



Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ciência da Computação

Métodos de Descoberta Adaptativa de Subconsultas para Busca Diversificada de Imagens

Noberto Pires Maciel

Feira de Santana

2024



Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ciência da Computação

Noberto Pires Maciel

Métodos de Descoberta Adaptativa de Subconsultas para Busca Diversificada de Imagens

Dissertação apresentada à Universidade
Estadual de Feira de Santana como parte
dos requisitos para a obtenção do título de
Mestre em Ciência da Computação.

Orientador: Rodrigo Tripodi Calumby

Feira de Santana

2024

Ficha catalográfica - Biblioteca Central Julieta Carteado – SISBI UEFS

M139 Maciel, Noberto Pires

Métodos de descoberta adaptativa de subconsultas para busca diversificada de imagens / Noberto Pires Maciel. – 2024.
158 f.: il.

Orientador: Rodrigo Tripodi Calumby.

Dissertação (mestrado) – Programa de Pós-Graduação em Ciência da Computação, Universidade Estadual de Feira de Santana, Feira de Santana, 2023.

1. Imagens – recuperação. 2. Busca diversificada. 3. Algoritmos – agrupamento. 4. Clusterização. 5. CBIR. 6. DTRS. I. Título. II. Calumby, Rodrigo Tripodi, orient. III. Programa de Pós-Graduação em Ciência da Computação. IV. Universidade Estadual de Feira de Santana.

CDU 004.65

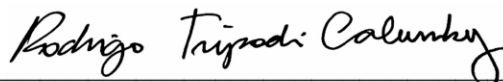
Noberto Pires Maciel

Métodos de Descoberta Adaptativa de Sub-consultas para Busca Diversificada de Imagens

Dissertação apresentada à Universidade Estadual de Feira de Santana como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Feira de Santana, 11 de junho de 2024

BANCA EXAMINADORA



Rodrigo Tripodi Calumby (Orientador(a))
Universidade Estadual de Feira de Santana



Tiago Amador Coelho
Universidade Estadual de Feira de Santana



Matheus Giovani Pires
Universidade Estadual de Feira de Santana

Abstract

Searching images for content in a data collection, whether through social media mechanisms or free web search tools, is a complex task where results based on similarity alone often present relevance problems such as unrepresentative items and near-duplicates. Commonly, search engines try to perform a broad coverage based on implicit subtopics of the query in order to serve the user as completely as possible. In this sense, the approach based on content diversification using data clustering algorithms has been widely used. In this approach, each group identified by the algorithm in the search results is treated as a subtopic. These groups are used to extract representative images that together bring diversity to the result presented to the user. However, the effectiveness of the approach depends on choosing a good clustering scheme, something that is directly linked to the number of groups generated by the algorithm, a task that has been an immense challenge. This work aims to evaluate the possible gains in terms of efficiency in the task of retrieving diverse images by selecting the best grouping schemes generated by clustering algorithms, dynamically searching for the ideal number of groups for each query. In addition, we intend to extend the literature by carrying out an experimental evaluation of the DTRS method for estimating the quality of clusters, as well as developing an efficient auxiliary method for determining the stopping criteria for clustering algorithms and, consequently, reducing the computational costs of the results diversification procedure. To this end, we conducted experiments using the K-Medoids and Hierarchical Agglomerative algorithms, employing different validation methods, exploring variations in the number of clusters and adopting different auxiliary approaches for selecting the best *clusters* schemes, such as the *Elbow's method*. The results showed gains in terms of efficiency in retrieving diverse images and a significant reduction in the running time of the CBIR system used in this work.

Keywords: clustering, cbir, dtrs, diversity

Resumo

A busca de imagens por conteúdo em uma coleção de dados, seja através de mecanismos de mídia social ou em ferramentas de busca livre na web, é uma tarefa complexa onde resultados baseados apenas em similaridade frequentemente apresentam problemas de relevância como itens pouco representativos e quase-duplicatas. Comumente, ferramentas de busca tentam realizar uma ampla cobertura baseada em subtópicos implícitos da consulta para atender ao usuário de forma mais completa possível. Neste sentido, a abordagem baseada em diversificação de conteúdo utilizando algoritmos de agrupamento de dados tem sido bastante utilizada. Nesta abordagem, cada grupo identificado pelo algoritmo nos resultados da busca é tratado como um subtópico. Estes grupos são utilizados para extrair imagens representativas que juntas tragam diversidade ao resultado apresentado ao usuário. Contudo, a eficácia da abordagem depende da escolha de um bom esquema de agrupamento, algo que está diretamente ligado ao número de grupos gerados pelo algoritmo, tarefa que tem sido um imenso desafio. Este trabalho tem como objetivo avaliar os possíveis ganhos em termos de eficácia na tarefa de recuperação de imagens diversificadas, através da seleção dos melhores esquemas de grupos gerados por algoritmos de agrupamento, buscando dinamicamente um número de grupos ideal para cada consulta. Adicionalmente, pretende-se estender a literatura realizando a avaliação experimental do método DTRS para estimativa da qualidade dos agrupamentos, bem como desenvolver um método auxiliar eficiente para determinação de critério de parada para algoritmos de agrupamento e, consequentemente, reduzir os custos computacionais do procedimento de diversificação de resultados. Para isso, conduzimos experimentos utilizando os algoritmos K-Medoids e Hierárquico Aglomerativo, empregando diferentes métodos de validação, explorando variações na quantidade de agrupamentos e adotando diferentes abordagens auxiliares para seleção dos melhores esquemas de *clusters*, como o método *Elbow*. Os resultados demonstraram ganhos em termos de eficácia na recuperação de imagens diversificadas e significativa redução do tempo de execução do sistema CBIR empregado neste trabalho.

Palavras-chave: clusterização, cbir, dtrs, diversificação

Prefácio

Esta dissertação de mestrado foi submetida à Universidade Estadual de Feira de Santana (UEFS) como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

A dissertação foi desenvolvida no Programa de Pós-Graduação em Ciência da Computação (PGCC), tendo como orientador o Prof. Dr. **Rodrigo Tripodi Calumby**.

Agradecimentos

Agradeço ao Professor **Dr. Rodrigo Tripodi Calumby** pela oportunidade dada e pela competente (além de paciente) orientação durante esses anos de muito estudo, dedicação e esforço; ao **Dr. Antônio Maurício Santana de Carvalho** por ter acreditado e oportunizado este momento; ao Professor **Dr. Angelo Conrado Loula** pelas breves e elucidadoras conversas; aos Professores **Dr. Matheus Giovani Pires** e **Dr. Tiago Amador Coelho** pela participação na banca examinadora, com suas valiosas sugestões e correções; a todos os professores das disciplinas cursadas no Programa de Pós-graduação em Ciência da Computação desta respeitada Universidade, os quais foram essenciais no desenvolvimento deste trabalho; aos colegas **Solenir Figueredo** e **Luciano Dourado** pela grande participação e pelas conversas de laboratório; à minha esposa **Thayse Bevenuto**, à minha filha **Nayla Melissa**, aos meus pais, à minha família e a todos que participaram desta jornada de alguma maneira.

Seguirei na busca pelo amadurecimento científico e acadêmico, bem como pela contínua e estimulante busca por soluções científicas na área de computação.

*“A perseverança não é uma longa
corrida; ela é muitas corridas
curtas, uma depois da outra.”*

– Walter Elliott

Sumário

Abstract	i
Resumo	ii
Prefácio	iii
Agradecimentos	iv
Alinhamento com a Linha de Pesquisa	ix
Lista de Tabelas	xiii
Lista de Figuras	xxi
Lista de Abreviações	xxii
Lista de Símbolos	xxiii
1 Introdução	1
1.1 Visão Geral	1
1.2 Objetivos	8
2 Fundamentação Teórica	9
2.1 Clusterização em CBIR	9
2.1.1 Métodos particionais	10
2.1.2 Métodos hierárquicos	11
2.1.3 Descritores e features	12
2.2 Índices de validação dos agrupamentos	13
2.2.1 Silhouette	15
2.2.2 Dunn Index	17
2.2.3 Davies-Bouldin	18
2.2.4 Outros índices de validação	20
2.2.5 Método do Cotovelo	21
2.3 Métodos de avaliação da eficácia	23
2.3.1 Precision	23

2.3.2	Cluster Recall	23
2.3.3	F-Measure	24
2.4	Decision-theoretic rough set models	24
2.4.1	Definição	24
2.4.2	Aproximação algébrica	24
2.4.3	Aproximação probabilística	25
2.5	Análise estatística dos resultados	27
2.5.1	Teste de Shapiro-Wilk	28
2.5.2	Valor-p (<i>p-value</i>)	28
2.5.3	Teste de Wilcoxon	29
2.6	Trabalhos relacionados	30
2.6.1	Recuperação de imagem baseada em conteúdo por clusterização	30
2.6.2	Diversificação visual dos resultados da busca de imagens . . .	32
2.6.3	Descoberta não supervisionada de subtópico implícito adaptável a consulta para recuperação de imagens diversificada com base na qualidade intrínseca do <i>cluster</i>	35
2.6.4	Um método automático para determinar o número de <i>clusters</i> usando <i>decision-theoretic rough sets</i>	37
3	Metodologia	41
3.1	O arcabouço	42
3.1.1	O dataset	43
3.2	Os experimentos	43
3.2.1	Os descritores utilizados	45
3.2.2	Critério de interrupção do algoritmo	45
3.2.3	Avaliação da performance	49
3.2.4	Observações estatísticas	49
3.2.5	Parametrização	50
4	Resultados	52
4.1	Avaliação do método DTRS no sistema CBIR (experimento 01)	52
4.1.1	Avaliação dos resultados de performance	53
4.1.2	Avaliação dos ganhos alcançados	62
4.2	Avaliação da abordagem dinâmica (experimentos 02 e 03)	65
4.2.1	Avaliação dos resultados de Performance	66
4.2.2	Avaliação dos ganhos alcançados	83
4.2.3	Avaliação do tempo de execução	87
5	Discussão	92
5.1	O método DTRS e o IASMIN	92
5.2	Resposta às questões de pesquisa	102
5.3	Ameaças à validade	103

6	Considerações finais	105
6.1	Trabalhos Futuros	106
	Referências	107
A	Imagens reais dos experimentos	113
A.0.1	Experimento 01: K-Medoids com $k = [15, 25]$	113
A.0.2	Experimento 02: Hierárquico Aglomerativo com $k = [1, 150]$.	115
B	Análise estatística dos experimentos	116
B.0.1	Avaliação descritiva dos resultados alcançados no experimento 01	116
B.0.2	Avaliação descritiva dos resultados alcançados nos experimen- tos 02 e 03	123
C	Resultados dos experimentos adicionais	125

Alinhamento com a Linha de Pesquisa

Linha de Pesquisa: Computação Inteligente

Neste trabalho são utilizados os conceitos e técnicas em aprendizado de máquina para tratar o problema da recuperação/obtenção de imagens, com base no conteúdo intrínseco, em um banco de dados. São tratados temas como representação de conteúdo em imagens e algoritmos de agrupamento.

Lista de Tabelas

2.1	<i>Relação intercluster</i> (entre elementos de <i>clusters</i> diferentes).	14
2.2	<i>Relação intrcluster</i> (entre elementos do mesmo <i>cluster</i>).	15
2.3	Desempenho médio em todos os tópicos, algoritmos e avaliadores humanos (Van Leuken et al., 2009)	35
2.4	Resultado do processo hierárquico aglomerativo para o <i>dataset</i> X11, ACA-DTRS (Yu et al., 2014)	39
2.5	Resultado comparativo ACA-DTRS (Yu et al., 2014) realizado para os <i>datasets</i> do repositório UCI (Asuncion e Newman, 2007). Coluna $ U $ representa o número de elementos do <i>dataset</i> . A última coluna indica o número real de agrupamentos. Demais colunas mostram o número de agrupamentos sugerido pelo respectivo índice de validação.	40
2.6	Resultado do experimento comparativo ACA-DTRS <i>vs</i> FACA-DTRS (Yu et al., 2014). A coluna $ U $ representa o número de elementos do <i>dataset</i> . A coluna $ C $ indica o número real de agrupamentos. A coluna $ C' $, o número de agrupamentos encontrado pelo respectivo algoritmo.	40
3.1	Parâmetros básicos adotados na geração do <i>baseline</i>	50
3.2	Parâmetros básicos adotados no experimento para avaliação do DTRS (item a., Seção 1.2, experimento 01)	50
3.3	Parâmetros básicos adotados nos experimentos 02 e 03 para avaliação do IASMIN (item b., Seção 1.2), eficácia do DTRS e tempo de execução.	51
4.1	Experimento 01: resultado comparativo entre todos os descritores visuais e textuais apenas para o método DTRS sem teste de significância estatística, seleção do número de grupos ideal sem uso de métodos auxiliares (pela forma natural do método, ou seja, risco mínimo). Melhores resultados em negrito.	54
4.2	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - TAMURA (Tamura et al., 1978)	55
4.3	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - SCH (Cinque et al., 2001)	55
4.4	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - PHOG (Bia-sotti et al., 2013)	56

4.5	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - LUM (Lux e Chatzichristofis, 2008)	56
4.6	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - JCD (Zagoris et al., 2010)	57
4.7	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - JACCARD (Lewis et al., 2006)	57
4.8	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - GABOR (Manjunath e Ma, 1996)	58
4.9	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - FCTH (Chatzichristofis e Boutalis, 2008b)	58
4.10	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - DICE (Dice, 1945)	59
4.11	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - COSINE . . .	59
4.12	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CNN_AD (Ionescu et al., 2014)	60
4.13	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CN3X3 (Ionescu et al., 2014)	60
4.14	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CEDD (Chatzichristofis e Boutalis, 2008a)	61
4.15	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - BM25 (Baeza-Yates et al., 1999)	61
4.16	Experimento 01: resultados K-Medoids - $k = [15, 25]$ - ACC (Penatti et al., 2012)	62
4.17	Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos gráficos 4.3 a 4.17.	62
4.18	Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica <i>Precision</i>	63
4.19	Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica <i>Cluster Recall</i>	63
4.20	Ganho médio relativo: DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica <i>F1-Measure</i>	64
4.21	Descritores e índices de validação de agrupamento aos quais o DTRS obteve ganho consistente (ganhos positivos simultaneamente nas métricas <i>Precision</i> e <i>Cluster Recall</i>).	65
4.22	Resultado comparativo entre todos os descritores visuais e textuais apenas para o método DTRS sem teste de significância estatística. Experimentos 02 e 03. Algoritmos Hierárquico Aglomerativo e K-Medoids, respectivamente. Com uso do IASMIN, melhores resultados em negrito.	67

4.23	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor TAMURA (Tamura et al., 1978)	68
4.24	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor SCH (Cinque et al., 2001)	69
4.25	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor PHOG (Biasotti et al., 2013)	70
4.26	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor LUM (Lux e Chatzichristofis, 2008)	71
4.27	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor JCD (Zagoris et al., 2010)	71
4.28	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor JACCARD (Lewis et al., 2006)	73
4.29	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor GABOR (Manjunath e Ma, 1996)	74
4.30	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor FCTH (Chatzichristofis e Boutalis, 2008b)	75
4.31	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor DICE (Dice, 1945)	76
4.32	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor COSINE	77
4.33	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CNN_AD (Ionescu et al., 2014)	78
4.34	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CN3X3 (Ionescu et al., 2014)	79
4.35	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CEDD (Chatzichristofis e Boutalis, 2008a)	80
4.36	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor BM25 (Baeza-Yates et al., 1999)	81
4.37	Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor ACC (Penatti et al., 2012)	82
4.38	Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem considerar o teste de significância estatística. Baseado nas observações dos gráficos das Figuras 4.23(a) a 4.37(a) e tabelas de resultado do experimento 02.	83
4.39	Superioridade do método DTRS por descritor	84
4.40	Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos gráficos 4.23(a) a 4.37(a) e tabelas de resultado do experimento.	86
4.41	Superioridade do método DTRS por descritor	86

4.42	Experimento 02: Hierárquico Aglomerativo, $k = [1, 150]$. Tempo médio de execução (milissegundos) para todos os descritores e 222 tópicos de consulta. Comparativo entre os índices de validação sem uso de método auxiliar (apenas a forma natural do <i>cluster validity</i>) e os métodos auxiliares para seleção do número k de grupos ideal (Método do Cotovelo e IASMIN). O valor 0,00* no DTRS indica que o índice encontrou o número de grupos ideal precocemente, ou seja, no início das iterações do algoritmo.	90
4.43	Experimento 03: K-Medoids, $k = [1, 150]$. Tempo médio de execução (milissegundos) para todos os descritores e 222 tópicos de consulta. Comparativo entre os índices de validação sem uso de método auxiliar e os métodos auxiliares para seleção do número de grupos ideal (Método do Cotovelo e IASMIN).	91
C.1	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor TAMURA	125
C.2	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor SCH . .	126
C.3	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor PHOG .	126
C.4	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor LUM . .	126
C.5	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor JCD . .	127
C.6	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor JACKARD	127
C.7	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor GABOR	127
C.8	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor FCTH .	128
C.9	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor DICE . .	128
C.10	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor COSINE	128
C.11	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CNN_AD	129
C.12	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CN3X3 .	129
C.13	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CEDD .	129
C.14	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor BM25 . .	130
C.15	Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor ACC . .	130

Lista de Figuras

1.1	Exemplo de metadado: resultado de consulta pela descrição textual para imagem extraída do <i>Flickr</i>	2
1.2	Resultado de consulta usando ferramenta de busca do <i>Google</i> pela palavra-chave ambígua <i>jaguar</i> . Houve o retorno do mamífero e do veículo no mesmo resultado, ambos têm a mesma palavra-chave nos metadados.	3
1.3	Exemplo ilustrativo de consulta por similaridade apenas com resultados relevantes.	3
1.4	Esquema básico da abordagem CBIR, adaptado de Singh e Batra (2020)	4
1.5	(a) Exemplo ilustrativo representando o resultado de uma consulta baseada apenas em similaridade. (b) Resultado de consulta após aplicação de técnicas de diversificação de conteúdo.	5
1.6	Problema da definição do número de <i>clusters</i> (Figuerêdo e Calumby, 2022)	6
2.1	Representação do resultado de um algoritmo hierárquico. a) representação de conjuntos. b) representação por dendrograma	12
2.2	Processo de extração simples baseada em histograma local (LIH) (Shimada et al., 2010)	13
2.3	Aspectos das relações intracluster e intercluster para diferentes esquemas de agrupamento.	14
2.4	Silhouette: em (a), a distância <i>intercluster</i> b_i é calculada em relação ao centro do <i>cluster</i> C_j ; em (b), a distância <i>intercluster</i> b_i é tomada em relação ao elemento mais externo, ou seja, a sua forma natural. . .	16
2.5	<i>Silhouette</i> : um valor próximo de 1 indica um bom esquema de agrupamento, ao passo que um valor -1 indica um esquema ruim. Gráficos obtidos através da biblioteca <i>SciKit Learn</i> (Buitinck et al., 2013) com dados fictícios. A linha pontilhada em vermelho é o <i>score</i> médio global.	16
2.6	Nesta Figura podemos observar alguns erros do <i>Silhouette Score</i> : em (a) o <i>score</i> obtido indica que a configuração encontrada pelo K-means (0.352) foi melhor que a configuração encontrada pelo DBscan (0.111). O mesmo acontece para a geometria em semi-círculo em (c), <i>score</i> 0.495, e em (d), <i>score</i> 0.391. Agrupamentos e <i>scores</i> obtidos através da biblioteca <i>SciKit Learn</i> (Buitinck et al., 2013)	17

2.7	Dunn Index: a distância <i>intercluster</i> deve ser a maior possível, tomada entre os elementos mais externos dos agrupamentos 2.7(a) (numerador da Eq. 2.4); a distância <i>intraccluster</i> deve ser a menor entre todos os elementos do mesmo agrupamento 2.7(b) (denominador da Equação 2.4).	18
2.8	Davies Bouldin: a relação <i>intraccluster</i> é tomada pela variância 2.8(a), enquanto que a distância <i>intercluster</i> deve ser calculada para cada centroide 2.8(b).	19
2.9	Davies Bouldin: o índice falha ao não considerar a orientação espacial da dispersão dos dados. A Figura mostra duas dispersões idênticas com orientações espaciais distintas, porém, o DB index apresenta o mesmo valor para ambos os casos (Thomas et al., 2013). A dispersão 2.9(b) parece ter características melhores, coesão e separabilidade, e deveria pontuar melhor que a 2.9(a).	20
2.10	Método do Cotovelo (<i>Elbow's method</i>) aplicado à curva de risco gerada por um <i>cluster validity</i>	22
2.11	Forma matemática do método do Cotovelo (<i>Elbow's method</i>): maior distância perpendicular entre um ponto $p(k,C)$ e a reta secante à curva do <i>cluster validity</i>	22
2.12	Figura ilustrativa exemplificando o cálculo do <i>Cluster Recall</i> para h subtópicos da consulta Q , onde foram retornados x elementos (imagens) relevantes dentre n selecionados. Fonte: autor.	24
2.13	Espaço de aproximação: divisão das regiões no modelo DTRS.	25
2.14	Gráfico teórico de uma distribuição normal (Ross, 2017). Cálculo unilateral ou unicaudal. Exemplo com nível de significância (α) de 5% (nível de confiança de 95%)	29
2.15	Diagrama do sistema CBICR (<i>Content-based Image CLUSTER Retrieval</i>) geral, uma simplificação adotada pelos autores para CBIR utilizando a abordagem CLUE (Chen et al., 2003)	31
2.16	Resultado obtido com uso do algoritmo CLUE. A imagem de consulta pode ser vista no canto superior esquerdo. A borda azul indica que a imagem foi a mais representativa do <i>cluster</i>	31
2.17	Esquema geral do algoritmo <i>Reciprocal election</i> (Van Leuken et al., 2009)	33
2.18	Resultado do agrupamento gerado pelo algoritmo <i>Reciprocal Election</i> para a busca por <i>jaguar</i>	34
2.19	Fluxo de trabalho (Figuerêdo e Calumby, 2022)	36
2.20	Fluxo de trabalho (Figuerêdo e Calumby, 2022).	36
2.21	Fluxo de trabalho, descoberta de subtópico implícito, exemplo de seleção do número de grupos ideal (<i>best K</i>) através do <i>Sillhouette Score</i> exemplificado no artigo (Figuerêdo e Calumby, 2022)	37
2.22	Curva da função de análise de risco, ponto de parada (Yu et al., 2014)	38
2.23	Curva da função de análise de risco para o <i>dataset</i> X11, ACA-DTRS (Yu et al., 2014)	39

3.1	Diagrama do arcabouço proposto. Baseado no trabalho relacionado da Subseção 2.6.3 (Figuerêdo e Calumby, 2022)	42
3.2	Critério de interrupção do algoritmo hierárquico aglomerativo baseado no comportamento do coeficiente angular da curva da média móvel extraída dos valores gerados pelo <i>cluster validity</i> , pontos de inflexão positiva (Figura 3.2(a)) e negativa (Figura 3.2(b)) para os casos de curvas convexas e côncavas.	46
3.3	Curva gerada pelos valores obtidos pelo índice de validação de agrupamento <i>Xie Beni</i> (número de grupos ideal pelo valor mínimo) para o primeiro tópico de consulta do experimento e k partindo de 100 <i>clusters</i> . <i>Best k</i> em 97. Algoritmo Hierárquico Aglomerativo.	47
3.4	Gráfico ilustrativo do cálculo para um ponto de inflexão negativa pela abordagem proposta.	48
3.5	Diagrama de fluxo da etapa de execução do critério de interrupção (IASMIN).	49
4.1	Diagrama em blocos dos experimentos: parâmetros básicos e objetivos.	52
4.2	Diagrama: objetivo e parâmetros do experimento 01	53
4.3	Performance: descritor TAMURA. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i>	55
4.4	Performance: descritor SCH. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i>	55
4.5	Performance: descritor PHOG. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n = 30$ da métrica <i>Precision</i>	56
4.6	Performance: descritor LUM. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i>	56
4.7	Performance: descritor JCD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n \geq 30$ da métrica <i>Precision</i>	57
4.8	Performance: descritor JACCARD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n \geq 40$ da métrica <i>Precision</i>	57
4.9	Performance: descritor GABOR. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n = 5$ da métrica <i>Precision</i>	58
4.10	Performance: descritor FCTH. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n = 30$ da métrica <i>Precision</i>	58
4.11	Performance: descritor DICE. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS nas faixas de ranking $n = 10$, $n = 20$ e $n = 40$ da métrica <i>Precision</i> . .	59

4.12	Performance: descritor COSINE. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n = 10$ da métrica <i>Precision</i>	59
4.13	Performance: descritor CNN_AD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS na faixa de ranking $n = 5$ da métrica <i>Precision</i>	60
4.14	Performance: descritor CN3x3. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i>	60
4.15	Performance: descritor CEDD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i> . Destaque para o método DTRS entre as faixas de ranking $n \geq 20$ da métrica <i>Precision</i>	61
4.16	Performance: descritor BM25. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do <i>cluster validity</i>	61
4.17	Performance: descritor ACC. Algoritmo K-Medoids, $k = [15, 25]$. Mé- todo natural do <i>cluster validity</i>	62
4.18	Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica <i>Precision</i> . Em verde, superioridade do DTRS.	63
4.19	Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica <i>Cluster Recall</i> . Em verde, superioridade do DTRS.	64
4.20	Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica <i>F1-Measure</i> . Em verde, superioridade do DTRS.	64
4.21	Diagrama: objetivo e parâmetros do experimento 02	65
4.22	Diagrama: objetivo e parâmetros do experimento 03	66
4.23	Performance: descritor TAMURA. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	68
4.24	Performance: descritor SCH. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	69
4.25	Performance: descritor PHOG. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	70
4.26	Performance: descritor LUM. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	71
4.27	Performance: descritor JCD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	72
4.28	Performance: descritor JACCARD. Experimentos 02 e 03, $k =$ $[1, 150]$ e uso do IASMIN.	73
4.29	Performance: descritor GABOR. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	74
4.30	Performance: descritor FCTH. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	75
4.31	Performance: descritor DICE. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	76

4.32	Performance: descritor COSINE. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	77
4.33	Performance: descritor CNN_AD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	78
4.34	Performance: descritor CN3X3. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	79
4.35	Performance: descritor CEDD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	80
4.36	Performance: descritor BM25. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	81
4.37	Performance: descritor ACC. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.	82
4.38	Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para as métricas <i>Precision</i> , <i>Cluster Recall</i> e <i>F1-Measure</i> (todas as faixas de ranking). Em verde, superioridade do DTRS. Experimento 02 (algoritmo Hierárquico Aglomerativo).	85
4.39	Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para as métricas <i>Precision</i> , <i>Cluster Recall</i> e <i>F1-Measure</i> (todas as faixas de ranking). Em verde, superioridade do DTRS. Experimento 03 (algoritmo K-Medoids).	87
4.40	Experimento 02: curva gerada pelos resultados do método DTRS (curva de risco em azul) para o tópico de consulta <i>angel_of_the_north</i> . Achatamento com demais curvas de média móvel (vermelho) e coeficiente angular (verde). Descritor ACC. O número de grupos ideal ($k = 147$) foi encontrado em 4151829 nanosegundos (0,04151829 segundos).	88
4.41	Experimento 02: comportamento da média móvel e do coeficiente angular da reta secante à média móvel para valores da curva de risco gerada pelo método DTRS. Tópico $ID = 1$ (<i>angel_of_the_north</i>) e descritor ACC. A interrupção do algoritmo logo nos primeiros esquemas de <i>cluster</i> fez com que o tempo de execução do sistema CBIR fosse o menor possível.	89
5.1	Nível de eficácia médio para a métrica <i>Precision</i> , comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o TAMURA, SCH, LUM e GABOR.	93

5.2	Nível de eficácia médio para a métrica <i>Cluster Recall</i> , comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o TAMURA, SCH, LUM e GABOR.	94
5.3	Nível de eficácia médio para a métrica <i>F1-Measure</i> , comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o LUM.	94
5.4	Nível de eficácia médio para a métrica <i>Precision</i> , comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 02 para todos os descritores do estudo.	95
5.5	Nível de eficácia médio para a métrica <i>Cluster Recall</i> , comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 03 para todos os descritores do estudo.	96
5.6	Nível de eficácia médio para a métrica <i>F1-measure</i> , comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 02 nos descritores visuais TAMURA, SCH, LUM, CNN_AD, CN3X3, ACC e textua COSINE. Para o experimento 03, superioridade do DTRS nos descritores visuais PHOG, JCD, GABOR, FCTH, CEDD e textuais DICE, BM25 e JACCARD.	96
5.7	Diagrama: objetivo e parâmetros dos experimentos adicionais	97
5.8	Gráfico comparativo de performance dos experimentos adicionais. Algoritmo Hierárquico Aglomerativo com $k = [1, 150]$ e as três abordagens para seleção do número de grupos ideal: apenas com uso dos índices de validação de agrupamento, em sua forma natural (critérios dos índices de <i>cluster validity</i> , baseados em valores mínimos ou máximos); com índices de validação e auxílio do método do Cotovelo (<i>Elbow's method</i>); com índices de validação e auxílio do IASMIN. Valores médios das faixas $n = 5$ a $n = 50$ do ranking.	99

5.9	Comportamento das distribuições de resultados e coeficiente angular da reta secante à média móvel para o índice de validação de agrupamento <i>Dunn Index</i> , tópico $ID = 4$ (<i>la_madeleine</i>) e descritor ACC. Número de grupos ideal em $k = 96$	101
5.10	Comportamento das distribuições de resultados e coeficiente angular da reta secante à média móvel para o índice de validação de agrupamento <i>Dunn Index</i> , tópico $ID = 3$ (<i>la_madeleine</i>) e descritor ACC. Comportamento atípico, impossibilidade de detecção do ponto de inflexão.	102
A.1	Imagens reais do experimento 01 (K-Medoids, $k = [15, 25]$, usando apenas os índices de validação sem método auxiliar. Resultados para o descritor JCD em duas consultas (<i>Angel_of_north</i> e <i>Palacio_barolo</i>) nas métricas <i>Precision</i> e <i>Cluster Recall</i> para a faixa de ranking $n = 10$. Comparativo entre o <i>baseline</i> e o DTRS. No quadro, a colocação da imagem na faixa de ranking, o id da imagem e o grupo semântico pertencente. Falsos positivos marcados em vermelho. Tópicos de consulta com melhores performances para o descritor.	114
A.2	Imagens reais do experimento 02 (Hierárquico Aglomerativo, $k = [1, 150]$, DTRS e IASMIN). Resultados para o descritor JCD em duas consultas (<i>Angel_of_north</i> e <i>place_de_la_republique</i>) nas métricas <i>Precision</i> e <i>Cluster Recall</i> para a faixa de ranking $n = 10$. Comparativo entre o <i>baseline</i> e o DTRS. No quadro, a colocação da imagem na faixa de ranking, o id da imagem e o grupo semântico pertencente. Falsos positivos marcados em vermelho. Tópicos de consulta com melhores performances para o descritor.	115
B.1	Comportamento dos resultados para todos os descritores e índices de validação agrupados para as métricas <i>Precision</i> e <i>Cluster Recall</i> . (a) e (b) Q.Q. Plot. (c) e (d) histogramas das distribuições.	118
B.2	Comportamento dos resultados por descritores e índices de validação na métrica <i>Precision</i> . O gráfico B.2(a) mostra a distribuição geral dos resultados no ranking $n = 5$. O gráfico B.2(b) (<i>Q.Q. Plot</i>) indica distribuição não-normal e assimetria negativa.	119
B.3	Comportamento dos resultados para todos os descritores e métodos de validação, métrica <i>Precision</i> , segmento final do ranking. O gráfico B.3(a) mostra a distribuição geral dos resultados no ranking $n = 50$. O gráfico B.3(b) (<i>Q.Q. Plot</i>) e o boxplot em B.3(d) indicam distribuição não-normal e assimétrica positiva.	120
B.4	Comportamento dos resultados para todos os descritores e métodos de validação, métrica <i>Cluster Recall</i> , primeiro segmento de ranking. O gráfico B.4(a) mostra a distribuição geral dos resultados no ranking $n = 5$. O gráfico B.4(b) (<i>Q.Q. Plot</i>) indica distribuição não-normal e assimétrica negativa, de acordo com o histograma B.4(c).	121

B.5	Análise para a métrica <i>Cluster Recall</i> no segmento $n = 50$. O gráfico B.5(a) mostra a distribuição geral para esta faixa e ranking. O gráfico B.5(b) (<i>Q.Q. Plot</i>) indica distribuição potencialmente normal, com assimetria negativa. O boxplot em B.5(d) com mediana tendente ao terceiro quartil e sem a presença de <i>outliers</i> . Teste de normalidade realizado com valores médios.	122
C.1	Gráfico comparativo da performance dos experimentos adicionais. Algoritmo K-Medoids, $k = [15, 25]$ e as três abordagens para seleção do número de grupos ideal: apenas com uso dos índices de validação de agrupamento, em sua forma natural (critérios dos índices de <i>cluster validity</i> , baseados em valores mínimos ou máximos); com índices de validação e auxílio do método do Cotovelo; com índices de validação e auxílio do IASMIN. Valores médios das faixas $n = 5$ a $n = 50$ do ranking.	131

Lista de Abreviações

Abreviação	Descrição
<i>AHC</i>	Agglomerative Hierarchical Clustering
<i>CBICR</i>	Content-based Image CLUSTER Retrieval
<i>CBIR</i>	Content-based Image Retrieval
<i>CLUE</i>	CLUster-based rETrieval of images by unsupervised learning
<i>DBSCAN</i>	Density-Based Spatial Clustering of Applications with Noise
<i>DHC</i>	Divisive Hierarchical Clustering
<i>DTRS</i>	Decision-theoretic rough sets
<i>IASMIN</i>	Interrupção pela Análise da reta Secante da Média móvel do INdice
<i>QBIC</i>	Query by image content
<i>RF</i>	Relevance Feedback
<i>SSE</i>	Sum of Squared Error
<i>TBIR</i>	Text-based image retrieval

Lista de Símbolos

Símbolo	Descrição
λ	Função de perda (<i>loss function</i>)

Capítulo 1

Introdução

“O que sabemos é uma gota; o que ignoramos é um oceano.”

– Isaac Newton

Diante da evolução das tecnologias de captura e armazenamento de dados multimídia, da sua recente popularização e do crescente volume que esses dados representam no tráfego das redes por computador, a implementação de ferramentas para recuperação, classificação e recomendação de conteúdo multimídia tem se mostrado cada vez mais necessária (Calumby et al., 2017). As aplicações e as áreas do conhecimento favorecidas pelo uso de tais ferramentas são as mais diversas: diagnóstico de doenças na medicina; análise de dados de satélite; busca em repositórios de vídeo e imagens; na análise forense digital; sistemas de vigilância; dentre outros (Tyagi, 2017). Ou seja, pode trazer benefícios a toda e qualquer aplicação ou área cuja utilização de dados multimídia se faça necessária através da recuperação e/ou classificação por conteúdo semântico, sobretudo, quando em grandes coleções.

1.1 Visão Geral

A tarefa de recuperação de imagens pode ser definida como a coleta e apresentação de imagens semanticamente relevantes a partir de uma consulta realizada sobre um *dataset*. Neste sentido, podemos destacar algumas abordagens comumente utilizadas para o desenvolvimento de ferramentas de recuperação como a *Text-based image retrieval* (TBIR) . Esta abordagem (que é a mais comum) é baseada em anotações textuais de descrição do conteúdo multimídia em um banco de dados (metadados). Os metadados podem ser definidos como dados sobre o dado (informações acerca dos dados armazenados) e podem ser de diferentes tipos, como: coordenadas de GPS, data e hora ou palavras-chave. Tipicamente, utiliza-se metadados obtidos de

forma manual ou automática. Na Figura 1.1 podemos ver um exemplo desse tipo de abordagem.

Os sistemas TBIR são focados em texto conectado a imagens como forma de representá-las (Tyagi, 2017). Costumam apresentar os melhores resultados de consultas e um custo computacional baixo devido ao seu caráter textual. No entanto, descrever imagens manualmente é uma tarefa complicada e cara para conjuntos de dados muito extensos. Algo que pode tornar essa abordagem muitas vezes inviável. Além disso, expressar o conteúdo visual das imagens em palavras pode ser muito difícil, o que pode fazer com que a TBIR produza resultados irrelevantes ou mesmo inesperados (Tyagi, 2017). A performance de um sistema de recuperação TBIR depende diretamente da comparação de palavras-chave nos metadados das imagens de retorno e do texto da consulta realizada pelo usuário.

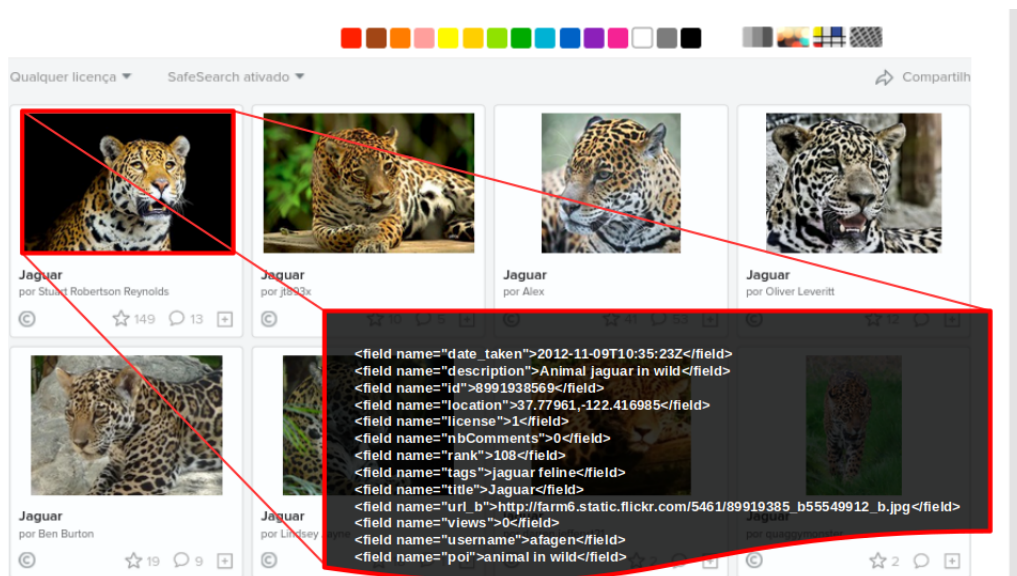


Figura 1.1: Exemplo de metadado: resultado de consulta pela descrição textual para imagem extraída do *Flickr*.

Todavia, usuários apresentam uma grande subjetividade na representação de suas necessidades ao especificar uma consulta. O mesmo se aplica em relação às anotações para composição de metadados. Neste último caso, imaginemos um grupo de anotadores responsável por catalogar uma grande quantidade de fotografias de pontos turísticos de uma determinada cidade. Eles devem vincular palavras-chave a cada uma das imagens (manualmente e de acordo com o que é visto nelas) para uma posterior consulta em um sistema TBIR. Não é difícil perceber que cada usuário terá sua percepção própria ao catalogar as fotografias e isso influenciará na definição das palavras-chave de cada imagem e nos resultados apresentados aos usuários de consulta. A Figura 1.2 mostra um exemplo de ambiguidade nos resultados da consulta textual.

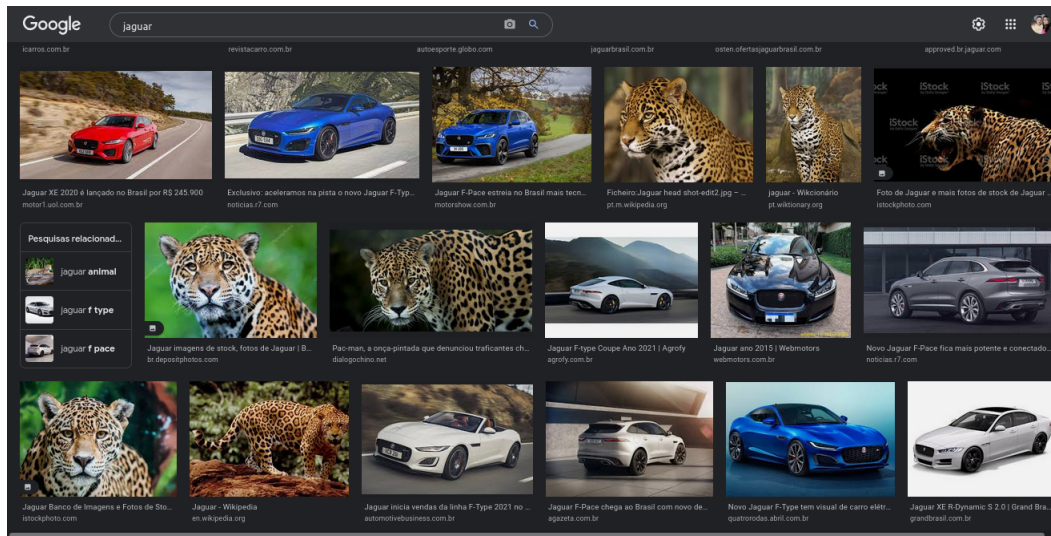


Figura 1.2: Resultado de consulta usando ferramenta de busca do *Google* pela palavra-chave ambígua *jaguar*. Houve o retorno do mamífero e do veículo no mesmo resultado, ambos têm a mesma palavra-chave nos metadados.

Então, como uma tentativa de solucionar algumas das dificuldades encontradas na recuperação textual de imagens, surge a CBIR (*Content-based Image Retrieval*) (Kato, 1992), também conhecida como *Query by image content* (QBIC), é uma abordagem de recuperação que se baseia em características intrínsecas das imagens (*features*) e não apenas em metadados. Nela, a recuperação de imagens (ou vídeos) é feita por similaridade e os elementos do conjunto de dados (instâncias) mais similares ao elemento da consulta são retornados. A análise dessa similaridade é feita através de atributos como cores, formato, textura, etc, que são agrupados em vetores empregados na obtenção das matrizes de distância (ou similaridade) geradas com o uso de alguma função como, por exemplo, a Euclidiana. A Figura 1.3 ilustra uma consulta por similaridade em CBIR.



Figura 1.3: Exemplo ilustrativo de consulta por similaridade apenas com resultados relevantes.

Um esquema básico de uma abordagem em CBIR pode ser visto na Figura 1.4. Neste esquema, é feito um pré-processamento na etapa inicial do fluxo de trabalho. O pré-processamento envolve a extração das características intrínsecas das imagens. Nesta etapa também pode ser realizada uma filtragem para eliminação de imagens pouco relevantes de acordo com critérios menos intrínsecos como, por exemplo, o nome do arquivo.

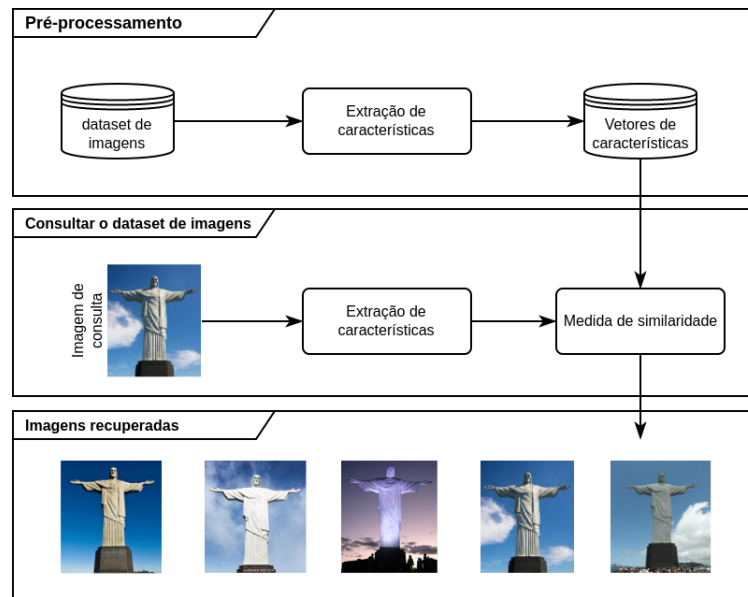


Figura 1.4: Esquema básico da abordagem CBIR, adaptado de Singh e Batra (2020)

Os resultados da etapa de pré-processamento são os vetores de características. Na etapa seguinte (etapa de consulta) a extração de características é feita para a imagem submetida pelo usuário e a matriz de distâncias é criada. As imagens mais relevantes são apresentadas na etapa final, a qual poderá apresentar o *ranking* de similaridade entre a imagem de consulta e as imagens retornadas da coleção.

Porém, consultas baseadas apenas em similaridade podem apresentar resultados pouco satisfatórios com itens não representativos ou mesmo duplicados (Calumby et al., 2017). Esses itens representativos podem ser definidos como elementos de retorno que estão fortemente ligados ao elemento da consulta mas, não necessariamente, iguais a ele.

Uma alternativa, então, é a aplicação de técnicas de diversificação de conteúdo para que resultados mais satisfatórios sejam obtidos, oferecendo uma melhor experiência ao usuário com a apresentação de um retorno livre de elementos quase-duplicatas e rico em elementos variados dentro do contexto da busca, como pode ser visto na Figura 1.5(b) (Zhao e Glotin, 2009).



Figura 1.5: (a) Exemplo ilustrativo representando o resultado de uma consulta baseada apenas em similaridade. (b) Resultado de consulta após aplicação de técnicas de diversificação de conteúdo.

Técnicas em aprendizado de máquina têm sido utilizadas para promover a diversificação na recuperação de conteúdo multimídia, a exemplo dos algoritmos de agrupamento de dados e de aprendizado profundo (Rohm et al., 2018), além de técnicas como Fusão de Dados (Zhang et al., 2012) e *Relevance Feedback (RF)* (Calumby et al., 2017). Na abordagem *RF*, a performance está diretamente ligada à avaliação do usuário após cada retorno de consulta, ou seja, existe a participação humana no processo de seleção. Por outro lado, notadamente, muitos trabalhos têm utilizado algoritmos de agrupamento de dados para a tarefa de recuperação de imagens diversificadas como em Figuerêdo e Calumby (2022), Yu et al. (2014) e Van Leuken et al. (2009) de forma automatizada ou semi-automatizada.

Os algoritmos de agrupamento destacam-se por permitir a descoberta de padrões ocultos em um conjunto de dados. Elementos com características similares são atribuídos a um mesmo grupo e, portanto, elementos de grupos diferentes podem ser ditos dissimilares. Além disso, algoritmos de agrupamento oferecem boa performance em bancos de dados extensos, não estruturados, com alta dimensionalidade e com múltiplos tipos de dados (Rasyid e Andayani, 2018). Contudo, a utilização desses algoritmos traz alguns problemas relacionados ao tipo de método escolhido

como, por exemplo, a definição do número de *clusters* nos algoritmos baseados em particionamento ou a região de aglutinação (limiar) nos algoritmos baseados em densidade, algo que, normalmente, é definido no início da sua execução e pode mudar drasticamente o esquema de agrupamentos final. A escolha desses parâmetros iniciais de forma a obter um esquema de agrupamentos ótimo tem sido um imenso desafio (Rodriguez et al., 2019).

Em algoritmos particionais como o *k-Means* (Jain, 2010) e o *k-Medoids* (Kaufman e Rousseeuw, 2009), os centros de *cluster* iniciais podem ser escolhidos aleatoriamente baseados na quantidade de agrupamentos pretendida e definida no início da sua execução. A qualidade desses agrupamentos dependerá, então, do valor definido (arbitrariamente) para um K número de grupos e da distribuição aleatória dos centros desses grupos. Nesses algoritmos, os elementos vizinhos ao atual centro espacial (centroide) ou o elemento mais central existente no grupo (medoide) são selecionados e agrupados para calcular um novo centro. Este processo é repetido várias vezes para cada um dos agrupamentos anteriormente definidos até que uma determinada condição de parada seja atingida. A Figura 1.6 mostra um exemplo onde o conjunto de dados possui cinco grupos conhecidos, destacados em cores, mas o algoritmo de clusterização é capaz de encontrar várias outras composições de acordo com os parâmetros escolhidos.

Numa abordagem *CBIR*, o elemento central do *cluster* ao qual o elemento submetido à consulta possua a maior similaridade e, portanto, a menor dissimilaridade no final da execução do algoritmo pode ser definido como o elemento mais representativo do esquema de recuperação. Ao passo que, os elementos centrais dos demais *clusters* podem representar a diversidade, pois, são proporcionalmente dissimilares. Dentro de cada *cluster*, os demais elementos podem ser ordenados em nível decrescente de similaridade. Neste caso, cada agrupamento gerado pode ser entendido como um subtópico da consulta e os seus elementos como variações.

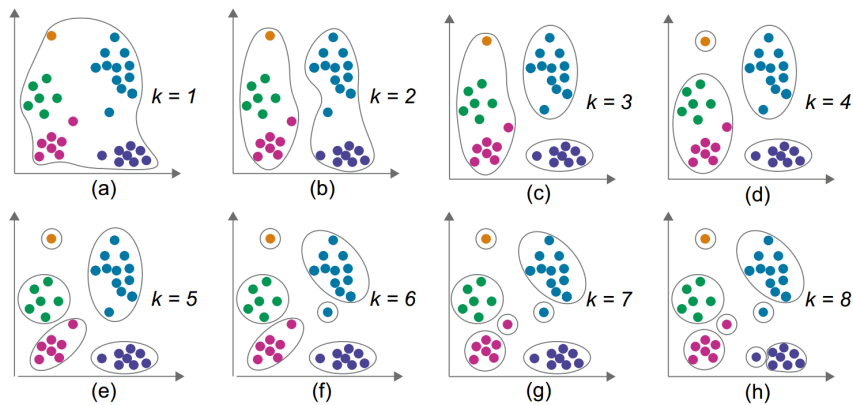


Figura 1.6: Problema da definição do número de *clusters* (Figuerêdo e Calumby, 2022)

No contexto da CBIR, estudos indicam que os usuários se concentram na diversificação dos primeiros 20 elementos da lista retornada pela consulta (Zhao e Glotin, 2009). Desta forma, algumas abordagens têm adotado essa quantidade de *clusters* ($k = 20$) como sendo a ideal. Contudo, essa definição não pode ser encarada como regra, uma vez que algoritmos de clusterização dependem de um conjunto de parâmetros complexos, testes de validação, além da composição do conjunto de dados utilizado, o que pode levar a soluções diferentes para o mesmo número K de agrupamentos. Portanto, um valor fixo de *clusters* para todas as consultas, na prática, não tem se mostrado como o ideal.

Além disso, validar a qualidade dos esquemas de agrupamento apresentados pelos algoritmos também tem sido um grande desafio. Diversos métodos (índices de validação de agrupamento) como o *Silhouette* (Rousseeuw, 1987), o *Dunn Index* (Dunn, 1973) e o *Davies-Bouldin* (Davies e Bouldin, 1979) podem ser utilizados para realizar a análise da qualidade dos agrupamentos em termos de homogeneidade e dispersão intra e inter-grupo dos esquemas gerados. Contudo, estudos têm demonstrado a necessidade da adoção de índices de validação que possam estimar a qualidade dos agrupamentos levando em consideração critérios geométricos e não apenas medidas de homogeneidade e dispersão (Kovács et al., 2005) (Rojas-Thomas et al., 2017) (Bezdek et al., 1997). É comum que métodos de validação diferentes apresentem resultados divergentes para um mesmo esquema de agrupamentos. Neste aspecto, o *Decision-theoretic Rough Sets* (DTRS) (Yu et al., 2014) surge como uma alternativa. Utilizando a teoria dos conjuntos aproximados de Pawlak (1982) em uma abordagem estendida ao teorema probabilístico de Bayes e Price (1763), o DTRS busca estimar o risco de um esquema de agrupamento, avaliando a sua qualidade através da função de perda (*loss function*) e pela probabilidade de pertencimento dos elementos do conjunto de dados aos grupos encontrados pelo algoritmo de agrupamento. No DTRS, um bom esquema de agrupamentos é aquele cujo risco calculado é o menor possível.

Assim, podemos ver que a tarefa de recuperar imagens representativamente diversificadas em uma base de dados está diretamente ligada ao método de agrupamento escolhido, à sua correta parametrização e à validação dos esquemas de *cluster* obtidos.

1.2 Objetivos

O objetivo deste trabalho é apresentar uma abordagem fundamentada no método *Decision-theoretic Rough Sets* (DTRS) como índice de validação de agrupamento (Yu et al., 2014), ou seja, na seleção do número de grupos ideal obtido pelo ajuste dos parâmetros em algoritmos de agrupamento, como o intervalo de *clusters* k , a fim de aprimorar a tarefa de recuperação e diversificação de imagens, utilizando, além da adaptação do processo do *framework* proposto por Figuerêdo e Calumby (2022), uma abordagem sem qualquer interferência manual, com validação de agrupamento automatizado, de forma adaptável à consulta, visando a obtenção de resultados superiores em CBIR.

Mais especificamente:

- a. Aplicar o método DTRS estendido (Yu et al., 2014), como índice de validação de agrupamentos (*cluster validity*), ao *framework* proposto por Figuerêdo e Calumby (2022) com uso do algoritmo *K-Medoids*;
- b. Implementar uma abordagem dinâmica (automatizada) através da análise do comportamento da curva gerada pelos resultados dos índices de validação de agrupamentos (*cluster validity*), como critério de parada, em um algoritmo de clusterização hierárquico aglomerativo;
- c. Realizar a análise comparativa entre o DTRS estendido e demais métodos bem estabelecidos na literatura.

Então, de acordo com os objetivos elencados, esperamos, com a conclusão deste trabalho, responder às seguintes questões de pesquisa:

- Q1:** Qual o desempenho do método DTRS quando utilizado como índice de validação de agrupamento no contexto da CBIR?
- Q2:** Qual o impacto do método DTRS no tempo de execução do sistema CBIR estudado?
- Q3:** Qual o impacto da adoção da abordagem automatizada no tempo de execução do sistema CBIR estudado?
- Q4:** Qual o impacto na eficácia do sistema CBIR estudado pela ampliação do intervalo de análise do número de agrupamentos?

Capítulo 2

Fundamentação Teórica

“Compreender nosso mundo requer conceituar as semelhanças e diferenças entre as entidades que o compõem.”

– Tyron and Bailey, 1970

Neste capítulo será apresentada a fundamentação teórica necessária à compreensão e execução deste trabalho com base numa revisão bibliográfica dentro do espectro temático e metodológico.

2.1 Clusterização em CBIR

A clusterização é uma técnica de aprendizado não-supervisionado que tem o objetivo de encontrar padrões ocultos em um conjunto de dados. Nesta técnica, os elementos de uma fonte de dados são agrupadas em subconjuntos por meio do cálculo da distância entre suas características (Rokach e Maimon, 2005). Existem inúmeros métodos para agrupamento de dados como:

- K-Means, que é baseado na média das distâncias dos elementos do conjunto, agrupando-os ao redor de centroides (Jain, 2010);
- K-Medoids, baseado na distância a medoides escolhidos entre os elementos do conjunto (Rousseeun e Kaufman, 1987);
- *DBSCAN*, baseado no pertencimento de um elemento à região definida (Ester et al., 1996);

- Hierárquicos Divisivos ou Aglomerativos, algoritmos baseados em conectividade (*Single-link*, *Complete-link* e *Average-link*) (Rokach e Maimon, 2005);
- Grid-based, que é baseado em separar o conjunto de dados com uma representação em grade (Rokach e Maimon, 2005);
- Graph-based, algoritmo baseado em grafos (Rokach e Maimon, 2005), dentre outros.

O cálculo da dissimilaridade $d(i,j)$, onde i e j são elementos pertencentes a um conjunto qualquer, pode ser feito com base em características (*features*) extraídas previamente desses elementos em composição com alguma função de distância como a *Euclidiana*, a *Manhattan* (casos lineares) ou uma *Geodésica* (casos não-lineares) (Kaufman e Rousseeuw, 2009).

2.1.1 Métodos particionais

Os métodos de particionamento (ou particionais) realocam instâncias movendo-as de um *cluster* para outro a partir de um particionamento inicial (Rokach e Maimon, 2005). De forma típica, requerem que um determinado número de agrupamentos seja definido inicialmente. Dentre os algoritmos baseados em métodos de particionamento, podemos citar o *K-Means* e o *K-Medoids*.

K-Means

O *K-Means* é um algoritmo de clusterização baseado em centroide. Isto significa que para encontrar padrões através do agrupamento de elementos do conjunto de dados ao redor de centros espaciais (até mesmo centros de massa) aproximados por alguma medida de distância ou dissimilaridade. Conforme descrito (Rokach e Maimon, 2005), esses *centroides* são o ponto médio entre os elementos pertencentes ao agrupamento em questão. Um novo elemento é escolhido para um grupo de acordo com a menor dissimilaridade entre ele e os centros de grupo até que não hajam mais elementos para agrupar ou que algum critério de parada estabelecido seja atingido pelo algoritmo.

A quantidade de grupos e os *centroides* são definidos no início da execução de forma arbitrária, sendo que a posição inicial pode ser totalmente aleatória. Isso faz com que performance do *K-Means* seja dependente desses dois atributos principais (posicionamento dos *centroides* e quantidade k de agrupamentos), algo que pode levar à uma convergência inesperada.

O *K-Means* não é indicado para tratar dados de diferentes tipos e possui uma baixa tolerância a *outliers*, a configuração dos agrupamentos pode ser bastante distorcida se houver a existência de pontos fora da curva. Isso acontece devido sua abordagem baseada em *centroides*, calculados através da distância média entre os pontos do conjunto.

K-Medoids

O *K-Medoids* possui o funcionamento semelhante ao *K-Means*. A diferença entre os dois reside na abordagem adotada para a escolha dos centros de cluster. Enquanto que o *K-Means* calcula o centro espacial pela média entre os elementos do conjunto definindo um elemento fictício central, o *K-Medoids* seleciona o novo centro (medoide) de acordo com o elemento cuja média das distâncias entre todos os outros elementos do conjunto seja a menor, ou seja, o centro de cluster é um elemento real pertencente ao conjunto. Essa abordagem faz com que o *K-Medoids* tenha uma boa tolerância a *outliers* diferentemente do *K-Means*.

2.1.2 Métodos hierárquicos

Os métodos hierárquicos constroem agrupamentos através da aglutinação ou particionamento recursivo das instâncias de um conjunto de elementos. Nesta categoria podemos ter os aglomerativos ou divisivos, *top-down* ou *bottom-up*, respectivamente. O resultado dos métodos hierárquicos é um dendrograma, que é um diagrama representativo do agrupamento aninhado de objetos e níveis de similaridade. Na Figura 2.1(b) temos um exemplo desse tipo de diagrama. Um agrupamento dos objetos de dados é obtido cortando o dendrograma no nível de similaridade desejado (Rokach e Maimon, 2005).

Em um algoritmo hierárquico aglomerativo, inicialmente, cada um dos elementos pertencentes ao *dataset* é considerado como um cluster; Em seguida, o algoritmo identifica os elementos mais próximos entre si e agrupa-os em um novo cluster. Para os elementos restantes é repetido o passo anterior. Os elementos mais próximos aos *clusters* são atribuídos em nível hierárquico superior. O algoritmo finaliza quando todos os elementos pertencerem a grupos ou quando for atingido um critério de parada.

O agrupamento hierárquico divisivo faz o caminho inverso ao aglomerativo. Inicialmente, todos os elementos pertencem a um único cluster. Gradualmente, o cluster é dividido em sub-clusters de forma recursiva até que o esquema de *clusters* desejado seja alcançado.

Para aglomerar ou dividir, o algoritmo pode utilizar alguma medida de similaridade entre os elementos do conjunto. Podemos ter ainda algumas maneiras para calcular essa similaridade como: *Single-link*, menor distância entre elementos de *clusters* diferentes; *Complete-link*, maior distância entre elementos de *clusters* diferentes; e *Average-link*, distância média entre quaisquer elementos de *clusters* diferentes (Rokach e Maimon, 2005).

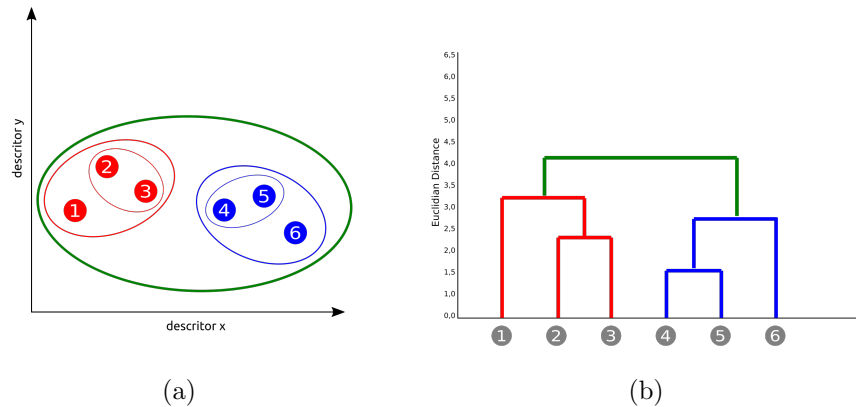


Figura 2.1: Representação do resultado de um algoritmo hierárquico. a) representação de conjuntos. b) representação por dendrograma

Neste trabalho, utilizaremos dois dos algoritmos de aprendizado não-supervisionado, para agrupamento de dados, amplamente utilizados na tarefa de recuperação de imagens baseada em conteúdo: o particional K-Medoids e o hierárquico Aglomerativo.

2.1.3 Descritores e features

As chamadas *Features*, são características de baixo nível das imagens (atributos) que, normalmente, são agrupadas em descritores (vetores de atributos) os quais podem carregar informações como densidade de cor, textura, tamanho, luminosidade, arranjo estrutural, etc. De maneira geral, esses descritores são obtidos através de algoritmos de extração de *features* aplicados às imagens. Uma vez extraídas essas *features* do conjunto de imagens, poderemos utilizá-las para realizar uma análise comparativa.

Há duas maneiras básicas de se extrair atributos: de forma global, na qual o conteúdo inteiro da imagem é analisado, extraído, e representado pelo descritor; ou de forma local, onde a imagem é dividida em sub-imagens e cada região poderá ter seu conteúdo representado pelo descritor separadamente. A Figura 2.2 mostra um exemplo de *features* (dados de histograma de intensidade) extraídas de várias subdivisões da imagem e concatenadas em descritor.

Em CBIR, a extração de atributos é a etapa preliminar do processo de recuperação de imagens em um banco. Já a etapa de análise comparativa se dá pelo emprego de algoritmos e medidas que nos permitam mensurar o quão similares são as características obtidas para duas imagens distintas.

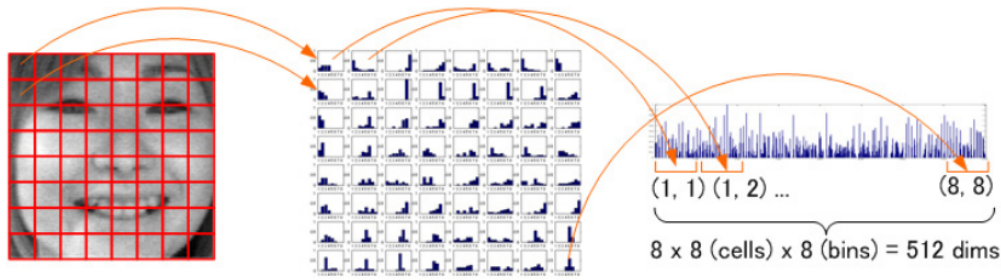


Figura 2.2: Processo de extração simples baseada em histograma local (LIH) (Shimada et al., 2010)

2.2 Índices de validação dos agrupamentos

A tarefa de identificar qual dos esquemas de agrupamento de imagens em conjuntos de similaridade gerados por um determinado algoritmo de clusterização tem sido um grande desafio. Índices de validação podem ser empregados para verificar o quão bons foram os agrupamentos feitos pelo algoritmo (Halkidi et al., 2002), permitindo a seleção do melhor esquema (Kovács et al., 2005). Um bom esquema de agrupamentos deve apresentar grupos compactos, com poucos elementos dispersos (relação *intracluster*), e com clara separação entre cada um dos *clusters* (relação *intercluster*), algo que pode influenciar diretamente sobre o retorno da consulta em um sistema CBIR.

Diferentes aspectos relacionados às medidas *intracluster* e *intercluster* são vistos na Figura 2.3. Em (a), o esquema de *clusters* pode ser representado como um grupo único ($k = 1$) e em (d) por um conjunto onde o número total de grupos é igual ao número de elementos do conjunto ($k = U$). No primeiro caso a medida *intercluster* é, logicamente, nula, algo que se inverte no segundo caso onde a medida *intracluster* não pode ser obtida (cada *cluster* possui apenas um elemento). Já em (b) e (c), diferenças nas relações *intercluster* e *intracluster* são observadas conforme o número de grupos varia.

