aprofundada no Capítulo 4, mas as Figuras 2.6 e 2.7 oferecem uma percepção do impacto da distribuição dos dados e da calibração do ruído na solução de agrupamento obtida.

Em específico, essas figuras apresentam comparação entre as soluções obtidas através do algoritmo de Lloyd [131] – que corresponde a uma das variantes do k-means – com e sem perturbação dos centroides, respectivamente, para identificação de 3 grupos de um conjunto hipotético de objetos descritos por 2 atributos, gerados segundo distribuição normal com $\sigma = 1$ e $\mu = \{0, 5, 10\}$, normalizados no intervalo $[-1, 1]$. Os centroides puros de cada grupo são identificados por algarismos em preto. Os algarismos em vermelho indicam a posição resultante do respectivo centroide após aplicação de PD através do mecanismo de Laplace. Por se tratar de grupos esféricos e bem separados⁴⁸, a perturbação

⁴⁸A presença de estruturas de grupos nos conjuntos de dados pode ser previamente identificada, i.e., antes da execução de um algoritmo de agrupamento, através de técnicas de avaliação de tendência de agrupamento [171]. Mais especificamente, a estatística de Hopkins [15, 167][94, p.485] oferece critério objetivo para avaliar se a distribuição dos objetos de um conjunto de dados é aleatória ou não.

aplicada não comprometeu a convergência do algoritmo, no sentido de que as atualizações sucessivas dos centroides ainda foram capazes de conduzir o processo iterativo na direção da solução de agrupamento esperada. Assim, mesmo com a introdução de ruído, os objetos continuaram sendo predominantemente atribuídos aos grupos corretos ao longo das iterações, preservando a estrutura de grupos e resultando em uma solução final semelhante à obtida sem perturbação.

A depender da estrutura de grupos e da magnitude do ruído adicionado, a alocação dos objetos pode não estabilizar, i.e., os objetos podem ser alocados em grupos diferentes ao longo de todas as iterações do algoritmo. Na Figura 2.6b, é possível ver que os centroides foram deslocados marginalmente e a alocação dos objetos aos grupos foi a mesma daquela obtida pela versão original do algoritmo, apresentada na Figura 2.6a, sugerindo que a perturbação dos centroides não comprometeu o processo de convergência e, consequentemente, não interferiu significativamente na solução de agrupamento obtida ao final das iterações realizadas. No entanto, para conjuntos de dados com estrutura de grupos menos favorável, apresentando sobreposição, por exemplo, como ilustrado nas Figuras 2.6c e 2.6d, o desempenho do método pode ficar comprometido. Nesse caso, os valores dos atributos foram gerados também segundo distribuição normal com $\mu = \{0, 5, 10\}$, porém com $\sigma = 3$. A solução apresentada na Figura 2.6c, sem perturbação, é visivelmente mais adequada do que aquela apresentada na Figura 2.6d, uma vez que os grupos são sabidamente esféricos e centrados nas mesmas coordenadas nas quatro figuras. Na Figura 2.6d, a magnitude do ruído adicionado comprometeu a convergência e dificultou a obtenção de uma solução adequada, no que se refere à identificação da estrutura de grupos presente no conjunto de dados. Tal situação evidencia a necessidade de definir uma estratégia adequada para distribuir o *privacy budget* ao longo das iterações.

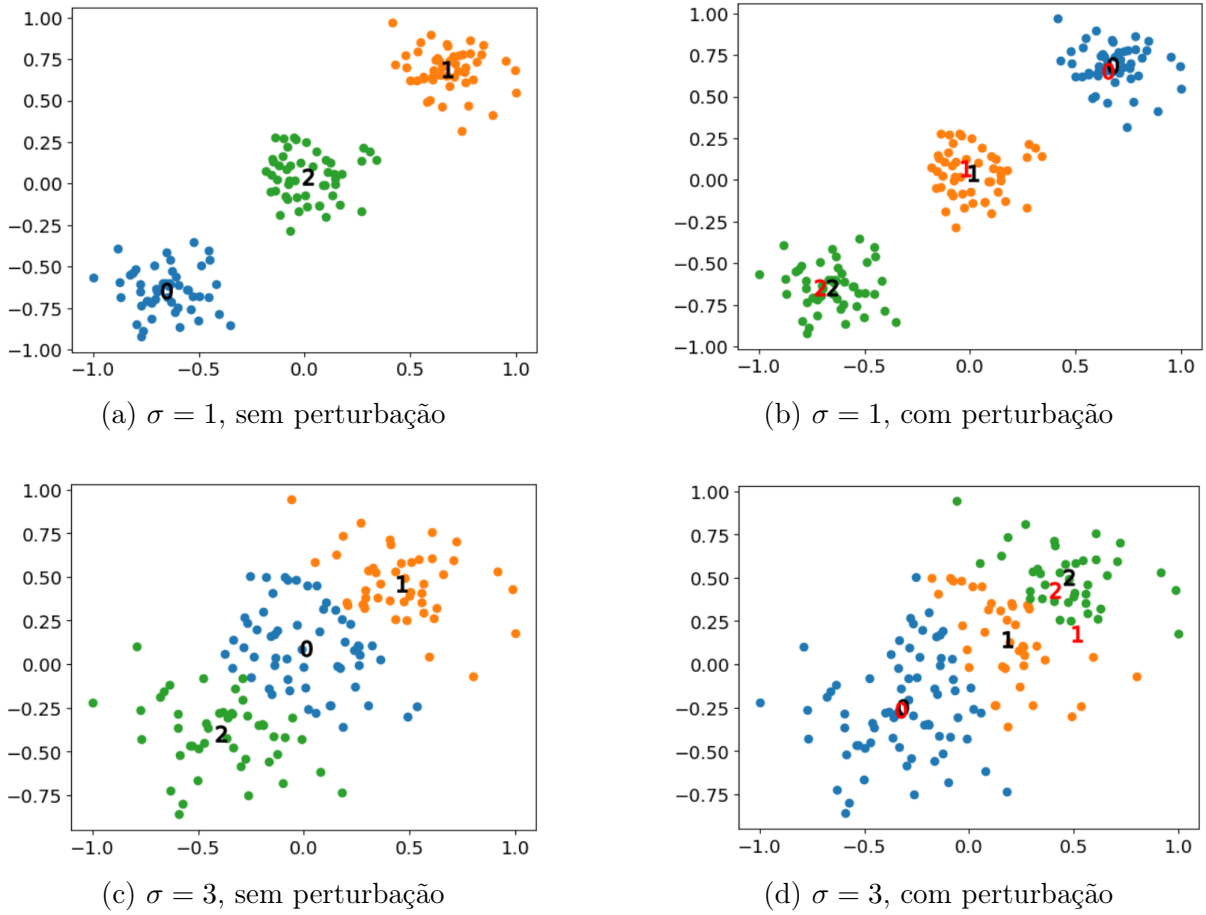


Figura 2.6: Exemplos de soluções do k-means com e sem aplicação de PD para perturbação dos centroides.

As Figuras 2.7a e 2.7b apresentam a primeira iteração de duas execuções do k-means com perturbação dos centroides através do mecanismo de Laplace adotando, respectivamente, $b = 0,1$ e $b = 0,5$. Na Figura 2.7b, é possível ver que os centroides perturbados identificados pelos algoritmos em vermelho, foram excessivamente deslocados. Se a mesma intensidade de ruído for aplicada nas iterações seguintes, a convergência, no que se refere à estabilização dos centroides, é pouco provável.

Considerando a alocação de *privacy budget* mais simples possível, distribuída de maneira uniforme por todas as iterações, i.e., $\epsilon^t = \epsilon/T$, $\forall t \in \{1, \dots, T\}$, estratégias que adotem parcelas pequenas do ϵ em cada iteração, embora permitam mais iterações, tendem a distorcer excessivamente a posição dos centroides. Isso prejudica a convergência do método – em especial a estabilização dos centroides – e resulta em soluções de agrupamento de baixa qualidade, nas quais os grupos formados não maximizam a similaridade intragrupos, comprometendo, por consequência, a utilidade das soluções obtidas. Assim, encontrar estratégia ótima para alocação de ϵ é o principal desafio na utilização do DPk-means [129, p.3].

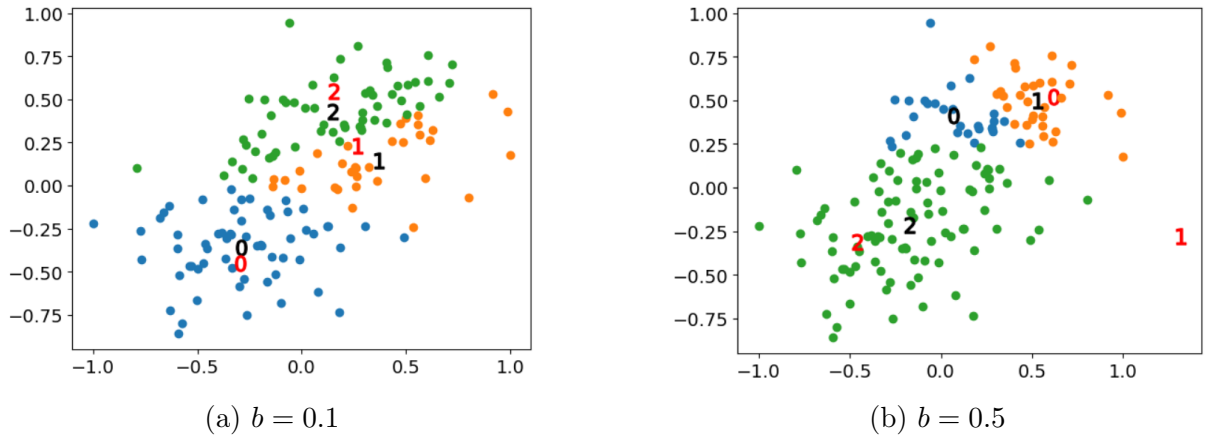


Figura 2.7: Exemplo da primeira iteração do k-means com perturbação aplicada através do mecanismo de Laplace com $b = 0,1$ e $b = 0,5$.

A aplicação real dessa abordagem é ilustrada na Figura 2.8. O analista pode controlar toda a execução do algoritmo de agrupamento, o que inclui, por exemplo, a seleção e a normalização de atributos, a definição do número de grupos⁴⁹ k , da estratégia de inicialização dos centroides, da função objetivo (que direciona a composição dos grupos) e do critério de parada⁵⁰. O curador é responsável por fazer a atribuição dos objetos aos k grupos, minimizando a função objetivo, e retornar o resultado perturbado das consultas realizadas pelo analista. Sejam então q_t^s e q_t^c as consultas submetidas pelo analista⁵¹, que aplicam, respectivamente, as funções para cálculo da soma dos atributos, f^s , e frequência dos objetos, f^c , referentes à atualização dos centroides na iteração t . O curador é responsável, primeiramente, por verificar se o *privacy budget* acumulado ao longo de todas as consultas solicitadas pelo analista ao conjunto de dados D se encontra dentro do limite ϵ previamente estabelecido. Se for o caso, o curador procede com as consultas solicitadas, adiciona o ruído através de um mecanismo de PD e retorna os resultados perturbados para o analista. O analista então calcula os centroides atuais através da razão entre a soma dos atributos e frequência de objetos e avalia se finaliza o processo ou segue para mais uma iteração.

Ao conduzir uma análise de agrupamento no modelo de PD global interativo, o analista não tem acesso a nenhuma informação individualizada dos objetos (registros) de D ,

⁴⁹O curador deve estabelecer um limite superior para o número de grupos a fim de que o analista não adote um valor próximo ou igual ao número de objetos.

⁵⁰O modelo de PD permite oferecer essa liberdade ao analista de forma segura, mas quais parâmetros serão de livre escolha do analista ou previamente estabelecidos pelo curador dependem das suas políticas e critérios de privacidade.

⁵¹Em uma implementação real, a interação do analista com o conjunto de dados D em cada iteração t pode se dar através de uma única consulta q_t , que encapsula todo o algoritmo necessário para realizar a atribuição dos objetos aos grupos e aplicar as funções de soma de atributos e frequência de objetos.

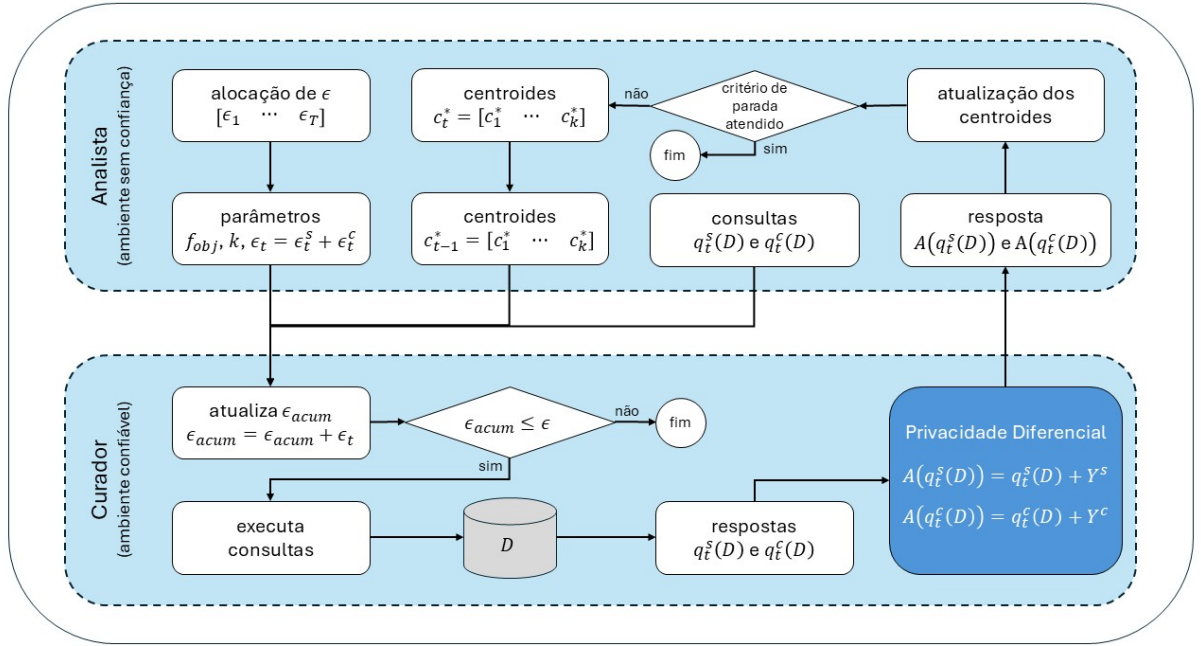


Figura 2.8: Diagrama de execução da iteração t do DPK-means no modelo de PD global iterativo.

nem mesmo aos rótulos dos grupos aos quais foram atribuídos. Embora esse cenário ofereça limitações importantes em relação à abordagem clássica (sem PD), os resultados das consultas q^s e q^c ao longo das T iterações permitem, por exemplo, obter estatísticas agregadas, identificar grupos populacionais, oferecer inicialização para a aplicação de outros métodos, realizar classificação baseada em protótipos ou monitoramento e compartilhamento dos centroides entre organizações.

Capítulo 3

Revisão da Literatura

A fim de verificar o estado da arte e identificar oportunidades concretas de contribuição, foi conduzida uma Revisão Sistemática da Literatura (RSL). O planejamento envolveu a definição de protocolo, critérios de seleção e critérios de avaliação de qualidade, de forma a identificar trabalhos relevantes sobre estratégias de alocação de *privacy budget* no contexto de agrupamento diferencialmente privado. Neste sentido, foram consideradas três bases principais, quais sejam: IEEE Xplore; ScienceDirect; e SpringerLink, abrangendo o período de janeiro de 2005 a abril de 2026.

O processo resultou na identificação inicial de 1058 artigos, dos quais 47 foram selecionados após as etapas de seleção e avaliação de qualidade. Os trabalhos selecionados foram classificados segundo o modelo de privacidade diferencial adotado (global ou local, iterativo ou não), o tipo de método de agrupamento (protótipos, densidade, grade, modelo ou hierárquico) e o tipo de processamento (centralizado ou distribuído). Os detalhes da metodologia empregada, bem como os resultados das etapas de seleção e avaliação da qualidade se encontram descritos no Apêndice C.

3.1 Abordagens de PD para o problema de agrupamento

Neste capítulo é apresentada uma revisão de algoritmos de agrupamento diferencialmente privado e estratégias de alocação de *privacy budget* propostas na literatura, conforme RSL realizada. Os artigos selecionados foram classificados de acordo com (i) o modelo de privacidade diferencial (global ou local e iterativo ou não-iterativo); (ii) o tipo de método de agrupamento (baseado em protótipos, densidade, grade, modelo ou hierárquico); e (iii) se sua implementação considerou dados distribuídos ou centralizados.

No que se refere ao modelo de PD, conforme detalhado na Seção 2.4.2, algoritmos diferencialmente privados operam segundo os modelos local ou global. No modelo global, há um curador confiável, que centraliza os dados. Na variante interativa desse modelo, o curador intermedia todas as consultas realizadas, perturbando os resultados na saída, i.e., antes de permitir acesso ao analista ou usuário. Já na variante não interativa, um conjunto de dados é disponibilizado para livre consulta do analista sem intermediação (interação) do curador. No modelo local, o curador não é considerado confiável e, portanto, o dado é perturbado localmente, antes de ser enviado para ele consolidar ou processar.

Em termos do método de agrupamento adotado, as abordagens foram classificadas como baseadas em protótipos, densidade, grade, modelo ou hierárquico. As versões clássicas dessas abordagens, i.e., sem incluir PD, são amplamente descritas na literatura sobre o tema. Em particular, uma revisão recente e abrangente sobre métodos de agrupamento é apresentada em [66]. Finalmente, no que diz respeito à implementação, as abordagens foram classificadas como centralizada ou distribuída, quando as execuções ocorrem de forma distribuída e cada cliente realiza o agrupamento, perturbando os resultados no nível local.

3.1.1 Métodos baseados em protótipos

Os métodos baseados em protótipos adotam duas estratégias para satisfazer PD [203]. A primeira delas, proposta originalmente no *framework* SuLQ [23], consiste, essencialmente, em perturbar os valores (coordenadas) da atualização dos protótipos em cada iteração. Mais especificamente, foi proposta uma variante do algoritmo de Lloyd [131] para o k-means, denominada DPk-means [23, p.5-6]. Na implementação proposta, o mecanismo de Laplace é aplicado, em cada iteração, às consultas referentes às operações de soma dos atributos dos objetos $x_i \in C_j$ em relação ao protótipo (centroide) c_j , denotada por S_j , e contagem (frequência) de objetos em C_j , denotada por $|C_j|$, resultando nos valores perturbados dessas consultas, S_j^* e $|C_j^*|$, respectivamente. O cálculo do protótipo (centroide) perturbado c_j^* é dado então por $c_j^* = \frac{S_j^*}{|C_j^*|}$, para cada grupo C_j , $j = 1, \dots, k$. Sendo assim, duas consultas são realizadas em cada iteração do DPk-means, a fim de obter os valores de S_j^* e $|C_j^*|$ e, posteriormente, de c_j^* , para cada grupo $j = 1, \dots, k$. Portanto, uma parcela do *privacy budget* total ϵ deve ser alocada a cada iteração t , $t = 1, \dots, T$, e distribuída entre as consultas S_j^* e $|C_j^*|$, sendo ϵ^s a parcela destinada à consulta de soma e ϵ^c à consulta de contagem, de forma que $\epsilon_t = \epsilon^s + \epsilon^c$ e $\epsilon = \sum_{t=1}^T \epsilon_t$ (composição sequencial). No entanto, a forma como essa alocação é feita tem grande impacto na performance do

método¹ e a definição de uma estratégia adequada não é trivial², dando então origem ao problema de alocação de *privacy budget* no k-means diferencialmente privado.

Em [54] foi proposta estratégia de alocação denominada *Dichotomy Allocation Strategy* (DAS), na qual $\epsilon_t = \frac{\epsilon}{2^t}$, de maneira a priorizar as primeiras iterações. Contudo, essa abordagem consome o *privacy budget* rapidamente e aumenta o número de iterações necessárias, o que pode introduzir a adição de ruído excessivo, comprometendo a atualização dos centroides nas iterações seguintes e impactando drasticamente a utilidade, o que se traduz em uma solução de agrupamento final de baixa qualidade.

Em [163] foi apresentado comparativo entre a aplicação do k-means clássico e do DPK-means no conjunto de dados Iris [75], no qual a qualidade da solução obtida na versão diferencialmente privada, em termos do índice de Calinski-Harabasz (CH) [30], foi inferior àquela obtida na versão clássica (sem perturbação). Segundo os autores, isso foi consequência da escolha dos centroides iniciais que, ao serem perturbados, se desviaram excessivamente dos centroides originais. Considerando tal fato, foi então proposto o algoritmo DPLk-means, baseado no DPK-means, com uma estratégia aprimorada de seleção dos centroides iniciais. Primeiramente, o conjunto de dados é dividido aleatoriamente em p subconjuntos, e o DPK-means é executado em cada um, gerando $p \times k'$ grupos, com $k' \gg k$. Em seguida, são escolhidos k centroides iniciais com base em critérios de diversidade e distância entre os centroides. Na segunda fase, é realizado o agrupamento com esses centroides, aplicando o mecanismo de Laplace para adição de ruído na atualização dos centroides em cada iteração, até que ocorra convergência do algoritmo, i.e., que o erro quadrático total seja inferior a um valor limite estabelecido previamente. No que se refere à alocação de *privacy budget*, considerou-se que o impacto das iterações iniciais nos resultados do agrupamento é maior do que o das iterações finais. Por isso, foi adotada estratégia em que ϵ tem o seu valor gradualmente reduzido ao longo das iterações, seguindo uma progressão geométrica de termo inicial $a = \frac{\epsilon}{2}$ e razão $r = 1/2$. Mais especificamente, $\epsilon_t = \frac{\epsilon}{2} \left(\frac{1}{2}\right)^{t-1} = \frac{\epsilon}{2^t}$, i.e., a primeira iteração consome $\frac{\epsilon}{2}$ e cada iteração seguinte consome metade do valor alocado à iteração anterior, até o final do processo.

O algoritmo EUGkM (Extended Uniform Grid k-means) proposto em [179, 180] seguiu premissa semelhante e adotou um critério para aprimorar a seleção dos centroides iniciais, a fim de reduzir o número de iterações. Com isso, independentemente da estratégia de alocação, uma parcela maior de *privacy budget* pôde ser alocada a cada iteração,

¹A alocação inadequada de ϵ faz com que a quantidade de ruído introduzido seja mal distribuída e a quantidade de ruído é diretamente associada à qualidade da solução obtida [216, p.490].

²Adicionar muito ruído não só leva à distorção dos dados, mas também faz com que os resultados finais do agrupamento sejam distorcidos ou o algoritmo não consiga convergir [129, p.3].

na comparação com o DPk-means original, reduzindo o ruído adicionado e favorecendo a convergência e a qualidade da solução final. O trabalho propôs ainda abordagem no modelo não iterativo através do algoritmo DP-Modes-Synopsis, no qual é gerado um resumo (*synopsis*) dos dados originais adicionando ruído de Laplace às contagens de células em uma grade que particiona o espaço de atributos. O algoritmo de agrupamento é então aplicado diretamente sobre esse resumo, e não nos dados originais, o que reduz o consumo de *privacy budget* e favorece a obtenção de uma solução de agrupamento melhor equilibrada em termos de utilidade e privacidade.

Em [152] foi abordado o método k-modes [109], proposto para superar a limitação do k-means para tratar apenas atributos numéricos, adaptando-o ao contexto de atributos categóricos. O k-modes se distingue do k-means em três aspectos essenciais, quais sejam: (i) adota medida de dissimilaridade baseada na distância de Hamming; (ii) adota modas em vez de médias como protótipos dos grupos; e (iii) utiliza estratégia baseada em frequência para atualizar os protótipos e minimizar a função de custo. Os autores propuseram abordagens adotando os modelos de PD local e global, bem como iterativo e não iterativo. No que se refere aos modelos local e iterativo, foi proposto o algoritmo DP-Modes-Lloyd, baseado no algoritmo de Lloyd [131], adotando o mecanismo de Laplace, além de duas variantes com diferentes estratégias de inicialização, denominadas DP-Modes-FI e DP-ModesSVT. A primeira adota estratégia de inicialização baseada em frequência e a segunda, baseada em *Sparse Vector Technique* (SVT) [59, p.55-64]. Ainda no modelo iterativo, foram propostas também duas alternativas que adotam o modelo global e o mecanismo exponencial, quais sejam: DP-Modes-EM, voltada para conjuntos de dados pequenos, e DP-Modes-MCMC, para conjuntos de dados grandes, que simula a amostragem de k modas perturbadas no espaço de atributos por meio da execução de um algoritmo de *Markov Chain Monte Carlo* (MCMC). No que se refere ao modelo não iterativo, foi adotado o algoritmo DP-Modes-Synopsis, baseado na técnica *synopsis* [179, p.30]. Os experimentos realizados demonstraram que as abordagens DP-Modes-Lloyd e DP-Modes-MCMC, baseadas no modelo de PD iterativo, apresentaram melhor desempenho do que a abordagem não interativa DP-Modes-Synopsis em termos do índice NICV³.

Em [70] foi proposta estratégia de alocação denominada *Series Sum Allocation Strategy* (SSAS), que oferece como principal vantagem não demandar a definição prévia do

³O índice NICV, dado por $NICV = \frac{1}{N} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2$, representa a distância quadrática média entre qualquer objeto e o centroide do grupo ao qual foi alocado. Como o NICV mede a variabilidade intra-cluster, quanto menor o valor do índice, melhor é a solução no que se refere à coesão interna dos grupos, configurando um problema de minimização.

número total de iterações T . Considerando $\sum_{t=1}^{\infty} \frac{\epsilon}{(t(t+1))} = \epsilon$, os autores propuseram que o *privacy budget* na iteração t fosse dado por $\epsilon_t = \frac{\epsilon}{(t(t+1))}$. No entanto, como não havia garantia de convergência, a adição de ruído se tornou excessiva quando o número de iterações aumentou, perturbando cada vez mais a atualização dos centroides e dificultando a convergência. Os autores propuseram então a estratégia denominada *Arithmetic Progression Allocation Strategy* (APAS), na qual ϵ_t é dado por

$$\epsilon_t = \epsilon_m + \frac{(T - t) \left(\frac{2\epsilon}{T} - 2\epsilon_m \right)}{T - 1}$$

sendo ϵ_m correspondente ao *privacy budget* mínimo [179] necessário para garantir a convergência do DPk-means. A estratégia adotada consistiu na alocação decrescente do *privacy budget* ao longo das iterações, com o objetivo de promover convergência mais rápida, ao mesmo tempo em que controlava o valor de ϵ_m para assegurar a convergência do algoritmo.

Em [135], a ideia central foi aproveitar a propriedade de convergência do algoritmo de Lloyd como parâmetro para limitar a magnitude do ruído adicionado às coordenadas dos centroides pelo mecanismo de Laplace. Mais especificamente, o ruído é limitado de forma a garantir que o centroide perturbado permaneça inscrito em uma circunferência determinada ao redor do centroide não perturbado em cada iteração, garantindo a convergência. Além disso, é aplicado o mecanismo exponencial⁴ em uma direção controlada da atualização dos centroides, de forma que mantenha-o na direção de convergência do algoritmo. Essa direção corresponde ao sentido do deslocamento do centroide entre duas iterações consecutivas do algoritmo, i.e., o vetor que aponta do centroide da iteração anterior para o centroide não perturbado da iteração atual, refletindo, portanto, o movimento natural (não perturbado) de convergência que o centroide teria na ausência de ruído. Isso difere de outras abordagens, uma vez que, geralmente, não há controle na direção da perturbação dos centroides.

Em [155] foi proposto anular parcialmente o ruído adicionado por meio da fusão de grupos ruidosos adjacentes e otimizar a alocação de *privacy budget*, a fim de adicionar ruído de forma adaptativa durante as iterações. Quanto à fusão de grupos, foi comprovado empiricamente que os ruídos se anulavam mutuamente [155, p.208]. No que se refere à alocação de *privacy budget*, a motivação por trás da proposta apresentada é de que os centroides mudam mais nas primeiras iterações e, portanto, a adição de ruído de maior amplitude não teria muito impacto. À medida que as iterações avançam, as mudanças

⁴O mecanismo exponencial se encontra descrito no apêndice B.2 (p.153).

nos centroides ficam menos significativas em relação à mudança de posição, e a adição de ruído deve ser mais cautelosa. Então, a ideia foi aumentar ϵ gradativamente conforme o número de iterações aumenta e os centroides se estabilizam. Os autores, no entanto, propuseram um esquema de alocação bastante simples, qual seja: definido número de iterações $T = 12$, alocar $\frac{\epsilon}{24}$ às primeiras 4 iterações, $\frac{\epsilon}{12}$ da quinta à oitava iteração e, da nona à última, alocar $\frac{\epsilon}{8}$. O trabalho adotou a estratégia do algoritmo EUGkM [179, 180] para a seleção dos centroides iniciais, de maneira a fazer com que eles sejam tão distantes entre si quanto possível.

Em [95] foi proposto um método adaptativo de alocação de *privacy budget* com base nos resultados do agrupamento. A abordagem é baseada na estratégia DAS [54], mas a alocação é ponderada com base na coesão e separação dos grupos em cada iteração, a fim de controlar o afastamento do centroide e melhor balancear utilidade e privacidade. Mais especificamente, ao final de cada iteração, a coesão e separação dos grupos C_j , $j = \{1, \dots, k\}$, é dada por [95, p.222]:

$$\text{CHS}_j = \frac{\sum_{i=1}^{|C_j|} d_{L2}(x_i, c_j)}{|C_j| \times d_{L2}(c_j, \bar{c})}$$

sendo $\bar{c}$ correspondente ao centroide global, i.e., a média de todos os objetos do conjunto de dados. Quanto menor o valor de CHS_j , mais coeso e bem separado é C_j . A alocação do *privacy budget* no cálculo do centroide do grupo C_j , na iteração t , é dada por [95, p.222]:

$$\epsilon_t^j = \frac{\epsilon}{2^t} \times \frac{\min_{1 \leq j \leq k} (\text{CHS}_j)}{\text{CHS}_j}$$

Além da estratégia de alocação de *privacy budget*, o trabalho também adotou uma variante do DPK-means similar ao EUGkM [179, 180], a fim de aprimorar a seleção dos centroides iniciais, reduzindo o número total de iterações.

Em [216] foi proposta alocação de *privacy budget* baseada em progressões aritméticas. Primeiramente é estabelecido o *privacy budget* mínimo para cada iteração. Quando ϵ é suficiente para garantir o mínimo em cada iteração, utiliza-se uma progressão aritmética decrescente de razão r para determinar a alocação, i.e., $\epsilon_t = \epsilon_1 + (t-1)(-r)$ e $\frac{T}{2}(\epsilon_1 + \epsilon_T) = \epsilon$. O *privacy budget* mínimo é definido como o primeiro termo da sequência, sendo a alocação de cada iteração feita em ordem decrescente, de maneira a alocar as parcelas maiores às iterações iniciais, mas garantindo o mínimo até a última iteração.

Em [63] foi proposto o método dist-DP k-means, que adota uma formulação de PD baseada em distância. Conforme descrito na definição de conjuntos de microdados vizinhos ou adjacentes (seção 2.4.1, p.23), dois conjuntos de microdados D_1 e D_2 são p -adjacentes em termos da métrica M se $d_M(D_1, D_2) \leq p$ [44, p.57]. Essa definição é formalizada tipicamente adotando a distância simétrica, $d_{\text{sim}}(D_1, D_2) = |D_1 \triangle D_2|$ (Eq. 2.7, p.23), quando a diferença entre os conjuntos de microdados se dá na presença (ou ausência) de registros (e.g., indivíduos), ou a distância de Hamming, $d_{\text{ham}}(D_1, D_2) = \sum_{i=1}^n \mathbb{I}[D_{1i} \neq D_{2i}]$ (Eq. 2.8, p.23), quando a diferença entre os conjuntos de microdados se dá na alteração/atualização dos valores dos atributos. Nesse trabalho, os autores consideraram uma definição ligeiramente diferente, na qual os conjuntos são ρ -adjacentes se tiverem, exatamente, o mesmo número de registros, porém diferirem em um e apenas um registro, para o qual a diferença (distância) é dada por ρ . A distância de Hamming quantifica o número de registros que apresenta alguma diferença, independentemente da magnitude dessa diferença. A definição adotada nesse trabalho considera $d_{\text{ham}}(D_1, D_2) = 1$ e quantifica a magnitude da diferença através de métricas de distância entre pares de objetos (e.g., distância euclidiana – Eq. 2.6, p.23). Além disso, consideraram ainda a relaxação (ϵ, δ) -PD (apêndice B.3.3, p.156) em vez da PD pura. O algoritmo proposto, dist-DP k-means, atende então (ϵ, δ) -PD, para $\epsilon, \delta \leq \frac{1}{2}$, em contextos de conjuntos de dados ρ -adjacentes.

Em [93] foi proposta uma variante diferencialmente privada do k-means++ [121, 142], denominada DP PSO k-means++. A principal diferença em relação ao k-means original é o processo de seleção dos k centroides iniciais, aspecto que é especialmente relevante em variantes diferencialmente privadas desse método. Para a seleção do primeiro centroide, a variância de cada objeto em relação a todos os outros é calculada e usada como medida de densidade, considerando que objetos com menor variância estão em regiões mais densas. Como primeiro centroide, é então selecionado o objeto com menor variância. Para os demais centroides é adotado critério que busca priorizar a seleção de objetos distantes dos centroides já selecionados. No que se refere à PD, além de perturbar o cálculo dos centroides, também foi adicionado ruído no valor da distância total calculada no momento da inicialização dos centroides, utilizando o mecanismo de Laplace em ambos os casos. O kernel Normal foi usado para definir a alocação de *privacy budget*. Por fim, foi utilizada abordagem de otimização baseada na metaheurística enxame de partículas (do inglês, *Particle Swarm Optimization* - PSO) para obter a solução final, tanto em termos da adição do ruído quanto da alocação dos objetos aos grupos.

Em [129, p.4], os autores argumentaram que estratégias de alocação que utilizam funções matemáticas específicas para definir ϵ_t , como algumas das propostas anteriormente,

são empíricas e que as soluções de agrupamento resultantes podem ser melhoradas em termos de convergência e utilidade. Os autores propuseram então abordar o problema de alocação de *privacy budget* como um problema de otimização combinatória e adotaram uma estratégia baseada em algoritmo genético, denominada GAPBAS, para propor uma solução. Primeiramente, foi adotada abordagem para aprimorar a seleção dos centroides iniciais e o controle do *privacy budget* mínimo, similar à estratégia proposta no EUGkM [179]. Um algoritmo genético foi então utilizado para otimizar a alocação de *privacy budget* nas iterações do DPk-means. Os experimentos computacionais realizados demonstraram melhores resultados da GAPBAS na comparação com DAS [54], SSAS e APAS [70]. A partir dos resultados obtidos, os autores propuseram ainda uma regra geral para alocação eficiente de *privacy budget* no DPk-means.

As abordagens anteriores adotam, predominantemente, o modelo global de PD, i.e., aplicam o agrupamento em contextos onde os dados são processados de forma centralizada. Em contextos em que os dados são processados de forma distribuída e posteriormente consolidados por um servidor central não confiável, o modelo de PD local é utilizado para perturbar os protótipos antes do compartilhamento. Uma abrangente revisão sobre aplicações desse modelo de PD foi publicada recentemente [207], na qual uma subseção foi dedicada a modelos de aprendizagem não supervisionados [207, p.11]. Os autores apontaram o k-means como um dos métodos de agrupamento mais fundamentais e que recentemente despertou interesse no contexto de PD local e, portanto, direcionaram o levantamento para estudos que aplicaram k-means e PD local em ambientes distribuídos. No que se refere à alocação de *privacy budget*, os autores mencionaram que os estudos desenvolvidos geralmente consideram minimizar o número de iterações ou estabelecer limites para a perda de privacidade. Porém, independentemente dos avanços alcançados, o problema de otimizar a alocação de *privacy budget* mediante a aplicação de algoritmos de agrupamento em ambientes distribuídos continua sendo um desafio.

Em [177], os autores destacam que, em algoritmos k-means diferencialmente privados, as versões disponíveis alcançavam apenas fatores de aproximação multiplicativos super constantes, em contraste com o problema clássico, para o qual existem algoritmos de tempo polinomial com erro multiplicativo constante e erro aditivo nulo. Com base nessa lacuna, o trabalho propõe o primeiro algoritmo diferencialmente privado para o k-means com erro multiplicativo constante e erro aditivo comparável ao de abordagens anteriores.

No LDPKmeans [205], o cliente (ou agente) é o responsável por adicionar o ruído e

determinar a qual grupo cada objeto pertence. Em seguida, enviam essas informações perturbadas ao servidor central, que reúne os dados perturbados, realiza o agrupamento, atualiza os protótipos e envia os novos protótipos aos clientes. Esse processo se repete até que o algoritmo convirja. Em [170], no entanto, os autores simulam um ataque capaz de comprometer o LDPKmeans distribuído, mesmo que o servidor central obtenha apenas as informações perturbadas dos grupos e protótipos a cada iteração, enviadas pelos clientes, permitindo inferir valores muito próximos dos reais.

Em [208] foi considerado o modelo de PD local. O trabalho adotou uma generalização do conceito de PD proposta inicialmente em [36], denotada como d_χ -privacy. Essa abordagem considera a sensibilidade dos dados com base em uma métrica definida sobre o espaço dos possíveis valores, permitindo um controle mais refinado sobre o grau de privacidade. O espaço métrico adotado foi o euclidiano, logo foi proposto método de perturbação para o k-means que atende ϵd_E -privacy. No cenário considerado, o servidor central não é confiável e os clientes não se comunicam. Os clientes aplicam um mecanismo de PD em seus conjuntos de dados originais de forma independente e os submetem, já perturbados, ao servidor central, que consolida os dados de todos os clientes e executa o k-means. Dessa forma, há apenas uma iteração de comunicação entre clientes e o servidor central, fazendo com que o modelo de PD adotado, além de local, seja não iterativo, favorecendo a redução do consumo de *privacy budget*. Entretanto, isso traz mais desafios à etapa de perturbação, pois é necessário garantir a privacidade dos registros, porém preservando utilidade suficiente para que o k-means possa ser executado.

O k-means distribuído, no contexto de *federated learning* horizontal,⁵ foi abordado com profundidade em [6]. O autor propôs o que argumenta ser a primeira integração formal de PD no k-means federado horizontal, resultando em um algoritmo mais rápido, preciso e com garantias formais de privacidade através de PD. As principais contribuições mencionadas incluem um protocolo eficiente de agregação segura, com ganhos de desempenho de até três ordens de magnitude em relação a soluções anteriores e a aplicação de restrições de tamanhos de grupo no DPK-means [18], permitindo aprimorar a utilidade em relação ao estado-da-arte do k-means no modelo de DP global. Além disso, o estudo propôs uma variante do DPK-means no que se refere à adição do ruído, aplicando o mecanismo de Laplace ao resultado final do cálculo do centroide, em vez de aplicar à

⁵O *federated learning* horizontal é uma categoria de aprendizado distribuído na qual os clientes possuem diferentes amostras de indivíduos/registros, mas com os mesmos atributos. É comum em modelos treinados sobre o mesmo esquema, como regressão e classificação. No *federated learning* vertical, em contrapartida, os clientes são detentores de diferentes atributos dos mesmos registros, como, por exemplo, no caso de colaboração entre entidades que atuam em áreas distintas, como uma empresa de telefonia e uma instituição financeira.

soma das distâncias e à frequência de objetos separadamente. O autor ponderou que essa abordagem introduz maior ruído, uma vez que a sensibilidade aumenta, mas pode ser útil em casos em que publicar uma contagem diferencialmente privada não seja aceitável. No que diz respeito à alocação de *privacy budget*, foi adotada a estratégia [6, p.42] proposta em [179].

Em [214], os autores abordaram o problema dos k -centros (*k-centers problem*), similar aos problemas k -means e k -medoids. A principal diferença reside na função objetivo. Enquanto k -means e k -medoids buscam minimizar uma soma de distâncias entre os objetos e os protótipos de seus respectivos grupos, o k -centers minimiza a distância máxima. Os autores propuseram algoritmos de aproximação baseados no mecanismo exponencial para resolver a versão clássica do problema, bem como duas variações, quais sejam: k -centers distribuído e k -centers distribuído com *outliers*. O trabalho realiza análise aprofundada dos algoritmos propostos, apresentando limites teóricos para cada variante. Para o k -centers clássico, é proposto um algoritmo que assegura $(2\epsilon\Delta k)$ -PD, onde $\Delta = \max_{i,j \in X} d(x_i, x_j)$ é o diâmetro do espaço métrico, e custo máximo de $2\text{OPT} + 3(\ln n)/\epsilon$, onde OPT representa o custo da solução ótima e n é o número de objetos em X . Para o k -centers distribuído, o algoritmo proposto atende $(2\epsilon\Delta k)$ -PD e apresenta custo máximo de $4\text{OPT} + 3(\ln n)/\epsilon$, e para o k -center distribuído com *outliers*, o custo máximo é $13\text{OPT} + 15(\ln n)/\epsilon$, atendendo $(2\epsilon\Delta(k+z))$ -DP, em que z representa o número de *outliers*.

Em [211] foi proposta abordagem não interativa voltado a dados categóricos de alta dimensionalidade, com ênfase na redução de custo de comunicação e riscos de violação da privacidade associados a abordagens interativas. O método, denominada FskLDP k -mode, combina redução de dimensionalidade local via algoritmo Fsketch com um único passo de perturbação utilizando o mecanismo de Laplace, garantindo ϵ -PD antes do envio dos dados ao servidor centralizador. A partir dos dados perturbados, o servidor realiza o agrupamento por meio da estimação da distância de Hamming, utilizando procedimento para mitigar os efeitos do ruído, o que permite a alocação de objetos aos grupos e a atualização iterativa dos centroides (modas) sem novas interações com os clientes.

Em [122], os autores propõem um aprimoramento do k -means distribuído com agregação segura [215] (DistK), denominado DistK-GS. Eles identificaram como limitação central do DistK o uso de limites de sensibilidade globais, definidos de forma fixa e independente dos dados (e.g., a partir de um intervalo máximo $[-B, B]$ por dimensão), o que tende a introduzir ruído em excesso e, conseqüentemente, a degradar a qualidade do agrupamento. Como solução, introduzem um procedimento adaptativo de calibração

da sensibilidade. Esse procedimento se baseia em uma avaliação geométrica do impacto que a modificação de um único objeto pode causar nos centroides. Mais especificamente, cada centroide define uma célula de Voronoi, e a sensibilidade é determinada pelo maior deslocamento possível do centroide quando um ponto situado na fronteira dessa célula é alterado. Para isso, em cada dimensão, é calculado o desvio máximo entre o centroide e os vértices da célula de Voronoi, que representam os pontos extremos de variação [122, p.9-10].

Esse procedimento produz limites de sensibilidade ajustados à configuração atual dos dados e da estrutura de grupos, significativamente mais restritos do que os limites de sensibilidade global, que, por sua vez, são baseados no pior caso. Os resultados experimentais indicam ganhos substanciais em termos de variância intra-cluster e estabilidade dos centroides [122, p.12], porém não exploram o aspecto de otimização da alocação de *privacy budget*. Os experimentos realizados adotaram estratégia uniforme [122, p.7] na alocação de *privacy budget* e o mecanismo Normal.

O processo iterativo dos algoritmos do tipo k-means diferencialmente privado apresenta desafios devido à adição excessiva de ruído, levando a problemas de convergência e utilidade reduzida [129, p.1]. Embora esses algoritmos utilizem esquemas de alocação de *privacy budget*, a magnitude do ruído adicionado ainda depende do número de iterações, fazendo com que o aumento do ruído degrade a qualidade do agrupamento [203, p.2] – traduzida em centroides excessivamente deslocados da posição real e, conseqüentemente, em objetos alocados a grupos menos representativos da estrutura de agrupamento presente nos dados. Em estratégias como essas, nas quais os protótipos são perturbados a cada iteração do método de agrupamento, a intensidade do ruído é função do total de iterações realizadas. A principal limitação desse tipo de abordagem se refere ao critério de parada que é, geralmente, o número máximo de iterações. Critérios mais robustos, baseados, por exemplo, na diferença entre coordenadas dos centroides em duas iterações consecutivas, não podem ser implementados a partir de estratégias desse tipo no contexto diferencialmente privado, uma vez que os centroides observados são perturbados pelo mecanismo de PD e, portanto, a variação entre iterações passa a refletir não apenas o processo do algoritmo na direção da função objetivo, mas também as perturbações aleatórias introduzidas pelo ruído, além de demandar consultas adicionais que implicariam consumo cumulativo do *privacy budget*.

Uma estratégia disponível na literatura para métodos baseados em protótipos, que contorna essa limitação, utiliza o conceito de *coresets* [4], descrito no Apêndice E. Em

linhas gerais, no contexto de agrupamento clássico, um *coreset* é um subconjunto de objetos do conjunto original que permite aproximar a função objetivo do conjunto original para qualquer escolha de centros (protótipos) [108]. O objetivo é reduzir o custo computacional de algoritmos de agrupamento aplicando-os a um conjunto reduzido de objetos, porém de maneira que a qualidade da solução obtida, no que se refere ao valor da função objetivo, seja próxima daquela obtida para o conjunto completo. *Coresets* são usados para encontrar agrupamentos aproximadamente ótimos em grandes conjuntos de dados. De forma geral, um *coreset* é subconjunto de objetos do conjunto original que permite aproximar a função objetivo do conjunto original para qualquer escolha de centros (protótipos) [108]. Para isso, pesos são atribuídos aos objetos selecionados para compor o *coreset*. Neste sentido, há diversas abordagens na literatura que aplicam *coresets* ao problema de agrupamento clássico [98, 108, 42, 40].

No contexto do problema de agrupamento diferencialmente privado, o uso de *coresets* oferece alternativa promissora ao modelo de PD iterativo, pois uma vez que o *coreset* satisfaz ϵ -PD, deixa de ser necessária a alocação de frações de ϵ a cada iteração do algoritmo de agrupamento, permitindo não apenas o uso de critérios de parada mais sofisticados do que o número máximo de iterações, mas também múltiplas execuções de variantes do k-means ou outros tipos de algoritmo de agrupamento. Nesta abordagem, a garantia da privacidade não é propriedade do algoritmo⁶ e sim do *coreset*, oferecendo mais flexibilidade [72, p.362].

Em [74] foi apresentada a primeira aplicação real conhecida que combina *coresets* com o problema de agrupamento diferencialmente privado baseado em protótipos. Os autores propuseram uma abordagem para agrupar dados de localização gerados por dispositivos móveis, dando ao usuário controle sobre o nível de ruído adicionado⁷. Mais especificamente, os autores desenvolveram o primeiro algoritmo k-means para agrupar dados de GPS gerados por dispositivos móveis, com foco em maximizar a proteção dos movimentos de localizações nas trajetórias.

⁶Por exemplo, no *framework* SuLQ [23] é proposta uma implementação diferencialmente privada do algoritmo de Lloyd.

⁷Usuários podem ficar mais confortáveis em compartilhar dados de localização do trajeto até o trabalho do que a localização onde residem. A motivação para desenvolvimento do método foi permitir, para cada dispositivo móvel, o ajuste do seu nível de privacidade, possibilitando a contribuição para estatísticas globais sem comprometer sua privacidade.

3.1.2 Métodos baseados em densidade

Assim como os métodos baseados em protótipos, os baseados em densidade também produzem soluções vulneráveis em termos de privacidade. O DBSCAN é um dos métodos de agrupamento baseado em densidade mais consagrado [107]. Nele se pode identificar grupos através da concentração de objetos em áreas de alta densidade, que são separadas por áreas de baixa densidade. No contexto da análise de dados privados, um adversário pode então explorar as informações sobre raio e densidade para fazer inferências em relação aos atributos dos objetos, comprometendo assim a privacidade [203, p.4].

Os métodos baseados em densidade apresentam vantagens em relação aos métodos iterativos baseados em protótipos no contexto da privacidade diferencial por geralmente exigirem menos iterações para convergirem para uma boa solução ou serem executados de forma não iterativa, reduzindo o consumo cumulativo de *privacy budget* decorrente da composição sequencial. Além disso, esses métodos dependem mais da estrutura relativa da densidade local dos objetos do que de métricas globais, o que reduz a sensibilidade das operações realizadas e, conseqüentemente, o impacto do ruído adicionado. Nesse caso, duas alternativas são tipicamente adotadas para adicionar o ruído, quais sejam: (i) perturbando a matriz de distâncias, i.e., perturbando o valor das distâncias entre todos os pares de objetos, o que pode influenciar na inclusão de um objeto no raio de abrangência de outro objeto; ou (ii) na frequência de objetos dentro de um raio de abrangência, o que pode influenciar na classificação do objeto como núcleo, borda ou ruído.

DP-MCDBSCAN [154, 153] é uma variante do DP-DBScan [204] que otimiza a inicialização através da seleção de objetos núcleo iniciais pelo critério da maior distância (*farthest distance*). Essa abordagem apresentou vantagens substanciais, na comparação com o DP-DBScan, quando adotados pequenos valores de ϵ , especialmente em conjuntos de dados maiores. A adição de ruído é feita na matriz de distâncias, aplicando o mecanismo de Laplace.

Em [38], foi proposta uma abordagem baseada no algoritmo de agrupamento de pico de densidade (do inglês, *Density Peak Clustering* - DPC) [165], que parte da ideia de que o objeto central de cada grupo seja o ponto de máxima densidade local do grupo e que a distância entre diferentes centros seja relativamente grande. Cada objeto x_i é avaliado em termos da sua densidade local, ρ_i , e da distância mínima até um objeto de maior densidade local, δ_i . Esses valores são combinados no produto $\gamma_i = \rho_i \delta_i$, utilizado para selecionar os centros, uma vez que centros tendem a ter valores altos para ambos os fatores e pontos de ruído podem ter para um deles. O algoritmo então identifica centros

de grupos como objetos que possuem alta densidade local e estão distantes de outros objeto de alta densidade. Cada ponto é atribuído ao centro mais próximo em termos de distância, seguindo o gradiente de densidade.

Além de questões relativas à privacidade, os autores apontaram que o DPC tende a performar mal em conjuntos de dados de elevada dimensionalidade e propuseram uma variante dele na qual a identificação dos centros é feita a partir da ordenação decrescente dos valores de γ_i e da determinação do ponto crítico, que corresponde ao ponto de maior inflexão na curva formada por esses valores. Os pontos de centro são, portanto, aqueles anteriores ao ponto crítico. No que se refere a PD, o mecanismo de Laplace é utilizado para adicionar ruído aos valores de densidade local⁸. Outra diferença relevante em relação ao DPC original é que a variante proposta adota a distância cosseno como métrica de similaridade, em lugar da distância euclidiana. O trabalho propôs os algoritmos DP-ADPC e DP-UDPC, que diferem no que diz respeito à identificação do ponto crítico na curva de γ , sendo o primeiro considerado mais apropriado em contextos com menor número de grupos (k) e o segundo em contextos com mais grupos.

Em [203, p.2], os autores ponderaram que a adição de ruído na matriz de distâncias introduz imprecisões na métrica de distância entre os objetos, o que pode levar a soluções de agrupamento insatisfatórias, devido à formação incorreta de vizinhanças⁹, i.e., regiões densas, comprometendo a formação dos grupos e, conseqüentemente, a utilidade da solução de agrupamento obtida. Os autores propuseram então o DBDP, que aplica o mecanismo de Laplace para perturbar a densidade estimada¹⁰ da vizinhança de cada objeto, evitando distorção na classificação dos objetos e permitindo melhor equilíbrio entre privacidade e utilidade. O método consiste de três estágios: (i) estimação de densidade, i.e., cálculo da frequência de objetos dentro do raio de abrangência (*distance threshold, eps*) de cada objeto do conjunto de dados; (ii) identificação de objetos núcleo, no qual são rotulados como “objetos núcleo” aqueles que têm densidade, i.e., número de objetos dentro do seu raio de abrangência, superior a *minPts*; e (iii) construção dos grupos, no qual são alocados em um mesmo grupo todos os objetos núcleo dentro do raio de abrangência

⁸Uma vez que a densidade local de x_i , dada por ρ_i , é utilizada no cálculo de δ_i e γ_i , não é necessário adicionar ruído a esses termos

⁹Métodos de agrupamento baseados em densidade, como o DBSCAN, classificam os objetos como núcleo, borda ou ruído, utilizando como critério a densidade de objetos na sua vizinhança, ou seja, número de objetos dentro de um determinado raio de abrangência. Uma vez que as distâncias são distorcidas pelo ruído, objetos podem ser classificados incorretamente, implicando na formação de grupos que não refletem a estrutura de agrupamento presente nos dados.

¹⁰Os autores do método chamam de “densidade estimada” associada a cada objeto, o número de objetos dentro do seu raio de abrangência (*distance threshold, eps*), representando a densidade de objetos na sua vizinhança.

de outros objetos núcleo, bem como os objetos não-núcleo dentro do raio de abrangência deles. Uma perturbação é aplicada em cada etapa, porém com alocação de *privacy budget* diferenciada, a fim de que ϵ decaia exponencialmente.

Em [128] foi proposto método chamado DCT-DP para proteção de dados de trajetória de dispositivos móveis no contexto do IoT. Tem como principal vantagem considerar atributos temporais, preservando correlações temporais e espaciais presentes originalmente no dado e cruciais para análise de trajetórias dinâmicas. O cenário considerado foi de confiança plena no servidor central e todo o processamento ocorreu de forma centralizada, a partir do recebimento dos dados brutos coletados pelos dispositivos móveis. O método consiste de quatro etapas: (i) o k-means é utilizado para generalizar os atributos temporais do conjunto de trajetórias original, agrupando usuários com tempos (*timestamps*) semelhantes; (ii) o método DENCLUS¹¹ é aplicado em cada grupo e as trajetórias são generalizadas em localizações núcleo; (iii) registros são mesclados e os dados de trajetórias anormais são removidos; e (iv) o mecanismo de Laplace é aplicado para adicionar ruído previamente à publicação.

Em [81] foi proposto o algoritmo DC-LDP, voltado para cenários com dados descentralizados. O método de agrupamento adotado é baseado em densidade e aplicado localmente. A abordagem utiliza então PD local e combina técnicas como *locality-sensitive hashing*, representação vetorial binária e transformadas wavelet para mitigar o impacto do ruído introduzido pelos mecanismos de privacidade.

Em [219], foi proposto o método DP-ADC, derivado do DBSCAN, que incorpora procedimentos adaptativos tanto na calibração dos parâmetros de raio de vizinhança e critério de densidade, quanto na alocação de *privacy budget*. O método utiliza o mecanismo de Laplace para aplicar PD em duas etapas críticas: (i) avaliação da densidade (identificação de pontos de núcleo/centro) e (ii) fusão dos grupos. No que se refere à PD, a principal contribuição reside em uma estratégia adaptativa de alocação do *privacy budget*, na qual ϵ é distribuído dinamicamente entre os grupos de acordo com seu tamanho (número de objetos) e densidade, de modo que regiões menos densas, cujos objetos estão sujeitos a maior risco de violação, recebem parcelas menores de *privacy budget*, i.e., ruído de maior magnitude, enquanto regiões densas recebem menos perturbação.

¹¹O método DENCLUS [128, p.8-10] é baseado em abordagem denominada TRACCLUS, na qual trajetórias são inicialmente divididas em subtrajetórias representadas por segmentos de reta, que, por sua vez, são descritos por um par de pontos. Em seguida, o DBSCAN é aplicado para agrupar esses pontos.

3.1.3 Métodos baseados em grade

Em [80] foi adotado método de agrupamento baseado em grade, cuja vantagem, segundo os autores, é que a utilidade depende das características da grade e não de registros individuais, o que favorece trabalhar o *trade-off* entre a utilidade (qualidade da solução de agrupamento) e a proteção da privacidade. O trabalho adota o modelo local de PD e propõe o algoritmo GC-LDP, que consiste em, primeiramente, construir uma grade não uniforme com base no coeficiente de variação (CV) de cada atributo. Mais especificamente, o número de divisões em cada dimensão é determinado com base no CV do respectivo atributo, de modo que dimensões com maior variabilidade recebam mais divisões.

