

Submission date: dd/month/year  
Resubmission date: dd/month/year  
Camera ready submission date: dd/month/year

1<sup>st</sup> round notification: dd/month/year  
2<sup>nd</sup> round notification: dd/month/year  
Available online: dd/month/year  
Publication date: dd/month/year

Section: [editorial | regular article | thematic issue | survey | innovation case | application | extended version]  
[If thematic version, then add the title of the thematic issue]

## Métodos de Descoberta Adaptativa de Subconsultas para Busca Diversificada de Imagens

# Adaptive Subquery Discovery Methods for Diversified Image Search

Noberto Pires Maciel<sup>1</sup>, Rodrigo Tripodi Calumby<sup>2</sup>

<sup>1</sup>Department of Exact Sciences – State University of Feira de Santana (UEFS)  
Feira de Santana, Bahia – Basil

<sup>2</sup>Department of Exact Sciences – State University of Feira de Santana (UEFS)  
Feira de Santana, Bahia – Basil

nobertomacie1@hotmail.com, rtcalumby@uefs.br

**Abstract.** *Searching for images in large collections, like social media, is challenging due to irrelevant items and near-duplicates. An effective solution involves content diversification using data clustering algorithms. Each cluster represents a subtopic of the query, and the most representative images are selected. The effectiveness depends on choosing an appropriate clustering scheme and the number of clusters generated. This work evaluates the effectiveness of retrieving diversified images, exploring clustering schemes, and proposing the DTRS method to estimate cluster quality, while also reducing computational costs.*

**Keywords.** Clustering; Cbir; Dtrs; Diversity.

**Resumo.** *A busca de imagens por conteúdo em grandes coleções, como redes sociais, é desafiadora devido a itens irrelevantes e quase-duplicatas. Uma solução eficaz envolve a diversificação de conteúdo, utilizando algoritmos de agrupamento de dados. Cada grupo representa um subtópico da consulta, e as imagens mais representativas são selecionadas. A eficácia depende da escolha de um esquema de agrupamento adequado e do número de grupos gerados. Este trabalho avalia a eficácia na recuperação de imagens diversificadas, explorando esquemas de agrupamento e propondo o método DTRS para estimar a qualidade dos agrupamentos, além de reduzir custos computacionais.*

**Palavras-Chave.** Clustering; Cbir; Dtrs; Diversificação.

### 1. Introdução

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) e a *Content-based Image Retrieval* (CBIR). A primeira abordagem, a mais comum, é focada em texto conectado a imagens como forma de representá-las [Tyagi 2017], ou seja, é baseada em anotações textuais de descrição do conteúdo multimídia em um banco de dados (metadados). Essas anotações são obtidas de forma manual ou automática e podem ser de diferentes tipos: coordenadas de GPS, data e hora ou palavras-chave.

No entanto, descrever imagens de forma manual é uma tarefa complicada e cara quando se trata de conjuntos de dados muito extensos, o que pode tornar a TBIR uma abordagem, muitas vezes, inviável. Aliado a este fato, expressar o conteúdo visual das imagens em palavras não é uma tarefa fácil; anotações incorretas podem fazer com que a TBIR produza resultados irrelevantes ou mesmo inesperados [Tyagi 2017]. Além disso, usuários apresentam uma grande subjetividade na representação de suas necessidades ao especificar uma consulta e isso se aplica também em relação às anotações para a composição manual de metadados.

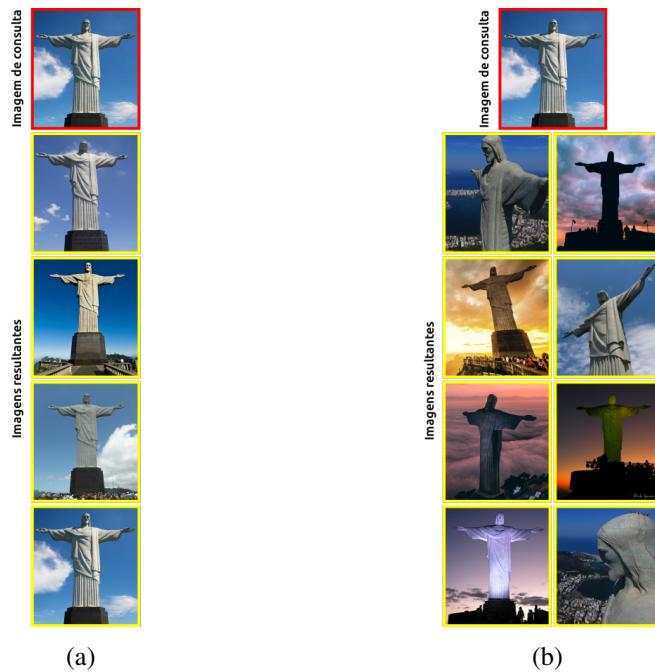
Assim, como uma tentativa de solucionar algumas das dificuldades encontradas na recuperação textual de imagens, surge a CBIR (*Content-based Image Retrieval*) [Kato 1992], também conhecida como *Query by image content* (QBIC), que é uma abordagem de recuperação baseada em características intrínsecas das imagens (*features*) e não apenas em metadados. Nela, a recuperação de imagens 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 ilustra uma consulta por similaridade em CBIR.



**Figura 1. Exemplo ilustrativo de consulta por similaridade apenas com resultados relevantes.**

Porém, consultas baseadas apenas em similaridade podem apresentar resultados pouco satisfatórios com itens não representativos ou mesmo duplicados [Calumby et al. 2017]. Esses podem ser definidos como elementos de retorno fortemente ligados ao elemento da consulta que não são, necessariamente, semanticamente iguais a ele. Este tipo de sistema desconsidera a existência de subgrupos semânticos no conjunto de imagens e tal fato acaba deixando itens pertencentes ao contexto da consulta fora dos resultados apresentados ao usuário. Uma alternativa, então, é a aplicação de técnicas de diversificação de conteúdo para que resultados mais satisfatórios sejam obtidos, considerando a existência de subtópicos implícitos na consulta e subgrupos semânticos no conjunto de imagens para oferecer 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 [Zhao and Glotin 2009], como pode ser visto na Figura 2(b).



**Figura 2. (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 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]. Muitos trabalhos têm utilizado algoritmos de agrupamento de dados para a tarefa de recuperação de imagens diversificadas como em [Figueiredo and Calumby 2022], [Yu et al. 2014] e [Van Leuken et al. 2009] de forma automatizada ou semi-automatizada.

Neste sentido, 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. Isso é, particularmente, interessante para sistemas CBIR, uma vez que a tarefa de recuperação de imagens diversificadas demanda ampla cobertura [Figueiredo and Calumby 2022] dos aspectos semânticos da consulta e das imagens do dataset. Assim, podemos relacionar os padrões de agrupamento nesses algoritmos aos subtópicos implícitos e às variações dos aspectos semânticos das imagens.

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 and Andayani 2018]. Contudo, a utilização desses algoritmos traz alguns problemas relacionados ao 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. Dentro desse contexto, o número  $k$  ideal de grupos (*clusters*) a serem gerados pelo algoritmo visando obter um esquema de agrupamentos ótimo tem sido um imenso desafio [Rodriguez et al. 2019]. Em relação ao sistemas CBIR, estudos indicam que os usuários se concentram na diversificação dos primeiros 20 elementos da lista retornada pela consulta [Zhao and Glotin 2009]. Desta forma, algumas abordagens têm adotado essa quantidade de grupos ( $k = 20$ ) 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.

Outra questão importante é a validação da qualidade dos esquemas de agrupamento apresentados por esses algoritmos, algo que 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 and 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 and Price 1763], o DTRS tenta 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.