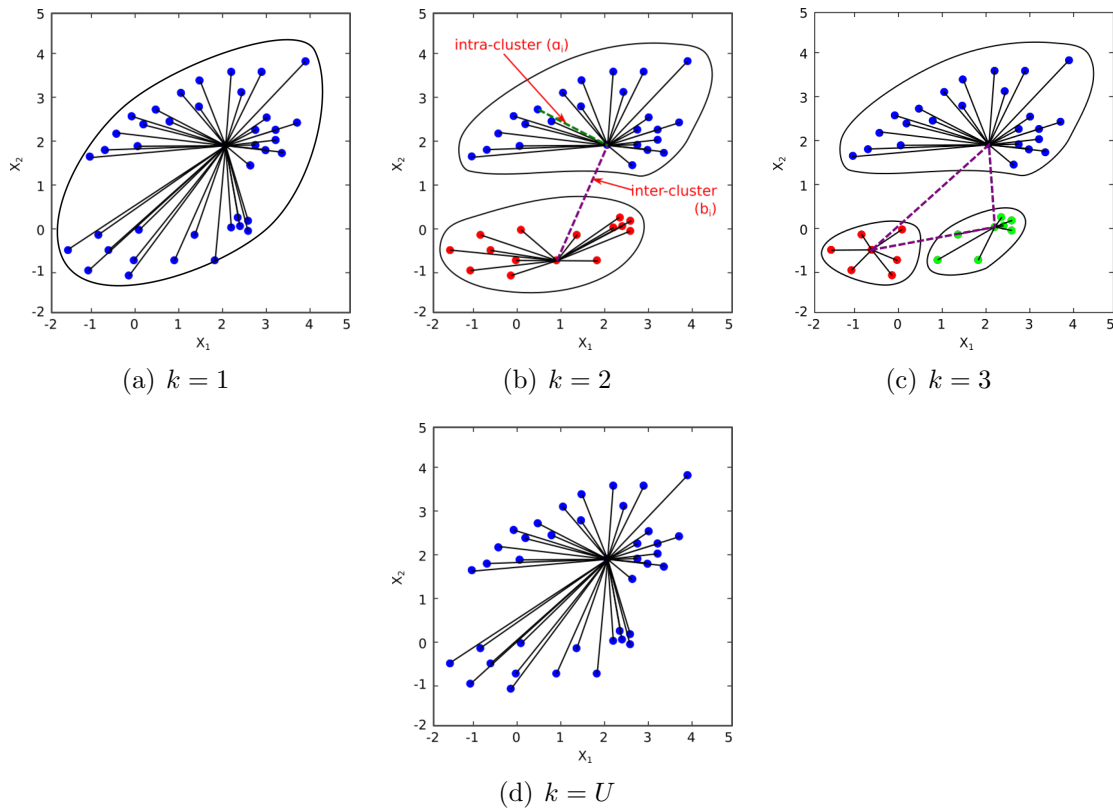


Figura 2.3: Aspectos das relações intracluster e intercluster para diferentes esquemas de agrupamento.

Essas relações são chamadas por alguns autores de *Medidas de Validação Relativas de Cluster*. As características intrínsecas na formação dos padrões de agrupamento se refletem nos índices de validação obtidos de acordo com cada método de cálculo elaborado e que servirá como parâmetro para a seleção do número de grupos ideal (melhor conjunto de padrões a ser alcançado). Tais relações podem ser calculadas de diferentes formas como pode ser visto nas Tabelas 2.1 e 2.2.

Medida	Relação
Single-linkage	Distância entre os dois elementos mais próximos pertencentes a dois <i>clusters</i> K_i e K_j
Complete-linkage	Distância entre os dois elementos mais distantes pertencentes a dois <i>clusters</i> K_i e K_j
Average-linkage	Média da distância entre todos os elementos pertencentes a dois <i>clusters</i> K_i e K_j
Centroid-linkage	Distância entre os centros dos <i>clusters</i> K_i e K_j

Tabela 2.1: Relação intercluster (entre elementos de *clusters* diferentes).

Medida	Relação
Complete-diameter	Distância entre os dois elementos mais distantes pertencentes ao mesmo <i>cluster</i> K_i
Average-diameter	Distância média entre todos os elementos pertencentes ao mesmo <i>cluster</i> K_i
Centroid-diameter	O dobro da distância média entre todos os elementos e o centro do <i>cluster</i> K_i

Tabela 2.2: *Relação intracluster* (entre elementos do mesmo *cluster*).

Existem inúmeros métodos de validação de agrupamentos explorando as relações *intercluster* e *intracluster* (Halkidi et al., 2002), contudo, estudos têm mostrado a necessidade da adoção de métodos que possam estimar a qualidade dos agrupamentos através de critérios que considerem a geometria observada nos dados e não apenas as medidas de homogeneidade e dispersão (Kovács et al., 2005).

2.2.1 Silhouette

O *Silhouette Score* (Rousseeuw, 1987) é um índice de validação amplamente utilizado na avaliação de agrupamentos que analisa a coesão (relação *intracluster*) e a separação (relação *intercluster*) de um esquema de agrupamentos encontrado por um algoritmo de clusterização. Ele fornece um valor entre $[-1, 1]$, onde um *score* igual a 1 representa um bom esquema de agrupamento e -1 indica um esquema ruim. O *Silhouette* é calculado individualmente para cada um dos elementos do conjunto de dados, atendendo aos seguintes critérios (Rousseeuw, 1987):

$$S_i = \begin{cases} 1 - a_i/b_i, & \text{se } a_i < b_i \\ 0, & \text{se } a_i = b_i \\ b_i/a_i - 1, & \text{se } a_i > b_i \end{cases} \quad (2.1)$$

Onde, a_i é a distância d média do i -ésimo elemento a todos os outros elementos do seu próprio *cluster* (relação *intracluster*), ou seja, $a_i = \text{average}_d(i, C_i)$, e b_i é a distância d mínima do i -ésimo elemento pertencente ao *cluster* C_i a um elemento do *cluster* C_j qualquer, ou seja, $b_i = \min_d(i, C_j)$ (relação *intercluster*). De forma genérica, podemos reescrever o sistema 2.1 como a seguinte equação:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (2.2)$$

A equação 2.2 sintetiza, portanto, a razão da distância *intercluster*, subtraída pela distância *intracluster* (a distância *intracluster* é, normalmente, menor que a distância *intercluster* para um bom esquema de agrupamentos) e a máxima dissimilaridade calculada entre essas duas relações (*inter* e *intra*) para o i -ésimo elemento do conjunto de dados. Como pode ser visto na Figura 2.4, a distância b_i pode ser obtida

de duas maneiras conforme o tipo de abordagem pretendida: a) partindo do i -ésimo elemento, pertencente a um *cluster* C_i qualquer, ao centro do *cluster* C_j ; b) partindo do i -ésimo elemento do *cluster* C_i ao elemento mais externo do *cluster* C_j . Esta última é a forma natural do método (Rousseeuw, 1987). Em ambos os casos, o *cluster* C_j será o mais próximo possível ao elemento i do *cluster* C_i . Por fim, um valor médio é calculado a partir de todos os *scores* individuais. Este valor médio representará o *score* global do esquema de K *clusters*. A Figura 2.5 mostra um exemplo de validação pelo *Silhouette* usando a implementação feita pela biblioteca *SciKit Learn* (Python).

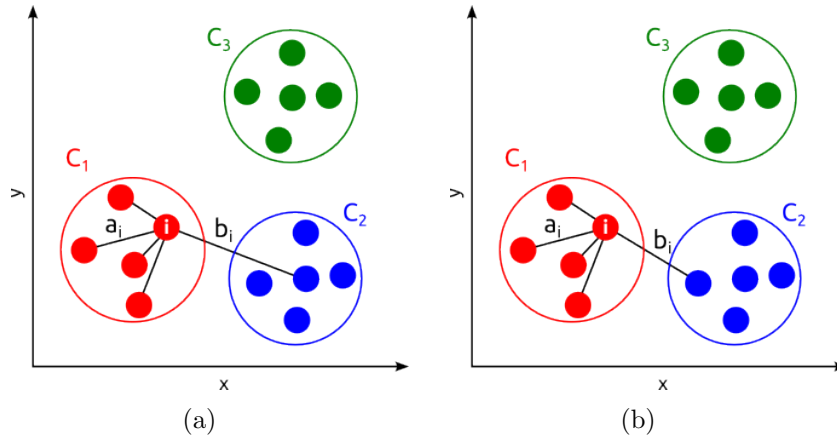


Figura 2.4: Silhouette: em (a), a distância *intercluster* b_i é calculada em relação ao centro do *cluster* C_j ; em (b), a distância *intercluster* b_i é tomada em relação ao elemento mais externo, ou seja, a sua forma natural.

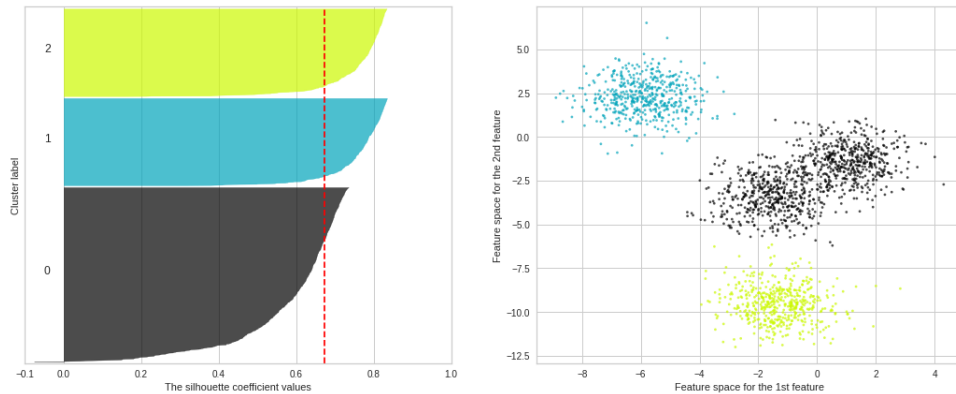


Figura 2.5: *Silhouette*: um valor próximo de 1 indica um bom esquema de agrupamento, ao passo que um valor -1 indica um esquema ruim. Gráficos obtidos através da biblioteca *SciKit Learn* (Buitinck et al., 2013) com dados fictícios. A linha pontilhada em vermelho é o *score* médio global.

Algumas particularidades precisam ser levadas em conta ao calcularmos o *Silhouette Score* (Garcia-Dias et al., 2020):

- o esquema de agrupamentos deve ter ao menos dois *clusters*, $k \geq 2$;
- requer o cálculo das matrizes de distância de todo o conjunto de dados, o que pode ser impraticável para grandes conjuntos de dados;
- quando *clusters* com um elemento forem formados, o *score* deverá adotar o valor $S_i = 0$ ($a_i = b_i$);
- o *score* pode apresentar distorções em conjuntos de dados cuja distribuição evidencie tipos de geometria esférica ou circular, resultando em *scores* altos para agrupamentos mal-formados, como apresentado na Figura 2.6.

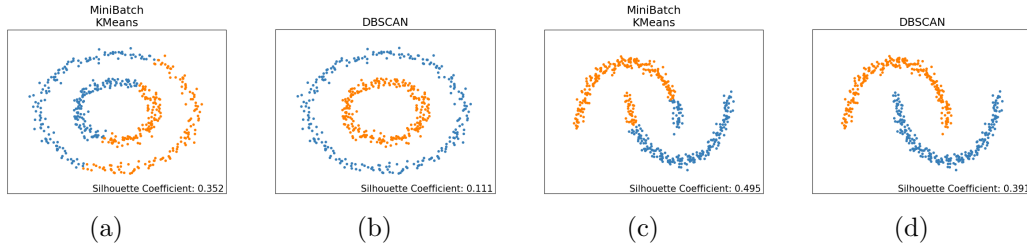


Figura 2.6: Nesta Figura podemos observar alguns erros do *Silhouette Score*: em (a) o *score* obtido indica que a configuração encontrada pelo K-means (0.352) foi melhor que a configuração encontrada pelo DBscan (0.111). O mesmo acontece para a geometria em semi-círculo em (c), *score* 0.495, e em (d), *score* 0.391. Agrupamentos e *scores* obtidos através da biblioteca *SciKit Learn* (Buitinck et al., 2013)

2.2.2 Dunn Index

O *Dunn Index* (Dunn, 1973) é definido como a razão entre a mínima distância *intracluster* (dois elementos de um mesmo *cluster*) e a máxima distância *intercluster* (entre *clusters* distintos). A medida *intercluster*, representada na Equação 2.3 por $\delta(X_p, X_q)$, onde X_p e X_q são os respectivos agrupamentos p e q , pode ser calculada usando qualquer função como a Euclidiana aplicando relações como *Single-linkage*, *Complete-linkage*, *Average-linkage* ou *Centroid-linkage* (Tabela 2.1). A distância *intracluster*, representada por $\Delta(X_k)$, pode ser calculada usando alguma relação *intracluster* como *Complete-diameter*, *Average-diameter* ou *Centroid-diameter* (Tabela 2.2). Onde, p e q representam *clusters* de um determinado esquema com um número k de *clusters* (Equações 2.3 e 2.4). Para este índice, um valor alto indica um bom esquema de agrupamento em termos de coesão e separabilidade.

$$Dunn = \min_{1 \leq p \leq k} \left\{ \min_{1 \leq q \leq k, (p \neq q)} \left\{ \frac{\delta(X_p, X_q)}{\max\{\Delta(X_k)\}} \right\} \right\} \quad (2.3)$$

Ou a forma reduzida,

$$Dunn = \frac{\min_{1 \leq p \leq q \leq k} \{\delta(X_p, X_q)\}}{\max\{\Delta(X_k)\}} \quad (2.4)$$

Sendo $p, q = 1 \dots k$

Algumas desvantagens do *Dunn Index* incluem: o aumento excessivo do custo computacional com o incremento da dimensionalidade, ou com o aumento do número de elementos no conjunto de dados, e certa sensibilidade à ruído nos dados (*outliers*). Nas Figuras 2.7(a) e 2.7(b) temos a representação das relações *intracluster* e *intercluster*, as quais devem ser tomadas ao máximo e mínimo, respectivamente.

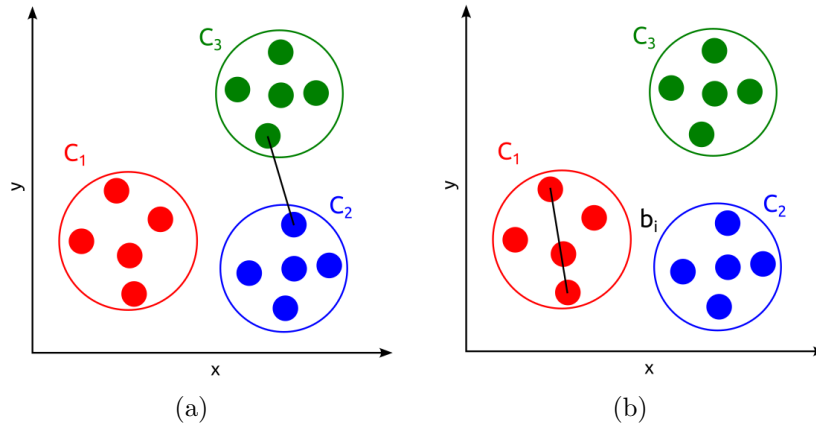


Figura 2.7: Dunn Index: a distância *intercluster* deve ser a maior possível, tomada entre os elementos mais externos dos agrupamentos 2.7(a) (numerador da Eq. 2.4); a distância *intracluster* deve ser a menor entre todos os elementos do mesmo agrupamento 2.7(b) (denominador da Equação 2.4).

2.2.3 Davies-Bouldin

O índice *Davies-Bouldin* (Davies e Bouldin, 1979) reflete a dispersão do esquema de *clusters* de forma inversa ao *Dunn Index*, ou seja, quanto menor o valor encontrado melhor o esquema de agrupamento em termos de coesão e separabilidade, relações *intracluster* e *intercluster*, respectivamente. O índice *Davies Bouldin* é dado pela Equação 2.5.

$$DBindex = \frac{1}{k} \sum_{p=1}^k R_p \quad (2.5)$$

Neste caso, R_p (Eq. 2.6) é chamado de função real valorada de p (Davies e Bouldin, 1979) e é dado pela razão entre a soma das dispersões dos *clusters* X_p e X_q

e a distância obtida pela relação *intercluster* de ambos, conforme pode ser visto na Equação 2.7, para um número k de *clusters* em um determinado esquema de agrupamentos.

$$R_p = \max R_{pq} \quad (2.6)$$

Sendo,

$$R_{pq} = \frac{\Delta(X_p) + \Delta(X_q)}{\delta(X_p, X_q)} \quad (2.7)$$

Onde $p, q = 1 \dots k$ e $p \neq q$

Portanto, $\Delta(X_p)$ e $\Delta(X_q)$ representam a dispersão (ou coesão) dos *clusters* p e q , respectivamente, enquanto que $\delta(X_p, X_q)$ é a distância entre esses *clusters* (separabilidade *intercluster*). A Figura 2.8 ilustra as relações *intracluster* e *intercluster* adotadas pelo *Davies Bouldin*.

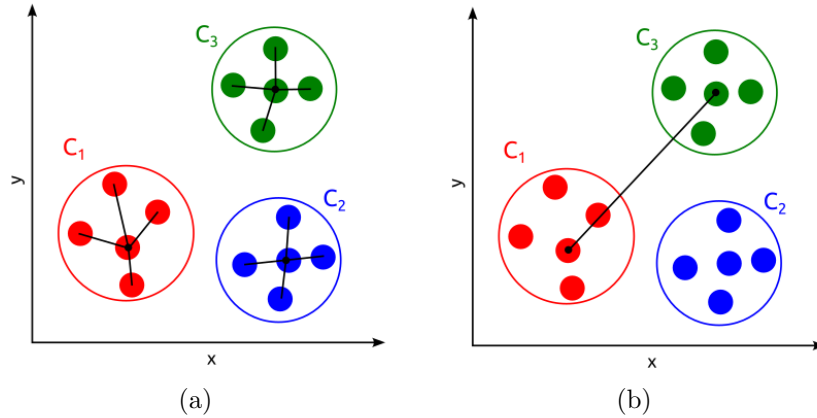


Figura 2.8: Davies Bouldin: a relação *intracluster* é tomada pela variância 2.8(a), enquanto que a distância *intercluster* deve ser calculada para cada centroides 2.8(b).

O *DB index* é, portanto, um valor médio da razão entre as relações *intracluster* e *intercluster* obtido para um esquema de *clusters* K qualquer, cuja principal limitação da proposta original do índice é o fato de não considerar a geometria espacial da distribuição dos *clusters*. Esta limitação está diretamente relacionada ao uso da média como medida de avaliação (Thomas et al., 2013). A Figura 2.9 mostra dispersões idênticas, porém, com orientação espacial diferente em que o *DB index* apresenta valores iguais.

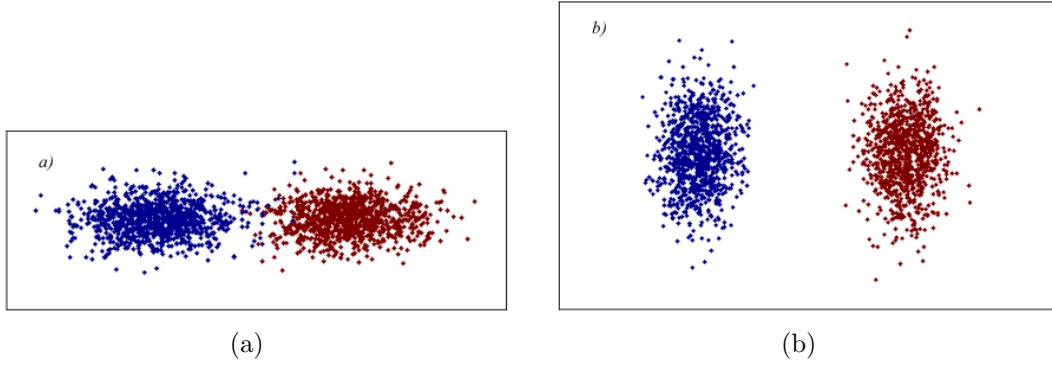


Figura 2.9: Davies Bouldin: o índice falha ao não considerar a orientação espacial da dispersão dos dados. A Figura mostra duas dispersões idênticas com orientações espaciais distintas, porém, o DB index apresenta o mesmo valor para ambos os casos (Thomas et al., 2013). A dispersão 2.9(b) parece ter características melhores, coesão e separabilidade, e deveria pontuar melhor que a 2.9(a).

2.2.4 Outros índices de validação

Existem diversos outros métodos para validação de agrupamentos (Rokach e Maimon, 2005) como a *Xie Beni Index* (Xie e Beni, 1991) que, da mesma forma que as anteriores, fornece um índice baseado na coesão e separação (separabilidade) dos *clusters* num esquema de agrupamentos. O *Xie Beni Index* é calculado através da equação 2.8:

$$XBindex = \frac{\sum_{i=1}^k \sum_{j=1}^n \mu_{ij}^2 \|V_i - X_j\|^2}{n * \min \|V_i - V_j\|^2} \quad (2.8)$$

Ou, simplificando:

$$XBindex = \frac{\sum_{i=1}^k \sum_{j=1}^n d(V_i - X_j)^2}{n * d_{min}(V_i - V_j)^2} \quad (2.9)$$

Onde, $d(V_i - X_j)$ é a distância do centro do *cluster* V_i ao elemento X_j pertencente à ele, $d_{min}(V_i - V_j)$ é a distância mínima entre os centros de cluster e n é o número de elementos do conjunto. Um valor baixo para o *Xie Beni Index* indica uma boa configuração de *clusters*. O XB Index possui três desvantagens conhecidas: o decréscimo do índice monotonicamente quando o número de *clusters* se torna muito grande, próximo ao número de elementos do conjunto de dados (n); o comportamento numericamente instável quando o expoente de ponderação *fuzzy* (m), μ_{ij}^m , quando $m > 2$, se aproxima de infinito (Tang et al., 2005); tendência de subagrupamento em que o número estimado de *clusters* é menor do que o número real (Kwon et al., 2021).

Outro método bastante utilizado como índice de validação de agrupamento é o *Sum of Squared Error - SSE* (Soma do Erro Quadrático), também conhecido por Soma Residual dos Erros Quadráticos, que é uma métrica a qual concentra as diferenças entre os valores previstos e aqueles encontrados pelo algoritmo:

$$SSE = \sum_{i=1}^n d(V_i - X_j)^2 \quad (2.10)$$

Onde, $d(V_i - X_j)^2$ é a distância Euclidiana entre o elemento médio do agrupamento (V_i), o centro do *cluster* k_i , e o elemento X_j pertencente a este *cluster*.

Como podemos perceber, a *SSE* não leva em conta critérios intrínsecos do esquema de agrupamentos como a separabilidade e compacidade. Um valor alto para este método indica um esquema de *clusters* ruim. Neste trabalho, faremos uma análise comparativa entre alguns dos métodos já bem estabelecidas pela comunidade científica e o recente *Decision-theoretic Rough Sets* (detalhado mais adiante, Seção 2.4.1), em sua abordagem estendida (Yu et al., 2014), o qual aplica conceitos probabilísticos da inferência bayesiana para a tomada de decisões (Hoff, 2009) à teoria dos conjuntos aproximados *Rough Sets* (Pawlak, 1982), como *cluster validity*, em um algoritmo hierárquico aglomerativo, no contexto da *Content-based Image Retrieval* (CBIR), sem interferência humana na seleção do *best K*.

2.2.5 Método do Cotovelo

O método do Cotovelo, *Elbow's method* (Thorndike, 1953), é uma heurística auxiliar para a escolha do esquema de *clusters* mais adequado em um algoritmo de aprendizado não-supervisionado. É um método de análise visual da curva gerada pelos valores de um índice de validação de agrupamento qualquer, onde o ponto de corte evidencia o início de uma possível estabilidade no comportamento do algoritmo, isto é, quando a curva de resultados começa a nivelar, indicando que o algoritmo de agrupamento alcançou o seu ponto de otimização. Uma representação do método pode ser vista na Figura 2.10.

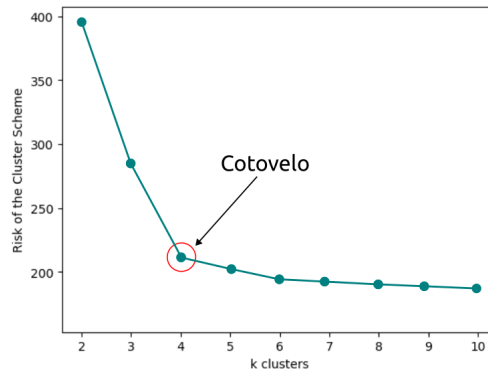


Figura 2.10: Método do Cotovelo (*Elbow's method*) aplicado à curva de risco gerada por um *cluster validity*.

Esse ponto de corte é denotado pelo ângulo acentuado em determinado local da curva, lembrando um cotovelo. Apesar de ser uma forma de análise intuitiva, visual, existem na literatura algumas formas de expressá-lo matematicamente (Sugar e James, 2003) (Onumanyi et al., 2022).

Contudo, o método do Cotovelo possui problemas e limitações (Schubert, 2023): não garante a escolha do número de grupos ideal; depende fortemente dos resultados do índice de validação de agrupamento; é subjetivo, quando em sua forma visual; não é aplicável em quantidades muito grandes de agrupamentos; algumas curvas desenhadas pelos índices de validação são bastante irregulares, o que dificulta a definição de um único ponto de corte.

Neste trabalho, a expressão matemática para o método do Cotovelo foi adotada como o ponto que apresenta a maior distância perpendicular à reta secante da curva de resultados do índice de validação de agrupamento analisado (Satopaa et al., 2011). A Figura 2.11 ilustra um exemplo de identificação do melhor esquema de agrupamentos ($k = 4$) através do método do Cotovelo.

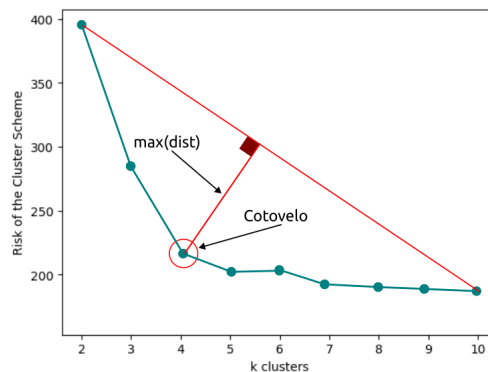


Figura 2.11: Forma matemática do método do Cotovelo (*Elbow's method*): maior distância perpendicular entre um ponto $p(k,C)$ e a reta secante à curva do *cluster validity*.

2.3 Métodos de avaliação da eficácia

Além da avaliação do quão bom foi o esquema de agrupamentos encontrado pelo algoritmo de clusterização, precisamos avaliar o quão assertiva foi a separação dos objetos de cada esquema (Zhai, 2008) (Gorunescu, 2011). Para isto, métricas de avaliação como o *Precision*, o *Cluster Recall* e o *F-Measure* comparam as relações entre acertos e erros, ou seja, a performance da ferramenta de recuperação.

2.3.1 Precision

A medida *Precision* quantifica o número de elementos classificados positivamente (*True Positive*) na classe pretendida em relação ao número de elementos totais que foram classificados positivamente mesmo que na classe não pretendida (*False Positive*), sendo matematicamente definida pela Equação 2.11.

$$Precision = \frac{TP}{TP + FP} \quad (2.11)$$

No contexto da CBIR, *Precision* é apresentada como o percentual de elementos relevantes dentre os retornados pelo sistema até uma determinada posição em um ranking de imagens. Um valor baixo na métrica *Precision* pode indicar que o algoritmo classificou um número de *FP* elevado, ou seja, marcou positivamente elementos que não existem na classe pretendida. Já um índice mais próximo de 1 indica maior assertividade.

2.3.2 Cluster Recall

Cluster Recall, também conhecida como *Subtopic Recall at Rank n* (Zhai et al., 2015), é uma adaptação baseada na métrica *Recall* (revocação) concentrada na cobertura parcial ou completa de subtópicos, ou seja, na capacidade de encontrar pelo menos um elemento relevante para o maior número possível de subtópicos ou vários elementos relevantes para qualquer subtópico (Carpineto et al., 2012).

Considere uma consulta Q com h subtópicos $S_1, \dots, S_h$ e uma classificação $d_1, \dots, d_m$ de m elementos. Seja $subtopics(d_i)$ o conjunto de subtópicos para os quais d_i é relevante. A recuperação de subtópicos na classificação n , denotada por $S-Rec@n$, é definida como a fração de subtópicos cobertos pelos primeiros n elementos, a saber:

$$S-Rec@n = \frac{\cup_{i=1}^n subtopics(d_i)}{h} \quad (2.12)$$

Para efeito de simplificação utilizaremos $CR@n$ ao invés de $S-Rec@n$ neste trabalho. Um exemplo ilustrativo para o cálculo do *Cluster Recall* pode ser visto na Figura 2.12.

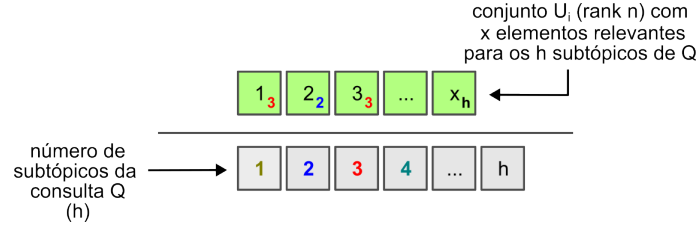


Figura 2.12: Figura ilustrativa exemplificando o cálculo do *Cluster Recall* para h subtópicos da consulta Q , onde foram retornados x elementos (imagens) relevantes dentre n selecionados. Fonte: autor.

2.3.3 F-Measure

Já a medida *F-Measure* (também conhecida como *F-Score* ou *F_β-Score*) é uma relação entre as métricas *Precision* e *Recall*. Eq. 2.13.

$$F_{\beta} = \frac{(1 + \beta^2)}{\frac{1}{Recall} + \frac{1}{Precision}} = (1 + \beta^2) * \frac{Precision * Recall}{(\beta^2 * Precision) + Recall} \quad (2.13)$$

Sendo β um valor positivo e real (fator de peso) dimensionador do equilíbrio entre *Precision* e *Cluster Recall*. Esta métrica pode ser considerada a mais adequada por concentrar um equilíbrio entre as duas medidas de avaliação de acurácia, sendo considerada média harmônica quando $\beta = 1$ (*F₁-Score*).

2.4 Decision-theoretic rough set models

2.4.1 Definição

Decision-theoretic rough set models são uma extensão probabilística do modelo algébrico de (Pawlak, 1982) onde os parâmetros para definição das aproximações *lower* e *upper* são baseados no cálculo do risco (custo) de uma ação α qualquer através do procedimento de inferência bayesiano (Yao, 2007).

Alguns trabalhos têm utilizado o DTRS como índice de validação para ajudar a resolver o problema da definição do melhor esquema de agrupamentos visando a obtenção do número K de *clusters* ideal (Yu et al., 2014) (Hoff, 2009) pelo cálculo do risco total do esquema de agrupamento e pela aplicação de métodos capazes de analisar o comportamento da curva gerada pelo par $(K_i, Risk_i)$ a exemplo do Método Cotovelo.

2.4.2 Aproximação algébrica

Seja um conjunto finito $U \neq \emptyset$ e R uma relação de equivalência em U , o par $apr = (U, R)$ é chamado de *Espaço de Aproximação* (Figura 2.13). *Espaço de Aproximação*

ou *Espaço Aproximado*, é a representação da região na qual a relação de equivalência R sobre U é dita indiscernível, ou relação de indiscernibilidade. Esta relação de equivalência R induz uma partição de U denotada por U/E . A classe de equivalência contendo um elemento x é dada por $[x] = \{y | xRy\}$. Essas classes de equivalência de R são os blocos básicos para construir aproximações algébricas dos *Rough Sets*. Então, para um subconjunto $A \subseteq U$, podemos ter as seguintes zonas de aproximação:

$$\underline{apr}(A) = \{x \in U \mid [x] \subseteq A\} \quad (2.14)$$

$$\overline{apr}(A) = \{x \in U \mid [x] \cap A \neq \emptyset\} \quad (2.15)$$

Com base nas aproximações *lower* (Equação 2.14) e *upper* (Equação 2.15) do conjunto A , podemos dividir o conjunto universo U em três regiões distintas, a região positiva $POS(A)$, a região negativa $NEG(A)$ e a região de fronteira ou limite $BND(A)$ conforme a Figura 2.13 (Yao, 2007).

$$POS(A) = \underline{apr}(A) \quad (2.16)$$

$$BND(A) = \overline{apr}(A) - \underline{apr}(A) \quad (2.17)$$

$$NEG(A) = U - POS(A) \cup BND(A) = U - \overline{apr}(A) \quad (2.18)$$

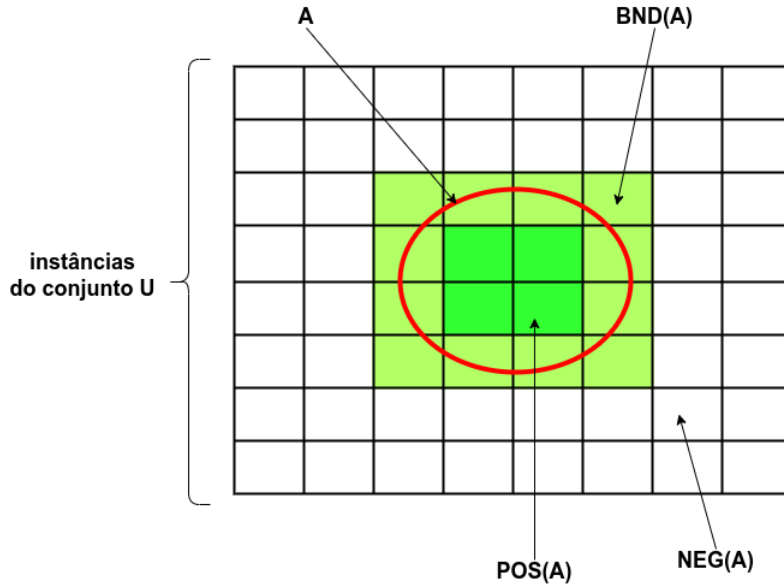


Figura 2.13: Espaço de aproximação: divisão das regiões no modelo DTRS.

2.4.3 Aproximação probabilística

Podemos definir um conjunto de estados $X, \neg X$ que indicam se um objeto está presente ou não está em X . O conjunto $A = a_P, a_N, a_B$ denota as possíveis ações

de classificação de um objeto: $POS(X)$, $NEG(X)$ e $BND(X)$. Onde $POS(X)$ é a região de classificação positiva de X , $NEG(X)$ é a região de classificação negativa e $BND(X)$ é a região de fronteira (*boundary*) entre $POS(X)$ e $NEG(X)$. Tomemos $i = P, N, B$, $\lambda_{iP}(a_i|X)$ e $\lambda_{iN}(a_i|\neg X)$ representando o custo para a uma ação a_i quando o estado é $X, \neg X$, respectivamente. Suponha que a ação seja tomada para um objeto com a descrição $r(x)$, como uma classe equivalente de x , nomeadamente $[x]$.

A perda esperada $R(a_i|r(x))$ associada com a tomada individual de ações pode ser expressa por:

$$R(a_P|r(x)) = \lambda_{PP}P(X|r(x)) + \lambda_{PN}P(\neg X|r(x)) \quad (2.19)$$

$$R(a_N|r(x)) = \lambda_{NP}P(X|r(x)) + \lambda_{NN}P(\neg X|r(x)) \quad (2.20)$$

$$R(a_B|r(x)) = \lambda_{BP}P(X|r(x)) + \lambda_{BN}P(\neg X|r(x)) \quad (2.21)$$

Onde $P(X|r(x))$ e $P(\neg X|r(x))$ são as probabilidades de um objeto com a descrição $r(x)$ pertencer a X e $\neg X$, respectivamente. Sendo, abaixo, o procedimento bayesiano para tomada de decisão baseada em risco mínimo:

- (P) Se $R(a_P|r(x)) \leq R(a_N|r(x))$ e $R(a_P|r(x)) \leq R(a_B|r(x))$, decida $x \in POS(X)$;
- (N) Se $R(a_N|r(x)) \leq R(a_P|r(x))$ e $R(a_N|r(x)) \leq R(a_B|r(x))$, decida $x \in NEG(X)$;
- (B) Se $R(a_B|r(x)) \leq R(a_P|r(x))$ e $R(a_B|r(x)) \leq R(a_N|r(x))$, decida $x \in BND(X)$;

Então, tomando $P(X|r(x)) + P(\neg X|r(x)) = 1$ é possível, então, simplificar essas regras e classificar qualquer objeto em $r(x)$ através das probabilidades $P(X|r(x))$ e das funções de perda (loss function) $\lambda_i (i = P, N, B; j = P, N)$

$$Risk(a_P|(x_i, x_j)) = P(\neg C|(x_i, x_j)) \quad (2.22)$$

$$Risk(a_N|(x_i, x_j)) = P(C|(x_i, x_j)) \quad (2.23)$$

$$Risk(CS_t|(x_i, x_j)) = \begin{cases} P(\neg C|(x_i, x_j)) & (x_i, x_j) \in C, \\ P(C|(x_i, x_j)) & (x_i, x_j) \in \neg C \end{cases} \quad (2.24)$$

Assim, $P(C|(x_i, x_j))$ é a probabilidade de x_i e x_j estarem no mesmo cluster. Generalizando, os autores assumem que um valor alto para $sim(x_i, x_j)$ significa alta probabilidade de que os elementos x_i e x_j estejam no mesmo cluster. Um $sim(x_i, x_j) = 0$ significa que os elementos definitivamente não estão no mesmo cluster, ou seja,

$P(C|(x_i, x_j)) = 0$. As funções de risco $Risk(a_P|(x_i, x_j))$ e $Risk(a_N|(x_i, x_j))$ são tomadas para valores dentro do intervalo de $[0,1]$. Essas funções são tratadas como representações das *loss functions* para os casos POS(x) e NEG(x). É, então, definido um $val \leq sim(x_i, x_j)$, considerando a matriz $sim(x_i, x_j)$, de tal modo que $P(C|(x_i, x_j)) \geq 0.5$ indicando que (x_i, x_j) estão, estatisticamente, muito mais inclinados a estarem no mesmo cluster, para extrair a média de similaridade (Equação 2.25).

$$val = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n sim(x_i, x_j) \quad (2.25)$$

E a Equação 2.27 como função de avaliação do risco total de um *cluster* CS_t , cujo resultado é dependente do sistema de equações 2.26 em que a avaliação probabilística do pertencimento do par (x_i, x_j) ao *cluster* C , e a relação entre a similaridade de (x_i, x_j) e val , geram o valor de $Risk(CS_t|(x_i, x_j))$.

$$Risk(CS_t|(x_i, x_j)) = \begin{cases} 0.5 - \frac{sim(x_i, x_j) - val}{2 - 2val}, sim(x_i, x_j) \geq val, (x_i, x_j) \in C \\ 0.5 + \frac{val - sim(x_i, x_j)}{2val}, sim(x_i, x_j) < val, (x_i, x_j) \in C \\ 0.5 + \frac{sim(x_i, x_j) - val}{2 - 2val}, sim(x_i, x_j) \geq val, (x_i, x_j) \in \neg C \\ 0.5 - \frac{val - sim(x_i, x_j)}{2val}, sim(x_i, x_j) < val, (x_i, x_j) \in \neg C \end{cases} \quad (2.26)$$

$$Risk(CS_t) = \sum_{i=1}^n \sum_{j=1}^n Risk(CS_t|(x_i, x_j)) \quad (2.27)$$

2.5 Análise estatística dos resultados

A estatística descritiva envolve técnicas que nos permitem resumir, descrever e interpretar um conjunto de dados. Ela fornece medidas que auxiliam na identificação de padrões e tendências. As relações de média aritmética, moda e mediana são conhecidas como medidas de tendência central. Gráficos como o *Boxplot*, o *Histograma* e o gráfico *Quantil-Quantil* facilitam a visualização em termos de dispersão, tendência dos dados e a identificação de *outliers*. Contudo, a escolha da metodologia e das ferramentas mais adequadas depende do tipo de dado e do objetivo da análise. Normalmente, quando se trata da análise comparativa entre os resultados de algoritmos de aprendizado de máquina, são utilizados, ainda, testes de hipótese (Corani et al., 2017). Os testes de hipótese podem determinar se os dados possuem

uma distribuição potencialmente normal (Gaussiana) ou não-normal. Um conjunto de resultados muito próximos da distribuição normal pode indicar que tais valores poderiam estar mais propensos à aleatoriedade do que um efetivo ganho do sistema (Ferreira e Patino, 2015). Essa observação implica na escolha do teste correto para a verificação da significância estatística entre os dados de dois grupos de resultados.

2.5.1 Teste de Shapiro-Wilk

O teste de *Shapiro-Wilk* (Shapiro e Wilk, 1965) é um teste de normalidade estatística que avalia se a distribuição dos dados é potencialmente normal através da verificação da hipótese nula (H_o), cuja afirmação diz que a diferença entre a amostra populacional e a distribuição normal não é grande o suficiente para ser estatisticamente significativa, ou seja, o teste verifica se a distribuição está próxima do que chamamos de curva de *Gauss*. A curva de *Gauss* é uma representação gráfica e teórica baseada nos parâmetros da média e do desvio padrão, sendo, por esse motivo, chamada de estatística paramétrica (Campos, 2002), ela representa as distribuições probabilísticas naturais e é simétrica em relação ao eixo das ordenadas. O teste é realizado encontrando-se o valor do W estatístico (Eq. 2.28) e comparando-o ao W_α tabelado de acordo com o nível de confiança escolhido.

$$W = \frac{b^2}{s^2} = \frac{\left(\sum_{i=1}^n a_i y_i\right)^2}{\sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (2.28)$$

O teste de *Shapiro-Wilk* possui como vantagem: a facilidade de uso, sua implementação é relativamente simples; o grande poder de detecção da normalidade se comparado a outros testes estatísticos; e sua boa aplicabilidade a pequenos espaços amostrais. Contudo, a eficácia de testes de normalidade sofre a influência da quantidade de elementos do espaço amostral. O *Shapiro-Wilk* é um teste recomendado para uma amostra com até 50 elementos, mas pode ser utilizado com espaços amostrais maiores. Porém, neste último caso, o teste se torna mais propenso a rejeitar a hipótese nula (H_o), com a consequente aceitação da hipótese alternativa (H_a).

Para uma quantidade de amostras acima de 500 unidades é aconselhável o uso do nível de significância menor que 1% ($\alpha < 0,01$, nível de confiança $> 99\%$) para prevenir o erro tipo II, falso-negativo, (Miot, 2017), isto é, quando hipótese nula é aceita mas deveria ser rejeitada (Banerjee et al., 2009).

2.5.2 Valor-p (*p-value*)

O *p-value* pode ser definido como o valor-probabilidade da média amostral de um determinado conjunto de dados. De acordo com Ferreira e Patino (2015): “O valor-*p* é definido como a probabilidade de se observar um valor da estatística de teste maior

ou igual ao encontrado. Tradicionalmente, o valor de corte para rejeitar a hipótese nula é de 0.05, o que significa que, quando não há nenhuma diferença, um valor tão extremo para a estatística de teste é esperado em menos de 5% das vezes.”

Na prática, o *p-value* é um teste de hipótese auxiliar baseado nas observações do que chamamos de *Região Crítica* (R_c). Ele indica a probabilidade de se observar uma diferença tão grande ou maior do que a observada sob a hipótese nula (Ferreira e Patino, 2015), ou seja, é a probabilidade de a média amostral ($\bar{x}$) estar situada dentro da área extrema do gráfico da distribuição normal teórica (com média $\bar{S}$) dada por $p\text{-value} = P(\bar{S} < \bar{x} \mid H_o)$. Neste caso, se o *p-value* encontrado for menor que o valor de α , rejeitamos a hipótese nula e consideramos que a média amostral encontra-se na região de rejeição. Caso contrário, aceitamos a hipótese nula e consideramos que a distribuição é potencialmente normal, uma vez que, sua média encontra-se mais próxima da região de aceitação. A Figura 2.14 mostra uma distribuição normal teórica para o cálculo unicaudal com nível de significância de 5% ($\alpha = 0,05$) na zona de rejeição da hipótese nula.

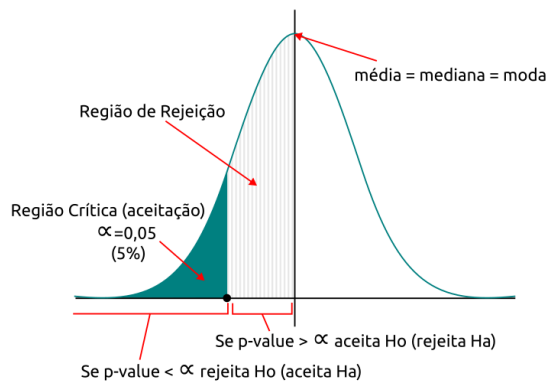


Figura 2.14: Gráfico teórico de uma distribuição normal (Ross, 2017). Cálculo unilateral ou unicaudal. Exemplo com nível de significância (α) de 5% (nível de confiança de 95%)

2.5.3 Teste de Wilcoxon

O teste de *Wilcoxon*, neste caso o *Wilcoxon Signed-Ranks Test* (também conhecido como *Wilcoxon Matched-Pairs*), é um método não-paramétrico (não é baseado nos parâmetros de normalidade) para comparação entre duas amostras pareadas (Wolson, 2007), isto é, entre distribuições amostrais que possuem alguma relação entre si. Ele é um teste de significância estatística no qual dois conjuntos amostrais são comparados através de um método que verifica as diferenças entre elementos desses conjuntos, identificando a preponderância de valores significativamente altos, indicando, ou não, a aleatoriedade postulada pela hipótese nula (H_o : a mediana das diferenças é igual a zero). Para a execução do teste de *Wilcoxon* não é necessário que as distribuições amostrais apresentem normalidade. A conclusão em relação à

hipótese nula se dá pela análise da estatística de teste, ou seja, a hipótese nula é rejeitada quando o valor estatístico obtido é maior que o valor de criticidade α , como visto na Figura 2.14, assim como feito para o *p-value*.

2.6 Trabalhos relacionados

Nesta seção faremos uma breve descrição de alguns dos trabalhos relacionados à pesquisa-foco deste estudo.

2.6.1 Recuperação de imagem baseada em conteúdo por clusterização

A escolha do número de grupos ideal é um dos maiores problemas em sistemas de recuperação de imagens baseados em conteúdo para abordagens que utilizam algoritmos de agrupamento. A correta parametrização e avaliação dos agrupamentos nesses algoritmos influencia diretamente na performance dos resultados em termos de representatividade e diversificação da consulta. Muitos trabalhos têm adotado um número de *clusters* fixo como em (Chen et al., 2003) e em torno de $k = 20$ *clusters*. Tal fato advém de estudos indicativos os quais afirmam que usuários normalmente se concentram nos primeiros 20 elementos da lista retornada pela consulta (Zhao e Glotin, 2009).

Chen et al. (2003) afirma que utilizar a matriz de confusão para medir a performance de um algoritmo de agrupamento é viável, mas que para fazer tal cálculo o número de *clusters* precisa ser igual ao número de semânticas distintas, algo que, na prática, é desconhecido. Assim, o CLUE foi testado, inicialmente, para um subconjunto de 10 categorias do *dataset* escolhido com $k = 10$ ($CS_t = 10$ de acordo com o artigo) para todas as consultas resolvidas por um esquema que tem o objetivo de tratar o problema do gap semântico com base na hipótese: *imagens semanticamente semelhantes tendem a ser agrupadas em algum espaço de características*. A performance do que os autores chamaram de CLUE (*CLUster-based rEtrieval of images by unsupervised learning*), uma abordagem não-supervisionada para realizar o agrupamento de imagens semanticamente semelhantes, foi avaliada em um *dataset* com 60.000 imagens (COREL database) para um sistema CBIR apelidado de CBICR (*Content-based Image CLUSTER Retrieval*).

O fluxo de trabalho do CBICR pode ser visto na Figura 2.15, onde as etapas iniciais são as mesmas de um esquema básico em CBIR com exceção da etapa de seleção de imagens que é feita pelo CLUE, que consiste num conjunto baseado na seleção de imagens vizinhas e num método não-supervisionado para agrupamento de dados. O CBICR pode adotar ainda uma etapa de *relevance feedback*. Os autores utilizaram o método de cortes normalizados (Ncut) (Shi e Malik, 2000) para segregar imagens individuais em grupos de características e assim realizar a busca por similaridade na

As consultas submetidas que obtiveram os melhores níveis de eficácia foram aquelas que apresentam imagens cujos grupos gerados pelo método Ncut puderam se mostrar mais bem definidos, ou seja, imagens cujas características intrínsecas apresentaram grupos com baixa dispersão, alta coesão e geometria bem definida. No entanto, o CLUE possui algumas limitações como: a tendência de bipartição de um grupo dito *puro* (bem definido) devido o método Ncut recursivo; encontrar uma imagem representativa com base em grupos de características pode retornar resultados imprecisos do ponto de vista semântico; a velocidade do sistema de recuperação é fortemente influenciada pela quantidade de imagens-alvo, um número grande de imagens pode tornar a aplicação demasiadamente lenta.

Além disso, a adoção de um K fixo ($k = 10$ no experimento inicial) pode não ser a escolha mais adequada para consultas diversificadas. Isso porque o número de grupos semânticos, algo que está diretamente ligado à quantidade de padrões encontrados pelo algoritmo de agrupamento, pode apresentar variabilidade de acordo com o conjunto de imagens escolhido para consulta. Sendo assim, não é difícil imaginar que melhores resultados podem encontrar-se dentro de uma faixa variada de K agrupamentos e não apenas em um esquema de *clusters* com tamanho fixo como foi adotado. De fato, os autores afirmam que o número de grupos fixado em $k = 10$, experimentalmente, pode ser bem diferente daquele encontrado em aplicações reais. Neste caso, medidas como *pureza* e *entropia* foram utilizadas para estimar a qualidade e a quantidade de grupos semânticos presentes no conjunto de imagens.

2.6.2 Diversificação visual dos resultados da busca de imagens

Neste artigo, Van Leuken et al. (2009) investigaram três métodos para diversificação visual de resultados em CBIR. Os métodos implementam técnicas de agrupamento em combinação com uma função de ponderação dinâmica das *features* para melhor capturar os aspectos discriminativos do conjunto resultante de imagens que é recuperado. Na abordagem, uma imagem representativa é selecionada de cada *cluster* e juntas formam um conjunto de resultados diversificado. Foram desenvolvidos três algoritmos no experimento (*Folding*, *Maxmin* e *Reciprocal election*).

Segundo os autores, os algoritmos *Folding* e o *Maxmin* são abordagens sem parâmetros, onde o número de *clusters* (k) é variável e definido dinamicamente através do ranking gerado em cada etapa da abordagem, sendo, para isto, calculado um valor limiar de ajuste baseado nos próprios resultados da consulta, com os pesos para as características intrínsecas das imagens ajustados automaticamente em cada um dos conjuntos resultantes. Em resumo, a quantidade de imagens selecionadas para o ranking numa etapa de execução do algoritmo definirá o número de agrupamentos a serem testados na próxima etapa. Desta forma, as imagens no topo do ranking, segundo os autores, teriam alta probabilidade de serem selecionadas como as mais representativas do *cluster*. Já o algoritmo *Reciprocal election* recebe um único parâmetro m definido arbitrariamente através de testes (melhores valores de m entre 3

e 8).

O primeiro algoritmo, *Folding*, seleciona as imagens levando em conta a classificação delas em um ranking obtido pelo modelo textual de recuperação. O resultado da busca I é armazenado num ranking $L = \{L_1, L_2..L_i\}$ em ordem decrescente do grau de relevância. O passo inicial da abordagem é selecionar as imagens representativas enquanto percorre o ranking L de cima para baixo. A primeira imagem selecionada L_i é sempre a mais representativa e é armazenada no conjunto R . O algoritmo então percorre o ranking L descendo e compara cada imagem ao conjunto de representativas (R) selecionadas no passo anterior. Quando uma imagem é suficientemente diferente de todas as já selecionadas, ela é adicionada ao conjunto de representativas (R). Após finalizar a passagem pelo ranking, os agrupamentos são formados em torno de cada representante usando a regra do vizinho mais próximo (*nearest neighbor*).

A chave desta abordagem é a definição do “suficientemente diferente” ao selecionar as imagens candidatas. O parâmetro é configurado de forma automática ou dinâmica. Ele é definido como a distância média de todas as imagens em I até a imagem média. A imagem média é uma imagem sintética que só existe no espaço de *features* e é construída agregando características de todas as imagens em uma imagem canônica. Já o algoritmo *Maxmin* tem um método um pouco diferente, ele não leva em consideração o ranking original como o *Folding*. O *Maxmin* tenta selecionar, por características visuais, os representantes mais diversos possíveis através da sua heurística maxmin, a qual se baseia na seleção de representantes em um conjunto não ordenado de imagens pela maior distância mínima entre pares do conjunto. O processo se inicia com a escolha aleatória do primeiro representante ($R1$) e o segundo ($R2$) cuja distância para $R1$ seja a maior. O processo para os demais representantes é contínuo e será interrompido quando a distância mínima máxima for menor que a média das distâncias entre todas as imagens do conjunto a uma imagem média anteriormente selecionada.

O terceiro algoritmo, *Reciprocal election*, promove uma espécie de auto seleção de diversidade baseada em uma eleição na qual cada imagem atribui pontos às suas vizinhas de acordo com a representatividade. Um esquema geral pode ser visto na Figura 2.17.

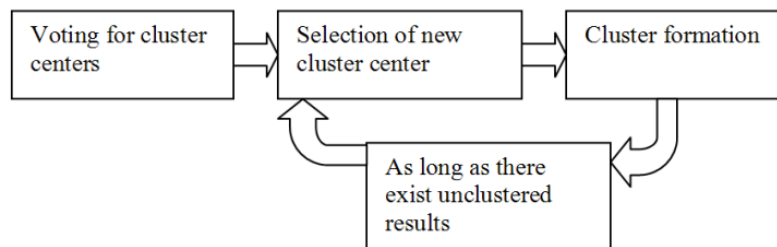


Figura 2.17: Esquema geral do algoritmo *Reciprocal election* (Van Leuken et al., 2009)

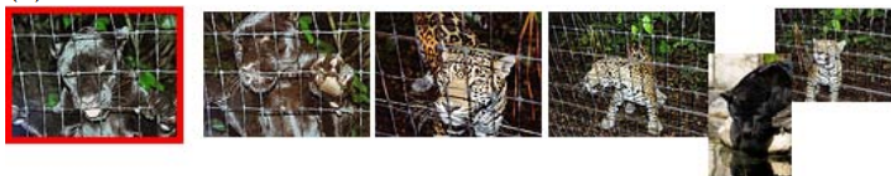
No experimento, os autores utilizaram um conjunto de 75 tópicos que foram selecionados aleatoriamente nos logs de pesquisa do *Flickr*. Para cada consulta foram recuperados os 50 principais resultados de uma fatia de 8,5 milhões de fotos no *Flickr*. Na Figura 2.18 é apresentado o resultado da busca por *jaguar* feita utilizando o algoritmo *Reciprocal Election*. Os elementos mais representativos de cada *cluster* estão marcados pela borda vermelha. Os demais elementos são variações de cada subtópico.



(a) Mamífero com pelagem de tigre



(b) Exposição de novo modelo de veículo



(c) Mamífero com pelagem de tigre enjaulado



(d) Veículo antigo na rua

(e) Modelo na expo

(f) Pantera negra

Figura 2.18: Resultado do agrupamento gerado pelo algoritmo *Reciprocal Election* para a busca por *jaguar*.

Conclusão

Em relação à diversificação, os autores em (Van Leuken et al., 2009) avaliaram que os métodos ficaram bem próximos do que seria a real percepção humana. Tal conclusão baseou-se, segundo os autores, no resultado de um extenso experimento de agrupamento realizado por avaliadores humanos, comparativamente. Já a avaliação

da qualidade dos esquemas de *clusters* é realizada através da heurística do próprio algoritmo desenvolvido pelos autores. A quantidade de grupos e o número de imagens de cada um deles são definidos de acordo com as etapas de ranqueamento de cada algoritmo, ou seja, o número K de *clusters* é definido pelas *top-n* imagens do ranking. A performance final foi avaliada pelo *Fowlkes–Mallows index* (FM) (Fowlkes e Mallows, 1983), que é um método de avaliação (ou validação) do quão similares são dois agrupamentos gerados por algoritmos diferentes, e pelo critério *Variation of Information* (VI) (Meilă, 2003). Um índice FM alto indica grande similaridade entre dois *clusters* enquanto que no critério VI isso ocorre de forma contrária (VI baixo indica grande similaridade) Tabela 2.3.

	Inter-assessor variability	Random	Election Reciprocal	Folding	Maxmin
FM	0.419	0.139	0.250	0.282	0.214
VI	1.463	2.513	1.975	2.081	2.129

Tabela 2.3: Desempenho médio em todos os tópicos, algoritmos e avaliadores humanos (Van Leuken et al., 2009)

2.6.3 Descoberta não supervisionada de subtópico implícito adaptável a consulta para recuperação de imagens diversificada com base na qualidade intrínseca do *cluster*

Nesta abordagem proposta por Figuerêdo e Calumby (2022), o número de *clusters* é variável e guiado por otimização não-supervisionada. No trabalho, os autores utilizaram o *dataset* de imagens da tarefa *Retrieving Diverse Social Images Task from MediaEval 2015* (Ionescu et al., 2014) (Müller et al., 2015) em um arcabouço que se inicia com a pré-seleção das imagens por relevância textual calculada pela média dos descritores *Cosine* e *Jaccard*, conforme podemos ver através da Figura 2.19, com exclusão de itens não-relevantes (Etapas 1 e 2). As imagens mais relevantes da etapa anterior passam pelo algoritmo de clusterização para descoberta de subtópicos implícitos (Etapas 3 e 4). Em seguida, as instâncias mais representativas são selecionadas por round-robin (Fürnkranz, 2002) nos *clusters* encontrados pelo algoritmo particional (Etapas 5 e 6). O conjunto de *clusters* de cada rodada de agrupamento (foi adotado um intervalo de *clusters* $k = [k_i, k_f]$ para o trabalho) é avaliada e a melhor configuração entre todas as outras é selecionada para a consulta. Esse é o fluxo de trabalho comum para a tarefa do *Diverse MediaEval* com a adoção, pelos autores, de um procedimento de otimização não supervisionado (etapa 4) e pela configuração do número de *clusters* de forma variável para cada consulta.

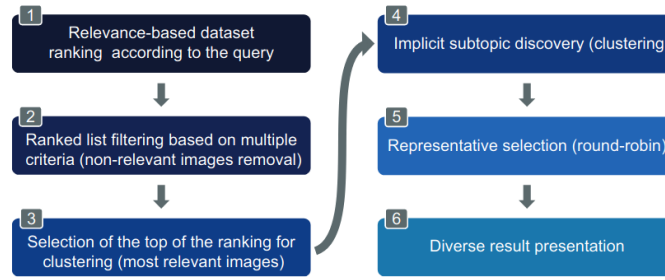


Figura 2.19: Fluxo de trabalho (Figuerêdo e Calumby, 2022)

Com o propósito de reduzir o custo computacional, os autores consideraram utilizar um intervalo entre um $K_{\min}$ e um $K_{\max}$ com variação dos esquemas de agrupamento em torno do número comumente adotado pela comunidade de pesquisa de $k = 20$ clusters. Os valores para $K_{\min}$ e $K_{\max}$ foram escolhidos arbitrariamente próximos do valor de K . Assim, foram definidos $K_{\min} = 15$ e $K_{\max} = 25$. A Figura 2.20 mostra o diagrama do fluxo de trabalho, detalhado, proposto pelos autores.

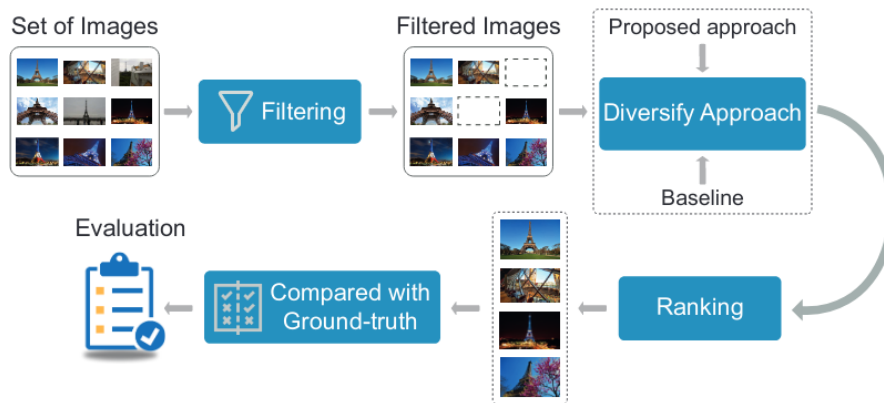


Figura 2.20: Fluxo de trabalho (Figuerêdo e Calumby, 2022).

A similaridade foi calculada com base em features visuais (ACC, CN3x3, LUM, SCH, Gabor, Tamura, CEDD, FCTH, JCD, PHOG e CNN AD) e textuais (Cosine, BM25, Dice e Jaccard). Foi aplicado o algoritmo *K-Medoids* para gerar os agrupamentos, e as medidas de eficácia *Precision* e *Cluster Recall*. As medidas avaliativas dos clusters foram obtidas pelos métodos *Silhouette*, *Dunn Index*, *Davies–Bouldin Index (DB)*, *Xie Beni Index (XB)* e *Sum of Squared Error (SSE)*. O ranking inicial das imagens recuperadas do *Flickr* foi gerado com base no descritor textual BM25 (Baeza-Yates et al., 1999). O dataset possuía 222 tópicos de consulta com cerca de 300 imagens cada. Na Figura 2.21 é apresentado um exemplo do esquema de seleção do número de grupos ideal (*best K*) utilizando o método do Cotovelo através da análise da curva gerada pelo *Silhouette*.

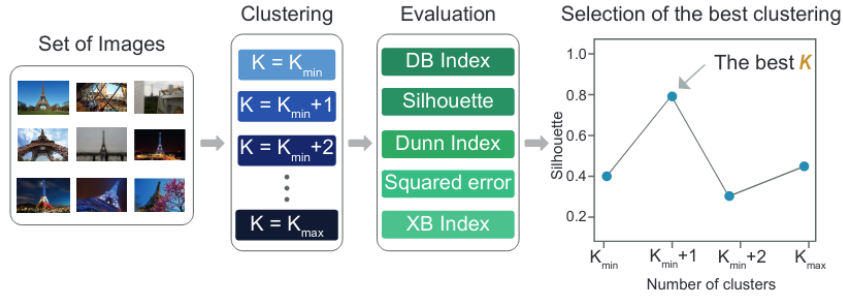


Figura 2.21: Fluxo de trabalho, descoberta de subtópico implícito, exemplo de seleção do número de grupos ideal (*best K*) através do *Silhouette Score* exemplificado no artigo (Figuerêdo e Calumby, 2022)

Conclusão

Adotar um k variando entre um número mínimo e um máximo, em torno de um valor de consenso na comunidade de pesquisa, como em Figuerêdo e Calumby (2022) (Subseção 2.6.3), mostra-se bastante promissor. O espectro de esquemas de *cluster* ficou em $15 \leq k \leq 25$. A utilização de uma ampla gama de descritores (textuais e visuais), como adotado pelos autores, também é algo a ser pontuado positivamente na análise da efetividade do sistema CBIR abordado. Além disso, o método adotado para a seleção de instâncias, com o uso do *Round Robin*, partindo do *cluster* mais relevante e da imagem mais relevante de cada *cluster* em ordem decrescente, corrobora a hipótese da variação semântica presente na variabilidade *intracluster* e *intercluster* vista em Chen et al. (2003). Contudo, não foi apresentada uma proposta para aplicações automatizadas como em Yu et al. (2014).

2.6.4 Um método automático para determinar o número de *clusters* usando *decision-theoretic rough sets*

As definições da Seção 2.4.3 foram utilizadas no algoritmo automático de clusterização ao qual os autores (Yu et al., 2014) chamaram de ACA-DTRS. O ACA-DTRS compara e avalia dois esquemas de agrupamento através da função de risco e interrompe o algoritmo em algumas situações:

- a) se todos os elementos em $\text{sim}(x_i, x_j)$ são iguais;
- b) se todos os elementos em $\text{sim}(x_i, x_j) \leq \text{val}$;
- c) se $\text{Risk}(CS_u) < \text{Risk}(CS_v)$ e $|CS_u| = 1$;
- d) se $\text{Risk}(CS_u) \geq \text{Risk}(CS_v)$.

Onde: $\text{sim}(x_i, x_j)$ é a distância (similaridade) entre os elementos x_i e x_j ; $\text{Risk}(CS_u)$ é o risco obtido pela mesclagem de dois esquemas de agrupamento C_i e C_j para um

novo esquema CS_u , ou seja, a etapa 5 do algoritmo ACA-DTRS descrito em Yu et al. (2014); $Risk(CS_v)$ é o risco atribuído por um esquema de agrupamento CS_v qualquer; e $|CS_u|$ é o número de elementos do *cluster* CS_u . O valor val foi definido na Seção 2.4.3.