Isso garante maior resolução onde há mais dispersão, o que contribui na preservação da utilidade mesmo após a introdução de ruído para privacidade. Uma vez construída a grade, um mecanismo de PD é aplicado. Por fim, para realizar o agrupamento em múltiplas densidades, é aplicado um método de agregação de células com base nas diferenças relativas de densidade entre as grades. Mais especificamente, o algoritmo GC-LDP é dividido em três etapas: (i) o curador (servidor central) define parâmetros como o número de clientes e o *privacy budget*, repassa essas informações aos clientes, que estimam seus dados com base em conhecimento prévio e os devolvem ao curador, que então constrói uma grade ideal G_{opt} usando um algoritmo de segmentação baseado no coeficiente de variação; (ii) o curador identifica células densas na grade G_{opt} usando uma busca aproximada e realiza a agregação rápida dessas células para formar os grupos iniciais; e (iii) o curador refina os grupos ajustando suas bordas (fronteiras) com informações adicionais fornecidas pelos indivíduos, aumentando assim a utilidade da solução final.

3.1.4 Métodos baseados em modelos

O agrupamento baseado em modelo de mistura de gaussianas (do inglês, *mixtures of Gaussians* - MoG) é um método de agrupamento fuzzy (i.e., não exclusivo) que assume que os dados seguem distribuições gaussianas distintas. Ele atribui a cada objeto uma probabilidade de pertencimento a diferentes grupos, correspondentes a diferentes gaussianas, em vez de um agrupamento rígido, e utiliza o método de maximização da expectativa para estimar os parâmetros do modelo da distribuição correspondente a cada grupo [213, p.487].

No algoritmo LDP-GMMC [213], os subgrupos construídos são unidos em k grupos com base no grau de sobreposição entre eles. Por se tratar de abordagem baseada em

cenário sem confiança mútua, é adotado o modelo local de PD. A estratégia consiste de duas etapas: (i) no cliente, os parâmetros são estimados com base nos dados locais, perturbados e enviados para o servidor central; (ii) no servidor, os dados são atualizados, grupos são fundidos conforme necessário, e os parâmetros atualizados são enviados de volta aos clientes até que os rótulos dos grupos se estabilizem, i.e., não ocorra mais atualizações dos grupos ou subgrupos. Esta estratégia superou o LDPk-means [205] na identificação de grupos não esféricos, mas não na de grupos esféricos.

O DPMoG-hard [29] utiliza mistura de gaussianas (do inglês, *mixtures of Gaussians* - MoG), que dispensam dados de treinamento após a estimação da função de densidade. A partir da função de densidade da MoG, estimada com base em resultados de agrupamentos diferencialmente privados, é construído um sistema dinâmico associado e um grafo ponderado é usado para conectar subgrupos, permitindo identificar grupos de formas arbitrárias por meio da fusão de subgrupos. Mais especificamente, o método proposto segue três etapas: (i) estima-se de forma diferencialmente privada os parâmetros de uma função de densidade modelada como uma mistura de gaussianas com k_0 grupos; (ii) os vetores de equilíbrio de transição (do inglês, *transition equilibrium vectors* - TEVs) são identificados e um grafo ponderado conectando os centros é construído; (iii) por fim, os centros adjacentes são unidos até atingir o número desejado de grupos k .

3.1.5 Métodos hierárquicos

Em [220], foi proposta abordagem baseada no algoritmo BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*) [218], que é um algoritmo de agrupamento hierárquico voltado ao processamento de grandes volumes de dados, uma vez que demanda pouco uso de memória. Basicamente, ele constrói incrementalmente uma árvore hierárquica chamada CF-tree (*Clustering Feature tree*), utilizando uma abordagem *bottom-up* para decompor iterativamente o conjunto de dados até atender a certos critérios de parada. A CF-tree é composta por três tipos de nós: (i) nós internos, denominados *nonleaf*, que armazenam uma quantidade limitada de subnós; (ii) nós folhas, denominados *leaf*, que, de maneira análoga, armazenam um número limitado de *minclusters*; e (iii) *minclusters*, que são agrupamentos mínimos, cuja diâmetro não ultrapassa um limiar pré-estabelecido. Por divulgar características como diâmetros e distâncias de grupos, o BIRCH é vulnerável a ataques de inferência, em cenários em que o adversário possua acesso a fontes de dados auxiliares. Assim, os autores adotaram PD para propor o algoritmo Diff-BIRCH e algumas variantes, dentre as quais a variante denominada DBSLN-Diff-BIRCH (Distance-

Based-Split-Leaf-Node-Diff-BIRCH) [220, p.50-51] apresentou os melhores resultados.

Em [19] foi apresentado estudo que analisou conjuntos de dados reais de consumo de tabaco na Índia usando métodos de agrupamento dentro de uma arquitetura de aprendizado federado (*federated learning*), enfatizando a preservação da privacidade. Foi realizada comparação envolvendo três métodos de agrupamento bem conhecidos: k-means, DBSCAN e agrupamento hierárquico aglomerativo, avaliando seu desempenho e *overhead* de comunicação. O modelo de PD local foi adotado, aplicando o mecanismo de Laplace antes do envio das soluções de agrupamento para o servidor central. Embora o estudo tenha como objetivo a análise resultante e não foque em aspectos relacionados à combinação dos métodos de agrupamento com a privacidade diferencial, apresentou resultados comparativos interessantes. Para o problema abordado, o DBSCAN superou a performance dos outros métodos em três índices de validação relativos considerados, quais sejam: silhueta, Calinski-Harabasz e Davies-Bouldin. Cabe destacar que as diferenças foram pequenas e provavelmente não significativas estatisticamente, embora nenhum teste de hipótese tenha sido conduzido. Além disso, com a aplicação de PD, a performance dos métodos, em termos do índice silhueta, reduziu conforme o ϵ foi reduzido, como esperado, e de maneira similar para os três métodos de agrupamento. Um resultado especialmente interessante indicou que o DBSCAN foi o método que teve a performance, em termos do índice silhueta, mais deteriorada com a adição do ruído, passando a ocupar a última posição entre os três métodos considerados. Novamente, as diferenças entre os métodos foram pequenas, mas resultados como esse demonstram as potencialidades do k-means em contextos nos quais a proteção de dados sensíveis se faz necessária.

3.2 Considerações finais relativas à RSL

Os métodos de agrupamento são ferramentas fundamentais em aprendizado não supervisionado, com ampla aplicação em diversas áreas. Entre esses métodos, o k-means se destaca pela simplicidade e eficiência, sendo amplamente adotado em contextos práticos [114]. No entanto, quando se trata da proteção de dados sensíveis, especialmente em cenários com múltiplas partes envolvidas, a integração com PD é promissora e o k-means diferencialmente privado surge como uma das abordagens mais estudadas e aplicadas. A Tabela 3.1 lista as abordagens identificadas nos trabalhos revisados.

Apesar de limitações conhecidas, como sensibilidade a valores extremos (*outliers*) e dificuldade em identificar corretamente grupos de formatos não esféricos, o k-means

Tabela 3.1: Comparação das abordagens identificadas na revisão da literatura

Classe	Método	Abordagem	Modelo de PD	Referência
protótipo	k-means	DPk-means	global interativo	[23]
protótipo	k-means	DPk-means + DAS	global interativo	[54]
protótipo	k-means	EUGkM	global interativo/não interativo	[179, 180]
protótipo	k-means	DPLk-means	local interativo	[163]
protótipo	k-means	Feldman et al.	local não interativo	[74]
protótipo	k-modes	DP-Modes	global/local interativo/não interativo	[152]
protótipo	k-means	LSH-GoodCenter	global/local não interativo	[177]
protótipo	k-means	DPk-means + SSAS	global interativo	[70]
protótipo	k-means	DPk-means + APAS	global interativo	[70]
protótipo	k-means	LDPk-means	local interativo	[205]
protótipo	k-means	DPk-means ¹²	global interativo	[135]
protótipo	k-means	DP-KCCM	global interativo	[155]
protótipo	k-means	ADPk-means	global interativo	[95]
protótipo	k-means	DKdP	local não interativo	[208]
protótipo	k-means	DPk-means + Zhang	global interativo	[216]
protótipo	k-means	dist-DP k-means	global interativo	[63]
protótipo	k-means++	DP PSO k-means++	global interativo	[93]
protótipo	k-means	Ahmed	global/local interativo	[6]
protótipo	k-centers	Yuan et al.	global/local interativo	[214]
protótipo	k-means	DPk-means + GAPBAS	global interativo	[129]
protótipo	k-means	dist-DP k-means	global não interativo	[63]
protótipo	k-modes	FskLDP k-mode	local não interativo	[211]
protótipo	k-means	distK-GS	local não interativo	[122]
densidade	DBSCAN	DP-MCDBSCAN	global não interativo	[154, 153]
densidade	DPC	DP-ADPC e DP-UDPC	global não interativo	[38]
densidade	DBSCAN	DBDP	global não interativo	[203]
densidade	DENCLUS	DCT-DP	global não interativo	[128]
densidade	DBSCAN	DC-LDP	local não interativo	[81]
densidade	DBSCAN	DP-ADC	global não interativo	[219]
grade		GC-LDP	local não interativo	[80]
modelo	MoG	LDP-GMMC	local não interativo	[213]
modelo	MoG	DPMoG-hard	local não interativo	[29]
hierárquico	BIRCH	Diff-BIRCH	global interativo	[220]

Tabela elaborada pelo autor.

permanece amplamente utilizado (conforme apresentado na Tabela 3.1) devido à sua simplicidade, eficiência computacional e versatilidade em diversas aplicações [161, p.2]. Nesse contexto, o contínuo desenvolvimento de pesquisas voltadas ao seu aprimoramento, especialmente por meio de abordagens híbridas e técnicas de otimização que visam aumentar sua robustez e escalabilidade se justificam [161, p.2]. Ademais, a crescente preocupação com aspectos relativos à privacidade individual e ao uso responsável de dados [76] reforça a necessidade de alinhar o desempenho algorítmico a aspectos como transparência e preservação da privacidade [161, p.18]. Nesse sentido, a integração do k-means com recursos de proteção de privacidade se torna fundamental em aplicações que envolvem informações sensíveis [161, p.15], permitindo extrair valor analítico sem comprometer a confidencialidade.

No que se refere especificamente ao uso de PD, apesar dos avanços significativos alcançados nos últimos anos em relação ao k-means diferencialmente privado, com propostas que vão desde mecanismos de ruído otimizados até estratégias adaptativas, não existe levantamento comparativo abrangente que sintetize e analise criticamente os principais avanços no campo, especialmente no que se refere à alocação de *privacy budget* no modelo de PD global iterativo. Nesse contexto, a presente pesquisa contribuiu primeiramente no sentido de oferecer uma visão consolidada das estratégias de alocação de *privacy budget* mais relevantes. Além disso, o cuidadoso levantamento realizado ao longo do desenvolvimento da tese mostra que há pouca investigação a respeito da aplicação de técnicas de otimização nesse contexto, apesar de seu potencial promissor, conforme demonstrado pelas poucas abordagens existentes, que apresentaram resultados encorajadores.

A Tabela 3.2 apresenta um comparativo dos trabalhos identificados na RSL que abordam o problema de alocação de *privacy budget* no k-means diferencialmente privado. Observa-se que predominam abordagens baseadas em heurísticas determinísticas derivadas de análises teóricas, como a alocação uniforme e as baseadas em progressões aritméticas, que fixam a sequência de ϵ_t a priori e não exploram o espaço de soluções de forma ampla e sistemática. Abordagens adaptativas ajustam localmente o *privacy budget* ao longo das iterações, em função de medidas de qualidade geralmente baseadas em índices de validação internos, mas ainda consideram regras de atualização relativamente simples, sem tratar explicitamente a alocação de ϵ como um problema de otimização. Foi identificada uma única iniciativa abordando o problema dessa forma, que apresentou resultados promissores em relação às demais abordagens.

Tabela 3.2: Consolidado dos trabalhos que abordam a alocação de *privacy budget* no DPk-means.

Referência	Estratégia de alocação	Tipo	Mecanismo	Ruído
[23]	Limita o número total de consultas T e calibra o ruído globalmente. ϵ_t definido implicitamente via teorema de composição.	Heurística	Normal	Adicionado às consultas
[54, 70]	A cada iteração, aloca uma fração proporcional do ϵ remanescente, produzindo uma sequência decrescente de ϵ_t que prioriza as primeiras iterações.	Heurística	Laplace	Adicionado às consultas
[180]	Divide ϵ entre abordagem não iterativa e refinamento iterativo.	Heurística	Laplace	Adicionado às consultas
[70]	Progressão aritmética decrescente de ϵ por iteração, baseada em $\epsilon_{\min}$.	Heurística	Laplace	Adicionado às consultas
[155]	Particiona ϵ em 12 iterações com parcelas crescentes.	Adaptativa	Laplace	Adicionado às consultas
[95]	Distribui o ϵ de cada iteração entre agregados de grupos usando um índice de qualidade (CHS), realocando a cada iteração.	Adaptativa	Laplace	Adicionado às consultas
[129]	Vetor $(\epsilon_1, \dots, \epsilon_T)$ otimizado por algoritmo genético, minimizando NICV sob restrição de $\epsilon_{\min}$.	Otimização	Laplace	Adicionado às consultas

Tabela elaborada pelo autor.

Com isso, parte-se da hipótese de pesquisa de que o uso de metaheurísticas para determinar a alocação dos valores de *privacy budget* no DPk-means tende a produzir soluções de agrupamento de boa qualidade maximizando a utilidade, preservando as garantias formais de privacidade. Nesse contexto, o objetivo da pesquisa e sua principal contribuição foi investigar, mais profundamente, o uso de técnicas de otimização para o problema de alocação de *privacy budget* no contexto do k-means diferencialmente privado, adotando o modelo de PD global iterativo. O problema foi formulado como um problema de otimização, no qual se busca encontrar, entre as possíveis distribuições de *privacy budget* ao longo das iterações, aquela que maximize a utilidade da solução sob restrições de privacidade. Nesse cenário, conforme será mostrado mais à frente, a partir da análise dos resultados dos experimentos realizados nesta tese, o desenvolvimento de algoritmos baseados em metaheurísticas se mostrou eficaz, tanto pela capacidade de explorar eficientemente grandes espaços de busca quanto pela flexibilidade para incorporar restrições complexas.

Capítulo 4

Abordagem Proposta

O presente capítulo apresenta a abordagem proposta nesta tese para a resolução do problema de alocação de *privacy budget* no contexto do k-means diferencialmente privado no modelo de PD interativo. A proposta tem por base a aplicação das metaheurísticas BRKGA, ILS e VNS para otimização da alocação de (*privacy budget*), visando maximizar a utilidade dos agrupamentos gerados sob restrições de privacidade. Ainda neste sentido, busca-se contribuir para o aprimoramento do atual estado da arte em termos do equilíbrio entre privacidade e utilidade, na alocação de *privacy budget* em métodos iterativos.

No que diz respeito à forma de interação do analista com o conjunto de dados, a implementação de uma solução baseada em Privacidade Diferencial (PD) pode se dar, essencialmente, através de dois modelos, quais sejam: (i) modelo não interativo ou *offline*; e (ii) modelo interativo ou *online*, conforme descrito na seção 2.4.2. Nesta pesquisa, optou-se por desenvolver o modelo interativo, mas foi também elaborada uma proposta de abordagem de otimização associada ao modelo não interativo, com base no uso de *coresets*, conforme detalhado no Apêndice E (p.178). Essa abordagem, no entanto, não foi implementada. Neste capítulo, apresenta-se uma abordagem para o modelo interativo de PD, cujos resultados foram comparados às abordagens recentes mais relevantes identificadas na RSL apresentada no Capítulo 3, conforme experimentos apresentados no Capítulo 5.

Na execução do k-means no modelo interativo de PD, a atualização do centroide em cada iteração corresponde à execução de duas consultas¹ ao conjunto de dados. Assim, o k-means clássico deve ser adaptado de forma a garantir proteção de privacidade em cada uma de suas iterações. Para isso, é necessário fazer uso de um mecanismo de PD e distribuir

¹As consultas são referentes às operações de soma dos atributos dos objetos $x_i \in C_j$ em relação ao protótipo (centroide) c_j , denotada por S_j , e contagem (frequência) de objetos em C_j , denotada por $|C_j|$.

ou alocar o *privacy budget* entre as T iterações, de modo a controlar o ruído introduzido em cada etapa. Mais especificamente, em cada iteração do k-means, os centroides são atualizados com base nas médias dos atributos dos objetos.

Não obstante, antes de serem divulgadas ou reutilizadas na próxima iteração, essas estatísticas recebem adição de um ruído gerado segundo o mecanismo de PD adotado. Como essas atualizações ocorrem múltiplas vezes, o uso de composição² sequencial é necessário para garantir que o *privacy budget* total não seja excedido ao longo das iterações do k-means. Assim, o modelo iterativo, quando bem sucedido, permite que o algoritmo convirja progressivamente para centroides suficientemente próximos dos reais (não perturbados), oferecendo, concomitantemente, uma utilidade adequada e garantindo a privacidade dos indivíduos representados nos objetos agrupados. Esse objetivo pode ser alcançado através de um controle cuidadoso da alocação do *privacy budget*, uma vez que ruído excessivo nas primeiras iterações pode impedir a convergência ou levar a uma convergência prematura com resultados ruins, i.e., soluções de baixa qualidade, caracterizadas por grupos pouco coesos e/ou mal separados, enquanto ruído excessivo nas últimas iterações pode comprometer a utilidade da solução de agrupamento obtida.

4.1 Formulação proposta

No modelo iterativo de PD, o problema de alocação de *privacy budget* no contexto do DPk-means [23, p.5-6] pode ser definido como a seguir.

Seja D um conjunto de dados contendo n objetos x_i , $i \in \{1, \dots, n\}$, descritos por seus h atributos, e k o número de grupos C_j , $j \in \{1, \dots, k\}$, aos quais esses objetos devem ser alocados por meio de um algoritmo k-means em T iterações, sendo cada iteração³ definida pela realização de duas etapas principais, quais sejam: (i) alocar cada objeto ao grupo no qual o centroide correspondente apresenta menor dissimilaridade; e (ii) atualizar os centroides a partir da nova composição dos grupos.

²Composição de privacidade [44, p.43-44] refere-se ao princípio segundo o qual a aplicação sucessiva de mecanismos de privacidade diferencial consome cumulativamente o orçamento total de privacidade ϵ . No modelo iterativo, o ruído é gerado na iteração t segundo o parâmetro ϵ_t e o efeito combinado das iterações deve respeitar o *privacy budget* total $\epsilon = \sum_{t=1}^T \epsilon_t$ (composição sequencial). Esse conceito é essencial para garantir que, ao fim de todas as iterações, a privacidade dos dados ainda esteja protegida de acordo com os parâmetros originalmente definidos.

³No que se refere ao modelo de PD **iterativo** ou *online*, as múltiplas **iterações** do k-means representam um contexto onde o analista (ou o método utilizado por ele) interage com o conjunto de dados através de consultas feitas por intermédio do curador confiável, que faz uso de um mecanismo de PD para adicionar o ruído necessário às saídas das consultas realizadas.

Em cada iteração do k-means, a atualização ou cálculo dos centroides é perturbada através da aplicação de um mecanismo de PD às consultas referentes às operações de soma dos atributos dos objetos $x_i \in C_j$, denotada por $f_j^s(D) = S_j$, e contagem (frequência) de objetos em cada grupo C_j , denotada por $f_j^c(D) = |C_j|$, resultando nos valores perturbados dessas consultas, S_j^* e $|C_j^*|$, respectivamente. O cálculo do centroide perturbado c_j^* é dado então por $c_j^* = \frac{S_j^*}{|C_j^*|}$, para cada grupo C_j , $j = 1, \dots, k$. A magnitude do ruído gerado pelo mecanismo⁴ A na iteração t ($t \in \{1, \dots, T\}$) é dada pelo parâmetro ϵ_t , e distribuída entre as duas consultas sucessivas, de forma que $\epsilon_t = \epsilon_t^s + \epsilon_t^c$, sendo $\sum_{t=1}^T \epsilon_t \leq \epsilon$, em que $\epsilon > 0$ é o *privacy budget* total. Portanto, em uma determinada iteração t , as consultas de soma de atributos e frequência de objetos são dadas, respectivamente, por $S_j^* = f_j^s(D) + r_t^s$ e $|C_j^*| = f_j^c(D) + r_t^c$, sendo $r_t^s \sim A(\Delta f^s, \epsilon_t^s)$ e $r_t^c \sim A(\Delta f^c, \epsilon_t^c)$. O Apêndice B.6 (p.165) define formalmente as especificidades da aplicação dos mecanismos de Laplace e Normal à atualização dos centroides no DPk-means sob o modelo global iterativo de PD.

Para fins de simplificação, foi considerado que $\epsilon_t^s = \epsilon_t^c = \frac{\epsilon_t}{2} \forall t$. O objetivo então é distribuir ϵ nas parcelas ϵ_t , $t = \{1, \dots, T\}$, de forma a garantir que o algoritmo satisfaça PD, maximizando a utilidade, i.e., maximizando a qualidade da solução de agrupamento em termos de uma métrica ou índice de validação previamente especificado. Tipicamente, seguindo o que está preconizado na literatura, são utilizados o índice de Calinski-Harabasz (CH) [30] ou *Normalized Intra-cluster Variance* (NICV).

Adotando o NICV como métrica de qualidade para as soluções de agrupamento, o problema de alocação de *privacy budget*, no contexto do k-means, consiste então em determinar uma sequência de valores $[\epsilon_1, \dots, \epsilon_t, \dots, \epsilon_T]$, tal que $\sum_{t=1}^T \epsilon_t \leq \epsilon$ e minimize o NICV, dado por:

$$\text{NICV} = \frac{1}{n} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2$$

Este problema pode ser abordado através de alocação adaptativa [155, 95], heurísticas [54, 70, 216] ou otimização [129]. Nesta pesquisa foi adotada abordagem baseada em otimização, formalizada a seguir.

$$\min f_{\text{obj}} = \frac{1}{n} \sum_{j=1}^k \sum_{x_i \in C_j} \left\| x_i - \frac{\sum_{x_i \in C_j} x_i + r_T^s}{|C_j| + r_T^c} \right\|^2 \quad (4.1)$$

⁴Os mecanismos utilizados foram definidos na Seção 2.4.3 (p.32).

Sendo r_T^s e r_T^c variáveis aleatórias⁵ tais que:

$$r_T^s \sim A(\Delta f^s, \epsilon_T/2); \quad r_T^c \sim A(\Delta f^c, \epsilon_T/2).$$

Sujeito a:

$$\sum_{t=1}^T \epsilon_t = \epsilon \quad (4.2)$$

Em que a Eq. 4.1 representa a função objetivo e Eq. 4.2 é a restrição que satisfaz PD⁶, sendo $\epsilon_t \in \mathbb{R}$, $t = \{1, \dots, T\}$, as variáveis de decisão.

Alguns dos trabalhos da literatura [129, 179, 54] consideram restrições adicionais, não intrínsecas ao problema. Essas restrições visam garantir ou priorizar aspectos apontados como relevantes no que diz respeito a preservar a utilidade da solução de agrupamento obtida. Na formulação proposta, a utilidade é garantida mediante a minimização do NICV, representado na função objetivo (Eq. 4.1), e o impacto de tais restrições sobre o espaço de busca é desconhecido, podendo acarretar a eliminação indevida de regiões potencialmente promissoras. Assim, apenas a restrição referente à garantia de PD, dada pela Eq. 4.2, foi considerada na formulação apresentada. No Apêndice D (p.174), as restrições adicionais foram discutidas e uma formulação alternativa, contemplando tais restrições, foi apresentada.

Na formulação proposta, o número máximo de iterações, T , deve ser determinado previamente. Na literatura, duas abordagens são tipicamente utilizadas. A primeira consiste em determinar um valor mínimo para o *privacy budget*, ϵ_m , em qualquer iteração [179, p.28-29][129], conforme detalhado no Apêndice B.3.6 (p.160), e, em seguida, determinar T em função desse parâmetro, conforme descrito na Eq. 4.3.

$$T = \begin{cases} \left\lfloor \frac{\epsilon}{\epsilon_m} \right\rfloor, & \text{se } \epsilon > 2\epsilon_m \\ 2, & \text{caso contrário} \end{cases} \quad (4.3)$$

⁵As componentes r^s e r^c representam ruídos gerados através do mecanismo de PD A , segundo uma distribuição de probabilidade parametrizada em função da sensibilidade da respectiva consulta ($\Delta f \cdot$) e da parcela de *privacy budget* alocada para sua execução. Os mecanismos utilizados foram definidos na Seção 2.4.3 (p.32).

⁶Essa restrição é baseada no conceito de composição sequencial de privacidade [44, p.43-44] e, a rigor, é satisfeita pela expressão $\sum_{t=1}^T \epsilon_t \leq \epsilon$. No entanto, a desigualdade permitiria que parte do *privacy budget* não fosse alocado, resultando no comprometimento da utilidade da solução final. Assim, a fim de garantir que todo o *privacy budget* disponível seja utilizado em cada execução, a expressão apresentada na Eq. 4.2 foi adotada.

Porém, ao permitir que T varie entre as execuções, corre-se o risco de obter um NICV elevado, i.e., uma solução de baixa qualidade, simplesmente em função de um número inadequado de iterações do DPk-means. A Eq. 4.3 estabelece como limite mínimo $T = 2$, o que é geralmente insuficiente para a execução do método de agrupamento [179, p.29], independentemente da adição de ruído. Por outro lado, não estabelece limite máximo, o que pode também ser prejudicial na obtenção de soluções de agrupamento de boa qualidade [179, p.28], uma vez que, se for adotado o número máximo de iterações como critério de parada, cada iteração contará com *privacy budget* igual a ϵ_m , o que tende a adicionar mais ruído do que o necessário em cada iteração. Além disso, o ruído adicionado em cada iteração pode dominar a melhoria do centroide [179, p.29], resultando em uma distorção cada vez maior. Já se outro critério de parada for adotado, é possível que parte do ϵ não seja utilizada, fazendo novamente com que mais ruído seja adicionado nas iterações concluídas e prejudique a maximização da utilidade.

Ao fixar T é possível controlar esse efeito, uma vez que o resultado do DPk-means passa a depender apenas da alocação de *privacy budget*, permitindo uma comparação adequada entre diferentes métodos. Assim, foi adotado número fixo de iterações. No que se refere à definição do valor de T , tomando como referência os limites propostos em [180, p.9] e [155, p.208], foi adotado $T = 10$.

A alocação de *privacy budget* na execução do k-means diferencialmente privado pode então ser definida como um problema de otimização contínua, em que as variáveis de decisão correspondem às parcelas de *privacy budget* atribuídas a cada iteração do algoritmo. Nesse contexto, metaheurísticas ganham destaque em razão das características da função-objetivo que, através da formulação proposta, relaciona a alocação de *privacy budget* à qualidade final do agrupamento, em termos do NICV, por meio de uma execução completa do DPk-means. Essa função apresenta não linearidade – decorrente da relação inversa entre ϵ_t e a magnitude do ruído adicionado –, descontinuidades, não convexidade e multimodalidade associadas à alocação iterativa de objetos aos grupos e comportamento estocástico oriundo do próprio mecanismo de PD. Uma vez que as soluções do problema podem ser convenientemente representadas através de chaves aleatórias, conforme descrito na Seção 4.3, foi utilizado o *framework* Random-Key Optimizer (RKO) [37], dotado de estrutura modular com recursos de decodificação, diversificação e intensificação, e que permitiu a aplicação direta de diferentes metaheurísticas à resolução da formulação proposta.

Considerando o rico arcabouço de metaheurísticas disponível na literatura, o Algoritmo

ritmo Genético de Chaves Aleatórias Viciadas (do inglês, *Biased Random Key Genetic Algorithm* - BRKGA) [89] foi identificado como uma estratégia particularmente promissora por combinar representação vetorial contínua, coerente com a natureza da variável de decisão do problema abordado, representada por $\epsilon_t \in \mathbb{R}$ ($t = \{1, \dots, T\}$), com mecanismos de seleção probabilísticos, que favorecem a convergência sem comprometer a diversidade da população constituída por um conjunto de soluções. Além disso, sua simplicidade de implementação e boa escalabilidade o tornam atrativo para variadas aplicações [133], incluindo aquelas relacionadas à otimização de parâmetros em espaços de busca contínuos [65, 69] ou mistos [115]. Adicionalmente, foram consideradas as metaheurísticas Busca Local Iterada (do inglês, *Iterated Local Search* - ILS) e Busca em Vizinhança Variável (do inglês, *Variable Neighborhood Search* - VNS), por sua capacidade de promover um equilíbrio entre intensificação e diversificação, através da combinação de busca local e perturbações controladas no caso do ILS, e da exploração sistemática de diferentes estruturas de vizinhança no caso do VNS, estratégias reconhecidas de forma consagrada como eficazes em problemas de otimização [134, p.322][27, p.2]. Com isso, foram implementados três algoritmos, descritos na Seção 4.5, cujos resultados dos experimentos computacionais são apresentados e analisados no Capítulo 5, no qual é realizada a sua comparação com as principais estratégias da literatura.

4.2 *Framework* Random-Key Optimizer (RKO)

O Random-Key Optimization (RKO) [37] é um *framework* genérico para a resolução de problemas de otimização, baseado na representação de soluções por vetores de chaves aleatórias. A abordagem propõe uma separação entre a representação da solução e sua interpretação no domínio do problema, realizada por meio da aplicação de um procedimento de decodificação. Essa separação confere maior flexibilidade e modularidade no desenvolvimento de algoritmos, permitindo a reutilização de componentes e a integração de diferentes metaheurísticas e procedimentos de busca disponíveis no RKO.

O procedimento de decodificação f_{decod} é um procedimento determinístico que recebe como entrada um vetor de chaves aleatórias e o transforma em uma solução viável para o problema de otimização em questão. Assim, dado um vetor u , de comprimento T , onde $u_i \in [0, 1] \forall i \in \{1, \dots, T\}$, a qualidade da solução é avaliada por uma função objetivo f_{obj} aplicada à solução decodificada, i.e., $f_{\text{obj}}(f_{\text{decod}}(u))$. Essa estrutura permite que diferentes metaheurísticas operem sobre o espaço contínuo das chaves aleatórias, enquanto o conhecimento específico do problema é encapsulado no decodificador. Para o problema

de alocação de *privacy budget* no contexto do DPK-means, foram propostos o decodificador e o procedimento para avaliação das soluções descritos, respectivamente, nas Seções 4.3 e 4.4.

No que se refere ao processo de otimização realizado pelo RKO, este se dá em torno de três operadores fundamentais reutilizáveis, além da geração de soluções iniciais, quais sejam: (i) perturbação (*shaking*), descrito na Seção 4.2.1; (ii) recombinação (*blending*), descrito na Seção 4.2.2; e (iii) busca local, descrito na Seção 4.2.3. Além disso, mantém um *pool* compartilhado de soluções elite. Esse *pool* atua como memória global, promovendo intensificação e diversificação ao longo do processo⁷.

Inicialmente, um conjunto de soluções (vetores de chaves aleatórias) é gerado e refinado por heurísticas de busca local. Ao longo da execução, novas soluções produzidas pelas metaheurísticas podem ser inseridas no *pool*, desde que não sejam duplicatas e apresentem qualidade competitiva em relação ao valor da função objetivo. Para preservar diversidade, soluções idênticas podem ser perturbadas via *shaking*.

A flexibilidade do RKO decorre da sua compatibilidade com diferentes metaheurísticas (como BRKGA, GRASP, ILS, VNS, PSO, entre outras), que operam sobre a mesma estrutura de representação de soluções e compartilham os mesmos operadores básicos. Além disso, o *framework* permite tanto o ajuste estático dos parâmetros de cada metaheurística quanto o uso de estratégias adaptativas, através do método Q-Learning [196][37, p.23-24]. Nas subseções a seguir, são descritos de forma resumida os operadores do RKO, bem como os procedimentos de decodificação e avaliação das soluções.

4.2.1 Perturbação

O operador de *shaking* [37, p.12] tem como objetivo introduzir diversidade na busca por meio de perturbações controladas em uma solução representada por um vetor de chaves aleatórias. Esse procedimento é particularmente importante para escapar de ótimos locais potencialmente de baixa qualidade.

A intensidade da perturbação é controlada por uma taxa $\beta \in [\beta_{\min}, \beta_{\max}]$, que determina o número de modificações aplicadas ao vetor u . Em cada modificação, um dos seguintes movimentos de vizinhança é selecionado aleatoriamente:

⁷As soluções armazenadas nesse *pool* podem ser reutilizadas por diferentes metaheurísticas executadas em paralelo, bem como por múltiplas execuções de uma mesma metaheurística, permitindo que informações de soluções promissoras sejam compartilhadas entre os procedimentos de busca.

- *Swap*: troca dos valores de duas posições aleatórias;
- *Swap Neighbor*: troca entre posições adjacentes;
- *Mirror*: substituição de um valor pelo seu complementar;
- *Random*: substituição por um novo valor aleatório (no intervalo $[0,1)$).

Ao final do processo, obtém-se um novo vetor perturbado u' , que deve ser posteriormente decodificado para avaliação. Esse operador é amplamente utilizado tanto para diversificação quanto como mecanismo auxiliar em outras etapas do processo.

4.2.2 Recombinação

O operador de recombinação (*blending*) [37, p.13] é responsável pela geração de novos vetores de chaves aleatórias u^c a partir da combinação de dois outros vetores u^a e u^b . Trata-se de uma generalização do operador de cruzamento uniforme, incorporando elementos estocásticos adicionais.

Para cada posição i do vetor u , decide-se, com probabilidade ρ_e , qual dos vetores ascendentes, u^a ou u^b , contribuirá com o valor correspondente para o vetor descendente u^c . Mais especificamente, um vetor auxiliar de mesmo comprimento dos ascendentes, contendo valores reais no intervalo $[0,1]$, é gerado e, para todas as posições do vetor auxiliar cujo valor for inferior a ρ_e , u^c recebe a posição correspondente de u^a , caso contrário, u^c recebe a posição correspondente de u^b . Além disso, um parâmetro adicional denominado *factor* permite modificar a contribuição do segundo ascendente (u^b): quando *factor* é igual a 1, o valor original do ascendente (u_i^b) é usado; quando *factor* é igual a -1, utiliza-se o valor complementar $(1 - u_i^b)$.

Adicionalmente, com uma pequena probabilidade $\mu = 0,05$, um novo valor aleatório é gerado para a posição i , promovendo diversidade na população. O vetor resultante u^c combina, portanto, herança e mutação.

4.2.3 Busca local

A busca local no RKO é realizada usando o método *Random Variable Neighborhood Descent* (RVND) [37, p.13-18], uma variante do método *Variable Neighborhood Descent* (VND) [97, p.762]. A principal ideia é explorar múltiplas estruturas de vizinhança de forma sistemática e adaptativa.

O RVND mantém uma lista de vizinhanças disponíveis e, a cada iteração, seleciona uma delas aleatoriamente. Caso uma solução melhor seja encontrada, ela substitui a solução atual e a lista de vizinhanças é reinicializada. Caso contrário, a vizinhança é removida da lista. O processo continua até que todas as vizinhanças tenham sido exploradas sem melhoria.

No contexto do RKO, estão disponibilizadas quatro heurísticas de busca local, cada uma definindo uma estrutura de vizinhança distinta, conforme descrito a seguir.

- *Swap Local Search*: realiza trocas entre pares de posições do vetor de chaves aleatórias, explorando permutações;
- *Mirror Local Search*: substitui valores por seus complementares, promovendo movimentos simétricos no espaço;
- *Farey Local Search*: ajusta valores com base em intervalos definidos pela sequência de Farey [92];
- *Nelder-Mead Local Search*: é uma adaptação da heurística Nelder-Mead [151], na qual o operador de recombinação é combinado com transformações simplex (reflexão, expansão, contração e redução) para explorar o espaço de busca de chaves aleatórias.

4.3 Decodificador proposto

Uma solução para o problema de alocação de *privacy budget* na execução do DPk-means consiste em um vetor de T posições, correspondendo às T iterações do método de agrupamento, no qual cada posição armazena a parcela de ϵ alocada à respectiva iteração. Por exemplo, considerando os parâmetros $\epsilon = 10$ e $T = 7$, uma solução viável é representada pelo vetor s com os termos s_t correspondentes aos termos ϵ_t .

$$s = [1,401, 1,824, 1,538, 1,390, 1,082, 1,648, 1,117]$$

Um vetor de chaves aleatórias, por sua vez, é um vetor de números reais u no intervalo $[0,1]$, gerado segundo uma distribuição uniforme, por exemplo, com o vetor a seguir.

$$u = [0,549, 0,715, 0,603, 0,545, 0,424, 0,646, 0,438]$$

Conforme já explicado, o procedimento de decodificação f_{decod} é um procedimento determinístico que recebe como entrada um vetor de chaves aleatórias e o transforma em uma solução viável para o problema de otimização em questão. Portanto, no problema estudado, sendo ϵ o *privacy budget* total e T o número de iterações, o decodificador deve produzir, a partir de um vetor de chaves aleatórias u , com T posições, um vetor resultante s , de mesmo comprimento, tal que $\sum_{t=1}^T s_t \leq \epsilon$, garantindo a restrição estabelecida na formulação, dada pela Eq. 4.2 (p.67).

Essa restrição pode ser garantida através da normalização de u , a fim de que a t -ésima posição de u represente a proporção de ϵ alocada a cada iteração t , ou seja,

$$u_t^N \leftarrow \frac{u_t}{\sum_{\tau=1}^T u_{\tau}}, \forall t \in \{1, \dots, T\}$$

Por exemplo, ainda considerando os parâmetros $\epsilon = 10$ e $T = 7$, o vetor de chaves aleatórias $u = [0,549, 0,715, 0,603, 0,545, 0,424, 0,646, 0,438]$ corresponde ao vetor normalizado $u^N = [0,1401, 0,1824, 0,1538, 0,1390, 0,1082, 0,1648, 0,1117]$ que, por sua vez, corresponde à solução

$$s = f_{\text{decod}}(\epsilon, u) = \epsilon \cdot u^N = [1,401, 1,824, 1,538, 1,390, 1,082, 1,648, 1,117]$$

Dessa forma, a saída produzida na etapa de decodificação é o vetor s contendo as parcelas (ϵ_t) do *privacy budget* total alocadas às T iterações do DPk-means.

4.4 Função de custo (avaliação)

Uma vez obtida uma alocação de *privacy budget*, é necessário avaliar o seu ajuste⁸. Para tal, o DP-kmeans foi executado para obter uma solução de agrupamento sob PD, considerando o valor de ϵ_t alocado a cada iteração para definir a magnitude do ruído gerado pelo mecanismo de PD adotado e esse ruído foi adicionado na atualização dos centroides. Mais especificamente, na iteração t do DPk-means, a soma de distâncias e frequência de objetos alocados ao grupo C_j é dada, respectivamente, por $S_j^* = S_j + \text{Lap}\left(\frac{\Delta f}{\epsilon_t/2}\right)$ e $|C_j^*| = |C_j| + \text{Lap}\left(\frac{\Delta f}{\epsilon_t/2}\right)$. O centroide do grupo C_j é então definido por $c_j^* = \frac{S_j^*}{|C_j^*|} \forall j \in \{1, \dots, k\}$.

Como descrito na Seção 4.1, o NICV foi adotado como métrica de qualidade para

⁸Na terminologia empregada originalmente pelos autores do RKO, a função que calcula a qualidade ou ajuste de uma solução é denominada função de custo (*cost function*).

as soluções de agrupamento, representando a função objetivo, dada pela Eq. 4.1 (p.66). Assim, para cada solução de agrupamento produzida pelo DPk-means, o NICV correspondeu ao valor da função de custo atribuído ao vetor de chaves aleatórias que definiu a alocação de *privacy budget* empregada na obtenção dessa solução.

4.5 Algoritmos propostos

Nesta seção são descritos os três algoritmos propostos para otimização da alocação de (*privacy budget*), tendo por base a aplicação das metaheurísticas BRKGA, ILS e VNS.

4.5.1 BRKGAPD

A metaheurística *Biased Random Key Genetic Algorithm* (BRKGA) [89, 90] utiliza vetores de chaves aleatórias u que, uma vez decodificados, representam cromossomos associados a soluções s de um problema de otimização. Na primeira iteração do BRKGA (geração $g = 1$), é gerada uma população inicial de p vetores de chaves aleatórias. Esses vetores são decodificados em soluções viáveis para o problema, que são avaliadas segundo uma função de *fitness* (ou aptidão), a fim de medir a aptidão dos vetores associados a cada solução. Os vetores de chaves aleatórias são então divididos em dois conjuntos, um denominado elite (E), reunindo os p_e vetores associados aos cromossomos com melhor valor de *fitness*, e outro, denominado não elite ($\bar{E}$), reunindo os $p - p_e$ restantes, sendo $p_e < p - p_e$.

Para as gerações seguintes, em particular na geração $g + 1$, os p_e vetores do conjunto elite da geração anterior (g) são copiados para a população da nova geração ($g + 1$) e p_m vetores mutantes são gerados através do mesmo procedimento adotado para gerar os vetores da população inicial. Os demais $p - p_e - p_m$ vetores de chaves aleatórias são descendentes produzidos através da aplicação do operador de cruzamento uniforme [175].

No BRKGA original, o cruzamento é feito sempre entre um vetor de chaves aleatórias pertencente ao conjunto elite e um pertencente ao conjunto não elite. Além disso, a definição do gene (chave aleatória) que será herdado pelo descendente é feita através de uma seleção viciada. Um vetor de números reais é gerado segundo uma distribuição uniforme no intervalo $[0,1]$ e quando o valor na i -ésima posição desse vetor é menor ou igual à probabilidade de cruzamento ρ_e , $\rho_e > 0,5$, o descendente herda, na posição i , o valor na i -ésima posição do vetor proveniente do conjunto elite. Caso contrário, herda o valor na i -ésima posição do vetor proveniente do conjunto não elite. O parâmetro ρ_e define, portanto, a probabilidade de o valor relativo a uma posição do vetor proveniente

do conjunto elite ser herdado pelo vetor descendente.

Uma vez completada a população da geração $g + 1$, seus vetores são novamente decodificados, avaliados segundo a função de *fitness* e mais uma vez divididos nos conjuntos elite e não elite. Esse procedimento se repete até que um critério de parada seja atingido. São critérios de parada comuns o número máximo de gerações, tempo máximo de execução e um valor alvo para a função de *fitness*. Por fim, também pode ser estabelecido como critério um número máximo de gerações sem melhoria, i.e., sem redução no valor de *fitness*, para um problema de minimização.

Como descrito na Seção 4.4, o valor de *fitness* atribuído ao vetor de chaves aleatórias que originou uma alocação de *privacy budget* é dado pelo NICV da solução de agrupamento resultante da aplicação do DPK-means considerando essa alocação. A Figura 4.1 apresenta o fluxo do algoritmo BRKGA clássico adaptado ao problema, com uma visão geral das etapas realizadas.

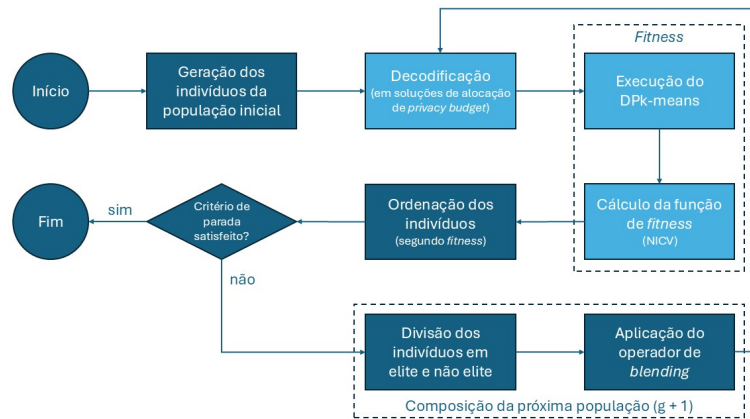


Figura 4.1: Fluxo de execução do algoritmo BRKGA proposto.

Por adotar abordagem evolutiva, já baseada em vetores de chaves aleatórias, o BRKGA é incorporado ao *framework* RKO preservando a sua lógica original. Adicionalmente, são introduzidos mecanismos de intensificação da busca, visando melhorar a exploração de regiões promissoras do espaço de soluções. Em particular, o mecanismo de divisão da população nos conjuntos elite e não elite é combinado com o operador *blending* (Seção 4.2.2), responsável pela geração de indivíduos mutantes (perturbação) e descendentes (intensificação).

Conforme já comentado, o *framework* conta ainda com o operador *shaking* (Seção 4.2.1) e o procedimento de busca local RVND (Random Variable Neighborhood Descent) (Seção 4.2.3), porém esses não foram incorporados na implementação utilizada no expe-

rimento. A justificativa para a não inclusão é que esses procedimentos representam um aumento significativo no custo computacional, em especial o RVND, cujo uso implica múltiplas avaliações adicionais da função de *fitness*, já reconhecida como o principal gargalo do problema. Além disso, buscou-se preservar a essência do BRKGA como uma metaheurística populacional baseada predominantemente em operadores genéticos, evitando a introdução de mecanismos intensivos de busca local que poderiam descaracterizar seu comportamento original. O Algoritmo 1 corresponde ao algoritmo BRKGA desenvolvido para o problema.

No que se refere à complexidade de tempo⁹, o Algoritmo BRKGAPD tem complexidade $O(G_{tmax}pnkhT)$, sendo G_{tmax} o número de gerações executadas até o limite de tempo, p o tamanho da população, n o número de objetos do conjunto de dados, k o número de grupos, h o número de atributos e T o comprimento de u (número de iterações executadas pelo DPk-means). Uma vez que k , h e T tendem a ser dominados por p e n , a complexidade de tempo do BRKGAPD pode ser dada por $O(G_{tmax}pn)$.

⁹A análise de complexidade de tempo do Algoritmo BRKGAPD é detalhada no Apêndice F.1, p. 186.

Algoritmo 1 BRKGAPD para otimização de alocação de privacy budget

Entrada: Tempo limite $t_{\max}$, parâmetros iniciais p , p_e , ρ_e , $pool$ de soluções

Saída: Melhor vetor de chaves u^*

```

1: Inicialize a população  $P$  com  $p$  vetores de chaves aleatórias
2: Inicialize a melhor solução  $(u^*, f^*) \leftarrow (\emptyset, +\infty)$ 
3: Inicialize  $q$ -table
4:  $t_{\text{acum}} \leftarrow 0$ 
5: enquanto  $t_{\text{acum}} < t_{\max}$  faça
6:   Atualize parâmetros  $p$ ,  $p_e$  e  $\rho_e$  via Q-Learning
7:   se  $t_{\text{acum}} > t_{\max}/2$  e sem reinicialização então
8:     Reinicialize população  $P$ 
9:   end se
10:   $P_{\text{eval}} \leftarrow \emptyset$ 
11:  para todo  $u \in P$  faça
12:     $s \leftarrow \text{DECODIFICADOR}(u)$ 
13:     $f \leftarrow \text{CUSTO}(s)$ 
14:    Adicione  $(u, f)$  em  $P_{\text{eval}}$ 
15:    se  $f < f^*$  então
16:       $(u^*, f^*) \leftarrow (u, f)$ 
17:      Atualize  $pool$  com  $(f^*, u^*)$ 
18:    end se
19:    se  $\text{STOP\_CONDITION}(f^*)$  então
20:      retorne  $u^*$ 
21:    end se
22:  end para
23:  Ordene  $P_{\text{eval}}$  por  $f$ 
24:   $E \leftarrow$  melhores  $p_e$  soluções de  $P_{\text{eval}}$ 
25:   $P_{\text{new}} \leftarrow$  cópia de  $E$ 
26:  enquanto  $|P_{\text{new}}| < p$  faça
27:    Selecione  $u_1$  do  $pool$  ou de  $P$  (probabilidade uniforme)
28:    Selecione  $u_2 \in E$ 
29:     $u' \leftarrow \text{BLENDING}(u_1, u_2, \rho_e)$ 
30:    Aplique mutação aleatória em  $u'$  com probabilidade  $\mu = 0,05$ 
31:    Adicione  $u'$  a  $P_{\text{new}}$ 
32:  end enquanto
33:   $P \leftarrow P_{\text{new}}$ 
34:  Atualize  $q$ -table
35:  Atualize  $t_{\text{acum}}$ 
36: end enquanto
37: retorne  $u^*$ 

```

4.5.2 ILSPD

A metaheurística *Iterated Local Search* (ILS) [178] é uma abordagem baseada na exploração iterativa do espaço de soluções por meio da aplicação repetida de um procedimento de busca local. Seu princípio fundamental consiste em gerar uma sequência de soluções, na qual cada nova solução é obtida a partir da perturbação de uma solução previamente encontrada, seguida da aplicação de um método de busca local. Isso permite refinar ou aprimorar as soluções ao longo das iterações.

O ILS consiste de três componentes principais, quais sejam: (i) procedimento de perturbação (exploração/diversificação), (ii) método de busca local (intensificação) e (iii) critério de aceitação. Inicialmente, uma solução é construída e refinada por meio de busca local. Em seguida, essa solução é perturbada, produzindo uma solução intermediária que é novamente submetida à busca local. O critério de aceitação determina se ela substituirá a solução corrente. A principal motivação do ILS é superar a limitação inerente aos métodos de busca local clássicos, que tendem a ficar presos em ótimos locais de baixa qualidade, explorando o espaço de soluções de forma mais eficiente do que estratégias baseadas em reinicialização aleatória.

O RKO adota uma implementação do ILS [37, p.21] na qual seus componentes são adaptados/reinterpretados para operar no espaço contínuo (relacionado ao vetor de chaves aleatórias). O processo iterativo alterna entre intensificação e diversificação por meio de dois mecanismos principais: (i) *shaking* (Seção 4.2.1), responsável pela perturbação da solução atual, cuja intensidade é controlada pelo parâmetro β , amostrado aleatoriamente dentro de um intervalo limitado pelos hiperparâmetros $\beta_{\min}$ e $\beta_{\max}$; e (ii) o método de busca local *Random Variable Neighborhood Descent* (RVND) (Seção 4.2.3). O Algoritmo 2 detalha o pseudocódigo do algoritmo ILS proposto.

No que se refere à complexidade de tempo¹⁰ do Algoritmo ILSPD, seu custo é dado por $O(G_{tmax}T^3nkh)$, sendo G_{tmax} o número de gerações executadas até o limite de tempo, n o número de objetos do conjunto de dados, k o número de grupos, h o número de atributos e T o comprimento de u (número de iterações executadas pelo DPk-means). Uma vez que k e h tendem a ser dominados por n , a complexidade de tempo do ILSPD pode ser dada por $O(G_{tmax}T^3n)$.

¹⁰A análise de complexidade de tempo do Algoritmo ILSPD é detalhada no Apêndice F.2, p. 188.

Algoritmo 2 ILSPD para otimização de alocação de privacy budget**Entrada:** Tempo limite $t_{\max}$, parâmetros $\beta_{\min}$, $\beta_{\max}$, $pool$ **Saída:** Melhor vetor de chaves u^*

```

1: Gere solução inicial  $u$ 
2:  $u \leftarrow \text{RVND}(u, \text{DECODIFICADOR}, pool)$ 
3:  $s \leftarrow \text{DECODIFICADOR}(u)$ 
4:  $f^* \leftarrow \text{CUSTO}(s)$ 
5:  $u^* \leftarrow u$ 
6: Atualize  $pool$  com  $(f^*, u^*)$ 
7: Inicialize  $q\text{-table}$ 
8:  $t_{\text{acum}} \leftarrow 0$ 
9: enquanto  $t_{\text{acum}} < t_{\max}$  faça
10:   Atualize  $\beta_{\min}, \beta_{\max}$  via  $Q\text{-Learning}$ 
11:    $u' \leftarrow \text{SHAKING}(u^*, \beta_{\min}, \beta_{\max})$ 
12:    $u'' \leftarrow \text{RVND}(u', \text{DECODIFICADOR}, pool)$ 
13:    $s'' \leftarrow \text{DECODIFICADOR}(u'')$ 
14:    $f'' \leftarrow \text{CUSTO}(s'')$ 
15:   se  $f'' \leq f^*$  então
16:      $(u^*, f^*) \leftarrow (u'', f'')$ 
17:     Atualize  $pool$  com  $(f^*, u^*)$ 
18:   end se
19:   se  $\text{STOP\_CONDITION}(f^*)$  então
20:     retorne  $u^*$ 
21:   end se
22:   Atualize  $q\text{-table}$ 
23:   Atualize  $t_{\text{acum}}$ 
24: end enquanto
25: retorne  $u^*$ 

```

4.5.3 VN SPD

A metaheurística *Variable Neighborhood Search* (VNS) [97] é baseada na ideia de explorar sistematicamente diferentes estruturas de vizinhança durante o processo de busca, com o objetivo de escapar de ótimos locais de qualidade inferior e aumentar a probabilidade de encontrar soluções de melhor qualidade. A variante básica do VNS envolve três componentes principais: perturbação (*shaking*), busca local e mudança de vizinhança. A partir de uma solução, um algoritmo VNS inicialmente aplica uma perturbação aleatória dentro de uma vizinhança específica, gerando uma nova solução. Em seguida, essa solução é refinada por meio de um procedimento de busca local. Por fim, um mecanismo de decisão avalia se a nova solução deve substituir a solução atual. Caso haja melhoria, a busca retorna à vizinhança inicial; caso contrário, o algoritmo passa a explorar uma vizinhança mais ampla.

No *framework* RKO, o VNS [37, p.22] é adaptado por meio da redefinição das estruturas de vizinhança no espaço das chaves aleatórias. As vizinhanças passam a ser definidas

pela intensidade da perturbação aplicada ao vetor de chaves aleatórias. Mais especificamente, o procedimento de *shaking* (Seção 4.2.1) utiliza o parâmetro β , definido pelo produto entre o índice da vizinhança atual k e o hiperparâmetro $\beta_{\min}$, i.e., $\beta = k \times \beta_{\min}$, de modo que vizinhanças mais distantes correspondem a perturbações mais intensas. Após cada perturbação, a solução é refinada utilizando o RVND (Seção 4.2.3). Caso uma solução melhor seja encontrada, a busca retorna à vizinhança inicial ($k = 1$), caso contrário avança para a próxima vizinhança ($k = k + 1$), até o limite definido pelo hiperparâmetro $k_{\max}$. O Algoritmo 3 detalha o pseudocódigo do algoritmo VNS proposto.

Algoritmo 3 VNSPD para otimização de alocação de privacy budget

Entrada: Tempo limite $t_{\max}$, parâmetros $k_{\max}$, $\beta_{\min}$, $pool$

Saída: Melhor vetor de chaves u^*

```

1: Gere solução inicial  $u$ 
2:  $u \leftarrow \text{RVND}(u, \text{DECODIFICADOR}, pool)$ 
3:  $s \leftarrow \text{DECODIFICADOR}(u)$ 
4:  $f_{\text{curr}} \leftarrow \text{CUSTO}(s)$ 
5:  $u_{\text{curr}} \leftarrow u$ 
6:  $(u^*, f^*) \leftarrow (u_{\text{curr}}, f_{\text{curr}})$ 
7: Atualize  $pool$ 
8: Inicialize  $q\text{-table}$ 
9:  $t_{\text{acum}} \leftarrow 0$ 
10: enquanto  $t_{\text{acum}} < t_{\max}$  faça
11:   Atualize  $k_{\max}, \beta_{\min}$  via  $Q\text{-Learning}$ 
12:    $k \leftarrow 1$ 
13:   enquanto  $k \leq k_{\max}$  faça
14:     se  $\text{STOP\_CONDITION}(f^*)$  então
15:       retorne  $u^*$ 
16:     end se
17:      $u' \leftarrow \text{SHAKING}(u_{\text{curr}}, k \cdot \beta_{\min}, (k + 1) \cdot \beta_{\min})$ 
18:      $u'' \leftarrow \text{RVND}(u', \text{DECODIFICADOR}, pool)$ 
19:      $s'' \leftarrow \text{DECODIFICADOR}(u'')$ 
20:      $f'' \leftarrow \text{CUSTO}(s'')$ 
21:     se  $f'' < f_{\text{curr}}$  então
22:        $(u_{\text{curr}}, f_{\text{curr}}) \leftarrow (u'', f'')$ 
23:        $k \leftarrow 1$ 
24:       se  $f'' < f^*$  então
25:          $(u^*, f^*) \leftarrow (u'', f'')$ 
26:         Atualize  $pool$ 
27:       end se
28:     senão
29:        $k \leftarrow k + 1$ 
30:     end se
31:   end enquanto
32:   Atualize  $q\text{-table}$ 
33: end enquanto
34: retorne  $u^*$ 

```

No que se refere à complexidade de tempo¹¹ do Algoritmo VN SPD, seu custo é dado por $O(G_{tmax}k_{max}T^3nkh)$, sendo G_{tmax} o número de gerações executadas até o limite de tempo, k_{max} o número máximo de estruturas de vizinhança, n o número de objetos do conjunto de dados, k o número de grupos, h o número de atributos e T o comprimento de u (número de iterações executadas pelo DPK-means). Uma vez que k e h tendem a ser dominados por n , a complexidade de tempo do VN SPD pode ser dada por $O(G_{tmax}k_{max}T^3n)$.