Neste artigo, apresentaremos 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 [Figueiredo and 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. 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?

## 2. Metodologia Proposta

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 and 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.

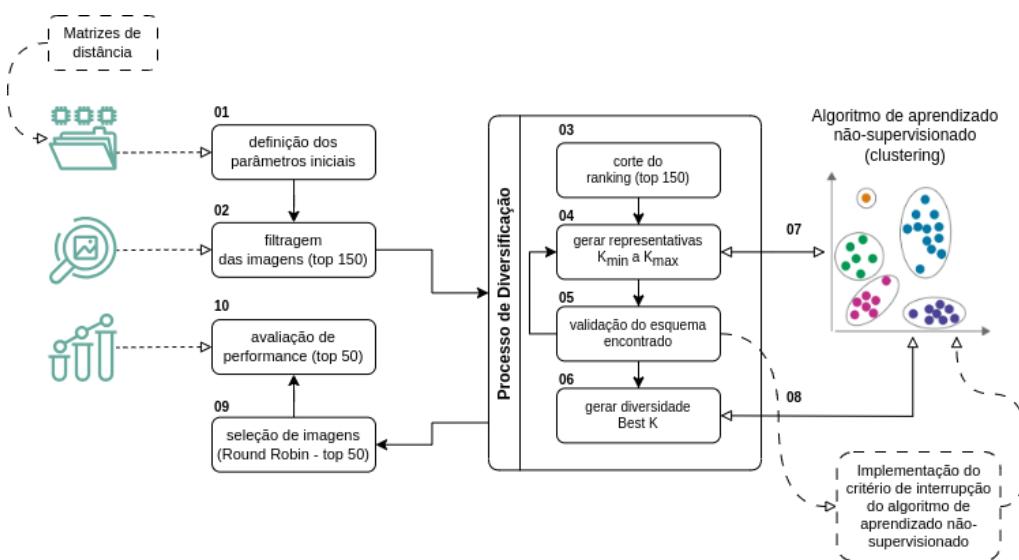
### 2.1. O arcabouço

O arcabouço, baseado no *framework* proposto por [Figueiredo and Calumby 2022], como pode ser visto na Figura 3, se inicia com a matriz de distância das características entre as imagens já extraída (*features*) e a definição dos parâmetros iniciais (etapa 1). Após essa parametrização inicial, é realizada uma filtragem de imagens (etapa 2), 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 (etapa 3), 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 2.4. Uma explicação detalhada sobre o DTRS pode ser vista nos artigos de [Yao et al. 1990] [Yao 2007] [Yu et al. 2014]. 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 (etapa 6).

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. Diagrama do arcabouço proposto, baseado no trabalho de [Figuerêdo and 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*.

## 2.2. 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.

### 2.3. Os experimentos

Para os experimentos, 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 inserção das matrizes de distância das imagens e 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) foi realizada com o propósito de reduzir o custo computacional da tarefa de recuperação. Para isso, utilizamos o descritor textual *Jaccard* como medida de determinação da relevância das imagens. As matrizes de distância foram extraídas dos descritores visuais ACC [Penatti et al. 2012], LUM [Lux and Chatzichristofis 2008], SCH [Cinque et al. 2001], GABOR [Manjunath and Ma 1996], TAMURA [Tamura et al. 1978], CN3x3, CNN\_AD [Ionescu et al. 2014], CEDD [Chatzichristofis and Boutalis 2008a], FCTH [Chatzichristofis and Boutalis 2008b], JCD [Zagoris et al. 2010] e PHOG [Biasotti et al. 2013]; textuais COSINE (Cosseno), BM25 [Baeza-Yates et al. 1999], DICE [Dice 1945] e JACCARD [Lewis et al. 2006]. Os índices de validação de agrupamentos foram o *Silhouette Score* [Rousseeuw 1987], *Dunn Index* [Dunn 1973], *Davies-Bouldin* [Davies and Bouldin 1979], *Xie Beni Index* [Xie and Beni 1991], *Sum of Squared Error (SSE)* e o DTRS [Yu et al. 2014].

Em relação ao processo de identificação dos padrões de relevância e diversificação nos conjuntos de imagens (seleção dos grupos semânticos), etapas 04 e 06 do diagrama da Figura 3, foram utilizados dois algoritmos de agrupamento diferentes (K-Medoids e Hierárquico Aglomerativo) em execuções distintas para cada experimento de acordo com os objetivos deste estudo. Assim, inicialmente (experimento 01), foi utilizado o algoritmo *K-Medoids*, com  $k$  variando entre [15, 25], em execuções individuais e independentes entre si, para cada um dos 15 descritores associados a um índice de validação (*cluster validity*) e tópico de consulta, objetivando a comparação da eficácia entre o DTRS no intervalo variado de  $k$ , uma execução com  $k$  fixo em 20 *clusters* (*baseline*) e os demais índices do estudo.

O segundo experimento foi realizado com o uso do algoritmo Hierárquico Aglomerativo (AHC) e número de *clusters* variando em  $k = [1, 150]$ . Aqui foi implementado o critério de interrupção (detalhado na Seção 2.4) para auxiliar o AHC na escolha do número de grupos ideal numa abordagem sem qualquer interferência humana. O objetivo, neste caso, é reduzir custo computacional do sistema CBIR e comparar o método

DTRS com o *baseline* e demais índices de validação, dentro da abordagem automatizada, em termos de eficácia e tempo de execução.

Um terceiro experimento, 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 (AHC 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.

Os resultados foram dispostos em tabelas e gráficos individualizados por descritor, 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 método do Cotovelo ou o critério de interrupção proposto. 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. Os resultados dos experimentos adicionais podem ser vistos no material [complementar](#).

## 2.4. 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), 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 validity*), individualmente, num processo dinâmico para a escolha do *best K*.

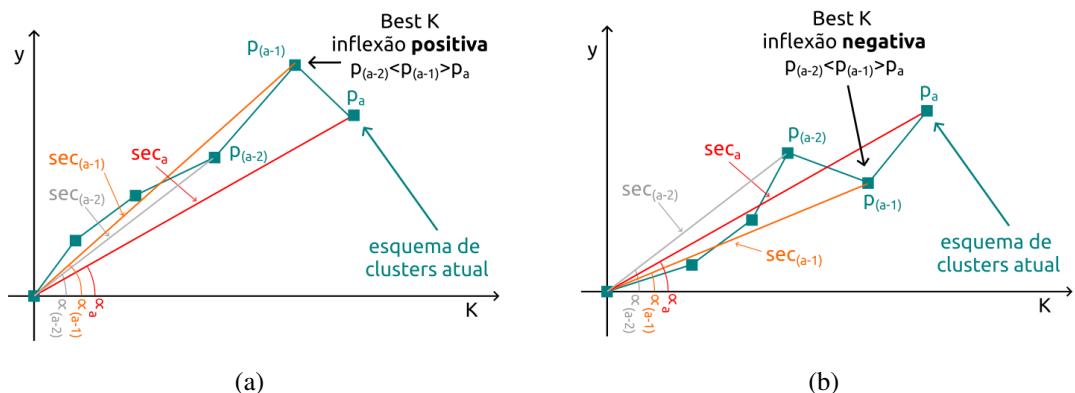
A adoção deste critério de interrupção teve como objetivo principal a redução do tempo de execução do sistema de recuperação e foi aplicado de duas formas, conforme pode ser visto na Figura 4, de acordo com o comportamento de cada índice de validação de agrupamento e suas particularidades:

- 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 4(a).