Partindo dos experimentos com o ACA-DTRS, os autores desenvolveram outro algoritmo (FACA-DTRS), que avalia o comportamento da curva de risco, ilustrada na Figura 2.22, cujo critério de parada está relacionado à perda da monotonicidade dessa função. Por exemplo, uma mudança de comportamento para ascensão na curva de risco do CS_1 , $k = 15$ (hipotético), para CS_2 , $k = 16$, indicaria que CS_1 é o esquema que tem a melhor configuração de clusters, pois foi aquele de menor risco. O algoritmo, então, terminaria a tarefa. Com essa abordagem, os autores conseguiram reduzir o custo computacional do sistema.

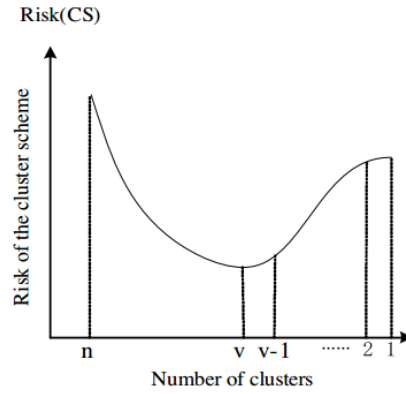


Figura 2.22: Curva da função de análise de risco, ponto de parada (Yu et al., 2014)

Tanto no ACA-DTRS quanto no FACA-DTRS, a escolha da melhor configuração de *clusters* foi atingida sem qualquer interferência manual. O experimento foi realizado com *datasets* do repositório de dados UCI (Asuncion e Newman, 2007). A Tabela 2.4 mostra, de forma exemplificativa, os resultados do processo aglomerativo, os esquemas de agrupamento e o risco de cada esquema (CS_t). A curva de risco gerada pelo método DTRS aplicado ao algoritmo ACA-DTRS pode ser vista na Figura 2.23.

CS_t - clustering results in the agglomerate processing	$Risk(CS_t)$
$CS_{11} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}\}$	52.8574
$CS_{10} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7, x_{10}\}, \{x_8\}, \{x_9\}, \{x_{11}\}\}$	50.8779
$CS_9 = \{\{x_1, x_3\}, \{x_2\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_7, x_{10}\}, \{x_8\}, \{x_9\}, \{x_{11}\}\}$	49.0436
$CS_8 = \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5\}, \{x_6\}, \{x_7, x_{10}\}, \{x_8\}, \{x_9\}, \{x_{11}\}\}$	47.2321
$CS_7 = \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5, x_6\}, \{x_7, x_{10}\}, \{x_8\}, \{x_9\}, \{x_{11}\}\}$	43.6750
$CS_6 = \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5, x_6\}, \{x_7, x_{10}\}, \{x_8, x_9\}, \{x_{11}\}\}$	41.9206
$CS_5 = \{\{x_1, x_3\}, \{x_2\}, \{x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}\}, \{x_{11}\}\}$	34.9554
$CS_4 = \{\{x_1, x_2, x_3\}, \{x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}\}, \{x_{11}\}\}$	31.5127
$CS_3 = \{\{x_1, x_2, x_3, x_{11}\}, \{x_4, x_5, x_6\}, \{x_7, x_8, x_9, x_{10}\}\}$	28.4527
$CS_2 = \{\{x_1, x_2, x_3, x_4, x_5, x_6, x_{11}\}, \{x_7, x_8, x_9, x_{10}\}\}$	39.4886
$CS_1 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}$	79.2677

Tabela 2.4: Resultado do processo hierárquico aglomerativo para o *dataset* X11, ACA-DTRS (Yu et al., 2014)

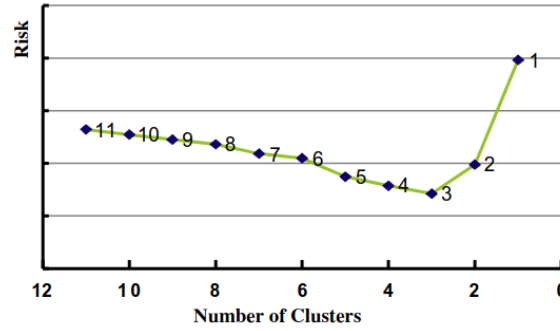


Figura 2.23: Curva da função de análise de risco para o *dataset* X11, ACA-DTRS (Yu et al., 2014)

Conclusão

A abordagem proposta em (Yu et al., 2014) (Subseção 2.6.4), baseada na aplicação probabilística dos conceitos da Teoria dos Conjuntos Aproximados de Pawlak, se mostrou bastante eficaz para determinar o melhor esquema de agrupamentos conforme podemos ver nas tabelas 2.5 e 2.6. A Tabela 2.5 mostra os resultados do algoritmo ACA-DTRS, onde a coluna $|U|$ representa o número de instâncias do *dataset*, as demais colunas o número de *clusters* encontrado por cada índice de validação e a última coluna o número real de *clusters* do respectivo *dataset*. O algoritmo hierárquico FACA-DTRS apresentou significativa redução no tempo de execução em relação ao ACA-DTRS devido à análise e uso da curva de risco como critério de parada. Contudo, a análise pura e simples da monotonicidade da curva de risco como critério para escolha do melhor CS_t pode levar a resultados distorcidos se o *dataset* possuir muitas classes ocultas. Ademais, um número de *clusters* muito alto tende a apresentar picos diferentes na curva de risco, o que pode levar a uma parada precoce do algoritmo devido a mudança de tendência momentânea.

Dataset	U	COPS	XB	Dunn	DTRS	True Cluster
IRIS	150	8	12	8	2	3
Wine	178	20	13	4	5	3
Ecoli	336	38	17	4	6	8
Vowel	528	37	22	5	4	11
Letter1(A,B,C)	300	35	15	9	3	3
Letter2(A,F,I,Z)	785	52	26	9	4	4
Poker(1-9)	938	20	15	11	8	9

Tabela 2.5: Resultado comparativo ACA-DTRS (Yu et al., 2014) realizado para os *datasets* do repositório UCI (Asuncion e Newman, 2007). Coluna |U| representa o número de elementos do *dataset*. A última coluna indica o número real de agrupamentos. Demais colunas mostram o número de agrupamentos sugerido pelo respectivo índice de validação.

A abordagem de interrupção pela análise da curva de risco no algoritmo FACA-DTRS não impactou negativamente de forma severa sobre a acurácia em relação ao algoritmo ACA-DTRS como pode ser visto na Tabela 2.6.

Database	U	C	FACA-DTRS			ACA-DTRS		
			Accuracy	CPU(s)	C'	Accuracy	CPU(s)	C'
X150	150	3	97	0.082	3	98	1.180	3
Wine	178	3	85	0.265	4	80	2.110	5
Ecoli	336	8	89	0.375	7	81	13.719	6
Vowel	528	11	77	0.734	4	78	52.203	4
Poker(1-9)	938	9	84	2.744	8	88	332.203	8
Letter1	1014	3	81.4	2.984	3	80.6	365.469	3
Letter2	785	4	81.3	1.654	3	88.1	179.550	4
Letter3	1124	3	85.4	4.279	3	87.4	505.922	3

Tabela 2.6: Resultado do experimento comparativo ACA-DTRS *vs* FACA-DTRS (Yu et al., 2014). A coluna |U| representa o número de elementos do *dataset*. A coluna |C| indica o número real de agrupamentos. A coluna |C'|, o número de agrupamentos encontrado pelo respectivo algoritmo.

Nenhuma comparação com algoritmos particionais como o *K-Means* ou o *K-Medoids* foi realizada pelos autores. O estudo apresenta uma proposta promissora para avaliação da eficácia dos esquemas de agrupamento em aprendizado não-supervisionado, algo que pode apresentar vantagem na solução de um dos problemas envolvidos na recuperação de imagens.

Capítulo 3

Metodologia

“A verdadeira questão não é se as máquinas pensam, mas se os homens fazem.”

– B.F. Skinner

Como vimos, a tarefa de encontrar o número de grupos ideal gerado por um algoritmo de *clusterização* depende da seleção de parâmetros e do método de *cluster validity* adotado na abordagem. No contexto da CBIR, onde o resultado do algoritmo é um conjunto de imagens a ser apresentado ao usuário, algo que normalmente está associado a mecanismos de consulta da web ou repositórios, é importante a observação aos aspectos da relevância e diversidade desses resultados, além do custo computacional e de uma abordagem sem interferência humana.

Este trabalho visa oferecer ganhos na tarefa de recuperação de imagens baseada em conteúdo através do uso de algoritmos de aprendizado não-supervisionado e da análise da qualidade dos agrupamentos pelo modelo estendido da teoria dos conjuntos aproximados de Pawlak (Decision-theoretic Rough Sets - DTRS), uma vez que, este modelo, baseado na teoria da inferência bayesiana, tem sido utilizado com sucesso na classificação e identificação de padrões (Yu et al., 2014). Além disso, a abordagem proposta explora o conceito de diversificação em CBIR tendo como base comparativa o arcabouço apresentado por (Figueiredo e Calumby, 2022) e a tarefa *Retrieving Diverse Social Images - MediaEval 2015*. Apresenta, ainda, um método matemático para otimização do tempo computacional pela interrupção do algoritmo de agrupamento e faz uma análise comparativa entre as performances obtidas para cada um dos métodos de avaliação de agrupamentos aplicados a descritores amplamente conhecidos por pesquisadores da área de visão computacional.

3.1 O arcabouço

O arcabouço, como pode ser visto na Figura 3.1, se inicia com a matriz de características das imagens já extraída (*features*). Para efeito comparativo, utilizamos os descritores presentes no *Dataset Retrieving Diverse Social Images Task from MediaEval 2015 - Devset* (Ionescu et al., 2014) e explorados no trabalho relacionado da Subseção 2.6.3 (Figuerêdo e Calumby, 2022) nas matrizes de distância. Os conjuntos de parâmetros iniciais (etapa 01) podem ser vistos nas tabelas 3.1 à 3.3. Após a parametrização inicial, é realizada uma filtragem de imagens, baseada nos descritores textuais *Cosine* ou *Jaccard*, visando a redução do custo computacional. O resultado é um ranking com as 150 imagens mais bem pontuadas, em relação a esses descritores, para cada tópico de consulta.

Na quarta etapa, o algoritmo de aprendizado não-supervisionado formará os esquemas de agrupamento que serão validados na etapa seguinte (etapa 05), dando origem a um ranking de relevância. O método estendido para o *Decision-theoretic rough set* (DTRS) foi acrescentado nesta etapa de validação (como índice de validação de agrupamento), bem como o critério matemático para interrupção do algoritmo de *clusterização*, objetivando a redução do custo computacional, a ser visto na Seção 3.2.2. O esquema de agrupamentos resultante (o número de grupos ideal K), apresentado pelo índice de validação de agrupamento nesta primeira execução do algoritmo de *clusterização*, é, então, posto em uma nova execução para que, desta vez, tenhamos o ranking de diversidade.

Ao final, serão selecionadas as 50 melhores imagens que irão para a avaliação da eficácia do método (nona e décima etapas), pelas métricas *Precision*, *Cluster Recall* e *F1-Measure*, em relação à base comparativa de imagens fornecida pela *Task MediaEval*.

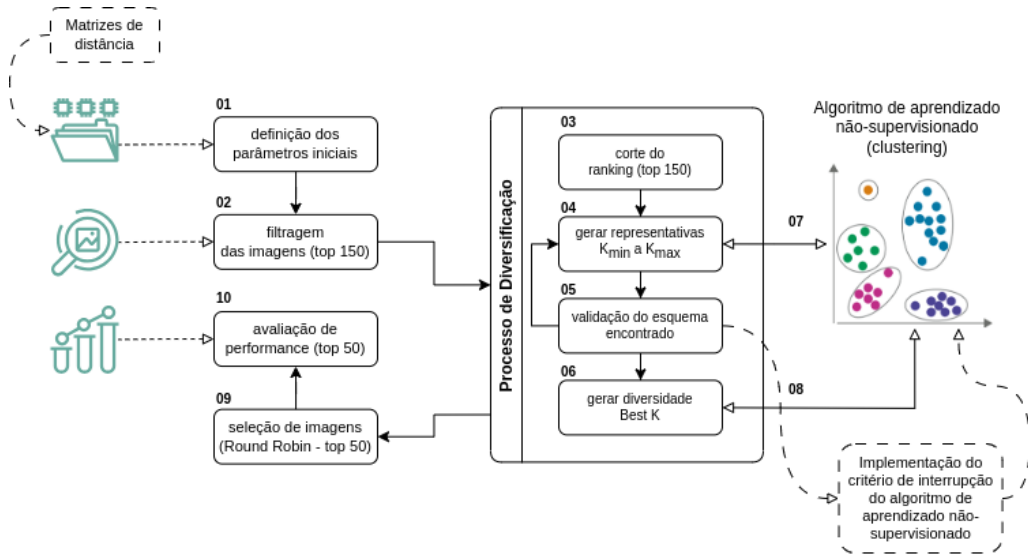


Figura 3.1: Diagrama do arcabouço proposto. Baseado no trabalho relacionado da Subseção 2.6.3 (Figuerêdo e Calumby, 2022)

Todo o processo é executado de forma individual para cada um dos 15 descritores da abordagem aplicados aos 6 métodos de *cluster validity*.

3.1.1 O dataset

Foi utilizado o dataset de imagens *Retrieving Diverse Social Images Task from MediaEval 2015* (Müller et al., 2015) que consiste em 150 consultas (*queries*) monoconceituais com aproximadamente 45000 imagens para desenvolvimento e 140 consultas com trechos textuais extraídos do *Flickr* (70 monoconceituais - 70 multiconceito), 42000 imagens para teste, mapeadas em descritores textuais e visuais. O contexto monoconceitual está relacionado à ideia de que algumas imagens são focadas em um único conceito ou tema, enquanto que multiconceitual incorpora múltiplos conceitos ou temas. Por exemplo, uma imagem monoconceitual pode ser uma fotografia que primariamente enfatiza uma ideia central, objeto ou assunto. Por outro lado, uma imagem multiconceitual pode envolver diversos elementos ou temas.

A base comparativa, ou rotulada (*ground-truth*), fornecida pela *Retrieving Diverse Social Images Task from MediaEval 2015*, para aplicação das medidas de avaliação da eficácia do sistema CBIR, foi produzida por anotadores humanos com critérios de representatividade e diversificação. Ela apresenta a classificação real das imagens do *dataset*, indicando tanto o tópico de consulta quanto o grupo semântico ao qual uma imagem pertence. O grupo semântico reflete a diversidade contextual associada à consulta, ou seja, as variações nas imagens de um determinado aspecto de consulta.

3.2 Os experimentos

Todos os experimentos realizados neste estudo foram baseados no arcabouço tratado no artigo da Seção 2.6.3, cuja adaptação pode ser vista no diagrama da Figura 3.1. Utilizamos um total de 222 consultas das 290 disponíveis no dataset da *Retrieving Diverse Social Images Task from MediaEval 2015*. Sendo que, para cada uma das 222 consultas, utilizamos cerca de 300 imagens extraídas do *Flickr*.

Após a definição dos parâmetros iniciais, a execução segue para uma etapa de filtragem das imagens, cujo processo é realizado com base em dados de descritores textuais, resultando em um ranking das 150 imagens mais relevantes entre as cerca de 300 mencionadas anteriormente. Essa etapa de filtragem inicial (pré-filtragem, descrita na Subseção 2.6.3) foi realizada com o propósito de reduzir o custo computacional da tarefa de recuperação. Neste sentido, Figuerêdo e Calumby (2022) utilizaram a média dos resultados de eficácia dos descritores textuais *Cosine* e *Jaccard* como medida para a determinação da relevância das imagens.

No entanto, neste estudo, após a execução de experimentos preliminares, observamos que a pré-filtragem utilizando apenas o descritor textual *Jaccard* demonstrou superioridade nos resultados gerais do sistema CBIR (eficácia em torno de 5%), independentemente do método de avaliação (*Precision*, *Cluster Recall*, *F1-Measure*),

em 70% dos casos, individualizados por descritor e índice de validação, se comparados aos experimentos com uso apenas do descritor *Cosine*. Tal fato tornou a adoção da média dos resultados da pré-filtragem entre os descritores *Cosine* e *Jaccard* pouco interessante, uma vez que a aplicação da média entre os dois descritores poderia provocar uma redução significativa nos valores finais de eficácia do sistema de recuperação.

Em relação ao processo de identificação dos padrões em termos de relevância e diversificação nos conjuntos de imagens (seleção dos grupos semânticos), etapas 04 e 06 do diagrama da Figura 3.1, foram utilizados dois algoritmos de agrupamento diferentes (K-Medoids e Hierárquico Aglomerativo) em execuções distintas para cada experimento e em acordo com os objetivos citados no Capítulo 1, Seção 1.2.

No caso do objetivo inicial (Item a. da Seção 1.2), foi utilizado o algoritmo *K-Medoids*, com k variando entre [15, 25], em execuções individuais e independentes entre si, realizadas por descritor, índice de validação (*cluster validity*) e tópico de consulta. Os resultados foram dispostos em tabelas e gráficos para os 15 descritores utilizados, individualmente, com os 6 índices de validação do estudo contrapostos ao *baseline* ($k = 20$) e em valores médios extraídos dos 222 tópicos de consulta do experimento. O *baseline* foi adotado para que pudéssemos realizar uma análise comparativa entre um número fixo de agrupamentos, algo comumente utilizado pela comunidade científica, e o intervalo adotado em cada um dos experimentos. Este *baseline* foi obtido pela execução do *K-Medoids* com $k = 20$ e sem a utilização das técnicas de filtragem textual, índices de validação de agrupamentos ou qualquer outro método auxiliar para seleção do número de grupos ideal como, por exemplo, o critério de interrupção que veremos na Seção 3.2.2.

Em relação ao segundo objetivo (item b. da Seção 1.2), adotamos uma abordagem dinâmica (automatizada) para o sistema de recuperação na qual fizemos uso de um critério para interrupção do algoritmo de agrupamento na etapa de geração de diversidade (etapa 06 do diagrama da Figura 3.1) e que pode ser visto na Seção 3.2.2. Neste caso, utilizamos o algoritmo Hierárquico Aglomerativo com k variando entre [1, 150] e com a aplicação da filtragem inicial baseada no descritor *Jaccard*. O *baseline* de performance utilizado neste segundo experimento foi o mesmo do objetivo anterior (*K-Medoids* com $k = 20$) e o algoritmo Hierárquico Aglomerativo foi adotado por considerarmos mais adequado para a implementação da abordagem dinâmica.

Um terceiro experimento principal, com os mesmos parâmetros do anterior, foi realizado com a utilização do algoritmo K-Medoids e com o objetivo de comparar a performance, em termos de tempo de execução e eficácia do sistema de recuperação de imagens, entre esses dois diferentes métodos de agrupamento (Hierárquico Aglomerativo e K-Medoids). Além disso, fizemos mais seis execuções adicionais com experimentações em torno do intervalo de k grupos e de diferentes métodos auxiliares para a seleção do número de grupos ideal: o método do Cotovelo, em sua forma calculada; e o método dinâmico, proposto neste trabalho, para interrupção

do algoritmo de agrupamento. Em todos os casos, as execuções ocorreram de forma distinta, ou seja, por tópico de consulta, por descritor e por índice de validação de agrupamento.

Adotamos o teste de significância estatística não-paramétrico (amostras pareadas) de *Wilcoxon* para a análise das diferenças amostrais nos resultados. Esta foi uma etapa não pertencente ao arcabouço e realizada após a confirmação da normalidade pelo teste de *Shapiro-Wilk*. Neste caso, tomamos o índice de validação DTRS no pós-intervenção do teste estatístico de *Wilcoxon* com o *baseline* e demais índices no pré-intervenção.

3.2.1 Os descritores utilizados

Os descritores utilizados foram os seguintes:

- Visuais
 - ACC (Penatti et al., 2012); LUM (Lux e Chatzichristofis, 2008), SCH (Cinque et al., 2001), GABOR (Manjunath e Ma, 1996) e TAMURA (Tamura et al., 1978); CN3x3 e CNN_AD (Ionescu et al., 2014); CEDD (Chatzichristofis e Boutalis, 2008a); FCTH (Chatzichristofis e Boutalis, 2008b); JCD (Zagoris et al., 2010); PHOG (Biasotti et al., 2013);
- Textuais
 - COSINE (Cosseno) e BM25 (Baeza-Yates et al., 1999); DICE (Dice, 1945) e JACCARD (Lewis et al., 2006).

3.2.2 Critério de interrupção do algoritmo

Adotamos um critério para interrupção do algoritmo de agrupamento na etapa de geração de diversidade das imagens (etapa 06 do diagrama da Figura 3.1), baseado na análise da monotonicidade da curva gerada pelos valores do coeficiente angular da reta secante (pontos inicial e final, ou atual) aplicada à curva da média móvel (aritmética) extraída dos valores obtidos por cada índice de validação de agrupamento (*cluster calidity*), individualmente, num processo dinâmico para a escolha do *best K*.

A adoção deste critério de interrupção teve como objetivo principal a redução do tempo de execução do sistema de recuperação e foi aplicado de duas formas, conforme pode ser visto na Figura 3.2, de acordo com o comportamento de cada índice de validação de agrupamento e suas particularidades:

- I. pela inflexão positiva do intervalo de análise da curva do coeficiente angular da média móvel, $f(k_{a-2}) < f(k_{a-1}) > f(k_a)$, supondo uma mudança de comportamento de crescente para decrescente;
- II. pela sua oposta, $f(k_{a-2}) > f(k_{a-1}) < f(k_a)$, inflexão negativa, quando a curva da média móvel mudou o seu comportamento de decrescente para crescente.

Onde, $f(k_a)$ é o valor resultante do método de validação em função de um esquema k_a de *clusters*. Sendo, portanto, $f(k_{a-1})$ e $f(k_{a-2})$ os respectivos valores para os dois esquemas de *clusters* anteriores ao ponto k_a .

O ponto de inflexão, neste caso, pode indicar o alcance máximo ou mínimo para os valores do índice de validação de agrupamento em análise e, portanto, a melhor configuração de grupos naquele instante. Tomemos um índice de validação cuja melhor configuração em um conjunto de *clusters* seja dada pelo maior valor obtido por este método (*Silhouette* e *Dunn Index*, por exemplo). Um ponto de inflexão para a curva do coeficiente angular da média móvel gerada por este índice de validação de agrupamento deverá, logicamente, indicar o ponto máximo desta curva cuja tendência de queda se inicie logo após uma sequência ascendente, levando-se em conta um determinado intervalo de análise. Assim, um ponto de inflexão positiva seria um possível critério de parada para este método de validação. Essa condição pode ser vista na Figura 3.2(a).

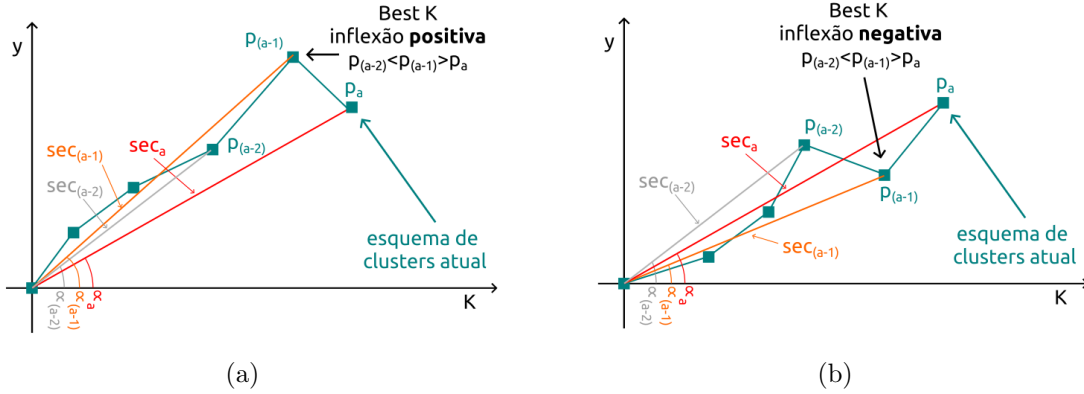


Figura 3.2: Critério de interrupção do algoritmo hierárquico aglomerativo baseado no comportamento do coeficiente angular da curva da média móvel extraída dos valores gerados pelo *cluster validity*, pontos de inflexão positiva (Figura 3.2(a)) e negativa (Figura 3.2(b)) para os casos de curvas convexas e côncavas.

O mesmo raciocínio pode ser utilizado para índices de validação de agrupamento cujo número de grupos ideal se apresente através de valores mínimos (*DTRS*, *SSE*, *Xie Beni*, *Davies Bouldin*). Sendo, portanto, de maneira inversa ao anteriormente adotado, um ponto de inflexão negativa o mais adequado como critério de parada

do algoritmo de agrupamento para esses métodos. Um exemplo ilustrativo para a inflexão negativa pode ser visto na Figura 3.2(b).

Essa estratégia foi necessária, pois, o comportamento da curva gerada pelos índices de validação de agrupamento nem sempre se apresenta de forma esperada como podemos ver através do gráfico da Figura 3.3. A presença de *outliers* e as características do dataset, ou do descritor, influenciam fortemente na tarefa de seleção do número de grupos ideal à medida em que a quantidade de agrupamentos para análise se estende entre um k_{min} e k_{max} . Assim, a adoção da média móvel aritmética no intervalo t teve o objetivo de promover a suavização da curva gerada pelos resultados do índice de validação de agrupamento, reduzindo valores de pico e diminuindo a possibilidade de interrupção prematura do algoritmo de *clusterização*.

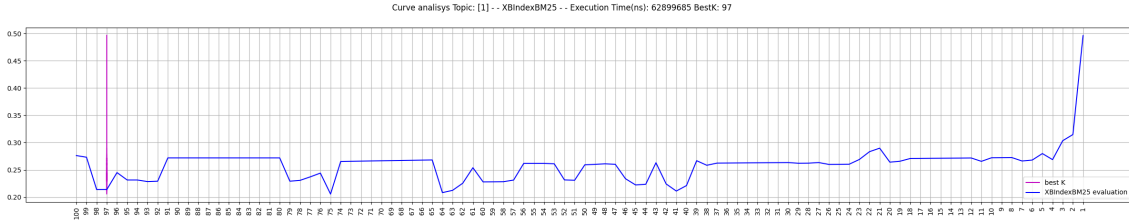


Figura 3.3: Curva gerada pelos valores obtidos pelo índice de validação de agrupamento *Xie Beni* (número de grupos ideal pelo valor mínimo) para o primeiro tópico de consulta do experimento e k partindo de 100 *clusters*. *Best k* em 97. Algoritmo Hierárquico Aglomerativo.

Adicionalmente, ainda com o propósito de evitar essa interrupção prematura, foi utilizada uma constante limitadora l (limiar) aplicada à distância de inflexão, medida do ponto $f(k_{a-1})$ ao ponto médio do intervalo de análise da curva do coeficiente angular da reta secante à curva da média móvel, fazendo com que variações muito baixas também não fossem detectadas. Neste caso, as condições para interrupção do algoritmo foram dadas de acordo com o tipo de inflexão (positiva ou negativa), garantindo que apenas mudanças de comportamento suficientemente grandes provocassem a interrupção do algoritmo. Assim, tivemos:

- **Cálculo do ponto médio no intervalo:**
 - $pontoMedio = [f(k_{a-2}) + f(k_a)]/2$;
- **Distância do ponto médio ao ponto de inflexão:**
 - $distInflexao = pontoMedio - f(k_{a-1})$;
- **Caso a inflexão seja positiva:**
 - Calcular: $pontoTeste = pontoMedio + (distInflexao \times l)$;
 - Interromper se: $f(k_{a-2}) < pontoTeste > f(k_a)$;
- **Caso a inflexão seja negativa:**
 - Calcular: $pontoTeste = pontoMedio - (distInflexao \times l)$;

- Interromper se: $f(k_{a-2}) > \text{pontoTeste} < f(k_a)$.

Caso não haja detecção do ponto de inflexão ao longo da curva do coeficiente angular com características suficientes para provocar a interrupção do algoritmo, uma última estratégia é adotada, baseada em resultados empíricos, apresentando o antepenúltimo esquema de agrupamentos (k_{a-2}) como o mais hábil (número de grupos ideal). A Figura 3.4 ilustra o cálculo para um ponto cuja inflexão negativa foi suficientemente grande para gerar a interrupção.

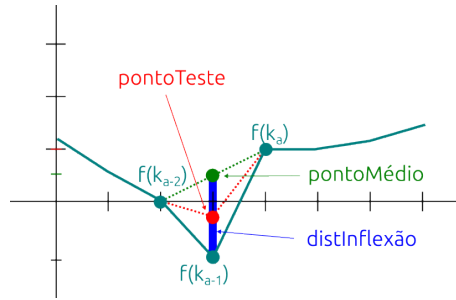


Figura 3.4: Gráfico ilustrativo do cálculo para um ponto de inflexão negativa pela abordagem proposta.

Nos experimentos, foram utilizados valores para o intervalo t (média móvel) entre 5 e 10 e, para o limiar l , entre 0,1 e 0,9, com melhores resultados em $t = 10$ e $l = 0,5$. Na Figura 3.5 temos um diagrama de fluxo mais detalhado do sistema de recuperação de imagens, com os pontos de tomada do tempo de execução e do critério de interrupção definidos.

Para efeito de simplificação, a partir deste ponto chamaremos o critério de interrupção, proposto por este trabalho, pelo acrônimo IASMIN (***I**nterrupção pela **A**nálise da Reta **S**ecante da **M**édia Móvel do **I**ndice*).

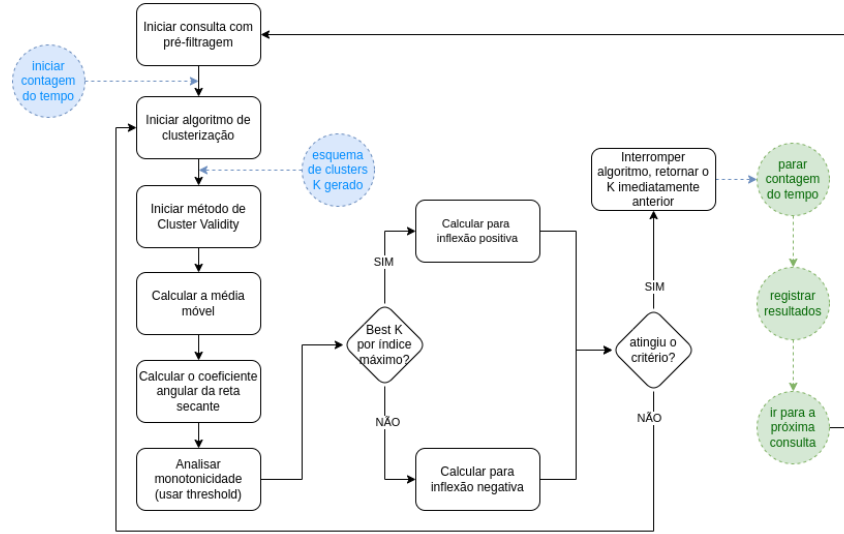


Figura 3.5: Diagrama de fluxo da etapa de execução do critério de interrupção (IASMIN).

3.2.3 Avaliação da performance

A performance do sistema de recuperação baseado no arcabouço proposto foi medida com a extração dos resultados de cada um dos índices de validação de agrupamento associados aos 15 descritores visuais e textuais do experimento. Para cada esquema de agrupamento resultante (o número de grupos ideal encontrado pelo índice de validação) associado ao seu respectivo descritor e tópico de consulta, foi preparado um ranking com as 50 primeiras imagens retornadas em ordem de similaridade (ou dissimilaridade).

Essas imagens ranqueadas foram avaliadas utilizando-se as bases comparativas (*Ground-truth*) de relevância e diversidade fornecidas com o *dataset* da tarefa *MediaEval 2015* e com o uso das métricas *Precision*, *Cluster-Recall* e *F1-Measure* ($P@n$, $CR@n$ e $F1@n$, respectivamente) em análise comparativa de subgrupos variando entre $5 \leq r \leq 50$ imagens (onde r é um subconjunto do ranking) para cada um dos 222 tópicos de consulta. Ao final, um valor médio de performance, baseado nesses 222 tópicos, foi calculado para cada um dos índices de validação de agrupamento associado ao seu respectivo descritor.

3.2.4 Observações estatísticas

A análise estatística descritiva foi realizada com os resultados médios de performance obtidos na etapa anterior para cada um dos índices de validação de agrupamento. Foi aplicado o teste de hipótese pareado e não-paramétrico de *Wilcoxon* (Woolson, 2007) além do teste de *Shapiro-Wilk* (Shapiro e Wilk, 1965) e representações gráficas como histogramas e o gráfico quantil-quantil (*Q. Q. Plot*) com o intuito de i) mostrar o comportamento desses resultados quanto à sua normalidade e ii) apresentar os valores de performance mais significativos em relação ao *baseline* ($k = 20$).

3.2.5 Parametrização

Os parâmetros básicos definidos para o experimento estão descritos na tabela 3.3. Como dito na Seção 3.2.3, o ranking das 150 melhores imagens, em termos de relevância e diversificação, foi ainda seccionado em subgrupos com as 5, 10, 20, 30, 40 e 50 melhores imagens para avaliação da performance.

Tabela 3.1: Parâmetros básicos adotados na geração do *baseline*

Tipo de parâmetro	Valor adotado
Algoritmo de agrupamento	K-Medoids
Descritores visuais	ACC, LUM, SCH, Gabor, Tamura, CN3x3, CNN AD, CEDD, FCTH, JCD e PHOG
Descritores textuais	Cosseno, BM25, Dice e Jaccard
Índices de validação de agrupamento	Não necessário
Intervalo K	[20,20]
Método auxiliar seleção best K	Não necessário
Normalização dos dados	Não utilizado (Falso)
Rank textual	Não utilizado (Falso)
Re-ordenação	Não utilizado (Falso)
Limiar do IASMIN	Não utilizado
Intervalo média móvel	Não utilizado
Truncate size	150 (top-150 imagens)

Tabela 3.2: Parâmetros básicos adotados no experimento para avaliação do DTRS (item a., Seção 1.2, experimento 01)

Tipo de parâmetro	Valor adotado
Algoritmo de agrupamento	Experimento 01: K-Medoids
Descritores visuais	ACC, LUM, SCH, Gabor, Tamura, CN3x3, CNN AD, CEDD, FCTH, JCD e PHOG
Descritores textuais	Cosseno, BM25, Dice e Jaccard
Índices de validação de agrupamento	Dunn Index, Silhouette, Davies Bouldin, Xie Beni, SSE e DTRS
Intervalo K	[15,25]
Método auxiliar seleção best K	índice de validação
Normalização dos dados	Não utilizado (Falso)
Rank textual	Jaccard
Re-ordenação	Não utilizado (Falso)
Limiar do IASMIN	Não (Falso)
Intervalo média móvel	Não utilizado
Truncate size	150 (top-150 imagens)

Tabela 3.3: Parâmetros básicos adotados nos experimentos 02 e 03 para avaliação do IASMIN (item b., Seção 1.2), eficácia do DTRS e tempo de execução.

Tipo de parâmetro	Valor adotado
Algoritmo de agrupamento	Experimento 02: Hierárquico Aglomerativo Experimento 03: K-Medoids
Descritores visuais	ACC, LUM, SCH, Gabor, Tamura, CN3x3, CNN AD, CEDD, FCTH, JCD e PHOG
Descritores textuais	Cosseno, BM25, Dice e Jaccard
Índices de validação (inflexão positiva)	Dunn Index, Silhouette (best k pelo valor máximo)
Índices de validação (inflexão negativa)	Davies Bouldin, Xie Beni, SSE e DTRS (best k pelo valor mínimo)
Intervalo K	[1,150]
Método auxiliar seleção best K (ambos experimentos 02 e 03)	Execução 1: índice de validação Execução 2: índice de validação + Método do Cotovelo Execução 3: índice de validação + IASMIN
Normalização dos dados	Não utilizado (Falso)
Rank textual	Jaccard
Re-ordenação	Não utilizado (Falso)
Limiar do IASMIN	0,5
Intervalo média móvel	10 clusters
Truncate size	150 (top-150 imagens)
Relação intercluster	Complete Linkage (Tabela 2.1 - Hierárquico Aglomerativo)

Capítulo 4

Resultados

Neste capítulo, discutiremos os resultados alcançados pelos experimentos baseados nos conceitos e definições descritos na metodologia do Capítulo 3. Demonstraremos os resultados alcançados ao aplicar o método DTRS estendido (Yu et al., 2014) na validação dos esquemas de *cluster*, dentro do contexto da CBIR, assim como a eficácia da abordagem dinâmica (IASMIN) em relação à redução do tempo de execução e ao incremento de performance do sistema de recuperação de imagens baseado em conteúdo adotado. Na Figura 4.1 temos o diagrama em blocos dos experimentos executados neste estudo, resumindo os principais parâmetros adotados e os objetivos de cada experimento.

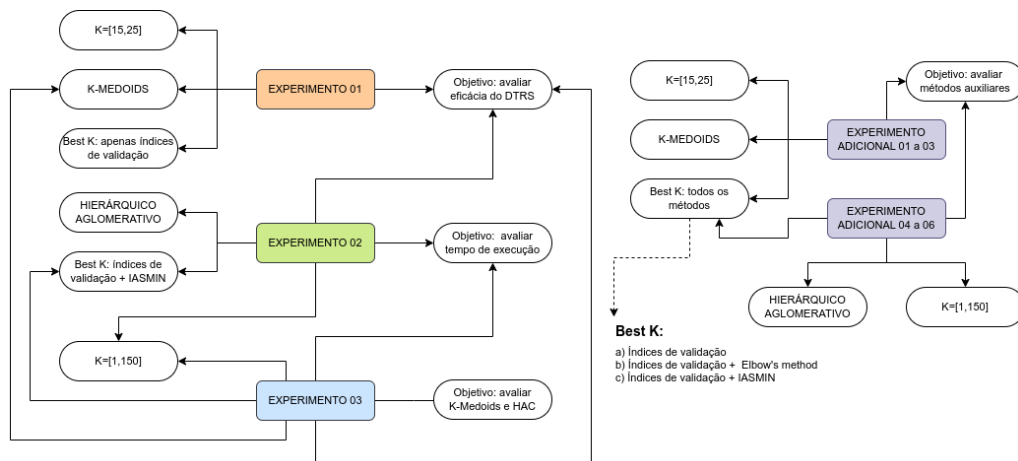


Figura 4.1: Diagrama em blocos dos experimentos: parâmetros básicos e objetivos.

4.1 Avaliação do método DTRS no sistema CBIR (experimento 01)

Nesta seção, demonstraremos a aplicabilidade do método DTRS no contexto da validação de agrupamentos, destacando os resultados alcançados entre os diferentes

descritores empregados nos experimentos. Além disso, realizaremos uma comparação detalhada desses resultados com aqueles obtidos por outros índices de validação de agrupamento. O algoritmo utilizado para esta etapa foi o K-Medoids, com seleção do número de grupos ideal pelo método natural do índice de validação de agrupamento (valores mínimos ou máximos, conforme o método). Demais parâmetros utilizados na execução deste primeiro experimento estão descritos na Tabela 3.2, cujo diagrama de blocos pode ser visto na Figura 4.2.

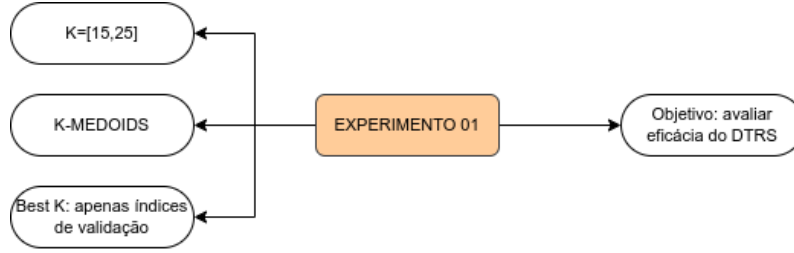


Figura 4.2: Diagrama: objetivo e parâmetros do experimento 01

4.1.1 Avaliação dos resultados de performance

A Tabela 4.1 apresenta os resultados alcançados pelo método DTRS no Experimento 01. Levando-se em conta a observação individualizada para este método e a análise comparativa entre descritores visuais e textuais, de acordo com o objetivo a. da Seção 1.2 (aplicabilidade do método DTRS com uso do algoritmo K-Medoids), temos que os melhores resultados, sem uso dos testes de significância estatístico e em termos de valores máximos atingidos em relação à medida *Precision*, concentraram-se entre os descritores visuais ACC ($P@5$), CEDD ($P@40$ e $P@50$), GABOR ($P@5$ à $P@30$), JCD ($P@40$ e $P@50$) e SCH ($P@5$ à $P@30$). Já entre os descritores textuais não houve destaque, em termos de valores máximos, para o método DTRS. Em relação à métrica *Cluster Recall*, apenas o descritor visual CN3x3 ($CR@20$ à $CR@50$) apresentou resultados superiores aos demais métodos de *cluster validity*. Enquanto que os descritores textuais BM25 ($CR@5$), COSINE ($CR@20$ à $CR@50$), DICE ($CR@10$) e JACCARD ($CR@5$ e $CR@10$) apresentaram superioridade do método DTRS frente aos demais.

Considerando que a métrica *Precision* representa a qualidade do ranking em termos de relevância e a métrica *Cluster Recall* representa diversidade, a observação concentrada no método DTRS (sem teste de significância) demonstrou que, em relação à relevância dos resultados da consulta, os descritores visuais apresentaram desempenho superior aos textuais, com os melhores resultados dispersos no ranking em todas as faixas ($5 \geq P@n \leq 50$). Em relação à diversidade, os melhores resultados, assim como na *Precision*, se concentraram em toda a profundidade do ranking ($5 \leq CR@n \leq 50$), porém, com mais ocorrências entre os descritores textuais.

		PRECISION					CLUSTER RECALL					F1 MEASURE																																		
		P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50																											
visual	ACC	0.8117	0.7878	0.7804	0.7730	0.7664	0.7594	0.1587	0.2613	0.4153	0.5154	0.5934	0.6511	0.2589	0.3842	0.5316	0.6067	0.6560	0.6877																											
	CEDD	0.7991	0.7595	0.7635	0.7737	0.7751	0.7709	0.1569	0.2658	0.4252	0.5284	0.5925	0.6451	0.2574	0.3868	0.5365	0.6170	0.6596	0.6899																											
	CN3x3	0.7730	0.7680	0.7653	0.7571	0.7521	0.7490	0.1591	0.2676	0.4351	0.5292	0.6044	0.6585	0.2569	0.3871	0.5428	0.6102	0.6569	0.6865																											
	CNN_AD	0.8090	0.7932	0.7766	0.7668	0.7596	0.7541	0.1560	0.2596	0.4090	0.5076	0.5911	0.6545	0.2559	0.3834	0.5262	0.5993	0.6526	0.6870																											
	FCHT	0.7865	0.7486	0.7588	0.7695	0.7709	0.7662	0.1568	0.2643	0.4254	0.5254	0.5902	0.6431	0.2567	0.3837	0.5347	0.6132	0.6562	0.6863																											
	GABOR	0.8117	0.8027	0.7905	0.7826	0.7730	0.7649	0.1593	0.2631	0.3997	0.5039	0.5815	0.6403	0.2599	0.3875	0.5208	0.6002	0.6496	0.6822																											
	JCD	0.7937	0.7581	0.7619	0.7746	0.7755	0.7704	0.1556	0.2671	0.4262	0.5281	0.5944	0.6443	0.2556	0.3882	0.5371	0.6172	0.6613	0.6891																											
	LUM	0.7982	0.7946	0.7757	0.7680	0.7588	0.7543	0.1519	0.2549	0.4112	0.5113	0.5913	0.6525	0.2490	0.3771	0.5267	0.6014	0.6508	0.6853																											
	PHOG	0.8063	0.7910	0.7818	0.7773	0.7682	0.7627	0.1570	0.2628	0.4132	0.5124	0.5928	0.6509	0.2574	0.3874	0.5297	0.6042	0.6552	0.6869																											
	SCH	0.8135	0.8014	0.7876	0.7805	0.7712	0.7630	0.1577	0.2607	0.4081	0.5131	0.5818	0.6416	0.2590	0.3838	0.5261	0.6067	0.6500	0.6830																											
textual	TAMURA	0.7892	0.7950	0.7858	0.7802	0.7729	0.7637	0.1533	0.2553	0.4062	0.5080	0.5830	0.6396	0.2523	0.3793	0.5243	0.6023	0.6515	0.6820																											
	BM25	0.7685	0.7586	0.7428	0.7380	0.7319	0.7326	0.1601	0.2731	0.4310	0.5234	0.5961	0.6508	0.2607	0.3945	0.5341	0.6004	0.6436	0.6754																											
	COSINE	0.7829	0.7802	0.7651	0.7593	0.7538	0.7516	0.1565	0.2745	0.4322	0.5330	0.6044	0.6565	0.2559	0.3996	0.5422	0.6155	0.6594	0.6879																											
	DICE	0.7541	0.7685	0.7448	0.7408	0.7404	0.7394	0.1524	0.2803	0.4301	0.5195	0.5899	0.6473	0.2491	0.4034	0.5333	0.5984	0.6432	0.6757																											
	JACCARD	0.7631	0.7572	0.7468	0.7407	0.7412	0.7401	0.1603	0.2770	0.4306	0.5238	0.5947	0.6497	0.2606	0.3982	0.5359	0.6004	0.6459	0.6775																											
	Max()	0.8135	0.8027	0.7905	0.7826	0.7755	0.7709	0.1603	0.2803	0.4351	0.5330	0.6044	0.6585	0.2607	0.4034	0.5428	0.6172	0.6613	0.6899																											
	Min()	0.7541	0.7486	0.7428	0.7380	0.7319	0.7326	0.1519	0.2549	0.3997	0.5039	0.5815	0.6396	0.2490	0.3771	0.5208	0.5984	0.6432	0.6754																											
		AVG 0.7699					Median 0.7684					Q 75% 0.7814					AVG 0.4336					Median 0.4695					Q 75% 0.5922					AVG 0.5200					Median 0.5706					Q 75% 0.6523				

Tabela 4.1: Experimento 01: resultado comparativo entre todos os descritores visuais e textuais apenas para o método DTRS sem teste de significância estatística, seleção do número de grupos ideal sem uso de métodos auxiliares (pela forma natural do método, ou seja, risco mínimo). Melhores resultados em negrito.

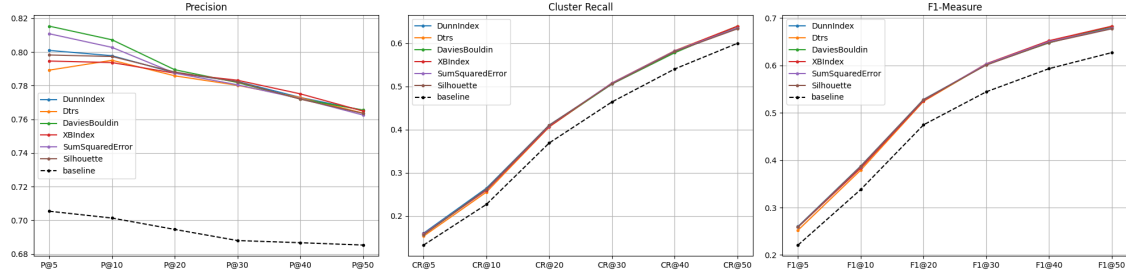
Nas Tabelas 4.2 à 4.16 são apresentados os resultados destacados com a aplicação do teste de significância estatística não-paramétrico de *Wilcoxon*. O nível de confiança escolhido foi de 95% (nível de significância $\alpha = 0,05$), com o teste pareado em relação ao método DTRS contraposto aos bem estabelecidos índices de validação de agrupamento *Davies Bouldin*, *Dunn Index*, *Silhouette*, *SSE* (*Sum of Squared Error*), *Xie Beni*, tomados um a um, bem como o *baseline* ($k = 20$, Tabela 3.1) implementado neste trabalho.

O teste estatístico foi aplicado individualmente nos N *subranks* retornados pela consulta ($n = \{5, 10, 20, 30, 40, 50\}$), para cada um dos 222 tópicos, adotando o DTRS como condição inicial (*before*) e o respectivo método de *cluster validity*, dentre os selecionados para o estudo, além do *baseline*, como condição final da experimentação para a obtenção do ranking de significância estatística. A escolha do teste de *Wilcoxon* se deu após observações em relação à normalidade dos dados resultantes através do teste estatístico de *Shapiro Wilk* (Seção B.0.1).

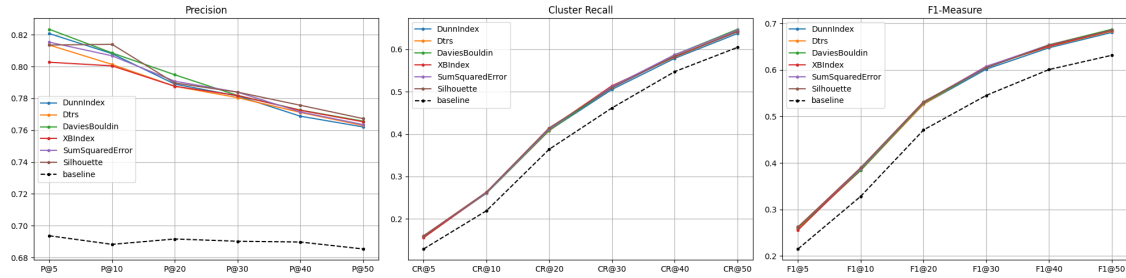
Os resultados do teste foram separados por cores, destacando valores estatisticamente significativos em **vermelho** para diferenças negativas, ou seja, quando o DTRS foi inferior ao método avaliado, e em **azul** para diferenças positivas, quando o método DTRS foi superior àquele estudado. Demais valores não destacados indicam equivalência estatística.

As Figuras de 4.3 a 4.17 trazem a representação visual dos resultados, previamente dispostos em forma tabular, exibindo as métricas de desempenho por faixa de ranking, distintamente, para cada método de validação, bem como o traçado dos valores obtidos para o *baseline*, em valores médios, com o número de grupos ideal selecionado pela forma natural de cada índice de validação de agrupamento (por máximos ou mínimos, sem uso de métodos auxiliares).

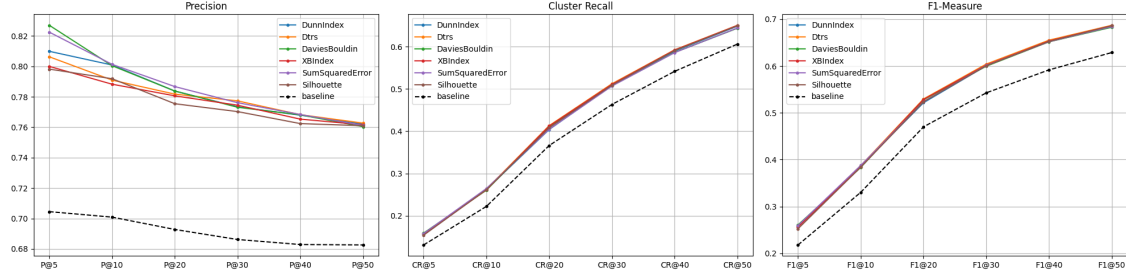
n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7892	0.7054	0.8009	0.8153	0.7946	0.8108	0.7982	0.1533	0.1325	0.1601	0.1574	0.1582	0.1577	0.1564	0.2523	0.2209	0.2602	0.2592	0.2589	0.2594	0.2591
10	0.7950	0.7014	0.7977	0.8072	0.7937	0.8027	0.7973	0.2553	0.2268	0.2643	0.2611	0.2595	0.2611	0.2620	0.3793	0.3383	0.3878	0.3849	0.3836	0.3868	0.3872
20	0.7858	0.6946	0.7876	0.7894	0.7874	0.7874	0.7881	0.4062	0.3691	0.4107	0.4078	0.4070	0.4094	0.4103	0.5243	0.4747	0.5283	0.5265	0.5256	0.5273	0.5278
30	0.7802	0.6880	0.7824	0.7818	0.7832	0.7805	0.7823	0.5080	0.4643	0.5058	0.5058	0.5082	0.5078	0.5061	0.6023	0.5444	0.6015	0.6017	0.6036	0.6027	0.6008
40	0.7729	0.6867	0.7730	0.7721	0.7751	0.7723	0.7720	0.5830	0.5407	0.5805	0.5782	0.5824	0.5824	0.5810	0.6515	0.5933	0.6489	0.6477	0.6523	0.6502	0.6482
50	0.7637	0.6853	0.7653	0.7655	0.7649	0.7625	0.7635	0.6396	0.6002	0.6358	0.6399	0.6398	0.6359	0.6338	0.6820	0.6271	0.6796	0.6824	0.6829	0.6785	0.6774

Tabela 4.2: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - TAMURA (Tamura et al., 1978)Figura 4.3: Performance: descriptor TAMURA. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

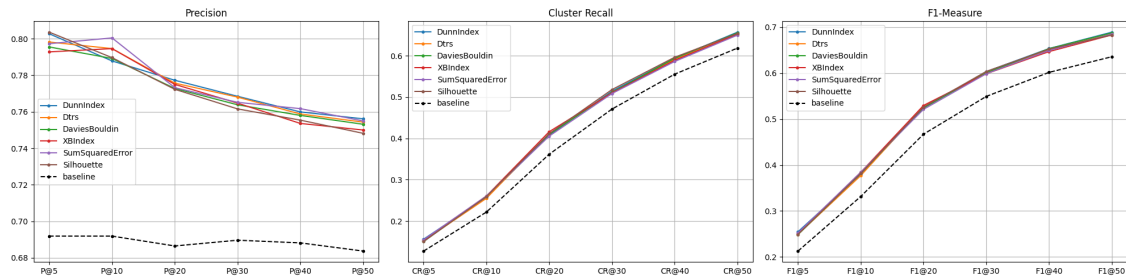
n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.8135	0.6937	0.8207	0.8234	0.8027	0.8153	0.8135	0.1577	0.1290	0.1577	0.1589	0.1556	0.1601	0.1600	0.2590	0.2151	0.2610	0.2609	0.2556	0.2626	0.2623
10	0.8014	0.6883	0.8081	0.8086	0.8005	0.8068	0.8140	0.2607	0.2190	0.2606	0.2604	0.2619	0.2610	0.2633	0.3838	0.3283	0.3882	0.3842	0.3874	0.3873	0.3900
20	0.7876	0.6917	0.7890	0.7948	0.7876	0.7908	0.7896	0.4081	0.3642	0.4088	0.4089	0.4124	0.4113	0.4147	0.5261	0.4708	0.5263	0.5283	0.5299	0.5303	0.5316
30	0.7805	0.6902	0.7815	0.7818	0.7818	0.7839	0.7839	0.5131	0.4618	0.5054	0.5098	0.5141	0.5121	0.5091	0.6067	0.5452	0.6014	0.6051	0.6075	0.6073	0.6043
40	0.7712	0.6898	0.7688	0.7726	0.7725	0.7715	0.7757	0.5818	0.5475	0.5788	0.5872	0.5825	0.5872	0.5838	0.6500	0.6007	0.6474	0.6537	0.6507	0.6537	0.6532
50	0.7630	0.6854	0.7621	0.7657	0.7653	0.7635	0.7673	0.6416	0.6046	0.6374	0.6475	0.6425	0.6450	0.6415	0.6830	0.6313	0.6804	0.6872	0.6843	0.6850	0.6854

Tabela 4.3: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - SCH (Cinque et al., 2001)Figura 4.4: Performance: descriptor SCH. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

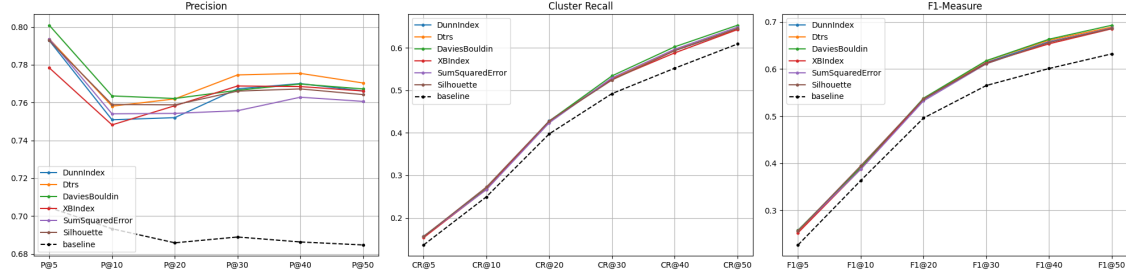
n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.8063	0.7045	0.8099	0.8270	0.8000	0.8225	0.7982	0.1570	0.1307	0.1564	0.1588	0.1564	0.1578	0.1539	0.2574	0.2180	0.2567	0.2608	0.2565	0.2594	0.2526
10	0.7910	0.7009	0.8009	0.8005	0.7883	0.8014	0.7919	0.2628	0.2220	0.2621	0.2603	0.2612	0.2643	0.2617	0.3874	0.3304	0.3867	0.3836	0.3853	0.3878	0.3842
20	0.7818	0.6928	0.7838	0.7838	0.7806	0.7867	0.7755	0.4132	0.3661	0.4040	0.4111	0.4122	0.4046	0.4081	0.5297	0.4701	0.5217	0.5280	0.5286	0.5239	0.5238
30	0.7773	0.6862	0.7734	0.7731	0.7745	0.7761	0.7703	0.5124	0.4634	0.5088	0.5077	0.5108	0.5069	0.5091	0.6042	0.5427	0.6005	0.5996	0.6021	0.6011	0.6005
40	0.7682	0.6829	0.7679	0.7681	0.7653	0.7684	0.7624	0.5928	0.5418	0.5898	0.5867	0.5923	0.5862	0.5911	0.6552	0.5918	0.6541	0.6517	0.6535	0.6518	0.6524
50	0.7627	0.6826	0.7622	0.7604	0.7616	0.7612	0.7610	0.6509	0.6064	0.6489	0.6439	0.6505	0.6444	0.6498	0.6869	0.6288	0.6861	0.6827	0.6860	0.6842	0.6860

Tabela 4.4: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - PHOG (Biasotti et al., 2013)Figura 4.5: Performance: descritor PHOG. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n = 30$ da métrica *Precision*.