4.6 Calibração de parâmetros

A aplicação desses três algoritmos, baseados nas metaheurísticas BRKGA, ILS e VNS, considerou a calibração de seus parâmetros através do método Q-Learning [196][37, p.23-24], que consiste em um algoritmo de aprendizado por reforço projetado para resolver problemas de decisão sequencial modelados como Processos de Decisão de Markov (MDP) [162]. Seu objetivo é aprender uma política que maximize a recompensa (ou *reward*) acumulada ao longo do tempo. No contexto do *framework* RKO, o Q-Learning é empregado como um mecanismo de controle adaptativo de parâmetros, permitindo ajustar dinamicamente a configuração de parâmetros associados às metaheurísticas (hiperparâmetros) ao longo da execução.

Para isso, é adotada modelagem que permite tratar o ajuste de parâmetros como um problema de aprendizado, no qual o algoritmo busca identificar automaticamente as configurações mais eficazes ao longo das iterações¹². O conhecimento adquirido pelo agente é armazenado em uma estrutura denominada Q-table. Essa tabela associa a cada par estado-ação (s_q, a_q) um valor $Q(s_q, a_q)$, denominado Q-value, que representa a estimativa da recompensa obtida ao executar a ação a_q no estado s_q . A recompensa é definida com base na evolução da melhor solução encontrada entre etapas (*steps*) consecutivas, de forma a priorizar configurações que contribuam para a melhoria contínua da qualidade das soluções. Em cada etapa (*step*), a Q-table é atualizada com base na recompensa imediata recebida. O valor aumenta quando a ação resulta em uma recompensa positiva e, quando o maior Q-value do estado seguinte supera o valor atual, indica que a ação foi mais valiosa do que previamente estimado. Por outro lado, ele diminui ao receber recompensas negativas, refletindo os resultados indesejáveis dessas ações.

A integração do Q-Learning ao RKO ocorre para as diferentes metaheurísticas se-

¹¹A análise de complexidade de tempo do Algoritmo VN SPD é detalhada no Apêndice F.3, p. 189.

¹²Essas são efetivamente iterações das metaheurísticas, nas quais combinações de parâmetros são escolhidas a partir da Q-table, avaliadas e, com base no desempenho obtido, têm seus valores atualizados, ajustando adaptativamente a probabilidade de seleção dessas configurações nas iterações subsequentes.

guindo três etapas principais quais sejam:

1. Inicialização: é definido o conjunto de estados (configurações de parâmetros) e ações possíveis e a Q-table é inicializada com valores arbitrários;
2. Seleção de parâmetros: em cada etapa (*step*), o agente seleciona uma ação, que determina a configuração de parâmetros a ser utilizada naquela etapa;
3. Aprendizado: a recompensa é calculada e a Q-table é atualizada de acordo.

Esse ciclo se repete ao longo da execução, permitindo que o algoritmo aprenda progressivamente quais configurações são mais eficazes em cada etapa do processo.

Capítulo 5

Experimentos Computacionais

Este capítulo traz o conjunto de resultados e análises obtidos a partir dos experimentos computacionais realizados com os algoritmos BRKGAPD, VNSPD e ILSPD (propostos na Seção 4.5) e os cinco algoritmos relacionados às principais estratégias de alocação disponíveis na literatura, quais sejam: Uniforme [23]; DAS [54]; SSAS [70]; APAS [70] e Escalonada [155]. Todos esses algoritmos – utilizados para obter soluções de alocação de *privacy budget* para o DPK-means –, foram implementados na linguagem Python¹. Os experimentos foram realizados em computador equipado com processador i9-13900F, 128GB de memória RAM e sistema operacional Windows 11. De forma a produzir uma análise mais ampla e robusta, permitindo uma melhor comparação entre os algoritmos propostos e os da literatura, foram utilizados 30 conjuntos de dados, considerando diferentes valores de *privacy budget* total.

As Seções 5.1 e 5.2 descrevem, respectivamente, os conjuntos de dados e as configurações de parâmetros adotadas. A análise dos resultados do experimento computacional realizado foi conduzida em duas partes. A primeira, apresentada na Seção 5.3, foi dedicada a comparar o desempenho dos algoritmos implementados com as cinco estratégias da literatura. A segunda parte foi dedicada a analisar diferentes aspectos dos algoritmos implementados. Na Seção 5.4, foi realizada uma análise dos tempos de execução necessários para que cada um dos algoritmos atingisse as melhores soluções, em termos do NICV, para cada conjunto de dados e em cada cenário. Na Seção 5.5 foi conduzida uma análise complementar com a intenção de avaliar a estabilidade do BRKGAPD, buscando investigar a robustez dos resultados obtidos a partir das alocações produzidas por ele. Finalmente, na Seção 5.6 buscou-se avaliar a contribuição da abordagem alternativa para o cálculo de Δf .

¹Python v3.11.

Adicionalmente, no Apêndice G, as Seções G.1 (p.192) e G.2 (p.195) apresentam, respectivamente, os resultados em termos do espaço de parâmetros considerado e da performance do Q-Learning na configuração dos parâmetros ao longo das execuções realizadas.

5.1 Conjuntos de dados

Para a realização dos experimentos computacionais, foram utilizados 30 conjuntos de dados, entre sintéticos e reais, com boa diversidade no que se refere ao número de objetos, entre 150 e 6000, e de classes (grupos), entre 2 e 16. Esses conjuntos também foram utilizados em outros experimentos recentes no contexto do problema de agrupamento diferencialmente privado [129, 213, 43, 199] ou são conjuntos de dados clássicos da literatura de análise de agrupamentos. Destaca-se que a maioria dos conjuntos de dados selecionados conta com dois atributos, facilitando a visualização da solução de agrupamento obtida. No entanto, também foram utilizados nos experimentos dois conjuntos de dados com quatro atributos e um com 30 atributos.

A Tabela 5.1 relaciona, para cada conjunto de dados, o número de objetos, atributos e classes, bem como a referência da fonte na qual foram disponibilizados e/ou utilizados anteriormente.

Tabela 5.1: Conjuntos de dados utilizados no experimento

Nome	Objetos	Atributos	Classes	Referência
banana	5300	2	2	[213, 8]
iris	150	4	3	[75]
transfusion	748	4	2	[129, 212]
wdbc	569	30	2	[201]
n6000k6	6000	2	6	^a
D1	1500	2	6	[199]
D2	1200	2	6	[199]
D3	1400	2	4	[199]
D4	2400	2	4	[199]
D5	350	2	5	[199]
D6	1100	2	5	[199]
D7	1500	2	6	[199]
D8	2000	2	6	[199]
D9	1000	2	3	[199]
D10	1250	2	5	[199]
D11	500	2	5	[199]
D12	1000	2	4	[199]
D13	1000	2	2	[199]
D14	1000	2	4	[199]
D15	1000	2	4	[199]
D16	425	2	6	[199]
D17	1750	2	5	[199]
D18	2250	2	2	[199]
R1	450	2	9	[199]
R2	1750	2	7	[199]
R3	1600	2	16	[199]
R4	1250	2	5	[199]
R5	1200	2	6	[199]
R6	1500	2	6	[199]
R7	1200	2	3	[199]

^a Gerado pelo autor.

5.2 Parâmetros

No que se refere aos parâmetros de PD, para os valores de *privacy budget* total, a ser distribuído entre as T iterações do DPk-means, seguindo o que está preconizado na literatura² [29, p.6], o seguinte conjunto de possíveis valores foi considerado: $\epsilon = \{10, 5, 3, 2, 1, 0,5\}$. O ruído adicionado na atualização dos centroides foi gerado utilizando os mecanismos de Laplace e Normal, descritos na Seção 2.4.3. A perturbação foi introduzida nos cálculos da soma dos atributos (S_j) e da frequência ($|C_j|$) dos objetos em cada grupo C_j . Em função disso, o *privacy budget* ϵ_t alocado a cada iteração t do k-means foi dividido entre as duas consultas realizadas para calcular S_j e $|C_j|$.

Em relação à sensibilidade global Δf (Definição 6, p.24) das consultas realizadas, duas abordagens foram consideradas. A primeira, mais recorrente na literatura [23, p.5][216, p.490], é dada simplesmente por $\Delta f^S + \Delta f^C$. Essa abordagem foi utilizada para determinar a sensibilidade global na aplicação de ambos os mecanismos, Laplace e Normal. Uma abordagem alternativa foi utilizada para o cálculo da sensibilidade global quando utilizado o mecanismo de Laplace, que foi identificado por Laplace* quando essa abordagem foi adotada. As especificidades referentes à aplicação de ambos os mecanismos na atualização dos centroides do DPk-means, bem como o detalhamento das abordagens adotadas para o cálculo da sensibilidade são apresentados no Apêndice B.6 (p.165).

Na execução do DPk-means, além dos parâmetros de PD, foram adotados os valores de k indicados na coluna Classes da Tabela 5.1, para cada conjunto de dados utilizado no experimento. A inicialização dos centroides foi realizada adotando a abordagem proposta no k-means++ [12]. Mais especificamente, com o objetivo de permitir uma comparação tão equânime quanto possível, o comparativo entre as alocações de *privacy budget* obtidas através das diferentes estratégias considerou os mesmos centroides iniciais, não apenas o mesmo método de inicialização.

Os três algoritmos propostos nesta tese foram implementados tendo, por base, o *framework* RKO [37]. No que diz respeito à configuração dos seus parâmetros, foi adotado o método Q-Learning [37, p.23-24], descrito na Seção 4.6. Os conjuntos de possíveis valores dos parâmetros (utilizados no Q-Learning) foram definidos com base nos intervalos de

²Embora a definição do parâmetro ϵ seja dependente do contexto, de maneira geral $0,01 \leq \epsilon \leq 1$ é considerado um intervalo adequado para dados sensíveis ou quando se deseja priorizar a proteção da privacidade, ao passo que $1 < \epsilon \leq 10$ é adotado em contextos menos sensíveis, quando se deseja priorizar a utilidade. A título de referência, para $\epsilon = 5$, a razão de chances dada pela Eq. 2.10 é superior a 140, indicando que o adversário pode se sentir extremamente confiante ao tentar distinguir entre dois conjuntos de dados vizinhos [44, p.33].

referência apresentados em [37, p.38] e [132, p.9], bem como nos resultados dos experimentos preliminares realizados durante a pesquisa. Os valores apresentados nas Tabelas 5.2, 5.3 e 5.4 são referentes, respectivamente, aos parâmetros dos algoritmos BRKGAPD, ILSPD e VNSPD propostos.

Tabela 5.2: Espaço de parâmetros do algoritmo BRKGAPD

Parâmetro	Valores
tamanho da população p	70, 100, 300, 500
tamanho do conjunto elite p_e (proporção de p)	0,1, 0,2, 0,3
probabilidade de cruzamento ρ	0,7, 0,8, 0,9

Tabela 5.3: Espaço de parâmetros do algoritmo ILSPD

Parâmetro	Valores
$\beta_{\min}$	0,010, 0,025, 0,050, 0,075, 0,100
$\beta_{\max}$	0,15, 0,20, 0,25, 0,30

Tabela 5.4: Espaço de parâmetros do algoritmo VNSPD

Parâmetro	Valores
$k_{\max}$	3, 4, 5, 6
$\beta_{\min}$	0,010, 0,025, 0,050, 0,075, 0,100, 0,200

5.3 Comparativo entre as estratégias da literatura e os algoritmos propostos

A fim de avaliar a efetividade dos algoritmos propostos para alocação de *privacy budget*, foi realizado um comparativo com as principais estratégias identificadas na literatura. São elas:

- Uniforme [23]
- DAS (*Dichotomy Allocation Strategy*) [54]
- SSAS (*Series Sum Allocation Strategy*) [70]
- APAS (*Arithmetic Progression Allocation Strategy*) [70]
- Escalonada [155]

Como mencionado na Seção 4.1, a qualidade de cada solução de alocação de *privacy budget* foi medida através do NICV resultante da aplicação do algoritmo DPK-means considerando a respectiva alocação.

Com o objetivo de oferecer uma comparação mais equitativa entre as abordagens, buscou-se compatibilizar o número de soluções avaliadas em cada caso. Nos algoritmos da literatura, cada execução produz uma única solução final, sendo realizadas 10 execuções independentes por instância, o que resulta em 10 réplicas para cada conjunto de dados, em cada cenário, sendo os cenários definidos pela combinação dos mecanismos e valores de *privacy budget*. Por outro lado, no caso do BRKGAPD, a cada execução, múltiplas soluções são geradas ao longo das gerações, sendo ao final selecionada aquela de melhor *fitness* (menor NICV) dentre as p soluções da população.

Para tornar a comparação mais alinhada com esse processo e com o número de réplicas dos demais algoritmos, cada estratégia da literatura foi executada 100 vezes para cada conjunto de dados e em cada cenário. Em seguida, foram selecionadas as 10 soluções com menor NICV, de modo a compor um conjunto de 10 réplicas comparáveis às produzidas pelos três algoritmos propostos.

Além disso, o mesmo centroide inicial³ foi adotado em cada condição comparada, ou seja, para um determinado conjunto de dados e valor de ϵ , as execuções do DPK-means com cada um dos algoritmos (propostos e baseadas nas estratégias da literatura) tiveram, como ponto de partida, os mesmos centroides iniciais. Os resultados dos valores medianos de NICV obtidos segundo cada um dos algoritmos propostos são apresentados no Anexo G (Tabela G.9, p.205).

Preliminarmente, foi realizada uma análise de consistência a fim de verificar se os três algoritmos propostos e as cinco estratégias obtiveram soluções para todos os conjuntos de dados e em todos os 18 cenários considerados, i.e., todas as combinações de mecanismos e valores de *privacy budget*.

Todos os algoritmos foram executados com sucesso e, no que diz respeito às estratégias, a APAS foi a única a não ser executada em todos os conjuntos de dados⁴ devido à restrição

³No que se refere à aplicação do DPK-means, a literatura evidencia que a estratégia adotada para a inicialização dos centroides tem impacto direto na obtenção de uma solução final de boa utilidade. Dessa forma, a estratégia de inicialização aleatória, adotada no algoritmo de Lloyd [131], foi substituída pela estratégia adotada no algoritmo k-means++ [12], indicada como bem sucedida em comparativo disponível na literatura [121].

⁴A Tabela G.7 (Apêndice G, p. 203) apresenta a frequência de conjuntos de dados com 10 repetições concluídas em cada cenário e a Tabela G.8 (Apêndice G, p. 204) lista os casos em que a estratégia APAS foi ou não executada com sucesso.

de ϵ mínimo, conforme descrito na Seção 3.1.1. Mais especificamente, essa estratégia estabelece um valor mínimo de *privacy budget*, ϵ_m (Eq. D.6, p. 175), a ser alocado a cada iteração. Porém, uma vez que o número de iterações T foi fixado previamente para todas as estratégias, a depender do ϵ total, o quociente $\frac{\epsilon}{T}$ pode ser inferior a ϵ_m , inviabilizando o uso dessa estratégia em alguns casos.

A fim de avaliar o desempenho da estratégia APAS, primeiramente foi conduzida uma análise parcial, considerando apenas os conjuntos de dados e valores de ϵ para os quais essa estratégia foi executada com sucesso. A comparação dos desempenhos dos algoritmos e das estratégias foi segmentada por cenários e baseada no *gap* percentual. Assim, foi calculado, em cada cenário, o *gap* percentual entre a melhor solução, i.e., o menor valor de NICV, denotado por $\text{NICV}_{\min}$, obtido a partir da aplicação dos diferentes algoritmos e estratégias avaliados, e aquele obtido através da solução de alocação de *privacy budget* gerada utilizando cada um dos algoritmos e estratégias naquele cenário, denotado por $\text{NICV}_{\text{solução}}$, conforme descrito na Eq. 5.1.

$$\text{gap}_{\text{solução}} = \frac{\text{NICV}_{\text{solução}} - \text{NICV}_{\min}}{\text{NICV}_{\min}} \times 100 \quad (5.1)$$

A Tabela 5.5 apresenta a mediana do *gap* percentual em relação à melhor estratégia por cenário. Para cada combinação de conjunto de dados, valor de ϵ e réplica, o *gap* percentual foi calculado em relação ao menor valor de NICV observado entre todas as estratégias. No entanto, como a estratégia APAS não foi executada com sucesso em todos os conjuntos de dados, a inclusão de todos os resultados das demais estratégias poderia introduzir viés na comparação, favorecendo artificialmente essa estratégia. Para mitigar esse efeito, a análise foi limitada apenas aos conjuntos de dados para as quais foi possível obter resultados para a APAS, garantindo assim comparabilidade consistente entre as estratégias. O valor sublinhado indica o algoritmo ou estratégia que apresentou o menor *gap* percentual mediano no respectivo cenário e o sublinhado duplo indica o segundo colocado.

É possível verificar que o algoritmo BRKGAPD apresentou o menor *gap* mediano para todos os cenários considerados. No geral, os *gaps* percentuais medianos dos três algoritmos propostos foi menor do que todas as estratégias da literatura, exceto nos cenários (Laplace*, $\epsilon = 0,5$), onde a estratégia Escalonada superou os algoritmos ILSPD e VNSPD, e (Normal, $\epsilon = 0,5$), onde a estratégia Escalonada superou o algoritmo ILSPD. Dentre as estratégias da literatura, a estratégia Escalonada teve o melhor desempenho, apresentando o menor *gap* percentual mediano em 16 (89%) dos cenários, seguida pela

Tabela 5.5: Mediana do gap percentual em relação à melhor estratégia por cenário, incluindo APAS.

cenários		conjuntos de dados	BRKGAPD	ILSPD	VNSPD	APAS	DAS	Escalonada	SSAS	Uniforme
mecanismo	ϵ									
Laplace	0,5	2	<u>0,06</u>	0,95	<u>0,70</u>	1,91	3,14	1,22	70,46	1,95
Laplace	1	4	<u>0,04</u>	0,72	<u>0,61</u>	1,70	7,69	0,79	72,20	1,25
Laplace	2	8	<u>0,07</u>	<u>1,13</u>	<u>1,53</u>	2,40	22,50	3,13	257,11	4,89
Laplace	3	14	<u>0,24</u>	<u>2,26</u>	<u>2,03</u>	3,32	28,03	2,86	319,31	6,06
Laplace	5	23	<u>0,20</u>	<u>1,19</u>	<u>1,22</u>	3,97	17,13	2,09	339,92	4,16
Laplace	10	24	<u>0,03</u>	<u>0,23</u>	0,24	3,61	4,24	0,53	154,07	1,01
Laplace*	0,5	2	<u>0,02</u>	<u>0,41</u>	0,55	0,78	4,31	<u>0,33</u>	55,98	0,86
Laplace*	1	4	<u>0,01</u>	0,29	<u>0,22</u>	0,51	2,29	<u>0,31</u>	77,31	0,58
Laplace*	2	8	<u>0,02</u>	<u>0,41</u>	<u>0,56</u>	1,08	6,99	0,73	175,80	1,55
Laplace*	3	14	<u>0,07</u>	<u>0,63</u>	<u>0,56</u>	1,08	8,70	1,09	161,70	2,02
Laplace*	5	23	<u>0,06</u>	0,40	<u>0,39</u>	1,41	6,10	0,74	195,80	1,72
Laplace*	10	24	<u>0,01</u>	<u>0,08</u>	<u>0,09</u>	1,32	1,45	0,20	58,03	0,37
Normal	0,5	2	<u>0,04</u>	<u>1,29</u>	<u>0,71</u>	1,40	2,56	0,80	64,97	1,40
Normal	1	4	<u>0,03</u>	0,79	<u>0,53</u>	1,38	5,03	0,88	72,34	1,45
Normal	2	8	<u>0,05</u>	1,12	<u>0,88</u>	1,63	16,63	1,95	231,64	4,25
Normal	3	14	<u>0,15</u>	1,23	<u>1,07</u>	2,30	17,20	2,39	272,23	3,81
Normal	5	23	<u>0,13</u>	0,76	<u>0,73</u>	2,57	12,28	1,54	309,23	3,03
Normal	10	24	<u>0,02</u>	0,17	<u>0,14</u>	2,60	2,75	0,38	102,53	0,79

estratégia APAS, que apresentou o menor *gap* mediano nos outros 3 (17%) cenários, quais sejam (Laplace, $\epsilon = 2$), (Laplace*, $\epsilon = 3$) e (Normal, $\epsilon = 2$), e o segundo menor em 12 (67%) cenários.

Embora a estratégia APAS tenha apresentado desempenho competitivo na comparação com as demais estratégias, o seu *gap* percentual mediano foi o menor dentre os das estratégias da literatura em apenas 3 cenários e, mesmo nesses casos, os algoritmos BRKGAPD, VNSPD e ILSPD apresentaram *gap* percentual mediano substancialmente menor. Com isso, a análise foi realizada com todos os conjuntos de dados, excluindo a estratégia APAS, e cujos resultados se encontram na Tabela 5.6. Novamente o algoritmo BRKGAPD apresentou o menor *gap* mediano para todos os cenários considerados. No geral, o *gap* percentual mediano dos algoritmos foi menor do que todas as estratégias da literatura, exceto no cenário (Laplace*, $\epsilon = 1$), onde a estratégia Escalonada superou ligeiramente os algoritmos ILSPD e VNSPD, enquanto, nos demais cenários, os algoritmos a superaram geralmente com folga. No que diz respeito apenas às estratégias da literatura, a estratégia Escalonada teve novamente o melhor desempenho absoluto, apresentando o menor *gap* percentual mediano em todos os cenários, seguida pela estratégia Uniforme, que apresentou o segundo menor *gap* mediano, também em todos os cenários.

A Figura 5.1⁵ compara a distribuição do *gap* percentual de 300 soluções (10 réplicas

⁵No que se refere à construção dos *boxplots*, foram rotuladas como *outliers* observações que ultrapas-

Tabela 5.6: Mediana do gap percentual em relação à melhor estratégia por cenário, excluindo APAS.

cenários		BRKGAPD	ILSPD	VNSPD	DAS	Escalonada	SSAS	Uniforme
mecanismo	ϵ							
Laplace	0,5	10,94	148,09	161,54	394,61	187,36	665,31	266,02
Laplace	1	3,23	81,99	85,00	349,95	99,57	767,57	197,99
Laplace	2	0,89	20,28	21,77	183,62	26,72	686,88	54,19
Laplace	3	0,45	7,13	8,09	85,74	8,56	575,44	18,54
Laplace	5	0,12	1,80	1,73	23,49	2,91	471,56	5,75
Laplace	10	0,02	0,29	0,35	5,65	0,68	197,98	1,40
Laplace*	0,5	5,91	95,51	108,06	345,77	110,77	742,74	182,83
Laplace*	1	1,66	48,11	45,75	222,42	44,95	766,11	94,65
Laplace*	2	0,42	6,68	6,25	69,45	7,24	595,65	13,91
Laplace*	3	0,11	1,82	2,14	24,63	2,87	479,38	5,72
Laplace*	5	0,03	0,51	0,55	8,77	1,02	279,22	2,22
Laplace*	10	0,00	0,12	0,12	1,95	0,24	73,23	0,48
Normal	0,5	7,49	128,73	123,62	397,16	154,44	726,64	231,52
Normal	1	2,95	61,24	56,12	340,98	72,26	776,27	136,95
Normal	2	0,78	12,39	11,66	125,44	13,76	686,89	28,23
Normal	3	0,20	3,50	3,68	52,74	5,46	570,24	11,76
Normal	5	0,05	0,91	1,12	17,52	1,91	410,42	3,91
Normal	10	0,01	0,20	0,21	3,81	0,49	141,59	0,93

em cada um dos 30 conjuntos de dados) obtidas a partir de cada um dos algoritmos e das estratégias da literatura nos seis cenários definidos pela combinação dos três mecanismos com os valores mínimo e máximo de *privacy budget*. Para todos os cenários, observa-se que o BRKGAPD apresentou a menor dispersão, enquanto que a estratégia SSAS apresentou a maior dispersão.

A Figura 5.2 compara a distribuição do *gap* percentual de 900 soluções (10 réplicas aplicando cada um dos 3 mecanismos em cada um dos 30 conjuntos de dados) obtidas a partir de cada um dos algoritmos e das estratégias da literatura nos seis valores de *privacy budget*. Para todos os valores de *privacy budget*, novamente observa-se que o BRKGAPD apresentou a menor dispersão, enquanto que a estratégia SSAS apresentou a maior dispersão. Além disso, é possível notar tendência decrescente do *gap* percentual conforme aumenta o *privacy budget*, como esperado.

saram os limites inferior ou superior de 1,5 vezes a distância interquartilica em relação ao primeiro e ao terceiro quartis, respectivamente.

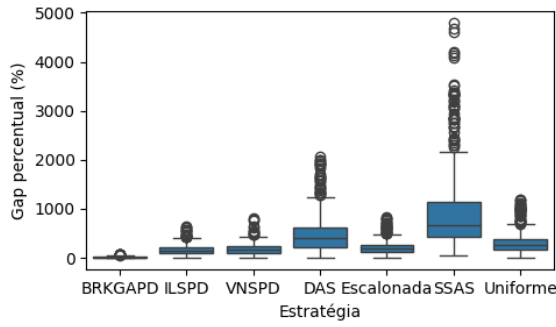
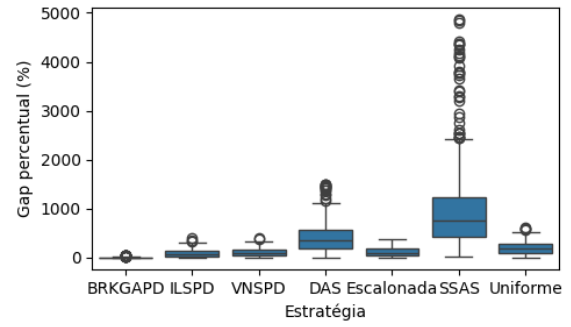
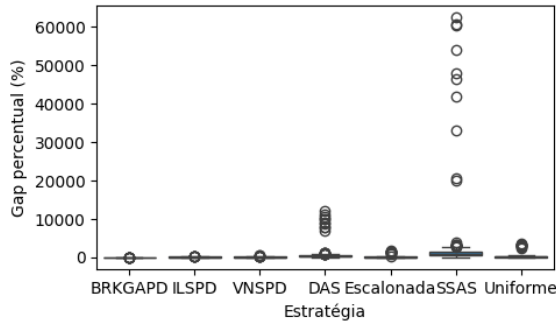
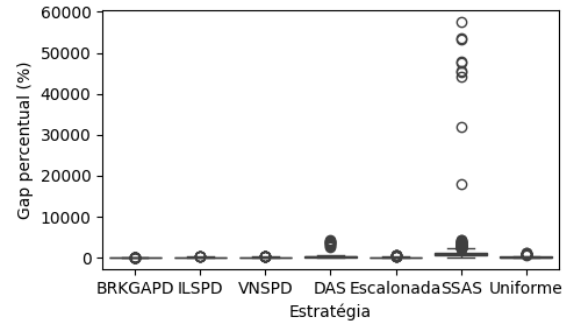
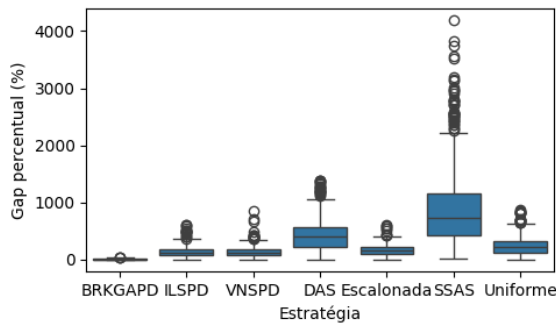
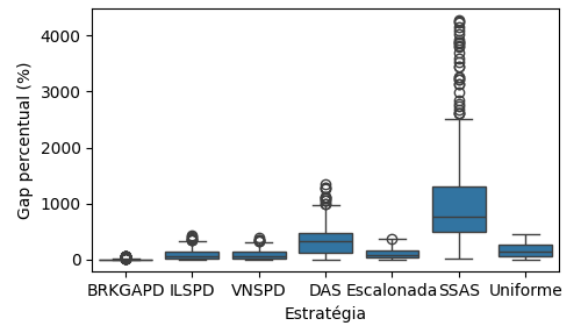
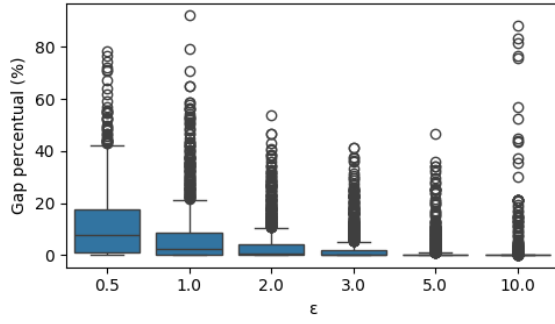
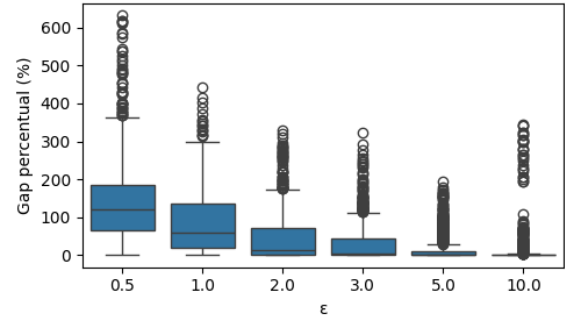
(a) Laplace, $\epsilon = 0,5$ (b) Laplace, $\epsilon = 10$ (c) Laplace*, $\epsilon = 0,5$ (d) Laplace*, $\epsilon = 10$ (e) Normal, $\epsilon = 0,5$ (f) Normal, $\epsilon = 10$

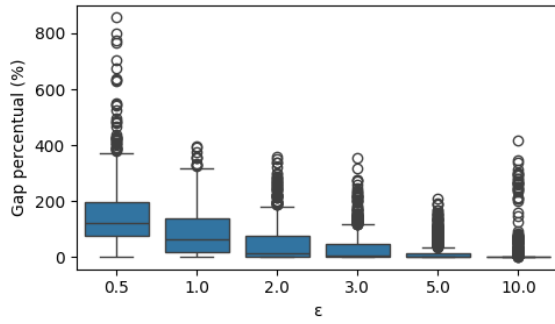
Figura 5.1: Distribuição do gap percentual por estratégias e cenários selecionados.



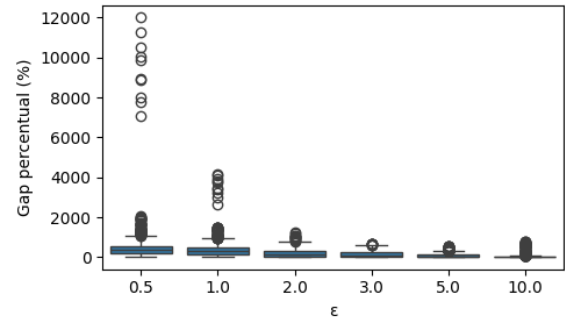
(a) BRKGAPD



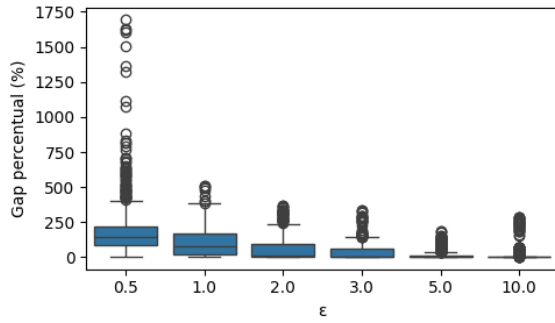
(b) ILSPD



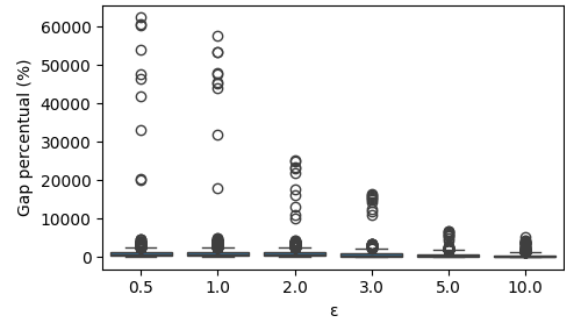
(c) VNSPD



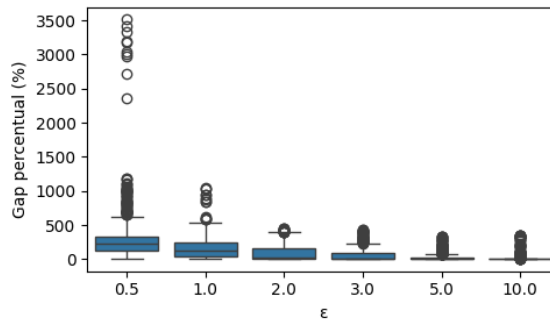
(d) DAS



(e) Escalonada



(f) SSAS



(g) Uniforme

Figura 5.2: Distribuição do gap percentual por valores de *privacy budget*, independente-mente do mecanismo adotado.

A Tabela 5.7 apresenta as estatísticas resumo do *gap* percentual correspondente aos resultados produzidos pelos algoritmos⁶, independentemente do mecanismo utilizado. De forma geral, o BRKGAPD apresentou *gaps* bem inferiores aos outros dois algoritmos.

Tabela 5.7: Gap percentual - estatísticas resumo.

ϵ	algoritmo	mínimo	q1	mediana	q3	máximo	média
0,5	BRKGAPD	0,00	0,95	7,98	17,49	78,48	12,01
0,5	ILSPD	0,03	66,91	122,07	186,79	632,10	139,59
0,5	VNSPD	0,06	76,98	124,21	195,08	856,83	144,81
1	BRKGAPD	0,00	0,24	2,50	8,70	92,05	7,99
1	ILSPD	0,01	19,32	60,89	136,13	443,21	86,67
1	VNSPD	0,01	18,64	65,29	138,32	395,93	87,42
2	BRKGAPD	0,00	0,04	0,62	4,31	53,78	4,44
2	ILSPD	0,00	2,75	12,56	71,82	329,23	45,69
2	VNSPD	0,00	2,77	13,08	75,78	361,16	47,14
3	BRKGAPD	0,00	0,01	0,20	2,15	41,27	3,19
3	ILSPD	0,00	0,99	4,03	45,71	323,22	30,69
3	VNSPD	0,00	1,01	4,24	47,82	353,87	33,31
5	BRKGAPD	0,00	0,00	0,05	0,39	46,46	1,78
5	ILSPD	0,00	0,27	1,01	11,73	193,18	17,69
5	VNSPD	0,00	0,27	1,10	14,79	209,33	18,31
10	BRKGAPD	0,00	0,00	0,01	0,06	88,32	1,63
10	ILSPD	0,00	0,06	0,21	1,45	344,43	14,60
10	VNSPD	0,00	0,06	0,22	1,53	417,88	14,58

Finalmente, a fim de avaliar se as diferenças de desempenho, em termos de NICV, apresentadas pelos métodos considerados são estatisticamente significativas, foi aplicado o teste de Friedman [79, 104], considerando como hipótese nula que todos os algoritmos e/ou estratégias têm desempenho equivalente e, como hipótese alternativa, que ao menos um par de algoritmo e/ou estratégia têm desempenho significativamente diferente.

Uma vez que o objetivo foi comparar algoritmos/estratégias, esses foram definidos como tratamentos. A fim de controlar a variação externa, os blocos foram definidos pela combinação de conjunto de dados, valor de *privacy budget*, mecanismo e repetições. Finalmente, o NICV foi definido como variável resposta, i.e., a métrica de desempenho entre os tratamentos.

O p-valor resultante do teste de Friedman [79, 104] foi aproximadamente zero, sugerindo a rejeição da hipótese nula, i.e., que há evidências de diferença significativa ao menos entre dois dos métodos considerados. A Tabela 5.8 apresenta os algoritmos e estratégias da literatura ordenados segundo rank médio global obtido por cada algoritmo/estratégia. Ressalta-se que, nesse contexto, menores valores de *rank* indicam melhor desempenho relativo, uma vez que correspondem a posições mais altas na ordenação média ao longo dos cenários avaliados.

⁶Resultados referentes a 900 soluções por algoritmo e valor de ϵ , obtidas a partir da combinação de 3 mecanismos, 30 conjuntos de dados e 10 repetições.

Tabela 5.8: Algoritmos propostos e estratégias da literatura ordenados segundo rank médio.

Tratamento	BRKGAPD	ILSPD	VNSPD	Escalonada	Uniforme	DAS	SSAS
Rank	1,00	2,82	2,88	3,52	4,82	5,97	6,98

Em função da rejeição da hipótese nula no teste de Friedman, foi aplicado o teste pareado de Wilcoxon [197], com correção de Holm [105, 7] para comparações múltiplas, a fim de identificar quais pares de algoritmos/estratégias apresentaram diferenças estatisticamente significativas em termos do valor de NICV obtido.

A Figura 5.3 mostra mapas de calor⁷ dos p-valores resultantes da aplicação do teste nos seis cenários definidos pela combinação dos mecanismos com valores mínimo e máximo de *privacy budget*. As tabelas com os resultados numéricos se encontram no Apêndice G (Tabelas G.10 a G.12, p. 220-221). De maneira geral, a hipótese nula não foi rejeitada apenas na comparação entre os algoritmos ILSPD e VNSPD. Dos seis cenários apresentados nesta análise, apenas no cenário (Laplace*, $\epsilon = 0,5$) a hipótese nula foi rejeitada para todos os pares de algoritmos/estratégias, inclusive para ILSPD e VNSPD, como mostra a Figura 5.3c. No entanto, os testes pareados de Wilcoxon rejeitaram a hipótese nula em todos os pares que incluíram o algoritmo BRKGAPD, sugerindo que as diferenças observadas no *gap* percentual em relação aos demais algoritmos foram estatisticamente significativa.

Em resumo, os resultados apresentados indicam que os algoritmos apresentam desempenho superior às estratégias da literatura, tanto em termos descritivos quanto estatísticos. Em particular, o algoritmo BRKGAPD se destacou com *gap* percentual mediano inferior em todos os cenários analisados. Os resultados dos testes de hipótese realizados suportam essa conclusão, uma vez que o teste de Friedman apontou diferenças significativas entre os algoritmos/estratégias considerados e os testes pareados de Wilcoxon rejeitaram a hipótese nula em todos os pares que incluíram o algoritmo BRKGAPD. Por fim, embora algumas estratégias da literatura, especialmente a estratégia Escalonada, tenham apresentado desempenho competitivo em cenários específicos, os três algoritmos mantiveram vantagem consistente na ampla maioria dos cenários e conjuntos de dados. Tais resultados sugerem que a abordagem proposta constitui uma alternativa eficaz para a alocação de *privacy budget* no DPK-means, oferecendo soluções de agrupamento com menor NICV em comparação com as estratégias da literatura.

⁷A escala de cores adotada contempla apenas o intervalo $[0, 0,05]$, limitado superiormente pelo nível de significância adotado nos testes de hipóteses. As células na cor cinza se referem a resultados fora desse intervalo, i.e, acima de 0,05, indicando que a hipótese nula não foi rejeitada.

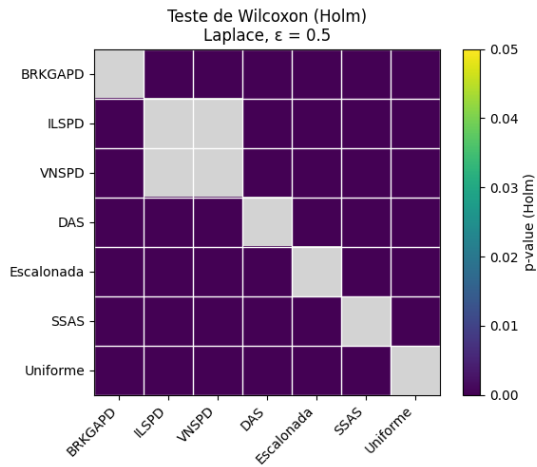
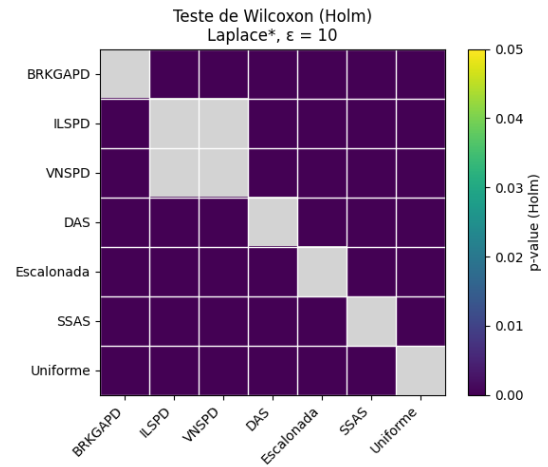
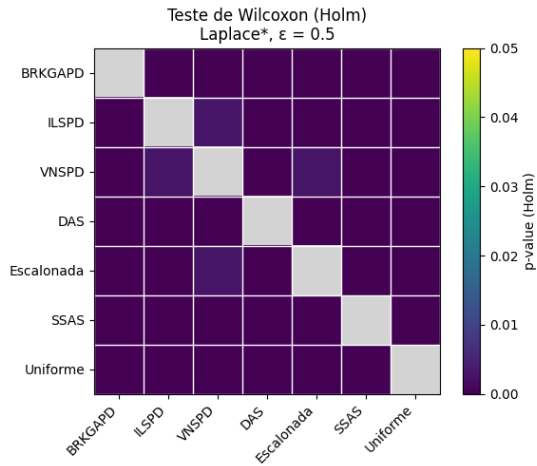
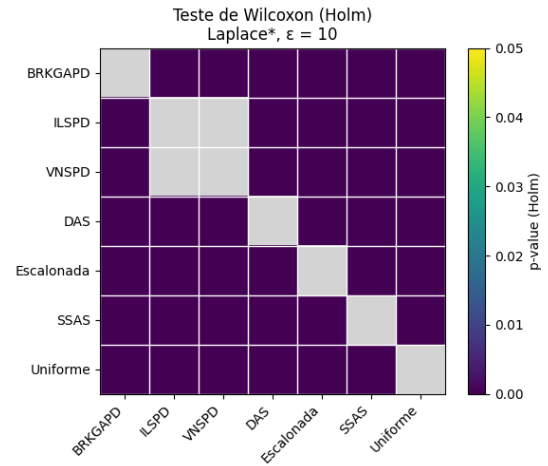
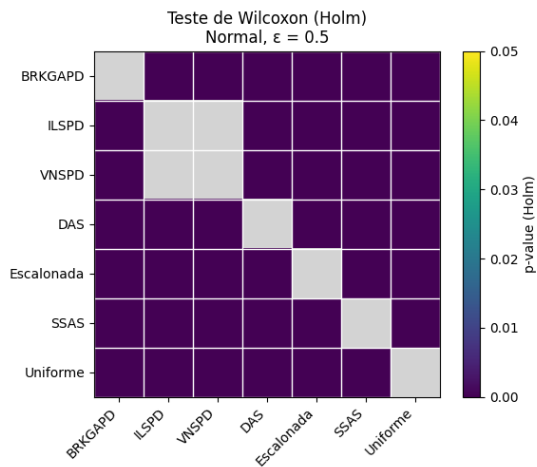
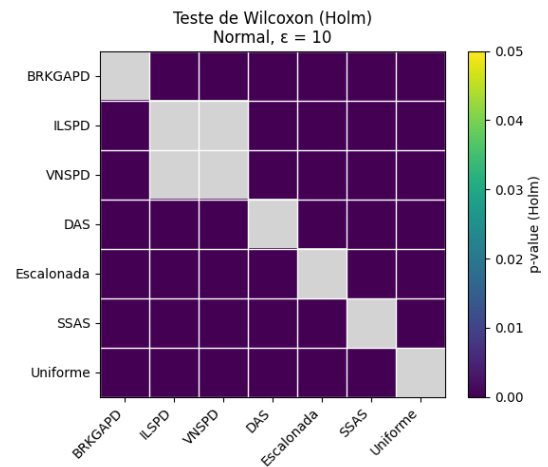
(a) Laplace, $\epsilon = 0,5$ (b) Laplace, $\epsilon = 10$ (c) Laplace*, $\epsilon = 0,5$ (d) Laplace*, $\epsilon = 10$ (e) Normal, $\epsilon = 0,5$ (f) Normal, $\epsilon = 10$

Figura 5.3: Resultados do teste pareado de Wilcoxon.

Por fim, é importante ressaltar que uma alocação mais eficiente de *privacy budget* permite mitigar o *trade-off* entre privacidade e utilidade, ao minimizar a degradação da qualidade dos agrupamentos sem comprometer a garantia oferecida pela privacidade diferencial. Em particular, os resultados indicam que os algoritmos são capazes de distribuir o ruído de forma mais adequada ao longo das iterações do DPk-means, preservando informação relevante para a formação dos grupos e, ao mesmo tempo, mantendo níveis de proteção compatíveis com os requisitos de PD.

Isso é especialmente relevante em aplicações em que conclusões ou decisões operacionais dependem diretamente da qualidade dos agrupamentos, como na segmentação de perfis de risco em serviços financeiros, na estratificação de pacientes em saúde pública ou na identificação de padrões de mobilidade em planejamento urbano, nas quais distorções excessivas podem comprometer a validade das conclusões e das políticas derivadas desses modelos. No contexto dos INE, uma alocação mais eficiente de *privacy budget* pode viabilizar análises mais informativas e desagregadas envolvendo, por exemplo, tipologias de municípios, agrupamentos socioeconômicos ou perfis de empresas, mantendo ao mesmo tempo níveis de proteção compatíveis com marcos legais e diretrizes internas de confidencialidade. Dessa forma, os resultados desta pesquisa contribuem para o avanço de métodos de agrupamento diferencialmente privados que conciliam, de maneira mais equilibrada, os dois objetivos conflitantes inerentes à análise de dados sensíveis, quais sejam, maximizar utilidade e privacidade simultaneamente.

5.4 Tempos de execução

Nesta seção foi realizada uma análise dos tempos de execução⁸ necessários para que cada um dos algoritmos atingisse as melhores soluções, em termos do NICV, para cada conjunto de dados e em cada cenário, considerando as 10 repetições⁹. A Figura 5.4 apresenta a distribuição do tempo de execução até a melhor solução ser produzida, em segundos, nos seis cenários definidos pela combinação dos três mecanismos com os valores mínimo e máximo de *privacy budget*. Verificou-se que os algoritmos apresentaram distribuições similares, com certa desvantagem para o BRKGAPD, que apresentou valores mais elevados para os três quartis nos seis cenários.

⁸Os algoritmos tiveram, como critério de parada, o tempo total de 540 segundos.

⁹Em cada cenário, cada um dos algoritmos produziu 10 soluções para cada um dos 30 conjuntos de dados. Portanto, os resultados compreenderam 300 registros de tempo para cada algoritmo em cada um dos cenários considerados.

Esse resultado está relacionado à natureza populacional do BRKGAPD. Diferentemente do ILSPD e do VNSPD, que operam sobre uma solução, explorando sua vizinhança, o BRKGAPD mantém e avalia uma população de soluções a cada iteração. Isso implica um custo computacional maior por iteração, sobretudo devido à necessidade de decodificar e avaliar múltiplos indivíduos, além das etapas adicionais de ordenação, seleção e recombinação. Como o cálculo do *fitness* é o principal gargalo computacional, esse custo se acumula.

Independentemente dos cenários, os algoritmos apresentaram tempo de execução mediano entre 100 e 150 segundos. As diferenças relativamente pequenas observadas entre os tempos de execução dos três algoritmos estão provavelmente associadas ao uso do RVND (Seção 4.2.3, p. 71). Enquanto os algoritmos ILSPD e VNSPD incorporam explicitamente o RVND como mecanismo de intensificação, a implementação do BRKGAPD utilizada neste experimento não contemplou esse operador. O RVND é capaz de promover melhorias nas soluções ao explorar diferentes estruturas de vizinhança, mas pode ser razoavelmente custoso, contribuindo para aproximar os tempos dos algoritmos baseados nas metaheurísticas ILS e VNS ao do algoritmo baseado na metaheurística BRKGA.

A dominância do BRKGAPD em produzir soluções de menor NICV em relação aos demais algoritmos, conforme apresentado na seção anterior, e a diferença não significativa entre os tempos até a melhor solução dos algoritmos, sugere que os operadores do BRKGAPD, foram, em alguma medida, capazes de compensar a ausência de busca local. De toda forma, esse resultado indica uma direção promissora para trabalhos futuros, com vistas a investigar a contribuição do RVND no BRKGAPD e avaliar se os ganhos em qualidade compensam um provável aumento no tempo de execução associado à busca local nos três algoritmos considerados.

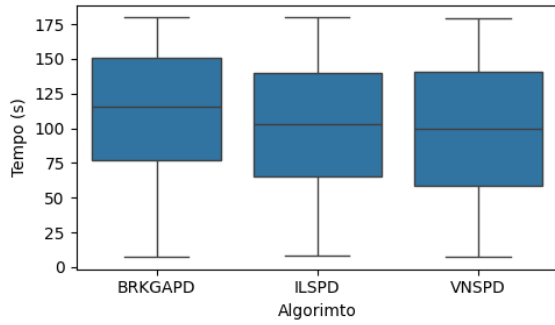
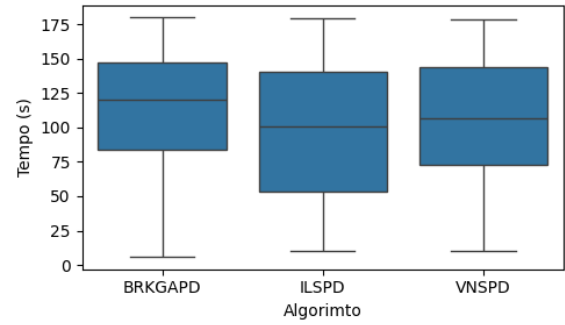
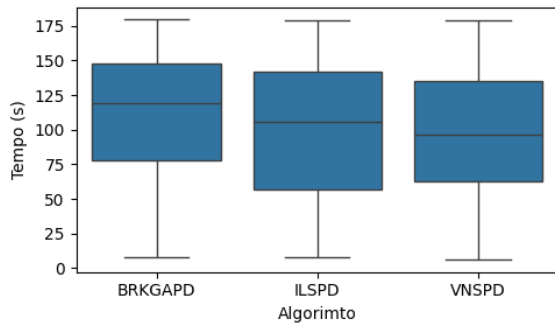
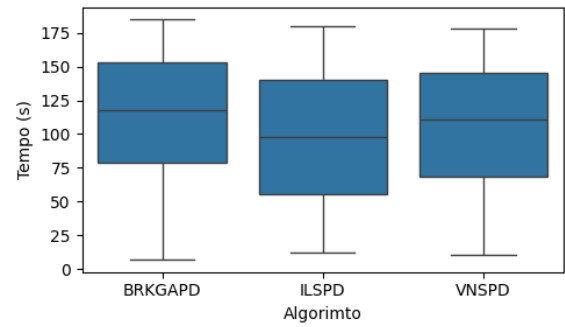
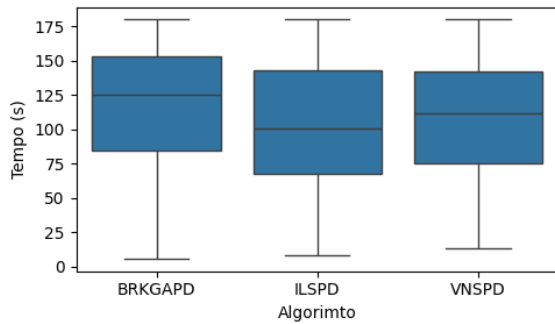
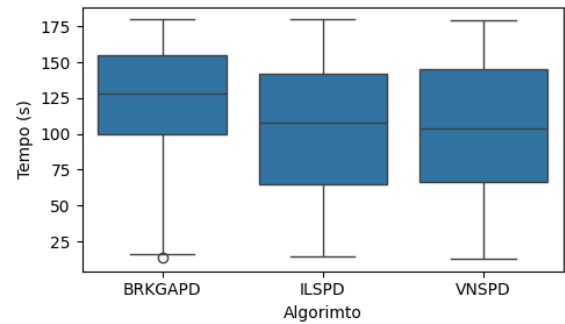
(a) Laplace, $\epsilon = 0,5$ (b) Laplace, $\epsilon = 10$ (c) Laplace*, $\epsilon = 0,5$ (d) Laplace*, $\epsilon = 10$ (e) Normal, $\epsilon = 0,5$ (f) Normal, $\epsilon = 10$

Figura 5.4: Distribuição do tempo de execução até a melhor solução, considerando, para cada cenário e algoritmo, os 300 tempos registrados na execução de 10 repetições em cada um dos 30 conjuntos de dados.

5.5 Avaliação de estabilidade do BRKGAPD

Considerando os resultados obtidos, observou-se que o algoritmo BRKGAPD apresentou desempenho superior em relação aos demais algoritmos, dominando os resultados em termos do NICV da solução de agrupamento resultante da aplicação do DPk-means. Diante desse cenário, foi conduzida uma análise complementar com a intenção de avaliar a estabilidade do BRKGAPD, buscando investigar a robustez dos resultados obtidos a partir das alocações produzidas por ele.

A análise de estabilidade é particularmente relevante no contexto de metaheurísticas, uma vez que o comportamento estocástico no processo de busca tende a produzir soluções distintas em execuções independentes de um mesmo algoritmo, mesmo quando mantidos fixos os parâmetros do problema e os parâmetros do algoritmo. Assim, além da qualidade das soluções obtidas, é relevante avaliar o grau de variabilidade associado aos resultados produzidos, permitindo analisar experimentalmente a confiabilidade prática do algoritmo. Nesse sentido, algoritmos capazes de produzir consistentemente soluções de boa qualidade tendem a apresentar maior robustez e previsibilidade, características desejáveis em qualquer aplicação, especialmente aquelas envolvendo privacidade diferencial e análise de dados sensíveis.

Para a realização dessa avaliação, foram considerados o mecanismo de Laplace e os seis valores de *privacy budget* utilizados ao longo dos experimentos computacionais. Adicionalmente, foi selecionado (dentre os 30) um subconjunto composto por três conjuntos de dados, definidos em função do número de objetos, correspondendo, respectivamente, às instâncias de menor porte, porte intermediário e maior porte. Para cada combinação entre instância e valor de ϵ , foram feitas 50 execuções independentes do BRKGAPD.

Com base nos resultados obtidos, foram calculadas estatísticas descritivas e o coeficiente de variação (CV) dos valores de NICV associados às soluções produzidas. O CV permite analisar a dispersão relativa dos resultados em torno da média, fornecendo uma medida padronizada de estabilidade do algoritmo. Nesse contexto, menores valores de CV indicam maior estabilidade e robustez do algoritmo.

A Tabela 5.9 apresenta as estatísticas resumo dos resultados obtidos para cada par de conjuntos de dados e valores de ϵ . De maneira geral, os resultados indicam comportamento bastante estável do BRKGAPD, sobretudo para valores mais elevados de ϵ , nos quais a intensidade do ruído introduzido pelos mecanismos de privacidade diferencial é reduzida. Observa-se que o conjunto n6000k6 apresentou os menores coeficientes de variação em

Tabela 5.9: Resultados do experimento de estabilidade do BRKGAPD.

conjunto de dados	epsilon	média	mediana	CV (%)
D4	0,5	0,050	0,050	3,946
D4	1	0,046	0,046	1,090
D4	2	0,044	0,044	0,130
D4	3	0,044	0,044	0,038
D4	5	0,044	0,044	0,010
D4	10	0,044	0,044	0,002
iris	0,5	0,406	0,399	10,934
iris	1	0,326	0,325	4,430
iris	2	0,301	0,307	5,487
iris	3	0,276	0,289	10,712
iris	5	0,236	0,231	7,830
iris	10	0,178	0,178	1,325
n6000k6	0,5	0,023	0,023	1,509
n6000k6	1	0,022	0,022	0,117
n6000k6	2	0,022	0,022	0,016
n6000k6	3	0,022	0,022	0,008
n6000k6	5	0,022	0,022	0,003
n6000k6	10	0,022	0,022	0,000

praticamente todos os cenários analisados, atingindo valores próximos de zero para $\epsilon \geq 2$, o que evidencia elevada robustez e forte consistência entre as execuções do algoritmo. O conjunto D4 também apresentou comportamento estável, com redução progressiva do CV à medida que o *privacy budget* aumenta, alcançando variabilidade praticamente nula para $\epsilon = 10$. Por outro lado, o conjunto iris apresentou maior dispersão relativa dos resultados, especialmente para valores menores de ϵ , indicando maior sensibilidade à aleatoriedade introduzida pelo mecanismo de privacidade. Esses resultados indicam que a robustez do BRKGAPD depende não apenas do valor de *privacy budget* adotado, mas também das características do conjunto de dados, como balanceamento (distribuição dos objetos), separabilidade e coesão dos grupos, por exemplo.