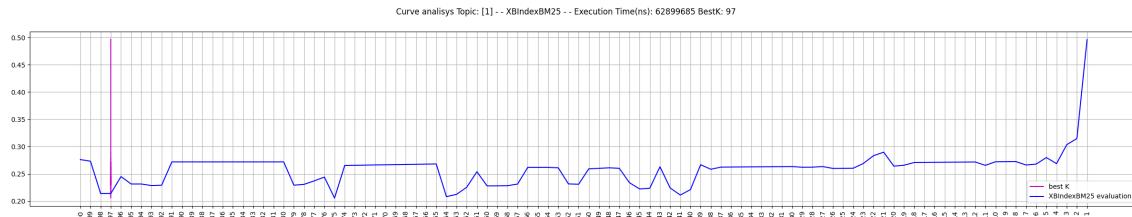


**Figura 4. Critério de interrupção do algoritmo hierárquico aglomerativo baseado no comportamento do coeficiente angular da curva da média móvel extraída dos valores gerados pelo *cluster validity*, pontos de inflexão positiva (Figura 4(a)) e negativa (Figura 4(b)) para os casos de curvas convexas e 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 4(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 5. A presença de *outliers* e as características do dataset, ou do descriptor, 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*.

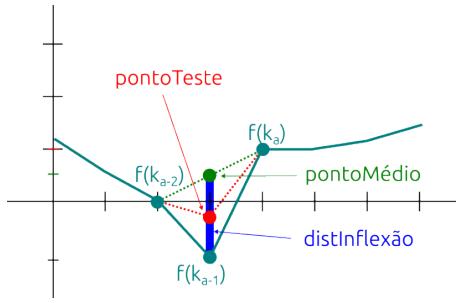


**Figura 5. Curva gerada pelos valores obtidos pelo índice de validação de agrupamento Xie Beni (número de grupos ideal pelo valor mínimo) para o primeiro tópico de consulta do experimento e  $k$  partindo de 100 *clusters*. Best  $k$  em 97.**

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

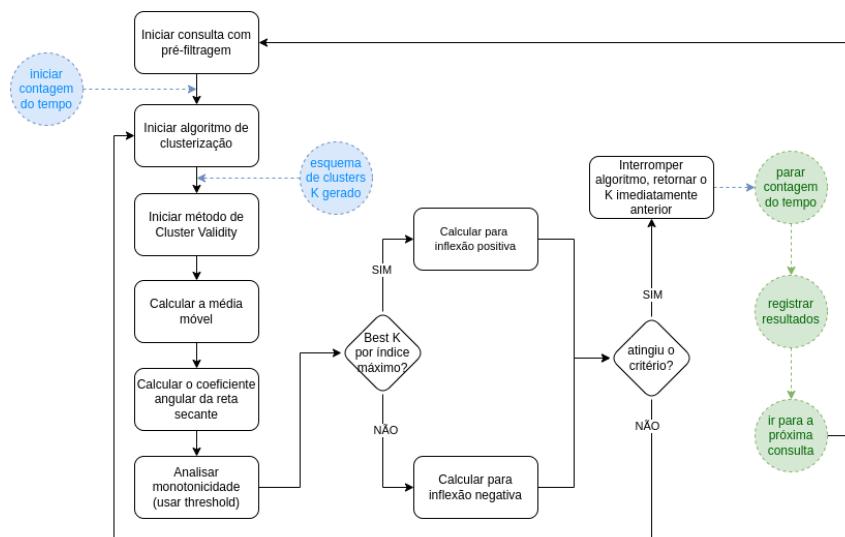
Caso não haja detecção do ponto de inflexão ao longo da curva do coeficiente angular com características suficientes para provocar a interrupção do algoritmo, uma última estratégia é adotada, baseada em resultados empíricos, apresentando o antepenúltimo esquema de agrupamentos ( $k_{a-2}$ ) como o mais hábil (número de grupos ideal). A Figura 6 ilustra o cálculo para um ponto cuja inflexão negativa foi suficientemente grande para gerar a interrupção.



**Figura 6. 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 7 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 (*Interrupção pela Análise da Reta Secante da Média Móvel do INdice*).



**Figura 7. Diagrama de fluxo da etapa de execução do critério de interrupção (IASMIN).**

## 2.5. 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. Resultados e discussão**

Nas próximas seções discutiremos os resultados alcançados pelos experimentos baseados nos conceitos e definições descritos na metodologia proposta. 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 primeira seção (Seção 3.1), será apresentada uma análise detalhada do desempenho do DTRS como índice de validação de agrupamentos, baseada nos resultados do experimento 01 (K-Medoids com  $k = [15, 25]$ ), com foco na eficácia do método na recuperação de imagens e nos ganhos obtidos em comparação com outros índices e o *baseline*. Na Seção 3.2, falaremos sobre a eficácia geral do sistema com uso do DTRS aliado ao IASMIN com base nos resultados dos experimentos 02 e 03 mostrando os ganhos alcançados da abordagem. Por fim (Seção 3.3), faremos a avaliação em relação ao tempo de execução do sistema CBIR, também baseada nos resultados dos experimentos 02 e 03.

#### **3.1. 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).

##### **3.1.1. Avaliação dos resultados de performance**

A Tabela 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 inicial (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

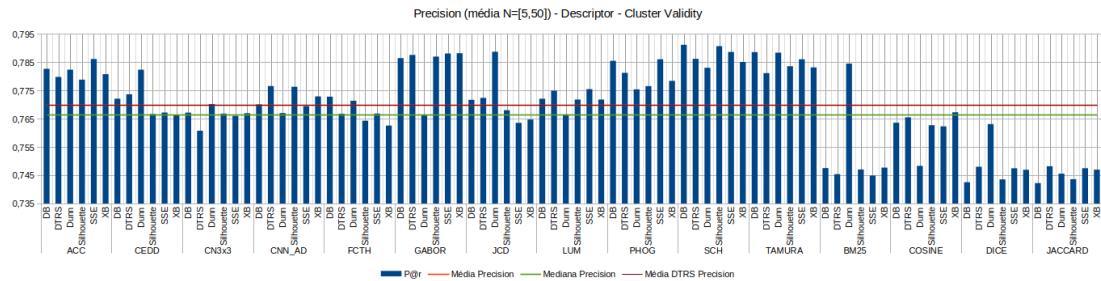
atingidos em relação à medida *Precision*, concentraram-se entre os descritores visuais ACC ( $P@5$ ), CEDD ( $P@40$  e  $P@50$ ), GABOR ( $P@5$  à  $P@30$ ), JCD ( $P@40$  e  $P@50$ ) e SCH ( $P@5$  à  $P@30$ ). Já entre os descritores textuais não houve destaque, em termos de valores máximos, para 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 centrada 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 ( $5 \leq CR@n \geq 50$ ), porém, com mais ocorrências entre os descritores textuais.