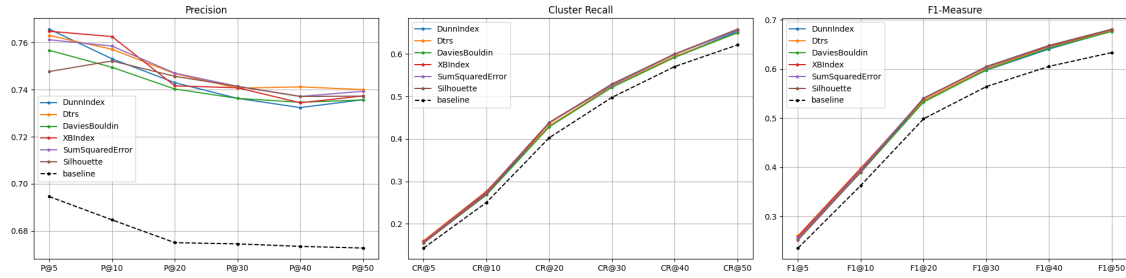
n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7982	0.6919	0.8027	0.7955	0.7928	0.7973	0.8036	0.1519	0.1272	0.1561	0.1504	0.1533	0.1546	0.1503	0.2490	0.2127	0.2548	0.2487	0.2505	0.2523	0.2491
10	0.7946	0.6919	0.7878	0.7892	0.7946	0.8005	0.7896	0.2549	0.2216	0.2574	0.2597	0.2604	0.2606	0.2592	0.3771	0.3316	0.3798	0.3818	0.3842	0.3844	0.3818
20	0.7757	0.6865	0.7773	0.7727	0.7750	0.7732	0.7723	0.4112	0.3616	0.4075	0.4095	0.4162	0.4055	0.4115	0.5267	0.4672	0.5240	0.5245	0.5298	0.5212	0.5261
30	0.7680	0.6896	0.7685	0.7638	0.7647	0.7652	0.7616	0.5113	0.4710	0.5143	0.5112	0.5101	0.5086	0.5182	0.6014	0.5494	0.6034	0.6006	0.5987	0.5988	0.6037
40	0.7588	0.6882	0.7599	0.7580	0.7536	0.7617	0.7554	0.5913	0.5556	0.5938	0.5940	0.5891	0.5863	0.5966	0.6508	0.6017	0.6535	0.6534	0.6469	0.6494	0.6523
50	0.7543	0.6837	0.7561	0.7532	0.7500	0.7550	0.7482	0.6525	0.6188	0.6573	0.6557	0.6524	0.6502	0.6540	0.6853	0.6357	0.6892	0.6877	0.6832	0.6849	0.6831

Tabela 4.5: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - LUM (Lux e Chatzichristofis, 2008)Figura 4.6: Performance: descritor LUM. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7937	0.7045	0.7928	0.8009	0.7784	0.7937	0.7928	0.1556	0.1364	0.1560	0.1568	0.1536	0.1567	0.1561	0.2556	0.2264	0.2560	0.2579	0.2522	0.2570	0.2564
10	0.7581	0.6932	0.7509	0.7635	0.7482	0.7541	0.7590	0.2671	0.2497	0.2681	0.2689	0.2677	0.2665	0.2726	0.3882	0.3634	0.3885	0.3910	0.3877	0.3872	0.3946
20	0.7619	0.6858	0.7520	0.7622	0.7583	0.7543	0.7588	0.4262	0.3976	0.4280	0.4277	0.4274	0.4236	0.4278	0.5371	0.4960	0.5353	0.5382	0.5361	0.5324	0.5368
30	0.7746	0.6889	0.7673	0.7665	0.7688	0.7557	0.7661	0.5281	0.4923	0.5274	0.5340	0.5259	0.5293	0.5243	0.6172	0.5650	0.6140	0.6178	0.6130	0.6112	0.6115
40	0.7755	0.6863	0.7700	0.7698	0.7685	0.7628	0.7672	0.5944	0.5521	0.5935	0.6024	0.5881	0.5966	0.5933	0.6613	0.6012	0.6583	0.6634	0.6536	0.6572	0.6564
50	0.7704	0.6847	0.7661	0.7672	0.7661	0.7606	0.7642	0.6443	0.6090	0.6437	0.6530	0.6429	0.6486	0.6457	0.6891	0.6323	0.6863	0.6930	0.6855	0.6873	0.6862

Tabela 4.6: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - JCD (Zagoris et al., 2010)Figura 4.7: Performance: descritor JCD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n \geq 30$ da métrica *Precision*.

n	Precision						Cluster-Recall (CR)						F1-Measure (F1)								
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7631	0.6946	0.7658	0.7568	0.7649	0.7613	0.7477	0.1603	0.1428	0.1557	0.1551	0.1584	0.1565	0.1542	0.2606	0.2352	0.2564	0.2550	0.2583	0.2555	0.2515
10	0.7572	0.6847	0.7532	0.7495	0.7626	0.7586	0.7523	0.2770	0.2502	0.2716	0.2680	0.2753	0.2728	0.2702	0.3982	0.3628	0.3937	0.3894	0.3969	0.3943	0.3902
20	0.7468	0.6750	0.7430	0.7403	0.7417	0.7471	0.7457	0.4306	0.4026	0.4305	0.4282	0.4388	0.4378	0.4372	0.5359	0.4985	0.5349	0.5326	0.5397	0.5410	0.5399
30	0.7407	0.6745	0.7363	0.7363	0.7408	0.7414	0.7414	0.5238	0.4979	0.5218	0.5216	0.5279	0.5263	0.5295	0.6004	0.5641	0.5975	0.5983	0.6037	0.6033	0.6051
40	0.7412	0.6734	0.7324	0.7347	0.7345	0.7372	0.7372	0.5947	0.5704	0.5928	0.5921	0.5999	0.5992	0.6000	0.6459	0.6053	0.6407	0.6429	0.6463	0.6476	0.6480
50	0.7401	0.6727	0.7358	0.7357	0.7374	0.7394	0.7373	0.6497	0.6214	0.6536	0.6501	0.6572	0.6560	0.6586	0.6775	0.6335	0.6773	0.6764	0.6801	0.6806	0.6811

Tabela 4.7: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - JACCARD (Lewis et al., 2006)Figura 4.8: Performance: descritor JACCARD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n \geq 40$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.8117	0.7081	0.8108	0.7919	0.8018	0.8027	0.7964	0.1593	0.1352	0.1546	0.1533	0.1521	0.1560	0.1541	0.2599	0.2229	0.2526	0.2500	0.2504	0.2542	0.2515
10	0.8027	0.7018	0.8068	0.8086	0.8095	0.8086	0.8005	0.2631	0.2242	0.2604	0.2595	0.2572	0.2623	0.2560	0.3875	0.3333	0.3844	0.3837	0.3827	0.3873	0.3793
20	0.7905	0.6957	0.7939	0.7955	0.7939	0.7955	0.7966	0.3997	0.3632	0.4021	0.4000	0.4009	0.4061	0.4051	0.5208	0.4685	0.5222	0.5208	0.5214	0.5265	0.5258
30	0.7826	0.6892	0.7832	0.7827	0.7859	0.7856	0.7874	0.5039	0.4535	0.4990	0.4968	0.5002	0.5068	0.5048	0.6002	0.5360	0.5973	0.5949	0.5986	0.6034	0.6027
40	0.7730	0.6842	0.7727	0.7748	0.7740	0.7738	0.7742	0.5815	0.5337	0.5780	0.5735	0.5777	0.5847	0.5840	0.6496	0.5867	0.6480	0.6449	0.6476	0.6527	0.6512
50	0.7649	0.6822	0.7647	0.7651	0.7640	0.7623	0.7668	0.6403	0.5920	0.6381	0.6322	0.6342	0.6432	0.6420	0.6822	0.6190	0.6804	0.6769	0.6780	0.6834	0.6831

Tabela 4.8: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - GABOR (Manjunath e Ma, 1996)

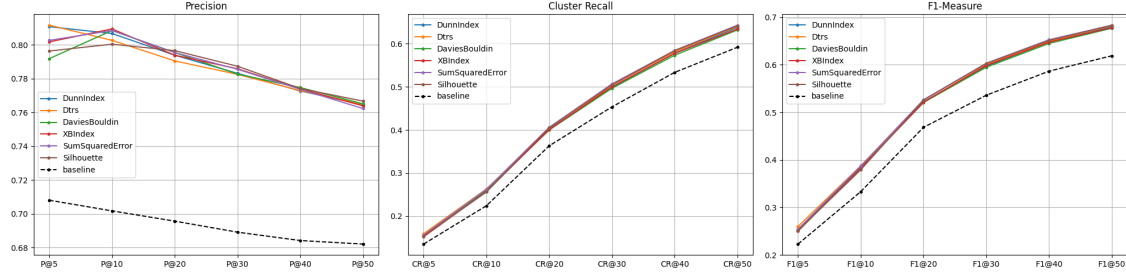


Figura 4.9: Performance: descritor GABOR. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n = 5$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7865	0.7009	0.7910	0.7982	0.7802	0.7991	0.7901	0.1568	0.1355	0.1583	0.1582	0.1567	0.1597	0.1584	0.2567	0.2251	0.2589	0.2595	0.2565	0.2615	0.2593
10	0.7486	0.6919	0.7486	0.7644	0.7450	0.7577	0.7500	0.2643	0.2485	0.2662	0.2703	0.2684	0.2709	0.2711	0.3837	0.3621	0.3859	0.3927	0.3880	0.3924	0.3918
20	0.7588	0.6872	0.7547	0.7658	0.7538	0.7568	0.7563	0.4254	0.4006	0.4261	0.4305	0.4266	0.4256	0.4264	0.5347	0.4989	0.5346	0.5413	0.5350	0.5344	0.5350
30	0.7695	0.6892	0.7668	0.7689	0.7661	0.7590	0.7634	0.5254	0.4944	0.5284	0.5353	0.5266	0.5316	0.5252	0.6132	0.5664	0.6148	0.6197	0.6126	0.6138	0.6104
40	0.7709	0.6870	0.7696	0.7713	0.7662	0.7655	0.7640	0.5902	0.5557	0.5900	0.5997	0.5879	0.5979	0.5910	0.6562	0.6032	0.6557	0.6623	0.6528	0.6589	0.6530
50	0.7662	0.6846	0.7662	0.7682	0.7641	0.7630	0.7621	0.6431	0.6112	0.6443	0.6529	0.6425	0.6518	0.6444	0.6863	0.6331	0.6868	0.6930	0.6846	0.6903	0.6846

Tabela 4.9: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - FCTH (Chatzichristofis e Boutalis, 2008b)

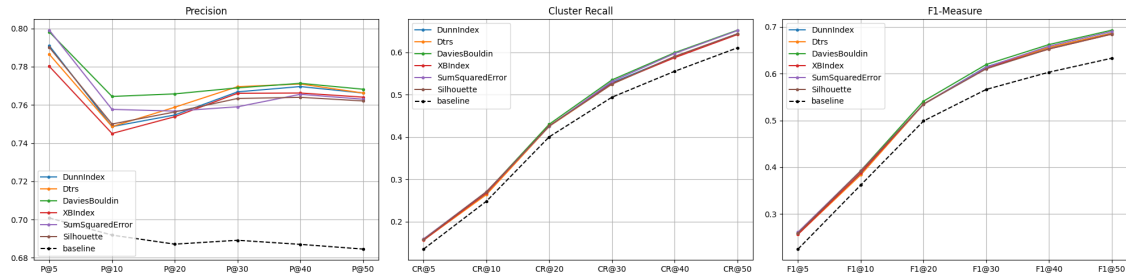


Figura 4.10: Performance: descritor FCTH. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n = 30$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7541	0.7009	0.7577	0.7595	0.7649	0.7595	0.7486	0.1524	0.1415	0.1551	0.1537	0.1563	0.1568	0.1539	0.2491	0.2339	0.2526	0.2511	0.2548	0.2554	0.2505
10	0.7685	0.6761	0.7595	0.7563	0.7613	0.7644	0.7586	0.2803	0.2486	0.2768	0.2750	0.2795	0.2774	0.2765	0.4034	0.3601	0.3985	0.3963	0.4016	0.3997	0.3976
20	0.7448	0.6764	0.7428	0.7376	0.7390	0.7439	0.7441	0.4301	0.3982	0.4332	0.4237	0.4362	0.4325	0.4332	0.5333	0.4947	0.5358	0.5279	0.5368	0.5353	0.5357
30	0.7408	0.6764	0.7404	0.7351	0.7422	0.7417	0.7402	0.5195	0.4936	0.5219	0.5227	0.5252	0.5230	0.5261	0.5984	0.5625	0.5994	0.5981	0.6027	0.6015	0.6028
40	0.7404	0.6764	0.7359	0.7339	0.7367	0.7369	0.7346	0.5899	0.5650	0.5962	0.5959	0.5983	0.5968	0.5924	0.6432	0.6058	0.6454	0.6447	0.6468	0.6466	0.6431
50	0.7394	0.6732	0.7371	0.7328	0.7376	0.7383	0.7350	0.6473	0.6162	0.6517	0.6537	0.6538	0.6547	0.6506	0.6757	0.6325	0.6773	0.6769	0.6787	0.6799	0.6757

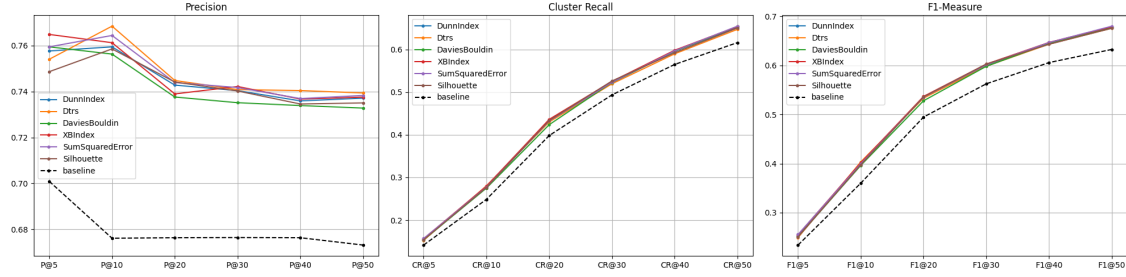
Tabela 4.10: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - DICE (Dice, 1945)

Figura 4.11: Performance: descritor DICE. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS nas faixas de ranking $n = 10$, $n = 20$ e $n = 40$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7829	0.6838	0.7919	0.7838	0.7919	0.7811	0.7829	0.1565	0.1369	0.1555	0.1532	0.1573	0.1556	0.1569	0.2559	0.2262	0.2572	0.2517	0.2596	0.2551	0.2569
10	0.7802	0.6775	0.7730	0.7725	0.7775	0.7739	0.7779	0.2745	0.2396	0.2726	0.2726	0.2760	0.2741	0.2736	0.3996	0.3495	0.3962	0.3962	0.4004	0.3977	0.3981
20	0.7651	0.6813	0.7586	0.7590	0.7651	0.7622	0.7581	0.4322	0.3923	0.4295	0.4339	0.4336	0.4352	0.4246	0.5422	0.4915	0.5384	0.5421	0.5436	0.5427	0.5337
30	0.7593	0.6827	0.7541	0.7578	0.7610	0.7530	0.7560	0.5330	0.4890	0.5329	0.5353	0.5335	0.5358	0.5293	0.6155	0.5610	0.6131	0.6157	0.6165	0.6141	0.6116
40	0.7538	0.6843	0.7516	0.7566	0.7570	0.7542	0.7519	0.6044	0.5625	0.6025	0.6074	0.6045	0.6106	0.5983	0.6594	0.6072	0.6566	0.6616	0.6605	0.6617	0.6544
50	0.7516	0.6861	0.7495	0.7517	0.7513	0.7495	0.7494	0.6565	0.6146	0.6584	0.6591	0.6565	0.6593	0.6501	0.6879	0.6369	0.6883	0.6897	0.6883	0.6882	0.6835

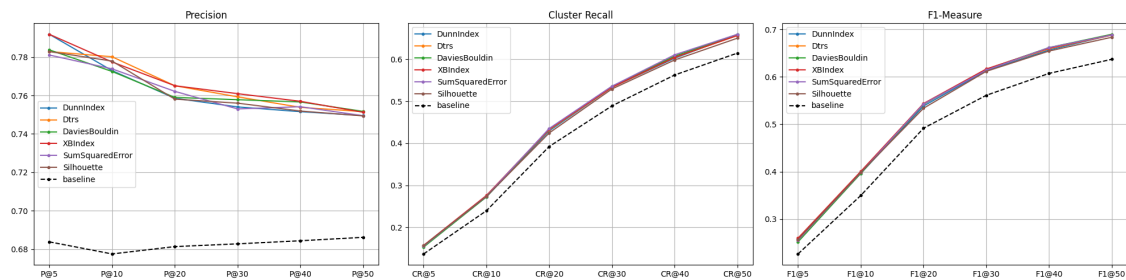
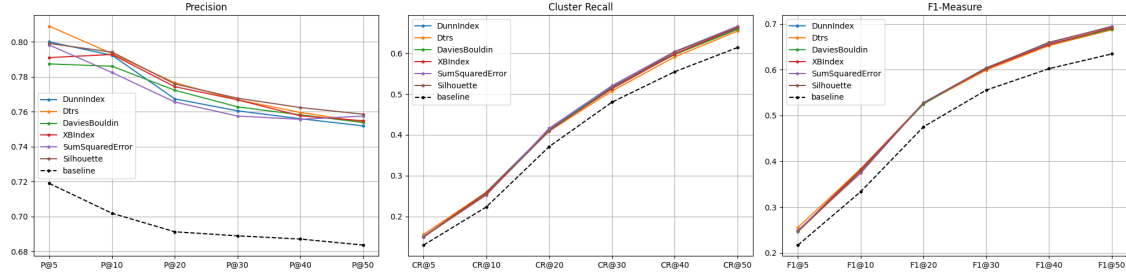
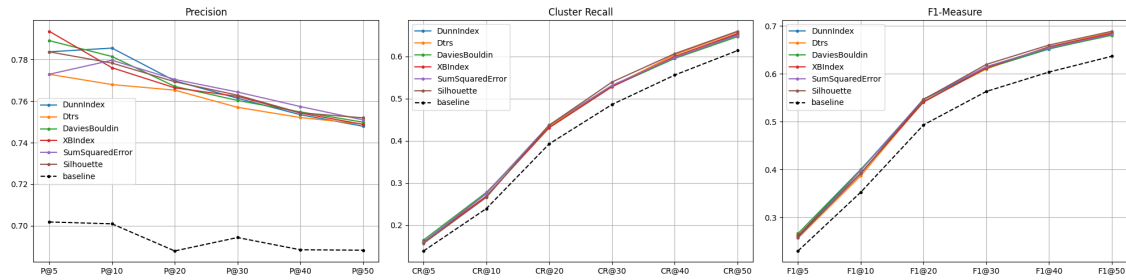
Tabela 4.11: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - COSINE

Figura 4.12: Performance: descritor COSINE. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n = 10$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0,8090	0,7189	0,8000	0,7874	0,7910	0,7982	0,7991	0,1560	0,1298	0,1495	0,1505	0,1511	0,1515	0,1495	0,2559	0,2169	0,2469	0,2474	0,2481	0,2493	0,2464
10	0,7932	0,7018	0,7923	0,7860	0,7928	0,7824	0,7941	0,2596	0,2233	0,2599	0,2586	0,2580	0,2524	0,2541	0,3834	0,3338	0,3835	0,3817	0,3819	0,3741	0,3776
20	0,7766	0,6912	0,7673	0,7723	0,7743	0,7655	0,7759	0,4090	0,3712	0,4128	0,4110	0,4114	0,4163	0,4097	0,5262	0,4754	0,5259	0,5250	0,5275	0,5282	0,5267
30	0,7668	0,6889	0,7605	0,7628	0,7668	0,7575	0,7677	0,5076	0,4801	0,5180	0,5171	0,5130	0,5207	0,5158	0,5993	0,5555	0,6040	0,6039	0,6018	0,6041	0,6037
40	0,7596	0,6870	0,7560	0,7582	0,7577	0,7556	0,7624	0,5911	0,5549	0,5987	0,5969	0,5983	0,6044	0,6032	0,6526	0,6022	0,6554	0,6547	0,6550	0,6581	0,6598
50	0,7541	0,6836	0,7519	0,7538	0,7548	0,7575	0,7586	0,6545	0,6141	0,6613	0,6586	0,6627	0,6666	0,6640	0,6870	0,6345	0,6899	0,6888	0,6913	0,6950	0,6938

Tabela 4.12: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CNN_AD (Ionescu et al., 2014)Figura 4.13: Performance: descritor CNN_AD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS na faixa de ranking $n = 5$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0,7730	0,7018	0,7838	0,7892	0,7937	0,7730	0,7838	0,1591	0,1390	0,1604	0,1651	0,1605	0,1588	0,1569	0,2569	0,2299	0,2615	0,2664	0,2618	0,2568	0,2586
10	0,7680	0,7009	0,7856	0,7815	0,7761	0,7797	0,7784	0,2676	0,2395	0,2750	0,2778	0,2691	0,2759	0,2663	0,3871	0,3526	0,3996	0,4003	0,3925	0,3983	0,3917
20	0,7653	0,6878	0,7698	0,7671	0,7664	0,7705	0,7694	0,4351	0,3928	0,4369	0,4314	0,4312	0,4376	0,4373	0,5428	0,4933	0,5465	0,5415	0,5410	0,5473	0,5464
30	0,7571	0,6943	0,7614	0,7604	0,7623	0,7644	0,7629	0,5292	0,4862	0,5289	0,5292	0,5282	0,5317	0,5395	0,6102	0,5632	0,6117	0,6116	0,6126	0,6151	0,6198
40	0,7521	0,6884	0,7535	0,7548	0,7543	0,7574	0,7544	0,6044	0,5561	0,5961	0,5956	0,6002	0,5975	0,6069	0,6569	0,6038	0,6520	0,6527	0,6563	0,6547	0,6603
50	0,7490	0,6882	0,7479	0,7499	0,7488	0,7510	0,7520	0,6585	0,6142	0,6520	0,6471	0,6545	0,6502	0,6600	0,6865	0,6367	0,6828	0,6806	0,6848	0,6825	0,6890

Tabela 4.13: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CN3X3 (Ionescu et al., 2014)Figura 4.14: Performance: descritor CN3x3. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7991	0.7081	0.7964	0.8009	0.7865	0.7964	0.7919	0.1569	0.1360	0.1585	0.1587	0.1559	0.1539	0.1589	0.2574	0.2260	0.2594	0.2601	0.2560	0.2536	0.2602
10	0.7595	0.6919	0.7509	0.7604	0.7491	0.7581	0.7532	0.2658	0.2476	0.2692	0.2670	0.2667	0.2664	0.2730	0.3868	0.3609	0.3892	0.3881	0.3868	0.3881	0.3941
20	0.7635	0.6860	0.7595	0.7624	0.7568	0.7574	0.7559	0.4252	0.3987	0.4301	0.4284	0.4263	0.4273	0.4316	0.5365	0.4969	0.5388	0.5388	0.5354	0.5367	0.5382
30	0.7737	0.6895	0.7721	0.7683	0.7698	0.7616	0.7664	0.5284	0.4951	0.5330	0.5324	0.5258	0.5361	0.5267	0.6170	0.5674	0.6195	0.6178	0.6135	0.6183	0.6127
40	0.7751	0.6866	0.7739	0.7720	0.7680	0.7651	0.7671	0.5925	0.5543	0.5983	0.6017	0.5888	0.5980	0.5954	0.6596	0.6027	0.6623	0.6639	0.6543	0.6591	0.6574
50	0.7709	0.6847	0.7686	0.7686	0.7682	0.7645	0.7656	0.6451	0.6101	0.6459	0.6541	0.6419	0.6520	0.6488	0.6899	0.6332	0.6887	0.6940	0.6860	0.6913	0.6885

Tabela 4.14: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - CEDD (Chatzichristofis e Boutalis, 2008a)

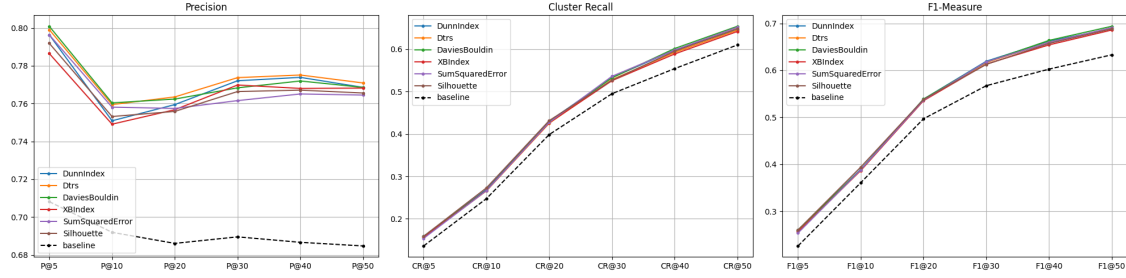


Figura 4.15: Performance: descritor CEDD. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*. Destaque para o método DTRS entre as faixas de ranking $n \geq 20$ da métrica *Precision*.

n	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.7685	0.7063	0.7712	0.7730	0.7631	0.7604	0.7622	0.1601	0.1468	0.1577	0.1602	0.1560	0.1572	0.1565	0.2607	0.2412	0.2578	0.2609	0.2545	0.2563	0.2553
10	0.7586	0.6869	0.7644	0.7595	0.7658	0.7563	0.7604	0.2731	0.2517	0.2776	0.2726	0.2734	0.2734	0.2708	0.3945	0.3649	0.4005	0.3947	0.3957	0.3946	0.3920
20	0.7428	0.6768	0.7450	0.7444	0.7455	0.7450	0.7462	0.4310	0.4074	0.4349	0.4306	0.4364	0.4337	0.4311	0.5341	0.5005	0.5380	0.5344	0.5389	0.5360	0.5337
30	0.7380	0.6794	0.7396	0.7408	0.7425	0.7377	0.7402	0.5234	0.5053	0.5244	0.5230	0.5305	0.5276	0.5235	0.6004	0.5686	0.6015	0.6011	0.6063	0.6020	0.5998
40	0.7319	0.6762	0.7347	0.7347	0.7350	0.7346	0.7373	0.5961	0.5764	0.5976	0.5954	0.6020	0.5956	0.5982	0.6436	0.6097	0.6454	0.6445	0.6479	0.6436	0.6460
50	0.7326	0.6757	0.7350	0.7329	0.7343	0.7355	0.7360	0.6508	0.6276	0.6521	0.6525	0.6575	0.6551	0.6547	0.6754	0.6380	0.6767	0.6764	0.6791	0.6782	0.6781

Tabela 4.15: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - BM25 (Baeza-Yates et al., 1999)

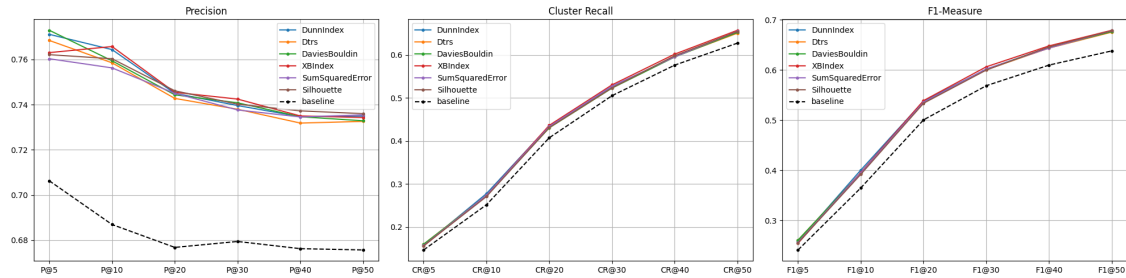
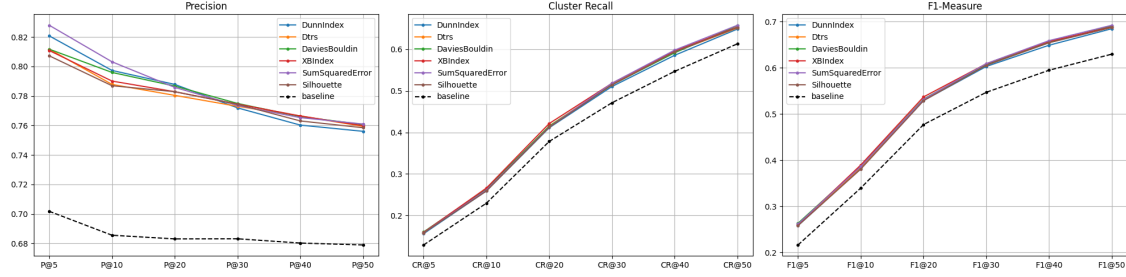


Figura 4.16: Performance: descritor BM25. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

n	DTRS	Precision						Cluster-Recall (CR)						F1-Measure (F1)							
		K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
5	0.8117	0.7018	0.8207	0.8117	0.8108	0.8279	0.8072	0.1587	0.1293	0.1564	0.1606	0.1595	0.1586	0.1583	0.2589	0.2161	0.2579	0.2631	0.2602	0.2614	0.2581
10	0.7878	0.6856	0.7973	0.7959	0.7901	0.8032	0.7869	0.2613	0.2297	0.2591	0.2639	0.2661	0.2614	0.2590	0.3842	0.3396	0.3822	0.3868	0.3895	0.3862	0.3806
20	0.7804	0.6831	0.7878	0.7867	0.7829	0.7858	0.7829	0.4153	0.3782	0.4111	0.4148	0.4213	0.4128	0.4124	0.5316	0.4771	0.5293	0.5325	0.5371	0.5316	0.5289
30	0.7730	0.6832	0.7719	0.7749	0.7745	0.7736	0.7743	0.5154	0.4713	0.5103	0.5144	0.5180	0.5183	0.5139	0.6067	0.5471	0.6032	0.6069	0.6090	0.6093	0.6056
40	0.7664	0.6803	0.7602	0.7660	0.7664	0.7653	0.7631	0.5934	0.5408	0.5855	0.5915	0.5961	0.5979	0.5937	0.6500	0.5950	0.6492	0.6552	0.6575	0.6591	0.6545
50	0.7594	0.6790	0.7560	0.7606	0.7598	0.7610	0.7585	0.6511	0.6131	0.6486	0.6569	0.6533	0.6576	0.6522	0.6877	0.6298	0.6850	0.6910	0.6888	0.6924	0.6878

Tabela 4.16: Experimento 01: resultados K-Medoids - $k = [15, 25]$ - ACC (Penatti et al., 2012)Figura 4.17: Performance: descritor ACC. Algoritmo K-Medoids, $k = [15, 25]$. Método natural do *cluster validity*.

4.1.2 Avaliação dos ganhos alcançados

Conforme evidenciado pelos gráficos nas Figuras 4.3 a 4.17, a abordagem utilizada neste estudo, detalhada na Seção 3.1, demonstrou superioridade estatística em relação ao *baseline* ($k = 20$) em termos de relevância (*Precision*) e diversidade (*Cluster Recall*) para todos os índices de validação de agrupamento e descritores estudados (ganho médio relativo aproximado de 9,7%). Essa superioridade foi corroborada pela média harmônica (*F1-Measure*) em todos os casos de estudo.

De acordo com as representações gráficas das Figuras B.1(a) e B.1(b) do Apêndice B, observamos que o método DTRS se manteve consistentemente próximo à média dos resultados, destacando-se, no entanto, quanto aos descritores PHOG, JCD, JACCARD, COSINE, GABOR, FCTH, DICE, CNN_AD e CEDD, sobre os quais o DTRS alcançou superioridade em termos de valores máximos. A Tabela 4.17 apresenta um resumo desses resultados por faixa de ranking e métrica de performance.

	Descritor	Precision	Cluster Recall	F1-Measure
visual	PHOG	$n = 30, n = 50$	$n \geq 20$	$n \geq 20$
	JCD	$n \geq 30$	-	-
	GABOR	$n = 5$	$n = 5$	-
	FCTH	$n = 5$	-	-
	CNN_AD	$n = 5, n = 20$	$n = 5$	$n = 5$
	CEDD	$n \geq 20$	-	-
textual	COSINE	$n = 10$	-	-
	DICE	$n = 10$	-	-
	JACCARD	$n = 40$	-	-

Tabela 4.17: Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos gráficos 4.3 a 4.17.

Já através da observação dos resultados destacados com a aplicação do teste de significância estatística (*Wilcoxon Signed Rank test*), Tabelas 4.2 à 4.16, temos que o método DTRS se mostrou superior ao *baseline* em todos os descritores com ganho relativo médio calculado em 9,7% (*Precision*: pior caso 8,3%; melhor caso 12,2%. *Cluster Recall*: pior caso 5,3%; melhor caso 11,4%).

Nas Tabelas 4.18 e 4.20 temos os resultados em ganho relativo para as métricas *Precision*, *Cluster Recall* e *F1-Measure*, respectivamente, assim como nas Figuras 4.18, 4.19 e 4.20 a representação gráfica desses resultados.

	Descritor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
visual	TAMURA	-0,424	-0,9187	-0,2562	-0,6054	-0,3042
	SCH	-0,2658	-0,6196	0,1342	-0,3075	-0,5652
	PHOG	-0,2202	-0,5133	0,3614	-0,5922	0,604
	LUM	-0,0585	0,3689	0,4108	-0,0655	0,4177
	JCD	0,7674	0,0938	0,9983	1,1649	0,5669
	GABOR	-0,1401	0,1448	-0,0763	-0,063	0,0707
	FCTH	0,0826	-0,7776	0,550	0,0027	0,326
	CNN_AD	0,6749	0,8292	0,4659	0,9148	0,0243
	CN3X3	-0,8013	-0,8219	-0,7919	-0,6833	-0,7832
	CEDD	0,4486	0,2035	0,9436	0,844	0,9096
	ACC	-0,3059	-0,3632	-0,1239	-0,7828	0,1239
	COSINE	0,3215	0,2486	-0,2293	0,4185	0,3671
textual	DICE	0,3269	0,7395	0,1487	0,073	0,5976
	JACCARD	0,5132	0,8029	0,1701	0,0958	0,6142
	BM25	-0,3907	-0,2864	-0,3107	0,057	-0,2275

Tabela 4.18: Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica *Precision*.

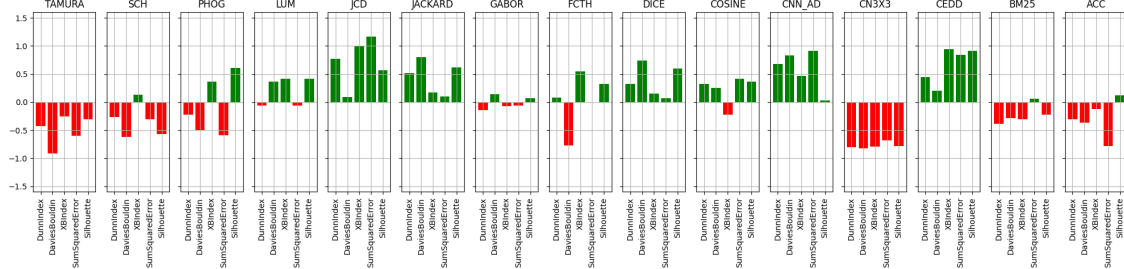


Figura 4.18: Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica *Precision*. Em verde, superioridade do DTRS.

	Descritor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
visual	TAMURA	-0,6132	-0,2826	-0,3584	-0,4625	-0,3851
	SCH	0,5083	-0,2533	-0,4031	-0,4338	-0,4267
	PHOG	0,8129	0,8979	0,261	0,9549	0,5533
	LUM	-0,359	-0,4661	-0,54	0,1925	-0,8343
	JCD	-0,0868	-0,9535	0,2419	-0,0816	-0,3381
	GABOR	0,4723	1,0853	0,8666	-0,5709	0,1117
	FCTH	-0,3258	-1,6689	-0,3135	-1,2618	-0,614
	CNN_AD	-1,0687	-0,7023	-0,688	-1,0855	-0,602
	CN3X3	-0,1228	0,0906	0,3712	-0,3212	-0,5065
	CEDD	-0,8728	-0,9757	0,2012	-0,8233	-0,9686
	ACC	0,9198	-0,2413	-0,9004	-0,3489	0,3343
	COSINE	0,2754	-0,2057	-0,1979	-0,5028	0,9658
textual	DICE	-0,3215	0,1675	-0,9226	-0,498	-0,3081
	JACCARD	0,4207	0,9445	-0,8106	-0,4633	-0,471
	BM25	-0,6374	0,0452	-0,9328	-0,4258	-0,0229

Tabela 4.19: Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica *Cluster Recall*.

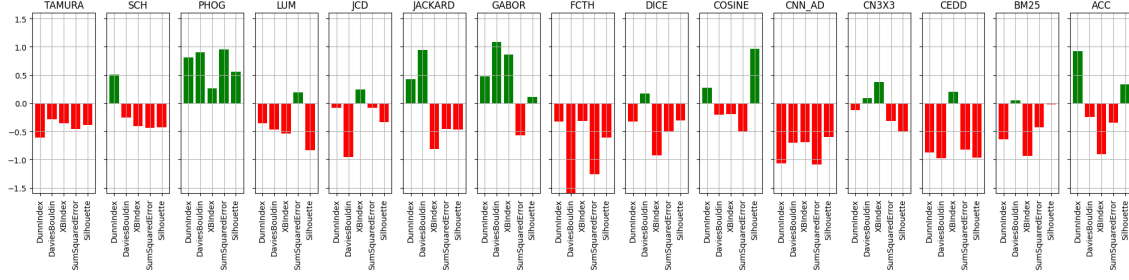


Figura 4.19: Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica *Cluster Recall*. Em verde, superioridade do DTRS.

	Descritor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
visual	TAMURA	-0.852	-0.6506	-0.7269	-0.7682	-0.6472
	SCH	-0.0517	-0.3656	-0.1305	-0.6718	-0.7248
	PHOG	0.4886	0.3235	0.3148	0.2769	0.8435
	LUM	-0.6283	-0.2113	-0.2785	-0.2366	-0.2407
	JCD	0.2515	-0.4641	0.6762	0.409	0.0406
	GABOR	0.7392	1.2284	1.0209	0.0052	0.6313
	FCTH	-0.2717	-1.2552	-0.0339	-0.8494	-0.326
	ACC	0.5655	-0.4673	-0.6068	-0.507	0.3662
	CNN_AD	0.3401	0.4627	0.3125	0.3239	0.4175
	CN3X3	-0.7575	-0.8936	-0.4913	-0.5811	-0.8172
textual	CEDD	-0.408	-0.5274	0.451	0.1303	-0.3314
	COSINE	0.3067	0.3232	-0.3799	0.0867	0.6097
	DICE	-0.2235	0.2744	-0.6898	-0.5935	-0.0397
	JACKARD	0.7149	1.0128	-0.0816	0.1384	0.545
	BM25	-0.29	-0.1024	-0.1573	0.1132	0.3598

Tabela 4.20: Ganho médio relativo: DTRS versus demais métodos, algoritmo K-Medoids, $k = [15, 25]$. Superioridade do DTRS em azul. Métrica *F1-Measure*.

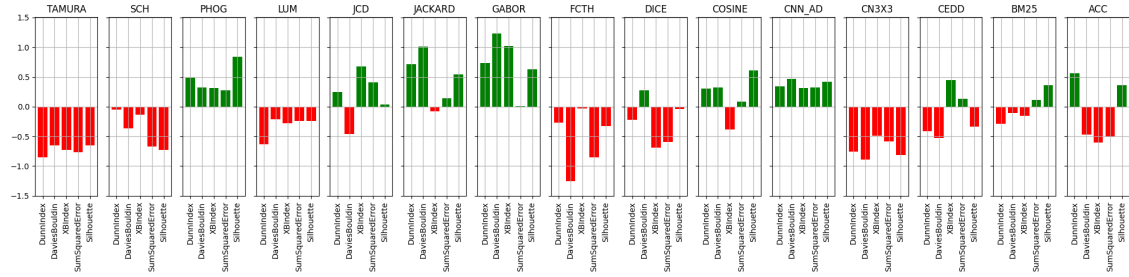


Figura 4.20: Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica *F1-Measure*. Em verde, superioridade do DTRS.

As representações para o ganho médio relativo revelaram que o método DTRS foi consistentemente superior nos casos apontados na Tabela 4.21. Imagens reais deste experimento podem ser vistas no Apêndice A, Subseção A.0.1.

	Descritor	Índices de Validação de Agrupamento
visual	PHOG	Xie Beni, Silhouette
	JCD	Xie Beni
	GABOR	Davies Bouldin, Silhouette
	LUM	Sum of squared error
	ACC	Silhouette
	CEDD	Xie Beni
textual	COSINE	Dunn Index, Silhouette
	DICE	Davies Bouldin
	JACCARD	Dunn Index, Davies Bouldin

Tabela 4.21: Descritores e índices de validação de agrupamento aos quais o DTRS obteve ganho consistente (ganhos positivos simultaneamente nas métricas *Precision* e *Cluster Recall*).

4.2 Avaliação da abordagem dinâmica (experimentos 02 e 03)

Nesta seção, trataremos dos resultados alcançados com o uso do IASMIN, abordado na Seção 3.2.2, cujos objetivos são a otimização do custo computacional, através da redução do tempo de execução, e a melhoria da performance geral do sistema CBIR.

Para o primeiro experimento desta seção (experimento 02, conforme detalhado na Tabela 3.3 e diagrama da Figura 4.21), optamos pelo uso do algoritmo Hierárquico Aglomerativo devido à sua característica de independência na inicialização dos centros de *cluster*, tendo em vista que, diferentemente dos algoritmos K-Means e K-Medoids, o algoritmo Hierárquico Aglomerativo não necessita da inicialização randômica ou de iterações para a redefinição dos seus centros de *cluster*, algo que reduziria, ao menos em tese, o tempo despendido na tarefa de encontrar o melhor esquema de agrupamentos.

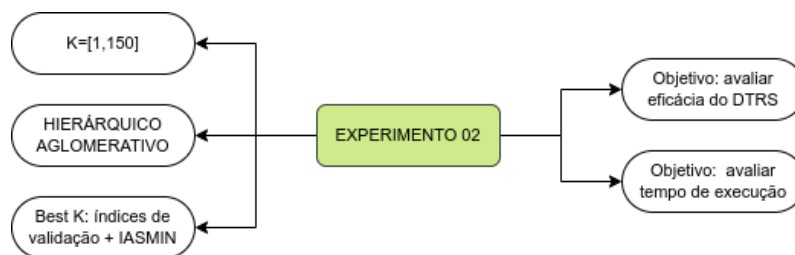


Figura 4.21: Diagrama: objetivo e parâmetros do experimento 02

No que diz respeito ao segundo experimento desta seção (experimento 03, detalhado na Tabela 3.3), utilizamos o algoritmo K-Medoids, com as mesmas condições do experimento anterior, objetivando realizar uma comparação do tempo de execução deste último com aquele obtido, anteriormente, pelo algoritmo Hierárquico Aglomerativo (experimento 02). É importante ressaltar que, para este caso, não foi implementada a interrupção das iterações para o ajuste dos centros de *cluster* do K-Medoids pelo IASMIN, as execuções ocorreram até o limite de iterações do próprio algoritmo e o tempo de execução foi tomado ao final de todo o processo de agrupamento para cada esquema *K clusters* testado. Essa estratégia foi necessária,

uma vez que o algoritmo K-Medoids difere do Hierárquico Aglomerativo no processo de geração dos agrupamentos, utilizando um K fixo. Em ambos os experimentos, decidimos pela ampliação da faixa de *clusters* a serem testados (de $k = [15, 25]$ para $k = [1, 150]$). Esta ampliação na faixa de *clusters* foi necessária para que pudéssemos testar o alcance do IASMIN, bem como para uma melhor adaptação à média móvel, presente no cálculo do IASMIN, por conta do intervalo de 10 *clusters* escolhido.

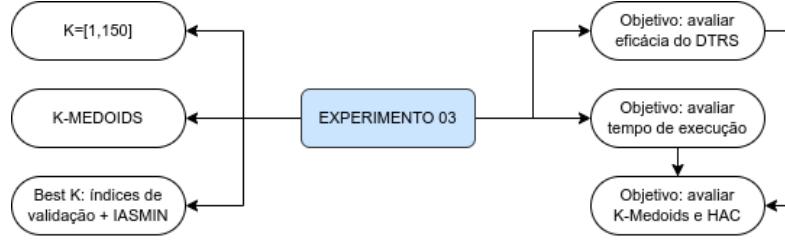


Figura 4.22: Diagrama: objetivo e parâmetros do experimento 03

Para além dos experimentos 02 e 03 mencionados anteriormente, foram realizadas outras três execuções, usando os dois algoritmos (Hierárquico Aglomerativo e K-Medoids), com as seguintes configurações para seleção do número de grupos ideal: 1) sem uso de métodos auxiliares, ou seja, apenas com uso dos índices de validação de agrupamento, pela forma natural de cada índice; 2) com uso dos índices de validação em conjunto com o método do Cotovelo (auxiliar); 3) com uso dos índices de validação em conjunto com o método auxiliar IASMIN. Todos com intervalo $k = [15, 25]$. O objetivo dessas execuções foi comparar a performance das três abordagens em condições idênticas. Os resultados podem ser vistos no Apêndice (Tabelas C.1 a C.15, gráficos da Figura C.1) com uma breve discussão acerca desses resultados na seção 4.2.2.

4.2.1 Avaliação dos resultados de Performance

A partir da observação individualizada para o método DTRS e análise comparativa entre descritores visuais e textuais (Tabela 4.23(a)), de acordo com o segundo objetivo (item b. da Seção 1.2), temos que os melhores resultados, sem uso dos testes de significância estatístico e em termos de valores máximos atingidos, para o algoritmo Hierárquico Aglomerativo (experimento 02) com uso do IASMIN, em relação à métrica *Precision*, concentraram-se entre os descritores visuais CEDD ($P@30$ à $P@50$), FCTH ($P@20$) e JCD ($P@10$ à $P@50$). Em relação aos descritores textuais, o método *DTRS* apresentou valores máximos com COSINE ($P@5$ e $P@10$) e JACCARD ($P@5$). Para a métrica *Cluster Recall*, os descritores visuais ACC ($CR@20$ e $CR@30$), CN3x3 ($CR@20$ à $CR@50$), CNN_AD ($CR@10$), LUM ($CR@5$ e $CR@40$) e SCH ($CR@10$), bem como os descritores textuais COSINE ($CR@5$) e DICE ($CR@50$), apresentaram resultados superiores aos demais *cluster validity*.

Quanto ao experimento com uso do K-Medoids (experimento 03), temos que os melhores resultados (Tabela 4.23(b)) para a métrica *Precision* concentraram-se en-

tre os descritores visuais ACC ($P@50$), CEDD ($P@20$ e $P@30$), CN3x3 ($P@40$), FTCH ($\geq P@10$) e JCD ($P@5$), e o descritor textual JACCARD ($\leq P@10$). Para a métrica *Cluster Recall*, os melhores resultados ficaram entre os descritores visuais ACC ($CR@50$), CEDD ($CR@5$ e $CR@50$) e PHOG ($CR@20$ e $CR@30$), textuais BM25 ($CR@20$), DICE ($CR@10$ e $CR@40$) e JACCARD ($CR@5$, $CR@10$, $CR@30$ e $CR@40$). Contudo, foi observada uma redução significativa nos valores de performance para a métrica *Precision* (relevância) e aumento da diversidade, conforme podemos observar pela métrica *Cluster Recall*, em relação ao experimento anterior.

		PRECISION					CLUSTER RECALL					F1 MEASURE							
		P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
visual	ACC	0.7658	0.7986	0.7840	0.7709	0.7566	0.7416	0.1471	0.2605	0.4064	0.5098	0.5912	0.6565	0.2444	0.3853	0.5247	0.6018	0.6501	0.6821
	CEDD	0.7964	0.8099	0.7883	0.7749	0.7667	0.7553	0.1389	0.2485	0.3875	0.4931	0.5760	0.6409	0.2330	0.3728	0.5094	0.5906	0.6441	0.6789
	CN3x3	0.7604	0.7932	0.7809	0.7664	0.7548	0.7428	0.1424	0.2581	0.4073	0.5135	0.5956	0.6636	0.2366	0.3818	0.5254	0.6033	0.6521	0.6861
	CNN_AD	0.7568	0.7914	0.7795	0.7665	0.7546	0.7441	0.1509	0.2609	0.4011	0.5052	0.5914	0.6556	0.2491	0.3855	0.5198	0.5963	0.6496	0.6828
	FTCH	0.7964	0.8104	0.7912	0.7743	0.7645	0.7529	0.1385	0.2456	0.3892	0.4923	0.5738	0.6357	0.2323	0.3692	0.5116	0.5897	0.6422	0.6750
	GABOR	0.7450	0.7847	0.7804	0.7643	0.7533	0.7428	0.1471	0.2598	0.4039	0.5072	0.5865	0.6478	0.2434	0.3837	0.5225	0.5978	0.6451	0.6772
	JCD	0.7964	0.8149	0.7946	0.7758	0.7666	0.7542	0.1362	0.2484	0.3917	0.4971	0.5783	0.6443	0.2290	0.3730	0.5144	0.5932	0.6462	0.6810
	LUM	0.7631	0.7928	0.7851	0.7712	0.7579	0.7444	0.1555	0.2591	0.4039	0.5075	0.5935	0.6596	0.2555	0.3842	0.5238	0.6002	0.6518	0.6849
	PHOG	0.7595	0.7914	0.7847	0.7712	0.7587	0.7457	0.1496	0.2587	0.4000	0.5022	0.5863	0.6520	0.2474	0.3828	0.5198	0.5969	0.6488	0.6821
	SCH	0.7694	0.7995	0.7824	0.7677	0.7557	0.7424	0.1508	0.2606	0.3983	0.5032	0.5903	0.6582	0.2496	0.3858	0.5165	0.5956	0.6485	0.6831
textual	TAMURA	0.7351	0.7851	0.7824	0.7680	0.7561	0.7439	0.1465	0.2574	0.3994	0.5037	0.5911	0.6543	0.2419	0.3803	0.5186	0.5970	0.6504	0.6826
	BM25	0.7090	0.7653	0.7687	0.7629	0.7508	0.7406	0.1296	0.2449	0.4006	0.5079	0.5864	0.6545	0.2168	0.3641	0.5176	0.5986	0.6453	0.6810
	COSINE	0.8243	0.8117	0.7847	0.7704	0.7557	0.7439	0.1562	0.2551	0.3999	0.5079	0.5910	0.6588	0.2559	0.3793	0.5187	0.5995	0.6492	0.6835
	DICE	0.8000	0.8009	0.7800	0.7700	0.7545	0.7445	0.1475	0.2520	0.4031	0.5084	0.5885	0.6605	0.2411	0.3736	0.5200	0.5989	0.6461	0.6837
	JACCARD	0.8279	0.8090	0.7800	0.7659	0.7533	0.7421	0.1478	0.2530	0.3966	0.5008	0.5836	0.6564	0.2457	0.3777	0.5156	0.5926	0.6433	0.6809
	Max()	0.8279	0.8149	0.7946	0.7758	0.7667	0.7553	0.1562	0.2609	0.4073	0.5135	0.5956	0.6636	0.2559	0.3858	0.5254	0.6033	0.6521	0.6861
	Min()	0.7090	0.7653	0.7687	0.7629	0.7508	0.7406	0.1296	0.2449	0.3875	0.4923	0.5738	0.6357	0.2168	0.3641	0.5094	0.5897	0.6422	0.6750
		AVG	0.7710	Median	0.7679	Q 75%	0.7850	AVG	0.4240	Median	0.4498	Q 75%	0.5880	AVG	0.5108	Median	0.5576	Q 75%	0.6480

(a) Resultados do experimento 02 (Hierárquico Aglomerativo) para o DTRS.

		PRECISION					CLUSTER RECALL					F1 MEASURE							
		P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
visual	ACC	0.7423	0.7239	0.7287	0.7287	0.7275	0.7310	0.1481	0.2584	0.4148	0.5298	0.6104	0.6731	0.2439	0.3754	0.5173	0.6006	0.6504	0.6870
	CEDD	0.7550	0.7356	0.7321	0.7296	0.7286	0.7259	0.1520	0.2580	0.4199	0.5295	0.6112	0.6729	0.2504	0.3768	0.5246	0.6009	0.6507	0.6837
	CN3x3	0.7423	0.7342	0.7291	0.7288	0.7288	0.7281	0.1456	0.2538	0.4122	0.5204	0.5996	0.6564	0.2388	0.3707	0.5178	0.5975	0.6460	0.6778
	CNN_AD	0.7216	0.7198	0.7259	0.7264	0.7256	0.7262	0.1432	0.2507	0.4148	0.5213	0.6056	0.6677	0.2367	0.3677	0.5203	0.5968	0.6479	0.6824
	FTCH	0.7559	0.7432	0.7372	0.7336	0.7301	0.7303	0.1509	0.2606	0.4183	0.5291	0.6043	0.6631	0.2488	0.3799	0.5255	0.6049	0.6502	0.6836
	GABOR	0.7459	0.7320	0.7243	0.7254	0.7256	0.7268	0.1485	0.2599	0.4175	0.5329	0.6082	0.6693	0.2454	0.3793	0.5223	0.6029	0.6490	0.6830
	JCD	0.7586	0.7284	0.7270	0.7282	0.7287	0.7257	0.1516	0.2593	0.4177	0.5299	0.6081	0.6660	0.2488	0.3760	0.5218	0.6024	0.6501	0.6818
	LUM	0.7324	0.7275	0.7259	0.7170	0.7182	0.7190	0.1458	0.2590	0.4217	0.5247	0.6061	0.6666	0.2410	0.3774	0.5261	0.5969	0.6466	0.6793
	PHOG	0.7387	0.7234	0.7311	0.7278	0.7264	0.7224	0.1514	0.2618	0.4270	0.5345	0.6075	0.6652	0.2490	0.3797	0.5306	0.6069	0.6517	0.6808
	SCH	0.7432	0.7329	0.7270	0.7269	0.7203	0.7198	0.1465	0.2604	0.4178	0.5248	0.6024	0.6584	0.2426	0.3798	0.5239	0.6009	0.6446	0.6744
textual	TAMURA	0.7396	0.7297	0.7198	0.7201	0.7170	0.7181	0.1458	0.2534	0.4092	0.5166	0.5915	0.6554	0.2408	0.3715	0.5138	0.5926	0.6372	0.6728
	BM25	0.7414	0.7194	0.7277	0.7273	0.7251	0.7245	0.1514	0.2597	0.4267	0.5304	0.6104	0.6684	0.2472	0.3750	0.5261	0.6010	0.6499	0.6816
	COSINE	0.7351	0.7185	0.7198	0.7218	0.7216	0.7214	0.1479	0.2576	0.4235	0.5272	0.6062	0.6686	0.2442	0.3739	0.5252	0.5989	0.6471	0.6807
	DICE	0.7315	0.7378	0.7273	0.7288	0.7250	0.7223	0.1498	0.2657	0.4163	0.5319	0.6113	0.6686	0.2462	0.3856	0.5210	0.6038	0.6511	0.6810
	JACCARD	0.7559	0.7396	0.7239	0.7248	0.7260	0.7253	0.1541	0.2703	0.4262	0.5344	0.6124	0.6668	0.2535	0.3900	0.5282	0.6051	0.6527	0.6819
	Max()	0.7586	0.7432	0.7381	0.7336	0.7301	0.7310	0.1541	0.2703	0.4270	0.5345	0.6124	0.6731	0.2535	0.3900	0.5306	0.6069	0.6527	0.6870
	Min()	0.7216	0.7185	0.7198	0.7170	0.7170	0.7181	0.1432	0.2507	0.4092	0.5166	0.5915	0.6554	0.2367	0.3677	0.5138	0.5926	0.6372	0.6728
		AVG	0.7292		Median	0.7273	Q 75%	0.7319	AVG	0.4378	Median	0.4718	Q 75%	0.6072	AVG	0.5126	Median	0.5616	Q 75%

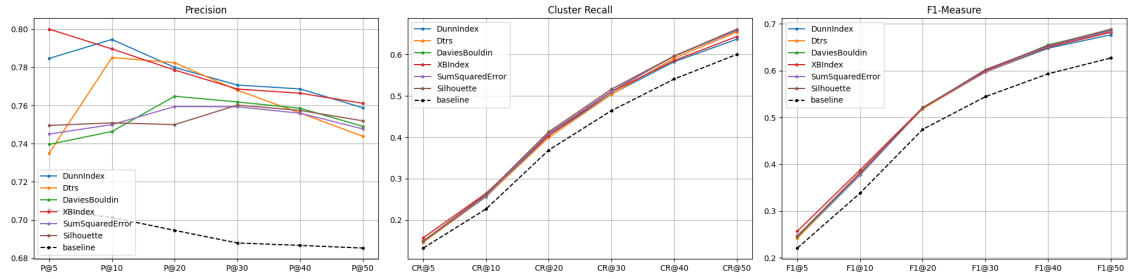
(b) Resultados do experimento 03 (K-Medoids) para o DTRS.

Tabela 4.22: Resultado comparativo entre todos os descritores visuais e textuais apenas para o método DTRS sem teste de significância estatística. Experimentos 02 e 03. Algoritmos Hierárquico Aglomerativo e K-Medoids, respectivamente. Com uso do IASMIN, melhores resultados em negrito.

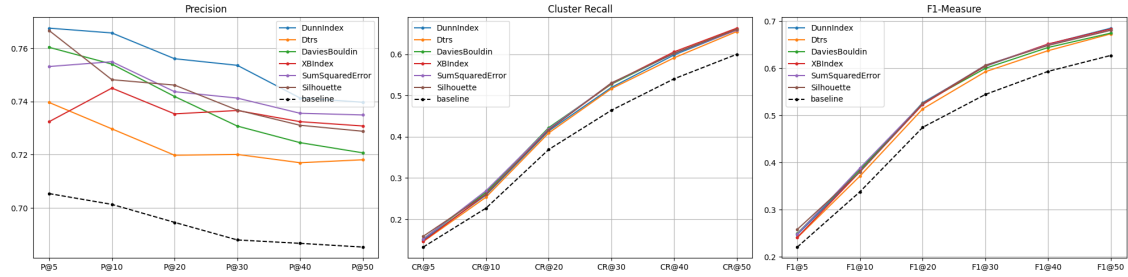
Nas Tabelas 4.23 a 4.37 temos os resultados das métricas *Precision*, *Cluster Recall* e *F1-Measure*, obtidas pelo experimento 02, com uso do IASMIN aplicado ao algoritmo Hierárquico Aglomerativo e com a realização do teste de significância estatística de *Wilcoxon*. Aqui, a superioridade do método DTRS aparece destacada em azul, com a inferioridade em vermelho, assim como no experimento da seção anterior.

	n	DTRS	K=20	Precision					DTRS	K=20	Dunn	Cluster-Recall (CR)				DTRS	K=20	Dunn	F1-Measure (F1)				DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
				Dunn	DB	XB	SSE	Silhouette				DB	XB	SSE	Silhouette				DB	XB	SSE	Silhouette							
EXPERIMENTO 02	5	0.7351	0.7054	0.7847	0.7396	0.8000	0.7450	0.7495	0.1465	0.1325	0.1473	0.1488	0.1574	0.1498	0.1494	0.2419	0.2209	0.2429	0.2453	0.2569	0.2471	0.2469	0.2419	0.2209	0.2429	0.2453	0.2569	0.2471	0.2469
	10	0.7851	0.7014	0.7946	0.7464	0.7896	0.7500	0.7509	0.2574	0.2268	0.2563	0.2596	0.2646	0.2586	0.2628	0.3803	0.3383	0.3771	0.3775	0.3877	0.3772	0.3822	0.3803	0.3383	0.3771	0.3775	0.3877	0.3772	0.3822
	20	0.7824	0.6946	0.7800	0.7649	0.7786	0.7595	0.7500	0.3994	0.3691	0.4047	0.4090	0.4047	0.4092	0.4131	0.5186	0.4747	0.5203	0.5213	0.5210	0.5206	0.5202	0.5186	0.4747	0.5203	0.5213	0.5210	0.5206	0.5202
	30	0.7680	0.6880	0.7707	0.7619	0.7686	0.7593	0.7602	0.5037	0.4643	0.5045	0.5108	0.5098	0.5111	0.5162	0.5970	0.5444	0.5971	0.5995	0.6002	0.5979	0.6022	0.5970	0.5444	0.5971	0.5995	0.6002	0.5979	0.6022
	40	0.7561	0.6867	0.7687	0.7586	0.7664	0.7560	0.7574	0.5911	0.5407	0.5824	0.5965	0.5854	0.5964	0.5963	0.6504	0.5933	0.6480	0.6550	0.6501	0.6521	0.6536	0.6504	0.5933	0.6480	0.6550	0.6501	0.6521	0.6536
	50	0.7439	0.6853	0.7588	0.7491	0.7612	0.7477	0.7520	0.6543	0.6002	0.6368	0.6583	0.6432	0.6579	0.6612	0.6826	0.6271	0.6766	0.6869	0.6820	0.6849	0.6887	0.6826	0.6271	0.6766	0.6869	0.6820	0.6849	0.6887
EXPERIMENTO 03	5	0.7396	0.7054	0.7676	0.7604	0.7324	0.7532	0.7667	0.1458	0.1325	0.1494	0.1525	0.1465	0.1519	0.1587	0.2408	0.2209	0.2473	0.2499	0.2415	0.2487	0.2584	0.2408	0.2209	0.2473	0.2499	0.2415	0.2487	0.2584
	10	0.7297	0.7014	0.7658	0.7541	0.7450	0.7550	0.7482	0.2534	0.2268	0.2584	0.2652	0.2607	0.2690	0.2611	0.3715	0.3383	0.3814	0.3852	0.3808	0.3887	0.3804	0.3715	0.3383	0.3814	0.3852	0.3808	0.3887	0.3804
	20	0.7198	0.6946	0.7561	0.7419	0.7354	0.7437	0.7462	0.4092	0.3691	0.4141	0.4221	0.4176	0.4194	0.4165	0.5138	0.4747	0.5270	0.5259	0.5231	0.5256	0.5251	0.5138	0.4747	0.5270	0.5259	0.5231	0.5256	0.5251
	30	0.7201	0.6880	0.7536	0.7308	0.7366	0.7413	0.7368	0.5166	0.4643	0.5196	0.5277	0.5293	0.5301	0.5307	0.5926	0.5444	0.6042	0.5998	0.6050	0.6064	0.6061	0.5926	0.5444	0.6042	0.5998	0.6050	0.6064	0.6061
	40	0.7170	0.6867	0.7413	0.7245	0.7324	0.7356	0.7311	0.5915	0.5407	0.5986	0.6026	0.6061	0.6022	0.6022	0.6372	0.5933	0.6509	0.6439	0.6514	0.6494	0.6483	0.6372	0.5933	0.6509	0.6439	0.6514	0.6494	0.6483
	50	0.7181	0.6853	0.7397	0.7207	0.7308	0.7350	0.7288	0.6554	0.6002	0.6594	0.6610	0.6635	0.6586	0.6605	0.6728	0.6271	0.6849	0.6744	0.6833	0.6808	0.6804	0.6728	0.6271	0.6849	0.6744	0.6833	0.6808	0.6804

Tabela 4.23: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor TAMURA (Tamura et al., 1978)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

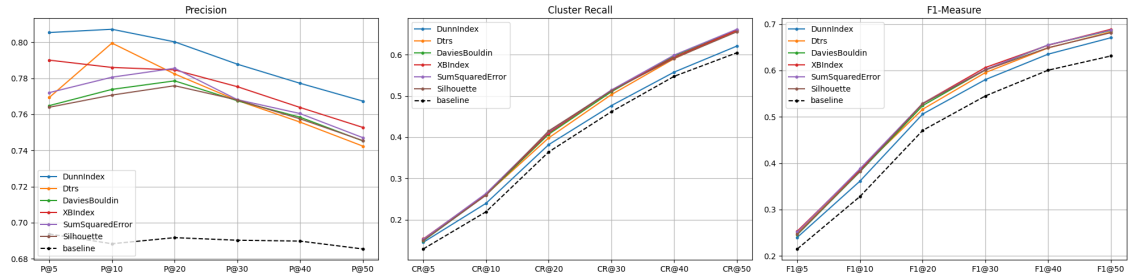


(b) Comportamento dos resultados experimento 03: K-Medoids.

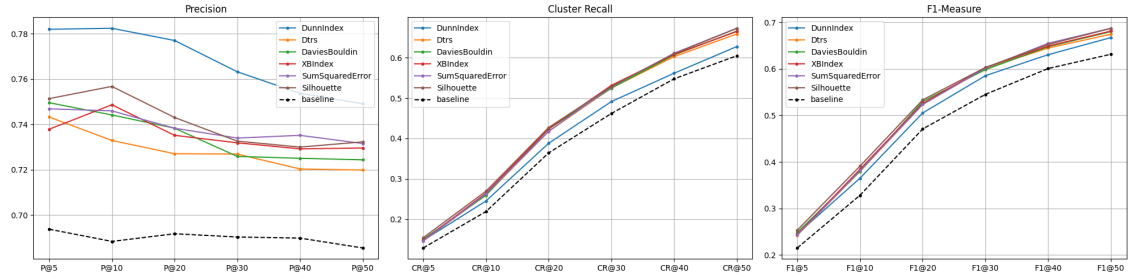
Figura 4.23: Performance: descritor TAMURA. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision					DTRS	K=20	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
				Dunn	DB	XB	SSE	Silhouette			Dunn	DB	XB	SSE	Silhouette														
experimento 02	5	0.7694	0.6937	0.8054	0.7649	0.7901	0.7721	0.7640	0.1508	0.1290	0.1451	0.1486	0.1533	0.1524	0.1493	0.2496	0.2151	0.2402	0.2460	0.2537	0.2516	0.2467	0.2496	0.2151	0.2402	0.2460	0.2537	0.2516	0.2467
	10	0.7995	0.6883	0.8072	0.7739	0.7860	0.7806	0.7707	0.2606	0.2190	0.2396	0.2599	0.2620	0.2636	0.2592	0.3858	0.3283	0.3616	0.3818	0.3860	0.3882	0.3827	0.3858	0.3283	0.3616	0.3818	0.3860	0.3882	0.3827
	20	0.7824	0.6917	0.8002	0.7786	0.7847	0.7856	0.7759	0.3983	0.3642	0.3820	0.4071	0.4103	0.4134	0.4152	0.5165	0.4708	0.5061	0.5243	0.5291	0.5290	0.5284	0.5165	0.4708	0.5061	0.5243	0.5291	0.5290	0.5284
	30	0.7677	0.6902	0.7878	0.7676	0.7754	0.7682	0.7680	0.5032	0.4618	0.4766	0.5101	0.5136	0.5143	0.5125	0.5956	0.5452	0.5804	0.6013	0.6065	0.6022	0.6011	0.5956	0.5452	0.5804	0.6013	0.6065	0.6022	0.6011
	40	0.7557	0.6898	0.7773	0.7584	0.7639	0.7605	0.7575	0.5903	0.5475	0.5578	0.5971	0.5937	0.5990	0.5913	0.6485	0.6007	0.6352	0.6549	0.6545	0.6552	0.6491	0.6485	0.6007	0.6352	0.6549	0.6545	0.6552	0.6491
	50	0.7424	0.6854	0.7674	0.7453	0.7528	0.7471	0.7455	0.6582	0.6046	0.6209	0.6602	0.6590	0.6616	0.6555	0.6831	0.6313	0.6706	0.6863	0.6889	0.6871	0.6815	0.6831	0.6313	0.6706	0.6863	0.6889	0.6871	0.6815
experimento 03	5	0.7432	0.6937	0.7820	0.7495	0.7378	0.7468	0.7514	0.1465	0.1290	0.1476	0.1504	0.1477	0.1465	0.1544	0.2426	0.2151	0.2451	0.2485	0.2440	0.2427	0.2534	0.2426	0.2151	0.2451	0.2485	0.2440	0.2427	0.2534
	10	0.7329	0.6883	0.7824	0.7441	0.7486	0.7459	0.7568	0.2604	0.2190	0.2451	0.2592	0.2647	0.2624	0.2695	0.3798	0.3283	0.3651	0.3796	0.3842	0.3814	0.3910	0.3798	0.3283	0.3651	0.3796	0.3842	0.3814	0.3910
	20	0.7270	0.6917	0.7770	0.7383	0.7351	0.7383	0.7430	0.4178	0.3642	0.3880	0.4235	0.4239	0.4174	0.4271	0.5239	0.4708	0.5052	0.5298	0.5257	0.5232	0.5333	0.5239	0.4708	0.5052	0.5298	0.5257	0.5232	0.5333
	30	0.7269	0.6902	0.7632	0.7258	0.7318	0.7339	0.7326	0.5248	0.4618	0.4912	0.5245	0.5314	0.5275	0.5286	0.6009	0.5452	0.5853	0.5983	0.6027	0.6016	0.6029	0.6009	0.5452	0.5853	0.5983	0.6027	0.6016	0.6029
	40	0.7203	0.6898	0.7537	0.7250	0.7292	0.7351	0.7300	0.6024	0.5475	0.5614	0.6066	0.6076	0.6108	0.6096	0.6446	0.6007	0.6304	0.6476	0.6490	0.6548	0.6524	0.6446	0.6007	0.6304	0.6476	0.6490	0.6548	0.6524
	50	0.7198	0.6854	0.7491	0.7243	0.7295	0.7315	0.7323	0.6584	0.6046	0.6274	0.6657	0.6648	0.6727	0.6725	0.6744	0.6313	0.6670	0.6799	0.6812	0.6871	0.6869	0.6744	0.6313	0.6670	0.6799	0.6812	0.6871	0.6869

Tabela 4.24: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor SCH (Cinque et al., 2001)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

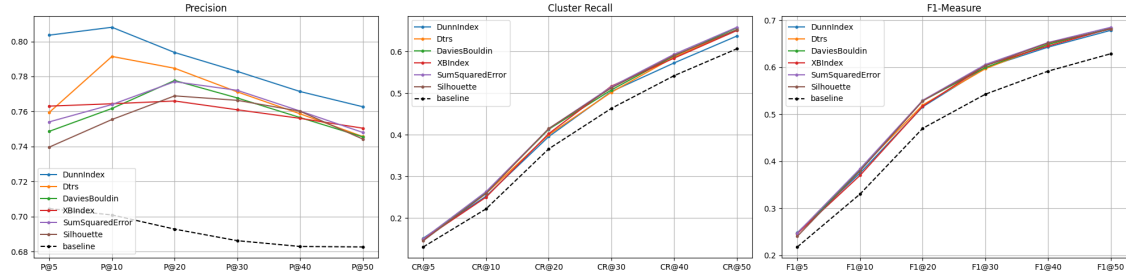


(b) Comportamento dos resultados experimento 03: K-Medoids.