A fim de ilustrar esses aspectos, a Figura 5.5 apresenta os diagramas de dispersão dos resultados da execução de 10 iterações do k-means clássico (sem PD) e do DPk-means nos conjuntos de dados iris¹⁰, D4 e n6000k6. O conjunto iris apresenta grupos não esféricos e evidentemente menos coesos e separados que os demais. Em particular, dois dos grupos apresentam sobreposição considerável. Essas características fazem com que mesmo uma pequena perturbação dos centroides ($\epsilon = 5$) altere a alocação dos objetos aos grupos de maneira significativa em cada iteração.

¹⁰O conjunto de dados iris conta com 4 atributos. A fim de permitir a visualização bidimensional, a Figura G.6 apresenta o diagrama referente à combinação dos dois primeiros atributos. A Figura G.6 (Seção G.3 do Apêndice G, p.198) apresenta todos os diagramas referentes às combinações dos atributos 2 a 2.

O conjunto D4, por sua vez, apresenta grupos esféricos, coesos e bem separados. Entretanto, eles são desbalanceados, sendo os dois grupos menores compostos por 200 objetos e os dois maiores por 1000 objetos. A proximidade e o desbalanceamento dos grupos fez com que os centroides perturbados dos grupos menores fossem atraídos na direção dos maiores. Isso se justifica pela função objetivo, que minimiza o somatório das distâncias dos objetos aos centroides e pelo reduzido número de iterações ($T = 10$). No conjunto n6000k6, embora alguns grupos apresentem maior dispersão em comparação com o conjunto D4 e até uma sutil sobreposição, todos os grupos são formados por 1000 objetos. Nesse caso, em função do baixo ruído adicionado ($\epsilon = 5$), os centroides perturbados convergiram sempre para próximo dos centros dos grupos, mesmo com poucas iterações ($T = 10$).

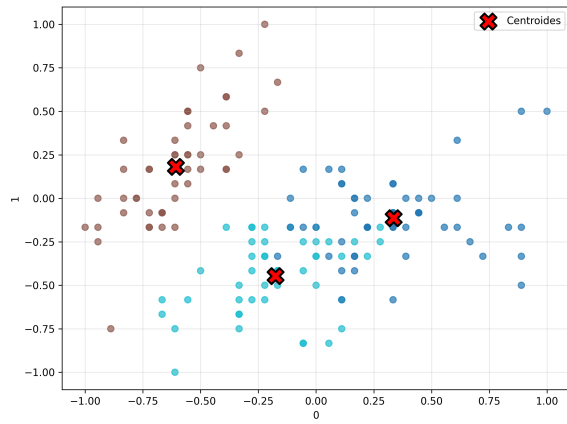
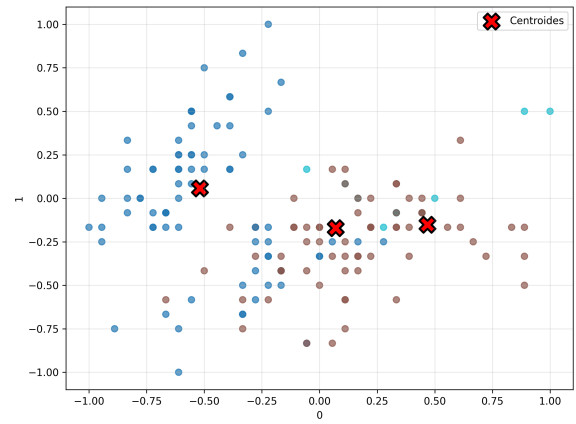
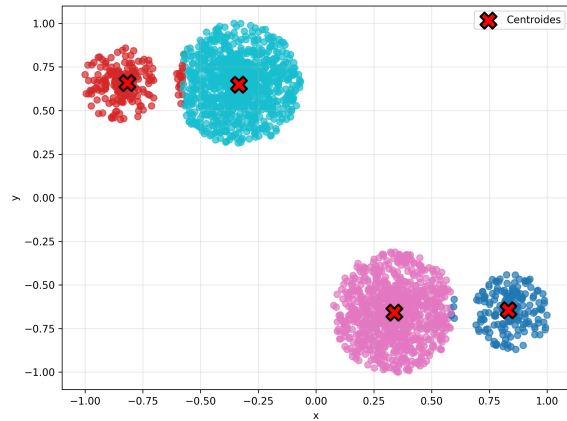
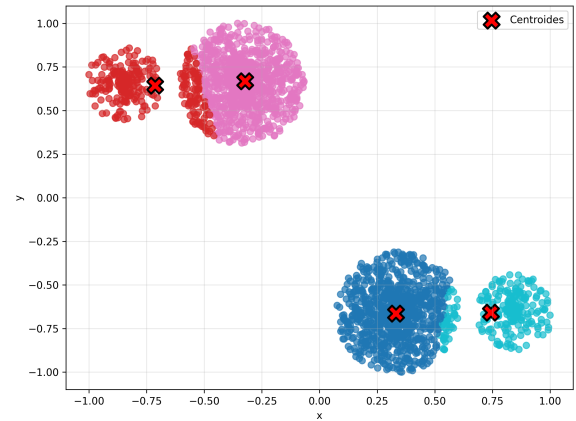
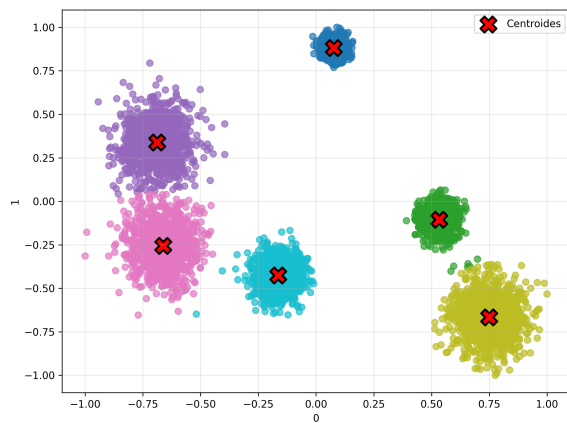
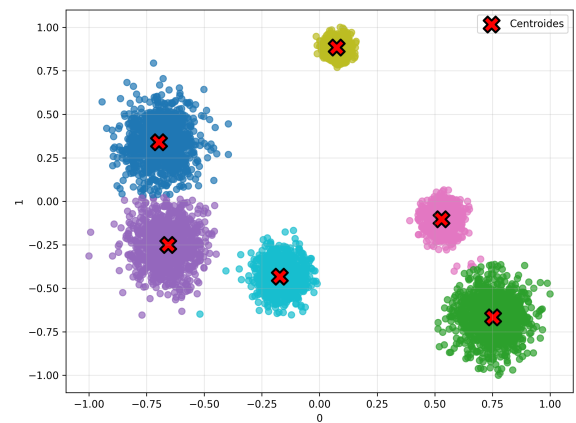
(a) iris ($k = 3$), sem PD(b) iris ($k = 3$), Laplace e $\epsilon = 5$ (c) D4 ($k = 4$), sem PD(d) D4 ($k = 4$), Laplace e $\epsilon = 5$ (e) n6000k6 ($k = 6$), sem PD(f) n6000k6 ($k = 6$), Laplace e $\epsilon = 5$

Figura 5.5: Diagramas de dispersão do resultado da aplicação do k-Means e do DPk-Means nos conjuntos de dados D4, iris e n6000k6.

5.6 Mecanismos

Como mencionado na Seção 5.2, em função das caudas mais pesadas da distribuição de Laplace, existe uma probabilidade não desprezível de que o ruído gerado segundo ela seja de grande intensidade. Assim, testou-se a hipótese de que uma abordagem alternativa para o cálculo de Δf poderia contribuir para mitigar esse efeito. Mais especificamente, duas abordagens foram consideradas para definir a sensibilidade global do cálculo dos centroides. A primeira, mais recorrente na literatura [23, p.5][216, p.490], é dada simplesmente por $\Delta f = h + 1$. Na segunda abordagem, a sensibilidade das funções referentes às consultas de soma de atributos e frequência dos objetos é calculada separadamente. Para o cálculo da soma dos atributos, apenas a variação máxima h é considerada e, para o cálculo da frequência dos objetos, a presença ou não do objeto, ou seja, 1. Nesta Seção, os resultados são avaliados a fim de elaborar conclusão acerca de tal hipótese.

A Tabela 5.10 apresenta, nas colunas Laplace, Laplace* e Normal, a mediana do NICV por cenário para cada um dos algoritmos. É possível ver que o mecanismo de Laplace* apresenta o menor valor em todos os casos.

Tabela 5.10: Comparativo dos mecanismos em termos da mediana do NICV para cada valor de ϵ e algoritmo, considerando, em cada caso, as 300 execuções referentes a 10 repetições para cada conjunto de dados.

algoritmo	epsilon	Laplace	Laplace*	Normal
BRKGAPD	0,5	0,15800	0,14480	0,15130
BRKGAPD	1	0,12420	0,11580	0,11830
BRKGAPD	2	0,10220	0,09995	0,10130
BRKGAPD	3	0,10010	0,09775	0,09957
BRKGAPD	5	0,09885	0,09675	0,09864
BRKGAPD	10	0,09615	0,09560	0,09598
ILSPD	0,5	0,34780	0,28290	0,33230
ILSPD	1	0,20400	0,16120	0,18070
ILSPD	2	0,13240	0,12110	0,12850
ILSPD	3	0,11780	0,10500	0,11250
ILSPD	5	0,10400	0,09901	0,10320
ILSPD	10	0,09881	0,09715	0,09870
VNSPD	0,5	0,35740	0,28500	0,31120
VNSPD	1	0,21050	0,16420	0,18040
VNSPD	2	0,13470	0,12190	0,12890
VNSPD	3	0,12290	0,10750	0,11400
VNSPD	5	0,10400	0,09917	0,09961
VNSPD	10	0,09882	0,09737	0,09851

No entanto, a fim de avaliar se essa diferença é estatisticamente significativa, foi aplicado o teste de Friedman [79, 104], tendo, como hipótese nula, a igualdade entre os três mecanismos, em termos do NICV da solução final de agrupamento obtida pelo DPK-means. Para cada combinação de algoritmo e valor de ϵ , adotando-se o mecanismo

como tratamento, o teste de Friedman foi aplicado considerando 300 observações (blocos), resultantes da produto entre os 30 conjuntos de dados e as 10 repetições.

Os resultados apresentados na Tabela 5.11 evidenciam a superioridade consistente do algoritmo BRKGAPD, que obteve o melhor rank médio em todos os cenários avaliados, independentemente do mecanismo, abordagem para cálculo da sensibilidade ou valor de ϵ adotados. Esse comportamento sugere maior robustez e capacidade de adaptação do método frente às diferentes configurações. Os algoritmos ILSPD e VNSPD alternaram entre a segunda e a terceira posição, com o ILSPD apresentando desempenho ligeiramente superior na maioria dos cenários (67%), especialmente quando adotado o mecanismo de Laplace. Para cada cenário, o valor sublinhado indica o algoritmo que apresentou o menor rank médio e o sublinhado duplo indica o segundo colocado.

Tabela 5.11: Ranks por cenário dos algoritmos propostos.

epsilon	mecanismo	BRKGAPD	ILSPD	VNSPD
0,5	Laplace	<u>1,000</u>	<u>2,490</u>	2,510
0,5	Laplace*	<u>1,000</u>	<u>2,400</u>	2,600
0,5	Normal	<u>1,003</u>	2,547	<u>2,450</u>
1	Laplace	<u>1,000</u>	<u>2,470</u>	<u>2,530</u>
1	Laplace*	<u>1,000</u>	<u>2,467</u>	2,533
1	Normal	<u>1,000</u>	2,520	<u>2,480</u>
2	Laplace	<u>1,003</u>	<u>2,437</u>	<u>2,560</u>
2	Laplace*	<u>1,007</u>	<u>2,440</u>	2,553
2	Normal	<u>1,000</u>	<u>2,550</u>	<u>2,450</u>
3	Laplace	<u>1,000</u>	<u>2,440</u>	2,560
3	Laplace*	<u>1,007</u>	<u>2,437</u>	2,557
3	Normal	<u>1,003</u>	<u>2,497</u>	2,500
5	Laplace	<u>1,000</u>	<u>2,443</u>	2,557
5	Laplace*	<u>1,000</u>	<u>2,490</u>	2,510
5	Normal	<u>1,000</u>	2,507	<u>2,493</u>
10	Laplace	<u>1,000</u>	2,507	<u>2,493</u>
10	Laplace*	<u>1,003</u>	<u>2,473</u>	2,523
10	Normal	<u>1,003</u>	2,533	<u>2,463</u>

Na Tabela 5.12, a coluna Friedman apresenta os p-valores obtidos, indicando a rejeição da hipótese nula em todos os casos. A partir desse resultado, foi aplicado o teste de Wilcoxon pareado [197], usando correção de Holm [105, 7], entre o mecanismo de Laplace* e os outros dois mecanismos considerados. Nesse caso, a hipótese nula considera que a mediana das diferenças pareadas dos resultados obtidos através do mecanismo de Laplace* e do mecanismo concorrente é igual a zero. As colunas Laplace* x Laplace e Laplace* x Normal, da Tabela 5.12 apresentam os p-valores de cada caso, sendo a hipótese nula

novamente rejeitada em todos eles.

Tabela 5.12: P-valores dos testes de hipótese de significância entre diferenças de medianas de NICV para os mecanismos considerados.

algoritmo	epsilon	Friedman	Wilcoxon	
			Laplace* x Laplace	Laplace* x Normal
BRKGAPD	0,5	6,138e-92	2,694e-48	3,552e-49
BRKGAPD	1	4,055e-98	2,09e-50	2,039e-46
BRKGAPD	2	7,478e-91	7,001e-30	4,853e-27
BRKGAPD	3	5,024e-88	1,452e-28	1,452e-28
BRKGAPD	5	8,089e-75	2,474e-24	5,187e-23
BRKGAPD	10	4,175e-77	6,909e-27	9,292e-26
ILSPD	0,5	1,433e-94	1,962e-49	1,962e-49
ILSPD	1	3,493e-106	4,81e-42	4,81e-42
ILSPD	2	2,542e-76	1,258e-26	9,852e-23
ILSPD	3	1,893e-96	3,636e-46	2,799e-27
ILSPD	5	1,932e-78	1,199e-36	1,019e-26
ILSPD	10	2,316e-98	2,686e-38	7,193e-33
VNSPD	0,5	9,711e-107	9,623e-46	3,248e-42
VNSPD	1	3,857e-98	1,315e-38	2,593e-24
VNSPD	2	1,398e-87	5,136e-36	7,811e-31
VNSPD	3	1,153e-82	8,79e-26	5,906e-17
VNSPD	5	4,23e-95	3,141e-37	9,762e-26
VNSPD	10	2,566e-86	5,448e-36	4,506e-31

Os resultados sugerem, portanto, que a abordagem alternativa para o cálculo de Δf contribuiu para mitigar o efeito das caudas pesadas da distribuição de Laplace na geração do ruído, fazendo com que a utilidade do mecanismo de Laplace superasse a do mecanismo Normal. É provável que essa abordagem, quando associada ao mecanismo Normal, produza resultados ainda melhores em termos de utilidade. No entanto, esse mecanismo satisfaz apenas a privacidade diferencial aproximada (ϵ, δ) -PD, ao passo que o mecanismo de Laplace satisfaz a privacidade diferencial pura (ϵ) -PD.

Capítulo 6

Conclusões

Algoritmos de agrupamento desempenham papel central em tarefas de aprendizado não supervisionado e são amplamente utilizados em análises exploratórias e apoio à decisão [186, p.1], sendo o k-means particularmente destacado por sua simplicidade, eficiência computacional e ampla adoção prática [161, p.2]. Sendo assim, a incorporação de privacidade diferencial (PD) a esses métodos se torna essencial para viabilizar a análise de dados sensíveis sem comprometer a confidencialidade dos indivíduos, especialmente em cenários com múltiplos usuários e finalidades. Apesar disso, observou-se que grande parte dos trabalhos disponíveis na literatura não adota técnicas de otimização para equilibrar o *trade-off* entre utilidade e privacidade, um desafio crítico nesse domínio.

Nesse contexto, a pesquisa desenvolvida nesta tese visou contribuir na resolução do problema de alocação de *privacy budget* no contexto do k-means diferencialmente privado, um desafio central na medida em que a distribuição inadequada desse parâmetro pode comprometer significativamente o equilíbrio entre privacidade e utilidade. Em particular, as estratégias de alocação disponíveis na literatura tendem a introduzir o ruído de forma pouco eficaz, resultando em perturbações que acabam por prejudicar a qualidade da solução de agrupamento resultante. Diante disso, foram implementados três algoritmos baseados nas metaheurísticas BRKGA, ILS e VNS para a otimização da alocação do *privacy budget*, com o objetivo de maximizar a utilidade da solução de agrupamento gerada sob restrições de privacidade.

Para elaboração da abordagem apresentada, foi realizado um estudo dos métodos clássicos de agrupamento (Seção 2.2, p. 9), bem como de técnicas voltadas ao aprimoramento da privacidade (Apêndice A, p. 135). Mais especificamente, aprofundou-se o estudo da técnica de Privacidade Diferencial (Seção 2.4, p. 14) e de sua contribuição para mitigar o risco de violação da privacidade inerente à aplicação de métodos de agrupamento a

conjuntos de dados com atributos sensíveis (Seção 2.5, p. 34).

Em seguida, foi conduzida uma revisão sistemática da literatura, com o objetivo de compreender o estado da arte do problema de agrupamento diferencialmente privado (Seção 3.1, p. 43), identificar oportunidades de contribuição e mapear problemas de pesquisa em aberto. Foi estabelecido um protocolo, descrito no Apêndice C (Seção C.1.1, p. 169), que resultou na identificação de 1058 trabalhos. A aplicação dos critérios de seleção e avaliação da qualidade previstos no protocolo resultou na seleção de 47 trabalhos considerados relevantes para a proposta de pesquisa. Esses trabalhos foram estudados e um resumo das contribuições identificadas foi apresentado na Seção 3.1 (p.43).

Esse levantamento fundamentou a escolha das principais estratégias de alocação de *privacy budget* e confirmou o potencial ainda pouco explorado da aplicação de técnicas de otimização no problema abordado. Com isso, foi proposta a formulação apresentada na Seção 4.1 (p.65), que foi objeto do experimento cuja descrição e resultados foram apresentados no Capítulo 5.

Para a realização do experimento, foram selecionados 30 conjuntos de dados, entre sintéticos e reais, com boa diversidade quanto ao número de observações, atributos e classes.

Os algoritmos apresentados na Seção 4.5 foram implementados em linguagem Python, utilizando o *framework* Random-Key Optimizer (RKO) [37] (Seção 4.2) e com base na implementação dele mantida pelos autores [174]. No que se refere à calibração de parâmetros, esta foi conduzida através do método Q-Learning [196][37, p.23-24] (Seção 4.6), abordagem baseada em aprendizado por reforço (*reinforcement learning*), empregada como um mecanismo de controle adaptativo de parâmetros no contexto do RKO, permitindo ajustar dinamicamente a configuração de parâmetros das metaheurísticas (hiperparâmetros) ao longo da execução.

Os algoritmos BRKGAPD, ILSPD e VNSPD foram aplicados aos 30 conjuntos de dados, variando os três mecanismos de PD e os valores de *privacy budget* total no conjunto $\epsilon = \{10, 5, 3, 2, 1, 0,5\}$. As alocações de *privacy budget* produzidas pelos algoritmos foram comparadas com aquelas produzidas pelas cinco principais estratégias identificadas na revisão da literatura, quais sejam: Uniforme [23], DAS (*Dichotomy Allocation Strategy*) [54], SSAS (*Series Sum Allocation Strategy*) [70], APAS (*Arithmetic Progression Allocation Strategy*) [70] e Escalonada [155]. A comparação entre as estratégias foi feita com base no NICV da solução final de agrupamento obtido pelo algoritmo DPK-means, segundo cada estratégia considerada. As soluções resultantes dos algoritmos propostos

dominaram os resultados de todas as abordagens da literatura avaliadas. Em particular, as medianas do *gap* para a melhor solução, obtidas a partir dos resultados produzidos pelo algoritmo BRKGAPD, foram inferiores às de todas as outras abordagens, incluindo os outros dois algoritmos, em todos os cenários considerados. Complementarmente, foram aplicados testes de hipótese que indicaram diferenças estatisticamente significativas em todos os casos.

No que se refere aos tempos de execução, os resultados apresentados na Seção 5.4 (p.97) indicam que o BRKGAPD foi competitivo na comparação com os outros dois algoritmos, embora com ligeira desvantagem.

Em relação à privacidade diferencial, foram adotados os mecanismos de Laplace e Normal, descritos na Seção 2.4.3 (p.32). No uso do mecanismo de Laplace, em função das caudas mais pesadas da distribuição a ele associada, em comparação com a distribuição Normal, existe uma probabilidade não desprezível de que o ruído gerado seja de grande intensidade.

Uma vez que o parâmetro de dispersão da distribuição é diretamente proporcional à sensibilidade global ($b = \frac{\Delta f}{\epsilon}$), foram avaliadas duas abordagens para o cálculo da sensibilidade global (ver Definição 6, p.24), conforme detalhado na Seção 5.2 (p.86). A primeira abordagem, mais recorrente na literatura [23, p.5][216, p.490], foi definida simplesmente por $\Delta f = \Delta f^S + \Delta f^C$, e foi aplicada a ambos os mecanismos, Laplace e Normal. A segunda abordagem, utilizada apenas no mecanismo de Laplace, considerou, para o cálculo da soma dos atributos, apenas a variação dos valores dos atributos, $\Delta f^S = h$, e, para o cálculo da frequência dos objetos, a presença ou não do objeto, i.e., $\Delta f^C = 1$. A abordagem alternativa foi aplicada apenas ao mecanismo de Laplace, a fim de avaliar seu efeito na comparação com o mecanismo Normal. Os resultados apresentados na Seção 5.6 (p.104) sugeriram que a abordagem alternativa para o cálculo de Δf contribuiu para mitigar o efeito das caudas pesadas da distribuição de Laplace na geração do ruído, fazendo com que a utilidade do mecanismo de Laplace superasse a do mecanismo Normal.

Independentemente do cenário, os resultados demonstram a superioridade do BRKGAPD em produzir estratégias de alocação de *privacy budget* para o DPk-means que minimizem o NICV da solução de agrupamento resultante, o que, no contexto da privacidade diferencial, significa maximizar a utilidade. Mais especificamente, os resultados apresentados na Tabela 5.6 mostram que a mediana do *gap* percentual do BRKGAPD em relação à melhor solução foi de cerca de 10% no pior caso, qual seja, o cenário em que foi adotado o mecanismo de Laplace com $\epsilon = 0,5$.

Para o mesmo cenário, o segundo menor *gap* mediano foi de 148%, obtido pelo ILSPD. Dentre as estratégias da literatura, o melhor desempenho nesse cenário foi obtido pela estratégia Escalonada, com *gap* mediano de 187%. Além disso, o BRKGAPD foi o único a obter *gap* mediano menor do que 1 em todos os cenários em que $1 < \epsilon \leq 10$.

Como mencionado anteriormente (Seção 2.4.1), embora a definição do parâmetro ϵ seja dependente do contexto, de maneira geral, $0,01 \leq \epsilon \leq 1$ é considerado um intervalo adequado para dados sensíveis ou para cenários em que se prioriza a privacidade, ao passo que $1 < \epsilon \leq 10$ é adotado em contextos menos sensíveis, nos quais a utilidade é priorizada. Nesse sentido, no experimento realizado, o BRKGAPD apresentou os melhores resultados em termos de NICV decorrente da aplicação do DPk-means, independentemente do objetivo priorizado, seja ele privacidade ou utilidade.

Em termos dos mecanismos e parametrização da privacidade diferencial, a abordagem alternativa para cálculo da sensibilidade global, associada ao mecanismo de Laplace, também contribuiu para soluções de melhor utilidade. Mais especificamente, os resultados obtidos no experimento forneceram evidências de que a metodologia proposta, especialmente quando combinados à abordagem alternativa para o cálculo da sensibilidade, o mecanismo de Laplace e o BRKGAPD, foi capaz de explorar a alocação de *privacy budget* de forma mais eficaz, reduzindo a degradação associada à adição de ruído nas iterações do DPk-means e resultando em soluções de agrupamento de melhor qualidade, em termos do NICV.

Em suma, a utilização de abordagens de otimização baseadas em metaheurísticas combinadas com análise de agrupamentos, aplicadas ao problema de privacidade diferencial, constitui-se como uma alternativa promissora para mitigar o *trade-off* entre utilidade e privacidade. Os resultados obtidos demonstraram a eficácia da alocação de *privacy budget* orientada por algoritmos baseados em metaheurísticas, que foram capazes de maximizar a utilidade, produzindo soluções de agrupamento de melhor qualidade quando comparadas às abordagens da literatura, enquanto preservaram as garantias formais de privacidade.

Os achados obtidos reforçam o potencial do uso de técnicas de otimização no contexto do problema de agrupamento diferencialmente privado, mais especificamente do método DPk-means, e evidenciam que abordagens capazes de integrar mecanismos de PD e métodos avançados de busca, representam um caminho promissor para o desenvolvimento de soluções mais eficazes, robustas e adequadas a aplicações reais envolvendo dados sensíveis. Nesse sentido, o avanço de métodos capazes de conciliar utilidade analítica e proteção

formal da privacidade contribui para reduzir, de maneira segura, barreiras de acesso a dados sensíveis, frequentemente indisponíveis devido aos riscos associados à sua divulgação. Como consequência, amplia-se o potencial de desenvolvimento de metodologias, modelos analíticos e estudos aplicados com capacidade de gerar impactos relevantes em diferentes domínios científicos, econômicos e sociais.

Como produtos diretos desta pesquisa, destaca-se a elaboração do artigo intitulado *Enhancing Clustering Utility in DPk-means through Metaheuristic Optimization of Privacy Budget Allocation*, atualmente em processo de revisão no periódico *Expert Systems with Applications*, no qual são detalhados os principais resultados e contribuições metodológicas desta tese. Adicionalmente, encontra-se em desenvolvimento um artigo complementar, a ser submetido a um periódico da área de estatística, com o objetivo de explorar aplicações e aprofundar aspectos metodológicos e analíticos, incluindo o estudo do conjunto abrangente de soluções de boa qualidade produzido pelos três algoritmos, a fim de investigar potenciais padrões estruturais que possam ser generalizados para diferentes conjuntos de dados e cenários de privacidade.

6.1 Trabalhos futuros

Apesar da comprovada eficácia da abordagem proposta, em especial do arranjo composto pelo algoritmo BRKGAPD, com o mecanismo de Laplace e o cálculo alternativo da sensibilidade, frente às principais estratégias identificadas na literatura, a natureza abrangente e complexa do problema e suas múltiplas vertentes sugerem diversos caminhos para estudos futuros.

Diversos desafios associados ao problema de agrupamento clássico são apontados na literatura [114, p.656] e representam desafios também no contexto do problema de agrupamento diferencialmente privado, devendo ser investigados em trabalhos futuros. Entre eles, destacam-se a seleção dos atributos relevantes [96], a identificação de valores atípicos (*outliers*) [103], a escolha de uma medida apropriada de similaridade ou dissimilaridade [169], a determinação do número ideal de grupos presentes nos dados [167], a seleção do algoritmo de agrupamento mais adequado [17], a verificação da existência de uma estrutura de agrupamento nos dados [94, p.484-486] e a validação dos grupos obtidos [186].

Em termos de índices e métricas utilizados para avaliar a qualidade da solução de agrupamento produzida pelo DPk-means, na presente pesquisa foi adotado o *Normalized Intra-cluster Variance* (NICV), em função de sua ampla utilização na literatura. Entre-

tanto, outros índices igualmente populares podem ser investigados [186], como o índice de Calinski-Harabasz (CH) [30], sendo para isso apenas necessário adaptar a função de custo (Seção 4.4, p. 73).

Ainda no que diz respeito à produção da solução de agrupamento, a definição do número de grupos k é um problema inerente a métodos de agrupamento baseados em protótipos, como o k-means. Uma escolha inadequada de k pode comprometer significativamente a utilidade da solução. No problema de agrupamento clássico (sem PD), uma abordagem típica para mitigar esse efeito consiste em executar o algoritmo de agrupamento com diferentes valores de k e selecionar aquele que produz a melhor solução segundo um índice de validação¹ (e.g., NICV ou CH). No entanto, no contexto do problema abordado, múltiplas execuções do algoritmo, com diferentes valores de k , aumentam substancialmente o risco de revelação. Embora essa questão não tenha sido abordada nesta pesquisa, pode-se incorporar abordagens como *Private Selection from Private Candidates* (PSPC) [130], conforme sugerido em [44, p.165–167], para abordar o problema.

Em relação à abordagem de otimização utilizada, o *framework* Random-Key Optimizer (RKO) [37] (Seção 4.2, p. 69) é dotado de estrutura modular, que permite, através de recursos de decodificação, diversificação e intensificação, a aplicação direta de diferentes metaheurísticas. Apesar dos bons resultados alcançados no experimento, alguns recursos disponíveis no *framework* não foram explorados, como o uso da busca local RVND e perturbação (*shaking*) no BRKGA e a implementação de algoritmos baseados em outras metaheurísticas. Além disso, sua execução mantém um *pool* compartilhado de soluções elite, que atua como memória global, promovendo intensificação e diversificação ao longo do processo. Esse *pool* é capaz de receber e combinar soluções provenientes de algoritmos baseados em qualquer uma das metaheurísticas implementadas no *framework*. Embora o experimento realizado nessa pesquisa tenha feito uso desse *pool*, sua utilização se deu de maneira isolada, i.e., os três algoritmos foram executados separadamente, com o *pool* restrito ao contexto interno de cada algoritmo, não havendo compartilhamento de soluções entre eles. Esses recursos não explorados do *framework* RKO têm grande potencial, e podem ser adotados em estudos futuros.

No que se refere à aplicação de privacidade diferencial ao k-means, na etapa de atualização dos centroides, o ruído foi gerado e introduzido de forma independente no cálculo da soma das distâncias e da frequência de objetos em cada grupo, como é feito predo-

¹Essa abordagem é geralmente denominada método ou critério do cotovelo (do inglês, *elbow*) e, embora muito difundida em aplicações relacionadas ao problema de agrupamento clássico, seu uso não é consenso nem mesmo nesse contexto [166].

minantemente na literatura sobre o tema. Todavia, há um estudo [6] que sugere que a perturbação das coordenadas do centroide pode ser mais eficiente em alguns casos. Apesar de levar a uma maior sensibilidade global e, portanto, a um aumento do ruído necessário para garantir o mesmo nível de privacidade, ao adicionar ruído diretamente à média (centroide), evita-se a propagação acumulada dos erros introduzidos na soma e na contagem, que podem ser amplificados quando combinados.

Outro aspecto relativo à atualização dos centroides que pode ser melhor explorado em abordagens baseadas em otimização, é a distribuição do valor de ϵ_t , i.e., do *privacy budget* alocado à iteração t do DPk-means, entre as consultas de soma de distâncias (S_j) e frequências ($|C_j|$). Conforme descrito na 4.1 (p.65), para fins de simplificação, em cada iteração t , metade de ϵ_t foi usada para cada consulta, i.e., foi considerado que $\epsilon^s = \epsilon^c = \frac{1}{2}\epsilon_t$. Todavia, é razoável supor que distribuições mais eficazes, em termos de produzir soluções de maior utilidade, possam ser propostas. Uma possível abordagem para isso é adaptar a representação de solução proposta na Seção 4.3 (p.72) para armazenar parcelas de *privacy budget* independentes para as consultas de soma de distâncias e frequências em cada uma das T iterações.

No que se refere ao cálculo da sensibilidade global, é provável que a abordagem alternativa, quando associada ao mecanismo Normal, produza resultados ainda melhores em termos de utilidade do que aqueles alcançados com o mecanismo de Laplace. No entanto, o mecanismo Normal satisfaz a privacidade diferencial aproximada (ϵ, δ) -PD, ao passo que o mecanismo de Laplace satisfaz a privacidade diferencial pura (ϵ) -PD. Por outro lado, o mecanismo de Laplace truncado [44, p.116-117], outra alternativa apontada na literatura para lidar com o problema das caudas pesadas da distribuição de Laplace, tem a característica de satisfazer (ϵ, δ) -PD, sendo comparável ao mecanismo Normal em termos da garantia de privacidade oferecida. Assim, um experimento promissor seria comparar ambos os mecanismos com as abordagens tradicional e alternativa para o cálculo da sensibilidade.

Uma vertente adicional da metodologia proposta a ser explorada em trabalhos futuros é sua flexibilidade e potencial de generalização para além do DPk-means. Mais especificamente, a formulação e os algoritmos propostos podem ser facilmente adaptados a outros métodos iterativos que operem sob o modelo DP interativo, desde que o número total de iterações seja previamente estabelecido. Para isso, a adaptação necessária consiste essencialmente em redefinir a função objetivo, o que é facilitado pela arquitetura modular do *framework* RKO.

No que se refere à privacidade diferencial em uma perspectiva mais ampla, é possível ainda abordar o problema de agrupamento diferencialmente privado adotando o modelo não interativo de PD, através do uso de *coresets* [72, 74, p.362], conforme discutido no Apêndice E (p.178).

Referências

- [1] ABE, Y.; MINAMI, K. Matching Attacks on Non-deterministic Algorithms for Cell Suppression Problem for Tabular Data. In *2022 IEEE International Conference on Big Data (Big Data)* (Dec. 2022), pp. 2169–2174.
- [2] ABOWD, J. M.; ADAMS, T.; ASHMEAD, R.; DARAIS, D.; DEY, S.; GARFINKEL, S. L.; GOLDSCHLAG, N.; KIFER, D.; LECLERC, P.; LEW, E.; MOORE, S.; RODRÍGUEZ, R. A.; TADROS, R. N.; VILHUBER, L. The 2010 Census Confidentiality Protections Failed, Here’s How and Why, Dec. 2023. arXiv:2312.11283 [cs, econ, stat].
- [3] ADAMS, C. *Introduction to Privacy Enhancing Technologies: A Classification-Based Approach to Understanding PETs*. Springer International Publishing, Cham, 2021.
- [4] AGARWAL, P. K.; HAR-PELED, S.; VARADARAJAN, K. R. Approximating extent measures of points. *J. ACM* 51, 4 (July 2004), 606–635.
- [5] AGRESTI, A. *Categorical Data Analysis*, 1 ed. Wiley Series in Probability and Statistics. Wiley, July 2002.
- [6] AHMED, A. *Differentially-private Multiparty Clustering*. Tese de Doutorado, University of Waterloo, Sept. 2023.
- [7] AICKIN, M.; GENSLER, H. Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *American Journal of Public Health* 86, 5 (May 1996), 726–728.
- [8] ALCALÁ-FDEZ, J.; FERNANDEZ, A.; LUENGO, J.; DERRAC, J.; GARCÍA, S.; SÁNCHEZ, L.; HERRERA, F. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Journal of Multiple-Valued Logic and Soft Computing* 17, 2-3 (2011), 255–287.
- [9] ALIZADEH, E. The ABCs of Differential Privacy, Apr. 2023.
- [10] ALOISE, D.; DESHPANDE, A.; HANSEN, P.; POPAT, P. NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning* 75, 2 (May 2009), 245–248.
- [11] ALZOUBI, Y. I.; MISHRA, A. Differential privacy and artificial intelligence: potentials, challenges, and future avenues. *EURASIP Journal on Information Security* 2025, 1 (May 2025), 18.

- [12] ARTHUR, D.; VASSILVITSKII, S. k-means++: the advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (USA, 2007), SODA '07, Society for Industrial and Applied Mathematics, pp. 1027–1035. event-place: New Orleans, Louisiana.
- [13] ASHRAF, M.; RADY, S.; ABDELKADER, T.; GHARIB, T. F. Efficient privacy preserving algorithms for hiding sensitive high utility itemsets. *Computers & Security* 132 (2023), 103360.
- [14] BALCAN, M.-F.; DICK, T.; LIANG, Y.; MOU, W.; ZHANG, H. Differentially private clustering in high-dimensional Euclidean spaces. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (2017), ICML'17, JMLR.org, pp. 322–331. event-place: Sydney, NSW, Australia.
- [15] BANERJEE, A.; DAVE, R. Validating clusters using the Hopkins statistic. In *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)* (July 2004), vol. 1, pp. 149–153 vol.1. ISSN: 1098-7584.
- [16] BARAK, B.; CHAUDHURI, K.; DWORK, C.; KALE, S.; MCSHERRY, F.; TALWAR, K. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (New York, NY, USA, June 2007), PODS '07, Association for Computing Machinery, pp. 273–282.
- [17] BELTRÃO, A. P. V.; OCHI, L. S.; BRITO, J. A. D. M.; SEMAAN, G. S.; FADEL, A. C. BRKGA applied to the cluster ensemble problem. *RAIRO - Operations Research* 60, 4 (July 2026), 1931–1957.
- [18] BENNETT, K.; BRADLEY, P.; DEMIRIZ, A. Constrained K-Means Clustering. Tech. Rep. MSR-TR-2000-65, Microsoft Research, May 2000.
- [19] BISWAS, R. H. K. P. Decentralized big data mining: federated learning for clustering youth tobacco use in India. *Journal of Big Data* 11, 1 (2024).
- [20] BKAKRIA, A.; TASIDOU, A.; CUPPENS-BOULAHIA, N.; CUPPENS, F.; BOUATTOUR, F.; BEN FREDJ, F. Optimal Distribution of Privacy Budget in Differential Privacy. In *Risks and Security of Internet and Systems* (Cham, 2019), A. Zemari, M. Mosbah, N. Cuppens-Boulahia, and F. Cuppens, Eds., Lecture Notes in Computer Science, Springer International Publishing, pp. 222–236.
- [21] BLAKLEY, G. R.; KABATIANSKY, G. Secret Sharing Schemes. In *Encyclopedia of Cryptography and Security*, H. C. A. van Tilborg and S. Jajodia, Eds. Springer US, Boston, MA, 2011, pp. 1095–1097.
- [22] BLANCO-JUSTICIA, A.; SÁNCHEZ, D.; DOMINGO-FERRER, J.; MURALIDHAR, K. A Critical Review on the Use (and Misuse) of Differential Privacy in Machine Learning. *ACM Computing Surveys* 55, 8 (Aug. 2023), 1–16.
- [23] BLUM, A.; DWORK, C.; MCSHERRY, F.; NISSIM, K. Practical privacy: the SuLQ framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (New York, NY, USA, June 2005), PODS '05, Association for Computing Machinery, pp. 128–138.

- [24] BLUM, A.; LIGETT, K.; ROTH, A. A learning theory approach to non-interactive database privacy. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2008), STOC '08, Association for Computing Machinery, pp. 609–618. event-place: Victoria, British Columbia, Canada.
- [25] BOGETOFT, P.; CHRISTENSEN, D. L.; DAMGÅRD, I.; GEISLER, M.; JAKOBSEN, T.; KRØIGAARD, M.; NIELSEN, J. D.; NIELSEN, J. B.; NIELSEN, K.; PAGTER, J.; SCHWARTZBACH, M.; TOFT, T. Secure Multiparty Computation Goes Live. In *Financial Cryptography and Data Security* (Berlin, Heidelberg, 2009), R. Dingledine and P. Golle, Eds., Springer, pp. 325–343.
- [26] BONDY, A.; MURTY, U. *Graph Theory*. Graduate Texts in Mathematics. Springer London, 2010.
- [27] BRIMBERG, J.; SALHI, S.; TODOSIJEVIĆ, R.; UROŠEVIĆ, D. Variable Neighborhood Search: The power of change and simplicity. *Computers & Operations Research* 155 (July 2023), 106221.
- [28] BUN, M.; STEINKE, T. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *Theory of Cryptography* (Berlin, Heidelberg, 2016), M. Hirt and A. Smith, Eds., Springer, pp. 635–658.
- [29] BYUN, J.; CHOI, Y.; LEE, J. Improving the utility of differentially private clustering through dynamical processing. *Pattern Recognition* 157 (2025), 110890.
- [30] CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods* 3, 1 (1974), 1–27.
- [31] CANONNE, C. L.; KAMATH, G.; STEINKE, T. The discrete Gaussian for differential privacy. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, Dec. 2020), NIPS '20, Curran Associates Inc., pp. 15676–15688.
- [32] CARVALHO, T.; MONIZ, N.; FARIA, P.; ANTUNES, L. Survey on Privacy-Preserving Techniques for Microdata Publication. *ACM Comput. Surv.* 55, 14s (July 2023).
- [33] CASTRO, J. Thirty years of optimization-based SDC methods for tabular data. *Trans. Data Priv.* 16 (2023), 3–13.
- [34] CASTRO, J.; GENTILE, C.; SPAGNOLO-ARRIZABALAGA, E. An algorithm for the microaggregation problem using column generation. *Computers & Operations Research* 144 (Aug. 2022), 105817.
- [35] CASTRO, J.; GONZÁLEZ, J. A. A linear optimization-based method for data privacy in statistical tabular data. *Optimization Methods and Software* 34, 1 (Jan. 2019), 37–61.
- [36] CHATZIKOKOLAKIS, K.; ANDRÉS, M. E.; BORDENABE, N. E.; PALAMIDESSI, C. Broadening the Scope of Differential Privacy Using Metrics. In *Privacy Enhancing Technologies*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C.

- Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, E. De Cristofaro, and M. Wright, Eds., vol. 7981. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 82–102.
- [37] CHAVES, A. A.; RESENDE, M. G. C.; SCHUETZ, M. J. A.; BRUBAKER, J. K.; KATZGRABER, H. G.; DE ARRUDA, E. F.; SILVA, R. M. A. A Random-Key Optimizer for Combinatorial Optimization, 2024.
- [38] CHEN, H.; ZHOU, Y.; MEI, K.; WANG, N.; CAI, G. A New Density Peak Clustering Algorithm With Adaptive Clustering Center Based on Differential Privacy. *IEEE Access* 11 (2023), 1418–1431.
- [39] CHEN, Y.; GAN, W.; WU, Y.; YU, P. S. Privacy-preserving federated mining of frequent itemsets. *Information Sciences* 625 (2023), 504–520.
- [40] CHENG, D.; LI, Y.; XIA, S.; WANG, G.; HUANG, J.; ZHANG, S. A Fast Granular-Ball-Based Density Peaks Clustering Algorithm for Large-Scale Data. *IEEE Transactions on Neural Networks and Learning Systems* 35, 12 (Dec. 2024), 17202–17215.
- [41] CHRISTOFIDES, T. C. Chapter 4 - The Classical Randomized Response Techniques: Reading Warner (1965) and Greenberg et al. (1969) 50 Years Later. In *Handbook of Statistics*, A. Chaudhuri, T. C. Christofides, and C. R. Rao, Eds., vol. 34 of *Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits*. Elsevier, Jan. 2016, pp. 29–41.
- [42] COHEN-ADDAD, V.; GREEN LARSEN, K.; SAULPIC, D.; SCHWIEGELSHOHN, C.; SHEIKH-OMAR, O. A. Improved Coresets for Euclidean k-Means. In *Advances in Neural Information Processing Systems* (2022), S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35, Curran Associates, Inc., pp. 2679–2694.
- [43] CONRADS, T.; DREXLER, L.; KÖNEN, J.; SCHMIDT, D. R.; SCHMIDT, M. Local Search k-means++ with Foresight. *LIPICs, Volume 301, SEA 2024* 301 (2024), 7:1–7:20.
- [44] COWAN, E.; SHOEMATE, M.; PEREIRA, M. *Hands-On Differential Privacy*, 1 ed. O'Reilly Media, Inc., 2024.
- [45] CUMMINGS, R.; DESFONTAINES, D.; EVANS, D.; GEAMBASU, R.; HUANG, Y.; JAGIELSKI, M.; KAIROUZ, P.; KAMATH, G.; OH, S.; OHRIMENKO, O.; PAPERNOT, N.; ROGERS, R.; SHEN, M.; SONG, S.; SU, W.; TERZIS, A.; THAKURTA, A.; VASSILVITSKII, S.; WANG, Y.-X.; XIONG, L.; YEKHANIN, S.; YU, D.; ZHANG, H.; ZHANG, W. Advancing Differential Privacy: Where We Are Now and Future Directions for Real-World Deployment. *Harvard Data Science Review* 6, 1 (Jan. 2024).
- [46] CYPHERS, B. Understanding differential privacy and why it matters for digital rights, Oct. 2017.
- [47] CZUMAJ, A.; SOHLER, C. Sublinear-time approximation algorithms for clustering via random sampling. *Random Struct. Algorithms* 30, 1–2 (Jan. 2007), 226–256.

- [48] DENNING, D. E. Secure statistical databases with random sample queries. *ACM Trans. Database Syst.* 5, 3 (Sept. 1980), 291–315.
- [49] DINUR, I.; NISSIM, K. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (New York, NY, USA, June 2003), PODS '03, Association for Computing Machinery, pp. 202–210.
- [50] DOMINGO-FERRER, J.; MATEO-SANZ, J. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and Data Engineering* 14, 1 (Jan. 2002), 189–201.
- [51] DOMINGO-FERRER, J.; SÁNCHEZ, D.; BLANCO-JUSTICIA, A. The Limits of Differential Privacy (and its Misuse in Data Release and Machine Learning), 2020.
- [52] DONG, J.; ROTH, A.; SU, W. J. Gaussian Differential Privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84, 1 (Feb. 2022), 3–37.
- [53] DRECHSLER, J. *Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation*. Springer New York, NY, 2011.
- [54] DWORK, C. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation* (Berlin, Heidelberg, 2008), M. Agrawal, D. Du, Z. Duan, and A. Li, Eds., Springer, pp. 1–19.
- [55] DWORK, C.; KOHLI, N.; MULLIGAN, D. Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality* 9, 2 (Oct. 2019).
- [56] DWORK, C.; LEI, J. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing* (New York, NY, USA, May 2009), STOC '09, Association for Computing Machinery, pp. 371–380.
- [57] DWORK, C.; MCSHERRY, F.; NISSIM, K.; SMITH, A. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography* (Berlin, Heidelberg, 2006), S. Halevi and T. Rabin, Eds., Lecture Notes in Computer Science, Springer, pp. 265–284.
- [58] DWORK, C.; NISSIM, K. Privacy-Preserving Datamining on Vertically Partitioned Databases. In *Advances in Cryptology – CRYPTO 2004* (Berlin, Heidelberg, 2004), M. Franklin, Ed., Springer, pp. 528–544.
- [59] DWORK, C.; ROTH, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.* 9, 3–4 (aug 2014), 211–407.
- [60] DWORK, C.; SMITH, A.; STEINKE, T.; ULLMAN, J. Exposed! A Survey of Attacks on Private Data. *Annual Review of Statistics and Its Application* 4, 1 (Mar. 2017), 61–84.
- [61] ELABD, E. Dynamic differential privacy technique for deep learning models. *Scientific Reports* 15, 1 (2025), 39353.

- [62] ELLIOT, M.; HUNDEPOOL, A.; NORDHOLT, E. S.; TAMBAY, J.-L.; WENDE, T. Glossary on statistical disclosure control. *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality* (2005).
- [63] EPASTO, A.; MIRROKNI, V.; NARAYANAN, S.; ZHONG, P. k-Means Clustering with Distance-Based Privacy. In *Advances in Neural Information Processing Systems* (2023), A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36, Curran Associates, Inc., pp. 19570–19593.
- [64] ERLINGSSON, U.; PIHUR, V.; KOROLOVA, A. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale Arizona USA, Nov. 2014), ACM, pp. 1054–1067.
- [65] ERZURUM CICEK, Z. I.; KAMISLI OZTURK, Z. Optimizing the artificial neural network parameters using a biased random key genetic algorithm for time series forecasting. *Applied Soft Computing* 102 (Apr. 2021), 107091.
- [66] EZUGWU, A. E.; IKOTUN, A. M.; OYELADE, O. O.; ABUALIGAH, L.; AGUSHAKA, J. O.; EKE, C. I.; AKINYELU, A. A. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence* 110 (Apr. 2022), 104743.
- [67] FADEL, A. C.; OCHI, L. S.; BRITO, J. A. D. M.; SEMAAN, G. S. Microaggregation heuristic applied to statistical disclosure control. *Information Sciences* 548 (Feb. 2021), 37–55.
- [68] FAHAD, A.; ALSHATRI, N.; TARI, Z.; ALAMRI, A.; KHALIL, I.; ZOMAYA, A. Y.; FOUFOU, S.; BOURAS, A. A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis. *IEEE Transactions on Emerging Topics in Computing* 2, 3 (Sept. 2014), 267–279.
- [69] FALLS, M.; BERNARDELLO, R.; CASTRILLO, M.; ACOSTA, M.; LLORT, J.; GALÍ, M. Use of genetic algorithms for ocean model parameter optimisation: a case study using PISCES-v2_rc for North Atlantic particulate organic carbon. *Geoscientific Model Development* 15, 14 (July 2022), 5713–5737.
- [70] FAN, Z.; XU, X. APDPk-Means: A New Differential Privacy Clustering Algorithm Based on Arithmetic Progression Privacy Budget Allocation. In *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (Aug. 2019), pp. 1737–1742.
- [71] FEDERRATH, H. Privacy-Enhancing Technologies (PET). In *Encyclopedia of Cryptography, Security and Privacy*, S. Jajodia, P. Samarati, and M. Yung, Eds. Springer Nature Switzerland, Cham, 2025, pp. 1932–1935.
- [72] FELDMAN, D.; FIAT, A.; KAPLAN, H.; NISSIM, K. Private coresets. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing* (New York,

- NY, USA, 2009), STOC '09, Association for Computing Machinery, pp. 361–370. event-place: Bethesda, MD, USA.
- [73] FELDMAN, D.; LANGBERG, M. A unified framework for approximating and clustering data. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2011), STOC '11, Association for Computing Machinery, pp. 569–578. event-place: San Jose, California, USA.
- [74] FELDMAN, D.; XIANG, C.; ZHU, R.; RUS, D. Coresets for differentially private k-means clustering and applications to privacy in mobile sensor networks. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks* (New York, NY, USA, 2017), IPSN '17, Association for Computing Machinery, pp. 3–15. event-place: Pittsburgh, Pennsylvania.
- [75] FISHER, R. A. Iris, 1936. Published: UCI Machine Learning Repository.
- [76] FLEMING, G.; BRUCE, P. *Responsible Data Science*. Wiley, 2021.
- [77] FORGY, E. W. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *Biometrics* (1965).
- [78] FREITAS, V. Perform Systematic Literature Reviews, 2014.
- [79] FRIEDMAN, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association* 32, 200 (Dec. 1937), 675–701.
- [80] FU, N.; NI, W.; HU, H.; ZHANG, S. Multidimensional grid-based clustering with local differential privacy. *Information Sciences* 623 (2023), 402–420.
- [81] FU, N.; WANG, J.; XU, Y.; NI, W.; HOU, L.; ZHANG, D.; YANG, Y. Distributed clustering under local differential privacy. *Knowledge-Based Systems* 333 (2026), 114994.
- [82] GAO, Y. Application of optimized differential privacy protection algorithm in data mining privacy protection. *International Journal of Data Science and Analytics* 21, 1 (June 2026), 19.
- [83] GENG, Q.; DING, W.; GUO, R.; KUMAR, S. Tight Analysis of Privacy and Utility Tradeoff in Approximate Differential Privacy. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (June 2020), PMLR, pp. 89–99.
- [84] GENG, Q.; VISWANATH, P. The optimal mechanism in differential privacy. In *2014 IEEE International Symposium on Information Theory* (June 2014), pp. 2371–2375. ISSN: 2157-8117.
- [85] GENG, Q.; VISWANATH, P. The Optimal Noise-Adding Mechanism in Differential Privacy. *IEEE Transactions on Information Theory* 62, 2 (Feb. 2016), 925–951. Conference Name: IEEE Transactions on Information Theory.
- [86] GENTRY, C. *A fully homomorphic encryption scheme*. Tese de Doutorado, Stanford University, 2009.

- [87] GHOSHAL, P.; DHAKA, M.; SAIRAM, A. S. On the effectiveness of differential privacy to continuous queries. *Service Oriented Computing and Applications* (May 2024).
- [88] GOLDBREICH, O.; MICALI, S.; WIGDERSON, A. How to play any mental game, or a completeness theorem for protocols with honest majority. In *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. Association for Computing Machinery, New York, NY, USA, Oct. 2019, pp. 307–328.
- [89] GONÇALVES, J. F.; RESENDE, M. G. C. Biased random-key genetic algorithms for combinatorial optimization. *Journal of Heuristics* 17, 5 (Oct. 2011), 487–525.
- [90] GONÇALVES, J. F.; RESENDE, M. G. C. Biased Random-Key Genetic Programming. In *Handbook of Heuristics*, R. Martí, P. M. Pardalos, and M. G. C. Resende, Eds. Springer International Publishing, Cham, 2018, pp. 23–37.
- [91] GOUWEELEEUW, J. M.; KOOIMAN, P.; WILLENBORG, L.; DE-WOLF, P. P. Post randomisation for statistical disclosure control: Theory and implementation. Tech. rep., Statistics Netherlands, 1997.
- [92] GRAHAM, R. L.; KNUTH, D. E.; PATASHNIK, O. *Concrete mathematics: a foundation for computer science*, 2. ed., 31. print ed. Addison-Wesley, Upper Saddle River, NJ, 2017.
- [93] GU, Y.; ZHANG, Y.; ZHANG, H. Particle Swarm K-Means++ Clustering Method Based on Multiple Differential Privacy Protection Mechanism. In *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)* (May 2023), vol. 3, pp. 769–773.
- [94] HAN, J.; KAMBER, M.; PEI, J. 10 - Cluster Analysis: Basic Concepts and Methods. In *Data Mining (Third Edition)*, J. Han, M. Kamber, and J. Pei, Eds., The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Boston, Jan. 2012, pp. 443–495.
- [95] HAN, L.; XIE, Y.; FAN, D.; LIU, J. Improved differential privacy K-means clustering algorithm for privacy budget allocation. In *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)* (July 2022), pp. 221–225.
- [96] HANCER, E.; XUE, B.; ZHANG, M. A survey on feature selection approaches for clustering. *Artificial Intelligence Review* 53, 6 (Aug. 2020), 4519–4545.
- [97] HANSEN, P.; MLADENović, N. Variable Neighborhood Search. In *Handbook of Heuristics*. Springer, Cham, 2018, pp. 759–787.
- [98] HAR-PELED, S.; MAZUMDAR, S. On coresets for k-means and k-median clustering. In *Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2004), STOC '04, Association for Computing Machinery, pp. 291–300. event-place: Chicago, IL, USA.
- [99] HARTIGAN, J. A.; WONG, M. A. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108.