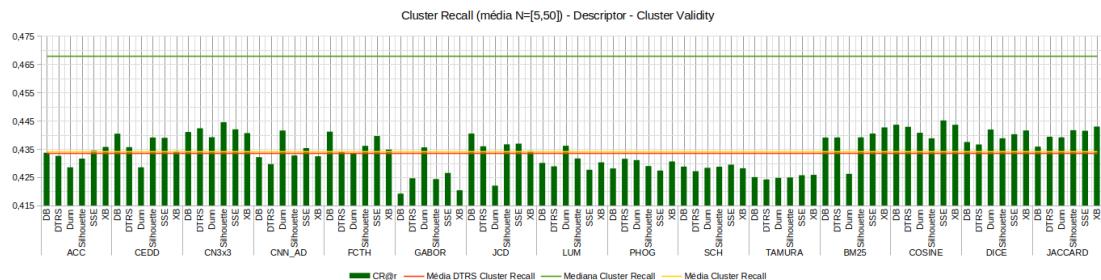
		PRECISION					CLUSTER RECALL					F1 MEASURE							
		P@5	P@10	P@20	P@30	P@40	P@50	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
visual	ACC	0.8117	0.7878	0.7804	0.7730	0.7664	0.7594	0.1587	0.2613	0.4153	0.5154	0.5934	0.6511	0.2589	0.3842	0.5316	0.6067	0.6560	0.6877
	CEDD	0.7991	0.7595	0.7635	0.7737	0.7751	<b>0.7709</b>	0.1569	0.2658	0.4252	0.5284	0.5925	0.6451	0.2574	0.3868	0.5365	0.6170	0.6596	<b>0.6899</b>
	CN3x3	0.7730	0.7680	0.7653	0.7571	0.7521	0.7490	0.1591	0.2676	<b>0.4351</b>	0.5292	<b>0.6044</b>	<b>0.6585</b>	0.2569	0.3871	<b>0.5428</b>	0.6102	0.6569	0.6865
	CNN_AE	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
	FCH	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	<b>0.8027</b>	<b>0.7905</b>	<b>0.7826</b>	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	<b>0.7755</b>	0.7704	0.1556	0.2671	0.4262	0.5281	0.5944	0.6443	0.2556	0.3882	0.5371	<b>0.6172</b>	<b>0.6613</b>	0.6891
	LUM	0.7982	0.7946	0.7757	0.7680	0.7586	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
textual	PHOG	0.8063	0.7910	0.7818	0.7773	0.7684	0.7627	0.1570	0.2628	0.4132	0.5124	0.5922	0.6501	0.2574	0.3874	0.5297	0.6042	0.6552	0.6869
	SCH	<b>0.8135</b>	0.8014	0.7876	0.7805	0.7712	0.7630	0.1577	0.2607	0.4081	0.5131	0.5818	0.6416	0.2590	0.3888	0.5261	0.6067	0.6500	0.6830
	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	<b>0.2607</b>	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	<b>0.5330</b>	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	<b>0.2803</b>	0.4301	0.5195	0.5899	0.6473	0.2491	<b>0.4034</b>	0.5333	0.5984	0.6432	0.6757
	JACCARD	0.7631	0.7572	0.7468	0.7407	0.7412	0.7401	<b>0.1603</b>	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	0.6858	0.2607	0.4034	0.5428	0.6172	<b>0.6613</b>	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	0.6399	0.2490	0.3771	0.5208	0.5984	0.6432	<b>0.6754</b>
		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 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.**

As Figuras 8 e 9 mostram a representação gráfica dos resultados médios de eficácia para as métricas *Precision* e *Cluster Recall*, separados por descritor e índice de validação. Esses valores de eficácia correspondem à média entre todas as faixas de ranking computadas ( $r = 5$  a  $r = 50$ ).



**Figura 8. 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).**



**Figura 9. 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*.**

Nas Tabelas 2 e 3 são apresentados os resultados destacados com a aplicação do teste de significância estatística não-paramétrico de *Wilcoxon* para os dois descriptores com melhores resultados. 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* implementado neste trabalho.

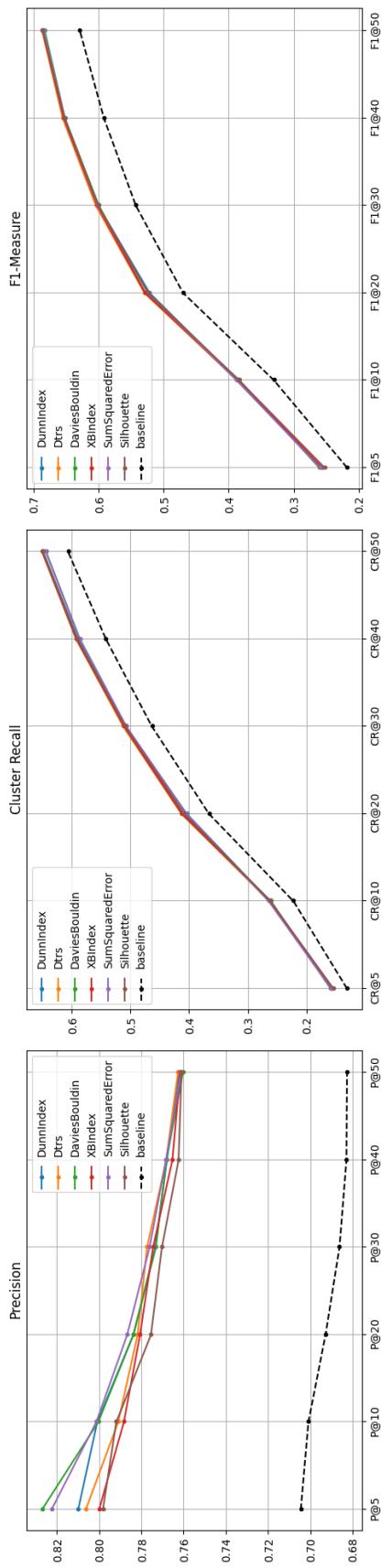
O teste estatístico foi aplicado individualmente nos  $N$  subrankings retornados pela consulta ( $n = \{5, 10, 20, 30, 40, 50\}$ ), para cada um dos 222 tópicos, adotando o DTRS como condição inicial (*before*) e o respectivo método de *cluster validity*, dentre os selecionados para o estudo, além do *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*.

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 10 e 11 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	DTRS	K=20	Dunn	Precision	DB	XB	SSE	Silhouette	Cluster-Recall (CR)			DTRS	K=20	Dunn	F1-Measure (F1)	XB	SSE	Silhouette
									DTRSS	DB	XB							
5	0.8063	<b>0.7045</b>	(0.8099	0.8270	<b>0.8000</b>	0.8225	<b>0.7983</b>	0.7982	0.1570	<b>0.1307</b>	<b>0.1564</b>	0.1588	<b>0.1564</b>	0.1578	<b>0.1539</b>	0.2574	<b>0.2180</b>	<b>0.2567</b>
10	0.7910	<b>0.7009</b>	(0.8009	0.8005	<b>0.7883</b>	0.8014	<b>0.7883</b>	0.7919	0.2628	0.2220	0.2621	0.2603	<b>0.2612</b>	0.2643	<b>0.2617</b>	0.3874	<b>0.3304</b>	<b>0.3867</b>
20	0.7818	<b>0.6928</b>	(0.7838	0.7838	<b>0.7805</b>	0.7838	<b>0.7805</b>	0.7767	0.4132	0.4040	0.4122	0.4111	<b>0.4081</b>	0.4046	<b>0.4081</b>	0.5297	<b>0.4701</b>	<b>0.5217</b>
30	0.7773	<b>0.6862</b>	(0.7734	0.7734	<b>0.7745</b>	0.7761	<b>0.7745</b>	0.7703	0.5124	0.4634	0.5088	0.5077	<b>0.5108</b>	0.5069	<b>0.5091</b>	0.6042	<b>0.5427</b>	<b>0.6005</b>
40	0.7682	<b>0.6829</b>	(0.7679	0.7679	<b>0.7653</b>	0.7684	<b>0.7653</b>	0.7624	0.5928	0.5418	0.5898	0.5867	<b>0.5923</b>	0.5862	<b>0.5911</b>	0.6552	<b>0.5918</b>	<b>0.6535</b>
50	0.7627	<b>0.6826</b>	(0.7622	0.7622	<b>0.7616</b>	0.7612	<b>0.7616</b>	0.7610	0.6509	0.6064	0.6489	0.6439	<b>0.6505</b>	0.6444	<b>0.6498</b>	0.6869	<b>0.6288</b>	<b>0.6860</b>