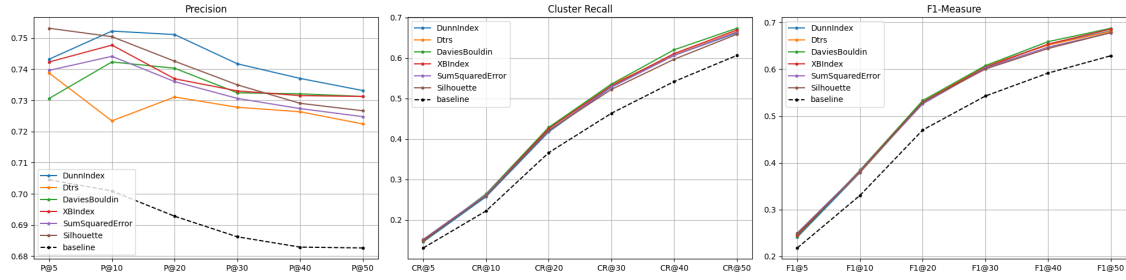
Figura 4.24: Performance: descritor SCH. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					SSE	Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					SSE	Silhouette	
			K=20	Dunn	DB	XB				K=20	Dunn	DB	XB	SSE			Silhouette	K=20	Dunn	DB	XB			SSE
experimento 02	5	0.7595	0.7045	0.8036	0.7486	0.7631	0.7541	0.7396	0.1496	0.1307	0.1518	0.1500	0.1487	0.1501	0.1456	0.2474	0.2180	0.2485	0.2467	0.2464	0.2476	0.2405	0.6488	0.5918
	10	0.7914	0.7009	0.8081	0.7617	0.7644	0.7640	0.7554	0.2587	0.2220	0.2522	0.2613	0.2501	0.2632	0.2590	0.3828	0.3304	0.3750	0.3817	0.3701	0.3842	0.3797	0.5198	0.4701
	20	0.7847	0.6928	0.7937	0.7777	0.7660	0.7770	0.7689	0.4000	0.3661	0.3955	0.4131	0.4025	0.4146	0.4149	0.5198	0.4701	0.5159	0.5289	0.5170	0.5297	0.5285	0.5969	0.5427
	30	0.7712	0.6862	0.7829	0.7677	0.7610	0.7721	0.7664	0.5022	0.4634	0.5038	0.5077	0.5169	0.5153	0.5126	0.6069	0.5427	0.5992	0.5994	0.6039	0.6059	0.6031	0.6488	0.5918
	40	0.7587	0.6829	0.7714	0.7565	0.7561	0.7602	0.7601	0.5863	0.5418	0.5725	0.5892	0.5839	0.5933	0.5905	0.6488	0.5918	0.6428	0.6486	0.6452	0.6525	0.6515	0.6821	0.6288
	50	0.7457	0.6826	0.7627	0.7455	0.7505	0.7505	0.7480	0.7442	0.6520	0.6064	0.6370	0.6568	0.6503	0.6581	0.6524	0.6821	0.6288	0.6785	0.6832	0.6823	0.6850	0.6811	0.6808
experimento 03	5	0.7387	0.7045	0.7432	0.7306	0.7423	0.7396	0.7532	0.1514	0.1307	0.1458	0.1474	0.1483	0.1519	0.1501	0.2490	0.2180	0.2411	0.2429	0.2452	0.2501	0.2484	0.6075	0.5418
	10	0.7234	0.7009	0.7523	0.7423	0.7477	0.7441	0.7505	0.2618	0.2220	0.2572	0.2641	0.2618	0.2620	0.2595	0.3797	0.3304	0.3792	0.3843	0.3831	0.3824	0.3805	0.6075	0.5418
	20	0.7311	0.6928	0.7511	0.7403	0.7369	0.7360	0.7426	0.4270	0.3661	0.4175	0.4286	0.4253	0.4198	0.4211	0.5306	0.4701	0.5282	0.5331	0.5295	0.5258	0.5292	0.6075	0.5418
	30	0.7278	0.6862	0.7417	0.7324	0.7330	0.7306	0.7350	0.5345	0.4634	0.5281	0.5356	0.5310	0.5286	0.5225	0.6069	0.5427	0.6065	0.6077	0.6039	0.6027	0.6005	0.6075	0.5418
	40	0.7264	0.6829	0.7370	0.7321	0.7315	0.7274	0.7291	0.6075	0.5418	0.6066	0.6203	0.6109	0.6065	0.5969	0.6517	0.5918	0.6527	0.6589	0.6536	0.6473	0.6443	0.6517	0.5918
	50	0.7224	0.6826	0.7332	0.7313	0.7313	0.7248	0.7267	0.6652	0.6064	0.6621	0.6735	0.6693	0.6638	0.6586	0.6808	0.6288	0.6819	0.6875	0.6859	0.6779	0.6782	0.6808	0.6288

Tabela 4.25: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor PHOG (Biasotti et al., 2013)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

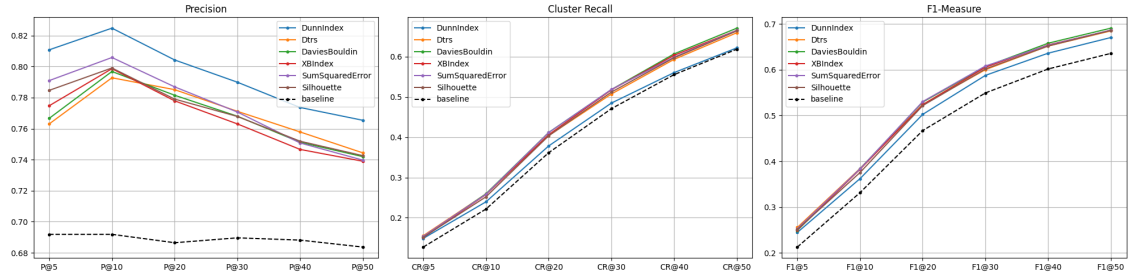


(b) Comportamento dos resultados experimento 03: K-Medoids.

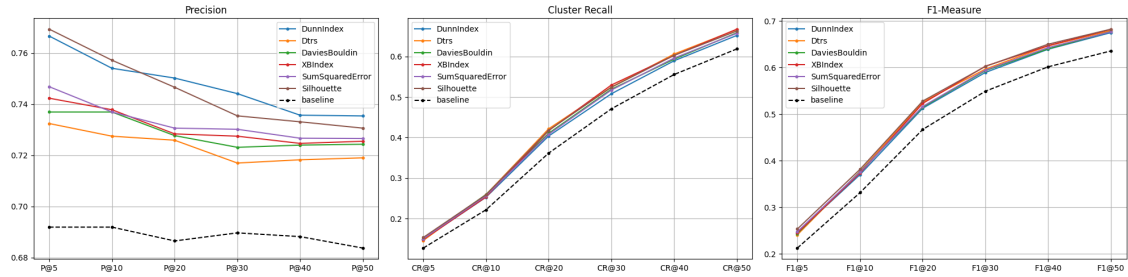
Figura 4.25: Performance: descritor PHOG. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					SSE	Silhouette
			K=20	Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB			SSE	DTRS	K=20	Dunn	DB		
experimento 02	5	0.7631	0.6919	0.8108	0.7667	0.7748	0.7910	0.7847	0.1555	0.1272	0.1488	0.1520	0.1508	0.1540	0.1517	0.2555	0.2127	0.2447	0.2490	0.2500	0.2529	0.2510
	10	0.7928	0.6919	0.8248	0.7968	0.7986	0.8059	0.7991	0.2591	0.2216	0.2396	0.2596	0.2579	0.2584	0.2520	0.3842	0.3316	0.3621	0.3838	0.3826	0.3835	0.3755
	20	0.7851	0.6865	0.8043	0.7815	0.7779	0.7869	0.7791	0.4039	0.3616	0.3784	0.4118	0.4063	0.4116	0.4036	0.5238	0.4672	0.5025	0.5294	0.5239	0.5305	0.5217
	30	0.7712	0.6896	0.7899	0.7680	0.7631	0.7707	0.7679	0.5075	0.4710	0.4849	0.5185	0.5179	0.5188	0.5122	0.6002	0.5494	0.5878	0.6069	0.6055	0.6082	0.6019
	40	0.7579	0.6882	0.7736	0.7512	0.7466	0.7508	0.7519	0.5935	0.5556	0.5607	0.6071	0.6036	0.5990	0.5975	0.6518	0.6017	0.6362	0.6580	0.6542	0.6535	0.6519
50	0.7444	0.6837	0.7654	0.7420	0.7390	0.7397	0.7426	0.6596	0.6188	0.6220	0.6710	0.6657	0.6656	0.6642	0.6849	0.6357	0.6704	0.6909	0.6862	0.6866	0.6859	
experimento 03	5	0.7324	0.6919	0.7667	0.7369	0.7423	0.7468	0.7694	0.1458	0.1272	0.1486	0.1493	0.1483	0.1501	0.1534	0.2410	0.2127	0.2460	0.2437	0.2452	0.2479	0.2536
	10	0.7275	0.6919	0.7541	0.7369	0.7378	0.7369	0.7572	0.2590	0.2216	0.2527	0.2595	0.2532	0.2561	0.2587	0.3774	0.3316	0.3703	0.3765	0.3736	0.3758	0.3817
	20	0.7259	0.6865	0.7502	0.7277	0.7284	0.7306	0.7466	0.4217	0.3616	0.4038	0.4105	0.4171	0.4081	0.4180	0.5261	0.4672	0.5128	0.5150	0.5231	0.5167	0.5282
	30	0.7170	0.6896	0.7441	0.7231	0.7275	0.7302	0.7354	0.5247	0.4710	0.5081	0.5193	0.5297	0.5172	0.5241	0.5969	0.5494	0.5897	0.5935	0.6031	0.5947	0.6025
	40	0.7182	0.6882	0.7357	0.7240	0.7247	0.7267	0.7331	0.6061	0.5556	0.5895	0.5935	0.6027	0.5962	0.6029	0.6466	0.6017	0.6394	0.6403	0.6471	0.6429	0.6502
50	0.7190	0.6837	0.7354	0.7243	0.7255	0.7266	0.7306	0.6666	0.6188	0.6521	0.6587	0.6676	0.6572	0.6631	0.6793	0.6357	0.6750	0.6773	0.6819	0.6769	0.6825	

Tabela 4.26: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor LUM (Lux e Chatzichristofis, 2008)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

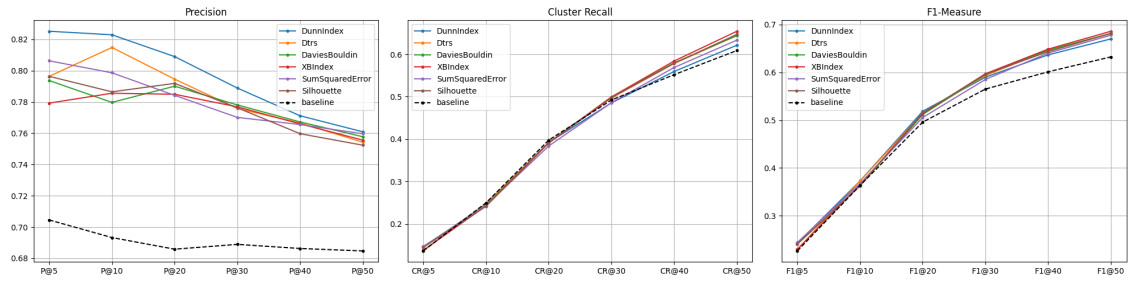


(b) Comportamento dos resultados experimento 03: K-Medoids.

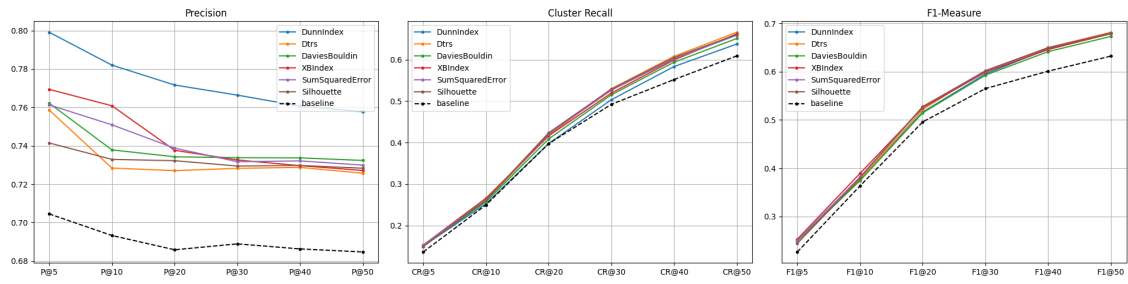
Figura 4.26: Performance: descritor LUM. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					SSE	Silhouette
			K=20	Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB			SSE	DTRS	K=20	Dunn	DB		
experimento 02	5	0.7964	0.7045	0.8252	0.7937	0.7793	0.8063	0.7964	0.1362	0.1364	0.1450	0.1463	0.1377	0.1464	0.1445	0.2290	0.2264	0.2425	0.2407	0.2301	0.2441	0.2407
	10	0.8149	0.6932	0.8230	0.7797	0.7856	0.7986	0.7865	0.2484	0.2497	0.2460	0.2442	0.2424	0.2437	0.2415	0.3730	0.3634	0.3729	0.3629	0.3646	0.3674	0.3630
	20	0.7946	0.6858	0.8090	0.7901	0.7849	0.7842	0.7919	0.3917	0.3976	0.3920	0.3912	0.3919	0.3837	0.3905	0.5144	0.4960	0.5187	0.5119	0.5140	0.5056	0.5153
	30	0.7758	0.6889	0.7889	0.7781	0.7769	0.7701	0.7763	0.4971	0.4972	0.4855	0.4992	0.4995	0.4859	0.4978	0.5932	0.5650	0.5899	0.5954	0.5972	0.5852	0.5961
	40	0.7666	0.6863	0.7712	0.7672	0.7660	0.7655	0.7597	0.5783	0.5521	0.5596	0.5800	0.5843	0.5690	0.5789	0.6462	0.6012	0.6366	0.6465	0.6486	0.6404	0.6435
50	0.7542	0.6847	0.7609	0.7577	0.7556	0.7596	0.7523	0.6443	0.6090	0.6215	0.6443	0.6545	0.6334	0.6470	0.6810	0.6323	0.6699	0.6815	0.6859	0.6777	0.6815	
experimento 03	5	0.7586	0.7045	0.7991	0.7622	0.7694	0.7613	0.7414	0.1516	0.1364	0.1506	0.1518	0.1531	0.1529	0.1493	0.2488	0.2264	0.2504	0.2490	0.2530	0.2500	0.2440
	10	0.7284	0.6932	0.7820	0.7378	0.7608	0.7509	0.7329	0.2593	0.2497	0.2526	0.2573	0.2659	0.2615	0.2617	0.3760	0.3634	0.3758	0.3742	0.3892	0.3819	0.3799
	20	0.7270	0.6858	0.7716	0.7342	0.7376	0.7387	0.7322	0.4177	0.3976	0.3979	0.4089	0.4164	0.4210	0.4237	0.5218	0.4960	0.5165	0.5147	0.5253	0.5278	0.5276
	30	0.7282	0.6889	0.7664	0.7338	0.7326	0.7317	0.7294	0.5299	0.4923	0.5034	0.5153	0.5195	0.5263	0.5289	0.6024	0.5650	0.5954	0.5929	0.5990	0.6009	0.6020
	40	0.7287	0.6863	0.7606	0.7337	0.7296	0.7321	0.7297	0.6081	0.5521	0.5834	0.5938	0.5993	0.6020	0.6051	0.6501	0.6012	0.6474	0.6416	0.6459	0.6491	0.6497
50	0.7257	0.6847	0.7577	0.7323	0.7271	0.7299	0.7283	0.6660	0.6090	0.6376	0.6510	0.6621	0.6589	0.6606	0.6818	0.6323	0.6786	0.6732	0.6792	0.6801	0.6803	

Tabela 4.27: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor JCD (Zagoris et al., 2010)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

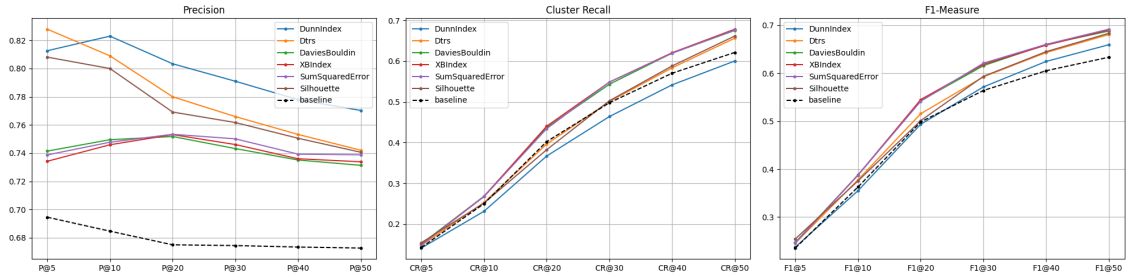


(b) Comportamento dos resultados experimento 03: K-Medoids.

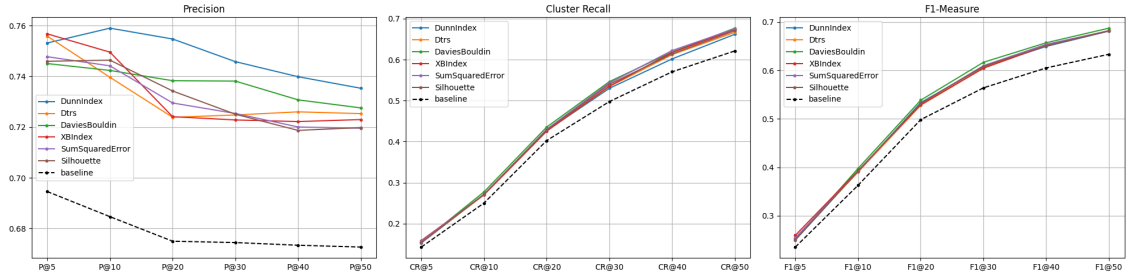
Figura 4.27: Performance: descritor JCD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					DTRS	Cluster-Recall (CR)					DTRS	F1-Measure (F1)					DTRS	K=20	Dunn	DB	SSE	Silhouette					
			K=20	Dunn	DB	XB	SSE		Silhouette	K=20	Dunn	DB	XB		SSE	Silhouette	K=20	Dunn	DB							XB	SSE	Silhouette		
experimento_02	5	0.8279	0.6946	0.8126	0.7414	0.7342	0.7387	0.8081	0.1478	0.1428	0.1414	0.1515	0.1495	0.1493	0.1544	0.2457	0.2352	0.2372	0.2473	0.2460	0.2465	0.2543	0.8090	0.6847	0.8230	0.7495	0.7459	0.7477	0.8000	
	10	0.8090	0.6847	0.8230	0.7495	0.7459	0.7477	0.8000	0.2530	0.2502	0.2316	0.2684	0.2681	0.2676	0.2518	0.3777	0.3628	0.3551	0.3879	0.3876	0.3875	0.3755	0.7800	0.6750	0.8034	0.7518	0.7532	0.7534	0.7691	
	20	0.7800	0.6750	0.8034	0.7518	0.7532	0.7534	0.7691	0.3966	0.4026	0.3671	0.4373	0.4404	0.4343	0.3826	0.5156	0.4985	0.4937	0.5429	0.5451	0.5416	0.5012	0.7659	0.6745	0.7910	0.7432	0.7461	0.7502	0.7617	
	30	0.7659	0.6745	0.7910	0.7432	0.7461	0.7502	0.7617	0.5008	0.4979	0.4640	0.5432	0.5486	0.5487	0.5028	0.5926	0.5641	0.5715	0.6159	0.6187	0.6211	0.5936	0.40	0.7533	0.6734	0.7779	0.7351	0.7360	0.7393	0.7506
	40	0.7533	0.6734	0.7779	0.7351	0.7360	0.7393	0.7506	0.5836	0.5704	0.5418	0.6198	0.6206	0.6193	0.5882	0.6433	0.6053	0.6247	0.6601	0.6588	0.6604	0.6451	50	0.7421	0.6727	0.7703	0.7314	0.7339	0.7388	0.7406
	50	0.7421	0.6727	0.7703	0.7314	0.7339	0.7388	0.7406	0.6564	0.6214	0.6004	0.6754	0.6785	0.6764	0.6616	0.6809	0.6335	0.6596	0.6882	0.6904	0.6921	0.6835								
experimento_03	5	0.7559	0.6946	0.7532	0.7450	0.7568	0.7477	0.7459	0.1541	0.1428	0.1530	0.1541	0.1586	0.1565	0.1534	0.2535	0.2352	0.2498	0.2515	0.2596	0.2545	0.2502	10	0.7396	0.6847	0.7590	0.7423	0.7495	0.7441	0.7464
	10	0.7396	0.6847	0.7590	0.7423	0.7495	0.7441	0.7464	0.2703	0.2502	0.2704	0.2775	0.2704	0.2711	0.2718	0.3900	0.3628	0.3916	0.3976	0.3930	0.3911	0.3920	20	0.7239	0.6750	0.7547	0.7383	0.7241	0.7295	0.7342
	20	0.7239	0.6750	0.7547	0.7383	0.7241	0.7295	0.7342	0.4262	0.4026	0.4268	0.4356	0.4256	0.4303	0.4281	0.5282	0.4985	0.5338	0.5387	0.5291	0.5321	0.5318	30	0.7248	0.6745	0.7458	0.7381	0.7228	0.7254	0.7251
	30	0.7248	0.6745	0.7458	0.7381	0.7228	0.7254	0.7251	0.5344	0.4979	0.5302	0.5465	0.5354	0.5432	0.5399	0.6051	0.5641	0.6076	0.6170	0.6047	0.6099	0.6088	40	0.7260	0.6734	0.7399	0.7307	0.7222	0.7200	0.7187
	40	0.7260	0.6734	0.7399	0.7307	0.7222	0.7200	0.7187	0.6124	0.5704	0.6019	0.6187	0.6176	0.6224	0.6139	0.6527	0.6053	0.6497	0.6573	0.6513	0.6540	0.6509	50	0.7253	0.6727	0.7353	0.7276	0.7230	0.7195	0.7198
	50	0.7253	0.6727	0.7353	0.7276	0.7230	0.7195	0.7198	0.6668	0.6214	0.6620	0.6769	0.6736	0.6751	0.6708	0.6819	0.6335	0.6814	0.6875	0.6818	0.6822	0.6821								

Tabela 4.28: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor JACCARD (Lewis et al., 2006)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

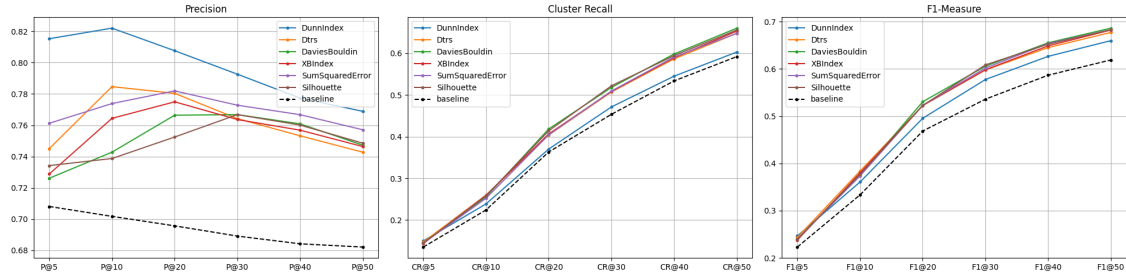


(b) Comportamento dos resultados experimento 03: K-Medoids.

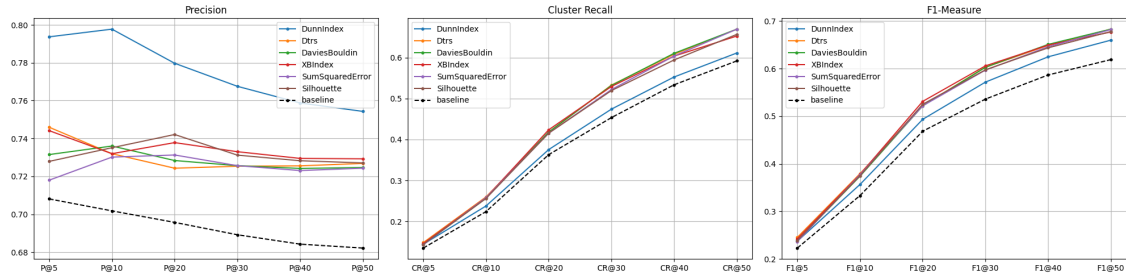
Figura 4.28: Performance: descritor JACCARD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision					Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					SSE	Silhouette
				Dunn	DB	XB	SSE	DTRS		K=20	Dunn	DB	XB	SSE			DTRS	K=20	Dunn	DB	XB		
experimento 02	5	0.7450	0.7081	0.8153	0.7261	0.7288	0.7613	0.7342	0.1471	0.1352	0.1500	0.1447	0.1436	0.1443	0.1445	0.2434	0.2229	0.2464	0.2389	0.2376	0.2399	0.2393	0.2393
	10	0.7847	0.7018	0.8221	0.7428	0.7644	0.7739	0.7387	0.2598	0.2242	0.2386	0.2556	0.2592	0.2522	0.2581	0.3837	0.3333	0.3606	0.3752	0.3799	0.3734	0.3770	
	20	0.7804	0.6957	0.8077	0.7664	0.7750	0.7820	0.7525	0.4039	0.3632	0.3702	0.4184	0.4065	0.4043	0.4140	0.5225	0.4685	0.4955	0.5305	0.5225	0.5229	0.5227	
	30	0.7643	0.6892	0.7926	0.7668	0.7637	0.7728	0.7668	0.5072	0.4535	0.4712	0.5180	0.5082	0.5093	0.5219	0.5978	0.5360	0.5776	0.6059	0.5982	0.6019	0.6085	
	40	0.7533	0.6842	0.7774	0.7609	0.7569	0.7668	0.7601	0.5865	0.5337	0.5451	0.5982	0.5893	0.5909	0.5944	0.6451	0.5867	0.6266	0.6554	0.6489	0.6539	0.6524	
	50	0.7428	0.6822	0.7689	0.7473	0.7464	0.7571	0.7486	0.6478	0.5920	0.6029	0.6602	0.6532	0.6478	0.6559	0.6772	0.6190	0.6598	0.6863	0.6822	0.6830	0.6830	
experimento 03	5	0.7459	0.7081	0.7937	0.7315	0.7441	0.7180	0.7279	0.1485	0.1352	0.1445	0.1430	0.1461	0.1432	0.1442	0.2454	0.2229	0.2373	0.2385	0.2423	0.2365	0.2388	0.2388
	10	0.7320	0.7018	0.7977	0.7360	0.7320	0.7302	0.7351	0.2599	0.2242	0.2379	0.2566	0.2582	0.2581	0.2562	0.3793	0.3333	0.3566	0.3740	0.3777	0.3769	0.3748	
	20	0.7243	0.6957	0.7797	0.7284	0.7378	0.7313	0.7421	0.4175	0.3632	0.3755	0.4185	0.4238	0.4157	0.4156	0.5223	0.4685	0.4935	0.5218	0.5312	0.5219	0.5251	
	30	0.7254	0.6892	0.7676	0.7257	0.7330	0.7257	0.7312	0.5329	0.4535	0.4742	0.5320	0.5295	0.5219	0.5195	0.6029	0.5360	0.5717	0.6028	0.6058	0.5970	0.5975	
	40	0.7256	0.6842	0.7588	0.7242	0.7295	0.7231	0.7283	0.6082	0.5337	0.5527	0.6105	0.6037	0.6039	0.5943	0.6490	0.5867	0.6248	0.6509	0.6494	0.6460	0.6437	
	50	0.7268	0.6822	0.7543	0.7247	0.7293	0.7243	0.7271	0.6693	0.5920	0.6113	0.6695	0.6527	0.6694	0.6567	0.6830	0.6190	0.6599	0.6830	0.6770	0.6820	0.6780	

Tabela 4.29: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor GABOR (Manjunath e Ma, 1996)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

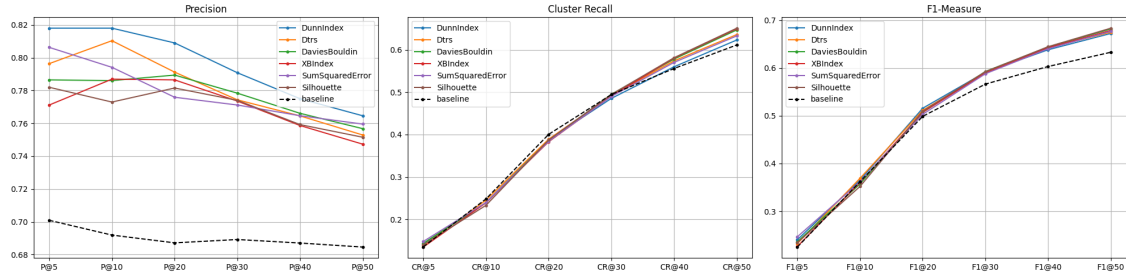


(b) Comportamento dos resultados experimento 03: K-Medoids.

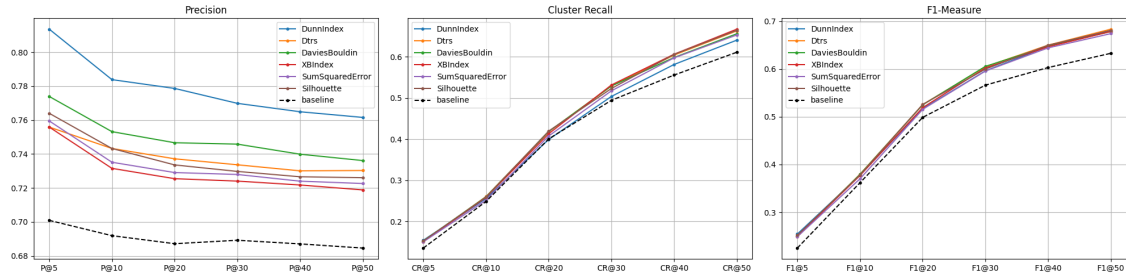
Figura 4.29: Performance: descritor GABOR. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Cluster-Recall (CR)					F1-Measure (F1)									
			K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
experimento 02	5	0.7964	0.7009	0.8180	0.7865	0.7712	0.8063	0.7820	0.1385	0.1355	0.1435	0.1437	0.1350	0.1482	0.1403	0.2323	0.2251	0.2404	0.2361	0.2262	0.2466	0.2343
	10	0.8104	0.6919	0.8180	0.7860	0.7869	0.7941	0.7730	0.2456	0.2485	0.2398	0.2396	0.2394	0.2400	0.2334	0.3692	0.3621	0.3654	0.3574	0.3607	0.3619	0.3520
	20	0.7912	0.6872	0.8090	0.7894	0.7865	0.7759	0.7815	0.3892	0.4006	0.3880	0.3856	0.3831	0.3841	0.3873	0.5116	0.4989	0.5159	0.5071	0.5075	0.5031	0.5096
	30	0.7743	0.6892	0.7908	0.7784	0.7736	0.7712	0.7740	0.4923	0.4944	0.4853	0.4949	0.4907	0.4900	0.4947	0.5897	0.5664	0.5914	0.5928	0.5896	0.5881	0.5926
	40	0.7645	0.6870	0.7750	0.7661	0.7587	0.7646	0.7592	0.5738	0.5557	0.5598	0.5780	0.5803	0.5701	0.5812	0.6422	0.6032	0.6383	0.6439	0.6425	0.6406	0.6447
	50	0.7529	0.6846	0.7646	0.7568	0.7473	0.7595	0.7515	0.6357	0.6112	0.6229	0.6466	0.6494	0.6325	0.6507	0.6750	0.6331	0.6724	0.6820	0.6789	0.6769	0.6835
experimento 03	5	0.7559	0.7009	0.8135	0.7739	0.7559	0.7595	0.7640	0.1509	0.1355	0.1544	0.1527	0.1514	0.1506	0.1529	0.2488	0.2251	0.2544	0.2503	0.2495	0.2489	0.2519
	10	0.7432	0.6919	0.7838	0.7532	0.7315	0.7351	0.7432	0.2606	0.2485	0.2558	0.2600	0.2585	0.2515	0.2594	0.3799	0.3621	0.3779	0.3795	0.3771	0.3702	0.3784
	20	0.7372	0.6872	0.7786	0.7466	0.7255	0.7291	0.7336	0.4183	0.4006	0.3994	0.4184	0.4131	0.4077	0.4197	0.5255	0.4989	0.5170	0.5253	0.5193	0.5158	0.5260
	30	0.7336	0.6892	0.7698	0.7458	0.7240	0.7279	0.7297	0.5291	0.4944	0.5037	0.5291	0.5316	0.5174	0.5230	0.6049	0.5664	0.5963	0.6057	0.6025	0.5959	0.6000
	40	0.7301	0.6870	0.7649	0.7399	0.7217	0.7240	0.7266	0.6043	0.5557	0.5815	0.5984	0.6064	0.5975	0.6060	0.6502	0.6032	0.6476	0.6478	0.6467	0.6444	0.6500
	50	0.7303	0.6846	0.7615	0.7361	0.7188	0.7226	0.7260	0.6631	0.6112	0.6405	0.6560	0.6672	0.6531	0.6645	0.6836	0.6331	0.6812	0.6787	0.6788	0.6743	0.6812

Tabela 4.30: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor FCTH (Chatzichristofis e Boutalis, 2008b)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

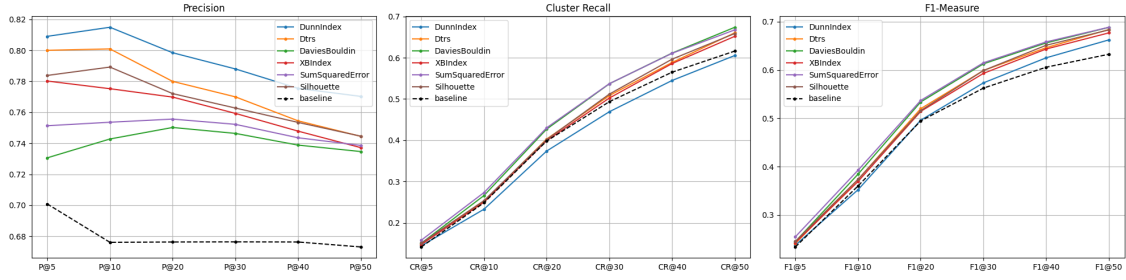


(b) Comportamento dos resultados experimento 03: K-Medoids.

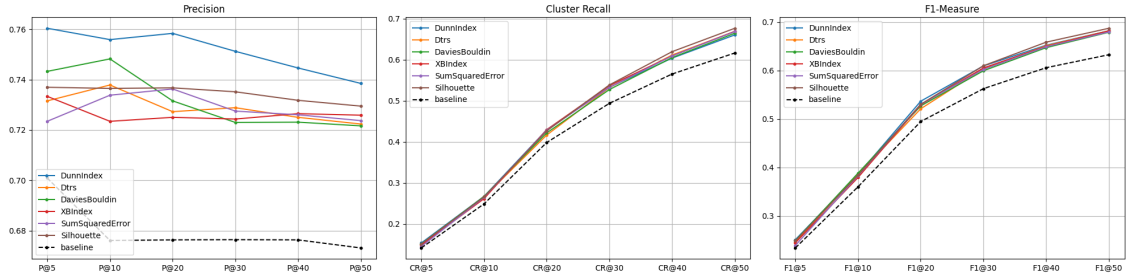
Figura 4.30: Performance: descritor FCTH. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					Silhouette	
			Dunn	DB	XB	SSE			Dunn	DB	XB	SSE				Dunn	DB	XB	SSE			
experimento 02	5	0.8000	0.7009	0.8090	0.7306	0.7802	0.7514	0.7838	0.1475	0.1415	0.1442	0.1505	0.1475	0.1578	0.1506	0.2411	0.2339	0.2376	0.2450	0.2428	0.2549	0.2455
	10	0.8009	0.6761	0.8149	0.7428	0.7752	0.7536	0.7892	0.2520	0.2486	0.2329	0.2661	0.2501	0.2733	0.2538	0.3736	0.3601	0.3523	0.3845	0.3694	0.3927	0.3737
	20	0.7800	0.6764	0.7984	0.7502	0.7698	0.7556	0.7721	0.4031	0.3982	0.3739	0.4271	0.3992	0.4303	0.4017	0.5200	0.4947	0.4962	0.5339	0.5144	0.5370	0.5161
	30	0.7700	0.6764	0.7880	0.7464	0.7593	0.7523	0.7628	0.5084	0.4936	0.4692	0.5370	0.5024	0.5372	0.5120	0.5989	0.5625	0.5734	0.6127	0.5930	0.6149	0.5989
	40	0.7545	0.6764	0.7755	0.7389	0.7479	0.7436	0.7535	0.5885	0.5650	0.5449	0.6113	0.5859	0.6108	0.5966	0.6461	0.6058	0.6252	0.6560	0.6434	0.6581	0.6511
	50	0.7445	0.6732	0.7703	0.7347	0.7372	0.7386	0.7446	0.6605	0.6162	0.6054	0.6739	0.6520	0.6676	0.6588	0.6837	0.6325	0.6622	0.6882	0.6768	0.6884	0.6829
experimento 03	5	0.7315	0.7009	0.7604	0.7432	0.7333	0.7234	0.7369	0.1498	0.1415	0.1535	0.1512	0.1483	0.1448	0.1511	0.2462	0.2339	0.2509	0.2490	0.2445	0.2390	0.2487
	10	0.7378	0.6761	0.7559	0.7482	0.7234	0.7338	0.7365	0.2657	0.2486	0.2669	0.2669	0.2614	0.2646	0.2658	0.3856	0.3601	0.3855	0.3883	0.3793	0.3828	0.3840
	20	0.7273	0.6764	0.7583	0.7315	0.7250	0.7363	0.7367	0.4163	0.3982	0.4290	0.4215	0.4294	0.4280	0.4265	0.5210	0.4947	0.5367	0.5270	0.5295	0.5302	0.5303
	30	0.7288	0.6764	0.7512	0.7230	0.7243	0.7275	0.7351	0.5319	0.4936	0.5326	0.5272	0.5367	0.5325	0.5384	0.6038	0.5625	0.6094	0.5995	0.6053	0.6022	0.6098
	40	0.7250	0.6764	0.7446	0.7231	0.7265	0.7260	0.7318	0.6113	0.5650	0.6037	0.6050	0.6092	0.6091	0.6195	0.6511	0.6058	0.6522	0.6471	0.6498	0.6491	0.6584
	50	0.7223	0.6732	0.7385	0.7216	0.7259	0.7237	0.7295	0.6686	0.6162	0.6602	0.6644	0.6691	0.6683	0.6749	0.6818	0.6325	0.6820	0.6789	0.6817	0.6796	0.6871

Tabela 4.31: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor DICE (Dice, 1945)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

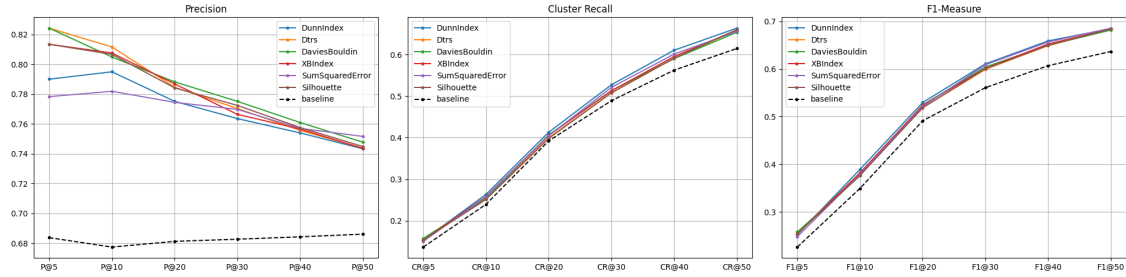


(b) Comportamento dos resultados experimento 03: K-Medoids.

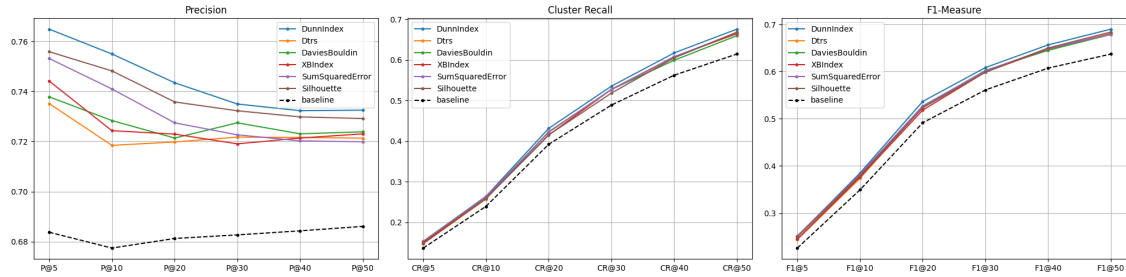
Figura 4.31: Performance: descritor DICE. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision					DTRS	K=20	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)				
				Dunn	DB	XB	SSE	Silhouette			Dunn	DB	XB	SSE	Silhouette			Dunn	DB	XB	SSE	Silhouette
experimento 02	5	0.8243	0.6838	0.7901	0.8243	0.8135	0.7784	0.8135	0.1562	0.1369	0.1532	0.1577	0.1536	0.1505	0.1539	0.2559	0.2262	0.2521	0.2579	0.2534	0.2478	0.2543
	10	0.8117	0.6775	0.7950	0.8050	0.8077	0.7820	0.8068	0.2551	0.2396	0.2636	0.2535	0.2586	0.2578	0.2514	0.3793	0.3495	0.3895	0.3765	0.3823	0.3791	0.3765
	20	0.7847	0.6813	0.7752	0.7883	0.7869	0.7745	0.7842	0.3999	0.3923	0.4132	0.4067	0.4054	0.4048	0.3982	0.5187	0.4915	0.5299	0.5248	0.5243	0.5216	0.5189
	30	0.7704	0.6827	0.7637	0.7752	0.7665	0.7698	0.7725	0.5079	0.4890	0.5273	0.5134	0.5128	0.5210	0.5079	0.5995	0.5610	0.6110	0.6044	0.6014	0.6091	0.6006
	40	0.7557	0.6843	0.7541	0.7610	0.7566	0.7571	0.7577	0.5910	0.5625	0.6104	0.5902	0.5954	0.6012	0.5914	0.6492	0.6072	0.6592	0.6502	0.6527	0.6569	0.6497
	50	0.7439	0.6861	0.7434	0.7478	0.7437	0.7517	0.7450	0.6588	0.6146	0.6636	0.6535	0.6591	0.6556	0.6600	0.6835	0.6369	0.6849	0.6820	0.6844	0.6860	0.6841
experimento 03	5	0.7351	0.6838	0.7649	0.7378	0.7441	0.7532	0.7559	0.1479	0.1369	0.1535	0.1487	0.1482	0.1534	0.1514	0.2442	0.2262	0.2511	0.2451	0.2449	0.2520	0.2501
	10	0.7185	0.6775	0.7550	0.7284	0.7243	0.7410	0.7482	0.2576	0.2396	0.2642	0.2592	0.2589	0.2627	0.2595	0.3739	0.3495	0.3852	0.3783	0.3764	0.3825	0.3799
	20	0.7198	0.6813	0.7435	0.7214	0.7230	0.7275	0.7358	0.4235	0.3923	0.4316	0.4222	0.4151	0.4224	0.4155	0.5252	0.4915	0.5367	0.5252	0.5183	0.5271	0.5233
	30	0.7218	0.6827	0.7350	0.7275	0.7191	0.7227	0.7323	0.5272	0.4890	0.5356	0.5274	0.5277	0.5270	0.5185	0.5989	0.5610	0.6084	0.6024	0.5988	0.6016	0.5985
	40	0.7216	0.6843	0.7323	0.7231	0.7214	0.7203	0.7298	0.6062	0.5625	0.6174	0.5993	0.6080	0.6079	0.6052	0.6471	0.6072	0.6563	0.6448	0.6481	0.6487	0.6498
	50	0.7214	0.6861	0.7325	0.7239	0.7231	0.7199	0.7292	0.6686	0.6146	0.6759	0.6598	0.6646	0.6671	0.6679	0.6807	0.6369	0.6897	0.6785	0.6795	0.6787	0.6837

Tabela 4.32: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor COSINE



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

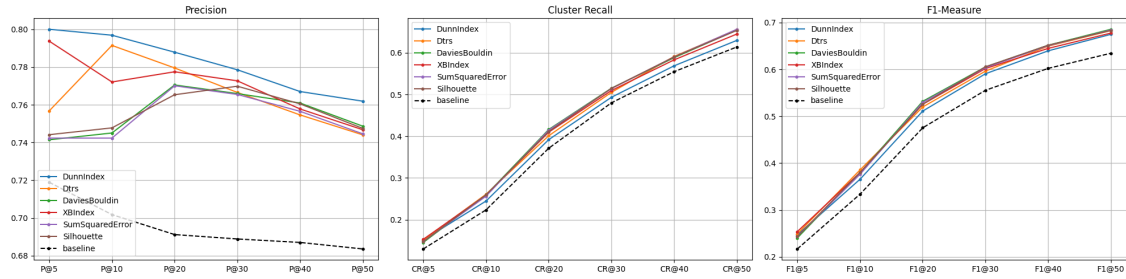


(b) Comportamento dos resultados experimento 03: K-Medoids.

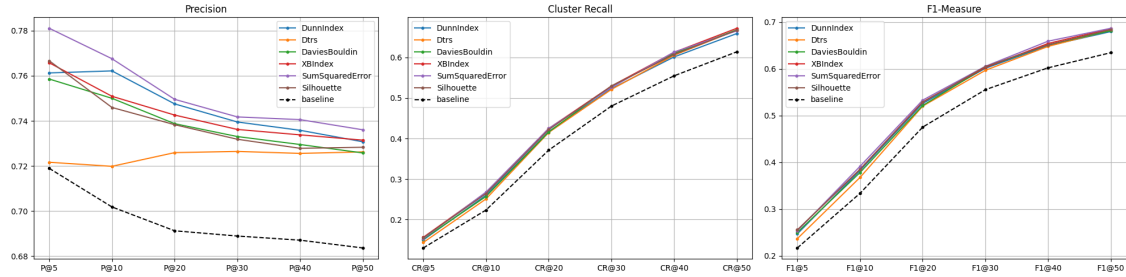
Figura 4.32: Performance: descritor COSINE. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Cluster-Recall (CR)					F1-Measure (F1)									
			K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
experimento 02	5	0.7508	0.7189	0.8000	0.7414	0.7937	0.7423	0.7441	0.1509	0.1298	0.1512	0.1453	0.1529	0.1475	0.1473	0.2491	0.2169	0.2470	0.2403	0.2540	0.2436	0.2434
	10	0.7914	0.7018	0.7968	0.7450	0.7721	0.7423	0.7477	0.2609	0.2233	0.2446	0.2590	0.2589	0.2561	0.2594	0.3855	0.3338	0.3656	0.3795	0.3799	0.3762	0.3800
	20	0.7795	0.6912	0.7878	0.7705	0.7775	0.7700	0.7653	0.4011	0.3712	0.3918	0.4162	0.4097	0.4147	0.4113	0.5198	0.4754	0.5111	0.5319	0.5255	0.5297	0.5260
	30	0.7665	0.6889	0.7785	0.7659	0.7727	0.7655	0.7698	0.5052	0.4801	0.4934	0.5151	0.5096	0.5142	0.5149	0.5963	0.5555	0.5907	0.6052	0.6018	0.6038	0.6055
	40	0.7546	0.6870	0.7670	0.7609	0.7578	0.7565	0.7605	0.5914	0.5549	0.5689	0.5883	0.5832	0.5902	0.5902	0.6496	0.6022	0.6396	0.6516	0.6448	0.6505	0.6515
	50	0.7441	0.6836	0.7619	0.7486	0.7468	0.7446	0.7475	0.6556	0.6141	0.6299	0.6547	0.6451	0.6565	0.6538	0.6828	0.6345	0.6748	0.6856	0.6775	0.6841	0.6831
experimento 03	5	0.7216	0.7189	0.7613	0.7586	0.7658	0.7811	0.7667	0.1432	0.1298	0.1494	0.1515	0.1560	0.1530	0.1551	0.2367	0.2169	0.2471	0.2501	0.2554	0.2536	0.2554
	10	0.7198	0.7018	0.7622	0.7500	0.7509	0.7676	0.7459	0.2507	0.2233	0.2599	0.2569	0.2665	0.2677	0.2639	0.3677	0.3338	0.3826	0.3782	0.3851	0.3915	0.3851
	20	0.7259	0.6912	0.7475	0.7387	0.7426	0.7495	0.7383	0.4148	0.3712	0.4172	0.4160	0.4247	0.4248	0.4219	0.5203	0.4754	0.5260	0.5215	0.5294	0.5330	0.5284
	30	0.7264	0.6889	0.7395	0.7330	0.7362	0.7417	0.7318	0.5213	0.4801	0.5230	0.5297	0.5295	0.5261	0.5295	0.5968	0.5555	0.6012	0.6032	0.6030	0.6054	0.6048
	40	0.7256	0.6870	0.7358	0.7295	0.7338	0.7405	0.7278	0.6056	0.5549	0.6016	0.6101	0.6129	0.6136	0.6076	0.6479	0.6022	0.6494	0.6514	0.6542	0.6591	0.6513
	50	0.7262	0.6836	0.7307	0.7258	0.7314	0.7360	0.7283	0.6677	0.6141	0.6592	0.6685	0.6723	0.6659	0.6674	0.6824	0.6345	0.6796	0.6814	0.6859	0.6867	0.6841

Tabela 4.33: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CNN_AD (Ionescu et al., 2014)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

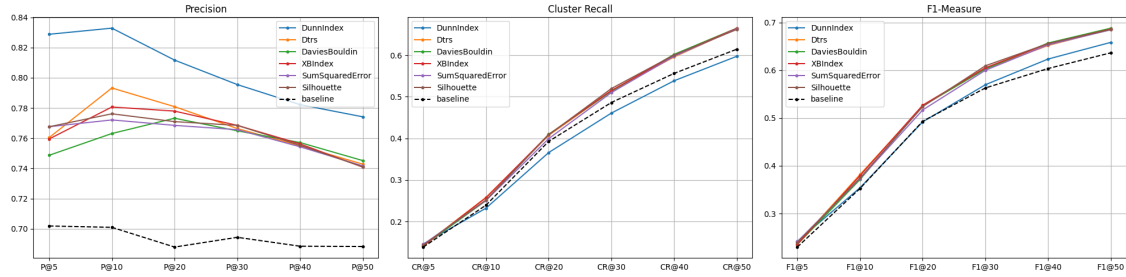


(b) Comportamento dos resultados experimento 03: K-Medoids.

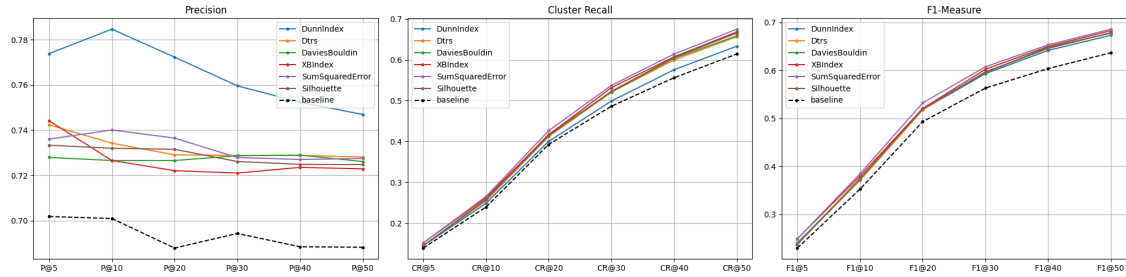
Figura 4.33: Performance: descritor CNN_AD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision				Silhouette	Cluster-Recall (CR)						DTRS	K=20	F1-Measure (F1)				SSE	Silhouette
				Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB	SSE			Dunn	DB	XB	SSE		
experimento 02	5	0.7604	0.7018	0.8288	0.7486	0.7595	0.7676	0.7676	0.1424	0.1390	0.1464	0.1418	0.1414	0.1453	0.1439	0.2366	0.2299	0.2419	0.2352	0.2354	0.2412	0.2394
	10	0.7932	0.7009	0.8329	0.7631	0.7806	0.7721	0.7761	0.2581	0.2395	0.2323	0.2505	0.2575	0.2509	0.2520	0.3818	0.3526	0.3540	0.3713	0.3795	0.3731	0.3744
	20	0.7809	0.6878	0.8117	0.7732	0.7779	0.7685	0.7709	0.4073	0.3928	0.3659	0.4098	0.4090	0.3993	0.4084	0.5254	0.4933	0.4922	0.5258	0.5271	0.5170	0.5241
	30	0.7664	0.6943	0.7955	0.7649	0.7685	0.7656	0.7685	0.5135	0.4862	0.4608	0.5132	0.5150	0.5009	0.5197	0.6033	0.5632	0.5699	0.6028	0.6054	0.6001	0.6096
	40	0.7548	0.6884	0.7822	0.7571	0.7562	0.7543	0.7553	0.5956	0.5561	0.5384	0.6018	0.5991	0.5971	0.6002	0.6521	0.6038	0.6237	0.6572	0.6557	0.6534	0.6560
	50	0.7428	0.6882	0.7741	0.7451	0.7408	0.7416	0.7410	0.6636	0.6142	0.5971	0.6647	0.6643	0.6629	0.6616	0.6861	0.6367	0.6586	0.6883	0.6865	0.6849	0.6855
experimento 03	5	0.7423	0.7018	0.7739	0.7279	0.7441	0.7360	0.7333	0.1456	0.1390	0.1444	0.1442	0.1515	0.1515	0.1443	0.2388	0.2299	0.2406	0.2386	0.2491	0.2489	0.2372
	10	0.7342	0.7009	0.7847	0.7266	0.7266	0.7401	0.7320	0.2537	0.2395	0.2484	0.2575	0.2621	0.2650	0.2584	0.3707	0.3526	0.3719	0.3736	0.3799	0.3847	0.3750
	20	0.7291	0.6878	0.7723	0.7266	0.7221	0.7365	0.7315	0.4122	0.3928	0.4002	0.4175	0.4164	0.4263	0.4142	0.5178	0.4933	0.5182	0.5206	0.5208	0.5324	0.5195
	30	0.7288	0.6943	0.7506	0.7287	0.7210	0.7279	0.7261	0.5204	0.4862	0.4989	0.5213	0.5318	0.5373	0.5228	0.5975	0.5632	0.5933	0.5959	0.6025	0.6076	0.5964
	40	0.7288	0.6884	0.7524	0.7289	0.7235	0.7270	0.7249	0.5996	0.5561	0.5757	0.6036	0.6072	0.6140	0.6056	0.6460	0.6038	0.6416	0.6483	0.6500	0.6532	0.6459
	50	0.7281	0.6882	0.7469	0.7261	0.7229	0.7275	0.7248	0.6564	0.6142	0.6333	0.6585	0.6683	0.6750	0.6653	0.6778	0.6367	0.6731	0.6778	0.6829	0.6859	0.6786

Tabela 4.34: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CN3X3 (Ionescu et al., 2014)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

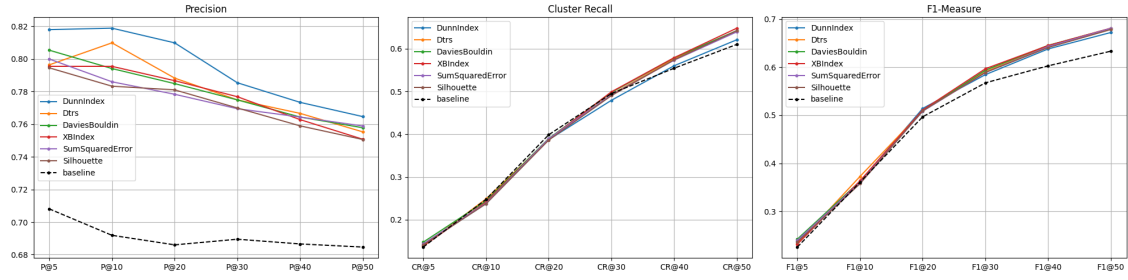


(b) Comportamento dos resultados experimento 03: K-Medoids.

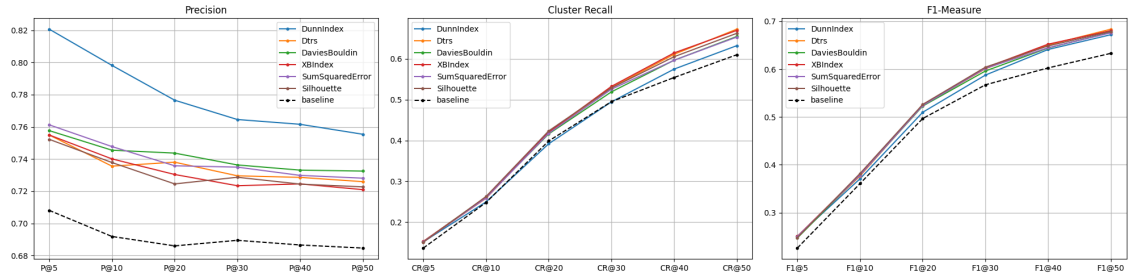
Figura 4.34: Performance: descritor CN3X3. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	Precision					Silhouette	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					Silhouette	
			Dunn	DB	XB	SSE	DTRS		K=20	Dunn	DB	XB	SSE			DTRS	K=20	Dunn	DB	XB		SSE
experimento 02	5	0.7964	0.7081	0.8180	0.8054	0.7955	0.8000	0.7946	0.1389	0.1360	0.1414	0.1476	0.1383	0.1436	0.1411	0.2330	0.2260	0.2379	0.2425	0.2319	0.2401	0.2364
	10	0.8099	0.6919	0.8189	0.7941	0.7955	0.7860	0.7833	0.2485	0.2476	0.2374	0.2438	0.2412	0.2386	0.2372	0.3728	0.3609	0.3623	0.3627	0.3636	0.3600	0.3577
	20	0.7883	0.6860	0.8099	0.7849	0.7867	0.7784	0.7811	0.3875	0.3987	0.3871	0.3899	0.3883	0.3892	0.3862	0.5094	0.4969	0.5141	0.5087	0.5110	0.5084	0.5091
	30	0.7749	0.6895	0.7854	0.7749	0.7769	0.7695	0.7700	0.4931	0.4951	0.4793	0.4978	0.4991	0.4923	0.4903	0.5906	0.5674	0.5844	0.5937	0.5968	0.5888	0.5889
	40	0.7667	0.6866	0.7734	0.7644	0.7628	0.7644	0.7590	0.5760	0.5543	0.5601	0.5786	0.5790	0.5731	0.5740	0.6441	0.6027	0.6378	0.6453	0.6448	0.6428	0.6408
	50	0.7553	0.6847	0.7647	0.7578	0.7598	0.7589	0.7506	0.6409	0.6101	0.6209	0.6424	0.6481	0.6394	0.6423	0.6789	0.6332	0.6720	0.6805	0.6810	0.6808	0.6782
experimento 03	5	0.7550	0.7081	0.8207	0.7577	0.7550	0.7613	0.7523	0.1520	0.1360	0.1512	0.1520	0.1522	0.1513	0.1502	0.2504	0.2260	0.2504	0.2470	0.2505	0.2501	0.2478
	10	0.7356	0.6919	0.7982	0.7455	0.7401	0.7477	0.7378	0.2580	0.2476	0.2491	0.2611	0.2607	0.2587	0.2624	0.3768	0.3609	0.3709	0.3791	0.3797	0.3786	0.3820
	20	0.7381	0.6860	0.7766	0.7437	0.7304	0.7358	0.7245	0.4199	0.3987	0.3918	0.4159	0.4201	0.4175	0.4233	0.5246	0.4969	0.5100	0.5229	0.5255	0.5241	0.5264
	30	0.7296	0.6895	0.7646	0.7363	0.7234	0.7350	0.7287	0.5295	0.4951	0.4946	0.5191	0.5325	0.5256	0.5287	0.6009	0.5674	0.5877	0.5963	0.6031	0.6018	0.6042
	40	0.7286	0.6866	0.7616	0.7331	0.7245	0.7298	0.7244	0.6112	0.5543	0.5749	0.5966	0.6147	0.5966	0.6053	0.6507	0.6027	0.6412	0.6444	0.6524	0.6450	0.6488
	50	0.7259	0.6847	0.7554	0.7325	0.7210	0.7281	0.7227	0.6729	0.6101	0.6324	0.6550	0.6699	0.6535	0.6627	0.6837	0.6332	0.6721	0.6770	0.6806	0.6760	0.6789

Tabela 4.35: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor CEDD (Chatzichristofis e Boutalis, 2008a)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

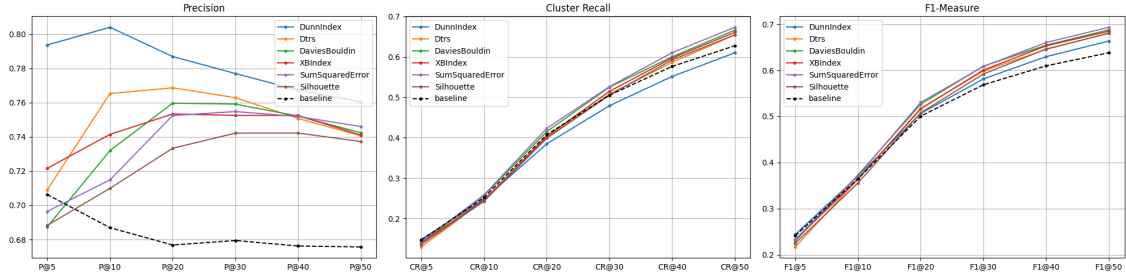


(b) Comportamento dos resultados experimento 03: K-Medoids.

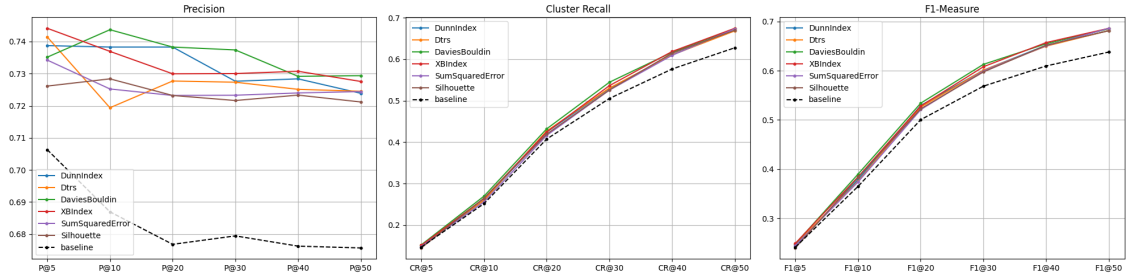
Figura 4.35: Performance: descritor CEDD. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision					DTRS	K=20	Cluster-Recall (CR)					DTRS	K=20	F1-Measure (F1)					DTRS	K=20	SSE	Silhouette
				Dunn	DB	XB	SSE	Silhouette			Dunn	DB	XB	SSE	Silhouette			Dunn	DB	XB	SSE	Silhouette				
experimento 02	5	0.7090	0.7063	0.7937	0.6874	0.7216	0.6964	0.6883	0.1296	0.1468	0.1482	0.1396	0.1400	0.1419	0.1358	0.2168	0.2412	0.2440	0.2297	0.2324	0.2334	0.2247				
	10	0.7653	0.6869	0.8041	0.7320	0.7414	0.7149	0.7099	0.2449	0.2517	0.2474	0.2568	0.2451	0.2584	0.2426	0.3641	0.3649	0.3704	0.3732	0.3638	0.3709	0.3557				
	20	0.7687	0.6768	0.7869	0.7597	0.7534	0.7525	0.7333	0.4006	0.4074	0.3847	0.4151	0.4033	0.4231	0.3986	0.5176	0.5005	0.5064	0.5265	0.5164	0.5305	0.5080				
	30	0.7629	0.6794	0.7770	0.7592	0.7526	0.7548	0.7422	0.5079	0.5053	0.4792	0.5252	0.5147	0.5263	0.5054	0.5988	0.5686	0.5814	0.6092	0.6005	0.6084	0.5920				
	40	0.7508	0.6762	0.7673	0.7521	0.7525	0.7519	0.7422	0.5864	0.5764	0.5516	0.6002	0.5970	0.6104	0.5924	0.6453	0.6097	0.6296	0.6546	0.6531	0.6603	0.6461				
	50	0.7406	0.6757	0.7605	0.7424	0.7411	0.7460	0.7373	0.6545	0.6276	0.6101	0.6664	0.6609	0.6734	0.6546	0.6810	0.6380	0.6634	0.6878	0.6848	0.6934	0.6802				
experimento 03	5	0.7414	0.7063	0.7387	0.7351	0.7441	0.7342	0.7261	0.1514	0.1468	0.1479	0.1523	0.1510	0.1482	0.1448	0.2472	0.2412	0.2429	0.2481	0.2491	0.2447	0.2398				
	10	0.7194	0.6869	0.7383	0.7437	0.7369	0.7252	0.7284	0.2597	0.2517	0.2600	0.2697	0.2641	0.2557	0.2650	0.3750	0.3649	0.3786	0.3898	0.3841	0.3739	0.3829				
	20	0.7277	0.6768	0.7383	0.7383	0.7300	0.7232	0.7232	0.4267	0.4074	0.4241	0.4321	0.4244	0.4163	0.4193	0.5261	0.5005	0.5294	0.5341	0.5283	0.5210	0.5226				
	30	0.7273	0.6794	0.7276	0.7374	0.7300	0.7233	0.7216	0.5304	0.5053	0.5295	0.5445	0.5363	0.5255	0.5259	0.6010	0.5686	0.6010	0.6130	0.6081	0.5995	0.5975				
	40	0.7251	0.6762	0.7284	0.7292	0.7307	0.7240	0.7233	0.6104	0.5764	0.6108	0.6169	0.6187	0.6094	0.6145	0.6499	0.6097	0.6517	0.6548	0.6572	0.6507	0.6518				
	50	0.7245	0.6757	0.7239	0.7294	0.7276	0.7244	0.7212	0.6684	0.6276	0.6693	0.6736	0.6747	0.6734	0.6700	0.6816	0.6380	0.6814	0.6853	0.6862	0.6863	0.6814				

Tabela 4.36: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor BM25 (Baeza-Yates et al., 1999)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.