- [100] HASTIE, T.; FRIEDMAN, J.; TIBSHIRANI, R. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2001.
- [101] HE, Z.; WANG, L.; CAI, Z. Clustered Federated Learning With Adaptive Local Differential Privacy on Heterogeneous IoT Data. *IEEE Internet of Things Journal* 11, 1 (Jan. 2024), 137–146.
- [102] HOCHBERG, Y.; TAMHANE, A. C. *Multiple Comparison Procedures*, 1 ed. Wiley Series in Probability and Statistics. Wiley, Sept. 1987.
- [103] HODGE, V.; AUSTIN, J. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* 22, 2 (Oct. 2004), 85–126.
- [104] HODGES, J. L.; LEHMANN, E. L. Rank Methods for Combination of Independent Experiments in Analysis of Variance. In *Selected Works of E. L. Lehmann*, J. Rojo, Ed. Springer US, Boston, MA, 2012, pp. 403–418.
- [105] HOLM, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2 (1979), 65–70.
- [106] HOSHINO, N. A firm foundation for statistical disclosure control. *Japanese Journal of Statistics and Data Science* 3, 2 (Dec. 2020), 721–746.
- [107] HOU, J.; YUAN, H.; PELILLO, M. Towards Parameter-Free Clustering for Real-World Data. *Pattern Recognition* 134 (Feb. 2023), 109062.
- [108] HUANG, L.; VISHNOI, N. K. Coresets for clustering in Euclidean spaces: importance sampling is nearly optimal. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (New York, NY, USA, 2020), STOC 2020, Association for Computing Machinery, pp. 1416–1429. event-place: Chicago, IL, USA.
- [109] HUANG, Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2, 3 (Sept. 1998), 283–304.
- [110] HUANG, Z.; LIU, J. Optimal Differentially Private Algorithms for k-Means Clustering. In *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems* (New York, NY, USA, May 2018), PODS '18, Association for Computing Machinery, pp. 395–408.
- [111] HUNDEPOOL, A.; DOMINGO-FERRER, J.; FRANCONI, L.; GIESSING, S.; NORDHOLT, E.; SPICER, K.; DE WOLF, P. *Statistical Disclosure Control*. Wiley Series in Survey Methodology. Wiley, 2012.
- [112] ISMAIL, A. P. A. B. F. Machine learning security and privacy: a review of threats and countermeasures. *EURASIP Journal on Information Security* 2024, 1 (2024).
- [113] JADON, A.; KUMAR, S. Leveraging Generative AI Models for Synthetic Data Generation in Healthcare: Balancing Research and Privacy. In *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)* (July 2023), pp. 1–4.

- [114] JAIN, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31, 8 (June 2010), 651–666.
- [115] JAPA, L.; SERQUEIRA, M.; MENDONÇA, I.; ARITSUGI, M.; BEZERRA, E.; GONZÁLEZ, P. H. A Population-Based Hybrid Approach for Hyperparameter Optimization of Neural Networks. *IEEE Access* 11 (2023), 50752–50768.
- [116] JIA, P.; ZHANG, J.; ZHAO, B.; LI, H.; LIU, X. Privacy-preserving association rule mining via multi-key fully homomorphic encryption. *Journal of King Saud University - Computer and Information Sciences* 35, 2 (2023), 641–650.
- [117] JIN, J.; MCMURTRY, E.; RUBINSTEIN, B. I. P.; OHRIMENKO, O. Are We There Yet? Timing and Floating-Point Attacks on Differential Privacy Systems. In *2022 IEEE Symposium on Security and Privacy (SP)* (San Francisco, CA, USA, May 2022), IEEE, pp. 473–488.
- [118] KAIROUZ, P.; MCMAHAN, H. B.; AVENT, B.; BELLET, A.; BENNIS, M.; BHAGOJI, A. N.; BONAWIT, K.; CHARLES, Z.; CORMODE, G.; CUMMINGS, R.; D’OLIVEIRA, R. G. L.; EICHNER, H.; EL ROUAYHEB, S.; EVANS, D.; GARDNER, J.; GARRETT, Z.; GASCÓN, A.; GHAZI, B.; GIBBONS, P. B.; GRUTESER, M.; HARCHAOUI, Z.; HE, C.; HE, L.; HUO, Z.; HUTCHINSON, B.; HSU, J.; JAGGI, M.; JAVIDI, T.; JOSHI, G.; KHODAK, M.; KONECNÝ, J.; KOROLOVA, A.; KOUSHANFAR, F.; KOYEJO, S.; LEPOINT, T.; LIU, Y.; MITTAL, P.; MOHRI, M.; NOCK, R.; ÖZGÜR, A.; PAGH, R.; QI, H.; RAMAGE, D.; RASKAR, R.; RAYKOVA, M.; SONG, D.; SONG, W.; STICH, S. U.; SUN, Z.; THEERTHA SURESH, A.; TRAMÈR, F.; VEPAKOMMA, P.; WANG, J.; XIONG, L.; XU, Z.; YANG, Q.; YU, F. X.; YU, H.; ZHAO, S. *Advances and Open Problems in Federated Learning*. Now Foundations and Trends, 2021.
- [119] KAISSIS, A. Z. T. M. S. F. F. B. B. R. Reconciling privacy and accuracy in AI for medical imaging. *Nature Machine Intelligence* 6, 7 (2024).
- [120] KANAGAVELU, R.; ANIL, C. G.; WANG, Y.; FU, H.; WEI, Q.; LIU, Y.; MONG GOH, R. S. Fed-SHARC: Resilient Decentralized Federated Learning based on Reward driven Clustering. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (June 2024), pp. 581–586.
- [121] KAPOOR, A.; SINGHAL, A. A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms. In *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)* (Ghaziabad, India, Feb. 2017), IEEE, pp. 1–6.
- [122] KIM, J. W.; CHO, S.-H. Enhancing differentially private distributed k-means clustering through data-driven sensitivity calibration and optimized dropout handling. *Information Sciences* 744 (2026), 123362.
- [123] KITCHENHAM, B.; CHARTERS, S.; OTHERS. Guidelines for performing systematic literature reviews in software engineering, 2007.
- [124] KONDOR, G. NP-hardness of m-dimensional weighted matching problems. *Theoretical Computer Science* 930 (Sept. 2022), 33–36.

- [125] LEE, J.-S.; JUN, S.-P. Privacy-preserving data mining for open government data from heterogeneous sources. *Government Information Quarterly* 38, 1 (2021), 101544.
- [126] LEMIEUX, V. L.; WERNER, J. Protecting Privacy in Digital Records: The Potential of Privacy-Enhancing Technologies. *Journal on Computing and Cultural Heritage* 16, 4 (Dec. 2023), 1–18.
- [127] LI, M.; ZHOU, Y.; TANG, W.; LU, L. K-modes Based Categorical Data Clustering Algorithms Satisfying Differential Privacy. In *2020 International Conference on Networking and Network Applications (NaNA)* (Dec. 2020), pp. 86–91.
- [128] LI, Q. Z. Z. C. Density clustering-based optimization model for trajectory data publication. *The Journal of Supercomputing* 81, 1 (2024).
- [129] LI, Y.; SONG, X.; TU, Y.; LIU, M. GAPBAS: Genetic algorithm-based privacy budget allocation strategy in differential privacy K-means clustering algorithm. *Computers & Security* 139 (Apr. 2024), 103697.
- [130] LIU, J.; TALWAR, K. Private selection from private candidates. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (New York, NY, USA, June 2019), STOC 2019, Association for Computing Machinery, pp. 298–309.
- [131] LLOYD, S. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (Mar. 1982), 129–137.
- [132] LONDE, M. A.; PESSOA, L. S.; ANDRADE, C. E.; GONÇALVES, J. F.; RESENDE, M. G. C. Random-key genetic algorithms: Principles and applications, 2025.
- [133] LONDE, M. A.; PESSOA, L. S.; ANDRADE, C. E.; RESENDE, M. G. Biased random-key genetic algorithms: A review. *European Journal of Operational Research* 321, 1 (Feb. 2025), 1–22.
- [134] LOURENÇO, H. R.; MARTIN, O. C.; STÜTZLE, T. Iterated Local Search. In *Handbook of Metaheuristics*, F. Glover and G. A. Kochenberger, Eds., vol. 57. Kluwer Academic Publishers, Boston, 2003, pp. 320–353.
- [135] LU, Z.; SHEN, H. Differentially Private kk-Means Clustering With Convergence Guarantee. *IEEE Transactions on Dependable and Secure Computing* 18, 4 (July 2021), 1541–1552.
- [136] MACHANAVAJJHALA, A.; KIFER, D.; GEHRKE, J.; VENKITASUBRAMANIAM, M. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data* 1, 1 (Mar. 2007), 3–es.
- [137] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium On Mathematical Statistics and Probabilities* (1967).
- [138] MACWAN, K. R.; PATEL, S. J. Node Differential Privacy in Social Graph Degree Publishing. *Procedia Computer Science* 143 (Jan. 2018), 786–793.

- [139] MCMAHAN, H. B.; MOORE, E.; RAMAGE, D.; HAMPSON, S.; ARCAS, B. A. Y. Communication-Efficient Learning of Deep Networks from Decentralized Data, 2016.
- [140] MCSHERRY, F.; TALWAR, K. Mechanism Design via Differential Privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (Oct. 2007), pp. 94–103. ISSN: 0272-5428.
- [141] MEISER, S. Approximate and Probabilistic Differential Privacy Definitions, 2018. Publication info: Preprint. MINOR revision.
- [142] MIN, Z.; KAI-FEI, D. Improved Research to K-means Initial Cluster Centers. In *2015 Ninth International Conference on Frontier of Computer Science and Technology* (Dalian, China, Aug. 2015), IEEE, pp. 349–353.
- [143] MIRONOV, I. Rényi Differential Privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)* (Aug. 2017), pp. 263–275. ISSN: 2374-8303.
- [144] MIRONOV, I.; TALWAR, K.; ZHANG, L. Rényi Differential Privacy of the Sampled Gaussian Mechanism, Aug. 2019. arXiv:1908.10530 [cs, stat].
- [145] MURTAZA, H.; AHMED, M.; KHAN, N. F.; MURTAZA, G.; ZAFAR, S.; BANO, A. Synthetic data generation: State of the art in health care domain. *Computer Science Review* 48 (May 2023), 100546.
- [146] MUTHUKRISHNAN, G.; KALYANI, S. Grafting Laplace and Gaussian Distributions: A New Noise Mechanism for Differential Privacy. *IEEE Transactions on Information Forensics and Security* 18 (2023), 5359–5374.
- [147] NALDI, M. C. *Técnicas de combinação para agrupamento centralizado e distribuído de dados*. text, Universidade de São Paulo, Jan. 2011.
- [148] NANAYAKKARA, P.; GHAZI, E.; VADHAN, S. Practitioners' Perspectives on a Differential Privacy Deployment Registry, 2025.
- [149] NARAYANAN, A.; SHMATIKOV, V. How To Break Anonymity of the Netflix Prize Dataset, 2006.
- [150] NATIONS, U. *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics*. United Nations Committee of Experts on Big Data and Data Science for Official Statistics, New York, 2023.
- [151] NELDER, J. A.; MEAD, R. A Simplex Method for Function Minimization. *The Computer Journal* 7, 4 (Jan. 1965), 308–313.
- [152] NGUYEN, H. H. Privacy-preserving mechanisms for k-modes clustering. *Computers & Security* 78 (2018), 60–75.
- [153] NI, L.; LI, C.; LIU, H.; BOURGEOIS, A. G.; YU, J. Differential Private Preservation Multi-core DBScan Clustering for Network User Data. *Procedia Computer Science* 129 (2018), 257–262.

- [154] NI, L.; LI, C.; WANG, X.; JIANG, H.; YU, J. DP-MCDBSCAN: Differential Privacy Preserving Multi-Core DBSCAN Clustering for Network User Data. *IEEE Access* 6 (2018), 21053–21063.
- [155] NI, T.; QIAO, M.; CHEN, Z.; ZHANG, S.; ZHONG, H. Utility-efficient differentially private K-means clustering based on cluster merging. *Neurocomputing* 424 (Feb. 2021), 205–214.
- [156] NÓBREGA, T.; PIRES, C. E. S.; NASCIMENTO, D. C.; MARINHO, L. B. Towards automatic Privacy-Preserving Record Linkage: A Transfer Learning based classification step. *Data & Knowledge Engineering* 145 (2023), 102180.
- [157] OSORIO-MARULANDA, P. A.; EPELDE, G.; HERNANDEZ, M.; ISASA, I.; REYES, N. M.; IRAOLA, A. B. Privacy Mechanisms and Evaluation Metrics for Synthetic Data Generation: A Systematic Review. *IEEE Access* 12 (2024), 88048–88074.
- [158] PAN, K.; GONG, M.; GAO, Y. Privacy-enhanced generative adversarial network with adaptive noise allocation. *Knowledge-Based Systems* 272 (July 2023), 110576.
- [159] PAN, K.; ONG, Y.-S.; GONG, M.; LI, H.; QIN, A. K.; GAO, Y. Differential privacy in deep learning: A literature survey. *Neurocomputing* 589 (July 2024), 127663.
- [160] PATIL, S.; PARMAR, K. Differential Privacy Mechanisms: A State-of-the-Art Survey. In *Futuristic Trends in Networks and Computing Technologies* (Singapore, 2022), P. K. Singh, S. T. Wierzchoń, J. K. Chhabra, and S. Tanwar, Eds., Springer Nature, pp. 1049–1060.
- [161] PRAMUDYA, R. I.; KURNIAWAN, T. A.; CANDRA, M. H.; ONN, C. W.; DISSANAYAKE, K. K. Advancing unsupervised clustering: A systematic review of hybrid K-means and metaheuristic optimization algorithms in data mining. *Computer Science Review* 61 (2026), 100972.
- [162] PUTERMAN, M. L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1 ed. Wiley Series in Probability and Statistics. Wiley, Apr. 1994.
- [163] REN, J.; XIONG, J.; YAO, Z.; MA, R.; LIN, M. DPLK-Means: A Novel Differential Privacy K-Means Mechanism. In *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)* (2017), pp. 133–139.
- [164] RICCIATO, F.; RICCIATO, F.; BUJNOWSKA, A.; WIRTHMANN, A.; HAHN, M.; BARREDO-CAPELOT, E. A reflection on privacy and data confidentiality in Official Statistics. In *Proceeding of the 62nd ISI World Statistics Congress 2019* (Feb. 2020).
- [165] RODRIGUEZ, A.; LAIO, A. Clustering by fast search and find of density peaks. *Science* 344, 6191 (June 2014), 1492–1496.
- [166] SCHUBERT, E. Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter* 25, 1 (June 2023), 36–42.

- [167] SEMAAN, G. S.; FADEL, A. C.; BRITO, J. A. D. M.; OCHI, L. S. A Hybrid Heuristic with Hopkins Statistic for the Automatic Clustering Problem. *IEEE Latin America Transactions* 17, 01 (Jan. 2019), 7–17.
- [168] SHAMIR, A. How to share a secret. *Commun. ACM* 22, 11 (Nov. 1979), 612–613.
- [169] SHAPCOTT, Z. An Investigation into Distance Measures in Cluster Analysis, 2024.
- [170] SHI, C.; HUANG, X.; YU, P. On the Security of Distributed Multi-Agent K-Means Clustering With Local Differential Privacy. *IEEE Access* 12 (2024), 124751–124763.
- [171] SHKABERINA, G.; REZOVA, N.; TOVBIS, E.; KAZAKOVTSSEV, L. Visual Assessment of Cluster Tendency with Variations of Distance Measures. *Algorithms* 16 (Dec. 2022), 5.
- [172] SINGEL, R. NetFlix Cancels Recommendation Contest After Privacy Lawsuit. *Wired* (Mar. 2010).
- [173] SMITH, A.; ULLMAN, J. Privacy in Statistics and Machine Learning Spring 2023 - Lecture 10: Advanced Composition, 2023.
- [174] SOLVER, R. K. O. RKO-solver/RKO_python, Mar. 2026. original-date: 2025-09-08T16:06:53Z.
- [175] SPEARS, W. M.; JONG, K. A. D. On the Virtues of Parameterised Uniform Crossover. In *Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991* (1991), R. K. Belew and L. B. Booker, Eds., Morgan Kaufmann, pp. 230–236.
- [176] STATE, F. D. S. Privacy-Preserving Behavioral Anomaly Detection in Dynamic Graphs for Card Transactions. In *Web Information Systems Engineering – WISE 2024* (2025).
- [177] STEMMER, U.; KAPLAN, H. Differentially Private k-Means with Constant Multiplicative Error. In *Advances in Neural Information Processing Systems* (2018), S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, Curran Associates, Inc.
- [178] STÜTZLE, T.; RUIZ, R. Iterated Local Search. In *Handbook of Heuristics*, R. Martí, P. M. Pardalos, and M. G. C. Resende, Eds. Springer International Publishing, Cham, 2018, pp. 579–605.
- [179] SU, D.; CAO, J.; LI, N.; BERTINO, E.; JIN, H. Differentially Private K-Means Clustering. In *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy* (New York, NY, USA, Mar. 2016), CODASPY '16, Association for Computing Machinery, pp. 26–37.
- [180] SU, D.; CAO, J.; LI, N.; BERTINO, E.; LYU, M.; JIN, H. Differentially Private K-Means Clustering and a Hybrid Approach to Private Optimization. *ACM Trans. Priv. Secur.* 20, 4 (Oct. 2017), 16:1–16:33.

- [181] SWANBERG, M.; DESFONTAINES, D.; HANEY, S. DP-SIPS: A simpler, more scalable mechanism for differentially private partition selection. *Proceedings on Privacy Enhancing Technologies* (2023).
- [182] SWEENEY, L. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 10, 5 (Oct. 2002), 557–570.
- [183] SZWARCFITER, J. *Grafos e algoritmos computacionais*. Campus, 1986.
- [184] TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining: Pearson New International Edition*. Pearson Education, Aug. 2013.
- [185] TAYEBI ARASTEH, S.; ZILLER, A.; KUHL, C.; MAKOWSKI, M.; NEBELUNG, S.; BRAREN, R.; RUECKERT, D.; TRUHN, D.; KAISSIS, G. Preserving fairness and diagnostic accuracy in private large-scale AI models for medical imaging. *Communications Medicine* 4, 1 (Mar. 2024), 1–12. Publisher: Nature Publishing Group.
- [186] TODESCHINI, R.; BALLABIO, D.; TERMOPOLI, V.; CONSONNI, V. Extended multivariate comparison of 68 cluster validity indices. A review. *Chemometrics and Intelligent Laboratory Systems* 251 (Aug. 2024), 105117.
- [187] UDDIN, M. A.; AHSAN, M. N.; DAS, M. A Comparative Study on Various ML Models Using Synthetic Data for Privacy Preservation. In *2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC)* (June 2024), pp. 1–6.
- [188] USTUNDAG SOYKAN, E.; KARAÇAY, L.; KARAKOÇ, F.; TOMUR, E. A Survey and Guideline on Privacy Enhancing Technologies for Collaborative Machine Learning. *IEEE Access* 10 (2022), 97495–97519. Conference Name: IEEE Access.
- [189] VIEIRA FILHO, J. E.; AZEREDO, P. Tecnologia, criptografia e matemática: da troca de mensagens ao suporte em transações econômicas. *Revista Desenvolvimento Socioeconômico em Debate* 2 (Dec. 2016), 22–31.
- [190] VU, D.-H. Privacy-preserving Naive Bayes classification in semi-fully distributed data model. *Computers & Security* 115 (2022), 102630.
- [191] WAAL, A.; WILLENBORG, L. Global recodings and local suppressions in microdata sets. In *Symposium 95: from data to information: methods and systems* (Ottawa, 1995), Statistics Canada, pp. 121–132.
- [192] WANG, X.; LIU, X.; XU, W. Personalized Differential Privacy for Support Vector Machines. *Journal of Systems Science and Complexity* 39, 1 (Feb. 2026), 180–202.
- [193] WANG, Y.; PANG, W.; PEDRYCZ, W. One-Shot Federated Clustering Based on Stable Distance Relationships. *IEEE Transactions on Industrial Informatics* 20, 11 (Nov. 2024), 13262–13272.
- [194] WANG, Y.; WANG, Q.; ZHAO, L.; WANG, C. Differential privacy in deep learning: Privacy and beyond. *Future Generation Computer Systems* 148 (Nov. 2023), 408–424.

- [195] WARNER, S. L. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60, 309 (Mar. 1965), 63–69. Publisher: Taylor & Francis _eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1965.10480775>.
- [196] WATKINS, C. J.; DAYAN, P. Technical Note: Q-Learning. *Machine Learning* 8, 3 (May 1992), 279–292.
- [197] WILCOXON, F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1, 6 (Dec. 1945), 80.
- [198] WILLENBORG, L.; DE WAAL, T. Application of Non-Perturbative SDC Techniques for Microdata. In *Elements of Statistical Disclosure Control*, L. Willenborg and T. de Waal, Eds. Springer, New York, NY, 2001, pp. 93–106.
- [199] WIROONSRI, N. Clustering performance analysis using a new correlation-based cluster validity index. *Pattern Recognition* 145 (Jan. 2024), 109910.
- [200] WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WES-SLÉN, A. *Experimentation in Software Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [201] WOLBERG, W.; MANGASARIAN, O.; STREET, N.; STREET, W. Breast Cancer Wisconsin (Diagnostic), 1993. Published: UCI Machine Learning Repository.
- [202] WOOD, A.; ALTMAN, M.; BEMBENEK, A.; BUN, M.; GABOARDI, M.; HONAKER, J.; NISSIM, K.; OBRIEN, D. R.; STEINKE, T.; VADHAN, S. Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment & Technology Law* 21, 1 (2018), 209–275.
- [203] WU, F.; DU, M.; ZHI, Q. Density-based clustering with differential privacy. *Information Sciences* 681 (2024), 121211.
- [204] WU, W.; HUANG, H. A DP-DBScan clustering algorithm based on differential privacy preserving. *Computer Engineering & Science* 37, 04 (2015), 830–834.
- [205] XIA, C.; HUA, J.; TONG, W.; ZHONG, S. Distributed K-Means clustering guaranteeing local differential privacy. *Computers & Security* 90 (2020), 101699.
- [206] XING, K.; HU, C.; YU, J.; CHENG, X.; ZHANG, F. Mutual Privacy Preserving k-Means Clustering in Social Participatory Sensing. *IEEE Transactions on Industrial Informatics* 13, 4 (Aug. 2017), 2066–2076.
- [207] YANG, M.; GUO, T.; ZHU, T.; TJUAWINATA, I.; ZHAO, J.; LAM, K.-Y. Local differential privacy and its applications: A comprehensive survey. *Computer Standards & Interfaces* 89 (Apr. 2024), 103827.
- [208] YANG, M.; TJUAWINATA, I.; LAM, K.-Y. K-Means Clustering With Local dx-Privacy for Privacy-Preserving Data Analysis. *IEEE Transactions on Information Forensics and Security* 17 (2022), 2524–2537.
- [209] YANG, Q.; LIU, Y.; CHEN, T.; TONG, Y. Federated Machine Learning: Concept and Applications. *ACM Trans. Intell. Syst. Technol.* 10 (Jan. 2019), 12:1–12:19.

- [210] YE, J. Z.-Y. Z.-L. S.-C. Entropy-driven differential privacy protection scheme based on social graphlet attributes. *The Journal of Supercomputing* 80, 6 (2024).
- [211] YE, X.; ZHU, Y.; ZHANG, S.; DENG, H.; YU, P. Non-interactive K-mode clustering of high-dimensional categorical data under local differential privacy. *Information Sciences* 718 (2025), 122417.
- [212] YEH, I.-C. Blood Transfusion Service Center, 2008. Published: UCI Machine Learning Repository.
- [213] YU, J. Research on Data Mining Technology of Internet Privacy Protection Based on Local Differential Privacy. In *2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)* (May 2024), pp. 486–491.
- [214] YUAN, F.; XU, D.; DU, D.; LI, M. Differentially private k-center problems. *Optimization Letters* 18, 8 (Nov. 2024), 1791–1809.
- [215] ZHANG, E.; LI, H.; HUANG, Y.; HONG, S.; ZHAO, L.; JI, C. Practical multi-party private collaborative k-means clustering. *Neurocomputing* 467 (2022), 256–265.
- [216] ZHANG, H. Research on k-Means Algorithm Based on Differential Privacy. In *2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA)* (Aug. 2023), pp. 487–492.
- [217] ZHANG, J.; XIAO, X.; YANG, Y.; ZHANG, Z.; WINSLETT, M. PrivGene: differentially private model fitting using genetic algorithms. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, June 2013), SIGMOD '13, Association for Computing Machinery, pp. 665–676.
- [218] ZHANG, T.; RAMAKRISHNAN, R.; LIVNY, M. BIRCH: an efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 1996), SIGMOD '96, Association for Computing Machinery, pp. 103–114. event-place: Montreal, Quebec, Canada.
- [219] ZHANG, W.; ZUO, X.; WANG, Y.; ZHANG, L. Adaptive density clustering with differential privacy. *Pattern Recognition* 179 (2026), 113652.
- [220] ZHANG, Y.; LI, S. Privacy Preserving BIRCH Algorithm under Differential Privacy. In *2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)* (Oct. 2017), pp. 48–53.
- [221] ZHONG, T. H. T. N. Practical Privacy-Preserving Community Detection in Decentralized Weighted Networks. In *Security and Privacy in Communication Networks* (2025).
- [222] ZHOU, I.; TOFIGH, F.; PICCARDI, M.; ABOLHASAN, M.; FRANKLIN, D.; LIPMAN, J. Secure Multi-Party Computation for Machine Learning: A Survey. *IEEE Access* 12 (2024), 53881–53899.

- [223] ZHU, L.; SONG, H.; CHEN, X. Dynamic privacy budget allocation for enhanced differential privacy in federated learning. *Cluster Computing* 28, 15 (Dec. 2025), 999.
- [224] ZIGOMITROS, A.; CASINO, F.; SOLANAS, A.; PATSAKIS, C. A Survey on Privacy Properties for Data Publishing of Relational Data. *IEEE Access* (2020).
- [225] ÁGOSTON, K. C.; E.-NAGY, M. Mixed integer linear programming formulation for K-means clustering problem. *Central European Journal of Operations Research* 32, 1 (Mar. 2024), 11–27.
- [226] ÁLVAREZ, I. A.; EHAUS, M.; FRANK, M.-L.; SEDLMEIR, J. Privacy-Enhancing Technologies. In *Decentralization Technologies*, G. Fridgen, T. Guggenberger, J. Sedlmeir, and N. Urbach, Eds. Springer Nature Switzerland, Cham, 2024, pp. 97–119.

Glossário

analista O analista é o agente interessado em extrair informações agregadas do conjunto de dados, submetendo consultas e recebendo respostas perturbadas pelo mecanismo de PD, sem nunca ter acesso diretamente registros individualizados.. [15](#)

conjunto de microdados Coleção estruturada de registros individuais contendo atributos sobre unidades de análise (e.g., indivíduos, empresas, domicílios, eventos etc.), onde cada unidade é representada em uma linha e cada atributo é representado em uma coluna. É um conjunto de tuplas. [14](#), [15](#), [22](#), [24](#), [28–32](#), [35](#), [136](#), [141](#), [145](#), [147](#), [148](#)

consulta Uma consulta é uma função ou conjunto de funções, i.e., um algoritmo determinístico, que tem como saída um agregado estatístico sobre características de interesse da população representada no conjunto de microdados. [18](#), [20](#), [148](#), [149](#)

curador O curador é a entidade com acesso completo ao conjunto de microdados sensíveis e que tem interesse em permitir a extração de informações a partir dele, zelando, porém, pela confidencialidade ou privacidade. [15](#)

grupo Conjunto de objetos (unidades de análise) em um espaço de dados que são similares entre si de acordo com algum critério estabelecido previamente, como métricas de distância (e.g., distância euclidiana). [8](#)

multiset Um *multiset* é uma generalização do conceito de conjunto em que elementos podem aparecer múltiplas vezes. Pode ser definido formalmente como uma função $m : X \rightarrow \mathbb{N}$, onde $m(x)$ indica a multiplicidade (número de ocorrências) do elemento $x \in X$. [26](#)

risco de revelação O risco de revelação refere-se à possibilidade de que informações sensíveis sobre indivíduos específicos sejam inferidas a partir de respostas (saídas) fornecidas às consultas, sendo formalmente limitado pelo parâmetro ϵ no contexto de privacidade diferencial, i.e., o mesmo que perda de privacidade (do inglês, *privacy loss*). [15](#), [140](#), [142](#)

sensibilidade No contexto de PD, a sensibilidade é a máxima variação possível em sua saída quando aplicada a dois conjuntos de dados adjacentes, isto é, que diferem em apenas um único registro. [21](#)

tabela Estrutura de dados representada por um conjunto de linhas e colunas. Pode ser utilizada para apresentar conjuntos de microdados (tuplas) ou agregados estatísticos gerados a partir de consultas a conjuntos de microdados. No segundo caso são também referidas como tabulações. [141](#)

utilidade A utilidade diz respeito ao grau em que os resultados obtidos preservam informações relevantes do conjunto de dados original, refletindo a qualidade ou precisão desses resultados, mesmo após a introdução do ruído. [15](#), [140](#)

APÊNDICE A – Tecnologias de Aprimoramento da Privacidade

As Tecnologias de Aprimoramento da Privacidade (do inglês, *Privacy Enhancing Technologies* - PET) [150, 71] compreendem um amplo conjunto de abordagens voltadas para a proteção da privacidade de dados pessoais, geralmente desenvolvidas por profissionais das áreas de ciência da computação, matemática, estatística ou engenharia [3, p.21]. Adicionalmente, somada a essas abordagens, a elaboração de leis e políticas com o mesmo objetivo são de relevância primordial para oferecer proteção da privacidade de dados pessoais na sociedade cada vez mais hiperconectada. Essa vertente está, no entanto, fora do escopo deste trabalho, cujo foco se dá nas áreas de conhecimento que constituem o subconjunto tecnológico daquelas envolvidas com o tema.

Mais especificamente, faz-se necessário definir três conceitos basilares, quais sejam: **confidencialidade, privacidade e sigilo**. A confidencialidade está associada a impedir a identificação do indivíduo – ou unidade de análise em questão –, mas permitir o acesso a informações, inclusive sensíveis, para fins estatísticos. Isto é, o analista tem acesso aos atributos dos indivíduos, porém sem saber quem esses indivíduos são. A garantia da privacidade consiste em, no contexto em que os indivíduos são ou podem ser identificados, impedir que suas informações individuais sensíveis sejam expostas. Finalmente, o conceito de sigilo está associado com a obrigação legal de não divulgar ou permitir o acesso a informações pessoais (e.g., sigilo bancário, telemático etc.), geralmente reservado ao âmbito legal ou de segurança da informação.

Cabe destacar que a exposição acerca das PET apresentada nesta tese pressupõe que os recursos típicos de segurança da informação (e.g., controle de acesso, segurança redes etc.) estão garantidos e, portanto, não serão objeto de discussão.

As PET podem ser incorporadas a estratégias de análise ou divulgação de dados e são

tipicamente classificadas em duas categorias [150, p.20], quais sejam: *output privacy* e *input privacy*, apresentadas no Apêndice A (p.135). A Figura A.1 apresenta uma ilustração da diferença essencial entre as duas categorias. As tecnologias categorizadas como *output privacy* são aplicadas no contexto em que o analista tem acesso a uma versão protegida do [conjunto de microdados](#), na qual o aspecto da confidencialidade foi tratado.

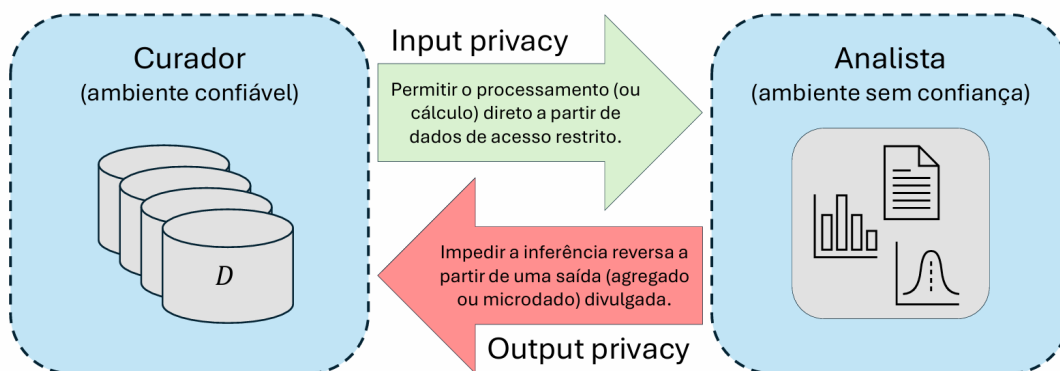


Figura A.1: Diferenças entre os problemas de *input* e *output privacy* – adaptado de [164].

O analista pode então utilizá-la livremente para extrair informações e o problema de *output privacy* é evitar que, nesse processo, ele seja capaz de fazer algum tipo de inferência reversa e recuperar informações consideradas sensíveis. Ou seja, no problema de *output privacy*, o analista tem acesso a todo o [conjunto de microdados](#) (ainda que modificado) e o objetivo é impedir que ele seja capaz de extrair ou recuperar informações indesejadas do ponto de vista da confidencialidade. A natureza do problema de *input privacy* é, até certo ponto, antagônica. Nesse caso, o analista é impedido de acessar o [conjunto de microdados](#), mas, através das tecnologias de *input privacy*, deseja-se habilitá-lo a executar determinado cálculo ou consulta, oferecendo assim a possibilidade de extrair informações do [conjunto de microdados](#) de maneira controlada.

As seções A.1 e A.2 não têm por objetivo fazer um levantamento exaustivo das PET, mas dar uma visão consubstanciada das principais abordagens em cada categoria. Ainda neste sentido, estão disponíveis na literatura levantamentos abrangentes concernentes a tal tema [111, 224, 106, 3, 150, 126, 226].

A.1 *Input Privacy*

As abordagens classificadas como *input privacy* são aquelas que visam permitir que duas ou mais partes participem ou interajam no processo de extração de informação, porém sem que haja compartilhamento de microdados entre elas [150, p.20]. São tipicamente

baseadas em arquiteturas tecnológicas específicas, ao passo que as abordagens classificadas como *output privacy* são tipicamente baseadas em algoritmos. Essa característica faz com que soluções baseadas em *input privacy* demandem mais recursos tecnológicos, financeiros e de pessoal. Além disso, podem trazer maiores dificuldades no que se refere à adaptação do usuário final, em função de necessidades de capacitação.

Algumas das abordagens mais comuns são brevemente descritas nas sessões a seguir, envolvendo partes que geralmente desempenham até três diferentes papéis, quais sejam: (i) parte proprietária do dado (*input party* - IP); (ii) parte que executa o processamento (*computing party* - CP); e (iii) parte que tem acesso ao resultado do processamento (*resulting party* - RP).

A.1.1 Computação segura entre múltiplas partes

A Computação Segura entre Múltiplas Partes (do inglês, *Secure Multi-Party Computation* - sMPC) [88] resolve essencialmente o problema de processar dados provenientes de diferentes fontes sem que nenhuma das partes envolvidas tenha acesso ao dado de entrada completo.

É comum adotar abordagens baseadas em compartilhamento linear de segredo (do inglês, *linear secret sharing*) [150, p.28]. Informalmente, um esquema de compartilhamento de segredo (do inglês, *secret sharing scheme* - SSS) permite compartilhar uma informação entre n participantes de tal forma que apenas um grupo predeterminado de participantes pode recuperar a informação completa, enquanto os demais participantes não podem obter nenhuma informação adicional além daquela informação parcial que lhes foi concedida [21]. Como exemplo, uma estratégia simples e efetiva pode ser baseada no esquema de Shamir [168].

Essa abordagem impõe um *overhead* intrínseco ao processamento distribuído, mas é viável em aplicações reais. A primeira aplicação prática e em larga escala de sMPC foi documentada em 2009 [25]. Há também diferentes cenários que podem permitir ganhos de eficiência. A aplicação de sMPC considerada padrão [150, p.29] consiste em *IP* exercendo também o papel de *CP*. Eventualmente podendo atuar ainda como *RP*. Nesse cenário a sMPC é aplicada apenas para evitar o compartilhamento efetivo de microdados.

Na área de aprendizado de máquina, devido à necessidade de alguns modelos de consumir grandes volumes de dados, o que pode implicar na dependência de envolver múltiplas entidades produtoras ou detentoras de dados. Nesse contexto, abordagens baseadas em

sMPC vêm sendo adotadas com sucesso [222].

A.1.2 Criptografia completamente homomórfica

Abordagens baseadas em sMPC e, portanto, em algum SSS, exigem ao menos duas *CP*. A Criptografia Completamente Homomórfica (do inglês, *Fully Homomorphic Encryption* - FHE) [86] é geralmente adotada quando existe apenas uma *CP* distinta da *IP* e não há confiança mútua.

A FHE faz uso de esquemas de criptografia que preservam propriedades aritméticas do dado original não criptografado (*plaintext*), permitindo que determinadas operações sejam realizadas nos dados criptografados. Em outras palavras, ela permite o processamento de dados enquanto eles permanecem criptografados, garantindo privacidade e segurança durante todo o processamento. Com a FHE, uma terceira parte (por exemplo, um provedor de serviço em nuvem) pode executar tarefas como operações matemáticas ou análise de dados sem nunca ter acesso aos dados originais. Uma vez que os cálculos são feitos, o resultado (e.g., totais, médias etc.) pode ser descriptografado apenas pelo proprietário dos dados (e.g., *IP*). Essa tecnologia tem implicações significativas para computação segura em nuvem, análise de dados sensíveis e vários campos onde dados confidenciais precisem ser manipulados com segurança. No entanto, a FHE é computacionalmente custosa, característica que tem motivado o desenvolvimento de pesquisas em busca de métodos mais eficientes.

A.1.3 Ambientes de execução confiáveis

Um ambiente de execução confiável (do inglês, *Trusted Execution Environment* - TEE) [150, p.51-52] é uma parte segura do processador de um computador que garante que dados e códigos confidenciais sejam protegidos durante o processamento. Consiste em um espaço isolado fisicamente no hardware, onde informações confidenciais podem ser manipuladas sem interferência do resto do sistema. Isso significa que, mesmo que o sistema operacional principal ou outros aplicativos sejam comprometidos, os dados confidenciais dentro do TEE permanecem seguros e inacessíveis.

Um dos principais recursos de um TEE é sua capacidade de garantir a integridade e o sigilo dos dados e códigos dentro dele. Somente aplicativos ou processos confiáveis podem ser executados dentro desse ambiente, e eles não podem ser modificados por programas

externos, sejam eles legítimos¹ ou não. Isso torna os TEE especialmente úteis em situações em que a segurança é crítica, como o processamento de transações financeiras, a proteção de registros médicos pessoais ou a segurança de canais de comunicação.

Exemplos de TEE incluem SGX (Software Guard Extensions) da Intel e TrustZone da ARM, que são amplamente usados em computação em nuvem e dispositivos móveis. Ao usar TEE, as organizações podem garantir que, mesmo em ambientes onde a segurança geral é questionável, os dados (ou algoritmos/códigos) mais sensíveis permaneçam seguros.

A.1.4 Aprendizagem distribuída

O aprendizado distribuído, particularmente o aprendizado federado (do inglês, *Federated Learning* - FL) [118, 209], é um método em que vários dispositivos ou servidores distribuídos são utilizados de forma combinada para treinar um modelo de aprendizado de máquina compartilhado sem compartilhar os microdados. Nessa abordagem, cada dispositivo (cliente ou nó) treina um modelo local em seus próprios microdados e compartilha apenas os parâmetros do modelo com um servidor central. O servidor então agrega essas informações para compor um modelo global, sem nunca acessar os microdados de nenhuma das partes. Essa técnica é especialmente útil em situações em que os dados são sensíveis ou regidos por regulamentações de privacidade rígidas, como em saúde, finanças ou dispositivos pessoais como *smartphones*. O FL ajuda a proteger a privacidade individual, mantendo os microdados armazenados localmente e cumprindo regulamentações legais.

Embora o FL aprimore a privacidade, também apresenta alguns desafios. A comunicação frequente entre dispositivos e o servidor pode introduzir altos custos em termos de tempo e consumo de bateria e banda de internet, sendo esses dois últimos particularmente relevantes em aplicações de internet das coisas (do inglês *Internet of Things* - IoT). Além disso, mesmo que os dados não sejam compartilhados diretamente, a privacidade ainda pode ser comprometida por ataques específicos. Portanto, camadas adicionais de privacidade, adotando técnicas como Privacidade Diferencial (Seção A.2.8), podem ser empregadas para mitigar esse risco [223]. O FL é altamente escalável, mas coordenar um grande número de dispositivos, cada um com sua própria distribuição de dados e capacidades computacionais exclusivas, é complexo. Apesar desses desafios, há grande potencial para aplicações onde tanto a privacidade quanto o uso de grandes conjuntos de

¹Programas legítimos externos ao TEE correspondem, por exemplo, ao sistema operacional ou aplicações autorizadas executadas fora do ambiente confiável, que, embora não sejam necessariamente maliciosos, também não possuem permissão para acessar ou modificar diretamente os dados e códigos protegidos.

dados descentralizados são críticos.

A.2 *Output Privacy*

As abordagens classificadas como *output privacy* são aquelas aplicadas à saída final, imediatamente antes da disponibilização ao usuário ou analista. Se destinam a evitar algum tipo de inferência² reversa [164, p.3], i.e., visa ocultar dados individuais sensíveis de serem identificados ou reidentificados da saída ou informação disseminada [150, p.20].

O conceito de *output privacy* é familiar à maioria dos Institutos Nacionais de Estatística (INE) e é geralmente conhecido como Controle Estatístico de Confidencialidade (do inglês, *Statistical Disclosure Control* - SDC) [150, p.20], noção que foi formalizada no final da década de 1970 [44, p.4]. No contexto da produção de estatísticas oficiais, a revolução tecnológica das últimas décadas vem permitindo a publicação de dados cada vez mais detalhados e a coleta de microdados extensivos de registros externos. Embora isso tenha melhorado significativamente a capacidade dos INE de oferecer informações de maior granularidade e tempestividade, auxiliando na formulação de políticas públicas e na tomada de decisão, também levanta preocupações no que se refere ao risco de violação da privacidade dos indivíduos e organizações representados nos dados coletados. Com isso, os INE precisam encontrar o equilíbrio entre os compromissos de oferecer o máximo de informação à sociedade e salvaguardar a privacidade desses indivíduos [111, p.xi].

Os métodos de SDC têm por objetivo minimizar o risco de revelar informações sobre as unidades de análise de uma população, sejam indivíduos, empresas, transações etc., ao mesmo tempo em que maximizam a quantidade de informações que podem ser divulgadas [111, p.3], i.e, minimizam a perda de informação. Em suma, tais métodos buscam tratar o *trade-off* entre o [risco de revelação](#) ou perda de privacidade e a perda de informação ou [utilidade](#).

As abordagens nas quais os métodos de SDC são baseados podem ser classificadas em duas categorias [111, p.3], são elas: (i) perturbativos, que engloba métodos que introduzem um ruído controlado no dado, perturbando a informação original; e (ii) não perturbativos, referente a métodos que reduzem a quantidade de informação através de agregação ou supressão dos dados, por exemplo.

²Neste contexto, inferência é a tarefa de extrair informações, mesmo que parciais, sobre um dos componentes de entrada ou registros com base no conhecimento da saída (e possivelmente outros dados externos) [164, p.3].

Em função da relevância do tema, muito esforço foi investido nas últimas duas décadas em propor métodos eficazes para proteger a confidencialidade, segundo uma ampla variedade de abordagens. Algumas das mais relevantes e frequentemente aplicadas são, dentre as não-perturbativas, agregação e supressão, e, dentre as perturbativas, geração de microdados sintéticos e privacidade diferencial. Há disponível na literatura diversos levantamentos e comparativos abrangentes no que se refere a métodos de SDC [224].

Além dos tipos de métodos, em SDC ou *output privacy* é importante classificar também o tipo de saída, uma vez que a escolha do método depende do tipo de saída em questão. Os tipos de saída são classificados em duas categorias [111, p.3], quais sejam: (i) [tabelas](#) (ou tabulações³); e (ii) [conjunto de microdados](#).

Tabelas são tipicamente utilizadas para comunicar estatísticas resumo sobre agregados populacionais. São dados estruturados, organizados em linhas e colunas, onde cada célula, composta pelo cruzamento de uma linha e uma coluna, armazena o agregado de um segmento da população de estudo. Esses agregados podem ser de magnitude, em que cada célula representa uma resposta numérica, tais como totais, médias ou proporções de determinada variável ou atributo de interesse, ou de frequência, onde cada célula representa o número de unidades de análise (indivíduos, empresas, etc.) que se enquadram no critério determinado pelas variáveis ou atributos de interesse.

Os microdados são também dados estruturados organizados em linhas e colunas, porém onde cada linha representa um registro de uma unidade de análise (indivíduos, empresas, etc.) e cada coluna representa uma variável ou atributo de interesse.

Em função da importância do problema e dos crescentes desafios associados, uma grande variedade de abordagens foi proposta na literatura nas últimas três décadas [111, 224]. Dada sua complexidade e as diferentes realidades em que se aplica, uma eventual expectativa de método dominante e de uso geral não seria realista. Em vez disso, a estratégia a ser adotada é abordar as vulnerabilidades identificadas como prioritárias, seja pelo risco ou pelo potencial da consequência, através da aplicação do método considerado mais adequado para mitigá-las.

A presente pesquisa considera a aplicação de métodos de agrupamento, o que envolve o uso de microdados. Considerando os riscos associados à sua utilização, as abordagens de *output privacy* mais comuns para lidar com o *trade-off* entre utilidade e privacidade

³No contexto de bancos de dados, uma tabela pode se referir a um conjunto de microdados. O termo tabulação pode, então, ser adotado para estabelecer a diferenciação necessária. Contudo, no contexto de análise de dados, os termos tabela e tabulação são tipicamente utilizados como sinônimos.

em microdados são apresentadas a seguir.

No que se refere a SDC aplicado a tabelas, há extensa literatura de referência abordando o tema. Em [33] é apresentada extensa revisão dos trabalhos seminais sobre métodos de otimização para SDC na divulgação de dados tabulados. O artigo explora ainda pesquisas subsequentes com foco em abordagens práticas e compara suas vantagens e desvantagens com técnicas mais recentes. Embora menos recente, em [111, p.131-207] os autores apresentam os principais conceitos e terminologias dos dados tabulados de forma didática, incluindo tabulações hierárquicas e vinculadas, abordando risco de divulgação, métodos de proteção e perda de informação, além de tratar de aspectos práticos, como softwares e estratégia para ganho de eficiência. Destacam-se ainda outros trabalhos de substancial relevância para o desenvolvimento do tema [1, 35].

Em função da grande variedade de métodos de *output privacy* para microdados disponíveis na literatura, as seções subsequentes descrevem sucintamente aqueles aplicados mais frequentemente no contexto dos produtores ou curadores de dados. A prevalência desses métodos se deve, além de sua eficácia, a fatores relacionados à explicabilidade, i.e., ao potencial de transparência junto aos usuários, e escalabilidade, que viabiliza sua aplicação em grandes bases de dados. Levantamento e comparativos abrangentes estão disponíveis na literatura [224].

No que se refere a microdados, o controle de confidencialidade se dá essencialmente em duas dimensões, a saber, registro e atributo. A confidencialidade de registro está relacionada com o conceito mais trivial de privacidade individual, que é dificultar a reidentificação do indivíduo. O primeiro passo para isso é remover os atributos identificadores (e.g., número do CPF), mas isso não é suficiente, uma vez que indivíduos com características raras na população podem ser reidentificados a partir de outros atributos, como, por exemplo, alguém com idade muito avançada em uma subpopulação específica. Há, ainda, a condição de indivíduos que não sobressaem na população por um único atributo, mas pela combinação de diferentes atributos, como idade, sexo e cor ou raça, por exemplo. Tais atributos, que embora não sejam identificadores diretos, podem permitir a reidentificação dos indivíduos quando combinados, são denominados atributos semi-identificadores.

Com a aplicação de métodos tradicionais de SDC, geralmente não é possível dar garantia absoluta da confidencialidade de registros. Além disso, a fim de manter determinado nível de utilidade, algum nível de [risco de revelação](#) deve ser geralmente admitido. Quando essa condição envolve também a presença de atributos sensíveis, tais como histórico de doenças, uso de drogas, violência, abuso etc., a confidencialidade de atributos

deve também ser considerada, como exemplificado nas seções A.2.3 e A.2.4. A Tabela A.1 traz uma visão geral das abordagens de *output privacy* consideradas nesta seção e suas características principais.

Tabela A.1: Comparação das principais abordagens de *output privacy*.

Método / Critério	Tipo	Atributos	Referência
supressão	não perturbativo	numéricos/categóricos	[111, p.53]
recodificação	não perturbativo	categóricos	[111, p.52]
k-anonimização	não perturbativo	categóricos	[182]
l-diversidade	não perturbativo	categóricos	[136]
PRAM	perturbativo	categóricos	[91]
microagregação	perturbativo	numéricos	[111, p.62]
geração sintética de microdados	perturbativo	numéricos/categóricos	[111, p.78-99][53]
privacidade diferencial	perturbativo	numéricos/categóricos	[59]

Tabela elaborada pelo autor.

A.2.1 Supressão

A supressão é uma abordagem não-perturbativa que consiste em identificar e ocultar informações consideradas de alto risco de reidentificação. Na supressão local [111, p.53], quando identificados indivíduos com combinações raras de atributos semi-identificadores, são suprimidos os valores dos atributos que mais contribuem para a reidentificação. No exemplo da Tabela A.2a, um indivíduo se destaca da população em questão pela sua idade. A supressão da idade desse indivíduo, conforme ilustrado na Tabela A.2b, reduz substancialmente o potencial de um intruso ou usuário mal intencionado reidentificá-lo.

Tabela A.2: Exemplo de supressão (local) - Microdados.

Idade	Sexo	Nível de Escolaridade
21	F	Médio
36	M	Fundamental
25	M	Superior
42	F	Superior
70	M	Superior

Tabela elaborada pelo autor.

(a) Original

Idade	Sexo	Nível de Escolaridade
21	F	Médio
36	M	Fundamental
25	M	Superior
42	F	Superior
-	M	Superior

Tabela elaborada pelo autor.

(b) Resultante

A supressão global se refere à supressão de todos os valores dos atributos semi-identificadores para um registro. O registro é mantido para garantir consistência na contagem do total populacional e em desagregados baseados nos demais atributos.

Naturalmente que, em uma população tão pequena como a do exemplo, a supressão, seja local ou global, não é muito eficaz. Para alguém que conheça todos os indivíduos

ali representados, é trivial saber a quem se refere o registro que teve suprimido o valor do atributo idade. Em populações maiores, no entanto, essa abordagem pode ser eficaz, pois ajuda a eliminar combinações raras de atributos semi-identificadores, aumentando o número de registros em cada grupo formado pela combinação de valores desses atributos [111, p.53] e, assim, reduzindo o risco de reidentificação. Adicionalmente, a depender do contexto, pode ser necessário suprimir informações de um registro adicional, a fim de evitar que seja possível recuperar informações suprimidas por diferença.

A.2.2 Recodificação

A recodificação global ou generalização é uma abordagem não-perturbativa que consiste em recodificar atributos categóricos (ou numéricos discretos) de forma a agregar duas ou mais categorias em uma única categoria mais geral [111, p.52], como apresentado no exemplo da Tabela A.3. Essa abordagem também é referida por agregação, geralmente quando se refere a discretizar/categorizar um atributo numérico contínuo. Nos dados originais (Tabela A.3a), há apenas uma ocorrência nas categorias "médio" e "fundamental" do atributo "nível de escolaridade". No microdado resultante (Tabela A.3b), ambos os registros foram agregados na nova categoria "educação básica".

Tabela A.3: Exemplo de recodificação - Microdados.

Idade	Sexo	Nível de Escolaridade	Idade	Sexo	Nível de Escolaridade
21	F	Médio	21	F	Educação básica
36	M	Fundamental	36	M	Educação básica
25	M	Superior	25	M	Educação superior
42	F	Superior	42	F	Educação superior
70	M	Superior	70	M	Educação superior
Tabela elaborada pelo autor.			Tabela elaborada pelo autor.		
(a) Original			(b) Resultante		

A recodificação pode ser aplicada com sucesso tanto na confidencialidade de registros como na de atributos. No exemplo, a recodificação da variável "nível de escolaridade" não foi muito eficaz no que se refere à confidencialidade de registros. Como os indivíduos têm sexo e idade (bem) diferentes, sua reidentificação permanece viável. Porém, considerando que um indivíduo de 36 anos que possua apenas o ensino fundamental pode ser alvo de preconceito em alguns contextos sociais, o segundo registro do conjunto de dados original (Tabela A.3a) apresentava uma vulnerabilidade no que se refere à confidencialidade de atributo. Nesse contexto, a recodificação foi eficaz ao produzir o conjunto de dados resultante apresentado na Tabela A.3b.

Tanto na supressão (Seção A.2.1) como na recodificação (Seção A.2.2), o objetivo de reduzir ou eliminar as combinações inseguras é minimizar o risco de reidentificação. No entanto, é necessário escolher adequadamente os valores a serem suprimidos ou categorias a serem recodificadas/agregadas, a fim de maximizar a utilidade do conjunto de dados resultante, sendo essa uma aplicação cujo equacionamento remete ao uso de métodos de otimização. Trabalhos recentes apontam questões em aberto nesse sentido [32, 33]. É possível também combinar a recodificação com a supressão [191, 198], a fim de obter melhores soluções em termos de ambos os objetivos.

A.2.3 k-Anonimização

A k-Anonimização⁴ [182] é, essencialmente, um critério para aplicação de recodificação e/ou supressão. É uma abordagem aplicada geralmente para dados categóricos e consiste em garantir que um indivíduo não possa ser distinguido de, pelo menos, $k - 1$ outros indivíduos nos microdados. Mais especificamente, garante que todas as combinações de atributos semi-identificadores tenham frequência igual ou superior a um valor k predefinido. Isto é, que cada registro torne-se indistinguível de pelo menos outros $k - 1$ registros com relação a esses atributos. Com isso, na hipótese de uma tentativa de pareamento⁵ com outro conjunto de microdados, cada par de registros terá probabilidade máxima de $1/k$ de ser o par correto e, portanto, quanto maior o valor de k menor é a confiança do invasor no sucesso do resultado da tentativa de pareamento.

No exemplo da Tabela A.3, mesmo considerando apenas os atributos sexo e nível de escolaridade, os microdados originais (Tabela A.3a) são 1-anonimizados, uma vez que há registros únicos ($k = 1$) em termos das combinações dos valores desses atributos. Nos microdados resultantes (Tabela A.3b), a condição se mantém, uma vez que há apenas 2 indivíduos com educação básica e eles não têm o mesmo sexo. Isso não é algo surpreendente, visto que garantir a confidencialidade de uma população tão reduzida é um grande desafio, seja qual for o método considerado.

No exemplo da Tabela A.4a, a recodificação foi aplicada para produzir microdados resultantes (Tabela A.4b) 2-anonimizados, nos quais todas as combinações dos valores dos

⁴O desenvolvimento dessa abordagem foi motivado pelos resultados de investigação realizada pela pesquisadora Latanya Sweeney, publicada sob o título "*Simple Demographics Often Identify People Uniquely*" (Carnegie Mellon University, 2000), revelando que muitas pessoas são identificadas exclusivamente usando apenas três informações: data de nascimento, gênero registrado pelo governo e código postal. Em particular, o então governador de Massachusetts foi reidentificado em uma base de registros médicos anonimizados.

⁵Neste contexto, pareamento se refere ao processo de vinculação de registros correspondentes em um ou mais conjuntos de microdados desidentificados, a partir de atributos (semi-identificadores) compatíveis.

atributos têm ao menos 2 ocorrências ($k = 2$).

Tabela A.4: Exemplo k-anonimização.

Sexo	Nível de Escolaridade
F	Médio
M	Fundamental
M	Superior
F	Superior
M	Superior

Tabela elaborada pelo autor.

(a) Original

Sexo	Nível de Escolaridade
F	Educação básica ou superior
M	Educação básica ou superior
M	Educação básica ou superior
F	Educação básica ou superior
M	Educação básica ou superior

Tabela elaborada pelo autor.

(b) Resultante

É importante destacar que, nesse exemplo, o critério de confidencialidade proposto foi alcançado, mas a utilidade foi substancialmente reduzida, uma vez que deixou de ser possível fazer distinção do nível de escolaridade dos indivíduos representados, implicando que o microdado resultante permite produzir estatísticas apenas sobre sexo. Novamente, esse é um exemplo com um número muito reduzido de registros e atributos semi-identificadores. Em casos reais, é pouco provável que a totalidade da população seja recodificada para uma única classe, ainda que isso possa se verificar em subpopulações específicas. Com maior variedade de registros e atributos semi-identificadores, é possível trabalhar uma solução que garanta um dado k -anonimizado, assumindo ser esse um critério de confidencialidade adequado, minimizando a perda de informação. No entanto, a principal fragilidade apontada na k -anonimização se deve a uma eventual violação de atributos sensíveis [111, p.6].

A.2.4 l-Diversidade

A l-Diversidade [136] foi a primeira proposta para tratar a fragilidade relacionada a atributos sensíveis apontada na Seção A.2.3. A abordagem é baseada essencialmente no fato de que tal fragilidade é advinda de diversidade insuficiente nos valores dos atributos sensíveis nos agrupamentos resultantes da k -anonimização. Portanto, a k -anonimização busca produzir grupos homogêneos quanto aos atributos semi-identificadores, enquanto a l-diversidade é um critério complementar que busca garantir heterogeneidade no que se refere aos atributos sensíveis desses grupos. Mais especificamente, a l-diversidade garante que ao menos l diferentes valores de um ou mais atributos sensíveis estejam presentes em cada grupo.