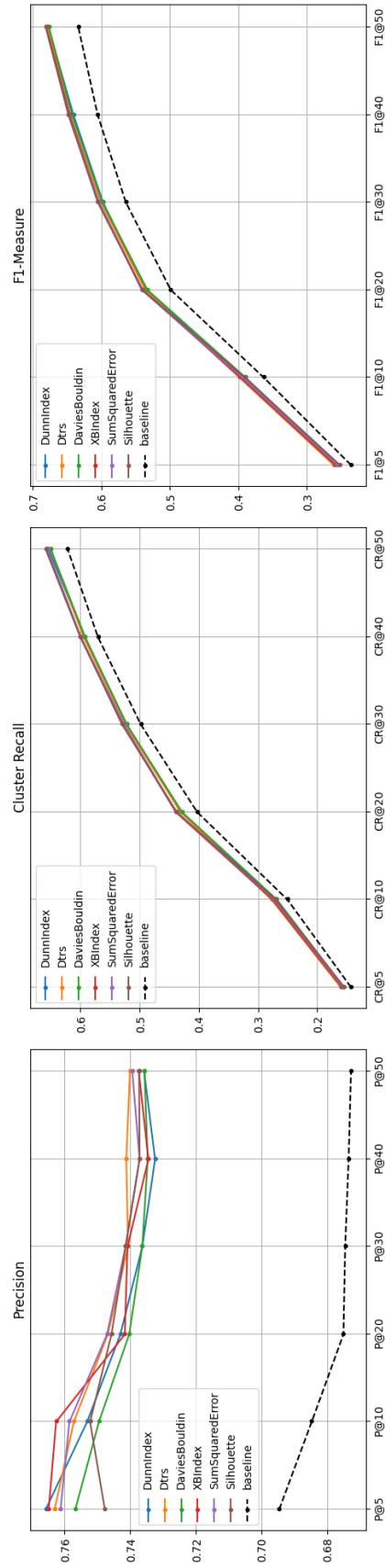
**Tabela 2. Experimento 01: resultados K-Medoids -  $k = [15, 25]$  - PHOG [Biasotti et al. 2013]**



**Figura 10. Performance: descriptor PHOG. Algoritmo K-Medoids,  $k = [15, 25]$ . M\'etodo natural do cluster validity. Destaque para o m\'etodo DTRS na faixa de ranking  $n = 30$  da m\'etria Precision.**

n	DTRS	K=20	Precision	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	Cluster-Recall (CR)	XB	SSE	Silhouette	DTRS	K=20	Dunn	F1-Measure (F1)	XB	SSE	Silhouette			
5	0.7631	<b>0.6946</b>	0.7658	<b>0.7568</b>	0.7649	0.7613	0.7477	0.1603	<b>0.1428</b>	0.1557	0.1584	0.1565	0.1542	0.2606	<b>0.2352</b>	<b>0.2564</b>	0.2583	0.2555	0.2515	0.3943	0.3969	0.3902		
10	0.7572	0.6847	0.7532	<b>0.7495</b>	0.7626	0.7586	0.7523	0.2770	<b>0.2502</b>	0.2716	0.2680	0.2753	0.2728	0.2702	0.3982	0.3937	<b>0.3894</b>	<b>0.3597</b>	<b>0.5326</b>	0.5349	0.5397	0.5410	0.5399	
20	0.7468	0.6750	0.7430	<b>0.7403</b>	0.7417	0.7411	0.7471	0.4306	<b>0.4026</b>	0.4305	0.4282	0.4388	0.4378	0.4372	0.4985	0.5349	0.5397	0.5341	0.5075	<b>0.5983</b>	0.6037	0.6033	0.6051	
30	0.7407	<b>0.6745</b>	0.7363	<b>0.7363</b>	0.7408	0.7414	0.7414	0.5238	<b>0.4979</b>	0.5218	0.5216	0.5279	0.5263	0.5295	0.6004	0.5641	0.6463	0.6429	0.6476	0.6480	0.6801	0.6806	0.6811	
40	0.7412	<b>0.6734</b>	0.7324	<b>0.7347</b>	0.7345	0.7372	0.7372	0.5947	<b>0.5704</b>	0.5928	0.5921	0.5999	0.5992	0.6000	0.6459	0.6053	0.6407	0.6429	0.6463	0.6773	0.6764	0.6801	0.6811	
50	0.7401	<b>0.6727</b>	(0.7358	<b>0.7357</b>	0.7374	0.7394	0.7373	0.6497	<b>0.6214</b>	0.6536	0.6501	0.6572	0.6560	0.6586	0.6775	0.6335	0.6773	0.6764	0.6801	0.6806	0.6811	0.6811	0.6811	0.6811

**Tabela 3. Experimento 01: resultados K-Medoids -  $k = [15, 25]$  - JACCARD [Lewis et al. 2006]**



**Figura 11. Performance: descriptor 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.**

### 3.1.2. Avaliação dos ganhos alcançados

Conforme evidenciado pelos gráficos nas Figuras 10 e 11, a abordagem utilizada neste estudo, detalhada na Seção 2.1, 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 (Figuras 8 e 9), 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 apresenta um resumo desses resultados por faixa de ranking e métrica de performance.

	Descriptor	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	$n = 40$	-	-

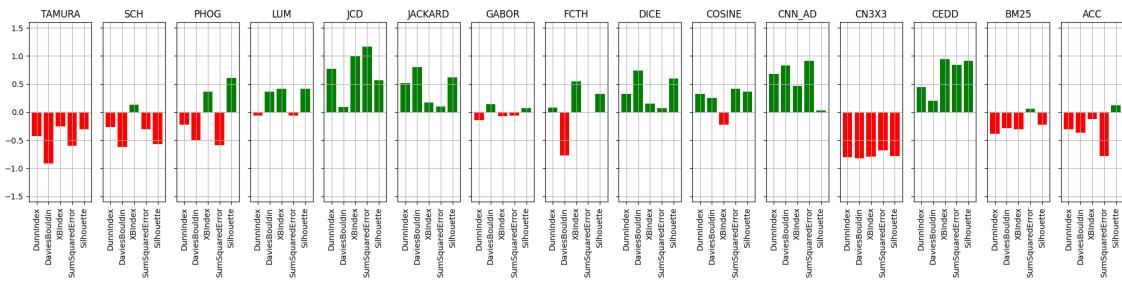
**Tabela 4. Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos dos resultados gerais do experimento 01.**

Já através da observação dos resultados destacados com a aplicação do teste de significância estatística (*Wilcoxon Signed Rank test*), Tabelas 2 e 3, bem como demais resultados não mostrados aqui por simplificação, 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 5 e 7 temos os resultados em ganho relativo para as métricas *Precision*, *Cluster Recall* e *F1-Measure*, respectivamente, assim como nas Figuras 12, 13 e 14 a representação gráfica desses resultados.

Descriptor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
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
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

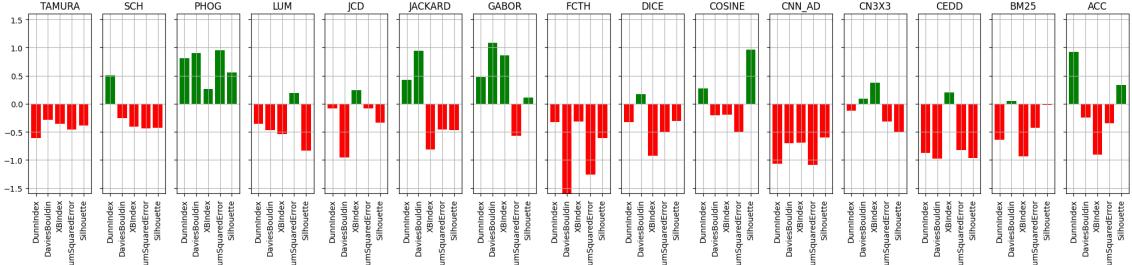
**Tabela 5.** Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids,  $k = [15, 25]$ . Superioridade do DTRS em azul. Métrica *Precision*.