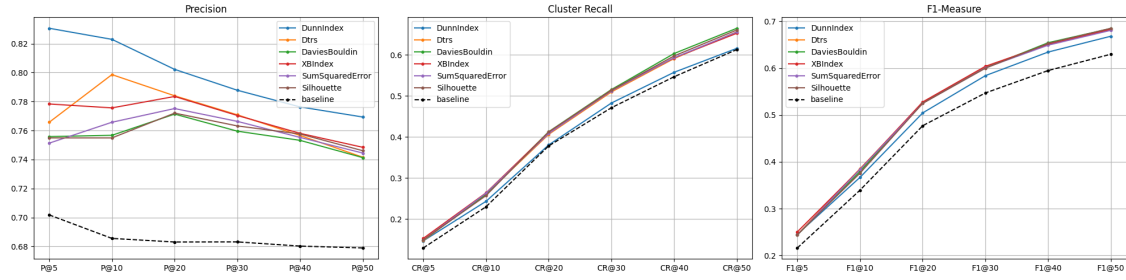


(b) Comportamento dos resultados experimento 03: K-Medoids.

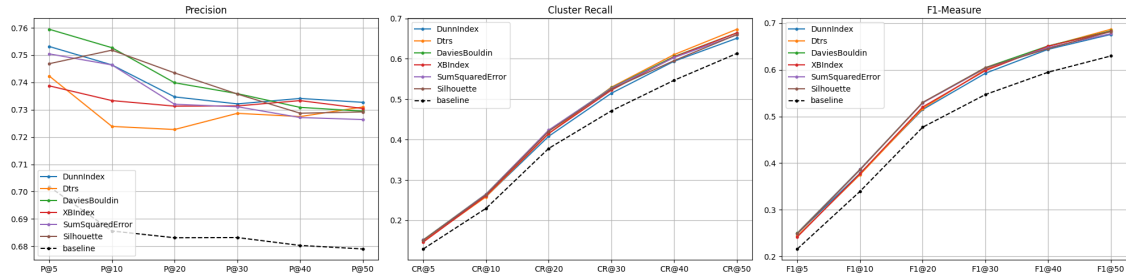
Figura 4.36: Performance: descritor BM25. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

	n	DTRS	K=20	Precision				SSE	Silhouette	Cluster-Recall (CR)						DTRS	K=20	F1-Measure (F1)				SSE	Silhouette
				Dunn	DB	XB				DTRS	K=20	Dunn	DB	XB				Dunn	DB	XB			
experimento 02	5	0.7658	0.7018	0.8306	0.7559	0.7784	0.7514	0.7550		0.1471	0.1293	0.1466	0.1483	0.1527	0.1485	0.1490	0.2444	0.2161	0.2453	0.2436	0.2504	0.2438	0.2445
	10	0.7986	0.6856	0.8230	0.7568	0.7757	0.7658	0.7550		0.2005	0.2297	0.2429	0.2596	0.2636	0.2627	0.2569	0.3853	0.3396	0.3668	0.3795	0.3841	0.3838	0.3759
	20	0.7840	0.6831	0.8023	0.7714	0.7836	0.7752	0.7721		0.4064	0.3782	0.3806	0.4128	0.4116	0.4083	0.4124	0.5247	0.4771	0.5045	0.5267	0.5279	0.5249	0.5251
	30	0.7709	0.6832	0.7878	0.7596	0.7704	0.7664	0.7632		0.5098	0.4713	0.4828	0.5156	0.5142	0.5123	0.5138	0.6018	0.5471	0.5841	0.6008	0.6040	0.6007	0.5999
	40	0.7566	0.6803	0.7764	0.7533	0.7581	0.7553	0.7577		0.5912	0.5468	0.5577	0.6035	0.5928	0.5931	0.5976	0.6501	0.5950	0.6344	0.6544	0.6508	0.6491	0.6528
	50	0.7416	0.6790	0.7694	0.7414	0.7485	0.7445	0.7461		0.6565	0.6131	0.6161	0.6651	0.6530	0.6560	0.6609	0.6821	0.6298	0.6682	0.6848	0.6830	0.6808	0.6852
experimento 03	5	0.7423	0.7018	0.7532	0.7505	0.7387	0.7505	0.7468		0.1481	0.1293	0.1512	0.1518	0.1462	0.1513	0.1511	0.2439	0.2161	0.2490	0.2503	0.2419	0.2496	0.2489
	10	0.7239	0.6856	0.7464	0.7527	0.7333	0.7464	0.7518		0.2584	0.2297	0.2589	0.2651	0.2607	0.2648	0.2634	0.3754	0.3396	0.3783	0.3868	0.3775	0.3866	0.3856
	20	0.7227	0.6831	0.7347	0.7399	0.7313	0.7320	0.7435		0.4148	0.3782	0.4082	0.4226	0.4152	0.4235	0.4206	0.5173	0.4771	0.5150	0.5298	0.5205	0.5295	0.5303
	30	0.7287	0.6832	0.7321	0.7359	0.7314	0.7311	0.7357		0.5298	0.4713	0.5148	0.5281	0.5224	0.5272	0.5252	0.6006	0.5471	0.5927	0.6048	0.5985	0.6032	0.6046
	40	0.7275	0.6803	0.7341	0.7309	0.7333	0.7271	0.7287		0.6104	0.5468	0.5939	0.6055	0.6044	0.6030	0.5955	0.6504	0.5950	0.6440	0.6512	0.6508	0.6478	0.6455
	50	0.7310	0.6790	0.7327	0.7295	0.7305	0.7264	0.7292		0.6731	0.6131	0.6508	0.6642	0.6640	0.6592	0.6598	0.6870	0.6298	0.6759	0.6836	0.6824	0.6775	0.6821

Tabela 4.37: Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com $k = [1, 150]$ - descritor ACC (Penatti et al., 2012)



(a) Comportamento dos resultados experimento 02: Hierárquico Aglomerativo.



(b) Comportamento dos resultados experimento 03: K-Medoids.

Figura 4.37: Performance: descritor ACC. Experimentos 02 e 03, $k = [1, 150]$ e uso do IASMIN.

4.2.2 Avaliação dos ganhos alcançados

Nesta seção faremos a avaliação da eficácia, em termos de relevância e diversidade, nos resultados dos experimentos 02 e 03 executados com o uso do índice de validação DTRS e do método auxiliar IASMIN e detalhados na seção 4.2.1.

Experimento 02 (algoritmo Hierárquico Aglomerativo - $k = [1, 150]$)

Assim como no experimento inicial, a abordagem detalhada na seção 3.1 demonstrou superioridade em relação ao *baseline* em termos de relevância e diversidade, bem como na média harmônica (*F1-Measure*), para todos os índices de validação de agrupamento (exceto nos descritores JCD, JACCARD, FCTH, CEDD e BM25, em relação ao método DTRS). No entanto, observou-se um ganho relativo menor no experimento atual (8,06%) em comparação com o anterior (9,7%), que utilizava o K-Medoids. Para as métricas *Precision*, *Cluster Recall* e *F1-Measure*, tivemos ganho relativo médio de 10,37%, 5,76% e 7,12%, respectivamente. Quanto ao método DTRS, os resultados concentraram-se em torno dos valores médios para os métodos de validação em todos os descritores do estudo, com ganho médio relativo de 8,16% quando em comparação ao *baseline*.

O ganho médio relativo geral do DTRS em relação aos demais índices de validação de agrupamento foi calculado em 0,38% na métrica *Precision*, -0,047% na *Cluster Recall* e 0,039% para a métrica *F1-Measure*. Com ganhos individualizados em torno de 1,0% frente aos métodos Davies Bouldin, Xie Beni, SSE e Silhouette. Podemos destacar, também a superioridade do DTRS frente aos demais métodos através da observação individualizada das representações gráficas (Figuras 4.23(a) a 4.37(a)) e tabelas de resultados, nos descritores TAMURA, JCD, JACCARD, GABOR, FCTH, COSINE, CNN_AD, CN3X3, CEDD e ACC, em valores máximos, para as faixas de ranking descritas na Tabela 4.38.

	Descritor	P@n	CR@n	F1@n
visual	TAMURA	20		
	JCD			10
	GABOR		10	10
	FCTH			10
	CNN_AD		10, 40	10
	CN3X3		10	10
	CEDD		10	10
	ACC			10
text	COSINE	5, 10		
	JACCARD	5		

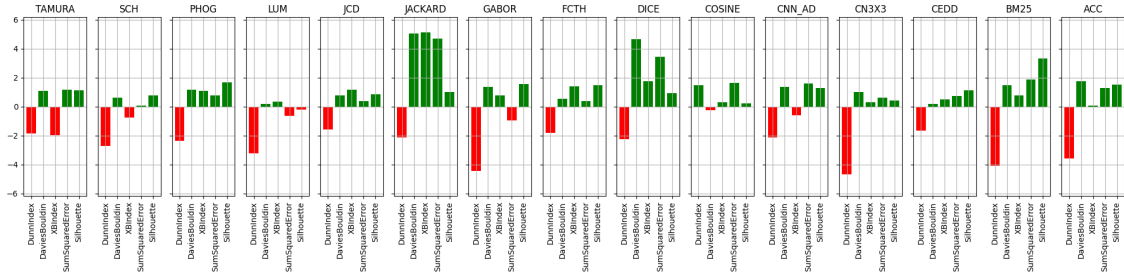
Tabela 4.38: Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem considerar o teste de significância estatística. Baseado nas observações dos gráficos das Figuras 4.23(a) a 4.37(a) e tabelas de resultado do experimento 02.

Os resultados com a aplicação do teste estatístico (Tabelas 4.23 a 4.37, experimento 02, algoritmo Hierárquico Aglomerativo) demonstraram superioridade estatística do DTRS nos casos específicos de acordo com a Tabela 4.39.

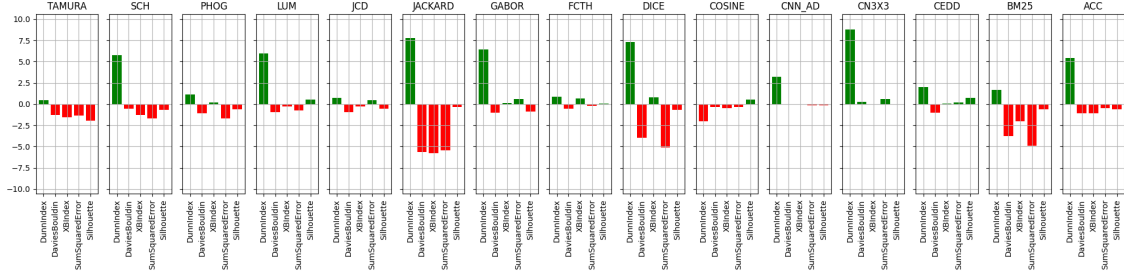
DESCRIPTOR	Precision	Cluster Recall	F1-Measure
SCH	-	Dunn Index	Dunn Index
PHOG	Davies Bouldin	-	-
LUM	-	Dunn Index	Dunn Index
JACCARD	Davies Bouldin, Xie Beni, SSE, Silhouette	Dunn Index	Dunn Index
FCTH	Xie Beni, Silhouette	-	-
DICE	Davies Bouldin, Xie Beni, SSE	Dunn Index	Dunn Index
COSINE	Dunn Index	-	-
CNN_AD	-	-	Dunn Index
BM25	Silhouette	-	-
ACC	Davies Bouldin	Dunn Index	-

Tabela 4.39: Superioridade do método DTRS por descritor

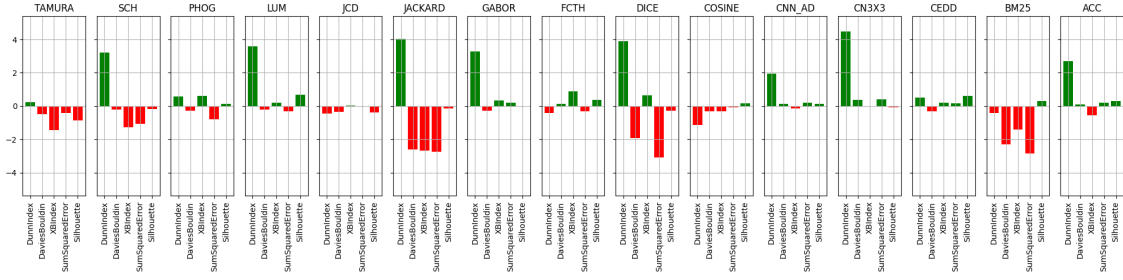
A Figura 4.38 mostra o ganho relativo para o método DTRS frente aos demais índices de validação de agrupamento nas métricas *Precision* e *Cluster Recall* em todos os descritores do estudo.



(a) Experimento 02: ganho médio do método DTRS, métrica *Precision*.



(b) Experimento 02: ganho médio do método DTRS, métrica *Cluster Recall*.



(c) Experimento 02: ganho médio do método DTRS, métrica *F1-Measure*.

Figura 4.38: Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para as métricas *Precision*, *Cluster Recall* e *F1-Measure* (todas as faixas de ranking). Em verde, superioridade do DTRS. Experimento 02 (algoritmo Hierárquico Aglomerativo).

Imagens reais deste experimento para o descritor JCD podem ser vistas no Apêndice A, Subseção A.0.2.

Experimento 03 (algoritmo K-Medoids - $k = [1, 150]$)

Os resultados do experimento 03 também demonstraram superioridade em relação ao *baseline* em todas as métricas para todos os descritores e métodos de *cluster validity* do estudo. O ganho médio relativo geral ficou em 7,07%, um pouco abaixo do experimento anterior. As métricas *Precision*, *Cluster Recall* e *F1-Measure* apresentaram ganho médio relativo de 6,61%, 8,59% e 7,88%, respectivamente, com uma perda de cerca de 3% de ganho na relevância em relação ao experimento 02.

Já o método DTRS teve ganho médio relativo de 7,11% na comparação com o *baseline*, bem próximo do ganho médio geral do experimento (ganho médio nas métricas *Precision*, *Cluster Recall* e *F1-Measure* em 5,57%, 8,65% e 7,6%, respectivamente). Em relação aos demais índices de validação de agrupamento foi calculado o ganho médio relativo geral do DTRS em -1,32% na métrica *Precision*, 0,074% na *Cluster Recall* e -0,37% para a métrica *F1-Measure*. Com ganho médio individualizado de 2,0% frente ao método Dunn Index na métrica *Cluster Recall*. Contudo, podemos destacar a superioridade do DTRS frente aos demais métodos de forma individualizada, de acordo com a observação das representações gráficas (Figuras 4.23(a) a 4.37(a)) e tabelas de resultados, nos descritores TAMURA, JCD, JACCARD, GABOR, FCTH, COSINE, CNN_AD, CN3X3, CEDD e ACC, em valores máximos, para as faixas de ranking descritas na Tabela 4.38.

Tabela 4.40: Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos gráficos 4.23(a) a 4.37(a) e tabelas de resultado do experimento.

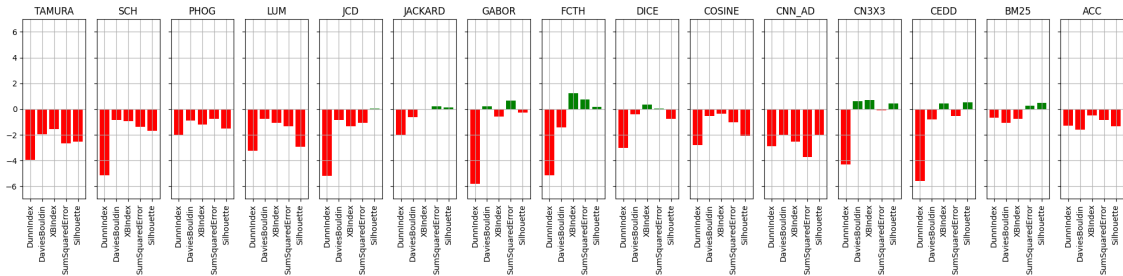
Descritor		P@n	CR@n	F1@n
visual	LUM		20, 40	
	JCD		≥ 30	50
	GABOR		5	10
	FCTH		30	50
	CEDD		50	50
	ACC		≥ 30	50

Os resultados com a aplicação do teste estatístico (Tabelas 4.23 a 4.37, experimento 03, algoritmo K-Medoids) demonstraram superioridade estatística do DTRS nos casos específicos de acordo com a Tabela 4.41.

Tabela 4.41: Superioridade do método DTRS por descritor

DESCRITOR	Precision	Cluster Recall	F1-Measure
PHOG		Dunn Index	
JCD		Dunn Index	
GABOR		Dunn Index	Dunn Index, SSE
FCTH			SSE
CN3X3		Dunn Index	
CEDD		Dunn Index	Dunn Index

A Figura 4.39 mostra a representação gráfica do ganho médio relativo para o método DTRS frente aos demais índices de validação de agrupamento nas métricas *Precision* e *Cluster Recall* em todos os descritores do estudo.



(a) Experimento 03: ganho médio do método DTRS, métrica *Precision*.

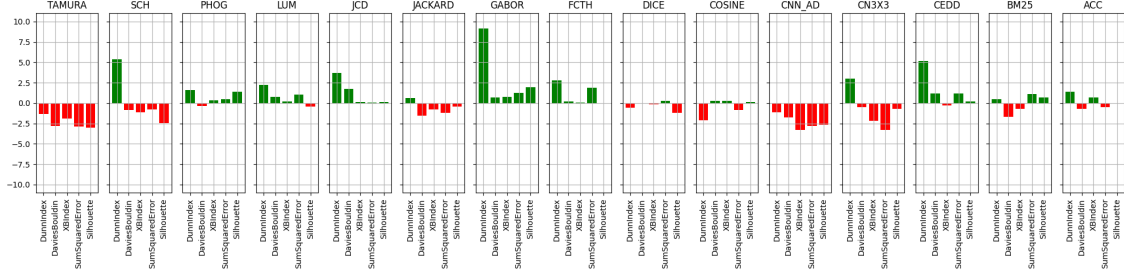
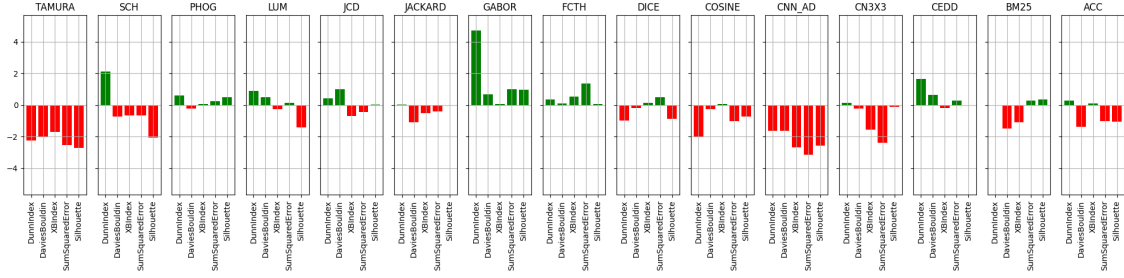
(b) Experimento 03: ganho médio do método DTRS, métrica *Cluster Recall*(c) Experimento 03: ganho médio do método DTRS, métrica *F1-Measure*

Figura 4.39: Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para as métricas *Precision*, *Cluster Recall* e *F1-Measure* (todas as faixas de ranking). Em verde, superioridade do DTRS. Experimento 03 (algoritmo K-Medoids).

4.2.3 Avaliação do tempo de execução

O objetivo principal desta etapa do estudo foi a redução do custo computacional associado ao processo de recuperação de imagens baseado em conteúdo. Normalmente, sistemas CBIR possuem um custo computacional elevado. A adoção de medidas para otimização desse custo, como a pré-filtragem ou o uso das matrizes de distância pré-calculadas, vistos no arcabouço apresentado no trabalho relacionado (Figuerêdo e Calumby, 2022), podem não ser medidas suficientes. Tal fato motivou a implementação de um método, intercalado na etapa de identificação dos padrões, que propiciasse a interrupção do algoritmo de aprendizado não-supervisionado dinamicamente, ou seja, sem a interferência humana, reduzindo o seu número de iterações.

Esse método, ou critério, de interrupção (IASMIN), detalhado na Seção 3.2.2, baseou-se na análise da monotonicidade da curva gerada pelos valores do coeficiente angular extraídos da reta secante traçada entre os pontos analíticos inicial e final ($k_{inicial}$ e k_{atual}) da curva traçada pelos valores da média móvel gerada pelos resultados de um índice de validação de agrupamento qualquer. Esta curva da média móvel de intervalo t (de acordo com a Tabela 3.3), como dito anteriormente, foi gerada a partir dos valores resultantes extraídos dos índices de validação de agrupamento utilizados no estudo, individualmente, para cada um dos 15 descritores usados na abordagem descrita na Seção 3.1.

Nas Figuras 4.40 e 4.41 podemos ver o comportamento dos índices de validação do método DTRS para o descritor ACC em relação à *Query ID = 1* (*angel_of_the_north*). No gráfico 4.40, temos a curva de risco gerada pelo método DTRS (azul) em conjunto com a média móvel (vermelho) e a curva do coeficiente angular da reta secante à média móvel (azul), curvas achatadas. O número de grupos ideal é representado pela reta vertical traçada no eixo das abscisas em $k = 147$.

A Figura 4.41(a) apresenta o traçado da média móvel, individualizado, baseado nos valores da curva de risco do método DTRS. Foi utilizado um intervalo igual a 10 pontos neste caso, conforme parametrizado para este experimento (Tabela 3.3). A curva gerada pelo coeficiente angular da reta secante aos pontos de teste da média móvel pode ser vista na Figura 4.41(b). Aqui podemos identificar o ponto de inflexão negativo no qual a interrupção do algoritmo foi alcançada. Neste ponto foram coletados o tempo de execução e o número k de grupos ideal anteriores ao ponto de inflexão. O tempo de execução ficou em 4151829 nanosegundos (0,04151829 segundos) para a amostra do tópico de consulta da Figura 4.41(b).

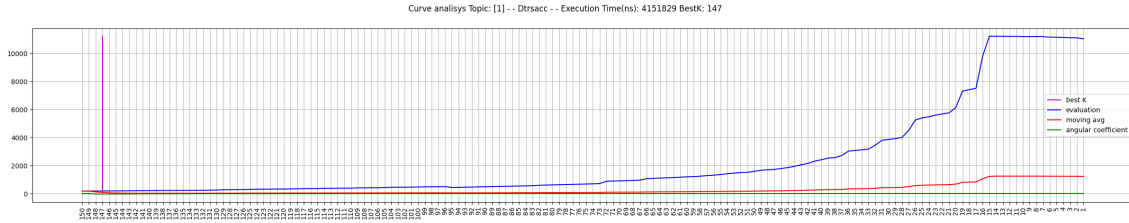
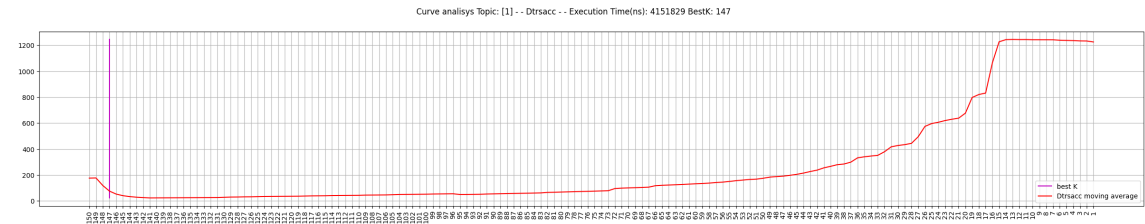
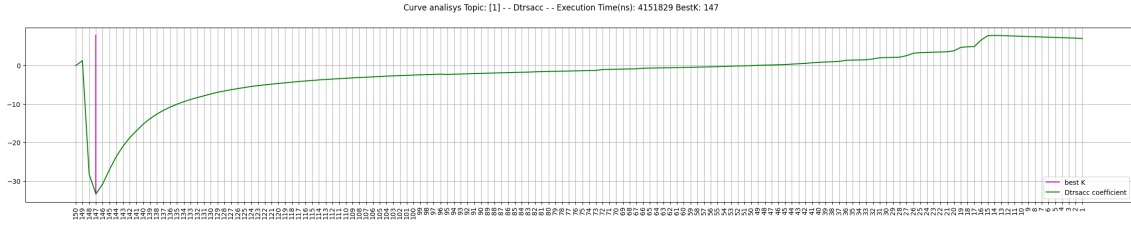


Figura 4.40: Experimento 02: curva gerada pelos resultados do método DTRS (curva de risco em azul) para o tópico de consulta *angel_of_the_north*. Achatamento com demais curvas de média móvel (vermelho) e coeficiente angular (verde). Descritor ACC. O número de grupos ideal ($k = 147$) foi encontrado em 4151829 nanosegundos (0,04151829 segundos).



(a) Curva da média móvel dos resultados do método DTRS (intervalo $t = 10$ pontos) para o tópico de consulta *angel_of_the_north*.



(b) Curva com os valores do coeficiente angular da reta secante aos pontos inicial e de teste baseados na média móvel. O corte em $k = 147$ correspondeu de forma assertiva com a estratégia de interrupção do algoritmo (inflexão negativa).

Figura 4.41: Experimento 02: comportamento da média móvel e do coeficiente angular da reta secante à média móvel para valores da curva de risco gerada pelo método DTRS. Tópico $ID = 1$ (*angle_of_the_north*) e descritor ACC. A interrupção do algoritmo logo nos primeiros esquemas de *cluster* fez com que o tempo de execução do sistema CBIR fosse o menor possível.

O tempo de execução do sistema CBIR foi tomado entre as etapas de pré-filtragem das imagens e a validação do agrupamento entregue pelo algoritmo de aprendizado não-supervisionado, com a consequente seleção do *best K*, de acordo com os critérios vistos na Seção 3.2.2. É importante destacar que o tempo de execução, por etapa, em um algoritmo Hierárquico Aglomerativo, diferentemente do K-Medoids, diminui à medida em que ele se aproxima do k final. Assim, o tempo necessário para que o sistema CBIR encontre um esquema de agrupamentos próximo do k inicial ($k = 150$, no caso do Aglomerativo) é bem mais elevado do que o tempo de um esquema encontrado em $k = 3$, por exemplo, neste tipo de algoritmo. Sendo, portanto, o tempo total de execução do sistema CBIR obtido pela soma das etapas do agrupamento, ou seja: $t_{total} = t_{k_i} + t_{k_{i-1}} + \dots + t_{bestK} \dots + t_{k_f}$, em nanosegundos, com k inicial de acordo com a Tabela 3.3, onde t_{bestK} foi o tempo de execução coletado para apresentação.

Para efeito comparativo, foram realizadas mais duas execuções dos experimentos 02 e 03, além daquela feita com o uso do IASMIN, para a coleta do tempo médio de execução do sistema CBIR. A primeira, sem o uso de método auxiliar para seleção do *best K*, ou seja, apenas a forma natural de cada índice de validação de agrupamento; e a segunda, com uso do método auxiliar do Cotovelo. Os resultados desta última avaliação (para o descritor ACC) podem ser vistos nas Tabelas 4.42 e 4.43.

Tempo de execução - experimento 02

Assim, na Tabela 4.42 temos o tempo médio de execução, transformado para milissegundos, do sistema CBIR em 222 tópicos de consulta usando o descritor visual ACC para cada um dos índices de validação de agrupamento avaliados no experimento 02 (Hierárquico Aglomerativo com $k = [1, 150]$). Neste caso, o método DTRS em conjunto com o IASMIN, no seu pior caso, alcançou uma redução de 64,3% no tempo médio de execução (em relação ao método SSE). No melhor caso, o DTRS alcançou 95,6% de redução do tempo médio de execução (em relação ao *Dunn Index*), considerando o mesmo método auxiliar (IASMIN). Em comparação ao método do

Cotovelo, o IASMIN conseguiu uma redução média no tempo de execução em torno de 71,1% para o algoritmo Hierárquico Aglomerativo.

Contudo, não tivemos redução no tempo médio de execução do sistema CBIR quando comparamos os valores do IASMIN com aqueles obtidos pelo uso da forma natural do *cluster validity*. De maneira oposta ao experimento anterior, tivemos um aumento significativo no valor médio das medições realizadas para o tempo de execução. A exceção ficou por conta do método Silhouette, cujo tempo médio foi reduzido em cerca de 84,3% com uso do IASMIN. Já para o DTRS, com uso do método natural de cada *cluster validity*, o tempo médio de execução ficou muito próximo de zero. Tal fato pode ser um indicativo de que houve interrupção (apenas a contagem do tempo, neste caso) precoce do sistema CBIR, ou seja, a seleção do número de grupos ideal ocorreu logo nas primeiras iterações do algoritmo Aglomerativo, na maioria das consultas submetidas (interrupção no k inicial).

	Tempo médio de execução (ms)		
	apenas índice de validação	Elbow + índice	IASMIN + índice
DB	2,58	139,11	14,29
DTRS	0,00*	102,43	4,81
Dunn	3,21	121,79	110,36
SSE	0,19	122,89	13,60
Silhouette	97,02	125,77	15,27
XB	1,04	112,69	51,21

Tabela 4.42: Experimento 02: Hierárquico Aglomerativo, $k = [1, 150]$. Tempo médio de execução (milissegundos) para todos os descritores e 222 tópicos de consulta. Comparativo entre os índices de validação sem uso de método auxiliar (apenas a forma natural do *cluster validity*) e os métodos auxiliares para seleção do número k de grupos ideal (Método do Cotovelo e IASMIN). O valor 0,00* no DTRS indica que o índice encontrou o número de grupos ideal precocemente, ou seja, no início das iterações do algoritmo.

Tempo de execução - experimento 03

Quando a observação é feita para o experimento 03 (com uso do algoritmo K-Medoids em conjunto com o IASMIN), Tabela 4.43, temos que o tempo médio de execução do sistema CBIR foi significativamente menor que aquele observado para o algoritmo Hierárquico Aglomerativo (experimento 02) em todos os casos. A redução no tempo médio de execução do sistema usando o K-Medoids alcançou 99,35% no melhor caso (Xie Beni) e 97,54% no pior caso (Silhouette), frente ao experimento usando o Hierárquico Aglomerativo com o mesmo método auxiliar (IASMIN).

A mesma observação pode ser feita com as execuções adicionais com a forma natural do índice de validação de agrupamento e com o método do Cotovelo. Nestas duas últimas situações, o experimento com K-Medoids alcançou 98,27% no melhor caso (Silhouette), 77,63% no pior caso (SSE), para a forma natural do índice de validação, e 98,55% no melhor caso (Silhouette), 97,36% no pior caso (SSE), para o método *Elbow*. Contudo, tivemos uma exceção com o DTRS quando os resultados do K-Medoids foram avaliados pela forma natural deste índice de validação de agrupamento (o tempo médio de execução do DTRS neste caso foi próximo de zero).

Na observação feita entre os índices de validação de agrupamento e os métodos auxiliares para seleção do número de grupos ideal (método do Cotovelo e IASMIN), comparativamente, para o experimento realizado com uso do K-Medoids, tivemos que o IASMIN foi superior ao método do Cotovelo, com tempo médio de execução significativamente menor, em torno de 93,41% para o melhor caso (DTRS), pior caso em 45,39% (Dunn Index), e inferior (tempo médio de execução alto) à forma natural de todos os índices de validação de agrupamento com exceção do Silhouette, onde o IASMIN alcançou um tempo 77,53% menor, similarmente ao que foi visto no experimento anterior.

	Tempo médio de execução (ms)		
	apenas índice de validação	Elbow + índice	IASMIN + índice
DB	0,0729	3,6758	0,3106
DTRS	0,0425	1,7577	0,1158
Dunn	0,6135	1,9391	1,0589
SSE	0,0425	2,7962	0,3004
Silhouette	1,6739	1,8261	0,3762
XB	0,0809	1,7873	0,3332

Tabela 4.43: Experimento 03: K-Medoids, $k = [1, 150]$. Tempo médio de execução (milissegundos) para todos os descritores e 222 tópicos de consulta. Comparativo entre os índices de validação sem uso de método auxiliar e os métodos auxiliares para seleção do número de grupos ideal (Método do Cotovelo e IASMIN).

O baixo desempenho em termos de tempo de execução pelo IASMIN em comparação com a abordagem natural dos índices de validação de agrupamento (sem uso do método auxiliar), conforme evidenciado pelas Tabelas 4.43 e 4.42 (por valores máximos e mínimos), pode ser explicado em virtude desta segunda abordagem ocorrer sem a implementação de qualquer outra heurística, sendo, portanto, o tempo de execução tomado sem a participação de métodos auxiliares como, por exemplo, o IASMIN ou o *Elbow*, medida que aumentaria o tempo de execução total do sistema, já que os índices de validação sempre fazem parte do fluxo de trabalho do sistema de recuperação proposto. Tal fato sugere que a seleção dos melhores esquemas de agrupamento usando a forma natural de cada *cluster validity* ocorre logo nas primeiras iterações do algoritmo de agrupamento. Neste sentido, o tipo de algoritmo utilizado, seja Hierárquico Aglomerativo ou K-Medoids, parece ser irrelevante.

Capítulo 5

Discussão

5.1 O método DTRS e o IASMIN

Neste capítulo discutiremos com mais profundidade os resultados dos experimentos realizados nos termos das questões de pesquisa. Examinaremos os ganhos alcançados e as ameaças à validade deste trabalho. Foram realizados três experimentos principais, sendo: o primeiro, utilizando o algoritmo K-Medoids com um número k de agrupamentos variando entre 15 e 25 *clusters*, com o intuito de testar a performance do método DTRS, no qual a seleção do número de grupos ideal se deu pela forma natural dos índices de validação de agrupamento; o segundo, utilizando Hierárquico Aglomerativo, onde introduzimos o IASMIN para redução do tempo de execução, com $k = [1, 150]$; e o terceiro experimento, usando novamente o algoritmo K-Medoids, nas mesmas condições do segundo experimento, para que pudéssemos comparar a performance do sistema CBIR, possíveis variabilidades do DTRS e o desempenho do IASMIN para os dois tipos de algoritmos de aprendizado não-supervisionado estudados. Os resultados desses três experimentos foram apresentados no capítulo anterior.

Comparação entre os experimentos 01 e 02

De acordo com os gráficos das Figuras 5.1 a 5.3, onde temos o comparativo dos resultados entre os experimentos 01 e 02, em termos de performance média do sistema CBIR, observamos que o experimento utilizando o algoritmo Hierárquico Aglomerativo, fazendo uso do método auxiliar IASMIN e com $k = [1, 150]$, apresentou grande variabilidade nos resultados de eficácia nos índices de validação de agrupamento em comparação ao experimento que utilizou o K-Medoids (com a seleção do *best K* pela forma natural dos índices de validação de agrupamento e $k = [15, 25]$), independentemente do método de avaliação da eficácia. Essa variabilidade pode estar relacionada com a diferença no intervalo de agrupamentos adotado, com o uso do critério de interrupção (IASMIN) e com as diferenças metodológicas entre os dois algoritmos de aprendizado não-supervisionado.

Em relação ao índices de validação de agrupamento, percebemos que o Dunn Index apresentou picos significativos de eficácia na observação feita para o experimento 02, métrica *Precision* (média dos segmentos de ranking $n = 5$ a $n = 50$), de acordo com o gráfico da Figura 5.1. Tal comportamento não foi observado no experimento 01, realizado com uso do K-Medoids. Observações posteriores, feitas nos resultados dos experimentos adicionais, sugerem que esse comportamento do Dunn Index pode estar ligado à abordagem adotada pelo algoritmo Hierárquico Aglomerativo, no nosso caso, *complete linkage*, uma vez que também não foram observadas anomalias na eficácia deste índice de validação nos experimentos adicionais com o K-Medoids.

Já em relação às métricas *Cluster Recall* e *F1-Measure*, o experimento usando o algoritmo Hierárquico Aglomerativo e $k = [1, 150]$ apresentou eficácia inferior ao realizado com o K-Medoids, em intervalo de *clusters* $k = [15, 25]$, em termos médios, para quase todos os índices de validação, conforme podemos ver através dos gráficos das Figuras 5.2 e 5.3. Não foram feitas comparações em relação ao tempo de execução dos experimentos 01 e 02 devido a adoção de diferentes intervalos de k *clusters* nos dois casos.

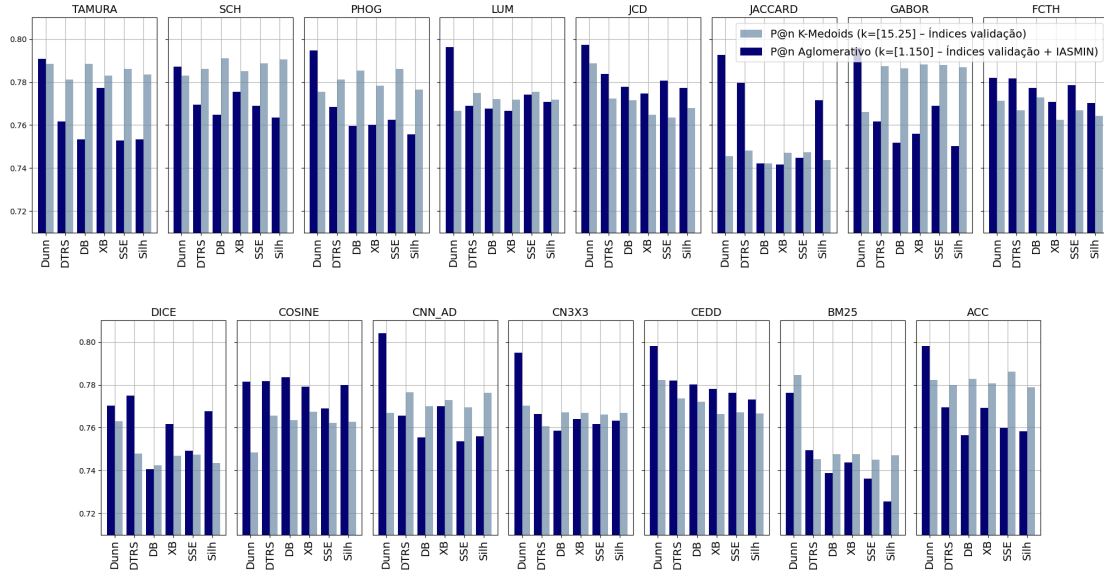


Figura 5.1: Nível de eficácia médio para a métrica *Precision*, comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o TAMURA, SCH, LUM e GABOR.

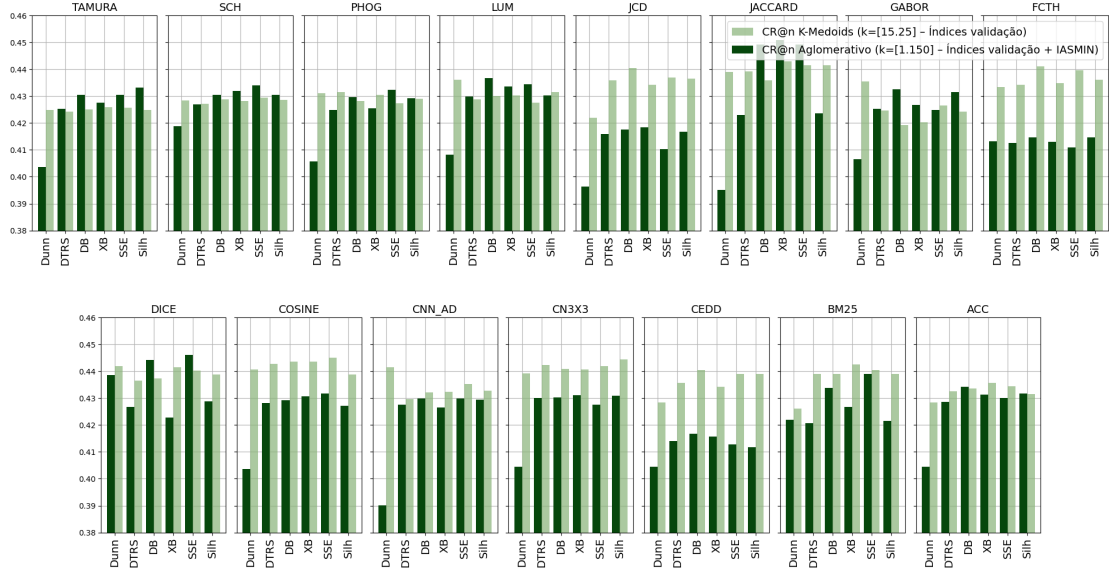


Figura 5.2: Nível de eficácia médio para a métrica *Cluster Recall*, comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o TAMURA, SCH, LUM e GABOR.

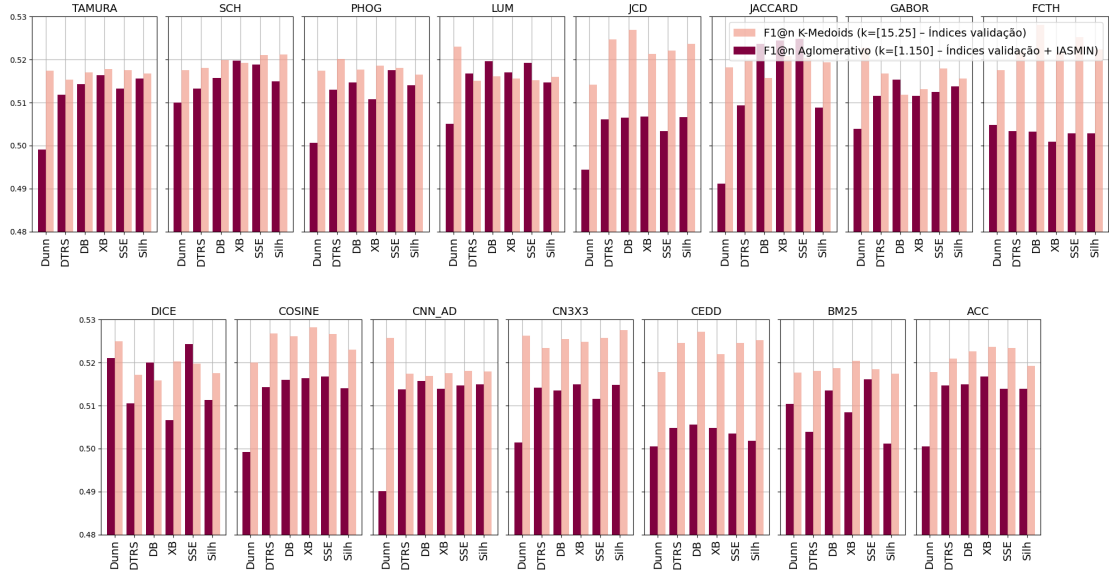


Figura 5.3: Nível de eficácia médio para a métrica *F1-Measure*, comparativo entre os experimentos 01 e 02. Algoritmo K-Medoids com $k = [15, 25]$, usando apenas índices de validação, versus Hierárquico Aglomerativo com $k = [1, 150]$, usando índices de validação e método auxiliar IASMIN (experimentos 01 e 02, respectivamente). Superioridade do DTRS em relação ao experimento 01 para todos os descritores, exceto o LUM.

Comparação entre os experimentos 02 e 03

Quando a comparação é feita em relação aos experimentos 02 e 03, os resultados para o algoritmo Hierárquico Aglomerativo, na média para a métrica *Precision* (Figura 5.4), se mostraram significativamente superiores àqueles obtidos para o K-Medoids. De fato, os valores médios dos resultados entre os dois experimentos, em relação à relevância, aparecem significativamente distantes (0,7688 e 0,7375) de acordo com a análise estatística feita na seção B.0.2. É possível perceber, ainda, que houve grande variabilidade nos dados de performance para o K-Medoids. Sugerindo que o fator primordial dessa variabilidade é o aumento do intervalo do número de *clusters* no sistema de recuperação. Em relação à métrica *Cluster Recall*, o experimento utilizando o algoritmo K-Medoids se mostrou superior ao que utilizou o Hierárquico Aglomerativo para gerar os agrupamentos. Na média harmônica (*F1-Measure*), o K-Medoids se mostrou mais estável com valores de performance concentrados e próximos à média.

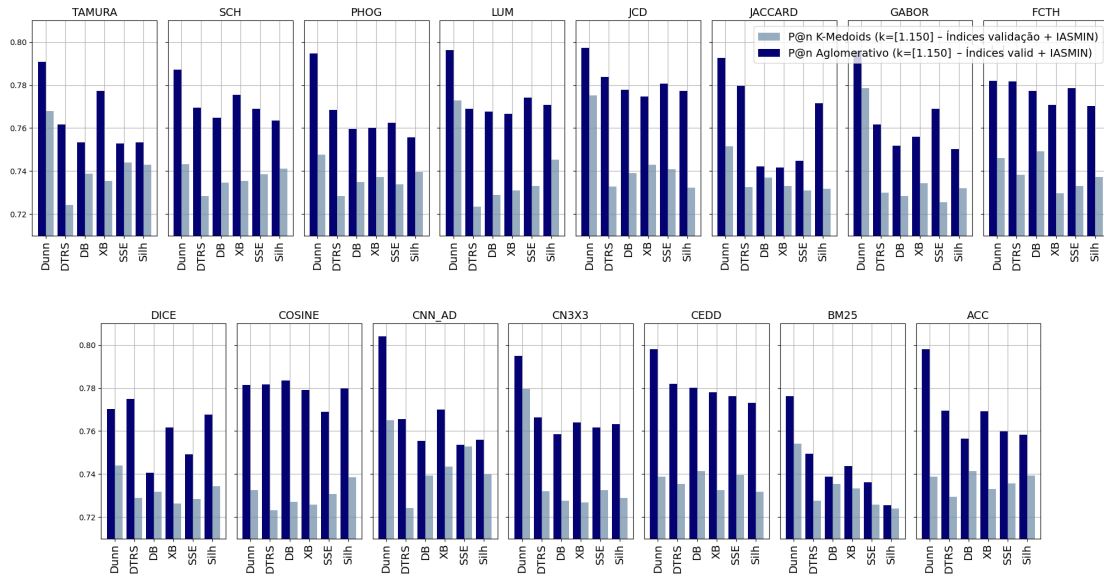


Figura 5.4: Nível de eficácia médio para a métrica *Precision*, comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 02 para todos os descritores do estudo.

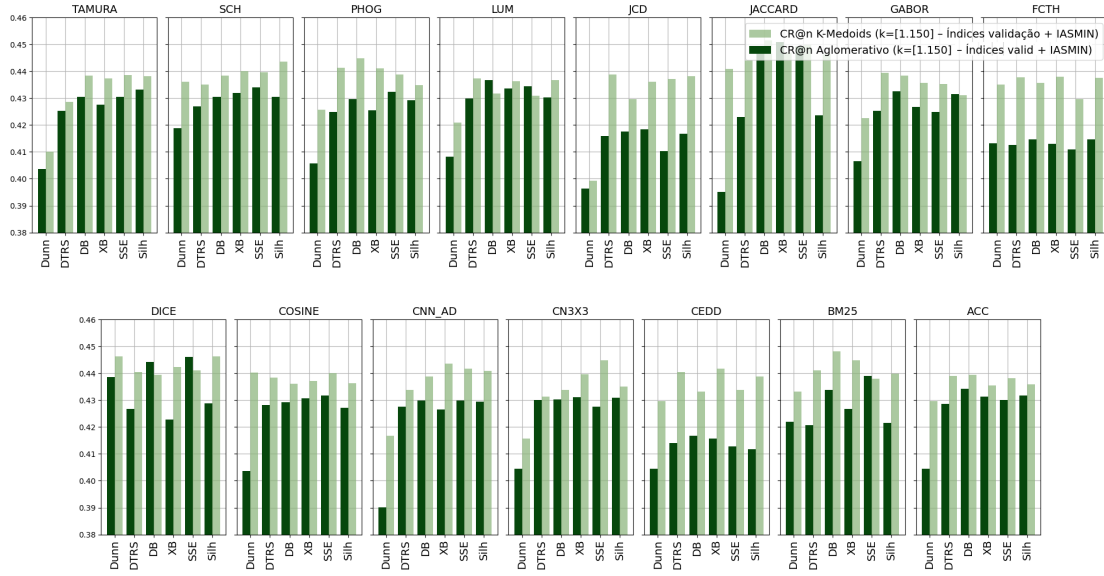


Figura 5.5: Nível de eficácia médio para a métrica *Cluster Recall*, comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 03 para todos os descritores do estudo.

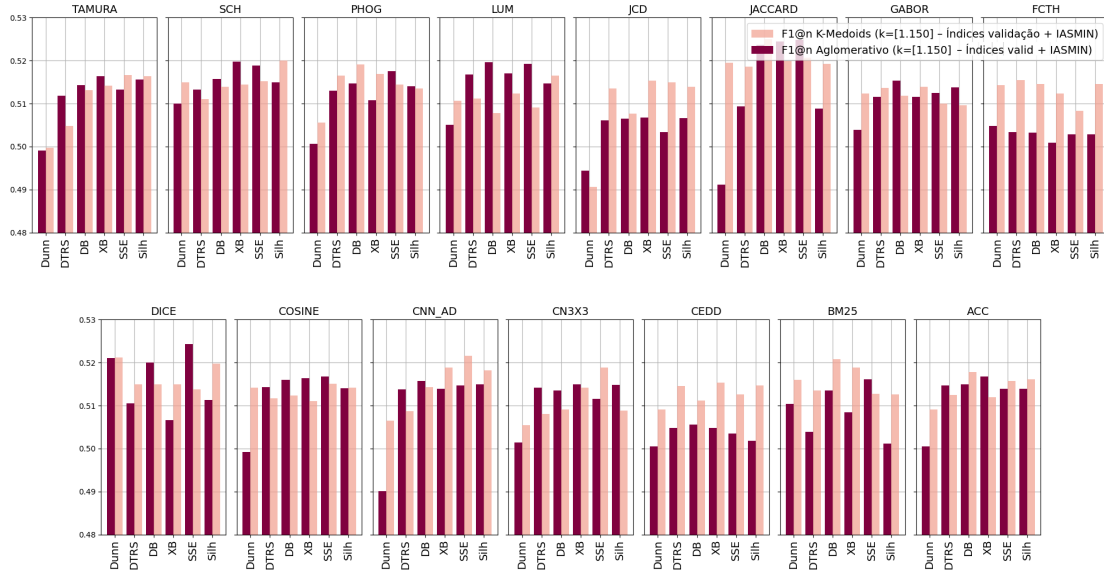


Figura 5.6: Nível de eficácia médio para a métrica *F1-measure*, comparativo entre os experimentos 02 e 03. Algoritmo Hierárquico Aglomerativo (experimento 02) versus K-Medoids (experimento 03), ambos com $k = [1, 150]$ e usando índices de validação com método auxiliar IASMIN. Superioridade do DTRS em relação ao experimento 02 nos descritores visuais TAMURA, SCH, LUM, CNN_AD, CN3X3, ACC e textuais COSINE. Para o experimento 03, superioridade do DTRS nos descritores visuais PHOG, JCD, GABOR, FCTH, CEDD e textuais DICE, BM25 e JACCARD.

Execuções adicionais

Além dos experimentos realizados anteriormente, foram feitas mais três execuções adicionais com parâmetros dados pelos experimentos 01 e 02 (Tabelas 3.2 e 3.3), onde utilizamos as três abordagens para seleção dos esquemas de *cluster* (apenas os índices de validação, na forma natural de cada índice; com uso dos índices de validação e do método auxiliar do Cotovelo; com uso dos índices de validação e do método auxiliar IASMIN), de forma que pudéssemos comparar o desempenho dessas três abordagens sob condições similares. A Figura 5.7 ilustra os parâmetros básicos e objetivos dos experimentos adicionais.

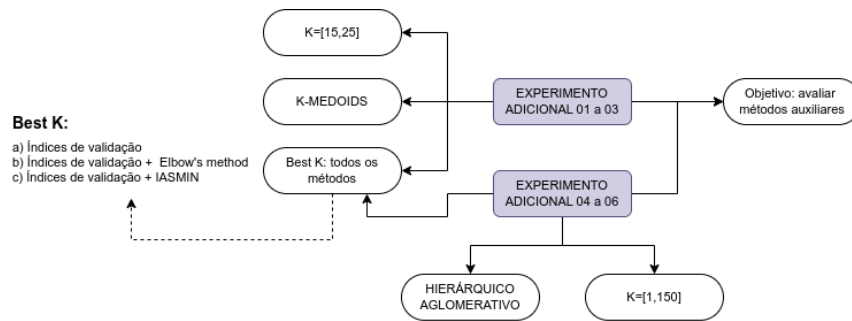
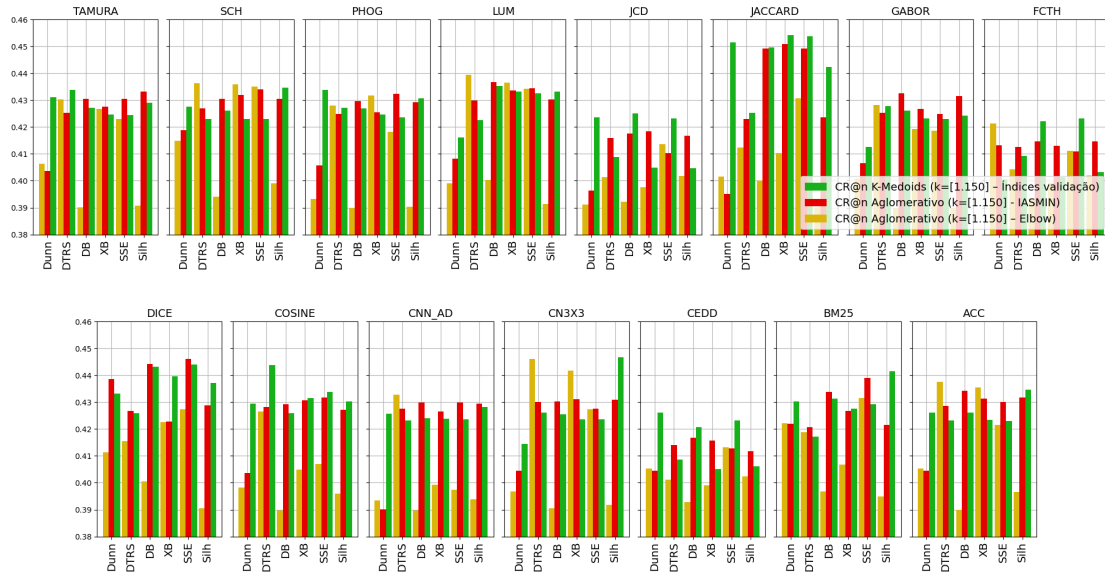
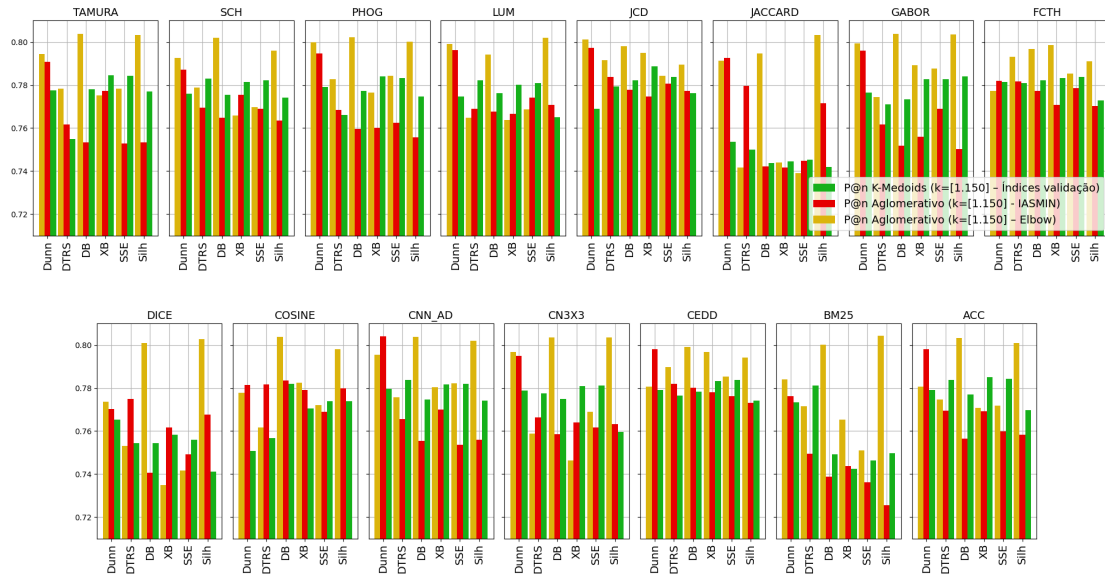


Figura 5.7: Diagrama: objetivo e parâmetros dos experimentos adicionais

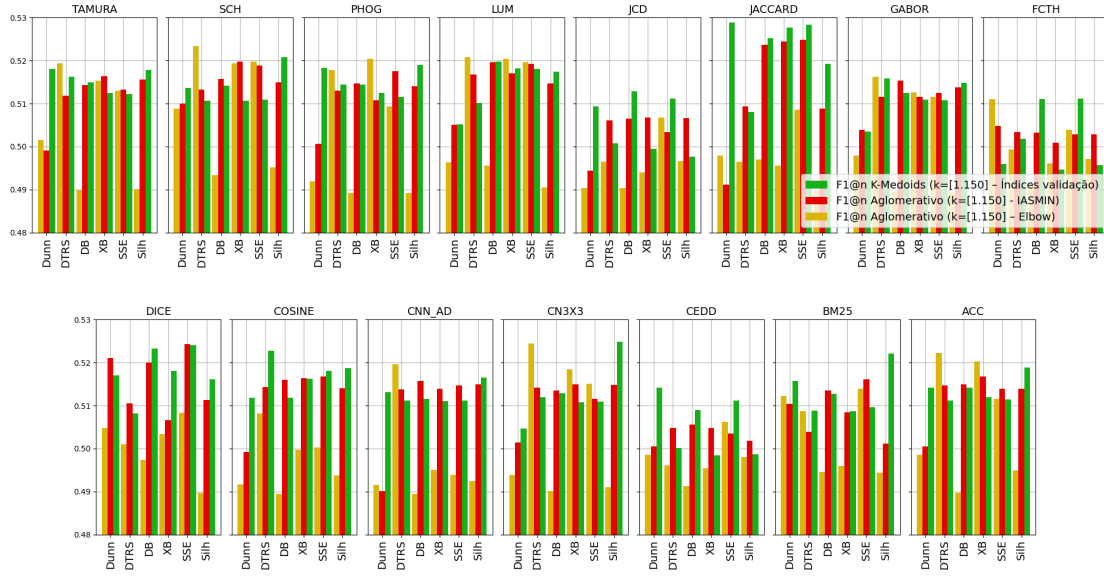
A Figura 5.8 traz os resultados de eficácia para as execuções adicionais do experimento 02. Aqui, o comparativo entre as três abordagens para seleção do número de grupos ideal: apenas os índices; índices com os dois métodos auxiliares. Valores médios, baseados em todos os descritores e ranking entre $n = 5$ e $n = 50$.



(a) Hierárquico Aglomerativo: comparativo entre diferentes abordagens para seleção do *best K*. Métrica *Precision* ($P@n$).



(a) Hierárquico Aglomerativo: comparativo entre diferentes abordagens para seleção do *best K*. Métrica *Cluster Recall* ($CR@n$)



(a) Hierárquico Aglomerativo: comparativo entre diferentes abordagens para seleção do *best K*. Métrica *F1-Measure* ($F1@n$)

Figura 5.8: Gráfico comparativo de performance dos experimentos adicionais. Algoritmo Hierárquico Aglomerativo com $k = [1, 150]$ e as três abordagens para seleção do número de grupos ideal: apenas com uso dos índices de validação de agrupamento, em sua forma natural (critérios dos índices de *cluster validity*, baseados em valores mínimos ou máximos); com índices de validação e auxílio do método do Cotovelo (*Elbow's method*); com índices de validação e auxílio do IASMIN. Valores médios das faixas $n = 5$ a $n = 50$ do ranking.

Os resultados da execução adicional para o experimento 01, utilizando o algoritmo K-Medoids com $k = [15, 25]$, estão detalhados no Apêndice deste trabalho. As tabelas C.1 a C.15 e Figura C.1 apresentam uma análise minuciosa desses resultados. É importante ressaltar que, neste último caso, a performance do sistema CBIR (Content-Based Image Retrieval) é significativamente influenciada pelo fator randômico na inicialização do algoritmo K-Medoids. Destacamos, ainda, que todos os experimentos e execuções deste trabalho foram conduzidos de maneira independente.

Tempo de execução e eficácia do sistema CBIR

Como visto nos resultados da seção 4.2.3, a abordagem utilizando o IASMIN demonstrou uma significativa vantagem sobre o método do Cotovelo no tempo de execução do sistema CBIR sem que houvesse comprometimento, ao menos, significativo, na performance geral em termos de relevância e diversificação. Neste sentido, ficou evidente que, quando as comparações são feitas dentro dos mesmos parâmetros, para um mesmo tipo de algoritmo, o IASMIN, no seu pior cenário, se mantém dentro da média de desempenho, em termos de eficácia do sistema CBIR, em relação aos demais métodos auxiliares (Figura C.1), sendo, portanto, extremamente vantajosa a adoção do IASMIN para a redução do tempo de execução do sistema de recuperação.

A mesma conclusão pode ser estendida ao DTRS, onde tempo médio de execução para este método utilizado na validação de agrupamentos foi consideravelmente inferior aos demais métodos, também, sem prejuízos na eficácia do sistema CBIR. Não ficando evidenciada nenhuma preferência deste método por descritores específicos, sejam eles textuais ou visuais, tanto em termos de custo computacional quanto em eficácia geral do sistema CBIR. Contudo, os resultados dos experimentos 01 e 02 sugerem que os descritores visuais alcançaram os maiores ganhos relativos para o método DTRS em todas as métricas de performance analisadas. A vantagem no tempo de execução sugere que tanto o DTRS quanto o método auxiliar IASMIN possuem uma tendência a encontrar os melhores esquemas de agrupamentos logo nas primeiras iterações do algoritmo de aprendizado não-supervisionado.

Vantagens e limitações

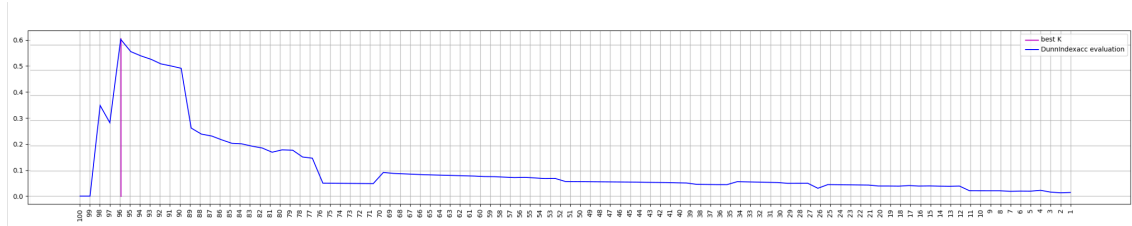
Em relação ao método DTRS, é necessário levar em consideração a natureza probabilística da solução na escolha do melhor esquema de *clusters*, além das características associadas aos conjuntos aproximados de Pawlak (1982), com suas regras de pertencimento, algo que pode ser considerado uma vantagem frente aos demais índices de validação de agrupamento que fazem uso das relações *intracluster* e *intercluster*. A natureza real dos dados processados em algoritmos de aprendizado não-supervisionado nem sempre é a ideal, ou seja, não se apresentam em agrupamentos fortemente coesos e distantes entre si. Por este motivo, o DTRS traz uma perspectiva mais próxima do mundo real na análise de dados. Um aspecto vantajoso é a capacidade de lidar com dados ruidosos (Li et al., 2016), que são dados sem sentido, dados que normalmente não podem ser interpretados pelo algoritmo.

Durante este estudo, observamos que o comportamento da curva de risco calculada pela função de perda (loss function) do DTRS, na maioria absoluta dos casos, apresenta pouca variabilidade. Esta característica pode ser considerada como uma vantagem do método para a análise dos resultados, tornando mais fácil a sua interpretação. O tempo de execução, apesar de estudos afirmarem que o DTRS pode ser custoso (Qian et al., 2017), se apresentou como outra vantagem frente aos demais métodos utilizados no estudo.