A.2.5 PRAM

O PRAM [91] ou método de pós aleatorização (do inglês, *post-randomization method*) é uma abordagem perturbativa para atributos categóricos, que consiste em alterar a classificação atribuída a alguns indivíduos de acordo com um mecanismo probabilístico estabelecido previamente, de forma a dificultar reidentificação de registros, mas permitir a produção de estatísticas agregadas.

O PRAM é uma abordagem perturbativa controversa, uma vez que altera classificações atribuídas originalmente e que não é trivial fazer inferências suficientemente precisas a partir de microdados a ele submetidos [111, p.77].

A.2.6 Microagregação

A Microagregação⁶ [50] é uma abordagem perturbativa proposta para dados contínuos, que consiste em particionar os n objetos de um [conjunto de microdados](#) com h atributos em k pequenos grupos homogêneos de tamanho n_i , estando n_i restrito ao intervalo definido por $l \leq n_i < 2l \ \forall i \in \{1, \dots, k\}$, sendo l um inteiro positivo correspondente ao nível de agregação e $\lceil n/(2l - 1) \rceil \leq k \leq \lfloor n/l \rfloor$. O critério de homogeneidade é dado por uma métrica (e.g. distância euclidiana), que permite avaliar a dissimilaridade entre os objetos do [conjunto de microdados](#). Os atributos dos objetos são então substituídos pelo agregado correspondente (e.g. média do grupo) e esses valores são divulgados em substituição aos valores originais [62].

A.2.7 Geração de microdados sintéticos

A geração de microdados sintéticos é uma abordagem perturbativa mais radical. Nela, microdados são gerados de maneira pseudo aleatória, considerando, no entanto, restrições que preservem determinadas propriedades estatísticas e associações do microdado original.

Essa abordagem é dividida em três níveis, quais sejam: (i) microdados totalmente sintéticos, onde um microdado inteiramente novo é gerado, não sendo preservada nenhuma informação original no microdado resultante; (ii) microdados parcialmente sintéticos, onde apenas registros ou atributos considerados mais sensíveis são gerados sinteticamente e substituídos para compor o microdado resultante; e (iii) microdados híbridos, onde mi-

⁶A microagregação multivariada ótima pode ser formulada como um problema de otimização combinatoria não linear [67, 34].

crodados sintéticos são combinados com dados originais para produzir o microdado resultante, geralmente com base em critérios de proporção.

O tema é extenso, tendo sido explorado pela academia intensamente nos últimos tempos [145, 157, 187, 113] e abordado em profundidade em algumas literaturas de referência da área de SDC [111, p.78-99][53].

A.2.8 Privacidade diferencial

A Privacidade Diferencial (PD) se distingue das abordagens anteriores de *output privacy* por ser idealizada para a realização de *consultas* (e.g., frequência de indivíduos com determinada característica de interesse, média salarial de uma determinada subpopulação, etc.) no modelo de bancos de dados estatísticos⁷. É, portanto, uma abordagem para microdados, pois permite ao usuário interagir com os registros individualizados, mas, diferentemente das demais abordagens apresentadas, não gera um *conjunto de microdados* resultante ao qual o usuário terá pleno acesso⁸.

É uma abordagem que pode ser aplicada também na produção de tabelas, através da execução de múltiplas *consultas*, cada uma correspondendo a uma célula da tabulação final. Nesse caso, contudo, é necessário avaliar o tratamento a ser dado aos totais marginais, uma vez que a simples totalização por linhas ou colunas resulta em valores não necessariamente em acordo com os critério de PD e *consultas* independentes para os totais marginais produzem tabelas inconsistentes.

A PD é uma abordagem baseada no conceito de que se um analista extrai as mesmas estatísticas de um conjunto de dados D_1 com n registros e de uma versão desse conjunto, D_2 , com um registro a menos, i.e., $n-1$ registros, é possível aprender algo sobre o indivíduo ao qual o registro ausente se refere [48].

Para ilustrar esse conceito, considere o exemplo⁹ de uma instituição de ensino que divulga mensalmente o seu número de calouros, bem como quantos deles têm renda familiar mensal superior a 20 salários mínimos. Em determinado mês, essas estatísticas

⁷Bancos de dados estatísticos só permitem a execução de *consultas* como “a idade média das pessoas neste banco de dados” e impedem *consultas* possíveis em bancos de dados relacionais tais como “SELECT * FROM users WHERE age = 32”[44, p.5]. Isto é, o usuário de um banco de dados estatístico pode interagir com os registros individualizados presentes no banco, mas nunca visualizá-los ou extraí-los. Apenas estatísticas resumo (agregados) podem ser extraídas.

⁸É possível adotar PD no modelo não interativo ou *offline*, como descrito na Seção 2.4.2, para produzir um conjunto de dados resultante autônomo, porém sem necessariamente reproduzir o layout do conjunto de dados original.

⁹Adaptado de [202, p.227-229].

indicavam, respectivamente, 3005 e 202 alunos. No mês seguinte, muito embora nenhum novo processo para ingresso de alunos tenha sido concluído, a instituição atualizou as estatísticas indicando 3004 calouros, dos quais 201 membros eram de famílias com renda superior a 20 salários mínimos. Um calouro, sabendo que um de seus colegas de turma abandonou o curso no final do mês anterior, conclui que a família desse colega tem renda mensal superior a 20 salários mínimos.

Esse exemplo anterior ilustra o chamado problema de composição [57], no qual divulgações aparentemente inofensivas podem levar a violações de privacidade quando combinadas. Outra vulnerabilidade relevante está relacionada ao chamado problema de inferência [48], quando é possível deduzir/inferir dados confidenciais a partir de estatísticas agregadas e outras informações disponíveis a priori. A vulnerabilidade na divulgação de resultados obtidos a partir da aplicação de métodos de agrupamento, descrita na seção 2.3, são um tipo de ataque de inferência.

Estratégias como essa podem ser aplicadas de maneira coordenada para realizar ataques de reconstrução de conjuntos de dados [49]. De maneira mais ampla, métodos de aprendizado de máquina estão sujeitos a diversos tipos de ataques, comumente denominados *adversarial machine learning attacks* [112]. Devido à sensibilidade à manipulação de dados de entrada e à dependência das escolhas feitas na seleção de atributos e arquitetura do modelo, os resultados produzidos por esses métodos podem apresentar fragilidades exploráveis, tornando-os suscetíveis a ataques como, por exemplo, engenharia reversa de modelos disponibilizados publicamente.

Por conta disso, muitos métodos de *output privacy* foram propostos para contornar essas vulnerabilidades. Inicialmente, eram pouco eficazes nesse propósito ou comprometiam de maneira substantiva a utilidade da saída produzida. Muitas abordagens mais recentes, que demonstram robustez em diversas situações, carecem de garantias formais, i.e., não oferecem limites matematicamente demonstráveis sobre o risco de divulgação, deixando-as permanentemente vulneráveis a potenciais novos recursos ou estratégias de ataque¹⁰. No entanto, a Privacidade Diferencial (PD) [57], proposta em 2006, conseguiu demonstrar garantias de privacidade ao adicionar ruído a qualquer consulta estatística.

Retomando o exemplo da divulgação do número de calouros da instituição de ensino, se as mesmas estatísticas mensais fossem perturbadas antes de serem divulgadas, i.e.,

¹⁰Métodos como k-anonimização e l-diversidade, por exemplo, dependem fortemente de hipóteses sobre o conhecimento prévio do adversário e sobre as características dos dados disponibilizados, o que os torna suscetíveis a ataques de reidentificação ou inferência. Dessa forma, tais abordagens permanecem potencialmente vulneráveis ao surgimento de novos recursos computacionais, disponibilização de novas bases auxiliares de dados ou estratégias de ataque mais sofisticadas.

tivessem adicionada uma quantidade cuidadosamente ajustada de ruído - gerado segundo uma distribuição estatística [202, p.229] -, as mesmas inferências sobre a renda familiar do aluno desistente não seriam possíveis.

A Privacidade Diferencial tem, como principal apelo, uma fundamentação matemática robusta e tem entre suas principais vantagens não depender de suposições quanto aos tipos de ataque ou informações às quais o adversário pode ter acesso e oferecer análise quantitativa para determinar o risco de reidentificação [128, p.2]. O método é descrito em maior profundidade e formalmente apresentado na Seção 2.4 (p.14).

APÊNDICE B – Mecanismos de PD

Este apêndice apresenta definições, propriedades e detalhamentos complementares dos mecanismos de Privacidade Diferencial mencionados ao longo deste trabalho. O objetivo é fornecer uma visão mais aprofundada dos fundamentos matemáticos associados a esses mecanismos, oferecendo suporte à compreensão das formulações e abordagens propostas. São descritos, em particular, os mecanismos de Laplace e Normal, incluindo aspectos relacionados à sensibilidade, calibração de ruído e garantias de privacidade, de modo a complementar as discussões apresentadas nos capítulos principais com vistas a não comprometer a fluidez da exposição metodológica.

B.1 Mecanismo de resposta aleatorizada

Para variáveis dicotômicas, é possível formalizar um mecanismo de PD a partir da abordagem baseada na resposta aleatorizada [195], descrita na Seção 2.4. A garantia da confidencialidade nessa abordagem se dá com base no argumento da "negação plausível" [59, p.30]. Por exemplo, considere que uma força de segurança pública tenha acesso a um conjunto de dados referente à prática de atividades ilegais em uma determinada população. Suponha que, após a aplicação do mecanismo de resposta aleatorizada, aproximadamente $\frac{1}{4}$ das respostas observadas sejam positivas, indicando que o respondente declarou exercer a atividade investigada. Nesse contexto, não é possível afirmar com certeza se uma resposta positiva corresponde a uma admissão verdadeira ou se foi produzida aleatoriamente pelo mecanismo, preservando a confidencialidade do indivíduo. Tal exemplo ilustra que o processo utilizado para obter as respostas afeta a forma como elas podem ser legitimamente interpretadas [59, p.30].

Mais especificamente, para um respondente da pesquisa que de fato exerceu a atividade ilegal (verdade = sim), o "sim" foi registrado no levantamento se o resultado obtido no primeiro lançamento da moeda foi "coroa" ($\frac{1}{2}$) ou o resultado de ambos os lançamentos foi "cara" ($(\frac{1}{2})^2 = \frac{1}{4}$), ou seja, a probabilidade da saída do mecanismo ser "sim" quando

a verdade foi "sim" é dada por $P[\text{resposta} = \text{sim} | \text{verdade} = \text{sim}] = \frac{3}{4}$. Já se o respondente não exerceu tal atividade (verdade = não), a probabilidade da saída do mecanismo ser "sim" é dada por $P[\text{resposta} = \text{sim} | \text{verdade} = \text{não}] = \frac{1}{4}$.

O mecanismo desconhece a verdade, ou seja, se o respondente de fato exerceu ou não a atividade, portanto essa informação é indiferente para sua resposta/saída. Entretanto, é possível aplicar raciocínio análogo ao anterior para quando a resposta registrada (ou a saída do mecanismo) foi "não". Temos então

$$\frac{P[\text{resposta} = \text{sim} | \text{verdade} = \text{sim}]}{P[\text{resposta} = \text{sim} | \text{verdade} = \text{não}]} = \frac{P[\text{resposta} = \text{não} | \text{verdade} = \text{não}]}{P[\text{resposta} = \text{não} | \text{verdade} = \text{sim}]} = \frac{3/4}{1/4} = 3$$

Da definição de PD, dada pela Eq. 2.11 (p. 26):

$$\frac{P[A(D_1) \in S]}{P[A(D_2) \in S] + \delta} \leq \exp(\epsilon)$$

Temos então que $\epsilon = \ln(3)$ e $\delta = 0$, logo o mecanismo de resposta aleatorizada é $(\ln(3), 0)$ -diferencialmente privado (ou satisfaz $(\ln(3), 0)$ -PD).

O mecanismo é baseado na distribuição de Bernoulli, o que permite generalizar a demonstração. Essa distribuição representa a probabilidade de uma variável aleatória discreta binária X assumir o valor $x = 1$ (sim ou verdadeiro) com probabilidade p e assumir o valor $x = 0$ (não ou falso) com probabilidade $1 - p$. Assim, se $X \sim \text{Bernoulli}(p)$, temos que

$$P[X = x] = \begin{cases} p, & \text{se } x = 1 \\ 1 - p, & \text{se } x = 0 \end{cases}$$

O controle da privacidade se dá através do parâmetro p . Se $p = 1$, nenhum ruído é introduzido. Por outro lado, se $p = 0,5$, a utilidade é nula, pois os valores 0 e 1 são selecionados em igual proporção.

Uma vez que $X \sim \text{Bernoulli}(p)$, $x \in S = \{0, 1\}$. Considerando novamente a Eq. 2.11 com $\delta = 0$ e incorporando os requisitos de PD, temos

$$\max_{D_1 \sim D_2} \max_{S=\{0,1\}} \ln \left(\frac{P[A(D_1) = S]}{P[A(D_2) = S]} \right) \leq \epsilon$$

Logo

$$\begin{aligned} \max_{D_1 \sim D_2} \max \left(\ln \left(\frac{P[A(D_1) = 1]}{P[A(D_2) = 1]} \right), \ln \left(\frac{P[A(D_1) = 0]}{P[A(D_2) = 0]} \right) \right) \\ = \max \left(\ln \left(\frac{p}{1-p} \right), \ln \left(\frac{1-p}{p} \right) \right) \end{aligned}$$

Considerando que o mecanismo deve oferecer utilidade não nula, restringe-se $p > 0,5$ e, portanto,

$$\max \left(\ln \left(\frac{p}{1-p} \right), \ln \left(\frac{1-p}{p} \right) \right) = \ln \left(\frac{p}{1-p} \right) = \epsilon$$

Com isso, demonstra-se que, no pior caso, a perda de privacidade é ϵ , logo o mecanismo é ϵ -diferencialmente privado (ou satisfaz ϵ -PD).

B.2 Mecanismo exponencial

Para casos de dados não numéricos, o mecanismo de privacidade diferencial geralmente adotado é o exponencial. Dada uma determinada consulta, esse mecanismo é baseado em uma função de ponderação $q(X, y)$, que expressa o quão adequada é a resposta y , dado o conjunto de dados X . Nessa formulação, y representa uma possível resposta a uma consulta feita sobre X .

O mecanismo exponencial funciona como uma amostragem ponderada [135, p.1543], na qual cada possível resposta recebe um peso proporcional à sua adequação, i.e., ao quão correta ou apropriada ela é. Assim, respostas com peso mais alto têm maior probabilidade de serem selecionadas, mas de forma controlada para garantir a privacidade diferencial. Isso permite preservar privacidade ao mesmo tempo em que prioriza respostas adequadas, mesmo em domínios com respostas relacionadas com variáveis qualitativas ou categóricas.

A fim de ilustrar esse conceito, considere o problema de descobrir a cidade favorita entre um grupo de usuários, de maneira que a privacidade individual seja preservada. Para isso, utilizamos o mecanismo exponencial, adequado para domínios não numéricos.

Seja o domínio de possíveis respostas:

$$\mathcal{V} = \{\text{São Paulo, Rio de Janeiro, Recife, Porto Alegre}\}.$$

Cada usuário informa sua cidade favorita e o conjunto de dados agregado é representado por X .

Uma possível consulta aplicada a esse conjunto pode ter o objetivo de responder qual foi a cidade mais votada pelo grupo. Deseja-se, no entanto, garantir privacidade diferencial, o que pode ser feito através da aplicação do mecanismo exponencial, conforme descrito a seguir.

1. Define-se uma consulta ou função de pontuação $q(X, y)$, que mede o quão apropriada é uma resposta $y \in \mathcal{V}$ para o conjunto X . Neste caso, poderia ser $q(X, y) =$ número de votos para a cidade.

Considere que os dados contidos em X sejam:

- São Paulo: 40 votos,
- Rio de Janeiro: 35 votos,
- Recife: 10 votos,
- Porto Alegre: 5 votos.

2. O mecanismo A pode ser aplicado ao realizar a consulta q em X de forma a garantir PD, fazendo com que cada possível saída $y \in \mathcal{V}$ seja selecionada com probabilidade $P[A(X) = y] \propto \exp\left(\frac{\epsilon q(X, y)}{2\Delta q}\right)$, onde ϵ é o *privacy budget* e Δq é a sensibilidade da consulta (neste caso, $\Delta q = 1$).
3. A saída final é obtida por amostragem ponderada sobre $\mathcal{V}$ conforme as probabilidades definidas.

Análise: Se ϵ for alto (pouco ruído), a cidade com mais votos (São Paulo) tem maior probabilidade de ser escolhida. Já com ϵ baixo (ruído elevado), as probabilidades entre as cidades se tornam mais próximas, preservando melhor a privacidade dos indivíduos, mas com menor fidelidade à distribuição real.

B.3 Definições complementares para o mecanismo de Laplace

B.3.1 Demonstração do mecanismo de Laplace

Para verificar se o mecanismo de Laplace satisfaz a condição ϵ -PD, primeiramente definimos a variável aleatória X tal que $X \sim \text{Laplace}(\mu, b)$. Temos assim que

$$P[X = x] = \frac{1}{2b} \exp\left(\frac{-|x - \mu|}{b}\right)$$

Considerando a Eq. 2.11 com $\delta = 0$ e incorporando os requisitos de PD, temos então

$$\max_{D_1 \sim D_2} \max_{S \subseteq \text{Im}(A)} \ln \left(\frac{P[A(D_1) \in S]}{P[A(D_2) \in S]} \right) \leq \epsilon$$

Logo

$$\begin{aligned} & \max_{D_1 \sim D_2} \max_{S \in \mathcal{R}} \ln \left(\frac{\frac{1}{2b} \exp\left(\frac{-|S - D_1|}{b}\right)}{\frac{1}{2b} \exp\left(\frac{-|S - D_2|}{b}\right)} \right) \\ &= \max_{D_1 \sim D_2} \max_{S \in \mathcal{R}} \ln \left(\exp\left(\frac{-|S - D_1|}{b}\right) - \frac{-|S - D_2|}{b} \right) \\ &= \max_{D_1 \sim D_2} \max_{S \in \mathcal{R}} \frac{|S - D_2| - |S - D_1|}{b} \leq \max_{D_1 \sim D_2} \frac{|D_1 - D_2|}{b} = \frac{\Delta}{b} \end{aligned}$$

Com isso, sempre que $\frac{\Delta f}{b} \leq \epsilon$, o mecanismo satisfaz ϵ -PD.

B.3.2 Demonstração alternativa do mecanismo de Laplace

A distribuição de Laplace centrada em zero ($\mu = 0$) e com parâmetro de escala b tem variância $\sigma^2 = 2b^2$ e função densidade de probabilidade descrita pela Eq. B.1.

$$\text{Lap}(b) = \text{Lap}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \quad (\text{B.1})$$

O mecanismo de Laplace pode então ser formalmente definido, para qualquer função $f : \mathbb{N}^{|X|} \rightarrow \mathcal{R}^k$, como [59, p.32]

$$A_L(x, f(\cdot), \epsilon) = f(x) + (Y_1, \dots, Y_k) \quad (\text{B.2})$$

sendo Y_i correspondentes a variáveis aleatórias independentes e identicamente distribuídas geradas segundo $\text{Lap}\left(\frac{\Delta f}{\epsilon}\right)$, i.e., distribuição de Laplace com parâmetros $\mu = 0$ e $b = \frac{\Delta f}{\epsilon}$.

Sejam então $x \in \mathbb{N}^{|X|}$ e $y \in \mathbb{N}^{|X|}$ tais que $\|x - y\| \leq 1$ e $f(\cdot)$ uma função $f : \mathbb{N}^{|X|} \rightarrow \mathcal{R}^k$. Sejam ainda p_x e p_y as funções densidade de probabilidade de $A_L(x, f, \epsilon)$ e $A_L(y, f, \epsilon)$,

respectivamente, comparando ambas em um ponto arbitrário $z \in \mathcal{R}^k$, temos [59, p.32]

$$\frac{p_x(z)}{p_y(z)} = \prod_{i=1}^k \left(\frac{\exp\left(-\frac{\epsilon|f(x)_i - z_i|}{\Delta f}\right)}{\exp\left(-\frac{\epsilon|f(y)_i - z_i|}{\Delta f}\right)} \right) = \prod_{i=1}^k \exp\left(\frac{\epsilon(|f(y)_i - z_i| - |f(x)_i - z_i|)}{\Delta f}\right)$$

Pela desigualdade triangular,

$$\frac{p_x(z)}{p_y(z)} \leq \prod_{i=1}^k \exp\left(\frac{\epsilon|f(x)_i - f(y)_i|}{\Delta f}\right) = \exp\left(\frac{\epsilon \cdot \|f(x) - f(y)\|_1}{\Delta f}\right)$$

Dado que $\|x - y\| \leq 1$ e pela definição de sensibilidade (Eq. 2.9), temos

$$\frac{p_x(z)}{p_y(z)} = \exp\left(\frac{\epsilon \cdot \|f(x) - f(y)\|_1}{\Delta f}\right) \leq \exp(\epsilon)$$

Finalmente, por simetria, temos ainda que $\frac{p_x(z)}{p_y(z)} \geq \exp(-\epsilon)$. Assim, o mecanismo de Laplace preserva $(\epsilon, 0)$ -privacidade diferencial.

B.3.3 Generalização (ϵ, δ) -PD

Conforme apresentado na Eq. 2.11 (p. 26), um mecanismo A satisfaz (ϵ, δ) -PD se, para todos os possíveis conjuntos de dados adjacentes D_1 e D_2 e qualquer subconjunto S de possíveis saídas, i.e., $S \subseteq \text{Im}(A)$:

$$P[A(D_1) \in S] \leq P[A(D_2) \in S] \cdot e^\epsilon + \delta$$

A divergência máxima pode então ser generalizada conforme Eq. B.3, sendo X e Y duas distribuições de probabilidade.

$$D_\infty^\delta(X||Y) = \max_{S \subseteq \text{Im}(A)} \ln \left(\frac{P[X \in S] + \delta}{P[Y \in S]} \right) \quad (\text{B.3})$$

Finalmente, a definição de (ϵ, δ) -PD via divergência é dada pela Eq. B.4.

$$\max_{D_1 \sim D_2} D_\infty^\delta(A(D_1), A(D_2)) \leq \epsilon \quad (\text{B.4})$$

O parâmetro δ suaviza a medida de divergência máxima (dada pela Eq. B.5 - p. 157),

permitindo que a diferença relativa entre distribuições seja grande, desde que a diferença absoluta permaneça pequena [44, p.114]. A recomendação da literatura é adotar valores bem pequenos, i.e., $\delta \ll 1/n$ [28], sendo n correspondente ao número de registros no conjunto de dados a ser protegido. Tipicamente, $10^{-9} \leq \delta \leq 10^{-6}$ [44, p.114].

$$D_{\max\text{-div}}(X||Y) = \max_{S \subseteq \text{Im}(A)} \ln \left(\frac{P[X \in S]}{P[Y \in S]} \right) = \max_{z \in \text{Im}(A)} \ln \left(\frac{P[X = z]}{P[Y = z]} \right) \quad (\text{B.5})$$

A generalização (ϵ, δ) -PD é também denominada PD aproximada [141], uma vez que representa uma relaxação da garantia absoluta oferecida pela PD pura, quando $\delta = 0$, ou ϵ -PD. É possível verificar que um mecanismo satisfaz a PD aproximada, demonstrando que ele satisfaz a PD probabilística [44, p.115].

A definição de PD probabilística adota a variável aleatória de perda de privacidade (do inglês, *Privacy Loss Random Variable* - PLRV), denotada por Z e definida pela Eq. B.6, que captura a perda de privacidade incorrida pela observação da saída y^1 , quando aplicado o mecanismo A nos conjuntos de dados adjacentes D_1 e D_2 [44, p.113].

$$Z = \ln \left(\frac{P[A(D_1) = y]}{P[A(D_2) = y]} \right) \quad (\text{B.6})$$

Uma vez que ϵ -PD garante que, para todos os conjuntos de dados adjacentes D_1 e D_2 , a perda de privacidade é superiormente limitada em ϵ , então, sob ϵ -PD, $P[|Z| > \epsilon] = 0$ para qualquer par de conjuntos de dados adjacentes.

$$P[|Z| > \epsilon] \leq \delta \quad (\text{B.7})$$

A Eq. B.7² estabelece que, sob PD probabilística, Z pode romper o limite ϵ com probabilidade máxima δ , i.e., que a probabilidade de o mecanismo falhar em satisfazer ϵ -DP é superiormente limitada por δ .

Há abordagens específicas para construir mecanismos apropriados a PD aproximada, tais como Propose-Test-Release [56], Zero-Concentrated Differential Privacy (zCDP) [28], Rényi Differential Privacy [143, 144] e Truncated Noise Mechanisms [83].

Independentemente do uso de um mecanismo específico para tratar PD aproximada, o

¹A PLRV quantifica a vantagem que um analista/intruso obtém ao observar uma saída específica y do mecanismo A [44, p.113].

²O requisito preciso da PD aproximada com base na PLRV (do inglês, *privacy loss random variable*) não é tão simples [44, p.115]. Uma derivação pode ser encontrada em [31, p.7].

parâmetro δ pode ser utilizado para avaliar o risco de revelação ou violação da privacidade, uma vez que representa a probabilidade de exceder o limite ϵ . Um exemplo de abordagem nesse sentido adota o mecanismo *name-and-shame*[173], que não é usado na prática, mas propõe o exercício de publicar os dados de cada indivíduo com probabilidade δ [44, 114].

B.3.4 Acurácia (α, β)

A descrição do mecanismo até aqui abordou o aspectos relacionados ao risco de violação da privacidade, mais especificamente demonstrando que o mecanismo de Laplace satisfaz os critérios da PD. Entretanto, a utilidade é uma dimensão que deve ser necessariamente levada em conta, uma vez que não há benefício na divulgação de estatísticas seguras se essas não permitem qualquer conclusão relevante³.

Com esse intuito será adotada a acurácia (α, β) (do inglês, (α, β) -accuracy) [44, p.39-40]. Dados uma entrada qualquer D , uma consulta $q(D)$, um mecanismo A e a saída correspondente $A(D) = y$, a acurácia (α, β) estabelece que a probabilidade da diferença entre o valor divulgado ($A(D)$) e o valor real não perturbado ($q(D)$) ser superior a α é limitada superiormente por β , conforme descrito na Eq. B.8.

$$P[|q(D) - y| \geq \alpha] \leq \beta \quad (\text{B.8})$$

Para o mecanismo de Laplace, isso implica que:

$$\alpha = b \cdot \ln\left(\frac{1}{\beta}\right) \quad (\text{B.9})$$

Por ser simétrica, a distribuição de Laplace permite construir, a partir de uma estimativa de precisão, um intervalo de confiança em torno da estimativa do parâmetro de interesse ou, no contexto de PD, em torno de saídas produzidas pelo mecanismo. Com isso, é possível construir um intervalo $[y - \alpha, y + \alpha]$ com nível de confiança de $100(1 - \beta)\%$ em torno de uma saída do mecanismo de Laplace.

Portanto, para $\beta = 0,05$, por exemplo, espera-se que 95% dos intervalos de confiança centrados em saídas de $A(D)$ contenham o valor real (não perturbado) da consulta correspondente $q(D)$, ou seja, se α for considerado grande, a utilidade de A está comprometida.

³Estatísticas divulgadas sob PD com parâmetros muito conservadores, mais especificamente, valores elevados de ϵ e extremamente pequeno (ou mesmo zero) de δ , podem ter ruído adicionado de ordem de grandeza próxima ou até superior ao real valor calculado, oferecendo resultado excessivamente impreciso.

É possível ainda fazer interpretação análoga ao p-valor, fixando α , i.e., um desvio máximo aceitável/desejável do valor de $q(D)$, e avaliar o β resultante. Quanto mais próximo de 1 for β , menor é a utilidade de A .

B.3.5 Demonstração do mecanismo de Laplace vetorial

O mecanismo de Laplace vetorial satisfaz ϵ -DP, onde b_{out} corresponde ao parâmetro de privacidade associado à divergência máxima entre distribuições de saída [44, p.92]. Seja h a dimensionalidade dos vetores de entrada e saída do mecanismo, i.e., o número de componentes perturbadas independentemente pelo ruído de Laplace. Sejam $v_1, v_2 \in \mathbb{R}^h$ dois vetores provenientes de conjuntos de dados adjacentes (e.g., D_1 e D_2) e denote por v^i a i -ésima componente de um vetor $v \in \mathbb{R}^h$, de modo que $v = (v^1, \dots, v^i, \dots, v^h)$. A demonstração a seguir baseia-se na fatoração da densidade conjunta do mecanismo em função da independência entre as componentes e na aplicação da desigualdade triangular reversa⁴, utilizada para limitar a diferença entre as distâncias absolutas observadas em cada dimensão:

$$\begin{aligned}
 \max_{v_1 \sim v_2} D_{\infty}(A(v_1), A(v_2)) &= \max_{v_1 \sim v_2} \max_{S \subseteq \text{Im}(A)} \ln \left(\frac{P[A(v_1) \in S]}{P[A(v_2) \in S]} \right) \\
 &= \max_{v_1 \sim v_2} \max_{y \in \mathbb{R}^h} \ln \left(\frac{P[VLap(v_1, b) = y]}{P[VLap(v_2, b) = y]} \right) \\
 &= \max_{v_1 \sim v_2} \max_{y \in \mathbb{R}^h} \ln \left(\frac{\prod_{i=1}^h \frac{1}{2b} \exp \left(-\frac{|v_1^i - y^i|}{b} \right)}{\prod_{i=1}^h \frac{1}{2b} \exp \left(-\frac{|v_2^i - y^i|}{b} \right)} \right) \\
 &= \max_{v_1 \sim v_2} \max_{y \in \mathbb{R}^h} \sum_{i=1}^h \ln \left(\frac{\exp \left(-\frac{|v_1^i - y^i|}{b} \right)}{\exp \left(-\frac{|v_2^i - y^i|}{b} \right)} \right) \\
 &= \max_{v_1 \sim v_2} \max_{y \in \mathbb{R}^h} \sum_{i=1}^h \frac{|v_1^i - y^i| - |v_2^i - y^i|}{b} \\
 &\leq \max_{v_1 \sim v_2} \frac{\sum_{i=1}^h |v_1^i - v_2^i|}{b} \\
 &= \frac{\max_{v_1 \sim v_2} \|v_1 - v_2\|_1}{b}
 \end{aligned}$$

Fazendo $\max_{v_1 \sim v_2} \|v_1 - v_2\|_1 = b_{\text{in}}$, temos

⁴A desigualdade triangular reversa estabelece que, para quaisquer $a, b \in \mathbb{R}$, vale $||a| - |b|| \leq |a - b|$. No contexto da demonstração, ela permite limitar a diferença entre as distâncias absolutas observadas em cada componente dos vetores adjacentes.

$$\max_{D_1 \sim D_2} D_\infty(A(D_1), A(D_2)) \leq \frac{b_{\text{in}}}{b}$$

Com isso, demonstra-se que a divergência entre as distribuições de saída é superiormente limitada por $b_{\text{out}} = \frac{b_{\text{in}}}{b}$. Assim, adotando $b = \frac{b_{\text{in}}}{\epsilon}$, obtém-se $b_{\text{out}} \leq \epsilon$, garantindo que o mecanismo satisfaz ϵ -DP.

B.3.6 Cálculo do *privacy budget* mínimo (ϵ_m) por iteração do DPk-means

No contexto do k-means diferencialmente privado, o valor mínimo de *privacy budget* a ser alocado em cada iteração pode ser calculado a partir do erro quadrático médio (do inglês, *Mean Squared Error* - MSE) entre o centroide perturbado e o centroide real [179, p.28][180, p.7][129, p.6].

Em uma determinada iteração, o centroide do grupo C_j é dado por $c_j = \frac{S_j}{|C_j|}$, sendo S_j a soma dos atributos dos objetos $x_i \in C_j$ e $|C_j|$ contagem (frequência) de objetos em C_j . No cálculo do centroide perturbado, um ruído é adicionado na soma das distâncias e na frequência dos objetos, denotados, respectivamente, por Λ_S e Λ_C ⁵. Assim, temos $c_j^* = \frac{S_j^i + \Lambda_S}{|C_j| + \Lambda_C}$.

O MSE entre os dois centroides é então calculado fazendo

$$\text{MSE}(c_j^*) = E \left[\sum_{l=1}^h \left(\frac{S_{jl} + \Lambda_S}{|C_j| + \Lambda_C} - \frac{S_{jl}}{|C_j|} \right)^2 \right]$$

A fim de generalizar o resultado para conjuntos de dados com qualquer distribuição de objetos, assume-se que $|C_j| = |C| = \frac{n}{k} \forall j \in \{1, \dots, k\}$. Temos então que

$$\begin{aligned} \text{MSE}(c_j^*) &\approx \text{MSE}(c^*) = E \left[\sum_{l=1}^h \left(\frac{S_l + \Lambda_S}{|C| + \Lambda_C} - \frac{S_l}{|C|} \right)^2 \right] \\ &= \sum_{l=1}^h E \left[\left(\frac{S_l + \Lambda_S}{|C| + \Lambda_C} - \frac{S_l}{|C|} \right)^2 \right] \\ &= \sum_{l=1}^h E \left[\left(\frac{C\Lambda_S - S_l\Lambda_C}{|C|(|C| + \Lambda_C)} \right)^2 \right] \end{aligned}$$

⁵Considerando mecanismo de Laplace, $\Lambda_S = \text{Lap}\left(\frac{\Delta f(s)}{\epsilon_s}\right)$ e $\Lambda_C = \text{Lap}\left(\frac{\Delta f(c)}{\epsilon_c}\right)$.

Considerando que, idealmente, o ruído adicionado é uma pequena perturbação, $\Lambda_C \ll |C|$, então

$$\begin{aligned} & \sum_{l=1}^h E \left[\left(\frac{C\Lambda_s - S_l\Lambda_C}{|C|(|C| + \Lambda_C)} \right)^2 \right] \\ & \approx \sum_{l=1}^h E \left[\left(\frac{C\Lambda_s - S_l\Lambda_C}{(|C|)^2} \right)^2 \right] \\ & = \sum_{l=1}^h \left(\frac{E[(\Lambda_S)^2]}{(|C|)^2} + \frac{E[S_l^2(\Lambda_C)^2]}{(|C|)^4} + \frac{2|C|S_lE[\Lambda_S\Lambda_C]}{(|C|)^4} \right) \end{aligned}$$

Uma vez que Λ_S e Λ_C são variáveis aleatórias independentes segundo uma distribuição de Laplace com parâmetros $\mu = 0$ e $b = \frac{h+1}{\epsilon_t}$, $E[\Lambda_S\Lambda_C] = E[\Lambda_S] + E[\Lambda_C] = 0$. Além disso, a variância de uma variável aleatória X é dada por $V[X] = E[(X - E[X])^2] = E[X^2] - E^2[X]$, logo

$$\begin{aligned} & \sum_{l=1}^h \left(\frac{E[(\Lambda_S)^2]}{(|C|)^2} + \frac{E[S_l^2(\Lambda_C)^2]}{(|C|)^4} + \frac{2|C|S_lE[\Lambda_S\Lambda_C]}{(|C|)^4} \right) \\ & = \sum_{l=1}^h \left(\frac{V[\Lambda_S]}{(|C|)^2} + \frac{S_l^2V[\Lambda_C]}{(|C|)^4} \right) \\ & = \sum_{l=1}^h \left(1 + \frac{S_l^2}{(|C|)^2} \right) \left(\frac{2(h+1)^2}{\epsilon_t^2(|C|)^2} \right) \\ & = 2 \sum_{l=1}^h \left(1 + \frac{S_l^2}{(|C|)^2} \right) \left(\frac{k(h+1)}{n\epsilon_t} \right)^2 \\ & \approx 2h \left(1 + \left(\frac{1}{2} \right)^2 \right) \left(\frac{k(h+1)}{n\epsilon_t} \right)^2 \\ & = \frac{5h(k(h+1))^2}{2(n\epsilon_t)^2} \end{aligned}$$

Adotando critério proposto em [129, p.6] de limitar o MSE em 0,01, a parcela mínima de *privacy budget* em um iteração qualquer do DPk-means é dada então por:

$$\begin{aligned}\frac{5h(k(h+1))^2}{2(n\epsilon_t)^2} &\leq 0,01 \\ \epsilon_t^2 &\geq \frac{5h(k(h+1))^2}{0,01 \times 2n^2} \\ \epsilon_t &\geq \sqrt{\frac{250h(k(h+1))^2}{n^2}}\end{aligned}$$

Logo,

$$\epsilon_m = \sqrt{\frac{250h(k(h+1))^2}{n^2}}$$

B.4 Definições complementares para o mecanismo Normal

Diferentemente do mecanismo de Laplace, no qual é possível limitar uniformemente a divergência máxima entre as distribuições de saída para quaisquer valores possíveis, o mecanismo Normal não admite tal limitação. Em particular, a divergência máxima

$$D_{\max\text{-div}}(A(D_1) || A(D_2)) = \max_{z \in \text{Im}(A)} \ln \left(\frac{P[A(D_1) = z]}{P[A(D_2) = z]} \right) \quad (\text{B.10})$$

não é finita no caso Normal, uma vez que existem valores extremos de z para os quais a razão entre as densidades pode crescer arbitrariamente. Isso decorre da presença de termos lineares no expoente da razão de verossimilhança, que não podem ser uniformemente limitados. Como consequência, não é possível estabelecer um limite superior uniforme ϵ para essa divergência. Embora o mecanismo Normal não admita um limite finito para a divergência máxima, essa limitação não implica ausência de garantia de privacidade. É possível caracterizar de forma precisa o grau de indistinguibilidade entre suas distribuições de saída por meio de uma interpretação baseada em testes de hipótese [52, p.9]. Mais especificamente, pode-se analisar o problema como um teste de hipóteses entre duas distribuições:

$$H_0 : Z \sim \mathcal{N}(f(D_1), \sigma^2 I)$$

$$H_1 : Z \sim \mathcal{N}(f(D_2), \sigma^2 I)$$

Nesse contexto, a capacidade de distinguir entre H_0 e H_1 é completamente caracterizada por uma função de trade-off $T(\alpha)$, que descreve o menor erro do tipo II possível para um dado nível de erro do tipo I α .

No caso Normal, essa função depende apenas da distância entre as médias normalizada pelo desvio padrão do ruído, i.e.,

$$\mu = \frac{\|f(D_1) - f(D_2)\|_2}{\sigma}$$

Assim, embora a divergência máxima seja infinita, o mecanismo Normal apresenta um trade-off finito e bem definido entre erros do tipo I e tipo II, o que permite quantificar de forma precisa o grau de indistinguibilidade entre as distribuições de saída. Com isso, adota-se uma relaxação da divergência máxima, permitindo que a condição de indistinguibilidade falhe em um conjunto de baixa probabilidade. Mais especificamente, introduz-se o parâmetro adicional δ , que limita a probabilidade desses eventos extremos, resultando na seguinte condição:

$$P[A(D_1) \in S] \leq e^\epsilon P[A(D_2) \in S] + \delta,$$

para todo conjunto mensurável $S \subseteq \text{Im}(A)$.

Assim, enquanto o mecanismo de Laplace garante um limite uniforme para a divergência máxima entre distribuições de saída, o mecanismo Normal garante esse limite apenas com alta probabilidade, sendo essa relaxação formalizada pela noção de (ϵ, δ) -PD.

A Figura B.1 ilustra a distinção fundamental entre as garantias de privacidade oferecidas pelos mecanismos de Laplace e Normal ao perturbar a resposta de uma consulta com sensibilidade $\Delta f = 1$ e $\epsilon = 1$. Os painéis superiores tratam do mecanismo de Laplace. O primeiro exibe as densidades $p_{D_1}(z)$ e $p_{D_2}(z)$, centradas nos valores verdadeiros da consulta sobre os conjuntos de dados adjacentes D_1 e D_2 , com escala $b = \Delta f / \epsilon = 1$. O segundo mostra que a perda de privacidade $L_{D_1, D_2}(z) = \log[p_{D_1}(z)/p_{D_2}(z)]$ permanece estritamente limitada ao intervalo $[-\epsilon, \epsilon]$ para todo z , confirmando numericamente ($\max |L| \approx 1,0$) que o mecanismo de Laplace satisfaz ϵ -PD pura, sem qualquer violação residual ($\delta_\epsilon \approx 3,16 \times 10^{-17}$, numericamente equivalente a zero). Os painéis inferiores repetem a análise para o mecanismo Normal, com $\sigma = 1,5$. O terceiro painel evidencia,

por meio da área hachurada, a região de saídas z em que $p_{D_1}(z)$ excede $e^\epsilon p_{D_2}(z)$, i.e., precisamente a massa de probabilidade que viola a garantia de ϵ -DP pura. O quarto painel mostra que a perda de privacidade $L_{D_1, D_2}(z)$ da Normal não é limitada efetivamente por $\pm\epsilon$. Essa violação se traduz em um δ_ϵ efetivo de aproximadamente 0,031 (idêntico em ambas as direções, por simetria), evidenciando que o mecanismo Normal não oferece ϵ -PD pura para qualquer escolha arbitrária de σ . Sua garantia é aproximada, do tipo (ϵ, δ) -PD, com $\delta > 0$ refletindo a probabilidade residual de que a perda de privacidade exceda o limiar ϵ em uma fração não desprezível do espaço de saídas.

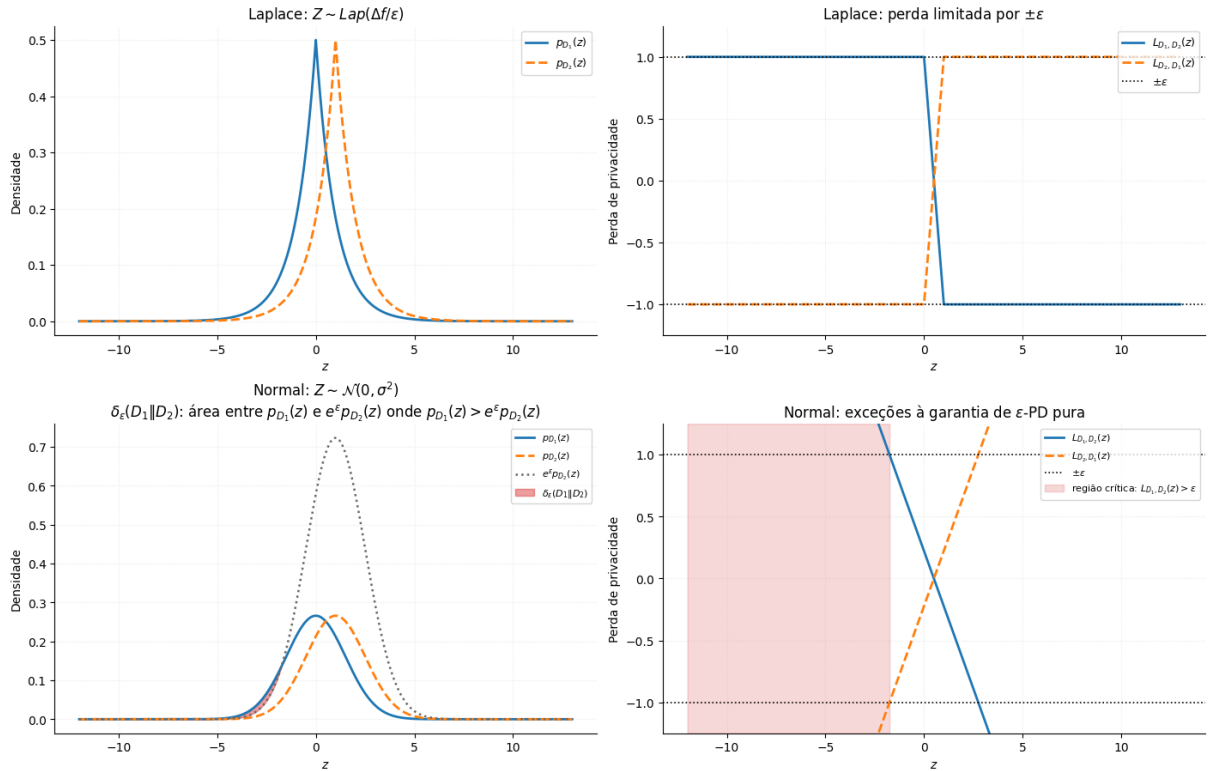


Figura B.1: Comparação entre as garantias de privacidade oferecidas pelos mecanismos de Laplace e Normal.

B.5 Sensibilidade

Uma questão central na formulação de mecanismos de privacidade diferencial se deve ao fato de que a sensibilidade da função f é medida em normas distintas no mecanismo de Laplace ℓ_1 e Normal ℓ_2 . Essa diferença não é arbitrária, mas decorre diretamente das propriedades das distribuições utilizadas [59].

No mecanismo de Laplace, o ruído é adicionado de forma independente a cada elemento da saída. A distribuição de Laplace multivariada pode ser vista como um produto

de distribuições unidimensionais, cuja densidade decai exponencialmente com a norma ℓ_1 . Em particular, a razão de verossimilhanças entre duas saídas possíveis depende linearmente da distância ℓ_1 entre elas. Por isso, para garantir que essa razão seja limitada por $\exp(\epsilon)$, conforme estabelecido na definição de PD, o ruído deve ser calibrado à sensibilidade medida na respectiva norma.

Formalmente, essa compatibilidade pode ser entendida pelo fato de que o mecanismo precisa controlar o termo:

$$\frac{P[A(D_1) = y]}{P[A(D_2) = y]}$$

e, no caso Laplaciano, essa razão envolve exponenciais da forma $\exp\left(\frac{|f(D_1) - f(D_2)|_1}{b}\right)$. Assim, a norma ℓ_1 surge naturalmente como a métrica relevante.

Por outro lado, o mecanismo Normal utiliza ruído com distribuição normal multivariada, cuja densidade depende da norma ℓ_2 . Mais especificamente, a densidade de uma distribuição Normal envolve termos do tipo:

$$\exp\left(-\frac{|z - f(x)|_2^2}{2\sigma^2}\right)$$

Ao analisar a razão de verossimilhanças entre os conjuntos de dados adjacentes, surgem termos que dependem da distância ℓ_2 entre $f(D_1)$ e $f(D_2)$. Consequentemente, a sensibilidade relevante passa a ser aquela medida nessa norma, pois é ela que controla a variação da densidade da distribuição Normal.

B.6 Mecanismos aplicados ao DPk-means

Esta seção apresenta definições complementares dos mecanismos de Laplace e Normal direcionadas especificamente à sua aplicação ao Problema de Agrupamento Diferencialmente Privado, mais especificamente na atualização dos centroides do DPk-means sob o modelo de PD global iterativo.

Em cada iteração do k-means, a atualização ou cálculo dos centroides é perturbada através da aplicação de um mecanismo de PD às consultas referentes às operações de soma dos atributos dos objetos $x_i \in C_j$, denotada por $f_j^s(D) = S_j$, e contagem (frequência) de objetos em cada grupo C_j , denotada por $f_j^c(D) = |C_j|$, resultando nos valores perturbados dessas consultas, S_j^* e $|C_j^*|$, respectivamente. No DPk-means, o cálculo do

centroide perturbado c_j^* é dado então por $c_j^* = \frac{S_j^*}{|C_j^*|}$, para cada grupo C_j , $j = 1, \dots, k$. A magnitude do ruído gerado pelo mecanismo de PD na iteração t ($t \in \{1, \dots, T\}$) é dada pelo parâmetro ϵ_t , e distribuída entre as duas consultas sucessivas, de forma que $\epsilon_t = \epsilon_t^s + \epsilon_t^c$, sendo $\sum_{t=1}^T \epsilon_t \leq \epsilon$, em que $\epsilon > 0$ é o *privacy budget* total.

Mais especificamente, seja a soma dos atributos dada por $f_j^s(D) = S_j = \sum_{x_i \in C_j} x_i \in \mathcal{R}^h$ e a frequência dos atributos dada por $f_j^c(D) = |C_j| \in \mathbb{N}$, aplicando-se os mecanismos de Laplace e Normal, os centroides perturbados c_j^* são dados pelas Eq. B.11 e B.12, respectivamente.

$$A_L(c_{jt}) = \frac{f_{jt}^s(D) + (Y_1, \dots, Y_l, \dots, Y_h)}{f_{jt}^c(D) + Z} \quad (\text{B.11})$$

Sendo B_1^s e B_1^c as sensibilidades em norma ℓ_1 de f^s e f^c , respectivamente, $Y_l \sim \text{Lap}(B_1^s/\epsilon_t^s) \forall l \in \{1, \dots, h\}$ e $Z \sim \text{Lap}(B_1^c/\epsilon_t^c)$.

$$A_N(c_{jt}) = \frac{f_{jt}^s(D) + \mathcal{N}(0, \sigma_s^2 I_h)}{f_{jt}^c(D) + \mathcal{N}(0, \sigma_c^2)} \quad (\text{B.12})$$

Sendo B_2^s e B_2^c as sensibilidades em norma ℓ_2 de f^s e f^c , respectivamente, $\sigma_s \geq \frac{c \cdot B_2^s}{\epsilon_t^s}$ e $\sigma_c \geq \frac{c \cdot B_2^c}{\epsilon_t^c}$.

Em relação à sensibilidade global Δf (Definição 6, p.24) das consultas realizadas, considerando os objetos descritos por seus h atributos normalizados no intervalo $[0,1]$, a variação máxima na soma dos atributos em norma ℓ_1 e ℓ_2 corresponde a $\Delta_1 f^S = \max_x \|x\|_1 = h$ e $\Delta_2 f^S = \max_x \|x\|_2 = \sqrt{h}$, respectivamente. Como o cálculo da frequência depende simplesmente da presença ou não do objeto, cada objeto pode contribuir com até uma unidade, i.e., $\Delta f^C = 1$.

Duas abordagens foram consideradas para definir a sensibilidade global do cálculo dos centroides. A primeira, mais recorrente na literatura [23, p.5][216, p.490], é dada simplesmente por $\Delta f^S + \Delta f^C$. Com isso, em cada iteração t , adotando-se o mecanismo de Laplace, a soma de atributos e frequência de objetos alocados ao grupo C_j é dada, respectivamente, por $S_j^* = S_j + \text{Lap}\left(\frac{h+1}{\epsilon_t/2}\right)$ e $|C_j^*| = |C_j| + \text{Lap}\left(\frac{h+1}{\epsilon_t/2}\right)$. O centroide do grupo C_j foi então definido por $c_j^* = \frac{S_j^*}{|C_j^*|} \forall j \in \{1, \dots, k\}$. Essa abordagem foi utilizada para a sensibilidade global na aplicação de ambos os mecanismos, Laplace e Normal.

No uso do mecanismo de Laplace, em função das caudas mais pesadas da distribuição a ele associada, em comparação com a distribuição Normal, existe uma probabilidade não

desprezível de que o ruído gerado seja muito distante do centro da distribuição, i.e., de grande intensidade ou amplitude. Uma vez que o parâmetro de dispersão da distribuição é diretamente proporcional à sensibilidade global ($b = \frac{\Delta f}{\epsilon}$), formulou-se a hipótese de que uma abordagem alternativa para o cálculo de Δf poderia contribuir para mitigar esse efeito. Com isso, uma segunda abordagem foi utilizada para o cálculo da sensibilidade global quando utilizado o mecanismo de Laplace. Essa abordagem considera, para o cálculo da soma dos atributos, apenas a variação máxima do total de atributos, h , e, para o cálculo da frequência dos objetos, a presença ou não do objeto. Assim, em cada iteração t , adotando-se o mecanismo de Laplace, a soma de atributos e frequência de objetos alocados ao grupo C_j é dada, respectivamente, por $S_j^* = S_j + \text{Lap}\left(\frac{h}{\epsilon_t/2}\right)$ e $|C_j^*| = |C_j| + \text{Lap}\left(\frac{1}{\epsilon_t/2}\right)$. O centroide do grupo j foi então dado por $c_j^* = \frac{S_j^*}{|C_j^*|} \forall j \in \{1, \dots, k\}$. O mecanismo de Laplace foi identificado por Laplace* quando essa segunda abordagem foi adotada no cálculo da sensibilidade.

Outra alternativa para tratar essa questão consiste em adotar um mecanismo truncado [44, p.116-117]. No mecanismo de Laplace truncado, utiliza-se a mesma distribuição, porém com as caudas limitadas a um intervalo $[-r, r]$. Isso é implementado através de *rejection sampling*, onde o ruído é selecionado da distribuição de Laplace e o valor é aceito apenas se sua magnitude for menor ou igual ao raio r , o que garante que o erro absoluto introduzido pelo mecanismo nunca exceda r . Entretanto, o truncamento altera as garantias de privacidade. Enquanto o mecanismo de Laplace padrão satisfaz $(\epsilon, 0)$ -PD, o truncamento cria regiões do espaço de saída que podem ocorrer para um conjunto de dados, mas não para um conjunto adjacente a ele. Nesses casos, um adversário poderia distinguir perfeitamente entre as duas bases de dados, o que corresponde a um evento de perda de privacidade infinita. A probabilidade de ocorrência desse evento é limitada pelo parâmetro δ , de modo que o mecanismo satisfaz (ϵ, δ) -PD aproximada. Intuitivamente, δ corresponde à probabilidade associada às regiões das caudas removidas da distribuição original.

APÊNDICE C – Metodologia Empregada na Revisão da Literatura

A partir da hipótese formulada e dos objetivos dela derivados, apresentados na Seção 1.1, uma revisão sistemática da literatura (RSL) foi planejada com base na metodologia proposta em [123]. As plataformas Parsifal [78], implementada para este fim, e Zotero¹ foram adotadas como ferramentas de apoio.

No presente capítulo, a Seção C.1 descreve todas as etapas e critérios adotados na RSL apresentada no Capítulo 3, enquanto a Seção C.2 relata o seu processo de execução.

C.1 Planejamento

A etapa de planejamento consistiu na definição e detalhamento do protocolo e dos critérios de avaliação de qualidade. O protocolo, descrito na seção C.1.1, tem como elementos:

- Objetivo;
- PICOC²;
- Perguntas de pesquisa;
- *String* de busca;
- Fontes (bases ou repositórios de conteúdo científico); e
- Critérios de seleção.

¹<https://www.zotero.org/>

²PICOC é um acrônimo composto pelos termos *population* (quem), *intervention* (o quê ou como), *comparison* (comparado a que), *outcomes* (o que se está tentando alcançar ou aprimorar) e *context* (em que tipo de organização ou circunstâncias). É um método usado para descrever os cinco elementos de uma pergunta "pesquisável" que teve origem na área médica e tem sido adotado por outras áreas, como a computação [200].

Os critérios de avaliação de qualidade, descritos na Seção C.1.2, compreendem um conjunto de perguntas ponderadas que permitem avaliar, objetivamente, os trabalhos selecionados.

C.1.1 Protocolo

Como objetivo, foi estabelecido "identificar abordagens para alocação/determinação de *privacy budget* em aplicações de privacidade diferencial para problemas de agrupamento".

Para o PICOC, os seguintes termos foram definidos:

- População³: métodos de agrupamento diferencialmente privados *differentially private clustering methods*;
- Intervenção⁴: estratégia de alocação de *privacy budget* (*privacy budget allocation strategy*);
- Comparação⁵: não aplicável;
- Resultado⁶ (*Outcome*): alocação ótima de *privacy budget* (*privacy budget optimal allocation*);
- Contexto⁷: agrupamento de dados sensíveis (*clustering sensitive data*).

Como perguntas de pesquisa, a RSL realizada buscou responder:

- Quais são as abordagens disponíveis para otimização da alocação de *privacy budget* no contexto do problema de agrupamento diferencialmente privado?
- A aplicação de técnicas de otimização nesse contexto é promissora?

³A população na qual as evidências são coletadas, ou seja, qual grupo de pessoas, programas ou empresas são de interesse para a revisão?

⁴A intervenção aplicada no estudo empírico, ou seja, qual tecnologia, ferramenta ou procedimento está sendo estudado?

⁵A comparação com a qual a intervenção é comparada, ou seja, como o tratamento de controle é definido? Em particular, a intervenção "placebo" é crítica, pois "não usar a intervenção" geralmente não é uma ação válida em engenharia de software.