**Figura 12.** Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica *Precision*. Em verde, superioridade do DTRS.

Descriptor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
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
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

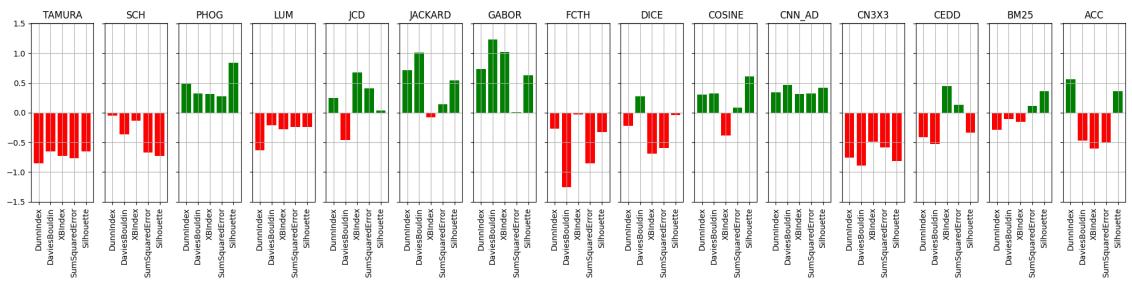
**Tabela 6.** Ganho médio relativo do DTRS versus demais métodos, algoritmo K-Medoids,  $k = [15, 25]$ . Superioridade do DTRS em azul. Métrica *Cluster Recall*.



**Figura 13.** Comportamento do método DTRS em relação aos demais métodos de validação, ganho médio relativo para a métrica *Cluster Recall*. Em verde, superioridade do DTRS.

Descriptor	Dunn Index	Davies Bouldin	Xie Beni	SSE	Silhouette
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
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 7.** Ganho médio relativo: DTRS versus demais métodos, algoritmo K-Medoids,  $k = [15, 25]$ . Superioridade do DTRS em azul. Métrica *F1-Measure*.



**Figura 14.** 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 8. Imagens reais deste experimento podem ser vistas na Figura 15.

Descriptor	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

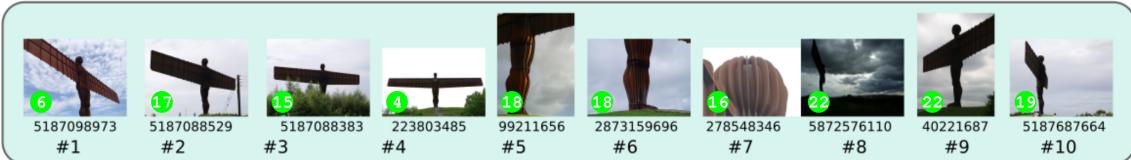
**Tabela 8. Descritores e índices de validação de agrupamento aos quais o DTRS obteve ganho consistente (ganhos positivos simultaneamente nas métricas *Precision* e *Cluster Recall*).**

**Baseline**



(a) Imagens reais: resultados do *baseline* na métrica *Precision* (query: angel\_of\_north, id: 1).

**Dtrs**



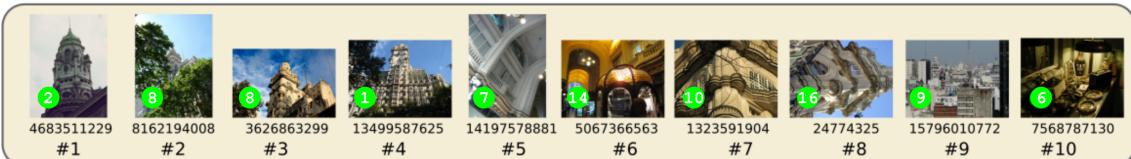
(b) Imagens reais: resultados do *DTRS* na métrica *Precision* (query: angel\_of\_north, id: 1).

**Baseline**



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

**Dtrs**



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

**Figura 15. Imagens reais do experimento 01 (K-Medoids,  $k = [15, 25]$ , usando apenas os índices de validação sem método auxiliar. Resultados para o descriptor 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 descriptor.**

### 3.2. Avaliação da abordagem dinâmica (experimentos 02 e 03)

Nesta seção, trataremos dos resultados alcançados com o uso do IASMIN, abordado na Seção 2.4, 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), 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.

No que diz respeito ao segundo experimento desta seção (experimento 03), 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.

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 dos experimentos adicionais e uma breve discussão podem ser vistos no material [complementar](#).

#### 3.2.1. Avaliação dos resultados de Performance

A partir da observação individualizada para o método DTRS e análise comparativa entre descritores visuais e textuais (Tabela 16(a)), com utilização do método auxiliar IASMIN, 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 16(b)) para a métrica *Precision* concentraram-se entre 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	<b>0.7667</b>	<b>0.7553</b>	0.1389	0.2485	0.3875	0.4931	0.5760	0.6405	0.2330	0.3728	0.5094	0.5906	0.6441	0.6789
	CN3x3	0.7604	0.7932	0.7809	0.7664	0.7548	0.7426	0.1424	0.258	<b>0.4073</b>	<b>0.5135</b>	<b>0.5956</b>	<b>0.6636</b>	0.2366	0.3818	<b>0.5254</b>	<b>0.6033</b>	<b>0.6521</b>	<b>0.6861</b>
	CNN_AD	0.7568	0.7914	0.7795	0.7665	0.7546	0.7441	0.1505	<b>0.2609</b>	0.4011	0.5052	0.5914	0.6556	0.2491	0.3855	0.5198	0.5963	0.6496	0.6828
	FCTH	0.7964	0.8104	0.7912	0.7743	0.7645	0.7528	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	<b>0.8149</b>	<b>0.7946</b>	<b>0.7758</b>	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	0.5228	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
textual	SCH	0.7694	0.7995	0.7824	0.7677	0.7557	0.7424	0.1505	0.2606	0.3983	0.5032	0.5903	0.6582	0.2496	<b>0.3858</b>	0.5185	0.5956	0.6485	0.6831
	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	0.1294	0.2449	0.4006	0.5079	0.5864	0.6545	0.2168	0.3641	0.5176	0.5982	0.6453	0.6810
	COSINE	0.8243	0.8117	0.7847	0.7704	0.7557	0.7439	<b>0.1562</b>	0.2551	0.3999	0.5079	0.5910	0.6588	<b>0.2559</b>	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	0.5986	0.6461	0.6837
	JACCARD	<b>0.8279</b>	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	<b>0.8149</b>	<b>0.7946</b>	<b>0.7758</b>	<b>0.7667</b>	<b>0.7553</b>	<b>0.1562</b>	<b>0.2609</b>	<b>0.4073</b>	<b>0.5135</b>	<b>0.5956</b>	<b>0.6636</b>	<b>0.2559</b>	<b>0.3858</b>	<b>0.5254</b>	<b>0.6033</b>	<b>0.6521</b>	<b>0.6861</b>
	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	0.7227	0.7287	0.7275	<b>0.7310</b>	0.1481	0.2584	0.4148	0.5298	0.6104	<b>0.6731</b>	0.2439	0.3754	0.5173	0.6006	0.6504	<b>0.6870</b>
	CEDD	0.7550	0.7366	<b>0.7381</b>	0.7296	0.7266	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_ADM	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	0.5207	0.5968	0.6479	0.6824
	FCTH	0.7559	<b>0.7432</b>	0.7372	<b>0.7336</b>	<b>0.7301</b>	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	<b>0.7586</b>	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
textual	PHOG	0.7387	0.7234	0.7311	0.7278	0.7264	0.7224	0.1514	0.2618	<b>0.4270</b>	<b>0.5345</b>	0.6075	0.6652	0.2490	0.3797	<b>0.5306</b>	<b>0.6069</b>	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
	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
textual	COSINE	0.7351	0.7185	0.7198	0.7218	0.7216	0.7214	0.1479	0.2576	0.4235	0.5272	0.6062	0.6696	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	0.6818
	JACCARD	0.7559	0.7396	0.7239	0.7248	0.7260	0.7253	<b>0.1541</b>	<b>0.2703</b>	0.4262	0.5344	<b>0.6124</b>	0.6668	<b>0.2535</b>	<b>0.3900</b>	0.5282	0.6051	<b>0.6527</b>	0.6819
textual	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	0.6089	0.6527	0.6870
	Min()	<b>0.7216</b>	<b>0.7185</b>	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%	0.6497	