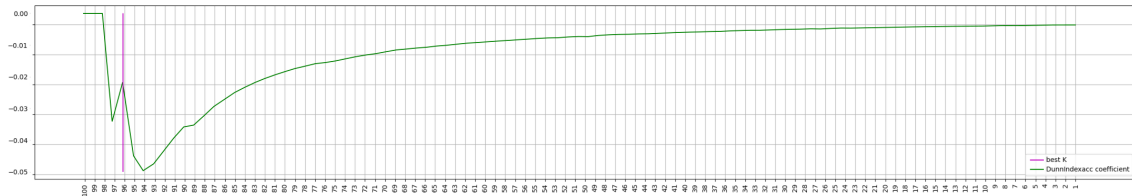
Já o método auxiliar IASMIN demonstrou vantagem em comparação ao método do Cotovelo em termos de tempo computacional, alcançando até 99,35% de redução no tempo médio de execução do sistema CBIR quando em conjunto com o DTRS, de acordo com a Tabela 4.43 do Capítulo 4. Preliminarmente, podemos destacar a flexibilidade do IASMIN como uma característica vantajosa. O ajuste de parâmetros como o intervalo t da média móvel e o limiar l para detecção da inflexão tornam o método adaptável ao comportamento do conjunto de dados analisado. Essa característica permite que o IASMIN possa ser utilizado em grandes conjuntos de dados, desde que feitos ajustes no intervalo da média móvel. O IASMIN é baseado na análise do comportamento da reta secante à média móvel dos resultados do *cluster validity*, tal característica poder ser uma vantagens no que se refere à adaptabilidade do mé-

todo em sistemas sujeitos a grande variabilidade nos dados. Contudo, o IASMIN é altamente dependente dos métodos de *cluster validity*, uma vez que a abordagem descrita se utiliza de tais métodos avaliativos para estimar a boa qualidade dos esquemas de agrupamentos gerados pelos algoritmos. Ademais, o IASMIN pode não ser adequado para conjuntos de dados muito pequenos, pois sua abordagem utiliza a média móvel para suavizar as variações nos índices de validação. Empiricamente, observado que o intervalo ideal para o cálculo da média móvel se concentra em torno de $t = 10$. Isso pode inviabilizar o uso do IASMIN com conjuntos de dados menores que o intervalo t . Além disso não há previsibilidade comportamental, devido à sua natureza bidimensional, em aplicações multidimensionais.

Em alguns casos observados neste estudo, o IASMIN selecionou esquemas de *cluster* em contradição à natureza do índice de validação de agrupamento como foi o caso, por exemplo, do Dunn Index (gráficos das Figuras 5.9 e 5.10). Aqui, tivemos dois comportamentos distintos para o mesmo índice de validação de agrupamento. No primeiro caso, gráfico da Figura 5.9 (mais precisamente, no gráfico da Figura 5.9(b)), tivemos que o ponto de inflexão foi detectado logo nas primeiras iterações do algoritmo Hierárquico Aglomerativo com a interrupção do processo de forma assertiva, de acordo com o que preconiza o método Dunn Index, ou seja, a seleção do melhor esquema de *clusters* pelo valor máximo atingido pelo índice de validação.



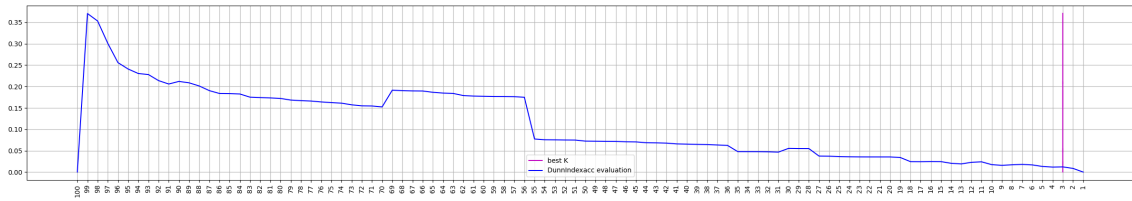
(a) Curva de distribuição dos resultados do método *Dunn Index* para o descritor ACC e tópico $ID = 4$ (*la_madeleine*). O valor máximo atingido ficou próximo ao ponto de inflexão da curva do coeficiente angular da reta secante à média móvel em 5.9(b). A interpretação do IASMIN é que, possivelmente, teríamos uma sequência de queda nos valores do índice a partir daquele ponto.



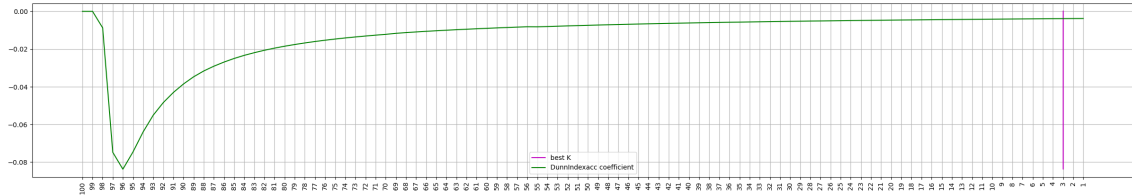
(b) Curva com os valores do coeficiente angular da reta secante à média móvel dos resultados para o método *Dunn Index*. Critério de interrupção (IASMIN) baseado na inflexão positiva em $k = 96$ e tópico *la_madeleine* (*Query ID = 4*).

Figura 5.9: Comportamento das distribuições de resultados e coeficiente angular da reta secante à média móvel para o índice de validação de agrupamento *Dunn Index*, tópico $ID = 4$ (*la_madeleine*) e descritor ACC. Número de grupos ideal em $k = 96$.

O segundo caso (gráfico da Figura 5.10(b)) mostra que o IASMIN não detectou nenhuma inflexão positiva na curva do coeficiente angular como convencionado para o método Dunn Index na seção 3.2.2, levando as iterações do algoritmo até finalizar em $k = 1$, selecionando o número de grupos ideal em $k = 3$ (o IASMIN seleciona o antepenúltimo valor para k caso o algoritmo não seja interrompido até o final). A curva do coeficiente angular em questão (gráfico da Figura 5.10(b)) apresentou um único ponto de inflexão negativo. Neste caso, o melhor esquema de agrupamentos encontrado não foi o mais adequado em termos naturais do índice de validação de agrupamento (o *Dunn Index* deve ser máximo), uma vez que o valor do índice correspondente ao *best K* foi tomado pelo mínimo e não pelo máximo como pressupõe o método.



(a) Curva dos resultados do método da *Dunn Index* com interrupção tardia provocada pela escolha errada do tipo de inflexão para o ponto de análise. Neste caso, o melhor esquema de *clusters*, segundo critérios naturais do índice, encontra-se em $k = 99$.



(b) Curva do coeficiente angular para o caso da figura anterior (5.10(a)). Esta distribuição apresentou comportamento atípico no tópico $ID = 3$. Curva com ponto de inflexão negativo em $k = 96$, nesse caso o algoritmo é finalizado no antepenúltimo valor da distribuição ($k = 3$). O algoritmo não encontrou nenhum ponto de inflexão positivo suficientemente grande.

Figura 5.10: Comportamento das distribuições de resultados e coeficiente angular da reta secante à média móvel para o índice de validação de agrupamento *Dunn Index*, tópico $ID = 3$ (*la_madeleine*) e descritor ACC. Comportamento atípico, impossibilidade de detecção do ponto de inflexão.

5.2 Resposta às questões de pesquisa

Q1: Qual o desempenho do método DTRS quando utilizado como índice de validação de agrupamento no contexto da CBIR?

Para a análise do desempenho geral do sistema CBIR foi necessário, inicialmente, a definição do referencial comparativo. Desta forma, adotamos uma execução preliminar com $k = 20$ como *baseline* para a aferição do quão o DTRS foi bom (ou ruim).

Assim, os resultados vistos no capítulo anterior mostraram que o desempenho do DTRS foi satisfatório em todos os casos, com melhores resultados no experimento inicial, com uso do algoritmo K-Medoids e intervalo $k = [15, 25]$ (ganho médio total de 9.7% frente ao *baseline*). Nos demais experimentos, o DTRS alcançou ganhos de 8.16% (Hierárquico Aglomerativo) e 7.11% (K-Medoids), ambos com intervalo estendido para $k = [1, 150]$.

Q2: Qual o impacto do método DTRS no tempo de execução do sistema CBIR?

O método DTRS apresentou os melhores resultados em termos de tempo de execução dentre todos os índices de validação de agrupamento do estudo. A coleta do tempo de execução do sistema CBIR foi medida em dois experimentos (02 e 03) com diferentes algoritmos (K-Medoids e Hierárquico Aglomerativo) e em termos médios. O percentual de redução no tempo médio de execução para o experimento com o K-Medoids, num intervalo $k = [1, 150]$, ficou entre 97.5% e 98.6%. O tempo médio de execução do sistema CBIR utilizando o algoritmo Hierárquico Aglomerativo, nas mesmas condições do experimento anterior, atingiu percentuais um pouco menores (entre 64.3% e 95.6%). Isto sugere que, independentemente do tipo de algoritmo escolhido, o método DTRS sempre terá impacto positivo sobre o tempo de execução do sistema CBIR.

Q3: Qual o impacto da adoção da abordagem automatizada no tempo de execução do sistema CBIR?

O IASMIN, cuja abordagem foi descrita no capítulo 3, apresentou redução significativa no tempo de execução frente ao método do Cotovelo, com percentuais de redução variando entre 45.4% e 93.4% de acordo com o índice de validação de agrupamento. As maiores reduções foram observadas em conjunto com o método DTRS. Porém, o IASMIN não apresentou resultados satisfatórios quando comparado com a forma natural dos índices de validação de agrupamento, uma vez que esta última abordagem ocorre sem a implementação de qualquer outra heurística sendo, portanto, o tempo de execução tomado sem a participação de quaisquer métodos auxiliares como o IASMIN ou o *Elbow*.

Q4: Qual o impacto na eficácia do sistema CBIR pela ampliação do intervalo de análise do número de agrupamentos?

Em termos gerais, houve redução na eficácia do sistema com o aumento do intervalo quando a observação é feita com o algoritmo K-Medoids (experimentos 01 e 02). Não foi realizado experimento com intervalos diferentes para o algoritmo Hierárquico Aglomerativo.

5.3 Ameaças à validade

Consideramos algumas ameaças à validade em torno da conclusão para este estudo que devem ser observadas:

- Em menos de 1% dos casos não foi possível a rejeição da hipótese nula, indicando uma provável normalidade dos dados. Isto constitui uma ameaça à validade em termos de baixo poder estatístico (erro do tipo 2) e violação das suposições. Contudo, consideramos que este fato não foi suficiente para descartarmos o teste de significância não-paramétrico de *Wilcoxon* na análise dos resultados. Além disso, o teste de *Wilcoxon* fornece inferências robustas mesmo quando as suposições de normalidade são violadas;
- Confiabilidade da implementação de tratamentos: a troca do K-Medoids pelo algoritmo Hierárquico Aglomerativo pode ter interferido nos resultados, tornando o DTRS e o IASMIN superiores por este motivo e não pela natureza deles. Este também é uma ameaça que pode ser descartada, os experimentos adicionais corroboraram com os resultados iniciais;
- O fator randômico na inicialização do K-Medoids pode ter interferido nos resultados. Este fato pode caracterizar duas possíveis ameaças à validade do estudo como em termos de irrelevância e heterogeneidade randômica. Contudo, os valores de performance e tempo de execução apresentados nos resultados representam a média dos valores coletados nos experimentos como forma de minimizar os possíveis efeitos randômicos.

Capítulo 6

Considerações finais

A descoberta de subtópicos implícitos nas consultas em sistemas de recuperação de imagens baseados em conteúdo é uma tarefa complexa, a qual tem por objetivo oferecer ao usuário uma experiência satisfatória, com resultados relevantes, em grau de diversificação ideal, e com tempo de resposta o mais eficiente possível. Esta tarefa está intimamente ligada ao desafio da seleção dos melhores esquemas de grupos em algoritmos de aprendizado não-supervisionado baseados em agrupamento de dados. Ela depende do correto ajuste dos parâmetros básicos e iniciais do algoritmo escolhido, assim como do método de validação dos esquemas de agrupamento, da abordagem auxiliar para a seleção dos resultados da validação, além dos descritores empregados na extração das características intrínsecas das imagens.

Neste trabalho, abordamos o Modelo de Decisão Teórica dos Conjuntos Aproximados (*Decision-Theoretic Rough Sets* - DTRS) de Pawlak (1982) como índice de validação de agrupamento no contexto da recuperação de imagens baseada em conteúdo (CBIR), dentro dos aspectos de relevância e diversificação, na descoberta de subtópicos ocultos em uma consulta. Além disso, apresentamos um novo método auxiliar para a interrupção do algoritmo de aprendizado não-supervisionado (IAS-MIN), visando a redução do tempo de execução do sistema e, consequentemente, a otimização do custo computacional. Para isto, conduzimos três experimentos principais, dos quais dois utilizaram o algoritmo K-Medoids com diferentes parâmetros, enquanto um terceiro empregou o método Hierárquico Aglomerativo. Adicionalmente, foram realizadas mais execuções, utilizando diferentes metodologias para a seleção do *best K*, a fim de respondermos às questões de pesquisa.

Assim, após a análise dos resultados, consideramos que a abordagem utilizando o modelo DTRS (Yao, 2007) como índice de validação de agrupamento apresentou contribuições práticas para a eficácia de sistemas CBIR. Essa abordagem proporcionou ganhos significativos em comparação com os resultados obtidos com o uso de um número fixo de agrupamentos ($k = 20$), prática amplamente utilizada pela comunidade científica para a tarefa de recuperação de imagens. Além disso, a eficácia geral do sistema CBIR com uso do DTRS manteve-se dentro da média resultante

dos demais métodos de validação, com equivalência estatística na grande maioria dos casos e superioridade com alguns descritores como, por exemplo, o PHOG. Contudo, não foi observada predileção do DTRS por tipos determinados de descritor, sejam eles visuais ou textuais. Quanto ao tempo de execução do sistema CBIR, o método DTRS demonstrou superioridade quando comparado aos demais índices de validação de agrupamento estudados, obtendo uma redução superior a 90% no tempo médio de execução, tornando-o uma opção interessante para a redução do custo computacional.

O método auxiliar IASMIN (*Interrupção pela Análise da reta Secante da Média móvel do Índice*) para a seleção dos melhores esquemas de *cluster*, apresentado neste estudo, atingiu bons resultados. A abordagem utilizada no IASMIN, baseada na análise comportamental da curva gerada pelos resultados dos métodos de validação, incrementou positivamente a performance do sistema de recuperação e diminuiu o tempo de execução, quando em comparação com o método do Cotovelo, em todos os casos estudados. Trazendo também, como contribuições práticas, a melhoria da eficácia na pesquisa de subtópicos implícitos em uma consulta CBIR, permitindo a geração de resultados relevantes com bom grau de diversificação dentro do contexto esperado, e a flexibilidade em relação ao tamanho e comportamento dos dados.

6.1 Trabalhos Futuros

Desde a realização da tarefa *MediaEval 2015*, houve o surgimento de novos descritores que permitiram a extração de *features* das imagens através de técnicas com novos métodos, possivelmente mais eficazes, de extração de *features* baseados em *Deep Learning*. Esses novos descritores podem ser utilizados em testes de parametrização (Tabela 3.3) dos algoritmos e índices de validação de agrupamento estudados objetivando ganhos de performance do sistema CBIR. Mudanças nos parâmetros da função de risco do método DTRS (equações 2.25 a 2.27) também podem ser realizadas visando ganhos na identificação dos esquemas de agrupamento mais hábeis e o consequente aumento na performance do sistema de recuperação. Uma análise comparativa entre os vários tipos de algoritmos de aprendizado não-supervisionado com a aplicação do *DTRS* em contraponto aos índices tradicionais de validação de agrupamento e, sobretudo, testes de ajuste na parametrização do IASMIN podem ser adotados nos próximos trabalhos.

Referências

- Asuncion, A. e Newman, D. (2007). Uci machine learning repository.
- Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., e Chaudhury, S. (2009). Hypothesis testing, type i and type ii errors. *Industrial psychiatry journal*, 18(2):127.
- Bayes, T. e Price, R. (1763). Philosophical transactions of the royal society of london. *Phil Trans R Soc*, 53:370–418.
- Bezdek, J. C., Li, W., Attikiouzel, Y., e Windham, M. (1997). A geometric approach to cluster validity for normal mixtures. *Soft Computing*, 1:166–179.
- Biasotti, S., Cerri, A., Giorgi, D., e Spagnuolo, M. (2013). Phog: Photometric and geometric functions for textured shape retrieval. In *Computer Graphics Forum*, volume 32, páginas 13–22. Wiley Online Library.
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., e Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html#sphx-glr-auto-examples-cluster-plot-kmeans-silhouette-analysis-py.
- Calumby, R. T., Gonçalves, M. A., e da Silva Torres, R. (2017). Diversity-based interactive learning meets multimodality. *Neurocomputing*, 259:159–175. Multimodal Media Data Understanding and Analytics.
- Campos, G. M. (2002). Estatística prática para docentes e pós-graduandos.
- Carpineto, C., D’Amico, M., e Romano, G. (2012). Evaluating subtopic retrieval methods: Clustering versus diversification of search results. *Information Processing & Management*, 48(2):358–373.
- Chatzichristofis, S. A. e Boutalis, Y. S. (2008a). Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In *International conference on computer vision systems*, páginas 312–322. Springer.

- Chatzichristofis, S. A. e Boutalis, Y. S. (2008b). Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In *2008 Ninth international workshop on image analysis for multimedia interactive services*, páginas 191–196. IEEE.
- Chen, Y., Wang, J. Z., e Krovetz, R. (2003). Content-based image retrieval by clustering. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, páginas 193–200.
- Cinque, L., Ciocca, G., Levialdi, S., Pellicano, A., e Schettini, R. (2001). Color-based image retrieval using spatial-chromatic histograms. *Image and Vision Computing*, 19(13):979–986.
- Corani, G., Benavoli, A., Demšar, J., Mangili, F., e Zaffalon, M. (2017). Statistical comparison of classifiers through bayesian hierarchical modelling. *Machine Learning*, 106:1817–1837.
- Davies, D. L. e Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96-34, páginas 226–231.
- Ferreira, J. C. e Patino, C. M. (2015). What does the p value really mean? *Jornal Brasileiro de Pneumologia*, 41(5):485.
- Figuerêdo, J. S. L. e Calumby, R. T. (2022). Unsupervised query-adaptive implicit subtopic discovery for diverse image retrieval based on intrinsic cluster quality. *Multimedia Tools and Applications*, páginas 1–21.
- Fowlkes, E. B. e Mallows, C. L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American statistical association*, 78(383):553–569.
- Fürnkranz, J. (2002). Round robin classification. *The Journal of Machine Learning Research*, 2:721–747.
- Garcia-Dias, R., Vieira, S., Pinaya, W. H. L., e Mechelli, A. (2020). Clustering analysis. In *machine learning*, páginas 227–247. Elsevier.
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques*, volume 12. Springer Science & Business Media.

- Halkidi, M., Batistakis, Y., e Vazirgiannis, M. (2002). Cluster validity methods: part i. *ACM Sigmod Record*, 31(2):40–45.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.
- Ionescu, B., Gînsca, A.-L., Boteanu, B., Popescu, A., Lupu, M., e Müller, H. (2014). Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. *MediaEval*, 1263.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR).
- Kato, T. (1992). Database architecture for content-based image retrieval. In Jamberdino, A. A. e Niblack, C. W., editores, *Image Storage and Retrieval Systems*, volume 1662, páginas 112 – 123. International Society for Optics and Photonics, SPIE.
- Kaufman, L. e Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.
- Kovács, F., Legány, C., e Babos, A. (2005). Cluster validity measurement techniques. In *6th International symposium of hungarian researchers on computational intelligence*, volume 35.
- Kwon, S. H., Kim, J., e Son, S. H. (2021). Improved cluster validity index for fuzzy clustering. *Electronics Letters*, 57(21):792–794.
- Lewis, J., Ossowski, S., Hicks, J., Errami, M., e Garner, H. R. (2006). Text similarity: an alternative way to search medline. *Bioinformatics*, 22(18):2298–2304.
- Li, W., Huang, Z., Jia, X., e Cai, X. (2016). Neighborhood based decision-theoretic rough set models. *International Journal of Approximate Reasoning*, 69:1–17.
- Lux, M. e Chatzichristofis, S. A. (2008). Lire: lucene image retrieval: an extensible java cbir library. In *Proceedings of the 16th ACM International Conference on Multimedia*, MM '08, página 1085–1088, New York, NY, USA. Association for Computing Machinery.
- Manjunath, B. S. e Ma, W.-Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842.
- Meilă, M. (2003). Comparing clusterings by the variation of information. In *Learning theory and kernel machines*, páginas 173–187. Springer.
- Miot, H. A. (2017). Assessing normality of data in clinical and experimental trials.

- Müller, H., Ionescu, B., Gînsca, A. L., Boteanu, B., Popescu, A., e Lupu, M. (2015). Retrieving diverse social images at mediaeval 2015. In *Working Notes Proceedings of the MediaEval 2015 Workshop*. 14-15 September 2015.
- Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., e Abu-Mahfouz, A. M. (2022). Autoelbow: An automatic elbow detection method for estimating the number of clusters in a dataset. *Applied Sciences*, 12(15).
- Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, 11:341–356.
- Penatti, O. A., Valle, E., e Torres, R. d. S. (2012). Comparative study of global color and texture descriptors for web image retrieval. *Journal of visual communication and image representation*, 23(2):359–380.
- Qian, Y., Liang, X., Lin, G., Guo, Q., e Liang, J. (2017). Local multigranulation decision-theoretic rough sets. *International Journal of Approximate Reasoning*, 82:119–137.
- Rasyid, L. e Andayani, S. (2018). Review on clustering algorithms based on data type: towards the method for data combined of numeric-fuzzy linguistics. In *Journal of Physics: Conference Series*, volume 1097, página 012082. IOP Publishing.
- Rdusseeun, L. e Kaufman, P. (1987). Clustering by means of medoids. In *Proceedings of the statistical data analysis based on the L1 norm conference, neuchatel, switzerland*, volume 31.
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. d. F., e Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1):1–34.
- Rohm, M., Ionescu, B., Gînscă, A. L., Santos, R. L. T., e Müller, H. (2018). Subdiv17: A dataset for investigating subjectivity in the visual diversification of image search results. In *Proceedings of the 9th ACM Multimedia Systems Conference, MMSys '18*, página 444–449, New York, NY, USA. Association for Computing Machinery.
- Rojas-Thomas, J., Santos, M., e Mora, M. (2017). New internal index for clustering validation based on graphs. *Expert Systems with Applications*, 86:334–349.
- Rokach, L. e Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, páginas 321–352. Springer.
- Ross, S. M. (2017). Chapter 9 - testing statistical hypotheses. In Ross, S. M., editor, *Introductory Statistics (Fourth Edition)*, páginas 381–432. Academic Press, Oxford, fourth edition edição.

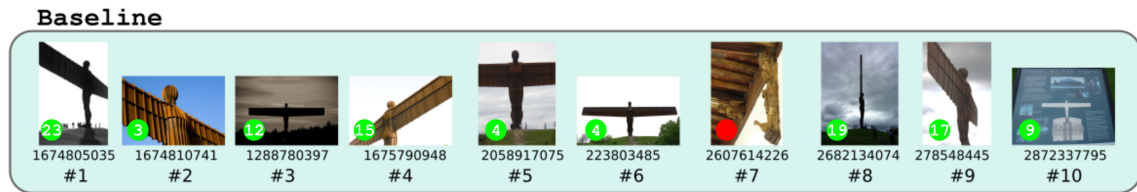
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
- Satopaa, V., Albrecht, J., Irwin, D., e Raghavan, B. (2011). Finding a "kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*, páginas 166–171.
- Schubert, E. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *SIGKDD Explor. Newsl.*, 25(1):36–42.
- Shapiro, S. S. e Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Shi, J. e Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905.
- Shimada, K., Matsukawa, T., Noguchi, Y., e Kurita, T. (2010). Appearance-based smile intensity estimation by cascaded support vector machines. In *Asian Conference on Computer Vision*, páginas 277–286. Springer.
- Singh, S. e Batra, S. (2020). An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications*, 79(25):17731–17759.
- Sugar, C. A. e James, G. M. (2003). Finding the number of clusters in a dataset: An information-theoretic approach. *Journal of the American Statistical Association*, 98(463):750–763.
- Tamura, H., Mori, S., e Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, man, and cybernetics*, 8(6):460–473.
- Tang, Y., Sun, F., e Sun, Z. (2005). Improved validation index for fuzzy clustering. In *Proceedings of the 2005, American Control Conference, 2005.*, páginas 1120–1125. IEEE.
- Thomas, J. C. R., Peñas, M. S., e Mora, M. (2013). New version of davies-bouldin index for clustering validation based on cylindrical distance. In *2013 32nd International Conference of the Chilean Computer Science Society (SCCC)*, páginas 49–53. IEEE.
- Thorndike, R. L. (1953). Who belongs in the family. In *Psychometrika*. Citeseer.
- Tyagi, V. (2017). *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. Springer.
- Van Leuken, R. H., Garcia, L., Olivares, X., e van Zwol, R. (2009). Visual diversification of image search results. In *Proceedings of the 18th international conference on World wide web*, páginas 341–350.

- Woolson, R. F. (2007). Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, páginas 1–3.
- Xie, X. L. e Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.
- Yao, Y. (2007). Decision-theoretic rough set models. In *International conference on rough sets and knowledge technology*, páginas 1–12. Springer.
- Yu, H., Liu, Z., e Wang, G. (2014). An automatic method to determine the number of clusters using decision-theoretic rough set. *International Journal of Approximate Reasoning*, 55(1):101–115.
- Zagoris, K., Chatzichristofis, S. A., Papamarkos, N., e Boutalis, Y. S. (2010). Automatic image annotation and retrieval using the joint composite descriptor. In *2010 14th Panhellenic Conference on Informatics*, páginas 143–147. IEEE.
- Zhai, C. (2008). Statistical language models for information retrieval. *Synthesis lectures on human language technologies*, 1(1):1–141.
- Zhai, C., Cohen, W. W., e Lafferty, J. (2015). Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Acm sigir forum*, volume 49, páginas 2–9. ACM New York, NY, USA.
- Zhang, S., Yang, M., Cour, T., Yu, K., e Metaxas, D. N. (2012). Query specific fusion for image retrieval. In *European conference on computer vision*, páginas 660–673. Springer.
- Zhao, Z. e Glotin, H. (2009). Diversifying image retrieval with affinity-propagation clustering on visual manifolds. *IEEE MultiMedia*, 16(04):34–43.

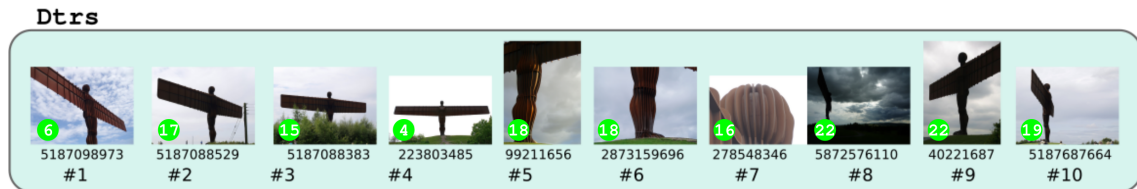
Apêndice A

Imagens reais dos experimentos

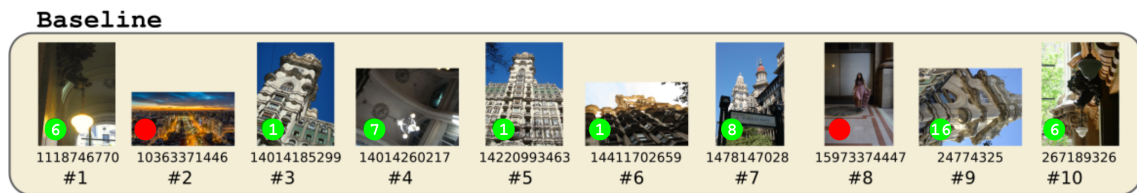
A.0.1 Experimento 01: K-Medoids com $k = [15, 25]$



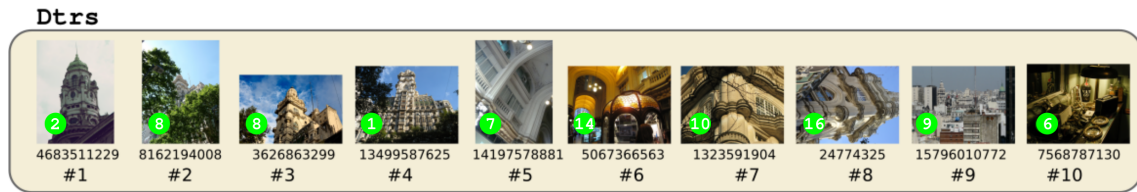
(a) Imagens reais: resultados do *baseline* na métrica *Precision* (query: angel_of_north, id: 1).



(b) Imagens reais: resultados do *DTRS* na métrica *Precision* (query: angel_of_north, id: 1).



(c) Imagens reais: resultados do *baseline* na métrica *Cluster Recall* (query: palacio_barolo, id: 216).

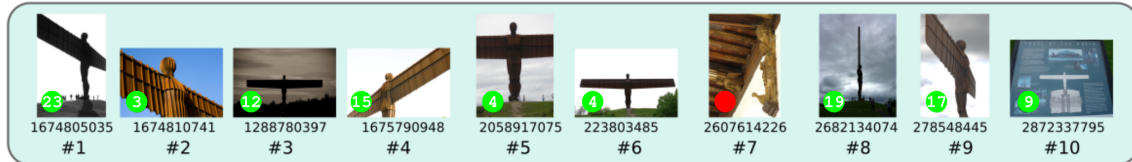


(d) Imagens reais: resultados do *DTRS* na métrica *Cluster Recall* (query: *palacio_barolo*, id: 216).

Figura A.1: Imagens reais do experimento 01 (K-Medoids, $k = [15, 25]$, usando apenas os índices de validação sem método auxiliar. Resultados para o descritor JCD em duas consultas (*Angel_of_north* e *Palacio_barolo*) nas métricas *Precision* e *Cluster Recall* para a faixa de ranking $n = 10$. Comparativo entre o *baseline* e o *DTRS*. No quadro, a colocação da imagem na faixa de ranking, o id da imagem e o grupo semântico pertencente. Falsos positivos marcados em vermelho. Tópicos de consulta com melhores performances para o descritor.

A.0.2 Experimento 02: Hierárquico Aglomerativo com $k = [1, 150]$

Baseline



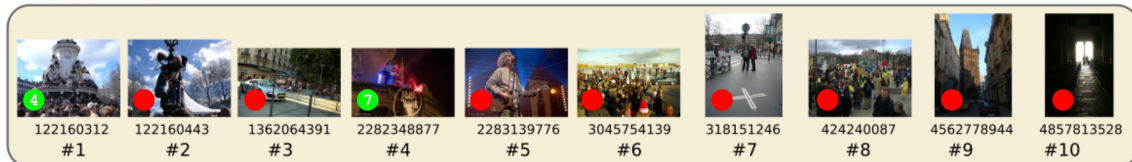
(a) Imagens reais: resultados do *baseline* na métrica *Precision* (query: angel_of_north, id: 1).

Dtrs + IASMIN



(b) Imagens reais: resultados do *DTRS* na métrica *Precision* (query: angel_of_north, id: 1).

Baseline



(c) Imagens reais: resultados do *baseline* na métrica *Cluster Recall* (query: place_de_la_republique, id: 57).

Dtrs + IASMIN



(d) Imagens reais: resultados do *DTRS* na métrica *Cluster Recall* (query: place_de_la_republique, id: 57).

Figura A.2: Imagens reais do experimento 02 (Hierárquico Aglomerativo, $k = [1, 150]$, DTRS e IASMIN). Resultados para o descritor JCD em duas consultas (Angel_of_north e place_de_la_republique) nas métricas *Precision* e *Cluster Recall* para a faixa de ranking $n = 10$. Comparativo entre o *baseline* e o DTRS. No quadro, a colocação da imagem na faixa de ranking, o id da imagem e o grupo semântico pertencente. Falsos positivos marcados em vermelho. Tópicos de consulta com melhores performances para o descritor.

Apêndice B

Análise estatística dos experimentos

B.0.1 Avaliação descritiva dos resultados alcançados no experimento 01

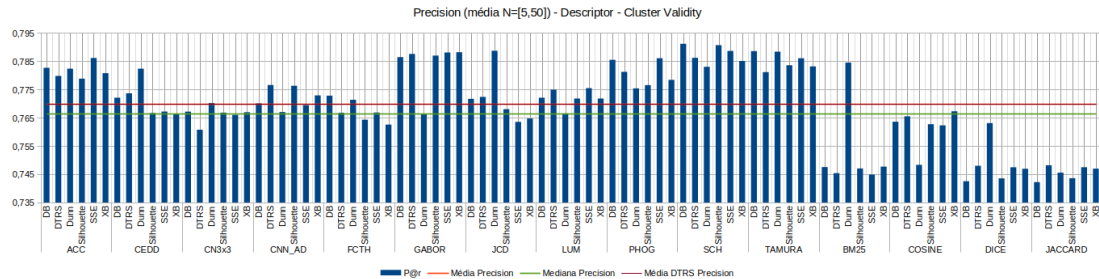
De acordo com os resultados gerais das métricas *Precision*, *Cluster-Recall* e *F1-Mesasure*, divididos em subgrupos ($P@n$, $CR@n$ e $F1@n$) de classificação das imagens para cada um dos descritores textuais e visuais, em análise preliminar, apresentaram valores dentro do esperado, em termos médios, para todos índices de validação de agrupamento experimentados, quando comparados com os resultados descritos em Figuerêdo e Calumby (2022).

Levando-se em conta a análise descritiva dos dados gerais (união de todos os resultados dos índices de validação de agrupamento para todos os descritores testados, em valores médios) e com a aplicação do teste de normalidade de Shapiro e Wilk (1965), obtivemos as seguintes conclusões:

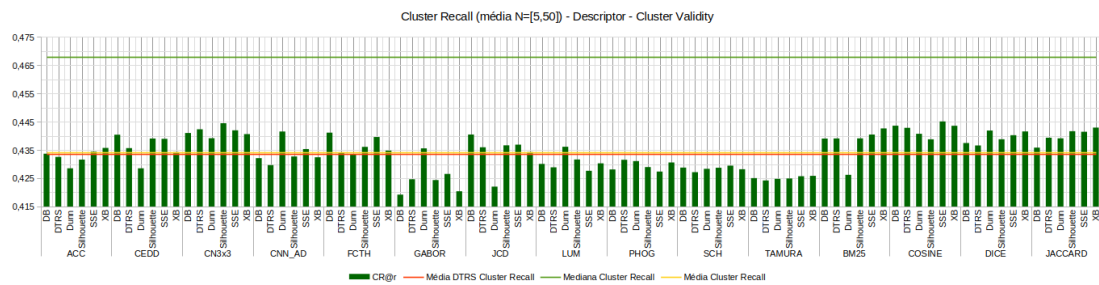
- Para a métrica *Precision* ($P@n$), os melhores valores resultantes, para o algoritmo K-Medoids, concentraram-se no intervalo $[0,7649, 0,8279]$, os piores valores no intervalo $[0,7319, 0,7491]$, a média $0,7699$, a mediana $0,7664$, o 3º quartil $Q(75\%) = 0,7838$ e a moda $0,7595$. O *p-value* encontrado foi de $2.477e-07$ ($0,0000002477$, teste de *Shapiro-Wilk*), bem abaixo do nível de significância ($\alpha = 0,05$), assumindo a não-normalidade da distribuição apesar da concentração de valores vista no gráfico *Q.Q. Plot* da Figura B.1(c). Observamos assimetria levemente positiva e com forte concentração nos dados gerais de acordo com o histograma da Figura B.1(e);
- Em relação à métrica *Cluster Recall* ($CR@n$), os melhores valores concentraram-se no intervalo $[0,6399, 0,6666]$, os piores valores em $[0,1495, 0,1569]$, a média $0,4343$, mediana $0,4678$, 3º quartil $Q(75\%) = 0,5937$, moda $0,1564$ e *p-value* $< 2.2e-16$ (ou seja, próximo de 0, rejeitando a hipótese nula, distribuição não-normal). Levando-se em conta apenas a média e medi-

ana, observamos assimetria à esquerda e grande dispersão, conforme pode ser visto nos gráficos B.1(e) e B.1(f) da Figura B.1;

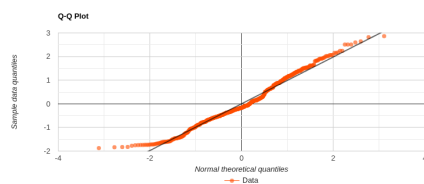
- Na média harmônica $F1-Measure$ ($F1@n$), os melhores valores se concentraram em $[0,6796, 0,6950]$, os piores valores $[0,2464, 0,2579]$, a média $0,5203$, a mediana $0,5711$, o 3º quartil em $Q(75\%) = 0,6528$, moda $0,2579$, $p-value < 2.2e-16$ e distribuição seguindo as características da *Cluster Recall*.



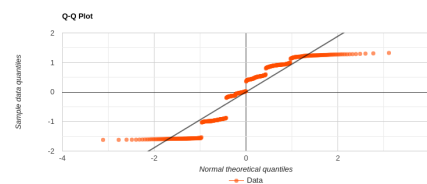
(a) Distribuição dos resultados em gerais da *Precision*. A média geral do DTRS foi muito próxima da média geral da *Precision* (faixas $n = 5$ a $n = 50$ do ranking).



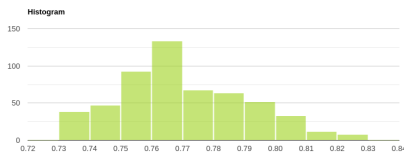
(b) Distribuição dos resultados em gerais da *Cluster Recall*. A média geral do DTRS, assim como na *Precision*, foi muito próxima da média geral da *Cluster Recall*.



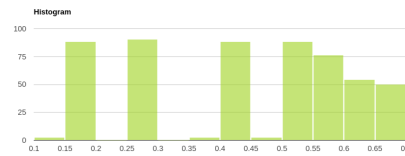
(c) Q.Q Plot gerado com os dados de resultado gerais do *Precision*. Formato levemente arqueado denota fuga de normalidade e assimetria positiva.



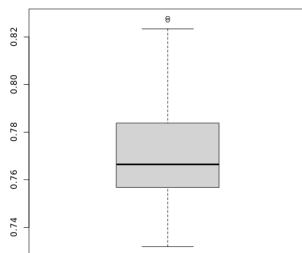
(d) Q.Q Plot gerado com os dados de resultado gerais do *Cluster Recall*. O efeito *escada* demonstra alta dispersão na distribuição.



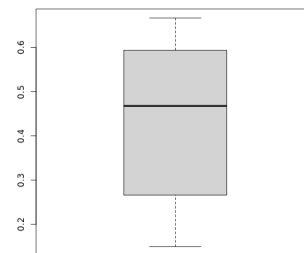
(e) Histograma dos dados de resultado gerais do Precision demonstrando leve assimetria positiva.



(f) Histograma: resultado gerais, *Cluster Recall* demonstra dispersão nos dados.



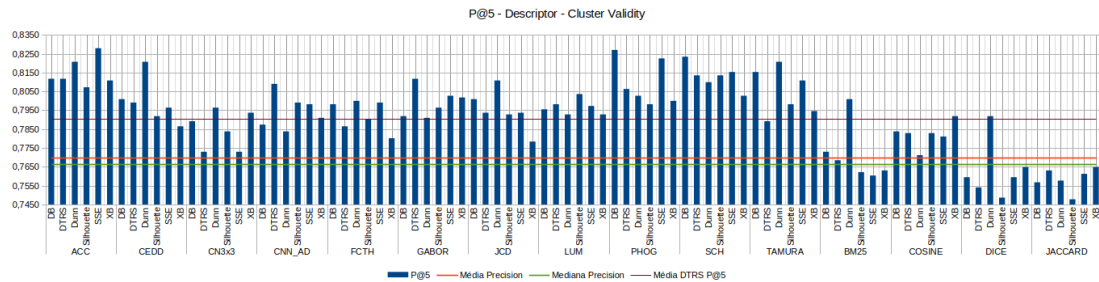
(g) Boxplot: distribuição da métrica *Precision*, presença de outliers e assimetria positiva.



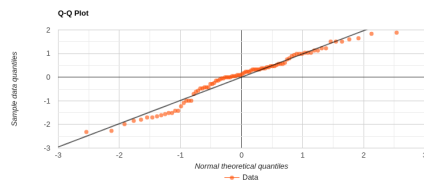
(h) Boxplot: distribuição do *Cluster Recall*, desvio assimétrico negativo.

Figura B.1: Comportamento dos resultados para todos os descritores e índices de validação agrupados para as métricas *Precision* e *Cluster Recall*. (a) e (b) Q.Q. Plot. (c) e (d) histogramas das distribuições.

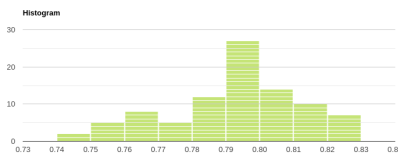
Na Figura B.1 temos as representações comportamentais dessas distribuições para todos os descritores e índices de validação de agrupamento em relação às métricas *Precision* e *Cluster Recall* agrupados em segmentos de ranking (sub-rankings $n = 5$ a $n = 50$). Em B.1(c), o *Q.Q. Plot* (análise de quartis) para a métrica *Precision* onde percebemos a assimetria positiva da distribuição. Em B.1(d), o *Q.Q. Plot* para a métrica *Cluster Recall*, demonstra alta dispersão entre todas as faixas dos sub-ranking e o *padrão de escada* que, normalmente, indica a presença de dados discretos. Os histogramas em B.1(e) e B.1(f), corroboram com a análise feita à partir do teste de *Shapiro-Wilk*, bem como com os gráficos *Quantil-Quantil* (*Q.Q. Plot*) e os *boxplot* em B.1(g) e B.1(h).



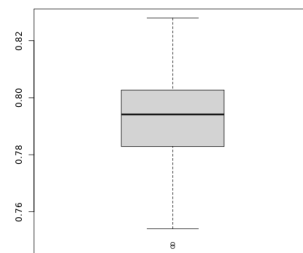
(a) Distribuição dos resultados em $P@5$, superioridade do DTRS nos descritores CNN_AD, GABOR (visuais) e JACCARD (textual). A média do DTRS para esta faixa de ranking foi superior à média geral do *Precision* (média extraída entre as faixas $n = 5$ e $n = 50$).



(b) Q.Q. Plot para $P@5$ mostra distribuição não-normal, assimétrica à direita (negativa).



(c) Histograma da distribuição $P@5$

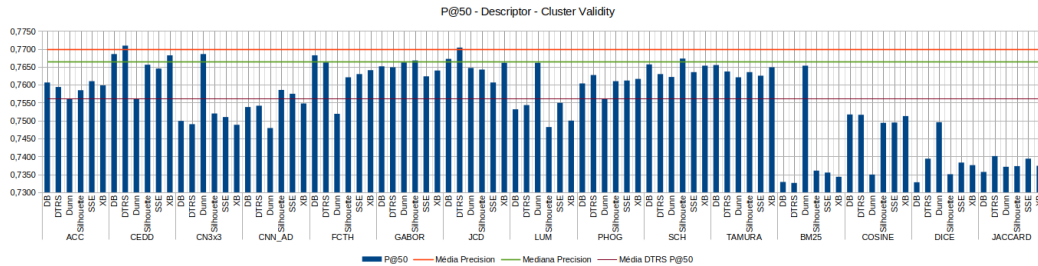


(d) O boxplot de $P@5$ mostra presença de *outliers* e desvio levemente assimétrico negativo.

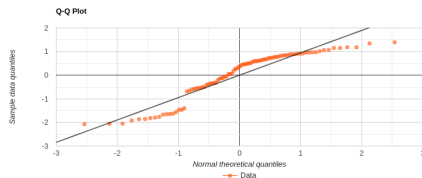
Figura B.2: Comportamento dos resultados por descritores e índices de validação na métrica *Precision*. O gráfico B.2(a) mostra a distribuição geral dos resultados no ranking $n = 5$. O gráfico B.2(b) (*Q.Q. Plot*) indica distribuição não-normal e assimetria negativa.

Nas Figuras B.2 e B.3, temos as distribuições dos resultados e testes estatísticos para as faixas extremas do ranking em relação à métrica *Precision* ($P@5$ e $P@50$). Os gráficos B.2(a) e B.3(a) mostram que o método DTRS apresentou as melhores performances em termos de valores máximos (sem teste de significância) entre os descritores CNN_AD, GABOR e JACCARD ($P@5$) e CEDD, FCTH, JCD, PHOG, COSINE, JACCARD ($P@50$). Em ambos os casos, e de acordo com os testes de normalidade além dos gráficos *Q.Q. Plot* e *histograma*, a tendência não-normal e

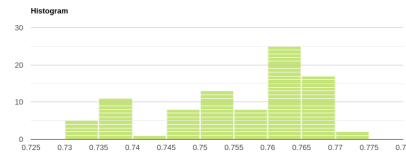
assimétrica foi observada. O p -value para $P@5$ ficou em 0,01256 e para a faixa $P@50$ em 0,000001526. O histograma B.2(c) revelou a presença de *outliers*.



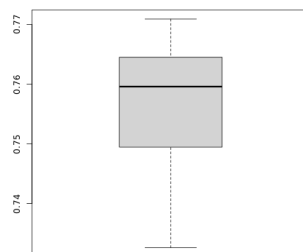
(a) $P@50$: o método DTRS foi melhor que os demais métodos nos descritores CEDD, FCTH, JCD, PHOG (visuais), COSINE e JACCARD (textuais). A média do DTRS para esta faixa de ranking foi inferior à média geral do *Precision* (média extraída entre as faixas $n = 5$ e $n = 50$).



(b) Q.Q. Plot em $P@50$ indicando distribuição não-normal, assimetria positiva e leve dispersão.



(c) Histograma da distribuição $P@50$ corroborando com o Q.Q. Plot.

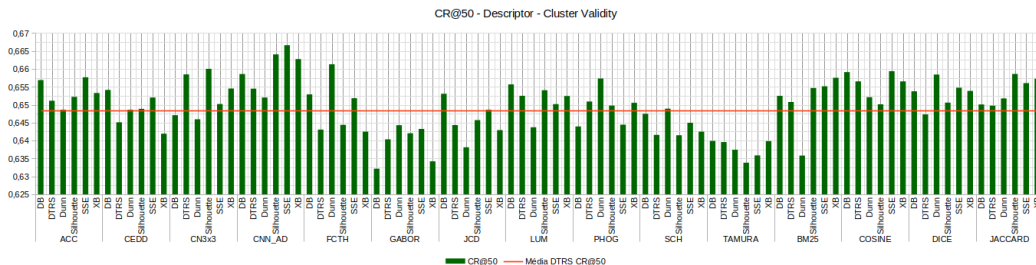


(d) Boxplot $P@50$: a distribuição mostra assimetria negativa.

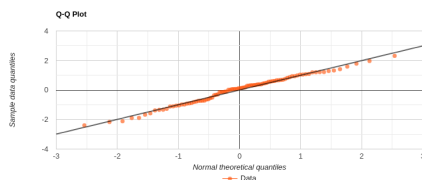
Figura B.3: Comportamento dos resultados para todos os descritores e métodos de validação, métrica *Precision*, segmento final do ranking. O gráfico B.3(a) mostra a distribuição geral dos resultados no ranking $n = 50$. O gráfico B.3(b) (*Q.Q. Plot*) e o boxplot em B.3(d) indicam distribuição não-normal e assimétrica positiva.

A Figura B.4 mostra os resultados gráficos do teste estatístico para a faixa do topo de ranking em relação à métrica *Cluster Recall* ($CR@5$ ou $n = 5$). Pelo gráfico B.4(a) temos que os melhores valores atingidos pelo método DTRS, em comparação aos demais métodos de *cluster validity*, foram GABOR (descritores visuais), BM25

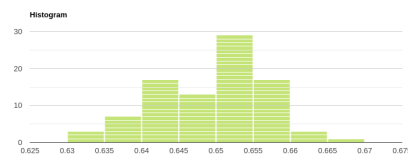
o gráfico da análise dos quartis (*Q.Q. Plot*, Figura B.5(b)). Contudo, o histograma B.5(c) e o boxplot B.5(d) indicam potencial assimetria negativa com o deslocamento da mediana em direção ao terceiro quartil.



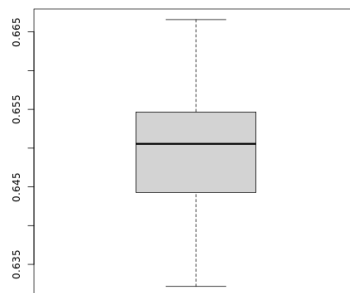
(a) Resultados do *Cluster Recall* para a faixa $n = 50$ do ranking ($CR@50$).



(b) Q.Q. Plot em $CR@50$ indica distribuição potencialmente normal.



(c) O histograma da distribuição em $CR@50$ indica assimetria positiva (tendência à direita).



(d) O boxplot mostra a mediana tendendo ao terceiro quartil. Assimetria negativa e sem presença de *outliers*.

Figura B.5: Análise para a métrica *Cluster Recall* no segmento $n = 50$. O gráfico B.5(a) mostra a distribuição geral para esta faixa e ranking. O gráfico B.5(b) (*Q.Q. Plot*) indica distribuição potencialmente normal, com assimetria negativa. O boxplot em B.5(d) com mediana tendendo ao terceiro quartil e sem a presença de *outliers*. Teste de normalidade realizado com valores médios.

O teste estatístico (*Shapiro-Wilk*) foi realizado, também, para os 222 tópicos de consulta em cada um dos descritores aplicados aos índices de validação de agrupamento e faixas de ranking dos valores de performance obtidos, com resultados semelhantes alcançados (rejeição da hipótese nula em torno de 99,7% dos casos para ambos os algoritmos testados).

B.0.2 Avaliação descritiva dos resultados alcançados nos experimentos 02 e 03

Assim como no experimento anterior (Seção B.0.1), a análise estatística descritiva, em torno dos resultados alcançados na etapa atual, revelou características de normalidade em percentuais compatíveis com a adoção do teste de significância de *Wilcoxon* (normalidade em 99,7% dos casos pelo teste de *Shapiro Wilk*) nos dois experimentos.

Assim, para o experimento 02, temos que:

- Em relação à métrica *Precision*, os melhores resultados concentraram-se no intervalo $[0,7946, 0,8306]$, os piores no intervalo $[0,6874, 0,7523]$, a média $0,7688$, a mediana $0,7665$, o 3º quartil $Q(75\%) = 0,7830$ e a moda $0,7847$. O *p-value*, pelo teste de *Shapiro-Wilk*, foi igual a $1.146e - 06$ ($0,000001146$), o que rejeita a hipótese nula e assume a não-normalidade do conjunto de dados;
- Para a métrica *Cluster Recall*, os melhores resultados concentraram-se no intervalo $[0,6481, 0,6785]$, os piores no intervalo $[0,1296, 0,1487]$, a média $0,4242$, a mediana $0,4506$, o 3º quartil $Q(75\%) = 0,5902$ e a moda $0,1466$. O *p-value*, pelo teste de *Shapiro-Wilk*, foi extremamente próximo de 0 ($< 2.2e-16$), o que rejeita a hipótese nula e assume a não-normalidade do conjunto de dados;
- Em relação à média harmônica (*F1-Measure*, $F1@n$), os melhores valores se concentraram em $[0,6810, 0,6934]$, os piores valores $[0,2168, 0,2459]$, a média $0,5108$, a mediana $0,5575$, o 3º quartil em $Q(75\%) = 0,6493$, moda $0,2452$, $p-value < 2.2e-16$.

Para o experimento 03, temos:

- Em relação à métrica *Precision*, os melhores resultados concentraram-se no intervalo $[0,7568, 0,8207]$, os piores no intervalo $[0,7169, 0,7257]$, a média $0,7375$, a mediana $0,7327$, o 3º quartil $Q(75\%) = 0,7441$ e a moda $0,7441$. O *p-value*, pelo teste de *Shapiro-Wilk*, foi menor que $2.2e - 16$, o que rejeita a hipótese nula e assume a não-normalidade do conjunto de dados;
- Para a métrica *Cluster Recall*, os melhores resultados concentraram-se no intervalo $[0,6635, 0,6769]$, os piores no intervalo $[0,1432, 0,1534]$, a média $0,4367$, a mediana $0,4549$, o 3º quartil $Q(75\%) = 0,6056$ e a moda $0,1511$. O *p-value*, pelo teste de *Shapiro-Wilk*, foi menor que $2.2e - 16$, o que rejeita a hipótese nula e assume a não-normalidade do conjunto de dados;
- Em relação à média harmônica (*F1-Measure*), os melhores valores se concentraram em $[0,6817, 0,6897]$, os piores valores $[0,2364, 0,2502]$, a média $0,5136$,

a mediana 0,5552, o 3º quartil em $Q(75\%) = 0,6492$, moda 0,2489, e *p-value* similar às métricas anteriores, menor que $2.2e - 16$.

A partir da análise estatística, percebemos que os valores dos resultantes de todos os descritores e índices de validação de agrupamento encontram-se muito próximos dos valores médios calculados para as três métricas de performance em todos os experimentos realizados.

Apêndice C

Resultados dos experimentos adicionais

Tabelas de resultado para as três abordagens de seleção do número de grupos ideal (apenas os índices de validação de agrupamento; índices de validação com o método do Cotovelo; e índices de validação com o IASMIN) com aplicação do teste de significância estatística de *Wilcoxon*, valores médios, baseados nos 222 tópicos de consulta da tarefa *MediaEval Diverse Social Media 2015*.

A superioridade do método DTRS (resultados positivos) é destacada em azul. Inferioridade, valores negativos, em vermelho. Demais resultados indicam equivalência estatística.

Tabela C.1: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor TAMURA

	N	Precision								Cluster-Recall (CR)								F1-Measure (F1)							
		DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette			
Elbow+Index	5	0.8171	0.7054	0.8027	0.8081	0.8036	0.8099	0.8099	0.1548	0.1325	0.1544	0.1582	0.1541	0.1539	0.1567	0.2573	0.2209	0.2537	0.2578	0.2540	0.2538	0.2577			
	10	0.8063	0.7014	0.7959	0.8077	0.8032	0.7968	0.8014	0.2611	0.2268	0.2606	0.2628	0.2600	0.2602	0.2606	0.5270	0.3383	0.5347	0.3879	0.3830	0.3851	0.3839			
	20	0.7878	0.6946	0.7892	0.7892	0.7874	0.7878	0.7914	0.4089	0.3691	0.4153	0.4052	0.4098	0.4081	0.4067	0.5270	0.4747	0.5331	0.5250	0.5274	0.5266	0.5249			
	30	0.7779	0.6880	0.7791	0.7793	0.7829	0.7809	0.7821	0.5041	0.4643	0.5069	0.5013	0.5037	0.5039	0.4983	0.5983	0.5444	0.6017	0.5979	0.6001	0.6004	0.5958			
	40	0.7694	0.6867	0.7705	0.7664	0.7726	0.7717	0.7721	0.5763	0.5407	0.5784	0.5709	0.5773	0.5770	0.5704	0.6446	0.5933	0.6469	0.6405	0.6470	0.6463	0.6419			
	50	0.7647	0.6853	0.7633	0.7605	0.7681	0.7659	0.7658	0.6396	0.6002	0.6409	0.6293	0.6402	0.6394	0.6356	0.6810	0.6271	0.6815	0.6741	0.6836	0.6821	0.6794			
apenas Index	5	0.7892	0.7054	0.8009	0.8153	0.7946	0.8108	0.7982	0.1533	0.1325	0.1601	0.1574	0.1582	0.1577	0.1564	0.2523	0.2209	0.2602	0.2592	0.2589	0.2594	0.2591			
	10	0.7950	0.7014	0.7977	0.8072	0.7937	0.8027	0.7973	0.2553	0.2268	0.2643	0.2611	0.2595	0.2611	0.2620	0.5293	0.3383	0.3878	0.3849	0.3836	0.3868	0.3872			
	20	0.7858	0.6946	0.7876	0.7894	0.7874	0.7874	0.7881	0.4062	0.3691	0.4107	0.4078	0.4070	0.4094	0.4103	0.5243	0.4747	0.5283	0.5265	0.5256	0.5273	0.5278			
	30	0.7802	0.6880	0.7824	0.7818	0.7832	0.7805	0.7823	0.5080	0.4643	0.5058	0.5058	0.5082	0.5078	0.5061	0.6023	0.5444	0.6015	0.6017	0.6036	0.6027	0.6008			
	40	0.7729	0.6867	0.7730	0.7721	0.7751	0.7723	0.7720	0.5830	0.5407	0.5805	0.5782	0.5824	0.5824	0.5810	0.6515	0.5933	0.6489	0.6477	0.6523	0.6502	0.6482			
	50	0.7637	0.6853	0.7653	0.7655	0.7649	0.7625	0.7635	0.6396	0.6002	0.6358	0.6399	0.6398	0.6359	0.6338	0.6820	0.6271	0.6796	0.6824	0.6829	0.6802	0.6785	0.6774		
IASMIN+Index	5	0.8099	0.7054	0.8117	0.8081	0.7964	0.8027	0.8126	0.1596	0.1325	0.1546	0.1575	0.1571	0.1554	0.1525	0.2598	0.2209	0.2548	0.2576	0.2558	0.2540	0.2520			
	10	0.8171	0.7014	0.8059	0.8059	0.8158	0.8149	0.8059	0.2694	0.2268	0.2574	0.2662	0.2675	0.2663	0.2562	0.3957	0.3383	0.3829	0.3914	0.3939	0.3930	0.3819			
	20	0.7950	0.6946	0.7966	0.7881	0.7894	0.7881	0.7959	0.4123	0.3691	0.4061	0.4095	0.4121	0.4094	0.4080	0.5316	0.4747	0.5263	0.5270	0.5305	0.5273	0.5280			
	30	0.7860	0.6880	0.7814	0.7811	0.7833	0.7823	0.7826	0.5081	0.4643	0.4993	0.5033	0.5087	0.5089	0.5019	0.6040	0.5444	0.5958	0.5991	0.6042	0.6032	0.5986			
	40	0.7720	0.6867	0.7696	0.7702	0.7707	0.7702	0.7708	0.5780	0.5407	0.5748	0.5796	0.5839	0.5822	0.5741	0.6467	0.5933	0.6434	0.6473	0.6508	0.6489	0.6437			
	50	0.7663	0.6853	0.7634	0.7642	0.7651	0.7644	0.7632	0.6388	0.6002	0.6323	0.6373	0.6406	0.6387	0.6320	0.6819	0.6271	0.6764	0.6798	0.6822	0.6808	0.6767			

Tabela C.2: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor SCH

	N	DTRS	Precision					Silhouette	Cluster-Recall (CR)					Silhouette	F1-Measure (F1)					Silhouette		
			K=20	Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB		SSE	DTRS	K=20	Dunn	DB		XB	SSE
Elbow+Index	5	0.8189	0.6937	0.8162	0.8162	0.8144	0.8153	0.8081	0.1562	0.1290	0.1567	0.1614	0.1582	0.1580	0.1556	0.2573	0.2151	0.2596	0.2649	0.2599	0.2613	0.2556
	10	0.8063	0.6883	0.7955	0.8050	0.8068	0.8018	0.8023	0.2607	0.2190	0.2540	0.2629	0.2654	0.2568	0.2572	0.3870	0.3283	0.3793	0.3888	0.3894	0.3835	0.3819
	20	0.7993	0.6917	0.7894	0.7876	0.7851	0.7842	0.7923	0.4105	0.3642	0.4040	0.4065	0.4107	0.4065	0.4096	0.5322	0.4708	0.5232	0.5249	0.5275	0.5246	0.5286
	30	0.7875	0.6902	0.7808	0.7830	0.7821	0.7778	0.7797	0.5075	0.4618	0.5039	0.5076	0.5069	0.5097	0.5071	0.6062	0.5452	0.6000	0.6038	0.6026	0.6033	0.6024
	40	0.7777	0.6898	0.7739	0.7751	0.7745	0.7712	0.7699	0.5858	0.5475	0.5797	0.5866	0.5812	0.5864	0.5848	0.6545	0.6007	0.6489	0.6547	0.6512	0.6528	0.6512
	50	0.7695	0.6854	0.7650	0.7664	0.7662	0.7650	0.7644	0.6408	0.6046	0.6460	0.6452	0.6417	0.6428	0.6428	0.6844	0.6313	0.6846	0.6864	0.6839	0.6839	0.6848
apenas Index	5	0.8135	0.6937	0.8207	0.8234	0.8027	0.8153	0.8135	0.1577	0.1290	0.1577	0.1589	0.1556	0.1601	0.1600	0.2590	0.2151	0.2610	0.2609	0.2556	0.2626	0.2623
	10	0.8014	0.6883	0.8081	0.8086	0.8005	0.8068	0.8140	0.2607	0.2190	0.2606	0.2604	0.2619	0.2610	0.2633	0.3838	0.3283	0.3882	0.3842	0.3874	0.3873	0.3900
	20	0.7876	0.6917	0.7890	0.7948	0.7876	0.7908	0.7896	0.4081	0.3642	0.4088	0.4089	0.4124	0.4113	0.4147	0.5261	0.4708	0.5263	0.5283	0.5299	0.5303	0.5316
	30	0.7805	0.6902	0.7815	0.7818	0.7818	0.7839	0.7839	0.5131	0.4618	0.5054	0.5098	0.5141	0.5121	0.5091	0.6067	0.5452	0.6014	0.6051	0.6075	0.6073	0.6043
	40	0.7712	0.6898	0.7688	0.7726	0.7725	0.7715	0.7757	0.5818	0.5475	0.5788	0.5872	0.5825	0.5872	0.5838	0.6500	0.6007	0.6474	0.6537	0.6507	0.6537	0.6532
	50	0.7630	0.6854	0.7621	0.7657	0.7653	0.7635	0.7673	0.6416	0.6046	0.6374	0.6475	0.6425	0.6450	0.6415	0.6830	0.6313	0.6804	0.6872	0.6843	0.6850	0.6854
IASMIN+Index	5	0.8099	0.6937	0.8081	0.7973	0.8036	0.8072	0.8063	0.1590	0.1290	0.1554	0.1552	0.1580	0.1577	0.1539	0.2606	0.2151	0.2559	0.2542	0.2587	0.2587	0.2537
	10	0.8068	0.6883	0.8032	0.7995	0.7959	0.7950	0.8032	0.2626	0.2190	0.2599	0.2580	0.2605	0.2585	0.2603	0.3889	0.3283	0.3848	0.3826	0.3856	0.3830	0.3851
	20	0.7833	0.6917	0.7926	0.7791	0.7777	0.7813	0.7917	0.4057	0.3642	0.4080	0.4037	0.4081	0.4068	0.4040	0.5230	0.4708	0.5266	0.5206	0.5248	0.5243	0.5227
	30	0.7802	0.6902	0.7845	0.7763	0.7740	0.7760	0.7814	0.5085	0.4618	0.5063	0.5032	0.5018	0.5001	0.5067	0.6028	0.5452	0.6031	0.5982	0.5970	0.5961	0.6025
	40	0.7729	0.6898	0.7743	0.7695	0.7693	0.7704	0.7723	0.5894	0.5475	0.5819	0.5817	0.5863	0.5826	0.5833	0.6559	0.6007	0.6512	0.6497	0.6524	0.6506	0.6515
	50	0.7646	0.6854	0.7681	0.7657	0.7649	0.7650	0.7669	0.6458	0.6046	0.6460	0.6404	0.6460	0.6432	0.6464	0.6866	0.6313	0.6878	0.6836	0.6862	0.6845	0.6874