⁶Os resultados do experimento não devem ser apenas estatisticamente significativos, mas também significativos do ponto de vista prático. Por exemplo, pode não ser interessante que um resultado seja 10% melhor em algum aspecto se ele consumir 10 vezes o tempo.

⁷O contexto do estudo deve ser definido, ou seja, uma visão ampliada da população, incluindo se ele é conduzido na academia ou na indústria, em qual segmento da indústria e também os incentivos para os participantes.

A *string* de busca foi definida a partir do PICOC e refinada considerando os termos presentes em um pequeno conjunto preliminar de trabalhos de referência, sendo adotada a *string* (cluster OR clustering) AND ("differential privacy" OR "differentially private") AND "privacy budget".

No que se refere às fontes, três bases foram adotadas, quais sejam: IEEE Xplore⁸, ScienceDirect⁹ e SpringerLink¹⁰.

Finalmente, foram adotados os seguintes critérios de seleção:

- Critérios de inclusão:
 - Estudo aborda alocação de *privacy budget* no contexto de privacidade diferencial para o problema de agrupamento.
- Critérios de exclusão:
 - Estudo aborda alocação de *privacy budget* fora do contexto do problema de agrupamento diferencialmente privado.
 - Estudo adota agrupamento apenas como etapa preliminar à aplicação de privacidade diferencial.
 - Estudo não aborda o problema de agrupamento diferencialmente privado.
 - Estudo em idioma diferente de português, espanhol ou inglês.
 - Literatura cinzenta.

C.1.2 Critérios de avaliação de qualidade

Como critérios de avaliação de qualidade, foram adotadas seguintes perguntas:

1. O objetivo da pesquisa está claramente descrito?
2. Alguma abordagem para alocação de *privacy budget* foi proposta?
3. Há comparação entre abordagens de alocação de *privacy budget*?
4. Há comparação entre métodos de agrupamento?

⁸<https://ieeexplore.ieee.org/Xplore/home.jsp>

⁹<https://www.sciencedirect.com/>

¹⁰<https://link.springer.com/>

5. Foi realizado experimento computacional com resultados e análises robustas que permitam fazer inferências?
6. Alguma (meta)heurística foi considerada?
7. Publicação tem 5 ou mais citações¹¹?

No que se refere aos pesos, estes foram atribuídos da seguinte forma:

- Se a resposta para a pergunta foi "sim", atribuído peso 1,0;
- Se a resposta para a pergunta foi "parcial/parcialmente", atribuído peso 0,5;
- Se a resposta para a pergunta foi "não/nenhum", atribuído peso 0.

Complementarmente, foi definido como critério mínimo de qualidade um peso superior a 2, i.e., foram desconsiderados os artigos selecionados cujo somatório dos pesos atribuídos foi igual ou inferior a 2.

C.2 Condução

No que diz respeito à condução da RSL, esta se deu nas seguintes etapas:

1. Busca;
2. Importação dos dados na plataforma Parsifal;
3. Seleção dos estudos;
4. Avaliação da qualidade;
5. Extração dos dados.

A etapa de busca foi realizada de forma manual nos *websites* das fontes selecionadas. Adicionalmente à *string* de busca, foram adotados alguns filtros a fim de

¹¹Para avaliação deste critério foi considerado o número de citações apresentado na plataforma Google Scholar (<https://scholar.google.com/>).

harmonizar os resultados obtidos. No IEEE Xplore, foram incluídas apenas publicações do tipo *conference* ou *journal*. No ScienceDirect, foram incluídas apenas publicações do tipo *review article* ou *research article*. No SpringerLink, foram incluídas apenas publicações do tipo *article* ou *conference paper*. Foram ainda incluídos manualmente alguns trabalhos considerados especialmente relevantes publicados em outras fontes [6, 14, 23, 177, 154, 180, 179, 135, 208, 63, 93, 214, 127, 206, 217, 110].

No que se refere ao período de referência, foram considerados todos os trabalhos publicados a partir de 2005¹² até abril de 2026, data da última atualização do levantamento realizado.

Tabela C.1: Número de trabalhos identificados na RSL.

Fonte	Tipo de publicação	Intervalo de data	Quantidade	Percentual
IEEE Xplore	conference, journal	2016 a 2026	52	4,92%
ScienceDirect	review/research article	2013 a 2026	492	46,50%
SpringerLink	article, conference paper	2012 a 2026	498	47,07%
Manual		2005 a 2024	16	1,51%
Total			1058	

Tabela elaborada pelo autor.

Todos os trabalhos identificados na RSL, conforme apresentado na Tabela C.1, foram importados a partir da plataforma Parsifal. Seguiu-se então com uma seleção preliminar com base no título, resumo e palavras-chave, na qual os trabalhos foram classificados segundo os critérios estabelecidos no protocolo da RSL (Seção C.1.1). Essa etapa resultou na seleção de 96 trabalhos classificados no critério de inclusão e, portanto, potencialmente relevantes para a pesquisa desenvolvida. Predominantemente, os trabalhos excluídos nessa etapa apresentam estudos que utilizam métodos de agrupamento e/ou privacidade diferencial, mas não abordam o problema de agrupamento diferencialmente privado, i.e., a aplicação do método de agrupamento tem caráter complementar.

Os trabalhos classificados no critério de inclusão foram então importados na plataforma Zotero¹³ a fim de realizar o refinamento da seleção preliminar¹⁴. Nessa etapa foram lidas as seções de introdução e conclusão de todos os trabalhos e, eventualmente, alguma seção adicional a fim de dirimir dúvidas relativas à relevância do trabalho para a pes-

¹²Em 2005 foi publicado o primeiro trabalho aplicando privacidade diferencial ao problema de agrupamento [23].

¹³<https://www.zotero.org/>

¹⁴A seleção preliminar resultou em 96 trabalhos classificados no critério de inclusão, porém 20 deles foram excluídos na etapa de refinamento da seleção preliminar em função de o texto completo não estar disponível.

quisa desenvolvida, resultando em 43 trabalhos identificados como aderentes à pesquisa desenvolvida.

Na etapa de refinamento foram predominantemente excluídas abordagens que utilizaram métodos de agrupamento em algum estágio complementar, mas não trataram o problema de agrupamento diferencialmente privado. Por exemplo, há abordagens baseadas em aprendizado federado (do inglês, *Federated Learning*, descrito na Seção A.1.4) nas quais, em vez da construção de um único modelo global resultante, foram aplicados métodos de agrupamento para identificar clientes similares no que diz respeito aos dados armazenados, a fim de construir múltiplos modelos subglobais agregando os modelos locais de maneira seletiva, obtendo maior poder de generalização do que o modelo global único [193, 101, 120]. Abordagens baseadas em aprendizado federado tipicamente adotam PD para proteger de ataques os parâmetros dos modelos locais na ocasião de compartilhamento entre clientes ou entre clientes e o servidor central. Tais trabalhos estão fora do escopo desta pesquisa.

Após a etapa de refinamento, os 45 trabalhos selecionados foram submetidos à etapa de seleção final, que contou com a leitura completa e com a avaliação da qualidade, conforme critérios descritos na Seção C.1.2. Finalmente, os trabalhos que atenderam aos critérios de qualidade propostos foram classificados quanto ao modelo de PD adotado, ao tipo método de agrupamento aplicado e ao tipo de processamento realizado, se distribuído ou centralizado. Esses trabalhos se encontram descritos na Seção 3.1.

APÊNDICE D – Formulação alternativa para o modelo iterativo

O problema de alocação de *privacy budget*, no contexto do k-means, consiste em determinar uma sequência de valores $[\epsilon_1, \dots, \epsilon_t, \dots, \epsilon_T]$, de forma a garantir que o algoritmo satisfaça ϵ -DP, maximizando a utilidade, i.e., maximizando a qualidade da solução de agrupamento em termos de uma métrica ou índice de validação previamente especificado¹.

Adotando o *Normalized Intra-cluster Variance* (NICV) como métrica de qualidade das soluções, o problema pode ser formulado como um problema de otimização, conforme apresentado na Seção 4.1.

$$\min f_{\text{obj}} = \frac{1}{n} \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j^*\|^2 \quad (\text{D.1})$$

Sujeito a:

$$\sum_{t=1}^T \epsilon_t \leq \epsilon \quad (\text{D.2})$$

Tais que:

- Eq. D.1 é a função objetivo.
- Eq. D.2 é a restrição que garante ϵ -DP.

No entanto, durante a revisão da literatura, foram identificadas restrições adicionais, definidas pelas Equações D.3, D.4 e D.5.

¹Tipicamente, seguindo o que está preconizado na literatura, são utilizados o índice de Calinski-Harabasz (CH) [30] ou *Normalized Intra-cluster Variance* (NICV).

$$\epsilon_t \geq \epsilon_m, \forall t \in \{1, \dots, T\} \quad (\text{D.3})$$

$$\epsilon_t \geq \epsilon_{t+1}, \forall t < T \quad (\text{D.4})$$

$$|\epsilon_t - \epsilon_{t+1}| \leq \xi, \forall t < T \quad (\text{D.5})$$

tais que:

- Eq. D.3 é a restrição de garantia de convergência, onde ϵ_m é o *privacy budget* mínimo a ser alocado a cada iteração.
- Eq. D.4 é a restrição que garante que ϵ_t será decrescente ao longo das iterações.
- Eq. D.5 é a restrição que determina a regularização entre iterações consecutivas.

Apenas a restrição D.2 é intrínseca ao problema. As demais restrições se referem a garantir ou priorizar aspectos identificados como relevantes no que diz respeito a preservar a utilidade da solução de agrupamento obtida, que terá a sua qualidade garantida mediante a minimização do NICV, representado na função objetivo. Mais especificamente, a restrição D.3 estabelece um valor mínimo de *privacy budget* a ser alocado a cada iteração², a fim de garantir estabilização da solução na iteração T [179]. A restrição D.4 está baseada nas conclusões obtidas em algumas pesquisas [54, 70, 216], indicando a efetividade de adotar-se uma estratégia de alocação decrescente em relação a t , a fim de priorizar as primeiras iterações. Por fim, a restrição D.5 diz respeito a evitar uma variação abrupta (superior a ξ) na magnitude do ruído adicionado entre iterações consecutivas [135]. A determinação do parâmetro de regularização ξ deve ser investigada através de experimentos.

O número máximo de iterações e, conseqüentemente, um valor mínimo para o *privacy budget* alocado em cada iteração foi determinado analiticamente em [179, p.28-29], dado por³:

$$\epsilon_m = \sqrt{\frac{250h(k(h+1))^2}{n^2}} \quad (\text{D.6})$$

sendo k correspondente ao número de grupos, n ao número total de objetos e h ao número de atributos.

²Quanto menor o *privacy budget* de cada iteração, maior é a magnitude do ruído adicionado, dificultando a estabilização ou convergência do algoritmo e, conseqüentemente, a obtenção da solução [129, p.5].

³A dedução completa é apresentada no Apêndice B.3.6, p.160.

Uma vez determinado ϵ_m , é possível determinar T em função desse parâmetro:

$$T = \begin{cases} \left\lfloor \frac{\epsilon}{\epsilon_m} \right\rfloor, & \text{se } \epsilon > 2\epsilon_m \\ 2, & \text{caso contrário} \end{cases} \quad (\text{D.7})$$

No segundo caso, o *privacy budget* total é insuficiente. Se ele for distribuído por mais de duas iterações, o ruído adicionado em cada iteração dominará a melhoria do centroide [179, p.29], o que apenas resultaria em uma distorção cada vez maior. Nesse caso, são executadas apenas duas iterações, com $\epsilon_t = \frac{\epsilon}{2}, t \in \{1, 2\}$, para fins de comparação.

Finalmente, a fim de converter vetores de chaves aleatórias em soluções viáveis para o problema, um procedimento de decodificação foi proposto.

Na formulação apresentada nesta Seção, a restrição D.2 pode ser garantida através da normalização de u , a fim de que a i -ésima posição de u represente a proporção de ϵ alocada a iteração i , ou seja,

$$u_i^N \leftarrow \frac{u_i}{\sum_{i=1}^T u_i}, \forall i \in \{1, \dots, T\}$$

Por exemplo, considerando os parâmetros $\epsilon = 10$, $\epsilon_m = 1,4$ e $T = 7$, o vetor de chaves aleatórias $u = [0,549, 0,715, 0,603, 0,545, 0,424, 0,646, 0,438]$ corresponde ao vetor normalizado $u^N = [0,1401, 0,1824, 0,1538, 0,1390, 0,1082, 0,1648, 0,1117]$ que, por sua vez, corresponde à solução (valores ϵ_t , $t = 1, \dots, T$) $s = \epsilon \cdot u^N = [1,400, 1,825, 1,538, 1,391, 1,081, 1,648, 1,117]$. Uma vez ordenado de forma decrescente, o vetor s atende simultaneamente às restrições D.2 e D.4.

Essa construção, no entanto, não necessariamente garante o cumprimento da restrição D.3, na qual $\epsilon_t \geq \epsilon_m$, $\forall t \in \{1, \dots, T\}$. Para tal foi proposta abordagem que consiste primeiramente em calcular a proporção excedente de ϵ , p_ϵ , a partir dos parâmetros ϵ_m , ϵ e T , fazendo:

$$p_\epsilon = 1 - \frac{\epsilon_m}{\epsilon} * T = 1 - \frac{1,4}{10} * 7 = 1 - 0,98 = 0,02$$

O vetor $u^N = \frac{u}{\sum u}$ é então utilizado para definir apenas como essa proporção excedente será distribuída, gerando o vetor de chaves aleatórias regenerado u_{reg}^N :

$$u_{reg}^N = \frac{\epsilon_m}{\epsilon} + u^N \cdot p_\epsilon$$

$$u_{reg}^N = 0,14 + [0,140, 0,182, 0,154, 0,139, 0,108, 0,165, 0,112] \times 0,02 \quad (D.8)$$

$$= 0,14 + [0,0028, 0,0036, 0,0031, 0,0028, 0,0022, 0,0033, 0,0022] \quad (D.9)$$

$$= [0,1428, 0,1436, 0,1431, 0,1428, 0,1422, 0,1433, 0,1422] \quad (D.10)$$

No que diz respeito à garantia restrição [D.4](#), o vetor u é previamente ordenado de forma decrescente em ambos os decodificadores. Finalmente, o valor de *privacy buget* a ser alocado é obtido multiplicando cada posição de u^N pelo *privacy budget* total, i.e.,

$$\epsilon_t = \epsilon \cdot u_{reg_t}^N, \forall t \in \{1, \dots, T\}$$

Portanto, o decodificador converte o vetor de chaves aleatórias u em um vetor que representa uma solução viável s para o problema, sendo o valor de cada posição de s correspondente à parcela de ϵ alocada à respectiva iteração. Por exemplo, seja $\epsilon = 10$, $\epsilon_m = 1,4$ e $T = 7$, o vetor de chaves aleatórias $u = [0,715, 0,646, 0,603, 0,549, 0,545, 0,438, 0,424]$ é convertido na solução viável $s = \epsilon \cdot u_{reg}^N = [1,576, 1,424, 1,400, 1,400, 1,400, 1,400, 1,400]$, atendendo simultaneamente as restrições [D.2](#), [D.3](#) e [D.4](#).

Dessa forma, a saída da etapa de decodificação é o vetor s , de comprimento T , contendo as parcelas (ϵ_t) do *privacy budget* total alocadas às iterações $t \in \{1, \dots, T\}$ do DPk-means.

APÊNDICE E – Abordagem proposta para o modelo não interativo

No modelo interativo de PD, o k-means diferencialmente privado tem como o principal desafio a alocação do *privacy budget* ao longo das iterações realizadas, de maneira a não comprometer a convergência do método e a utilidade da solução final, como descrito no Capítulo 4. A quantidade de ruído aumenta com o número de iterações, inviabilizando o uso de critérios de parada mais refinados, sendo geralmente necessário estabelecer o número máximo de iterações previamente. Como alternativa, o uso de *coresets* (Seção E.1) se mostra promissor, pois concentra a garantia de privacidade em um subconjunto de objetos e não no algoritmo em si. Isso permite aplicar algoritmos com múltiplas iterações ou diferentes estratégias de parada sem comprometer a privacidade, oferecendo mais flexibilidade e utilidade.

E.1 *Coreset*

Um *coreset* de um conjunto de pontos (ou objetos) D é um subconjunto, geralmente ponderado, desses pontos, que captura propriedades geométricas de D e permite obter respostas aproximadas para consultas em D [72, p.361]. Mais especificamente, o conceito de *coreset* clássico [4] consiste em, dado um conjunto de consultas Q e um conjunto de microdados D , construir um ϕ -*coreset* $D' \subset D$, tal que qualquer consulta $q \in Q$ que possa ser respondida por $q(D)$, possa ser respondida de maneira aproximada por $q(D')$, com um erro relativo dentro do intervalo¹ $(1 \pm \phi)$, i.e, $(1 - \phi) \cdot q(D) \leq q(D') \leq (1 + \phi) \cdot q(D)$.

¹Tradicionalmente, a produção acadêmica sobre *coresets* adota ϵ para denotar o fator de aproximação associado à preservação do custo (função objetivo). No entanto, no contexto de privacidade diferencial, ϵ é amplamente utilizado para representar o *privacy budget*. Assim, neste trabalho, optou-se por utilizar ϕ para representar o fator de aproximação, evitando ambiguidades e melhorando a legibilidade da notação matemática.

E.1.1 (k, z) -clustering

A aplicação de *coresets* ao problema de agrupamento adota uma formalização alternativa do problema. No contexto de *coresets*, a abordagem baseada em protótipos para o problema de agrupamento clássico pode ser definida como o problema (k, z) -clustering [108, p.3]. Considere o conjunto de dados $D = \{x_1, \dots, x_i, \dots, x_n\}$, $x_i \in \mathbb{R}^h \forall i \in \{1, \dots, n\}$. O objetivo é encontrar o conjunto C de k pontos, denominados centros (protótipos), tal que a soma das distâncias de cada objeto para o respectivo centro mais próximo seja minimizada, conforme Eq. E.1.

$$\text{custo}_z(D, C) := \sum_{x \in D} \min_{c \in C} \{d_{L2}^z(x, C)\} \quad (\text{E.1})$$

A distância d_{L2}^z corresponde à distância euclidiana elevada à potência z , $z \geq 1$, i.e., $d_{L2}(x_i, c_j) = \sqrt{\sum_{l=1}^h (x_{il} - c_{jl})^2}$. Na Eq. E.1, para cada $x \in D$, é somada a menor distância entre x e os k centros $c \in C$. Essa formulação abrange problemas clássicos como o k-medoids, quando $z = 1$, e o k-means², quando $z = 2$ [108, p.3].

E.1.2 *Coreset* aplicado ao problema de agrupamento

No contexto do problema de agrupamento clássico, *coresets* são usados para obter soluções de agrupamento cujo valor da função objetivo é aproximadamente igual ao da solução ótima em grandes conjuntos de dados³. De forma geral, um *coreset* é um subconjunto de objetos do conjunto original que permite aproximar a função objetivo do conjunto original para qualquer escolha de centros (protótipos) [108, p.3]. Para isso, pesos são atribuídos aos objetos selecionados para compor o *coreset*. Um estudo apresentado em [47] ilustra essa abordagem adotando amostragem aleatória simples na construção de *coresets* para quatro problemas de agrupamento, quais sejam: k-median, k-means, soma mínima e k-median balanceado.

Mais especificamente, a aplicação de *coresets* para o problema (k, z) -clustering [98, 108, 42, 40] consiste em, dado um conjunto D com n objetos ($D = \{x_1, \dots, x_i, \dots, x_n\}$, $x_i \in \mathbb{R}^h \forall i \in \{1, \dots, n\}$), um ϕ -coreset pode ser definido como [73, p.570] um subconjunto ponderado $D' \subseteq D$ com pesos $w : D' \rightarrow \mathbb{R}_+$ tal que, para qualquer conjunto C de k centros,

²Soma do quadrado das distâncias: $\sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - c_j\|^2$.

³O objetivo é reduzir o custo computacional de algoritmos de agrupamento aplicando-os a um conjunto reduzido de objetos, porém de maneira que a solução produzida possa ser generalizada para o conjunto completo.

o custo da solução de agrupamento em D' aproxima o custo correspondente em D dentro de um fator $(1 \pm \phi)$, com $\phi \in (0, 1)$ [108, p.3].

$$(1 - \phi) \cdot \text{custo}_z(D, C) \leq \sum_{x \in D'} w(x) \cdot \min_{c \in C} \{d_{L_2}^z(x, C)\} \leq (1 + \phi) \cdot \text{custo}_z(D, C) \quad (\text{E.2})$$

E.1.3 *Coreset* diferencialmente privado

O conceito de *coreset* diferencialmente privado [72] consiste em produzir um *coreset* que satisfaz ϵ -PD, i.e., em que os valores dos atributos dos registros selecionados foram perturbados por um mecanismo de PD, onde ϵ representa o *privacy budget* adotado. O *coreset* resultante pode ser explorado livremente pelo analista sem a necessidade do intermédio do curador, representando portanto o modelo não interativo ou *offline*, descrito na seção 2.4.2, que tende a oferecer resultados com menor perturbação [59, p.12].

A principal limitação dessa estratégia reside na necessidade de conhecer previamente o conjunto de consultas Q , o que inviabiliza⁴ adotá-la nos casos em que se pretende que o *coreset* sirva a aplicações ou análises mais gerais. No entanto, é uma estratégia adequada no caso de aplicações específicas, onde as consultas a serem realizadas já são conhecidas, como nos algoritmos de agrupamento⁵. Ademais, uma vez que o *coreset* satisfaz ϵ -PD, não se faz mais necessária a alocação de frações de ϵ a cada iteração do algoritmo de agrupamento, permitindo não apenas a execução de um número indeterminado de iterações, mas também múltiplas execuções do algoritmo.

O *coreset* diferencialmente privado tem duas diferenças relevantes em relação à abordagem clássica. A primeira delas diz respeito ao tamanho do *coreset* (i.e., seu número de objetos). Na abordagem clássica, a privacidade não é uma preocupação, e a motivação para o uso de *coresets* é o ganho de eficiência (redução do custo de processamento) obtido com o uso de um conjunto reduzido de objetos, às custas de uma imprecisão controlada, definida pelo parâmetro ϕ . Por outro lado, um *coreset* diferencialmente privado pode ter tamanho até mesmo maior que o do conjunto original [72, p.362], uma vez que privacidade

⁴Não é possível produzir conjuntos de microdados diferencialmente privados de uso geral [57, 72], i.e., para um conjunto de consultas Q indeterminado. No entanto, em [24] foi comprovada a existência de microdados diferencialmente privados para consultas baseadas em intervalo, como o número de indivíduos dentro de um determinado intervalo de idade ou renda [72, p.362].

⁵No k-means, por exemplo, as consultas realizadas pelos algoritmos se referem a um conjunto C de k centroides, cuja saída é a soma dos quadrados das distâncias de cada objeto no conjunto de dados até seu centroide mais próximo em C .

e utilidade passam a ser os objetivos⁶, i.e., a principal preocupação não é necessariamente minimizar o número de pontos do *coreset*, mas sim garantir a privacidade diferencial minimizando o erro de aproximação [74, p.7]. Outra diferença importante reside no fato de que *coresets* diferencialmente privados, além do fator de aproximação multiplicativo ϕ , exibem também um erro aditivo [72, p.362], tipicamente denotado por β .

Sendo assim, um esquema (ϵ, ϕ, β) -*coreset* [72, p.363] para um conjunto de consultas Q é um algoritmo que tem como entrada um conjunto de n pontos (ou objetos) D e, como saída, um conjunto $A(D) = D'$ de pares de pontos e respectivos pesos, de forma que:

$$(1 - \phi) \cdot q(D) - \beta \leq q(A(D)) \leq (1 + \phi) \cdot q(D) + \beta$$

Onde:

- D representa o conjunto de microdados original.
- q representa uma consulta que se deseja aplicar a D .
- A representa o mecanismo que produz *coresets* diferencialmente privados.
- $A(D) = D'$ representa o *coreset* diferencialmente privado resultante da aplicação de A em D .
- ϵ representa o *privacy budget*.
- ϕ representa o fator de aproximação multiplicativo.
- β representa a parcela de aproximação aditiva.

E.1.4 *Coreset* diferencialmente privado para o problema de agrupamento

No contexto do problema de agrupamento, considerando a formulação (k, z) -*clustering* apresentada na seção E.1.1 [72, 74, 203], uma consulta q é definida por um conjunto C de protótipos k e pelo parâmetro z , que corresponde ao expoente aplicado à distância euclidiana⁷. Quando aplicada ao conjunto de dados D , a consulta $q(D, C, z)$ tem como

⁶Se a redução do número de objetos for também relevante, é possível gerar um *coreset* convencional a partir de um *coreset* diferencialmente privado. Não há implicações à privacidade nessa estratégia, mas haverá um custo em termos de utilidade.

⁷Para o problema k-median, $z = 1$, i.e., $\sum_{x \in D} \min_{c \in C} \{\|x - C\|\}$. Para o problema k-means, $z = 2$, i.e., $\sum_{x \in D} \min_{c \in C} \{\|x - C\|^2\}$.

saída $\text{custo}_z(D, C)$ (Eq. E.1). A privacidade diferencial é integrada através do mecanismo A , com parâmetro de *privacy budget* ϵ , conforme apresentado na seção E.1.3, de forma que:

$$(1 - \phi) \cdot q(D, C, z) - \beta \leq q(A(D), C, z) \leq (1 + \phi) \cdot q(D, C, z) + \beta$$

ou, conforme notação da Eq. E.2,

$$(1 - \phi) \cdot \text{custo}_z(D, C) + \beta \leq \sum_{x \in D'} w(x) \cdot \min_{c \in C} \{d_{L_2}^z(x, C)\} \leq (1 + \phi) \cdot \text{custo}_z(D, C) + \beta$$

Onde:

- D representa o conjunto de microdados original.
- C representa o conjunto de k centros (ou protótipos).
- D' representa o *coreset* diferencialmente privado resultante da aplicação do mecanismo A em D .
- x é um objeto de D' .
- $w(x)$ é o peso associado ao objeto x .
- ϕ representa o fator de aproximação multiplicativo.
- β representa a parcela de aproximação aditiva.
- $d_{L_2}(x, c)$ representa a distância euclidiana entre o objeto x e o centro c .
- z representa o expoente aplicado à distância euclidiana.

E.2 Formulação do problema

O problema de construir um ϕ -*coreset* no contexto do problema de agrupamento clássico baseado em protótipos⁸ [98, 108, 42, 40] consiste em definir um subconjunto $X' \subset X$, $X = \{x_1, \dots, x_i, \dots, x_n\}$ ($x_i \in \mathbb{R}^h \forall i \in \{1, \dots, n\}$), e atribuir pesos $w(x)$ aos objetos de X' , de forma que, para qualquer conjunto C de k centros (protótipos), os valores da função de custo calculados a partir de X e X' , respeitem a desigualdade definida por $\text{custo}(X', C) \leq (1 \pm \phi) \cdot \text{custo}(X, C)$. Com isso, soluções de agrupamento produzidas para X' serão suficientemente próximas das produzidas para X , no que se refere aos valores da

⁸Seção E.1.2.

função de custo. Mais especificamente, o objetivo é que, para qualquer conjunto arbitrário de k centros, o valor da função de custo da solução obtida para um (k, ϕ) -coreset esteja dentro da tolerância de $(1 \pm \phi)$ vezes o valor da função de custo da solução obtida para o conjunto de dados original. Para tal, os objetos de X' têm um peso associado.

No contexto do agrupamento diferencialmente privado, o objetivo não é mais reduzir o número de objetos [72, p.362], passando a priorizar privacidade e utilidade. Outra diferença importante reside no fato de que em *coresets* diferencialmente privados, além do fator de aproximação multiplicativo ϕ , incluem também um erro aditivo [72, p.362], tipicamente denotado por β . Para formalizar essas garantias de aproximação no contexto da privacidade diferencial, pode-se adotar a formulação (k, z) -clustering⁹ [72, 74, 203] e considerando uma consulta q definida por um conjunto C de k centroides e pelo parâmetro z , que corresponde ao expoente aplicado à distância euclidiana¹⁰. Quando aplicada ao conjunto de dados X , a consulta $q(X, C, z)$ tem como saída $\text{custo}_z(X, C)$ (Eq. E.1, Apêndice E.1.1, p.179). A privacidade diferencial é integrada através do mecanismo A , com parâmetro de *privacy budget* ϵ , de forma que¹¹:

$$(1 - \phi) \cdot \text{custo}_2(X, C) + \beta \leq \text{custo}_2(X', C) \leq (1 + \phi) \cdot \text{custo}_2(X, C) + \beta$$

Onde:

- $\text{custo}_2(X, C) = \sum_{x \in D} \min_{c \in C} \{d_{L2}^2(x, C)\}$
- $\text{custo}_2(X', C) = \sum_{x \in X'} w(x) \cdot \min_{c \in C} \{d_{L2}^2(x, C)\}$
- X representa o conjunto de microdados original.
- C representa o conjunto de k centros (ou protótipos).
- X' representa o *coreset* diferencialmente privado resultante da aplicação do mecanismo A em X .
- x é um objeto de X' .
- $w(x)$ é o peso associado ao objeto $x \in X'$.
- ϕ representa o fator de aproximação multiplicativo.

⁹Ver Apêndice E.1.1 (p.179).

¹⁰Para o problema k-means, $z = 2$, i.e., $\sum_{x \in D} \min_{c \in C} \{\|x - C\|^2\}$.

¹¹Ver Apêndice E.1.3 (p.180).

- β representa a parcela de aproximação aditiva.
- $d_{L2}(x, c)$ representa a distância euclidiana entre o objeto x e o centroide c .

Considerando estas definições, o problema pode ser abordado como um problema de otimização, minimizando a perda de privacidade, dada pela Eq. E.3, como formalizado a seguir.

$$\min f_{\text{obj}} = L(X', X, y) \quad (\text{E.3})$$

Sujeito a:

$$L(X', X, y) \leq \epsilon \quad (\text{E.4})$$

$$\text{custo}_2(X', C) \geq (1 - \phi) \cdot \text{custo}_2(X, C) + \beta \quad (\text{E.5})$$

$$\text{custo}_2(X', C) \leq (1 + \phi) \cdot \text{custo}_2(X, C) + \beta \quad (\text{E.6})$$

$$w(x) \geq 0 \quad \forall x \in X' \quad (\text{E.7})$$

Em que:

- Eq. E.3 é a função objetivo a ser minimizada.
- Eq. E.4 é a restrição que garante ϵ -DP.
- Eq. E.5 e E.6 garantem a aproximação desejada em termos dos erros multiplicativo (ϕ) e aditivo (β).
- Eq. E.7 é a restrição de não negatividade dos pesos.

A função objetivo (Eq. E.3) é definida com base na perda de privacidade¹² (do inglês, *privacy loss*). No contexto de *coresets*, ela é quantificada por uma razão entre as probabilidades de dois eventos [44, p.28], conforme indicado na Eq. E.8. Seja y o resultado de uma consulta, o primeiro evento, $q(X') = y$, indica que o resultado y foi obtido a partir da aplicação da consulta¹³ q em X' e o segundo, $q(X) = y$, indica que foi obtido a partir de X .

¹²Ver definição 9, p.25.

¹³No k-means, por exemplo, as consultas realizadas pelos algoritmos se referem a um conjunto C de k centroides, cuja saída é a soma dos quadrados das distâncias de cada objeto no conjunto de dados até seu centroide mais próximo em C .

$$L(X', X, y) = \ln \left(\frac{P[q(X') = y]}{P[q(X) = y]} \right) \quad (\text{E.8})$$

As probabilidades de ambos os eventos devem ser suficientemente próximas a fim de que um intruso, de posse do resultado y , não possa inferir a partir de qual fonte, X ou X' , esse resultado foi obtido. Consequentemente, não sendo possível fazer nenhuma inferência suficientemente precisa sobre a composição, i.e., quais objetos estão presentes (ou ausentes), em cada caso. Quanto mais próxima de zero é a razão de chances dada por $L(X', X, y)$, mais forte é a garantia da privacidade. No entanto, a fim de garantir ϵ -DP, é necessário estabelecer a restrição dada pela Eq. [E.4](#), que limita a perda de privacidade em até ϵ , i.e., limita superiormente $L(X', X, y)$ em ϵ .

APÊNDICE F – Análises de complexidade dos algoritmos

Este apêndice apresenta a análise de complexidade de tempo dos algoritmos ILSPD, BRK-GAPD e VNSPD. Para cada método, são discutidos os principais componentes executados ao longo do processo, considerando tanto os operadores específicos de cada metaheurística quanto o custo associado à avaliação da função de avaliação, baseada na execução do algoritmo DPk-means. As análises são apresentadas em termos assintóticos, destacando os fatores dominantes na complexidade total de cada abordagem.

F.1 BRKGAPD

A complexidade de tempo do BRKGAPD pode ser analisada a partir das operações executadas em cada geração da população, sendo:

- $|u|$: tamanho do vetor de solução (número de iterações do DPk-means);
- p : tamanho da população;
- G_{tmax} : número de gerações/ iterações executadas até o limite de tempo t_{max} ;
- $T_{dec}(|u|)$: custo do decodificador;
- T_{fit} : custo da função de avaliação (*fitness*);
- $T_{blend}(|u|)$: custo do operador de *blending*.

Em cada geração, todas as soluções da população são decodificadas e avaliadas. Assim, o custo da etapa de avaliação é dado por:

$$O(p \cdot [T_{dec}(|u|) + T_{fit}])$$

Após a avaliação, a população é ordenada segundo os valores de *fitness*, o que requer $O(p \log p)$. Na etapa de cruzamento, $p - p_e$ descendentes são gerados por meio do operador de *blending* e mutação aleatória. Assumindo operações lineares sobre o vetor de chaves, obtém-se:

$$O(p \cdot T_{blend}(|u|)) = O(p|u|)$$

As atualizações da *q-table*, manipulação do *pool* e verificações de parada possuem custo assintoticamente inferior aos demais componentes. Portanto, o custo de uma geração pode ser expresso por:

$$T_{gen}(|u|) = O(p(T_{dec}(|u|) + T_{fit}) + p \log p + p|u|)$$

Assumindo custo linear para o decodificador e para o operador de *blending*, temos $T_{dec}(|u|) = O(|u|)$ e $T_{blend}(|u|) = O(|u|)$. Considerando que o custo da função de avaliação (*fitness*) domina os demais termos, obtém-se:

$$T_{gen}(|u|) = O(p \cdot T_{fit})$$

Assim, a complexidade temporal total do BRKGAPD é dada por $T_{BRKGAPD}(|u|) = O(G_{tmax} p T_{fit})$.

Uma vez que a função de avaliação (*fitness*) consiste na execução do algoritmo DPk-means, derivado do k-means, cuja complexidade clássica é:

$$T_{fit} = O(nkh|u|)$$

sendo:

- n : número de objetos do conjunto de dados;
- k : número de grupos;
- h : número de atributos dos dados;
- $|u|$: número de iterações executadas pelo DPk-means.

Substituindo T_{fit} na expressão anterior, temos:

$$T_{\text{BRKGAPD}}(|u|) = O(G_{t_{\max}} p n k h |u|)$$

F.2 ILSPD

A complexidade de tempo do algoritmo ILSPD pode ser analisada a partir dos principais componentes executados em cada iteração, sendo:

- $|u|$: tamanho do vetor de solução (número de iterações do DPk-means);
- $G_{t_{\max}}$: número de gerações/ iterações executadas até o limite de tempo $t_{\max}$;
- $T_{RVND}(|u|)$: custo do procedimento RVND;
- $T_{dec}(|u|)$: custo do decodificador;
- T_{fit} : custo da função de avaliação;
- $T_{shake}(|u|)$: custo da etapa de *shaking*.

Em cada iteração da estrutura de repetição principal são realizadas as seguintes operações: atualização dos parâmetros via *Q-Learning*, aplicação do *shaking*, execução do RVND, decodificação e cálculo da função de avaliação. Assim, o custo de uma iteração é dado por:

$$T_{\text{iter}}(|u|) = O(T_{shake}(|u|) + T_{RVND}(|u|) + T_{dec}(|u|) + T_{fit})$$

Como a atualização da *q-table* possui custo constante e o procedimento RVND domina os demais componentes, a complexidade total do algoritmo pode ser expressa por:

$$T_{\text{ILSPD}}(|u|) = O(G_{t_{\max}} \cdot T_{RVND}(|u|))$$

Assumindo um número constante de estruturas de vizinhança e que cada vizinhança avalia até $O(|u|^2)$ movimentos, obtém-se:

$$T_{RVND}(|u|) = O(|u|^2 \cdot (T_{dec}(|u|) + T_{fit}))$$

Assumindo ainda custo linear para o decodificador, $T_{dec}(|u|) = O(|u|)$, e considerando que o custo da função de avaliação domina os demais componentes, temos:

$$T_{RVND}(|u|) = O(|u|^2 \cdot T_{fit})$$

Portanto, a complexidade de tempo do ILSPD é dada por:

$$T_{ILSPD}(|u|) = O(G_{tmax}|u|^2 T_{fit})$$

Uma vez que a função de avaliação consiste na execução do algoritmo DPk-means, derivado do k-means, cuja complexidade clássica é dada por:

$$T_{fit} = O(nkh|u|)$$

em que:

- n : número de objetos do conjunto de dados;
- k : número de grupos;
- h : número de atributos dos dados;
- $|u|$: número de iterações executadas pelo DPk-means.

Substituindo T_{fit} na expressão anterior, temos:

$$T_{ILSPD}(|u|) = O(G_{tmax}|u|^3 nkh)$$

F.3 VN SPD

A complexidade de tempo do algoritmo VN SPD pode ser analisada a partir dos principais componentes executados em cada iteração, sendo:

- $|u|$: tamanho do vetor de solução (número de iterações do DPk-means);
- G_{tmax} : número de gerações/ iterações executadas até o limite de tempo t_{max} ;
- k_{max} : número máximo de estruturas de vizinhança;

- $T_{RVND}(|u|)$: custo do procedimento RVND;
- $T_{dec}(|u|)$: custo do decodificador;
- T_{fit} : custo da função de avaliação;
- $T_{shake}(|u|)$: custo da etapa de *shaking*.

Em cada iteração da estrutura de repetição principal, o algoritmo percorre até $k_{\max}$ estruturas de vizinhança. Para cada valor de k , são realizadas as seguintes operações: aplicação do *shaking*, execução do RVND, decodificação e cálculo da função de avaliação. Assim, o custo de uma iteração é dado por:

$$T_{\text{iter}}(|u|) = O(k_{\max} \cdot [T_{shake}(|u|) + T_{RVND}(|u|) + T_{dec}(|u|) + T_{fit}])$$

Como a atualização da *q-table* possui custo constante e o procedimento RVND domina os demais componentes, a complexidade total do algoritmo pode ser expressa por:

$$T_{VN SPD}(|u|) = O(G_{tmax} \cdot k_{\max} \cdot T_{RVND}(|u|))$$

Assumindo um número constante de estruturas de vizinhança no RVND e que cada vizinhança avalia até $O(|u|^2)$ movimentos, obtém-se:

$$T_{RVND}(|u|) = O(|u|^2 \cdot (T_{dec}(|u|) + T_{fit}))$$

Assumindo ainda custo linear para o decodificador, $T_{dec}(|u|) = O(|u|)$, e considerando que o custo da função de avaliação domina os demais componentes, temos:

$$T_{RVND}(|u|) = O(|u|^2 \cdot T_{fit})$$

Portanto, a complexidade de tempo do VN SPD é dada por:

$$T_{VN SPD}(|u|) = O(G_{tmax} k_{\max} |u|^2 T_{fit})$$

Uma vez que a função de avaliação consiste na execução do algoritmo DPk-means, derivado do k-means, cuja complexidade clássica é dada por:

$$T_{fit} = O(nkh|u|)$$

em que:

- n : número de objetos do conjunto de dados;
- k : número de grupos;
- h : número de atributos dos dados;
- s : número de iterações executadas pelo DPk-means.

Substituindo T_{fit} na expressão anterior, obtém-se:

$$T_{VN\text{SPD}}(|u|) = O(G_{tmax}k_{\max}|u|^3nkh)$$

APÊNDICE G – Resultados e análises complementares

Este apêndice reúne resultados e análises complementares que, embora relevantes para a compreensão mais aprofundada dos experimentos realizados, não integram diretamente a análise principal apresentada no Capítulo 5.

G.1 Análise das combinações de parâmetros dos algoritmos

O experimento contou com a aplicação dos três algoritmos aos 30 conjuntos de dados, considerando 18 cenários, definidos pela combinação dos três mecanismos – Laplace, Laplace* e Normal – e dos valores de *privacy budget* – 0,5, 1, 2, 3, 5, 10. Além disso, foram feitas 10 repetições para cada caso. Com isso, foram geradas 16200 soluções¹.

Preliminarmente, foi realizada uma análise de consistência a fim de verificar se os três algoritmos produziram soluções para todos os conjuntos de dados e em todos os cenários considerados, i.e., todas as combinações de mecanismos e valores de *privacy budget*. Uma vez verificado que sim, procedeu-se à análise dos resultados com base no NICV da solução final de agrupamento obtido pelo algoritmo DPk-means (Seção G.1) e na recompensa (ou *reward*), que orienta o processo de escolha dos parâmetros no Q-Learning (Seção G.2).

Os algoritmos foram também analisados separadamente, a fim de avaliar os resultados em termos do espaço de parâmetros considerado. Para isso, as soluções obtidas foram agregadas nos cenários considerados (mecanismos x valores de *privacy budget*) e foram totalizados o número de etapas (*steps*) do Q-Learning com cada combinação e o número de vezes que essa combinação obteve a melhor solução, em termos de NICV, no respectivo cenário. A razão entre essas duas quantidades representa a proporção de vitórias.

¹O número total de soluções se refere à execução de 3 algoritmos x 3 mecanismos x 6 valores de *privacy budget* x 30 conjuntos de dados x 10 repetições = 16200 soluções.

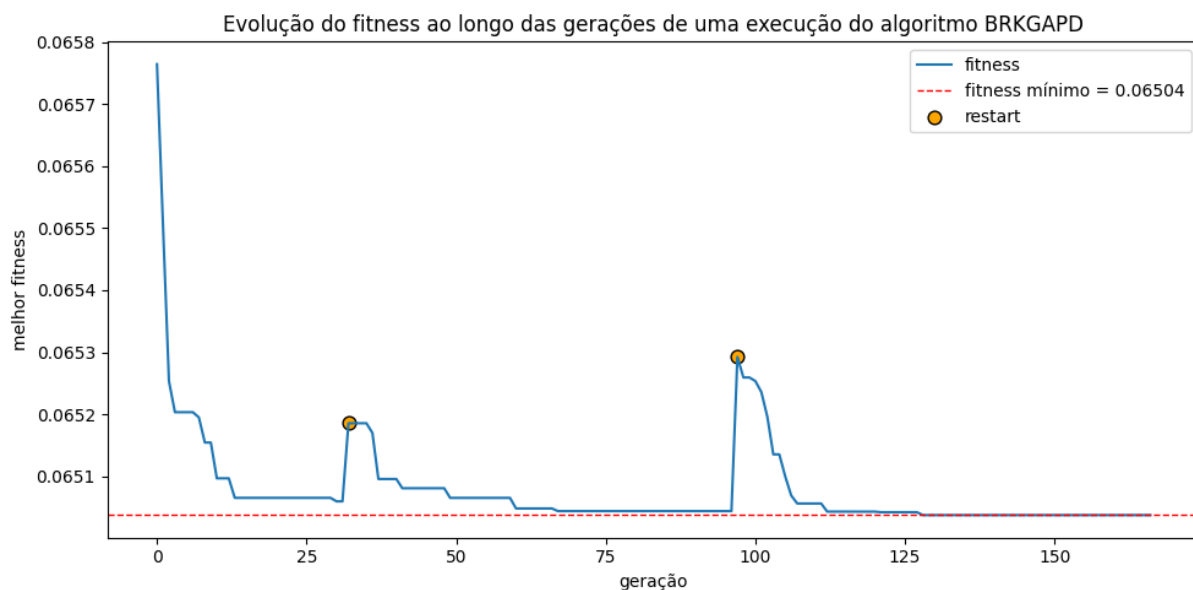


Figura G.1: Evolução do *fitness* ao longo das gerações de uma execução do algoritmo BRKGAPD (Cenário: Laplace, $\epsilon = 10$; Conjunto de dados: D1).

A título de ilustração, a Figura G.1 mostra a evolução do NICV (*fitness*) ao longo das gerações de uma execução do algoritmo BRKGAPD, utilizando o conjunto de dados D1, mecanismo de Laplace e $\epsilon = 10$. Cada ponto no eixo horizontal corresponde a uma geração do BRKGAPD e a uma etapa (*step*) do Q-Learning, em que uma ação² é executada. Cada ponto no eixo vertical mostra o melhor *fitness* da população correspondente à respectiva geração. Neste exemplo, foram executadas 167 gerações no total, das quais em 103 ocorreu uma transição, i.e., o Q-Learning efetivamente alterou o valor de um dos parâmetros do BRKGAPD. A sequência mais longa de gerações consecutivas com a mesma combinação de parâmetros foi de 13 gerações.

A Tabela G.1 apresenta, para os três algoritmos, as combinações de parâmetros com maior proporção de vitórias, para todos os conjuntos de dados e cenários considerados. Os resultados completos se encontram no Anexo G (Tabelas G.2 a G.4, p.199-201).

Tabela G.1: Combinações vencedoras - NICV

Metaheurística	Parâmetros	Total de vitórias	Total de <i>steps</i>	Proporção de vitórias
BRKGA	$p = 500, p_e = 0,3, \rho = 0,9$	2587	18940	0,1366
ILS	$\beta_{\min} = 0,1, \beta_{\max} = 0,15$	1743	9533	0,1828
VNS	$k_{\max} = 6, \beta_{\min} = 0,01$	2375	12901	0,1841

A fim de avaliar se o desempenho em termos da quantidade de melhores soluções (vitórias) apresentava diferença significativa entre as combinações de parâmetros utilizadas,

²Uma ação pode ou não corresponder a uma transição, i.e., a uma alteração na configuração de parâmetros.

alguns testes de hipótese foram realizados.

Primeiramente, o teste χ^2 de homogeneidade [5, p.78] foi aplicado para fazer uma avaliação preliminar global para as combinações de parâmetros adotada para cada algoritmo. As hipóteses foram formuladas considerando as m possíveis combinações de parâmetros e a proporção de vitórias correspondente, p_m .

H_0 : Todas as combinações apresentam a mesma proporção de vitórias.

$$p_1 = p_2 = \dots = p_m$$

H_1 : Pelo menos duas combinações apresentam proporções de vitórias diferentes.

$$\exists(i,j) \text{ tal que } p_i \neq p_j; \ i,j = 1, \dots, m$$

Considerando as soluções produzidas através dos algoritmos BRKGAPD, ILSPD e VNSPD, os p-valores foram, respectivamente, $3,37 \times 10^{-6}$, 0,2044 e $5,18 \times 10^{-32}$. A hipótese nula foi rejeitada para os algoritmos BRKGAPD e VNSPD, sugerindo que havia evidência de diferença significativa em ao menos um par de combinações. Já no algoritmo ILSPD, a hipótese nula não foi rejeitada, sugerindo que, para esse algoritmo, as diferenças entre as proporções de vitórias das combinações utilizadas não foram estatisticamente significativas.

Em seguida, visando avaliar se a combinação vencedora nos algoritmos BRKGAPD e VNSPD apresentou proporção de vitórias diferente das demais combinações, foi aplicado o teste χ^2 entre ela e todas as demais combinações. Nesse caso, as hipóteses foram:

H_0 : A proporção de vitórias da combinação vencedora (p_V) é igual à proporção de vitórias da combinação concorrente (p_C)

$$p_V = p_C$$

H_1 : A proporção de vitórias da combinação vencedora (p_V) é diferente da proporção de vitórias da combinação concorrente (p_C)

$$p_V \neq p_C$$

Adicionalmente, a fim de controlar o risco crescente de erro do tipo I em múltiplos

testes (*Familywise Error Rate*³ - FER), os p-valores foram ajustados através da correção de Holm [105, 7].

A hipótese nula foi rejeitada em 34% (12) e 70% (16) das combinações de parâmetros dos algoritmos BRKGAPD e VNSPD, respectivamente. Esse resultado sugere que a combinação vencedora superou algumas das outras combinações, mas não dominou de maneira sistemática. Os resultados completos se encontram no Anexo G (Tabelas G.5 e G.6, p.202-203).

G.2 Q-Learning - Análise da recompensa

Esta seção se destina a avaliar a performance do Q-Learning na configuração dos parâmetros ao longo das execuções realizadas. Como descrito na Seção 4.6, o Q-Learning é orientado pela recompensa (ou *reward*) e a análise foi conduzida com base nessa medida. Mais especificamente, os conjuntos de soluções foram separados por cenários e, dentro de cada cenário, foi calculada a recompensa média em cada ação. A Figura G.2 mostra a recompensa média acumulada ao longo de todas as ações na execução do algoritmo BRKGAPD, no cenário definido pelo mecanismo de Laplace, com $\epsilon = 10$.

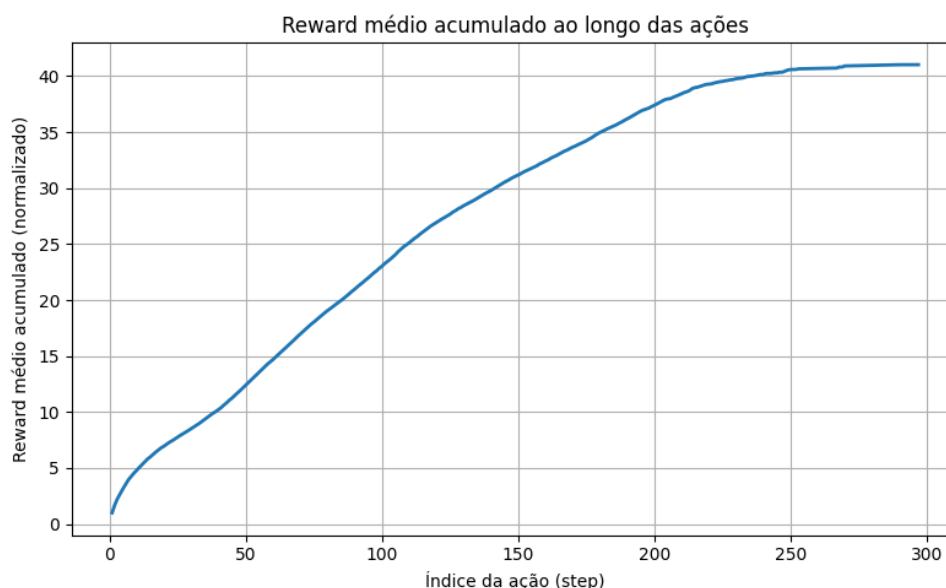


Figura G.2: Diagrama da recompensa média acumulada - BRKGAPD - Laplace, $\epsilon = 10$

É possível ver uma tendência predominantemente crescente ao longo de todo o processo, com a formação de um patamar ao final, sugerindo uma fase de convergência. A

³*Familywise Error Rates* (FER/FWER/FWE) é a probabilidade de cometer pelo menos um erro tipo I (falso positivo) dentro de um conjunto (família) de testes de hipótese realizados simultaneamente [102, p.3]. São também referenciadas na literatura como *Experimentwise Error Rates* (EER).

fim de tornar a análise mais objetiva e partindo da premissa de que quanto maior a recompensa acumulada, melhores foram as escolhas feitas pelo Q-Learning, a área sob a curva (do inglês, *Area Under the Curve* - AUC) foi adotada como medida de desempenho na configuração dos parâmetros dos 3 algoritmos ao longo dos 18 cenários considerados. As Figuras G.3, G.4 e G.5 apresentam os mapas de calor da AUC da recompensa média acumulada em cada cenário, para os algoritmos BRKGAPD, ILSPD e VNSPD, respectivamente.

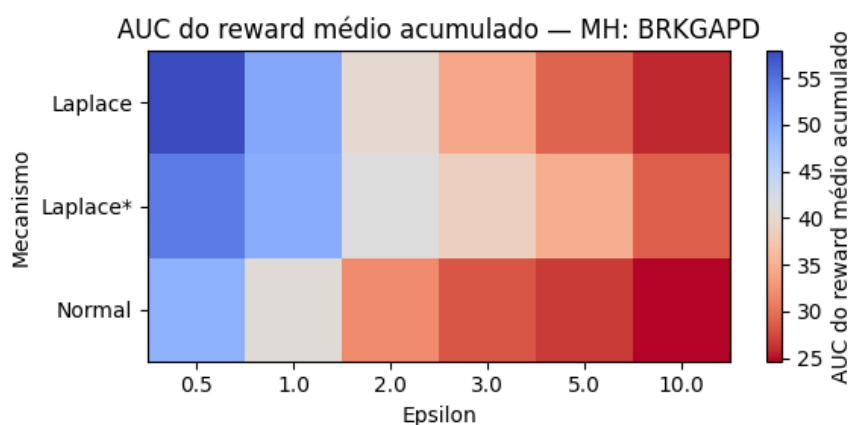


Figura G.3: AUC da recompensa média acumulada - BRKGAPD

Para os 3 algoritmos, o valor mínimo da AUC da recompensa média acumulada variou entre 20 e 60, indicando acúmulo relevante da recompensa em todos os casos. No que se refere aos valores de ϵ , há uma relação inversa bem definida com a AUC da recompensa média acumulada, indicando que, quanto menor a intensidade do ruído adicionado, menor é o acúmulo de recompensa média. Conclusão análoga pode ser feita em relação aos mecanismos, uma vez que o mecanismo Normal acumulou menos recompensa média do que os demais mecanismos em todos os casos. No entanto, em relação aos mecanismos de

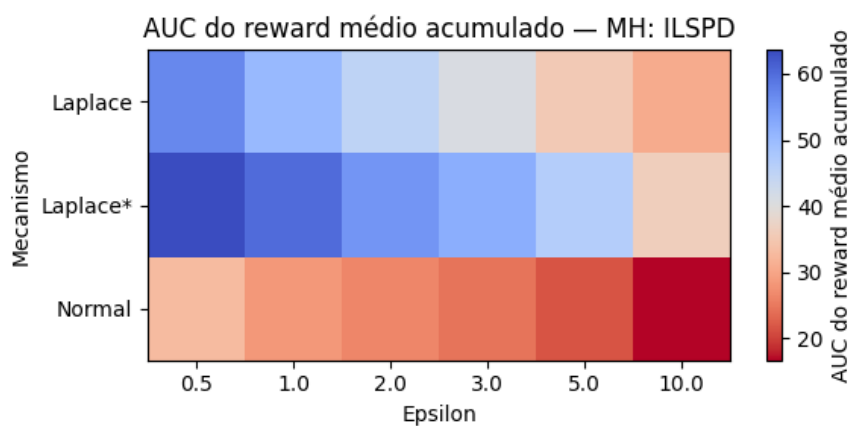


Figura G.4: AUC da recompensa média acumulada - ILSPD

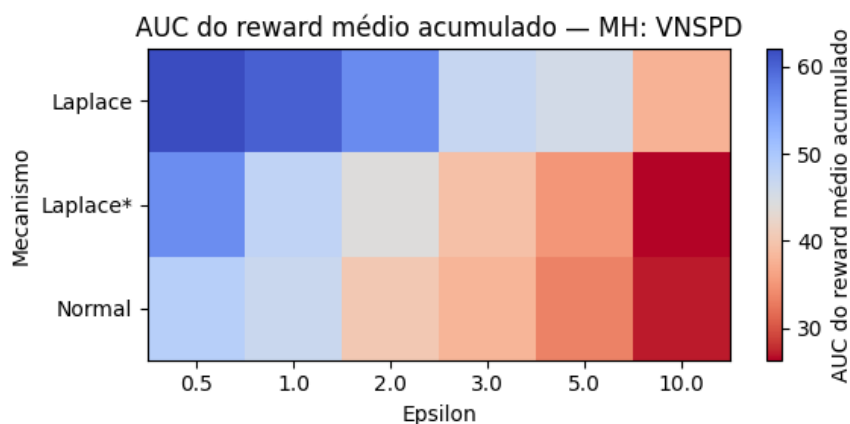


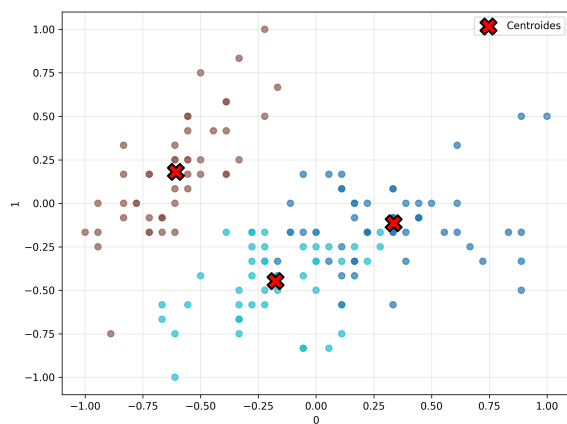
Figura G.5: AUC da recompensa média acumulada - VN SPD

Laplace e Laplace*, essa conclusão não foi consistente ao longo dos 3 algoritmos. Mais especificamente, o mecanismo de Laplace* acumulou menor recompensa nos algoritmos BRKGAPD e VN SPD, porém acumulou maior recompensa no algoritmo ILSPD.