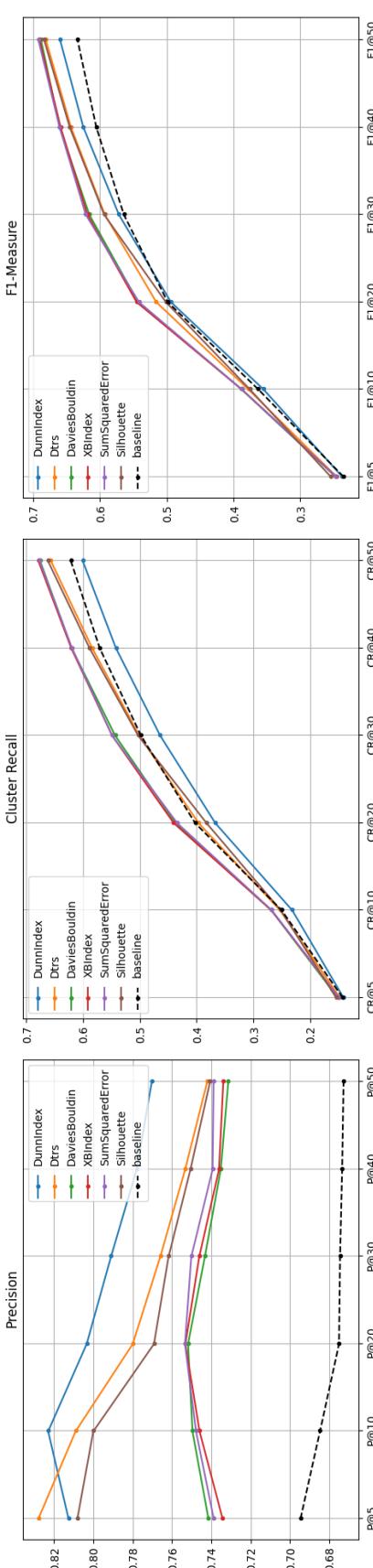
(b) Resultados do experimento 03 (K-Medoids) para o DTRS.

**Tabela 9. 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.**

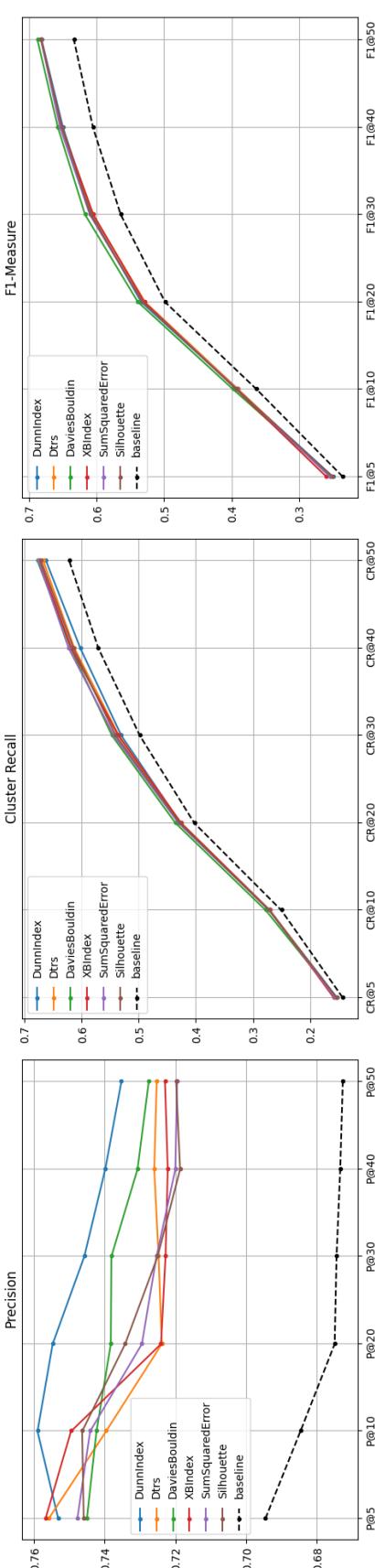
Na Tabela 10 temos os resultados das métricas *Precision*, *Cluster Recall* e *F1-Measure*, obtidas pelo experimento 02 para o descritor JACCARD, com uso do IASMIN aplicado ao algoritmo Hierárquico Aglomerativo e com a realização do teste de significância estatística de *Wilcoxon*. Aqui, a superioridade do método DTRS aparece destacada em azul, com a inferioridade em vermelho, assim como no experimento da seção anterior. As tabelas com os resultados dos 15 descritores do estudo podem ser vistos no material complementar.