Tabela C.3: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor PHOG

	N	DTRS	Precision					Silhouette	Cluster-Recall (CR)					Silhouette	F1-Measure (F1)					Silhouette		
			K=20	Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB		SSE	DTRS	K=20	Dunn	DB		XB	SSE
Elbow+Index	5	0.8144	0.7045	0.8072	0.8261	0.8108	0.8198	0.8198	0.1580	0.1307	0.1520	0.1591	0.1561	0.1545	0.1537	0.2594	0.2180	0.2503	0.2608	0.2563	0.2539	0.2530
	10	0.7977	0.7009	0.7977	0.8059	0.8014	0.7955	0.8014	0.2597	0.2220	0.2598	0.2605	0.2635	0.2597	0.2547	0.3828	0.3304	0.3824	0.3845	0.3873	0.3822	0.3776
	20	0.7824	0.6928	0.7858	0.7892	0.7847	0.7822	0.7878	0.4039	0.3661	0.4046	0.4062	0.4100	0.4058	0.4052	0.5217	0.4701	0.5229	0.5250	0.5274	0.5220	0.5239
	30	0.7733	0.6862	0.7772	0.7785	0.7721	0.7767	0.7761	0.5058	0.4634	0.5042	0.5065	0.5064	0.5113	0.5028	0.5993	0.5427	0.5994	0.6009	0.5987	0.6033	0.5981
	40	0.7705	0.6829	0.7715	0.7694	0.7670	0.7703	0.7690	0.5886	0.5418	0.5869	0.5845	0.5885	0.5872	0.5826	0.6540	0.5918	0.6535	0.6502	0.6526	0.6520	0.6495
	50	0.7604	0.6826	0.7635	0.7627	0.7591	0.7655	0.7647	0.6451	0.6064	0.6470	0.6487	0.6422	0.6508	0.6475	0.6839	0.6288	0.6856	0.6858	0.6816	0.6873	0.6867
apenas Index	5	0.8063	0.7045	0.8099	0.8270	0.8000	0.8225	0.7982	0.1570	0.1307	0.1564	0.1588	0.1564	0.1578	0.1539	0.2574	0.2180	0.2567	0.2608	0.2565	0.2594	0.2537
	10	0.7910	0.7009	0.8009	0.8005	0.7883	0.8014	0.7919	0.2628	0.2220	0.2621	0.2603	0.2612	0.2643	0.2617	0.3874	0.3304	0.3867	0.3836	0.3853	0.3878	0.3842
	20	0.7818	0.6928	0.7838	0.7838	0.7806	0.7867	0.7755	0.4132	0.3661	0.4040	0.4111	0.4122	0.4046	0.4081	0.5297	0.4701	0.5217	0.5280	0.5286	0.5239	0.5238
	30	0.7773	0.6862	0.7734	0.7731	0.7745	0.7761	0.7703	0.5124	0.4634	0.5088	0.5077	0.5108	0.5069	0.5091	0.6042	0.5427	0.6005	0.5996	0.6021	0.6011	0.6005
	40	0.7682	0.6829	0.7679	0.7681	0.7653	0.7684	0.7624	0.5928	0.5418	0.5898	0.5867	0.5923	0.5862	0.5911	0.6552	0.5918	0.6541	0.6517	0.6535	0.6518	0.6524
	50	0.7627	0.6826	0.7622	0.7604	0.7616	0.7612	0.7610	0.6509	0.6064	0.6489	0.6439	0.6505	0.6444	0.6498	0.6869	0.6288	0.6861	0.6827	0.6860	0.6842	0.6860
IASMIN+Index	5	0.8090	0.7045	0.8180	0.8207	0.8171	0.8189	0.8171	0.1534	0.1307	0.1541	0.1528	0.1542	0.1551	0.1535	0.2527	0.2180	0.2541	0.2525	0.2540	0.2553	0.2534
	10	0.8027	0.7009	0.8032	0.8041	0.8014	0.8027	0.8054	0.2615	0.2220	0.2600	0.2567	0.2589	0.2598	0.2595	0.3857	0.3304	0.3838	0.3800	0.3822	0.3832	0.3836
	20	0.7842	0.6928	0.7881	0.7851	0.7782	0.7791	0.7910	0.4102	0.3661	0.4035	0.4077	0.4120	0.4120	0.4047	0.5277	0.4701	0.5229	0.5257	0.5275	0.5278	0.5252
	30	0.7748	0.6862	0.7778	0.7773	0.7727	0.7742	0.7796	0.5092	0.4634	0.5083	0.5102	0.5109	0.5114	0.5100	0.6024	0.5427	0.6028	0.6040	0.6025	0.6031	0.6045
	40	0.7693	0.6829	0.7702	0.7698	0.7657	0.7669	0.7727	0.5857	0.5418	0.5869	0.5873	0.5860	0.5871	0.5882	0.6524	0.5918	0.6536	0.6536	0.6507	0.6519	0.6551
	50	0.7616	0.6826	0.7640	0.7614	0.7582	0.7589	0.7659	0.6445	0.6064	0.6452	0.6453	0.6458	0.6485	0.6480	0.6851	0.6288	0.6852	0.6851	0.6841	0.6859	0.6876

Tabela C.4: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor LUM

	N	DTRS	Precision					Silhouette	Cluster-Recall (CR)					Silhouette	F1-Measure (F1)					Silhouette		
			K=20	Dunn	DB	XB	SSE		DTRS	K=20	Dunn	DB	XB		SSE	DTRS	K=20	Dunn	DB		XB	SSE
Elbow+Index	5	0.8117	0.6919	0.8072	0.8162	0.8117	0.8090	0.8180	0.1542	0.1272	0.1516	0.1577	0.1554	0.1517	0.1543	0.2547	0.2127	0.2506	0.2600	0.2559	0.2507	0.2549
	10	0.8000	0.6919	0.7977	0.7914	0.7946	0.7937	0.7982	0.2595	0.2216	0.2594	0.2593	0.2606	0.2567	0.2608	0.3833	0.3316	0.3829	0.3824	0.3840	0.3791	0.3848
	20	0.7782	0.6865	0.7739	0.7795	0.7755	0.7788	0.7784	0.4101	0.3616	0.4064	0.4102	0.4089	0.4065	0.4051	0.5264	0.4672	0.5216	0.5261	0.5246	0.5230	0.5224
	30	0.7712	0.6896	0.7668	0.7685	0.7650	0.7671	0.7713	0.5157	0.4701	0.5141	0.5148	0.5162	0.5110	0.5121	0.6055	0.5494	0.6026	0.6031	0.6040	0.6010	0.6027
	40	0.7650	0.6882	0.7604	0.7607	0.7577	0.7578	0.7663	0.5917	0.5556	0.5926	0.5947	0.5975	0.5903	0.5926	0.6337	0.6017	0.6524	0.6543	0.6545	0.6497	0.6551
50	0.7586	0.6867	0.7562	0.7533	0.7513	0.7511	0.7587	0.6559	0.6188	0.6518	0.6556	0.6570	0.6503	0.6499	0.6886	0.6357	0.6857	0.6867	0.6870	0.6831	0.6859	
apenas_index	5	0.7982	0.6919	0.8072	0.7955	0.7928	0.7973	0.8036	0.1519	0.1272	0.1561	0.1504	0.1533	0.1546	0.1503	0.2490	0.2127	0.2548	0.2487	0.2505	0.2523	0.2491
	10	0.7946	0.6919	0.7878	0.7952	0.7946	0.8005	0.7896	0.2549	0.2216	0.2574	0.2597	0.2604	0.2566	0.2592	0.3771	0.3316	0.3798	0.3818	0.3842	0.3844	0.3818
	20	0.7757	0.6865	0.7773	0.7727	0.7750	0.7732	0.7723	0.4112	0.3616	0.4075	0.4095	0.4162	0.4055	0.4115	0.5267	0.4672	0.5240	0.5245	0.5298	0.5212	0.5261
	30	0.7680	0.6896	0.7688	0.7638	0.7647	0.7652	0.7616	0.5113	0.4701	0.5143	0.5112	0.5101	0.5086	0.5182	0.6014	0.5494	0.6034	0.6006	0.5987	0.5988	0.6037
	40	0.7588	0.6882	0.7599	0.7580	0.7536	0.7617	0.7554	0.5913	0.5556	0.5938	0.5940	0.5891	0.5863	0.5966	0.6508	0.6017	0.6535	0.6534	0.6469	0.6494	0.6523
50	0.7543	0.6837	0.7561	0.7532	0.7500	0.7550	0.7482	0.6525	0.6188	0.6573	0.6557	0.6521	0.6502	0.6540	0.6853	0.6357	0.6892	0.6877	0.6832	0.6849	0.6831	
IASMIN+index	5	0.8018	0.6919	0.8045	0.8153	0.8135	0.7982	0.8162	0.1566	0.1272	0.1543	0.1574	0.1606	0.1565	0.1535	0.2554	0.2127	0.2521	0.2571	0.2617	0.2571	0.2543
	10	0.8000	0.6919	0.8068	0.7995	0.8036	0.7982	0.8167	0.2630	0.2216	0.2626	0.2618	0.2634	0.2630	0.2607	0.3864	0.3316	0.3876	0.3813	0.3880	0.3873	0.3856
	20	0.7752	0.6865	0.7788	0.7739	0.7752	0.7761	0.7836	0.4099	0.3616	0.4037	0.4061	0.4095	0.4102	0.4040	0.5251	0.4672	0.5212	0.5208	0.5243	0.5252	0.5222
	30	0.7691	0.6896	0.7706	0.7673	0.7671	0.7677	0.7713	0.5090	0.4701	0.5124	0.5136	0.5096	0.5100	0.5083	0.6009	0.5494	0.6032	0.6037	0.6001	0.6002	0.6011
	40	0.7602	0.6882	0.7634	0.7614	0.7597	0.7598	0.7631	0.5878	0.5556	0.5920	0.5858	0.5885	0.5864	0.5879	0.6508	0.6017	0.6551	0.6494	0.6505	0.6489	0.6527
50	0.7519	0.6837	0.7586	0.7538	0.7513	0.7523	0.7577	0.6488	0.6188	0.6527	0.6533	0.6520	0.6520	0.6515	0.6837	0.6357	0.6896	0.6866	0.6845	0.6851	0.6888	

Tabela C.5: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor JCD

	N	DTRS	Precision						Silhouette	Cluster-Recall (CR)						DTRS	F1-Measure (F1)					
			K=20	Dunn	DB	XB	SSE	K=20		Dunn	DB	XB	SSE	K=20	Dunn		DB	XB	SSE	Silhouette		
Elbow+Index	5	0.7856	0.7045	0.7874	0.7847	0.7883	0.7937	0.7874	0.1546	0.1364	0.1556	0.1553	0.1565	0.1547	0.1555	0.2538	0.2264	0.2552	0.2548	0.2567	0.2544	0.2548
	10	0.7486	0.6932	0.7586	0.7563	0.7514	0.7536	0.7473	0.2648	0.2497	0.2701	0.2664	0.2699	0.2665	0.2649	0.3847	0.3634	0.3919	0.3877	0.3907	0.3874	0.3849
	20	0.7556	0.6858	0.7655	0.7599	0.7590	0.7604	0.7552	0.4222	0.3976	0.4293	0.4246	0.4305	0.4243	0.4275	0.5308	0.4960	0.5398	0.5344	0.5397	0.5341	0.5357
	30	0.7662	0.6889	0.7742	0.7707	0.7685	0.7700	0.7631	0.5223	0.4923	0.5294	0.5225	0.5330	0.5253	0.5259	0.6096	0.5650	0.6172	0.6122	0.6184	0.6134	0.6116
	40	0.7662	0.6863	0.7724	0.7708	0.7673	0.7679	0.7648	0.5912	0.5521	0.5966	0.5917	0.6005	0.5923	0.5930	0.6545	0.6012	0.6609	0.6569	0.6613	0.6558	0.6558
50	0.7647	0.6847	0.7685	0.7677	0.7650	0.7680	0.7632	0.6396	0.6090	0.6440	0.6437	0.6520	0.6474	0.6461	0.6827	0.6323	0.6872	0.6861	0.6908	0.6885	0.6867	
apenas Index	5	0.7937	0.7045	0.7928	0.8009	0.7784	0.7937	0.7928	0.1556	0.1364	0.1560	0.1568	0.1536	0.1567	0.1561	0.2556	0.2264	0.2560	0.2579	0.2522	0.2570	0.2564
	10	0.7581	0.6932	0.7509	0.7635	0.7482	0.7541	0.7590	0.2671	0.2497	0.2681	0.2689	0.2677	0.2665	0.2726	0.3882	0.3634	0.3885	0.3910	0.3877	0.3872	0.3946
	20	0.7619	0.6858	0.7520	0.7622	0.7583	0.7543	0.7588	0.4262	0.3976	0.4280	0.4277	0.4274	0.4236	0.4278	0.5371	0.4960	0.5353	0.5382	0.5361	0.5324	0.5368
	30	0.7746	0.6889	0.7673	0.7665	0.7688	0.7557	0.7661	0.5281	0.4923	0.5274	0.5340	0.5259	0.5293	0.5243	0.6172	0.5650	0.6140	0.6178	0.6130	0.6112	0.6115
	40	0.7755	0.6863	0.7700	0.7698	0.7685	0.7628	0.7672	0.5944	0.5521	0.5935	0.6024	0.5881	0.5966	0.5933	0.6613	0.6012	0.6583	0.6634	0.6536	0.6572	0.6564
50	0.7704	0.6847	0.7661	0.7672	0.7661	0.7606	0.7642	0.6443	0.6090	0.6437	0.6530	0.6429	0.6486	0.6457	0.6891	0.6323	0.6863	0.6930	0.6855	0.6873	0.6862	
IASMIN+Index	5	0.7919	0.7045	0.7847	0.7856	0.7802	0.7937	0.7829	0.1554	0.1364	0.1539	0.1537	0.1540	0.1574	0.1541	0.2551	0.2264	0.2526	0.2523	0.2525	0.2581	0.2529
	10	0.7590	0.6932	0.7505	0.7550	0.7468	0.7545	0.7545	0.2676	0.2497	0.2654	0.2693	0.2650	0.2682	0.2664	0.3886	0.3634	0.3850	0.3900	0.3851	0.3888	0.3871
	20	0.7590	0.6858	0.7590	0.7550	0.7518	0.7516	0.7597	0.4258	0.3976	0.4265	0.4257	0.4242	0.4231	0.4250	0.5350	0.4960	0.5356	0.5340	0.5324	0.5314	0.5350
	30	0.7667	0.6889	0.7718	0.7635	0.7619	0.7628	0.7706	0.5311	0.4923	0.5252	0.5292	0.5219	0.5314	0.5205	0.6153	0.5650	0.6143	0.6131	0.6082	0.6146	0.6106
	40	0.7677	0.6863	0.7696	0.7661	0.7644	0.7652	0.7711	0.5925	0.5521	0.5902	0.5921	0.5907	0.5955	0.5852	0.6563	0.6012	0.6562	0.6553	0.6543	0.6575	0.6534
50	0.7659	0.6847	0.7676	0.7654	0.7629	0.7626	0.7690	0.6454	0.6090	0.6449	0.6435	0.6428	0.6464	0.6415	0.6873	0.6323	0.6880	0.6861	0.6842	0.6866	0.6864	

Tabela C.6: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor JACKARD

	N	DTRS	Precision					Silhouette	Cluster-Recall (CR)					DTRS	F1-Measure (F1)							
			K=20	Dunn	DB	XB	SSE		K=20	Dunn	DB	XB	SSE		K=20	Dunn	DB	XB	SSE	Silhouette		
Elbow+Index	5	0.7532	0.6946	0.7532	0.7541	0.7595	0.7586	0.7559	0.1536	0.1428	0.1523	0.1532	0.1577	0.1558	0.1565	0.2528	0.2352	0.2511	0.2523	0.2569	0.2542	0.2550
	10	0.7459	0.6847	0.7441	0.7491	0.7482	0.7509	0.7464	0.2715	0.2502	0.2705	0.2689	0.2715	0.2717	0.2726	0.3923	0.3628	0.3915	0.3904	0.3912	0.3924	0.3918
	20	0.7358	0.6750	0.7367	0.7367	0.7410	0.7396	0.7428	0.4233	0.4026	0.4210	0.4179	0.4270	0.4261	0.4271	0.5269	0.4985	0.5260	0.5237	0.5294	0.5298	0.5299
	30	0.7317	0.6745	0.7323	0.7336	0.7377	0.7345	0.7369	0.5195	0.4979	0.5188	0.5125	0.5186	0.5198	0.5182	0.5943	0.5641	0.5949	0.5911	0.5959	0.5964	0.5955
	40	0.7324	0.6734	0.7324	0.7352	0.7355	0.7340	0.7337	0.5948	0.5704	0.5900	0.5877	0.5922	0.5941	0.5911	0.6421	0.6053	0.6406	0.6405	0.6429	0.6431	0.6409
50	0.7337	0.6727	0.7323	0.7341	0.7343	0.7329	0.7345	0.6486	0.6214	0.6420	0.6402	0.6495	0.6460	0.6479	0.6737	0.6335	0.6701	0.6708	0.6762	0.6724	0.6740	
apenas Index	5	0.7631	0.6946	0.7658	0.7568	0.7649	0.7613	0.7477	0.1603	0.1428	0.1557	0.1551	0.1584	0.1565	0.1542	0.2606	0.2352	0.2564	0.2550	0.2583	0.2555	0.2515
	10	0.7572	0.6847	0.7532	0.7495	0.7626	0.7586	0.7523	0.2770	0.2502	0.2716	0.2680	0.2753	0.2728	0.2702	0.3982	0.3628	0.3937	0.3894	0.3969	0.3943	0.3902
	20	0.7468	0.6750	0.7430	0.7403	0.7417	0.7471	0.7457	0.4306	0.4026	0.4305	0.4282	0.4388	0.4373	0.4372	0.5359	0.4985	0.5349	0.5326	0.5397	0.5410	0.5399
	30	0.7407	0.6745	0.7363	0.7363	0.7408	0.7414	0.7414	0.5238	0.4979	0.5218	0.5216	0.5279	0.5263	0.5295	0.6004	0.5641	0.5975	0.5983	0.6037	0.6033	0.6051
	40	0.7412	0.6734	0.7324	0.7347	0.7345	0.7372	0.7372	0.5947	0.5704	0.5928	0.5921	0.5999	0.5992	0.6000	0.6459	0.6053	0.6407	0.6429	0.6463	0.6476	0.6480
50	0.7401	0.6727	0.7358	0.7357	0.7374	0.7394	0.7373	0.6497	0.6214	0.6536	0.6501	0.6572	0.6560	0.6586	0.6775	0.6335	0.6773	0.6764	0.6801	0.6806	0.6811	
IASMIN+Index	5	0.7595	0.6946	0.7550	0.7613	0.7559	0.7667	0.7559	0.1545	0.1428	0.1542	0.1551	0.1537	0.1599	0.1550	0.2542	0.2352	0.2516	0.2534	0.2531	0.2602	0.2529
	10	0.7527	0.6847	0.7464	0.7518	0.7559	0.7581	0.7473	0.2724	0.2502	0.2686	0.2723	0.2737	0.2767	0.2698	0.3947	0.3628	0.3876	0.3923	0.3961	0.3979	0.3891
	20	0.7394	0.6750	0.7378	0.7392	0.7448	0.7446	0.7385	0.4257	0.4026	0.4152	0.4277	0.4308	0.4254	0.4156	0.5305	0.4985	0.5202	0.5303	0.5346	0.5299	0.5209
	30	0.7357	0.6745	0.7302	0.7374	0.7378	0.7392	0.7312	0.5187	0.4979	0.5071	0.5191	0.5214	0.5167	0.5064	0.5965	0.5641	0.5856	0.5973	0.5978	0.5952	0.5857
	40	0.7345	0.6734	0.7338	0.7378	0.7378	0.7412	0.7346	0.5878	0.5704	0.5808	0.5936	0.5896	0.5849	0.5786	0.6404	0.6053	0.6352	0.6453	0.6426	0.6403	0.6342
50	0.7349	0.6727	0.7330	0.7373	0.7364	0.7404	0.7345	0.6439	0.6214	0.6400	0.6463	0.6417	0.6417	0.6369	0.6727	0.6335	0.6691	0.6749	0.6719	0.6735	0.6681	

Tabela C.7: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor GABOR

	N	DTRS	Precision						Silhouette	Cluster-Recall (CR)						DTRS	F1-Measure (F1)	
--	---	------	-----------	--	--	--	--	--	------------	---------------------	--	--	--	--	--	------	-----------------	--

Tabela C.8: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor FCTH

		Precision							Cluster-Recall (CR)							F1-Measure (F1)							
		N	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7856	0.7009	0.7802	0.7919	0.7928	0.7874	0.7901	0.1567	0.1355	0.1564	0.1574	0.1590	0.1553	0.1553	0.1584	0.2566	0.2251	0.2557	0.2580	0.2602	0.2547	0.2551
	10	0.7482	0.6919	0.7495	0.7514	0.7550	0.7450	0.7514	0.2677	0.2485	0.2710	0.2698	0.2734	0.2663	0.2685	0.2711	0.3880	0.3621	0.3917	0.3907	0.3949	0.3861	0.3890
	20	0.7563	0.6872	0.7644	0.7579	0.7628	0.7538	0.7563	0.4257	0.4006	0.4296	0.4265	0.4347	0.4266	0.4260	0.4264	0.5338	0.4989	0.5396	0.5356	0.5431	0.5344	0.5349
	30	0.7653	0.6892	0.7727	0.7673	0.7698	0.7635	0.7655	0.5228	0.4944	0.5290	0.5278	0.5348	0.5246	0.5295	0.6091	0.5664	0.6159	0.6145	0.6193	0.6100	0.6142	
	40	0.7669	0.6870	0.7723	0.7702	0.7697	0.7652	0.7693	0.5863	0.5557	0.5939	0.5932	0.6004	0.5900	0.5915	0.6509	0.6032	0.6583	0.6573	0.6612	0.6526	0.6554	
	50	0.7637	0.6846	0.7668	0.7655	0.7651	0.7643	0.7650	0.6385	0.6112	0.6421	0.6441	0.6503	0.6452	0.6443	0.6812	0.6331	0.6852	0.6858	0.6892	0.6857	0.6859	
apenas Index	5	0.7865	0.7009	0.7910	0.7982	0.7802	0.7991	0.7901	0.1568	0.1355	0.1583	0.1582	0.1567	0.1597	0.1584	0.1584	0.2567	0.2251	0.2589	0.2595	0.2565	0.2615	0.2593
	10	0.7486	0.6919	0.7486	0.7644	0.7450	0.7577	0.7500	0.2643	0.2485	0.2662	0.2703	0.2684	0.2709	0.2711	0.2711	0.3837	0.3621	0.3859	0.3927	0.3880	0.3924	0.3918
	20	0.7588	0.6872	0.7547	0.7658	0.7538	0.7568	0.7563	0.4254	0.4006	0.4261	0.4305	0.4266	0.4256	0.4264	0.4264	0.5347	0.4989	0.5346	0.5413	0.5350	0.5344	0.5350
	30	0.7695	0.6892	0.7668	0.7689	0.7661	0.7590	0.7634	0.5254	0.4944	0.5284	0.5353	0.5266	0.5316	0.5252	0.6132	0.5664	0.6148	0.6197	0.6126	0.6138	0.6104	
	40	0.7709	0.6870	0.7696	0.7713	0.7662	0.7655	0.7640	0.5902	0.5557	0.5900	0.5997	0.5879	0.5979	0.5910	0.5910	0.6562	0.6032	0.6557	0.6623	0.6528	0.6589	0.6530
	50	0.7662	0.6846	0.7662	0.7682	0.7641	0.7630	0.7621	0.6431	0.6112	0.6443	0.6529	0.6425	0.6518	0.6444	0.6863	0.6331	0.6868	0.6930	0.6846	0.6903	0.6846	
IASMIN+Index	5	0.7847	0.7009	0.7811	0.7928	0.7856	0.7919	0.7838	0.1556	0.1355	0.1554	0.1549	0.1573	0.1579	0.1552	0.1552	0.2549	0.2251	0.2542	0.2541	0.2573	0.2585	0.2544
	10	0.7559	0.6919	0.7505	0.7523	0.7468	0.7532	0.7514	0.2688	0.2485	0.2652	0.2667	0.2665	0.2681	0.2652	0.2652	0.3897	0.3621	0.3849	0.3869	0.3867	0.3890	0.3853
	20	0.7563	0.6872	0.7586	0.7514	0.7545	0.7502	0.7574	0.4277	0.4006	0.4249	0.4231	0.4234	0.4236	0.4220	0.4236	0.5357	0.4989	0.5346	0.5304	0.5320	0.5314	0.5312
	30	0.7650	0.6892	0.7691	0.7607	0.7650	0.7590	0.7698	0.5321	0.4944	0.5227	0.5268	0.5277	0.5318	0.5192	0.6154	0.5664	0.6112	0.6104	0.6132	0.6140	0.6085	
	40	0.7675	0.6870	0.7708	0.7644	0.7667	0.7651	0.7707	0.5951	0.5557	0.5866	0.5914	0.5930	0.5951	0.5841	0.6572	0.6032	0.6537	0.6535	0.6559	0.6568	0.6515	
	50	0.7655	0.6846	0.7668	0.7623	0.7644	0.7623	0.7667	0.6484	0.6112	0.6414	0.6437	0.6432	0.6457	0.6408	0.6881	0.6331	0.6849	0.6844	0.6846	0.6856	0.6843	

Tabela C.9: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor DICE

		Precision							Cluster-Recall (CR)							F1-Measure (F1)							
		N	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7658	0.7009	0.7640	0.7586	0.7396	0.7486	0.7468	0.1577	0.1415	0.1563	0.1550	0.1532	0.1543	0.1545	0.1545	0.2568	0.2339	0.2549	0.2528	0.2492	0.2510	0.2513
	10	0.7599	0.6761	0.7554	0.7577	0.7514	0.7532	0.7527	0.2775	0.2486	0.2742	0.2726	0.2751	0.2767	0.2756	0.3989	0.3601	0.3954	0.3934	0.3952	0.3974	0.3957	
	20	0.7437	0.6764	0.7408	0.7403	0.7383	0.7399	0.7396	0.4225	0.3982	0.4258	0.4210	0.4257	0.4268	0.4249	0.5274	0.4947	0.5290	0.5253	0.5276	0.5307	0.5267	
	30	0.7369	0.6764	0.7363	0.7384	0.7371	0.7345	0.7368	0.5174	0.4936	0.5175	0.5161	0.5179	0.5186	0.5174	0.5949	0.5625	0.5958	0.5947	0.5954	0.5958	0.5941	
	40	0.7366	0.6764	0.7360	0.7380	0.7354	0.7318	0.7350	0.5930	0.5650	0.5925	0.5865	0.5899	0.5903	0.5906	0.6435	0.6058	0.6429	0.6398	0.6411	0.6406	0.6411	
	50	0.7357	0.6732	0.7340	0.7370	0.7335	0.7326	0.7343	0.6470	0.6162	0.6477	0.6447	0.6474	0.6482	0.6473	0.6744	0.6325	0.6734	0.6728	0.6734	0.6737	0.6731	
apenas Index	5	0.7541	0.7009	0.7577	0.7595	0.7649	0.7595	0.7486	0.1524	0.1415	0.1551	0.1537	0.1563	0.1568	0.1539	0.1539	0.2491	0.2339	0.2526	0.2511	0.2548	0.2554	0.2505
	10	0.7685	0.6761	0.7595	0.7563	0.7613	0.7644	0.7586	0.2803	0.2486	0.2768	0.2750	0.2795	0.2774	0.2765	0.4034	0.3601	0.3985	0.3963	0.4016	0.3997	0.3976	
	20	0.7448	0.6764	0.7428	0.7376	0.7390	0.7439	0.7441	0.4301	0.3982	0.4332	0.4237	0.4362	0.4325	0.4332	0.5333	0.4947	0.5358	0.5279	0.5368	0.5353	0.5357	
	30	0.7408	0.6764	0.7404	0.7351	0.7422	0.7417	0.7402	0.5195	0.4936	0.5219	0.5227	0.5252	0.5230	0.5261	0.5984	0.5625	0.5994	0.5981	0.6027	0.6015	0.6028	
	40	0.7404	0.6764	0.7359	0.7339	0.7367	0.7369	0.7346	0.5899	0.5650	0.5962	0.5959	0.5983	0.5968	0.5924	0.6432	0.6058	0.6454	0.6447	0.6468	0.6466	0.6431	
	50	0.7394	0.6732	0.7371	0.7328	0.7376	0.7383	0.7350	0.6473	0.6162	0.6517	0.6537	0.6538	0.6547	0.6506	0.6757	0.6325	0.6773	0.6769	0.6787	0.6799	0.6757	
IASMIN+Index	5	0.7607	0.7009	0.7523	0.7541	0.7640	0.7607	0.7514	0.1559	0.1415	0.1537	0.1529	0.1556	0.1566	0.1535	0.1535	0.2545	0.2339	0.2506	0.2496	0.2563	0.2554	0.2503
	10	0.7590	0.6761	0.7500	0.7545	0.7635	0.7541	0.7495	0.2754	0.2486	0.2722	0.2728	0.2778	0.2722	0.2718	0.3970	0.3601	0.3921	0.3931	0.3997	0.3925	0.3916	
	20	0.7455	0.6764	0.7396	0.7423	0.7466	0.7423	0.7390	0.4332	0.3982	0.4158	0.4237	0.4340	0.4261	0.4153	0.5367	0.4947	0.5212	0.5284	0.5374	0.5304	0.5206	
	30	0.7395	0.6764	0.7329	0.7362	0.7393	0.7365	0.7333	0.5242	0.4936	0.5061	0.5165	0.5202	0.5183	0.5064	0.6015	0.5625	0.5870	0.5945	0.5980	0.5957	0.5873	
	40	0.7346	0.6764	0.7368	0.7378	0.7390	0.7369	0.7377	0.5920	0.5650	0.5835	0.5878	0.5948	0.5883	0.5830	0.6418	0.6058	0.6381	0.6407	0.6460	0.6405	0.6381	
	50	0.7356	0.6732	0.7355	0.7362	0.7387	0.7370	0.7358	0.6485	0.6162	0.6415	0.6444	0.6502	0.6443	0.6406	0.6747	0.6325	0.6713	0.6731	0.6772	0.6730	0.6708	

Tabela C.10: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor COSINE

	N	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
		DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7901	0.6838	0.7928	0.7820	0.7775	0.7829	0.7874	0.1555	0.1369	0.1569	0.1560	0.1514	0.1561	0.1549	0.2567	0.2262	0.2592	0.2555	0.2513	0.2575	0.2560
	10	0.7752	0.6775	0.7681	0.7653	0.7676	0.7692	0.7752	0.2726	0.2396	0.2705	0.2715	0.2690	0.2703	0.2754	0.3981	0.4492	0.3924	0.3943	0.3917	0.3932	0.3950
	20	0.7583	0.6813	0.7547	0.7579	0.7568	0.7561	0.7615	0.4295	0.3923	0.4274	0.4314	0.4262	0.4312	0.4329	0.5368	0.5115	0.5350	0.5411	0.5350	0.5387	0.5414
	30	0.7533	0.6827	0.7524	0.7533	0.7565	0.7529	0.7545	0.5279	0.4890	0.5288	0.5338	0.5330	0.5285	0.5345	0.6090	0.5610	0.6095	0.6132	0.6123	0.6098	0.6145
	40	0.7524	0.6843	0.7528	0.7533	0.7564	0.7523	0.7517	0.5981	0.5625	0.5987	0.6027	0.6038	0.6003	0.6012	0.6546	0.6072	0.6543	0.6572	0.6594	0.6557	0.6563
50	0.7484	0.6861	0.7513	0.7502	0.7532	0.7496	0.7501	0.6508	0.6146	0.6509	0.6534	0.6548	0.6548	0.6581	0.6833	0.6369	0.6840	0.6852	0.6894	0.6860	0.6877	
apenas_index	5	0.7829	0.6838	0.7919	0.7838	0.7919	0.7811	0.7829	0.1565	0.1369	0.1555	0.1532	0.1573	0.1556	0.1569	0.2559	0.2262	0.2572	0.2571	0.2596	0.2551	0.2569
	10	0.7802	0.6775	0.7830	0.7725	0.7775	0.7739	0.7779	0.2745	0.2396	0.2726	0.2736	0.2760	0.2741	0.2736	0.3996	0.4495	0.3962	0.3962	0.4004	0.3977	0.3981
	20	0.7651	0.6813	0.7586	0.7590	0.7651	0.7622	0.7581	0.4322	0.3923	0.4295	0.4339	0.4336	0.4352	0.4246	0.5422	0.5115	0.5384	0.5421	0.5436	0.5427	0.5337
	30	0.7593	0.6827	0.7574	0.7578	0.7610	0.7530	0.7560	0.5330	0.4890	0.5329	0.5333	0.5335	0.5358	0.5293	0.6155	0.5610	0.6131	0.6127	0.6146	0.6141	0.6116
	40	0.7538	0.6843	0.7516	0.7566	0.7570	0.7542	0.7519	0.6044	0.5625	0.6025	0.6074	0.6045	0.6106	0.5983	0.6594	0.6072	0.6566	0.6616	0.6605	0.6617	0.6544
50	0.7516	0.6861	0.7495	0.7517	0.7513	0.7495	0.7494	0.6565	0.6146	0.6584	0.6591	0.6565	0.6593	0.6501	0.6879	0.6369	0.6883	0.6897	0.6883	0.6882	0.6833	
FASM-IN+Index	5	0.7811	0.6838	0.7838	0.7865	0.7868	0.7883	0.7784	0.1555	0.1369	0.1538	0.1543	0.1548	0.1582	0.1532	0.2566	0.2262	0.2546	0.2549	0.2559	0.2587	0.2535
	10	0.7703	0.6775	0.7777	0.7761	0.7667	0.7653	0.7595	0.2710	0.2396	0.2689	0.2690	0.2701	0.2689	0.2695	0.3939	0.4495	0.3900	0.3929	0.3928	0.3908	0.3908
	20	0.7541	0.6813	0.7552	0.7610	0.7604	0.7536	0.7554	0.4315	0.3923	0.4269	0.4300	0.4276	0.4263	0.4252	0.5386	0.5115	0.5359	0.5391	0.5369	0.5343	0.5344
	30	0.7532	0.6827	0.7563	0.7572	0.7575	0.7525	0.7566	0.5327	0.4890	0.5321	0.5333	0.5305	0.5308	0.5307	0.6126	0.5610	0.6142	0.6143	0.6128	0.6115	0.6132
	40	0.7528	0.6843	0.7533	0.7541	0.7546	0.7505	0.7534	0.6033	0.5625	0.5992	0.6044	0.6012	0.5985	0.5986	0.6577	0.6072	0.6553	0.6585	0.6563	0.6539	0.6549
50	0.7500	0.6861	0.7499	0.7513	0.7506	0.7484	0.7504	0.6557	0.6146	0.6524	0.6564	0.6556	0.6524	0.6521	0.6870	0.6369	0.6850	0.6876	0.6861	0.6843	0.685	

Tabela C.11: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CNN_AD

Precision										Cluster-Recall (CR)										F1-Measure (F1)									
	N	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette							
Elbow+Index	5	0.7928	0.7189	0.8099	0.8063	0.8045	0.8009	0.7991	0.1495	0.1298	0.1557	0.1534	0.1520	0.1549	0.1558	0.2469	0.2169	0.2539	0.2527	0.2506	0.2539	0.2536							
	10	0.7901	0.7018	0.7941	0.7964	0.7860	0.7838	0.7901	0.2557	0.2233	0.2626	0.2586	0.2625	0.2589	0.2641	0.3790	0.3338	0.3841	0.3830	0.3838	0.3815	0.3862							
	20	0.7682	0.6912	0.7741	0.7741	0.7759	0.7791	0.7806	0.4098	0.3712	0.4091	0.4106	0.4164	0.4117	0.4131	0.5214	0.4754	0.5229	0.5254	0.5302	0.5284	0.5282							
	30	0.7572	0.6889	0.7626	0.7634	0.7622	0.7689	0.7668	0.5106	0.4801	0.5101	0.5157	0.5210	0.5207	0.5192	0.5961	0.5555	0.5978	0.6034	0.6059	0.6087	0.6059							
	40	0.7541	0.6870	0.7553	0.7578	0.7570	0.7639	0.7591	0.5903	0.5549	0.5983	0.5967	0.5964	0.6034	0.5979	0.6481	0.6022	0.6540	0.6545	0.6528	0.6613	0.6553							
50	0.7524	0.6836	0.7509	0.7519	0.7541	0.7556	0.7525	0.6514	0.6141	0.6575	0.6594	0.6613	0.6616	0.6618	0.6833	0.6345	0.6858	0.6883	0.6890	0.6913	0.6898								
apenas Index	5	0.8090	0.7189	0.8000	0.7874	0.7910	0.7982	0.7991	0.1560	0.1298	0.1495	0.1505	0.1511	0.1515	0.1495	0.2559	0.2169	0.2469	0.2474	0.2481	0.2493	0.2464							
	10	0.7932	0.7018	0.7923	0.7860	0.7928	0.7824	0.7941	0.2596	0.2233	0.2599	0.2586	0.2580	0.2524	0.2541	0.3834	0.3338	0.3835	0.3817	0.3819	0.3741	0.3776							
	20	0.7766	0.6912	0.7673	0.7723	0.7743	0.7655	0.7759	0.4090	0.3712	0.4128	0.4110	0.4114	0.4163	0.4097	0.5262	0.4754	0.5259	0.5250	0.5275	0.5282	0.5267							
	30	0.7668	0.6889	0.7605	0.7628	0.7668	0.7575	0.7677	0.5076	0.4801	0.5180	0.5171	0.5130	0.5207	0.5158	0.5993	0.5555	0.6040	0.6039	0.6018	0.6041	0.6037							
	40	0.7596	0.6870	0.7560	0.7582	0.7577	0.7556	0.7624	0.5911	0.5549	0.5987	0.5969	0.5983	0.6044	0.6032	0.6526	0.6022	0.6554	0.6547	0.6550	0.6581	0.6598							
50	0.7541	0.6836	0.7519	0.7538	0.7548	0.7575	0.7586	0.6545	0.6141	0.6613	0.6586	0.6627	0.6666	0.6640	0.6870	0.6345	0.6899	0.6888	0.6913	0.6950	0.6938								
IASMIN+Index	5	0.7946	0.7189	0.8099	0.8099	0.7964	0.8009	0.8000	0.1548	0.1298	0.1513	0.1561	0.1526	0.1531	0.1509	0.2515	0.2169	0.2476	0.2564	0.2510	0.2521	0.2470							
	10	0.7901	0.7018	0.7838	0.7905	0.7851	0.7910	0.7878	0.2623	0.2233	0.2605	0.2627	0.2578	0.2634	0.2603	0.3833	0.3338	0.3818	0.3841	0.3784	0.3858	0.3820							
	20	0.7705	0.6912	0.7705	0.7718	0.7631	0.7676	0.7716	0.4120	0.3712	0.4044	0.4167	0.4076	0.4160	0.4080	0.5250	0.4754	0.5196	0.5294	0.5205	0.5286	0.5232							
	30	0.7625	0.6889	0.7638	0.7595	0.7536	0.7577	0.7644	0.5155	0.4801	0.5097	0.5174	0.5068	0.5127	0.5116	0.6015	0.5555	0.5992	0.6024	0.5936	0.5989	0.6010							
	40	0.7569	0.6870	0.7562	0.7546	0.7528	0.7571	0.7564	0.5957	0.5549	0.5977	0.5995	0.5926	0.5966	0.5971	0.6529	0.6022	0.6553	0.6542	0.6497	0.6539	0.6548							
50	0.7527	0.6836	0.7520	0.7510	0.7515	0.7536	0.7518	0.6522	0.6141	0.6584	0.6589	0.6507	0.6545	0.6600	0.6840	0.6345	0.6879	0.6874	0.6833	0.6868	0.6889								

Tabela C.12: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CN3X3

	N	Precision							Cluster-Recall (CR)							F1-Measure (F1)						
		DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7865	0.7018	0.7820	0.7829	0.7811	0.8063	0.7721	0.1628	0.1390	0.1564	0.1579	0.1590	0.1631	0.1560	0.2629	0.2299	0.2560	0.2563	0.2577	0.2667	0.2530
	10	0.7829	0.7009	0.7752	0.7811	0.7757	0.7815	0.7734	0.2777	0.2395	0.2707	0.2680	0.2711	0.2706	0.2692	0.4007	0.3526	0.3925	0.3899	0.3932	0.3947	0.3902
	20	0.7709	0.6878	0.7705	0.7700	0.7667	0.7703	0.7653	0.4358	0.3928	0.4350	0.4317	0.4308	0.4299	0.4310	0.5459	0.4933	0.5459	0.5431	0.5415	0.5420	0.5405
	30	0.7601	0.6943	0.7610	0.7616	0.7605	0.7596	0.7611	0.5273	0.4862	0.5336	0.5284	0.5307	0.5243	0.5283	0.6095	0.5632	0.6142	0.6119	0.6132	0.6074	0.6112
	40	0.7542	0.6884	0.7579	0.7584	0.7581	0.7553	0.7587	0.5936	0.5561	0.5989	0.5957	0.5988	0.5927	0.5983	0.6499	0.6038	0.6565	0.6544	0.6555	0.6509	0.6562
50	0.7486	0.6882	0.7525	0.7532	0.7497	0.7506	0.7513	0.6473	0.6142	0.6508	0.6442	0.6482	0.6433	0.6564	0.6791	0.6367	0.6835	0.6809	0.6803	0.6788	0.6865	
apenas Index	5	0.7730	0.7018	0.7838	0.7892	0.7937	0.7730	0.7838	0.1591	0.1390	0.1604	0.1651	0.1605	0.1588	0.1569	0.2569	0.2299	0.2615	0.2664	0.2618	0.2668	0.2586
	10	0.7680	0.7009	0.7856	0.7815	0.7761	0.7797	0.7784	0.2676	0.2395	0.2750	0.2778	0.2691	0.2759	0.2663	0.3871	0.3526	0.3996	0.4003	0.3925	0.3983	0.3917
	20	0.7653	0.6878	0.7698	0.7671	0.7664	0.7705	0.7694	0.4351	0.3928	0.4369	0.4314	0.4312	0.4376	0.4373	0.5428	0.4933	0.5465	0.5415	0.5410	0.5473	0.5464
	30	0.7571	0.6943	0.7614	0.7604	0.7623	0.7644	0.7629	0.5292	0.4862	0.5289	0.5292	0.5282	0.5317	0.5395	0.6102	0.5632	0.6117	0.6116	0.6126	0.6151	0.6198
	40	0.7521	0.6884	0.7535	0.7548	0.7543	0.7574	0.7544	0.6044	0.5561	0.5961	0.5956	0.6002	0.5975	0.6069	0.6569	0.6038	0.6520	0.6527	0.6563	0.6547	0.6603
50	0.7490	0.6882	0.7479	0.7499	0.7488	0.7510	0.7520	0.6585	0.6142	0.6520	0.6471	0.6545	0.6502	0.6600	0.6865	0.6367	0.6828	0.6806	0.6848	0.6825	0.6890	
IASMIN+Index	5	0.7901	0.7018	0.7928	0.7937	0.7856	0.7991	0.7928	0.1618	0.1390	0.1612	0.1586	0.1558	0.1633	0.1611	0.2642	0.2299	0.2636	0.2596	0.2616	0.2647	0.2633
	10	0.7820	0.7009	0.7937	0.7851	0.7892	0.7847	0.7901	0.2724	0.2395	0.2743	0.2685	0.2712	0.2727	0.2731	0.3967	0.3526	0.4008	0.3929	0.3958	0.3956	0.3985
	20	0.7707	0.6878	0.7755	0.7698	0.7741	0.7723	0.7757	0.4327	0.3928	0.4278	0.4336	0.4312	0.4360	0.4301	0.5434	0.4933	0.5412	0.5438	0.5430	0.5468	0.5432
	30	0.7640	0.6943	0.7656	0.7659	0.7676	0.7643	0.7653	0.5240	0.4862	0.5234	0.5263	0.5268	0.5266	0.5265	0.6096	0.5632	0.6107	0.6124	0.6128	0.6122	0.6125
	40	0.7591	0.6884	0.7627	0.7614	0.7614	0.7589	0.7607	0.5961	0.5561	0.5951	0.5981	0.5967	0.5969	0.5952	0.6553	0.6038	0.6560	0.6577	0.6564	0.6563	0.6551
50	0.7508	0.6882	0.7544	0.7531	0.7535	0.7513	0.7540	0.6460	0.6142	0.6458	0.6477	0.6497	0.6477	0.6456	0.6807	0.6367	0.6818	0.6827	0.6840	0.6818	0.6814	

Tabela C.13: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor CEDD

	Precision							Cluster-Recall (CR)							F1-Measure (F1)							
	N	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7901	0.7081	0.7928	0.7919	0.8018	0.7928	0.8009	0.1572	0.1360	0.1572	0.1577	0.1583	0.1556	0.1604	0.2576	0.2260	0.2576	0.2584	0.2596	0.2557	0.2624
	10	0.7495	0.6919	0.7572	0.7495	0.7568	0.7509	0.7541	0.2672	0.2476	0.2703	0.2657	0.2696	0.2631	0.2713	0.3876	0.3609	0.3918	0.3862	0.3909	0.3836	0.3919
	20	0.7595	0.6860	0.7622	0.7547	0.7642	0.7595	0.7608	0.4208	0.3987	0.4264	0.4218	0.4316	0.4222	0.4270	0.5350	0.4699	0.5364	0.5311	0.5415	0.5322	0.5369
	30	0.7692	0.6895	0.7712	0.7679	0.7709	0.7695	0.7685	0.5276	0.4951	0.5252	0.5245	0.5327	0.5225	0.5288	0.6143	0.5674	0.6133	0.6122	0.6185	0.6129	0.6148
	40	0.7687	0.6866	0.7070	0.7720	0.7717	0.7687	0.7686	0.5945	0.5543	0.5912	0.5925	0.6013	0.5897	0.5913	0.6570	0.6027	0.6564	0.6577	0.6632	0.6543	0.6558
50	0.7660	0.6847	0.7666	0.7675	0.7689	0.7675	0.7645	0.6452	0.6101	0.6408	0.6457	0.6534	0.6440	0.6430	0.6863	0.6332	0.6841	0.6872	0.6933	0.6867	0.6850	
apenas_index	5	0.7991	0.7081	0.7964	0.8009	0.7865	0.7964	0.7919	0.1569	0.1360	0.1585	0.1587	0.1559	0.1539	0.1580	0.2574	0.2260	0.2594	0.2601	0.2560	0.2536	0.2602
	10	0.7595	0.6919	0.7509	0.7604	0.7491	0.7581	0.7532	0.2658	0.2476	0.2692	0.2670	0.2607	0.2634	0.2730	0.3868	0.3609	0.3892	0.3881	0.3868	0.3881	0.3941
	20	0.7635	0.6860	0.7595	0.7624	0.7568	0.7574	0.7559	0.4252	0.3987	0.4301	0.4284	0.4263	0.4273	0.4316	0.5365	0.4699	0.5388	0.5388	0.5354	0.5367	0.5382
	30	0.7737	0.6895	0.7721	0.7683	0.7698	0.7616	0.7664	0.5284	0.4951	0.5330	0.5324	0.5258	0.5231	0.5267	0.6170	0.5674	0.6195	0.6178	0.6135	0.6183	0.6127
	40	0.7751	0.6866	0.7739	0.7720	0.7680	0.7651	0.7671	0.5925	0.5543	0.5983	0.6017	0.5888	0.5980	0.5954	0.6506	0.6027	0.6623	0.6639	0.6543	0.6591	0.6574
50	0.7709	0.6847	0.7686	0.7686	0.7682	0.7645	0.7656	0.6451	0.6101	0.6459	0.6451	0.6419	0.6250	0.6488	0.6899	0.6332	0.6887	0.6940	0.6860	0.6913	0.6885	
IASMIN+index	5	0.7928	0.7081	0.7883	0.7865	0.7892	0.7901	0.7928	0.1567	0.1360	0.1560	0.1550	0.1576	0.1568	0.1561	0.2570	0.2260	0.2555	0.2543	0.2579	0.2570	0.2562
	10	0.7590	0.6919	0.7545	0.7532	0.7477	0.7523	0.7554	0.2710	0.2476	0.2691	0.2673	0.2686	0.2682	0.2660	0.3925	0.3609	0.3896	0.3880	0.3889	0.3875	0.3865
	20	0.7622	0.6860	0.7610	0.7568	0.7529	0.7545	0.7604	0.4292	0.3987	0.4257	0.4253	0.4231	0.4249	0.4213	0.5390	0.4699	0.5363	0.5348	0.5322	0.5340	0.5323
	30	0.7676	0.6895	0.7727	0.7661	0.7634	0.7640	0.7721	0.5331	0.4951	0.5246	0.5300	0.5257	0.5239	0.5195	0.6176	0.5674	0.6141	0.6153	0.6118	0.6167	0.6100
	40	0.7677	0.6866	0.7717	0.7680	0.7667	0.7663	0.7715	0.5983	0.5543	0.5905	0.5959	0.5923	0.5968	0.5860	0.6597	0.6027	0.6571	0.6588	0.6558	0.6584	0.6539
50	0.7658	0.6847	0.7674	0.7657	0.7648	0.7644	0.7690	0.6489	0.6101	0.6424	0.6428	0.6432	0.6459	0.6414	0.6892	0.6332	0.6863	0.6895	0.6849	0.6873	0.6867	

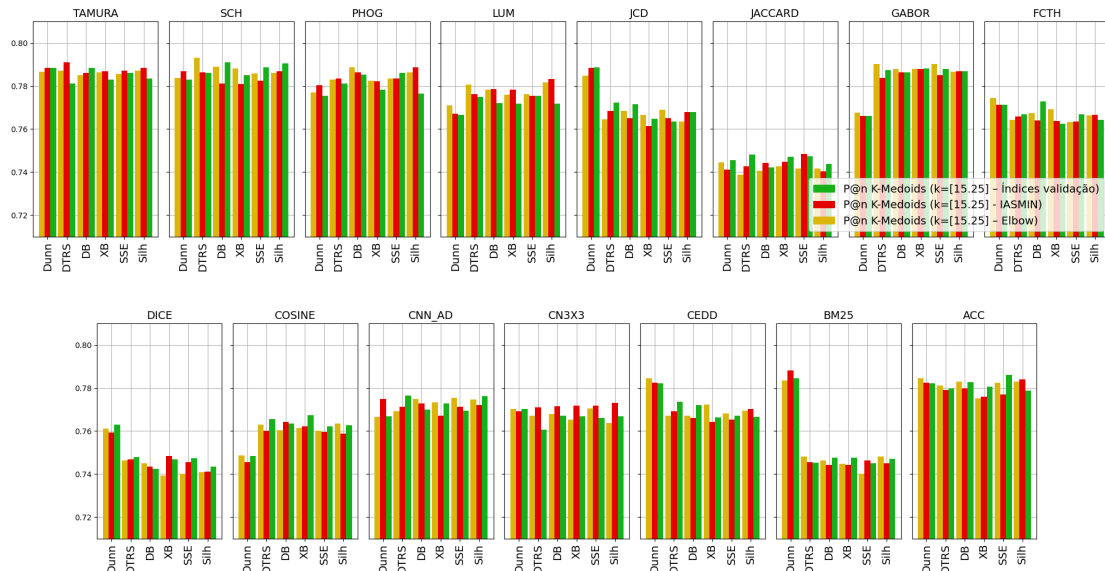
Tabela C.14: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor BM25

	N	DTRS	Precision						DTRS	Cluster-Recall (CR)						DTRS	F1-Measure (F1)					
			K=20	Dunn	DB	XB	SSE	Silhouette		K=20	Dunn	DB	XB	SSE	Silhouette		K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.7739	0.7063	0.7703	0.7631	0.7667	0.7586	0.7676	0.1597	0.1468	0.1606	0.1578	0.1612	0.1574	0.1602	0.2604	0.2412	0.2614	0.2569	0.2621	0.2561	0.2606
	10	0.7608	0.6869	0.7622	0.7572	0.7532	0.7518	0.7599	0.2745	0.2517	0.2725	0.2722	0.2736	0.2708	0.2766	0.3964	0.3649	0.3948	0.3934	0.3949	0.3913	0.3982
	20	0.7464	0.6768	0.7502	0.7480	0.7428	0.7390	0.7484	0.4267	0.4074	0.4306	0.4277	0.4264	0.4203	0.4320	0.5315	0.5005	0.5354	0.5329	0.5301	0.5243	0.5368
	30	0.7396	0.6794	0.7384	0.7404	0.7366	0.7317	0.7422	0.5201	0.5053	0.5226	0.5217	0.5223	0.5142	0.5243	0.5979	0.5686	0.5998	0.5987	0.5982	0.5913	0.6025
	40	0.7341	0.6762	0.7368	0.7360	0.7338	0.7298	0.7366	0.5966	0.5764	0.5948	0.5927	0.5938	0.5888	0.5951	0.6444	0.6097	0.6447	0.6422	0.6426	0.6378	0.6450
	50	0.7334	0.6757	0.7345	0.7339	0.7362	0.7296	0.7338	0.6494	0.6276	0.6528	0.6490	0.6512	0.6451	0.6481	0.6746	0.6380	0.6761	0.6742	0.6769	0.6707	0.6739
apenas Index	5	0.7685	0.7063	0.7712	0.7730	0.7631	0.7604	0.7622	0.1601	0.1468	0.1577	0.1602	0.1560	0.1572	0.1565	0.2607	0.2412	0.2578	0.2609	0.2545	0.2563	0.2553
	10	0.7586	0.6869	0.7644	0.7595	0.7658	0.7563	0.7604	0.2731	0.2517	0.2776	0.2726	0.2734	0.2734	0.2708	0.3945	0.3649	0.4005	0.3947	0.3957	0.3946	0.3920
	20	0.7428	0.6768	0.7450	0.7444	0.7455	0.7450	0.7462	0.4310	0.4074	0.4349	0.4306	0.4364	0.4337	0.4311	0.5341	0.5005	0.5380	0.5344	0.5389	0.5360	0.5337
	20	0.7380	0.6794	0.7396	0.7408	0.7425	0.7377	0.7402	0.5234	0.5053	0.5244	0.5230	0.5305	0.5276	0.5235	0.6004	0.5686	0.6015	0.6011	0.6063	0.6020	0.5998
	40	0.7319	0.6762	0.7347	0.7347	0.7350	0.7346	0.7373	0.5961	0.5764	0.5976	0.5954	0.6020	0.5956	0.5982	0.6436	0.6097	0.6454	0.6445	0.6479	0.6436	0.6460
	50	0.7326	0.6757	0.7350	0.7329	0.7343	0.7355	0.7360	0.6508	0.6276	0.6521	0.6525	0.6575	0.6551	0.6547	0.6754	0.6380	0.6767	0.6764	0.6791	0.6782	0.6781
IASMIN+Index	5	0.7685	0.7063	0.7793	0.7676	0.7604	0.7793	0.7775	0.1603	0.1468	0.1615	0.1592	0.1566	0.1606	0.1613	0.2611	0.2412	0.2633	0.2593	0.2552	0.2619	0.2629
	10	0.7509	0.6869	0.7509	0.7563	0.7550	0.7532	0.7495	0.2715	0.2517	0.2711	0.2747	0.2733	0.2733	0.2705	0.3924	0.3649	0.3918	0.3963	0.3945	0.3938	0.3909
	20	0.7444	0.6768	0.7432	0.7421	0.7432	0.7441	0.7432	0.4282	0.4074	0.4183	0.4277	0.4268	0.4243	0.4164	0.5328	0.5005	0.5245	0.5315	0.5317	0.5297	0.5229
	30	0.7387	0.6794	0.7338	0.7345	0.7377	0.7372	0.7341	0.5225	0.5053	0.5118	0.5218	0.5213	0.5156	0.5110	0.5998	0.5686	0.5905	0.5983	0.5992	0.5958	0.5900
	40	0.7348	0.6762	0.7327	0.7320	0.7345	0.7322	0.7331	0.5967	0.5764	0.5874	0.5973	0.5937	0.5890	0.5871	0.6451	0.6097	0.6386	0.6447	0.6440	0.6403	0.6386
	50	0.7353	0.6757	0.7326	0.7335	0.7353	0.7314	0.7332	0.6516	0.6276	0.6394	0.6506	0.6490	0.6436	0.6389	0.6771	0.6380	0.6682	0.6750	0.6755	0.6710	0.6684

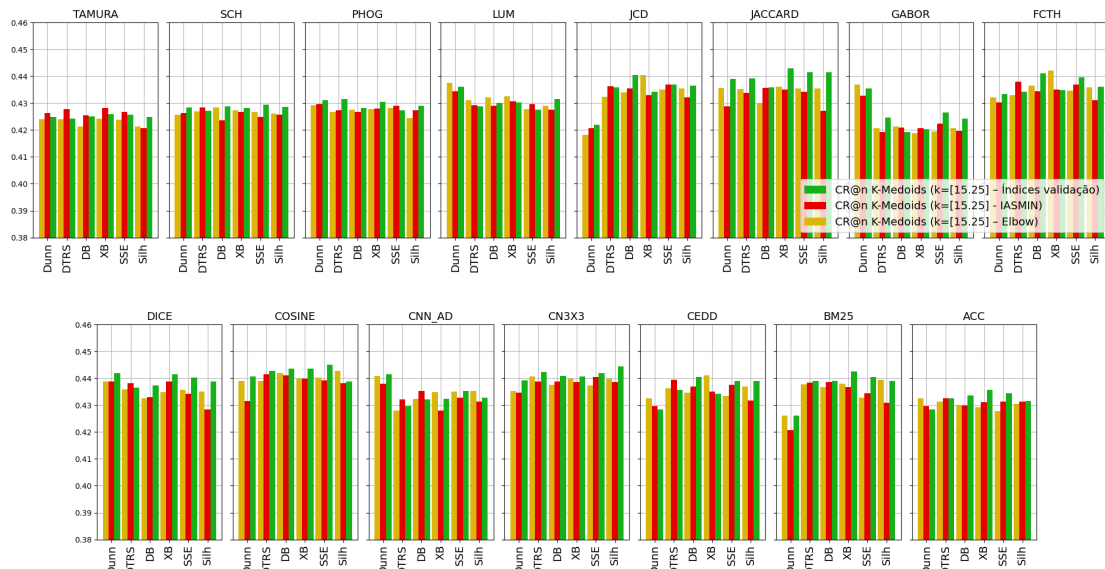
Tabela C.15: Resultados do experimento com K-Medoids, execuções para os três métodos de análise do cluster validity, $k = [15, 25]$, descritor ACC

	N	DTRS	Precision						DTRS	Cluster-Recall (CR)						DTRS	F1-Measure (F1)					
			K=20	Dunn	DB	XB	SSE	Silhouette		K=20	Dunn	DB	XB	SSE	Silhouette		K=20	Dunn	DB	XB	SSE	Silhouette
Elbow+Index	5	0.8126	0.7018	0.8189	0.8270	0.8117	0.8144	0.8216	0.1556	0.1293	0.1615	0.1579	0.1611	0.1586	0.1632	0.2563	0.2161	0.2636	0.2604	0.2620	0.2608	0.2653
	10	0.7914	0.6856	0.8041	0.8032	0.7896	0.8081	0.7991	0.2587	0.2297	0.2642	0.2604	0.2598	0.2657	0.2582	0.3805	0.3396	0.3894	0.3847	0.3818	0.3923	0.3817
	20	0.7827	0.6831	0.7831	0.7815	0.7732	0.7856	0.7845	0.4125	0.3782	0.4105	0.4091	0.4084	0.4066	0.4081	0.5294	0.4771	0.5279	0.5267	0.5244	0.5251	0.5270
	30	0.7754	0.6832	0.7748	0.7680	0.7662	0.7700	0.7724	0.5132	0.4713	0.5122	0.5161	0.5093	0.5077	0.5131	0.6059	0.5471	0.6048	0.6062	0.6006	0.6004	0.6056
	40	0.7653	0.6803	0.7667	0.7623	0.7565	0.7609	0.7648	0.5933	0.5468	0.5908	0.5892	0.5868	0.5823	0.5905	0.6557	0.5950	0.6549	0.6523	0.6481	0.6476	0.6535
	50	0.7596	0.6790	0.7599	0.7558	0.7545	0.7554	0.7565	0.6544	0.6131	0.6566	0.6484	0.6500	0.6456	0.6504	0.6888	0.6298	0.6906	0.6847	0.6847	0.6830	0.6854
apenas Index	5	0.8117	0.7018	0.8207	0.8117	0.8108	0.8279	0.8072	0.1587	0.1293	0.1564	0.1606	0.1595	0.1586	0.1583	0.2589	0.2161	0.2579	0.2631	0.2602	0.2614	0.2581
	10	0.7878	0.6856	0.7973	0.7959	0.7901	0.8032	0.7869	0.2613	0.2297	0.2591	0.2639	0.2661	0.2614	0.2590	0.3842	0.3396	0.3822	0.3868	0.3895	0.3862	0.3806
	20	0.7804	0.6831	0.7878	0.7867	0.7829	0.7858	0.7829	0.4153	0.3782	0.4111	0.4148	0.4213	0.4128	0.4124	0.5316	0.4771	0.5293	0.5325	0.5371	0.5316	0.5289
	20	0.7730	0.6832	0.7719	0.7749	0.7745	0.7736	0.7743	0.5154	0.4713	0.5103	0.5144	0.5180	0.5183	0.5139	0.6067	0.5471	0.6032	0.6069	0.6090	0.6093	0.6056
	40	0.7664	0.6803	0.7602	0.7600	0.7664	0.7653	0.7631	0.5934	0.5468	0.5855	0.5915	0.5961	0.5979	0.5937	0.6560	0.5950	0.6492	0.6552	0.6575	0.6591	0.6545
	50	0.7594	0.6790	0.7560	0.7606	0.7598	0.7610	0.7585	0.6511	0.6131	0.6486	0.6569	0.6533	0.6576	0.6522	0.6877	0.6298	0.6850	0.6910	0.6888	0.6924	0.6878
IASMIN+Index	5	0.8171	0.7018	0.8225	0.8117	0.8063	0.8099	0.8270	0.1606	0.1293	0.1582	0.1603	0.1598	0.1609	0.1612	0.2632	0.2161	0.2606	0.2623	0.2615	0.2631	0.2638
	10	0.7959	0.6856	0.8005	0.7946	0.7878	0.7910	0.8045	0.2624	0.2297	0.2580	0.2592	0.2579	0.2589	0.2611	0.3860	0.3396	0.3815	0.3821	0.3797	0.3813	0.3863
	20	0.7764	0.6831	0.7829	0.7791	0.7736	0.7757	0.7840	0.4180	0.3782	0.4097	0.4158	0.4133	0.4151	0.4128	0.5325	0.4771	0.5276	0.5312	0.5276	0.5296	0.5307
	30	0.7710	0.6832	0.7716	0.7739	0.7719	0.7718	0.7718	0.5160	0.4713	0.5138	0.5121	0.5137	0.5140	0.5141	0.6059	0.5471	0.6058	0.6036	0.6037	0.6041	0.6058
	40	0.7592	0.6803	0.7622	0.7631	0.7601	0.7597	0.7615	0.5876	0.5468	0.5895	0.5857	0.5884	0.5879	0.5883	0.6494	0.5950	0.6526	0.6493	0.6494	0.6495	0.6516
	50	0.7542	0.6790	0.7555	0.7570	0.7559	0.7548	0.7559	0.6503	0.6131	0.6493	0.6466	0.6531	0.6509	0.6504	0.6843	0.6298	0.6854	0.6836	0.6861	0.6850	0.6863

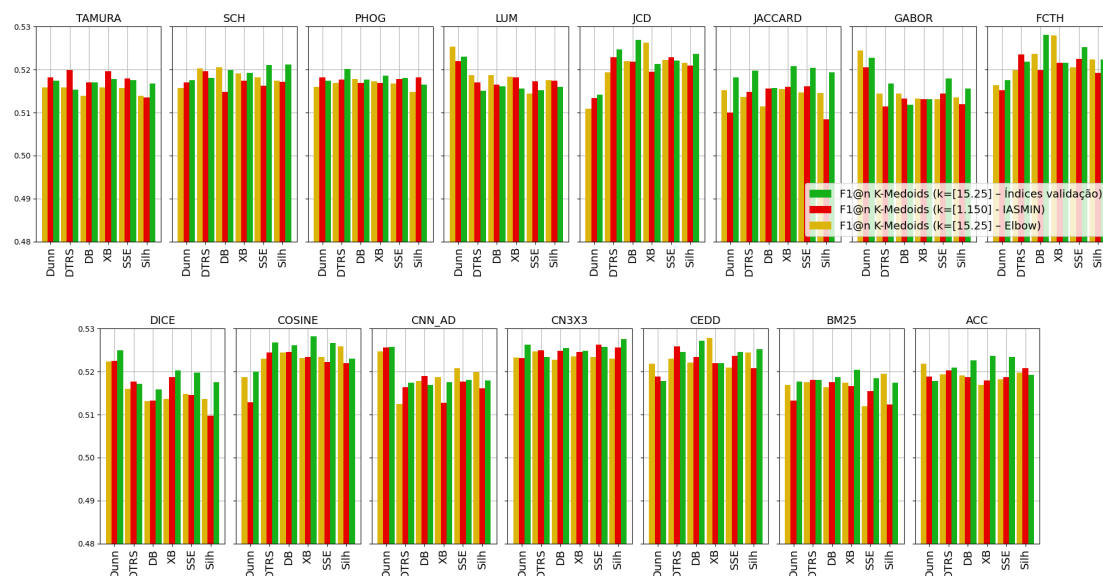
Figura C.1: Gráfico comparativo da performance dos experimentos adicionais. Algoritmo K-Medoids, $k = [15, 25]$ e as três abordagens para seleção do número de grupos ideal: apenas com uso dos índices de validação de agrupamento, em sua forma natural (critérios dos índices de *cluster validity*, baseados em valores mínimos ou máximos); com índices de validação e auxílio do método do Cotovelo; com índices de validação e auxílio do IASMIN. Valores médios das faixas $n = 5$ a $n = 50$ do ranking.



(a) Comportamento dos resultados métrica *Precision* ($P@n$)



(b) Comportamento dos resultados métrica *Cluster Recall* ($CR@n$)



(c) Comportamento dos resultados métrica $F1\text{-Measure}$ ($F1@n$)