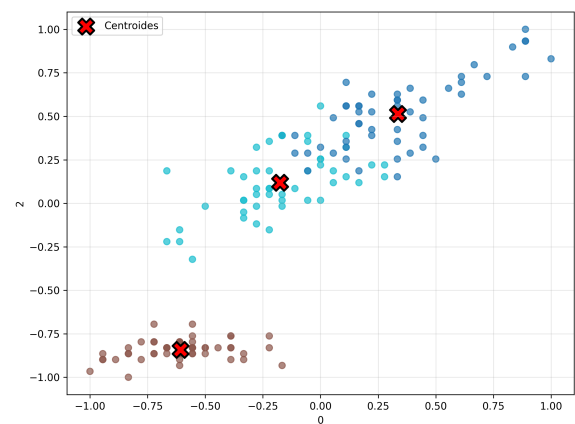
É importante destacar que a AUC da recompensa acumulada permite analisar a capacidade do Q-Learning de conduzir o processo de calibração em uma trajetória progressivamente favorável, mas não deve ser associada diretamente à qualidade final absoluta das soluções.

G.3 Visualização dos resultados do agrupamento

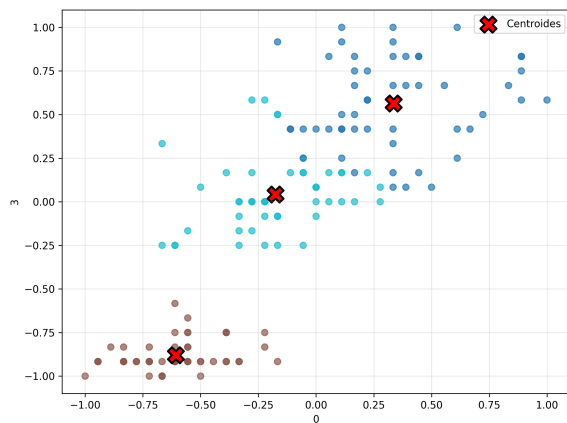
A Figura G.6 apresenta os diagramas de dispersão do resultado da aplicação do k-Means ao conjunto de dados `iris`, com $k = 3$ e sem o uso PD. A implementação utilizada adotou algoritmo de Lloyd e inicialização k-Means++. Uma vez que o conjunto `iris` conta com 4 atributos, esses foram combinados 2 a 2 a fim de permitir a visualização bidimensional.



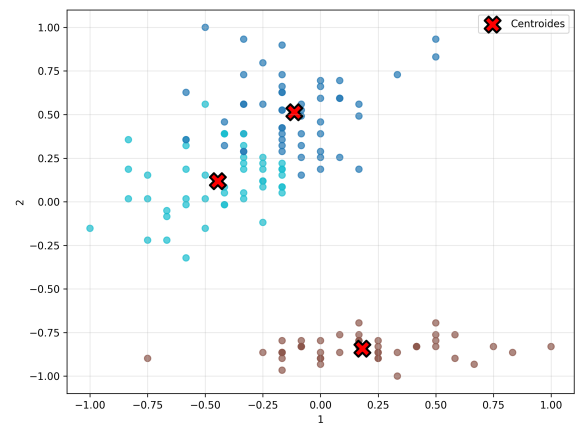
(a) atributos 0 e 1



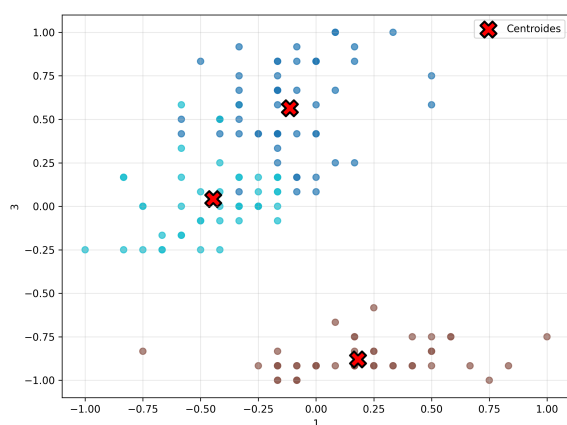
(b) atributos 0 e 2



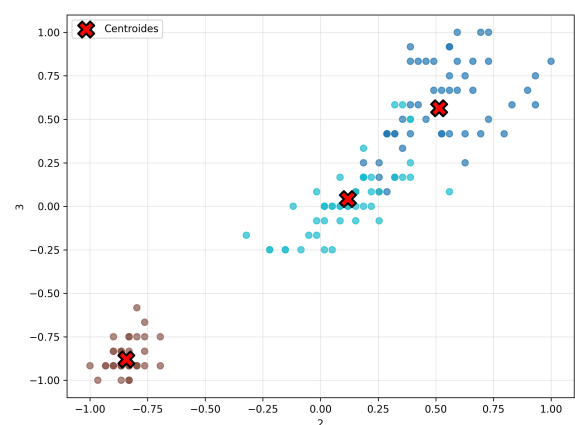
(c) atributos 0 e 3



(d) atributos 1 e 2



(e) atributos 1 e 3



(f) atributos 2 e 3

Figura G.6: Diagramas de dispersão do resultado da aplicação do k-Means ao conjunto de dados iris.

G.4 Tabelas completas de resultados

Tabela G.2: Parâmetros BRKGA

população	p_e	ρ	vitórias	execuções	proporção de vitórias
500	0,3	0,9	2587	18940	0,1366
500	0,2	0,8	2580	19487	0,1324
500	0,1	0,7	2501	18893	0,1324
100	0,2	0,7	2620	19851	0,1320
500	0,2	0,7	2520	19194	0,1313
500	0,1	0,9	2534	19305	0,1313
100	0,2	0,9	2625	20046	0,1309
300	0,2	0,8	2587	19791	0,1307
70	0,1	0,8	2683	20527	0,1307
500	0,2	0,9	2476	18996	0,1303
500	0,3	0,8	2518	19382	0,1299
70	0,3	0,8	2621	20248	0,1294
500	0,3	0,7	2457	18987	0,1294
100	0,2	0,8	2592	20058	0,1292
100	0,3	0,9	2622	20381	0,1286
300	0,2	0,9	2517	19611	0,1283
500	0,1	0,8	2443	19101	0,1279
300	0,3	0,7	2503	19609	0,1276
70	0,2	0,8	2607	20435	0,1276
70	0,3	0,7	2494	19602	0,1272
70	0,2	0,9	2581	20309	0,1271
300	0,1	0,8	2519	19839	0,1270
100	0,1	0,9	2575	20322	0,1267
300	0,1	0,7	2473	19550	0,1265
300	0,3	0,8	2503	19900	0,1258
300	0,1	0,9	2506	19935	0,1257
100	0,3	0,7	2539	20201	0,1257
300	0,3	0,9	2496	19909	0,1254
100	0,3	0,8	2522	20313	0,1242
70	0,1	0,9	2522	20403	0,1236
100	0,1	0,8	2504	20273	0,1235
70	0,2	0,7	2469	19996	0,1235
100	0,1	0,7	2453	19910	0,1232
70	0,3	0,9	2490	20369	0,1222
300	0,2	0,7	2385	19836	0,1202
70	0,1	0,7	2395	20150	0,1189

Tabela G.3: Parâmetros ILS

$\beta_{\min}$	$\beta_{\max}$	vitórias	execuções	proporção de vitórias
0,100	0,150	1743	9533	0,1828
0,050	0,200	1782	9749	0,1828
0,010	0,250	1704	9398	0,1813
0,025	0,150	1747	9722	0,1797
0,025	0,200	1735	9743	0,1781
0,075	0,150	1736	9768	0,1777
0,050	0,300	1727	9747	0,1772
0,100	0,200	1704	9650	0,1766
0,025	0,300	1692	9626	0,1758
0,075	0,200	1723	9804	0,1757
0,050	0,150	1710	9758	0,1752
0,100	0,300	1691	9650	0,1752
0,075	0,300	1734	9949	0,1743
0,025	0,250	1651	9481	0,1741
0,075	0,250	1705	9827	0,1735
0,010	0,200	1651	9516	0,1735
0,010	0,150	1636	9437	0,1734
0,100	0,250	1635	9528	0,1716
0,050	0,250	1629	9520	0,1711
0,010	0,300	1550	9463	0,1638

Tabela G.4: Parâmetros VNS

$k_{\max}$	$\beta_{\min}$	vitórias	execuções	proporção de vitórias
6	0,010	2375	12901	0,1841
6	0,050	2388	13022	0,1834
6	0,025	2288	12913	0,1772
4	0,025	1856	10475	0,1772
6	0,200	2214	12546	0,1765
5	0,010	2179	12434	0,1752
4	0,075	1772	10122	0,1751
5	0,050	2132	12303	0,1733
5	0,075	2009	11773	0,1706
6	0,075	2146	12696	0,1690
5	0,100	1966	11717	0,1678
4	0,010	1732	10324	0,1678
4	0,050	1738	10558	0,1646
5	0,200	1905	11611	0,1641
5	0,025	2005	12300	0,1630
6	0,100	2066	12679	0,1629
3	0,075	1409	8859	0,1590
3	0,100	1446	9147	0,1581
3	0,010	1419	9089	0,1561
4	0,100	1705	10928	0,1560
3	0,025	1376	8989	0,1531
3	0,050	1382	9043	0,1528
4	0,200	1508	10149	0,1486
3	0,200	1267	8824	0,1436

Tabela G.5: Teste de duas proporções - BRKGA

população	p_e	ρ	p-valor
70	0,1	0,7	0,00
300	0,2	0,7	0,00
70	0,3	0,9	0,00
100	0,1	0,7	0,00
70	0,2	0,7	0,00
100	0,1	0,8	0,00
70	0,1	0,9	0,00
100	0,3	0,8	0,00
300	0,3	0,9	0,03
100	0,3	0,7	0,04
300	0,1	0,9	0,04
300	0,3	0,8	0,04
300	0,1	0,7	0,08
100	0,1	0,9	0,08
300	0,1	0,8	0,11
70	0,2	0,9	0,11
70	0,3	0,7	0,13
70	0,2	0,8	0,15
300	0,3	0,7	0,16
500	0,1	0,8	0,20
300	0,2	0,9	0,25
100	0,3	0,9	0,28
100	0,2	0,8	0,42
70	0,3	0,8	0,45
500	0,3	0,7	0,45
500	0,3	0,8	0,55
500	0,2	0,9	0,66
500	0,2	0,8	0,69
500	0,1	0,9	0,69
500	0,1	0,7	0,69
100	0,2	0,7	0,69
500	0,2	0,7	0,69
70	0,1	0,8	0,69
300	0,2	0,8	0,69
100	0,2	0,9	0,69

Tabela G.6: Teste de duas proporções - VNS

$k_{\max}$	$\beta_{\min}$	p-valor
3	0,200	0,00
4	0,200	0,00
3	0,050	0,00
3	0,025	0,00
4	0,100	0,00
3	0,010	0,00
3	0,100	0,00
3	0,075	0,00
6	0,100	0,00
5	0,025	0,00
5	0,200	0,00
4	0,050	0,00
5	0,100	0,01
4	0,010	0,01
6	0,075	0,01
5	0,075	0,05
5	0,050	0,18
4	0,075	0,40
5	0,010	0,40
6	0,200	0,46
6	0,025	0,46
4	0,025	0,46
6	0,050	0,88

Tabela G.7: Frequência de conjuntos de dados com 10 repetições concluídas em cada cenário.

Cenário mecanismo	ϵ	BRKGAPD	ILSPD	VNSPD	APAS	DAS	Escalonada	SSAS	Uniforme
Laplace	0,5	30	30	30	2	30	30	30	30
Laplace	1	30	30	30	4	30	30	30	30
Laplace	2	30	30	30	8	30	30	30	30
Laplace	3	30	30	30	14	30	30	30	30
Laplace	5	30	30	30	23	30	30	30	30
Laplace	10	30	30	30	24	30	30	30	30
Laplace*	0,5	30	30	30	2	30	30	30	30
Laplace*	1	30	30	30	4	30	30	30	30
Laplace*	2	30	30	30	8	30	30	30	30
Laplace*	3	30	30	30	14	30	30	30	30
Laplace*	5	30	30	30	23	30	30	30	30
Laplace*	10	30	30	30	24	30	30	30	30
Normal	0,5	30	30	30	2	30	30	30	30
Normal	1	30	30	30	4	30	30	30	30
Normal	2	30	30	30	8	30	30	30	30
Normal	3	30	30	30	14	30	30	30	30
Normal	5	30	30	30	23	30	30	30	30
Normal	10	30	30	30	24	30	30	30	30

Tabela G.8: Conjuntos de dados em que a estratégia APAS foi executada com sucesso.

	0,5	1	2	3	5	10
banana	sim	sim	sim	sim	sim	sim
iris	-	-	-	-	-	-
n6000k6	-	sim	sim	sim	sim	sim
transfusion	-	-	-	sim	sim	sim
wdbc	-	-	-	-	-	-
D1	-	-	-	-	sim	sim
D2	-	-	-	-	sim	sim
D3	-	-	sim	sim	sim	sim
D4	-	-	sim	sim	sim	sim
D5	-	-	-	-	-	-
D6	-	-	-	-	sim	sim
D7	-	-	-	-	sim	sim
D8	-	-	-	sim	sim	sim
D9	-	-	sim	sim	sim	sim
D10	-	-	-	-	sim	sim
D11	-	-	-	-	-	sim
D12	-	-	-	sim	sim	sim
D13	-	sim	sim	sim	sim	sim
D14	-	-	-	sim	sim	sim
D15	-	-	-	sim	sim	sim
D16	-	-	-	-	-	-
D17	-	-	-	sim	sim	sim
D18	sim	sim	sim	sim	sim	sim
R1	-	-	-	-	-	-
R2	-	-	-	-	sim	sim
R3	-	-	-	-	-	-
R4	-	-	-	-	sim	sim
R5	-	-	-	-	sim	sim
R6	-	-	-	-	sim	sim
R7	-	-	sim	sim	sim	sim
Total	2	4	8	14	23	24

O hífen sinaliza os casos nos quais a estratégia APAS não foi executada.

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário.

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
banana	0,5	Laplace	0,5691	0,5692	0,5712	0,5708
banana	0,5	Laplace*	0,5691	0,5692	0,5699	0,5704
banana	0,5	Normal	0,5691	0,5692	0,5717	0,5702
banana	1	Laplace	0,5691	0,5691	0,5695	0,5694
banana	1	Laplace*	0,5691	0,5691	0,5692	0,5693
banana	1	Normal	0,5691	0,5691	0,5694	0,5695
banana	2	Laplace	0,5691	0,5691	0,5692	0,5693
banana	2	Laplace*	0,5691	0,5691	0,5691	0,5692
banana	2	Normal	0,5691	0,5691	0,5692	0,5692
banana	3	Laplace	0,5691	0,5691	0,5691	0,5691
banana	3	Laplace*	0,5691	0,5691	0,5691	0,5691
banana	3	Normal	0,5692	0,5691	0,5692	0,5691
banana	5	Laplace	0,5691	0,5691	0,5691	0,5691
banana	5	Laplace*	0,5691	0,5691	0,5691	0,5691
banana	5	Normal	0,5691	0,5691	0,5691	0,5691
banana	10	Laplace	0,5691	0,5691	0,5691	0,5691
banana	10	Laplace*	0,5691	0,5691	0,5691	0,5691
banana	10	Normal	0,5691	0,5691	0,5691	0,5691
iris	0,5	Laplace	0,1701	0,4441	1,0615	1,1013
iris	0,5	Laplace*	0,1701	0,4385	1,0834	1,0614
iris	0,5	Normal	0,1701	0,4689	1,1765	1,1708
iris	1	Laplace	0,1701	0,3373	0,9072	0,9699
iris	1	Laplace*	0,1701	0,3281	1,0090	1,0110
iris	1	Normal	0,1701	0,3498	0,9801	0,9021
iris	2	Laplace	0,1735	0,2785	0,5229	0,5564
iris	2	Laplace*	0,1701	0,3071	0,6264	0,5875
iris	2	Normal	0,1701	0,3093	0,5398	0,6009
iris	3	Laplace	0,1701	0,2384	0,4221	0,4334
iris	3	Laplace*	0,1701	0,2862	0,4314	0,4188
iris	3	Normal	0,1701	0,2290	0,4088	0,4568
iris	5	Laplace	0,1701	0,1914	0,3338	0,3314
iris	5	Laplace*	0,1701	0,2386	0,3391	0,3417
iris	5	Normal	0,1701	0,1977	0,3133	0,3289
iris	10	Laplace	0,1701	0,1771	0,2396	0,2319
iris	10	Laplace*	0,1701	0,1802	0,2395	0,2971
iris	10	Normal	0,1701	0,1780	0,2188	0,2313
n6000k6	0,5	Laplace	0,0220	0,0260	0,0434	0,0672
n6000k6	0,5	Laplace*	0,0220	0,0236	0,0315	0,0341

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
n6000k6	0,5	Normal	0,0220	0,0259	0,0439	0,0487
n6000k6	1	Laplace	0,0220	0,0227	0,0258	0,0257
n6000k6	1	Laplace*	0,0220	0,0221	0,0227	0,0228
n6000k6	1	Normal	0,0220	0,0223	0,0232	0,0241
n6000k6	2	Laplace	0,0220	0,0221	0,0224	0,0225
n6000k6	2	Laplace*	0,0220	0,0220	0,0221	0,0222
n6000k6	2	Normal	0,0220	0,0220	0,0223	0,0223
n6000k6	3	Laplace	0,0220	0,0220	0,0221	0,0222
n6000k6	3	Laplace*	0,0220	0,0220	0,0221	0,0221
n6000k6	3	Normal	0,0220	0,0220	0,0221	0,0221
n6000k6	5	Laplace	0,0220	0,0220	0,0220	0,0220
n6000k6	5	Laplace*	0,0220	0,0220	0,0220	0,0220
n6000k6	5	Normal	0,0220	0,0220	0,0220	0,0220
n6000k6	10	Laplace	0,0220	0,0220	0,0220	0,0220
n6000k6	10	Laplace*	0,0220	0,0220	0,0220	0,0220
n6000k6	10	Normal	0,0220	0,0220	0,0220	0,0220
transfusion	0,5	Laplace	0,4847	0,5244	0,9534	0,9426
transfusion	0,5	Laplace*	0,4847	0,5162	0,9554	1,0057
transfusion	0,5	Normal	0,4847	0,5216	0,9177	0,9626
transfusion	1	Laplace	0,4847	0,5027	0,7465	0,7269
transfusion	1	Laplace*	0,4847	0,4969	0,7699	0,7412
transfusion	1	Normal	0,4847	0,5020	0,6692	0,6131
transfusion	2	Laplace	0,4847	0,4922	0,5679	0,5647
transfusion	2	Laplace*	0,4847	0,4886	0,5935	0,5712
transfusion	2	Normal	0,4847	0,4910	0,5510	0,5447
transfusion	3	Laplace	0,4847	0,4886	0,5425	0,5198
transfusion	3	Laplace*	0,4847	0,4863	0,5384	0,5584
transfusion	3	Normal	0,4847	0,4881	0,5065	0,5083
transfusion	5	Laplace	0,4847	0,4861	0,4966	0,4940
transfusion	5	Laplace*	0,4847	0,4855	0,4964	0,4962
transfusion	5	Normal	0,4847	0,4858	0,4922	0,4950
transfusion	10	Laplace	0,4847	0,4850	0,4865	0,4869
transfusion	10	Laplace*	0,4847	0,4849	0,4859	0,4857
transfusion	10	Normal	0,4847	0,4849	0,4864	0,4862
wdbc	0,5	Laplace	0,6090	1,8777	4,5514	4,5314
wdbc	0,5	Laplace*	0,6090	1,3638	2,1027	3,0857
wdbc	0,5	Normal	0,6090	1,6671	3,8515	4,4078

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
wdbc	1	Laplace	0,6090	1,1231	2,4066	2,5118
wdbc	1	Laplace*	0,6090	1,0963	1,4744	1,5245
wdbc	1	Normal	0,6090	1,1433	1,9553	1,5452
wdbc	2	Laplace	0,6090	0,9622	1,2923	1,2545
wdbc	2	Laplace*	0,6090	1,0214	1,1240	1,1793
wdbc	2	Normal	0,6090	0,8928	1,1801	1,1550
wdbc	3	Laplace	0,6090	0,8269	1,1481	1,1294
wdbc	3	Laplace*	0,6090	1,0081	1,0720	1,0907
wdbc	3	Normal	0,6090	0,7714	1,0849	1,0566
wdbc	5	Laplace	0,6090	0,7056	1,0343	1,0385
wdbc	5	Laplace*	0,6090	0,9983	1,0130	1,0235
wdbc	5	Normal	0,6090	0,6999	1,0314	1,0223
wdbc	10	Laplace	0,6090	0,6465	1,0104	1,0146
wdbc	10	Laplace*	0,6090	1,0009	1,0038	1,0059
wdbc	10	Normal	0,6090	0,6461	0,9926	0,9343
D1	0,5	Laplace	0,0650	0,1150	0,2916	0,2915
D1	0,5	Laplace*	0,0650	0,0892	0,2015	0,2076
D1	0,5	Normal	0,0650	0,1043	0,2336	0,2426
D1	1	Laplace	0,0650	0,0740	0,1544	0,1686
D1	1	Laplace*	0,0650	0,0699	0,1049	0,1027
D1	1	Normal	0,0650	0,0738	0,1362	0,0983
D1	2	Laplace	0,0650	0,0670	0,0791	0,0815
D1	2	Laplace*	0,0650	0,0657	0,0696	0,0678
D1	2	Normal	0,0650	0,0668	0,0729	0,0732
D1	3	Laplace	0,0650	0,0658	0,0688	0,0716
D1	3	Laplace*	0,0650	0,0652	0,0663	0,0659
D1	3	Normal	0,0650	0,0655	0,0670	0,0676
D1	5	Laplace	0,0650	0,0652	0,0665	0,0663
D1	5	Laplace*	0,0650	0,0651	0,0653	0,0652
D1	5	Normal	0,0650	0,0651	0,0656	0,0656
D1	10	Laplace	0,0650	0,0650	0,0652	0,0652
D1	10	Laplace*	0,0650	0,0650	0,0651	0,0651
D1	10	Normal	0,0650	0,0650	0,0651	0,0651
D2	0,5	Laplace	0,0724	0,1791	0,3506	0,3764
D2	0,5	Laplace*	0,0724	0,1459	0,3045	0,2985
D2	0,5	Normal	0,0724	0,1701	0,3472	0,3389
D2	1	Laplace	0,0724	0,1042	0,2389	0,2361

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D2	1	Laplace*	0,0724	0,0928	0,1982	0,1898
D2	1	Normal	0,0724	0,0943	0,1906	0,1969
D2	2	Laplace	0,0724	0,0790	0,1232	0,1307
D2	2	Laplace*	0,0724	0,0762	0,0967	0,0965
D2	2	Normal	0,0724	0,0793	0,1027	0,1073
D2	3	Laplace	0,0724	0,0748	0,0891	0,0918
D2	3	Laplace*	0,0724	0,0728	0,0770	0,0759
D2	3	Normal	0,0724	0,0747	0,0825	0,0826
D2	5	Laplace	0,0724	0,0730	0,0756	0,0751
D2	5	Laplace*	0,0724	0,0725	0,0731	0,0729
D2	5	Normal	0,0724	0,0727	0,0739	0,0740
D2	10	Laplace	0,0724	0,0724	0,0729	0,0728
D2	10	Laplace*	0,0724	0,0724	0,0725	0,0725
D2	10	Normal	0,0724	0,0724	0,0726	0,0726
D3	0,5	Laplace	0,1223	0,1398	0,2938	0,2798
D3	0,5	Laplace*	0,1223	0,1402	0,2847	0,2653
D3	0,5	Normal	0,1223	0,1397	0,2746	0,2538
D3	1	Laplace	0,1223	0,1272	0,1724	0,2000
D3	1	Laplace*	0,1223	0,1259	0,1526	0,1524
D3	1	Normal	0,1223	0,1288	0,1480	0,1715
D3	2	Laplace	0,1223	0,1237	0,1309	0,1295
D3	2	Laplace*	0,1223	0,1228	0,1263	0,1259
D3	2	Normal	0,1223	0,1234	0,1269	0,1287
D3	3	Laplace	0,1223	0,1228	0,1254	0,1257
D3	3	Laplace*	0,1223	0,1224	0,1237	0,1232
D3	3	Normal	0,1223	0,1226	0,1240	0,1245
D3	5	Laplace	0,1223	0,1224	0,1231	0,1234
D3	5	Laplace*	0,1223	0,1223	0,1227	0,1226
D3	5	Normal	0,1223	0,1223	0,1228	0,1228
D3	10	Laplace	0,1223	0,1223	0,1224	0,1225
D3	10	Laplace*	0,1223	0,1223	0,1223	0,1223
D3	10	Normal	0,1223	0,1223	0,1224	0,1224
D4	0,5	Laplace	0,0441	0,0513	0,0866	0,0875
D4	0,5	Laplace*	0,0441	0,0496	0,0788	0,0782
D4	0,5	Normal	0,0441	0,0526	0,0841	0,0859
D4	1	Laplace	0,0441	0,0467	0,0743	0,0770
D4	1	Laplace*	0,0441	0,0461	0,0689	0,0700

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D4	1	Normal	0,0441	0,0468	0,0739	0,0722
D4	2	Laplace	0,0441	0,0451	0,0540	0,0542
D4	2	Laplace*	0,0441	0,0443	0,0453	0,0455
D4	2	Normal	0,0441	0,0449	0,0485	0,0488
D4	3	Laplace	0,0441	0,0445	0,0460	0,0469
D4	3	Laplace*	0,0441	0,0442	0,0447	0,0447
D4	3	Normal	0,0441	0,0443	0,0451	0,0451
D4	5	Laplace	0,0441	0,0442	0,0445	0,0445
D4	5	Laplace*	0,0441	0,0441	0,0442	0,0442
D4	5	Normal	0,0441	0,0442	0,0444	0,0444
D4	10	Laplace	0,0441	0,0442	0,0442	0,0442
D4	10	Laplace*	0,0441	0,0441	0,0441	0,0441
D4	10	Normal	0,0441	0,0441	0,0442	0,0442
D5	0,5	Laplace	0,0961	0,4474	0,8293	0,8118
D5	0,5	Laplace*	0,0961	0,4081	0,7560	0,7916
D5	0,5	Normal	0,0961	0,4388	0,9189	0,7804
D5	1	Laplace	0,0961	0,2926	0,7383	0,6595
D5	1	Laplace*	0,0961	0,2129	0,5711	0,6092
D5	1	Normal	0,0961	0,2592	0,7187	0,6271
D5	2	Laplace	0,0961	0,1547	0,4379	0,4989
D5	2	Laplace*	0,0961	0,1385	0,4184	0,3928
D5	2	Normal	0,0961	0,1474	0,4570	0,4403
D5	3	Laplace	0,0961	0,1308	0,3785	0,3823
D5	3	Laplace*	0,0961	0,1105	0,2821	0,3007
D5	3	Normal	0,0961	0,1212	0,3249	0,3461
D5	5	Laplace	0,0961	0,1086	0,2307	0,2411
D5	5	Laplace*	0,0961	0,1008	0,1727	0,1513
D5	5	Normal	0,0961	0,1045	0,1985	0,1953
D5	10	Laplace	0,0961	0,0985	0,1220	0,1198
D5	10	Laplace*	0,0961	0,0967	0,1025	0,1017
D5	10	Normal	0,0961	0,0980	0,1144	0,1153
D6	0,5	Laplace	0,1165	0,1767	0,4174	0,4226
D6	0,5	Laplace*	0,1165	0,1597	0,2655	0,3489
D6	0,5	Normal	0,1165	0,1718	0,3817	0,3346
D6	1	Laplace	0,1165	0,1325	0,2155	0,2326
D6	1	Laplace*	0,1165	0,1300	0,1878	0,1794
D6	1	Normal	0,1165	0,1325	0,2093	0,1924

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D6	2	Laplace	0,1165	0,1209	0,1358	0,1462
D6	2	Laplace*	0,1165	0,1190	0,1291	0,1258
D6	2	Normal	0,1165	0,1204	0,1352	0,1304
D6	3	Laplace	0,1165	0,1184	0,1257	0,1269
D6	3	Laplace*	0,1165	0,1173	0,1203	0,1221
D6	3	Normal	0,1165	0,1182	0,1243	0,1246
D6	5	Laplace	0,1165	0,1171	0,1203	0,1204
D6	5	Laplace*	0,1165	0,1168	0,1175	0,1179
D6	5	Normal	0,1165	0,1170	0,1182	0,1183
D6	10	Laplace	0,1165	0,1166	0,1170	0,1172
D6	10	Laplace*	0,1165	0,1166	0,1167	0,1169
D6	10	Normal	0,1165	0,1166	0,1169	0,1169
D7	0,5	Laplace	0,1110	0,1520	0,3423	0,3586
D7	0,5	Laplace*	0,1021	0,1413	0,2619	0,2699
D7	0,5	Normal	0,1054	0,1556	0,2615	0,2823
D7	1	Laplace	0,1067	0,1192	0,1828	0,1883
D7	1	Laplace*	0,1173	0,1148	0,1528	0,1566
D7	1	Normal	0,1021	0,1170	0,1575	0,1778
D7	2	Laplace	0,1021	0,1067	0,1254	0,1294
D7	2	Laplace*	0,1021	0,1043	0,1165	0,1163
D7	2	Normal	0,1021	0,1057	0,1215	0,1242
D7	3	Laplace	0,1021	0,1042	0,1133	0,1129
D7	3	Laplace*	0,1021	0,1030	0,1058	0,1066
D7	3	Normal	0,1021	0,1038	0,1076	0,1072
D7	5	Laplace	0,1021	0,1028	0,1049	0,1049
D7	5	Laplace*	0,1021	0,1023	0,1036	0,1032
D7	5	Normal	0,1021	0,1026	0,1044	0,1044
D7	10	Laplace	0,1027	0,1022	0,1027	0,1028
D7	10	Laplace*	0,1077	0,1021	0,1025	0,1025
D7	10	Normal	0,1021	0,1021	0,1025	0,1028
D8	0,5	Laplace	0,0948	0,1534	0,3081	0,3129
D8	0,5	Laplace*	0,0948	0,1294	0,2498	0,2455
D8	0,5	Normal	0,0948	0,1382	0,2418	0,2837
D8	1	Laplace	0,0948	0,1073	0,1789	0,1813
D8	1	Laplace*	0,0948	0,1033	0,1382	0,1437
D8	1	Normal	0,0948	0,1058	0,1607	0,1428
D8	2	Laplace	0,0948	0,0984	0,1088	0,1108

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D8	2	Laplace*	0,0948	0,0960	0,1031	0,1045
D8	2	Normal	0,0948	0,0976	0,1069	0,1061
D8	3	Laplace	0,0948	0,0964	0,1014	0,1008
D8	3	Laplace*	0,0948	0,0951	0,0964	0,0964
D8	3	Normal	0,0948	0,0957	0,0976	0,0988
D8	5	Laplace	0,0948	0,0950	0,0964	0,0960
D8	5	Laplace*	0,0948	0,0948	0,0952	0,0953
D8	5	Normal	0,0948	0,0950	0,0956	0,0958
D8	10	Laplace	0,0948	0,0948	0,0950	0,0950
D8	10	Laplace*	0,0948	0,0948	0,0949	0,0949
D8	10	Normal	0,0948	0,0948	0,0950	0,0949
D9	0,5	Laplace	0,1636	0,1678	0,2411	0,2386
D9	0,5	Laplace*	0,1636	0,1668	0,1885	0,1955
D9	0,5	Normal	0,1636	0,1691	0,2114	0,2172
D9	1	Laplace	0,1636	0,1648	0,1818	0,1840
D9	1	Laplace*	0,1636	0,1644	0,1697	0,1686
D9	1	Normal	0,1636	0,1645	0,1806	0,1759
D9	2	Laplace	0,1636	0,1638	0,1669	0,1666
D9	2	Laplace*	0,1636	0,1637	0,1647	0,1652
D9	2	Normal	0,1636	0,1638	0,1665	0,1657
D9	3	Laplace	0,1636	0,1636	0,1648	0,1647
D9	3	Laplace*	0,1636	0,1636	0,1640	0,1641
D9	3	Normal	0,1636	0,1637	0,1645	0,1644
D9	5	Laplace	0,1636	0,1636	0,1639	0,1640
D9	5	Laplace*	0,1636	0,1636	0,1637	0,1637
D9	5	Normal	0,1636	0,1636	0,1640	0,1639
D9	10	Laplace	0,1636	0,1636	0,1637	0,1637
D9	10	Laplace*	0,1636	0,1636	0,1636	0,1636
D9	10	Normal	0,1636	0,1636	0,1637	0,1637
D10	0,5	Laplace	0,0911	0,1348	0,3504	0,3456
D10	0,5	Laplace*	0,0911	0,1317	0,2506	0,2811
D10	0,5	Normal	0,0911	0,1262	0,3080	0,2932
D10	1	Laplace	0,0911	0,1029	0,2104	0,1934
D10	1	Laplace*	0,0911	0,0976	0,1434	0,1380
D10	1	Normal	0,0911	0,0982	0,1829	0,1621
D10	2	Laplace	0,0911	0,0936	0,1113	0,1041
D10	2	Laplace*	0,0911	0,0921	0,0948	0,0964

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D10	2	Normal	0,0911	0,0926	0,1035	0,0995
D10	3	Laplace	0,0911	0,0919	0,0973	0,0977
D10	3	Laplace*	0,0911	0,0914	0,0925	0,0925
D10	3	Normal	0,0911	0,0916	0,0947	0,0935
D10	5	Laplace	0,0911	0,0913	0,0922	0,0925
D10	5	Laplace*	0,0911	0,0912	0,0915	0,0915
D10	5	Normal	0,0911	0,0913	0,0919	0,0918
D10	10	Laplace	0,0911	0,0912	0,0914	0,0914
D10	10	Laplace*	0,0911	0,0911	0,0912	0,0912
D10	10	Normal	0,0911	0,0912	0,0913	0,0912
D11	0,5	Laplace	0,0544	0,1817	0,4379	0,4737
D11	0,5	Laplace*	0,0544	0,1604	0,3182	0,3530
D11	0,5	Normal	0,0544	0,1609	0,4151	0,3942
D11	1	Laplace	0,0544	0,1183	0,2737	0,2673
D11	1	Laplace*	0,0544	0,1001	0,2021	0,2499
D11	1	Normal	0,0544	0,1044	0,2678	0,2664
D11	2	Laplace	0,0544	0,0713	0,1568	0,1652
D11	2	Laplace*	0,0544	0,0691	0,1287	0,1428
D11	2	Normal	0,0544	0,0663	0,1428	0,1369
D11	3	Laplace	0,0544	0,0619	0,1066	0,1341
D11	3	Laplace*	0,0544	0,0584	0,0878	0,0855
D11	3	Normal	0,0544	0,0597	0,0996	0,1078
D11	5	Laplace	0,0544	0,0572	0,0793	0,0780
D11	5	Laplace*	0,0544	0,0554	0,0586	0,0605
D11	5	Normal	0,0544	0,0564	0,0670	0,0677
D11	10	Laplace	0,0544	0,0549	0,0579	0,0567
D11	10	Laplace*	0,0544	0,0545	0,0549	0,0552
D11	10	Normal	0,0544	0,0547	0,0555	0,0554
D12	0,5	Laplace	0,0383	0,0480	0,2031	0,1682
D12	0,5	Laplace*	0,0383	0,0423	0,0903	0,1131
D12	0,5	Normal	0,0383	0,0456	0,1154	0,1021
D12	1	Laplace	0,0383	0,0405	0,0738	0,0618
D12	1	Laplace*	0,0383	0,0390	0,0455	0,0471
D12	1	Normal	0,0383	0,0396	0,0503	0,0554
D12	2	Laplace	0,0383	0,0386	0,0429	0,0427
D12	2	Laplace*	0,0383	0,0384	0,0389	0,0390
D12	2	Normal	0,0383	0,0385	0,0402	0,0410

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D12	3	Laplace	0,0383	0,0384	0,0397	0,0396
D12	3	Laplace*	0,0383	0,0383	0,0386	0,0388
D12	3	Normal	0,0383	0,0384	0,0394	0,0390
D12	5	Laplace	0,0383	0,0383	0,0387	0,0386
D12	5	Laplace*	0,0383	0,0383	0,0384	0,0384
D12	5	Normal	0,0383	0,0383	0,0385	0,0385
D12	10	Laplace	0,0383	0,0383	0,0384	0,0384
D12	10	Laplace*	0,0383	0,0383	0,0383	0,0383
D12	10	Normal	0,0383	0,0383	0,0384	0,0383
D13	0,5	Laplace	0,4238	0,4246	0,4442	0,4438
D13	0,5	Laplace*	0,4238	0,4241	0,4314	0,4308
D13	0,5	Normal	0,4238	0,4243	0,4361	0,4368
D13	1	Laplace	0,4238	0,4239	0,4277	0,4267
D13	1	Laplace*	0,4238	0,4238	0,4247	0,4247
D13	1	Normal	0,4238	0,4239	0,4272	0,4262
D13	2	Laplace	0,4238	0,4238	0,4241	0,4243
D13	2	Laplace*	0,4238	0,4238	0,4240	0,4239
D13	2	Normal	0,4238	0,4238	0,4242	0,4241
D13	3	Laplace	0,4238	0,4238	0,4240	0,4240
D13	3	Laplace*	0,4238	0,4238	0,4239	0,4239
D13	3	Normal	0,4238	0,4238	0,4240	0,4239
D13	5	Laplace	0,4238	0,4238	0,4239	0,4239
D13	5	Laplace*	0,4238	0,4238	0,4238	0,4238
D13	5	Normal	0,4238	0,4238	0,4239	0,4239
D13	10	Laplace	0,4238	0,4238	0,4238	0,4238
D13	10	Laplace*	0,4238	0,4238	0,4238	0,4238
D13	10	Normal	0,4238	0,4238	0,4238	0,4238
D14	0,5	Laplace	0,1363	0,1630	0,3909	0,3831
D14	0,5	Laplace*	0,1363	0,1540	0,3018	0,2935
D14	0,5	Normal	0,1363	0,1572	0,4052	0,3209
D14	1	Laplace	0,1363	0,1427	0,2044	0,2162
D14	1	Laplace*	0,1363	0,1396	0,1596	0,1613
D14	1	Normal	0,1363	0,1418	0,1843	0,1733
D14	2	Laplace	0,1363	0,1378	0,1502	0,1471
D14	2	Laplace*	0,1363	0,1367	0,1393	0,1385
D14	2	Normal	0,1363	0,1372	0,1434	0,1412
D14	3	Laplace	0,1363	0,1367	0,1402	0,1395

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D14	3	Laplace*	0,1363	0,1364	0,1371	0,1369
D14	3	Normal	0,1363	0,1365	0,1382	0,1377
D14	5	Laplace	0,1363	0,1364	0,1371	0,1374
D14	5	Laplace*	0,1363	0,1363	0,1365	0,1365
D14	5	Normal	0,1363	0,1363	0,1369	0,1367
D14	10	Laplace	0,1363	0,1363	0,1364	0,1364
D14	10	Laplace*	0,1363	0,1363	0,1363	0,1363
D14	10	Normal	0,1363	0,1363	0,1364	0,1363
D15	0,5	Laplace	0,1919	0,2278	0,4598	0,4512
D15	0,5	Laplace*	0,1919	0,2203	0,3626	0,3948
D15	0,5	Normal	0,1919	0,2192	0,4386	0,3693
D15	1	Laplace	0,1919	0,2030	0,3065	0,3114
D15	1	Laplace*	0,1919	0,1994	0,2332	0,2259
D15	1	Normal	0,1919	0,2005	0,2492	0,2331
D15	2	Laplace	0,1919	0,1951	0,2073	0,2103
D15	2	Laplace*	0,1919	0,1930	0,1963	0,1975
D15	2	Normal	0,1919	0,1938	0,2039	0,2026
D15	3	Laplace	0,1919	0,1935	0,1983	0,2002
D15	3	Laplace*	0,1919	0,1923	0,1943	0,1939
D15	3	Normal	0,1919	0,1926	0,1955	0,1946
D15	5	Laplace	0,1919	0,1923	0,1934	0,1940
D15	5	Laplace*	0,1919	0,1920	0,1926	0,1925
D15	5	Normal	0,1919	0,1921	0,1935	0,1928
D15	10	Laplace	0,1919	0,1920	0,1923	0,1923
D15	10	Laplace*	0,1919	0,1919	0,1921	0,1920
D15	10	Normal	0,1919	0,1919	0,1921	0,1921
D16	0,5	Laplace	0,0531	0,2119	0,4816	0,5821
D16	0,5	Laplace*	0,0531	0,1765	0,3754	0,3565
D16	0,5	Normal	0,0531	0,1966	0,5833	0,5275
D16	1	Laplace	0,0531	0,1403	0,3349	0,3010
D16	1	Laplace*	0,0531	0,1179	0,2023	0,2092
D16	1	Normal	0,0531	0,1233	0,2901	0,2600
D16	2	Laplace	0,0531	0,0850	0,1723	0,1859
D16	2	Laplace*	0,0531	0,0799	0,1517	0,1317
D16	2	Normal	0,0531	0,0819	0,1711	0,1585
D16	3	Laplace	0,0531	0,0705	0,1307	0,1497
D16	3	Laplace*	0,0531	0,0631	0,1023	0,1029

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D16	3	Normal	0,0531	0,0656	0,1183	0,1127
D16	5	Laplace	0,0531	0,0584	0,0898	0,0942
D16	5	Laplace*	0,0531	0,0564	0,0724	0,0697
D16	5	Normal	0,0531	0,0565	0,0886	0,0822
D16	10	Laplace	0,0531	0,0545	0,0621	0,0653
D16	10	Laplace*	0,0531	0,0537	0,0556	0,0557
D16	10	Normal	0,0531	0,0540	0,0599	0,0576
D17	0,5	Laplace	0,0915	0,1426	0,3396	0,3950
D17	0,5	Laplace*	0,0915	0,1335	0,3137	0,2989
D17	0,5	Normal	0,0915	0,1351	0,3541	0,3545
D17	1	Laplace	0,0915	0,1066	0,1814	0,1818
D17	1	Laplace*	0,0915	0,0982	0,1492	0,1759
D17	1	Normal	0,0915	0,1016	0,1419	0,1597
D17	2	Laplace	0,0915	0,0948	0,1143	0,1155
D17	2	Laplace*	0,0915	0,0928	0,0978	0,0966
D17	2	Normal	0,0915	0,0942	0,1029	0,1052
D17	3	Laplace	0,0915	0,0926	0,0984	0,0993
D17	3	Laplace*	0,0915	0,0918	0,0932	0,0940
D17	3	Normal	0,0915	0,0922	0,0956	0,0953
D17	5	Laplace	0,0915	0,0919	0,0929	0,0929
D17	5	Laplace*	0,0915	0,0916	0,0921	0,0923
D17	5	Normal	0,0915	0,0917	0,0924	0,0925
D17	10	Laplace	0,0915	0,0915	0,0917	0,0918
D17	10	Laplace*	0,0915	0,0915	0,0916	0,0916
D17	10	Normal	0,0915	0,0915	0,0917	0,0917
D18	0,5	Laplace	0,6403	0,6411	0,6525	0,6484
D18	0,5	Laplace*	0,6402	0,6407	0,6523	0,6482
D18	0,5	Normal	0,6402	0,6411	0,6573	0,6498
D18	1	Laplace	0,6403	0,6405	0,6441	0,6436
D18	1	Laplace*	0,6402	0,6404	0,6425	0,6419
D18	1	Normal	0,6402	0,6406	0,6447	0,6436
D18	2	Laplace	0,6403	0,6403	0,6413	0,6414
D18	2	Laplace*	0,6536	0,6403	0,6408	0,6411
D18	2	Normal	0,6402	0,6403	0,6411	0,6413
D18	3	Laplace	0,6402	0,6403	0,6406	0,6407
D18	3	Laplace*	0,6402	0,6403	0,6406	0,6407
D18	3	Normal	0,6403	0,6403	0,6406	0,6411

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
D18	5	Laplace	0,6403	0,6403	0,6404	0,6406
D18	5	Laplace*	0,6420	0,6402	0,6406	0,6405
D18	5	Normal	0,6403	0,6403	0,6406	0,6403
D18	10	Laplace	0,6402	0,6402	0,6403	0,6403
D18	10	Laplace*	0,6403	0,6402	0,6403	0,6403
D18	10	Normal	0,6403	0,6402	0,6403	0,6404
R1	0,5	Laplace	0,0081	0,1117	0,4894	0,5382
R1	0,5	Laplace*	0,0081	0,1012	0,3448	0,3727
R1	0,5	Normal	0,0081	0,1166	0,5805	0,5241
R1	1	Laplace	0,0081	0,0884	0,1835	0,1994
R1	1	Laplace*	0,0081	0,0778	0,1431	0,1272
R1	1	Normal	0,0081	0,0892	0,2164	0,2559
R1	2	Laplace	0,0081	0,0648	0,1320	0,1088
R1	2	Laplace*	0,0081	0,0593	0,0981	0,0928
R1	2	Normal	0,0081	0,0649	0,1181	0,1139
R1	3	Laplace	0,0081	0,0514	0,0913	0,0949
R1	3	Laplace*	0,0081	0,0498	0,0808	0,0823
R1	3	Normal	0,0081	0,0540	0,0918	0,0955
R1	5	Laplace	0,0081	0,0395	0,0767	0,0781
R1	5	Laplace*	0,0081	0,0368	0,0728	0,0658
R1	5	Normal	0,0081	0,0370	0,0703	0,0741
R1	10	Laplace	0,0081	0,0245	0,0567	0,0539
R1	10	Laplace*	0,0081	0,0122	0,0353	0,0394
R1	10	Normal	0,0081	0,0148	0,0504	0,0445
R2	0,5	Laplace	0,0201	0,0526	0,1746	0,1724
R2	0,5	Laplace*	0,0201	0,0478	0,1025	0,1043
R2	0,5	Normal	0,0201	0,0501	0,1459	0,1499
R2	1	Laplace	0,0201	0,0294	0,0669	0,0835
R2	1	Laplace*	0,0201	0,0277	0,0487	0,0468
R2	1	Normal	0,0201	0,0288	0,0637	0,0609
R2	2	Laplace	0,0201	0,0228	0,0379	0,0370
R2	2	Laplace*	0,0201	0,0216	0,0292	0,0317
R2	2	Normal	0,0201	0,0224	0,0311	0,0409
R2	3	Laplace	0,0201	0,0213	0,0264	0,0299
R2	3	Laplace*	0,0201	0,0207	0,0221	0,0221
R2	3	Normal	0,0201	0,0212	0,0241	0,0265
R2	5	Laplace	0,0201	0,0205	0,0218	0,0225

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
R2	5	Laplace*	0,0201	0,0203	0,0205	0,0206
R2	5	Normal	0,0201	0,0203	0,0210	0,0216
R2	10	Laplace	0,0201	0,0202	0,0204	0,0204
R2	10	Laplace*	0,0201	0,0201	0,0203	0,0202
R2	10	Normal	0,0201	0,0202	0,0203	0,0203
R3	0,5	Laplace	0,0106	0,0842	0,2208	0,2881
R3	0,5	Laplace*	0,0106	0,0761	0,1367	0,1478
R3	0,5	Normal	0,0106	0,0968	0,2688	0,3429
R3	1	Laplace	0,0106	0,0640	0,1358	0,1474
R3	1	Laplace*	0,0106	0,0533	0,0920	0,0884
R3	1	Normal	0,0106	0,0660	0,1409	0,1456
R3	2	Laplace	0,0106	0,0399	0,0779	0,0914
R3	2	Laplace*	0,0106	0,0350	0,0562	0,0583
R3	2	Normal	0,0106	0,0375	0,0754	0,0747
R3	3	Laplace	0,0106	0,0337	0,0582	0,0608
R3	3	Laplace*	0,0106	0,0290	0,0419	0,0456
R3	3	Normal	0,0106	0,0307	0,0573	0,0524
R3	5	Laplace	0,0106	0,0267	0,0380	0,0415
R3	5	Laplace*	0,0106	0,0221	0,0304	0,0316
R3	5	Normal	0,0106	0,0227	0,0371	0,0400
R3	10	Laplace	0,0106	0,0186	0,0265	0,0261
R3	10	Laplace*	0,0106	0,0114	0,0143	0,0140
R3	10	Normal	0,0106	0,0136	0,0220	0,0230
R4	0,5	Laplace	0,0150	0,0433	0,1264	0,1063
R4	0,5	Laplace*	0,0150	0,0418	0,0739	0,0824
R4	0,5	Normal	0,0150	0,0431	0,1125	0,0990
R4	1	Laplace	0,0150	0,0230	0,0668	0,0657
R4	1	Laplace*	0,0150	0,0210	0,0465	0,0585
R4	1	Normal	0,0150	0,0245	0,0584	0,0564
R4	2	Laplace	0,0150	0,0172	0,0401	0,0391
R4	2	Laplace*	0,0150	0,0166	0,0249	0,0300
R4	2	Normal	0,0150	0,0177	0,0394	0,0315
R4	3	Laplace	0,0150	0,0159	0,0258	0,0265
R4	3	Laplace*	0,0150	0,0154	0,0171	0,0171
R4	3	Normal	0,0150	0,0162	0,0203	0,0204
R4	5	Laplace	0,0150	0,0153	0,0163	0,0175
R4	5	Laplace*	0,0150	0,0150	0,0154	0,0154

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
R4	5	Normal	0,0150	0,0151	0,0162	0,0156
R4	10	Laplace	0,0150	0,0150	0,0152	0,0152
R4	10	Laplace*	0,0150	0,0150	0,0150	0,0151
R4	10	Normal	0,0150	0,0150	0,0151	0,0151
R5	0,5	Laplace	0,0118	0,0402	0,1034	0,1003
R5	0,5	Laplace*	0,0118	0,0402	0,0640	0,0802
R5	0,5	Normal	0,0118	0,0420	0,0905	0,1163
R5	1	Laplace	0,0118	0,0316	0,0632	0,0659
R5	1	Laplace*	0,0118	0,0272	0,0513	0,0553
R5	1	Normal	0,0118	0,0319	0,0646	0,0607
R5	2	Laplace	0,0118	0,0236	0,0402	0,0426
R5	2	Laplace*	0,0118	0,0212	0,0324	0,0369
R5	2	Normal	0,0118	0,0194	0,0393	0,0397
R5	3	Laplace	0,0118	0,0167	0,0337	0,0345
R5	3	Laplace*	0,0118	0,0135	0,0228	0,0251
R5	3	Normal	0,0118	0,0158	0,0293	0,0313
R5	5	Laplace	0,0118	0,0135	0,0249	0,0250
R5	5	Laplace*	0,0118	0,0121	0,0128	0,0133
R5	5	Normal	0,0118	0,0127	0,0155	0,0157
R5	10	Laplace	0,0118	0,0120	0,0127	0,0125
R5	10	Laplace*	0,0118	0,0119	0,0120	0,0120
R5	10	Normal	0,0118	0,0119	0,0122	0,0122
R6	0,5	Laplace	0,1109	0,1705	0,3381	0,3444
R6	0,5	Laplace*	0,1109	0,1563	0,2713	0,2895
R6	0,5	Normal	0,1109	0,1664	0,2997	0,3015
R6	1	Laplace	0,1109	0,1286	0,2259	0,2348
R6	1	Laplace*	0,1109	0,1199	0,1741	0,1772
R6	1	Normal	0,1109	0,1227	0,2023	0,1936
R6	2	Laplace	0,1109	0,1147	0,1428	0,1378
R6	2	Laplace*	0,1109	0,1124	0,1181	0,1174
R6	2	Normal	0,1109	0,1143	0,1315	0,1292
R6	3	Laplace	0,1109	0,1128	0,1205	0,1224
R6	3	Laplace*	0,1109	0,1113	0,1129	0,1133
R6	3	Normal	0,1109	0,1121	0,1153	0,1152
R6	5	Laplace	0,1109	0,1113	0,1131	0,1128
R6	5	Laplace*	0,1109	0,1110	0,1115	0,1114
R6	5	Normal	0,1109	0,1111	0,1117	0,1125

Continua na página seguinte

Tabela G.9: Mediana do NICV obtido para cada conjunto de dados e cenário (continuação).

dataset	epsilon	mecanismo	sem PD	BRKGAPD	ILSPD	VNSPD
R6	10	Laplace	0,1109	0,1109	0,1111	0,1112
R6	10	Laplace*	0,1109	0,1109	0,1110	0,1110
R6	10	Normal	0,1109	0,1109	0,1111	0,1112
R7	0,5	Laplace	0,1300	0,1321	0,2142	0,2077
R7	0,5	Laplace*	0,1300	0,1313	0,1481	0,1524
R7	0,5	Normal	0,1300	0,1318	0,1737	0,1640
R7	1	Laplace	0,1300	0,1304	0,1396	0,1356
R7	1	Laplace*	0,1300	0,1301	0,1324	0,1326
R7	1	Normal	0,1300	0,1302	0,1377	0,1368
R7	2	Laplace	0,1300	0,1300	0,1310	0,1315
R7	2	Laplace*	0,1300	0,1300	0,1304	0,1305
R7	2	Normal	0,1300	0,1300	0,1313	0,1308
R7	3	Laplace	0,1300	0,1300	0,1305	0,1305
R7	3	Laplace*	0,1300	0,1300	0,1301	0,1301
R7	3	Normal	0,1300	0,1300	0,1303	0,1304
R7	5	Laplace	0,1300	0,1300	0,1301	0,1301
R7	5	Laplace*	0,1300	0,1300	0,1300	0,1300
R7	5	Normal	0,1300	0,1300	0,1301	0,1301
R7	10	Laplace	0,1300	0,1300	0,1300	0,1300
R7	10	Laplace*	0,1300	0,1300	0,1300	0,1300
R7	10	Normal	0,1300	0,1300	0,1300	0,1300

Tabela G.10: Resultados do teste pareado de Wilcoxon - Mecanismo de Laplace.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	0,88	Não rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,00	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(a) Laplace, $\epsilon = 0,5$.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	1,00	Não rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,00	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(b) Laplace, $\epsilon = 10$.

Tabela G.11: Resultados do teste pareado de Wilcoxon - Mecanismo de Laplace*.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	0,00	Rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,01	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(a) Laplace*, $\epsilon = 0,5$.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	0,50	Não rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,00	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(b) Laplace*, $\epsilon = 10$.

Tabela G.12: Resultados do teste pareado de Wilcoxon - Mecanismo Normal.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	0,18	Não rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,00	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(a) Normal, $\epsilon = 0,5$.

Estrategia 1	Estrategia 2	p-valor	H_0
BRKGA	ILS	0,00	Rejeita
BRKGA	VNS	0,00	Rejeita
BRKGA	DAS	0,00	Rejeita
BRKGA	Escalonada	0,00	Rejeita
BRKGA	SSAS	0,00	Rejeita
BRKGA	Uniforme	0,00	Rejeita
ILS	VNS	0,33	Não rejeita
ILS	DAS	0,00	Rejeita
ILS	Escalonada	0,00	Rejeita
ILS	SSAS	0,00	Rejeita
ILS	Uniforme	0,00	Rejeita
VNS	DAS	0,00	Rejeita
VNS	Escalonada	0,00	Rejeita
VNS	SSAS	0,00	Rejeita
VNS	Uniforme	0,00	Rejeita
DAS	Escalonada	0,00	Rejeita
DAS	SSAS	0,00	Rejeita
DAS	Uniforme	0,00	Rejeita
Escalonada	SSAS	0,00	Rejeita
Escalonada	Uniforme	0,00	Rejeita
SSAS	Uniforme	0,00	Rejeita

(b) Normal, $\epsilon = 10$.