n	DTRS	K=20	Precision	DB	XB	SSE	Silhouette	DTRS	K=20	Dunn	Cluster-Recall (CR)	XB	SSE	Silhouette	DTRS	K=20	Dunn	Cluster-Recall (CR)	XB	SSE	Silhouette	
5	0.8279	<b>0.6946</b>	0.8126	<b>0.7414</b>	<b>0.7342</b>	<b>0.7387</b>	<b>0.7495</b>	0.7459	0.7477	<b>0.1414</b>	<b>0.1515</b>	<b>0.1495</b>	<b>0.1493</b>	0.1544	0.2457	<b>0.2352</b>	<b>0.2372</b>	<b>0.2473</b>	<b>0.2465</b>	0.2545	0.2465	
10	0.8090	<b>0.6847</b>	0.8230	<b>0.7495</b>	<b>0.7459</b>	<b>0.7477</b>	<b>0.8080</b>	0.7530	0.2520	0.2316	<b>0.2684</b>	<b>0.2681</b>	<b>0.2676</b>	<b>0.2676</b>	0.2518	0.3777	<b>0.3628</b>	<b>0.3551</b>	<b>0.3879</b>	<b>0.3876</b>	0.3755	0.3755
20	0.7800	<b>0.6750</b>	0.8034	<b>0.7518</b>	<b>0.7532</b>	<b>0.7534</b>	<b>0.7691</b>	0.3966	0.4026	<b>0.3671</b>	<b>0.4373</b>	<b>0.4404</b>	<b>0.4343</b>	<b>0.4343</b>	0.3826	0.5156	<b>0.4985</b>	<b>0.4937</b>	<b>0.5429</b>	<b>0.5451</b>	0.5012	0.5012
30	0.7659	<b>0.6745</b>	0.7910	<b>0.7432</b>	<b>0.7461</b>	<b>0.7502</b>	<b>0.7617</b>	0.5008	0.4979	<b>0.4640</b>	<b>0.5432</b>	<b>0.5486</b>	<b>0.5487</b>	<b>0.5487</b>	0.5028	0.5226	<b>0.5641</b>	<b>0.5715</b>	<b>0.6159</b>	<b>0.6211</b>	0.5936	0.5936
40	0.7533	<b>0.6734</b>	0.7797	<b>0.7351</b>	<b>0.7360</b>	<b>0.7393</b>	<b>0.7506</b>	0.5836	0.5704	<b>0.5418</b>	<b>0.6198</b>	<b>0.6206</b>	<b>0.6193</b>	<b>0.6193</b>	0.5882	0.6433	<b>0.6053</b>	<b>0.6247</b>	<b>0.6601</b>	<b>0.6588</b>	0.6604	0.6604
50	0.7421	<b>0.6727</b>	0.7703	<b>0.7314</b>	<b>0.7339</b>	<b>0.7388</b>	<b>0.7406</b>	0.6536	0.6214	<b>0.6004</b>	<b>0.6754</b>	<b>0.6785</b>	<b>0.6764</b>	<b>0.6764</b>	0.6616	0.6809	<b>0.6335</b>	<b>0.6596</b>	<b>0.6882</b>	<b>0.6904</b>	0.6921	0.6835
5	0.7559	<b>0.6946</b>	0.7590	<b>0.7450</b>	<b>0.7495</b>	<b>0.7532</b>	<b>0.7568</b>	0.7477	0.7459	0.1541	<b>0.1428</b>	<b>0.1530</b>	<b>0.1541</b>	<b>0.1541</b>	0.1535	0.2355	<b>0.2352</b>	<b>0.2498</b>	<b>0.2536</b>	<b>0.2545</b>	0.2502	0.2502
10	0.7396	<b>0.6847</b>	0.7590	<b>0.7423</b>	<b>0.7495</b>	<b>0.7532</b>	<b>0.7568</b>	0.7441	0.7464	0.2703	<b>0.2502</b>	<b>0.2775</b>	<b>0.2704</b>	<b>0.2718</b>	0.3900	<b>0.3628</b>	<b>0.3916</b>	<b>0.3976</b>	<b>0.3930</b>	<b>0.3911</b>	0.3920	0.3920
20	0.7239	<b>0.6750</b>	0.7547	<b>0.7383</b>	<b>0.7241</b>	<b>0.7295</b>	<b>0.7342</b>	0.7254	0.7342	0.4262	<b>0.4026</b>	<b>0.4268</b>	<b>0.4256</b>	<b>0.4356</b>	0.4281	0.5282	<b>0.4985</b>	<b>0.5338</b>	<b>0.5387</b>	<b>0.5291</b>	0.5318	0.5318
30	0.7248	<b>0.6745</b>	0.7458	<b>0.7381</b>	<b>0.7288</b>	<b>0.7254</b>	<b>0.7351</b>	0.7254	0.7351	0.5344	<b>0.5465</b>	<b>0.5465</b>	<b>0.5465</b>	<b>0.5465</b>	0.5399	0.6051	<b>0.6047</b>	<b>0.6076</b>	<b>0.6099</b>	<b>0.6099</b>	0.6088	0.6088
40	0.7260	<b>0.6734</b>	0.7399	<b>0.7307</b>	<b>0.7222</b>	<b>0.7200</b>	<b>0.7187</b>	0.6124	0.5704	<b>0.5797</b>	<b>0.6019</b>	<b>0.6187</b>	<b>0.6176</b>	<b>0.6224</b>	0.6176	0.6497	<b>0.6513</b>	<b>0.6573</b>	<b>0.6590</b>	<b>0.6540</b>	0.6509	0.6509
50	0.7253	<b>0.6727</b>	0.7353	<b>0.7276</b>	<b>0.7230</b>	<b>0.7195</b>	<b>0.7198</b>	0.6668	0.6214	0.6620	<b>0.6769</b>	<b>0.6736</b>	<b>0.6751</b>	0.6708	0.6819	0.6335	<b>0.6814</b>	<b>0.6875</b>	<b>0.6818</b>	<b>0.6822</b>	0.6821	0.6821

**Tabela 10. Experimentos 02 (Aglomerativo) e 03 (K-Medoids): performance com  $k = [1, 150]$  - descriptor JACCARD**  
**[Lewis et al. 2006]**



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



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

**Figura 16. Performance: descriptor JACCARD. Experimentos 02 e 03,  $k = [1, 150]$  e uso do IASMIN.**

### 3.2.2. Ganhos alcançados - experimento 02 (algoritmo Hierárquico Aglomerativo - $k = [1, 150]$ )

Assim como no experimento inicial, a abordagem detalhada na seção 2.1 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 da representação gráfica da Figura 16(c) e tabelas de resultados do material complementar, 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 11.

	Descriptor	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		

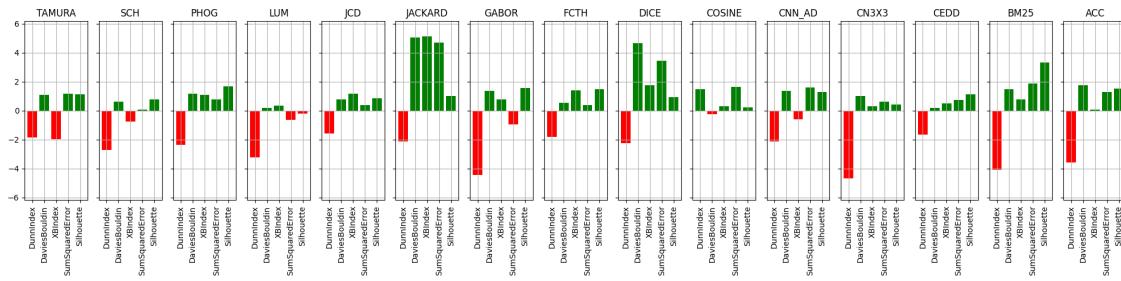
**Tabela 11. Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem considerar o teste de significância estatística. Baseado nas observações dos gráficos e tabelas de resultado do experimento 02 (material complementar).**

Os resultados com a aplicação do teste estatístico para o experimento 02, utilizando o algoritmo Hierárquico Aglomerativo, demonstraram superioridade estatística do DTRS nos casos específicos como pode ser visto na Tabela 12.

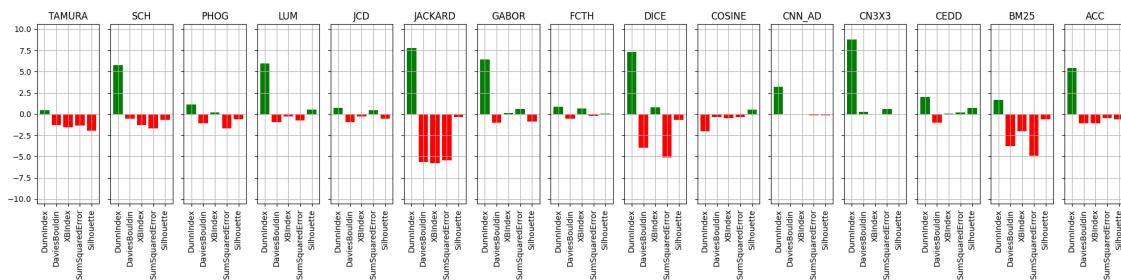
DESCRITOR	Precision	Cluster Recall	F1-Measure
SCH	-	Dunn Index	Dunn Index
PHOG	Davies Bouldin	-	-
LUM	-	Dunn Index	Dunn Index
JACCARD	Davies Bouldin, Xie Beni, SSE, Silhouette	Dunn Index	Dunn Index
FCTH	Xie Beni, Silhouette	-	-
DICE	Davies Bouldin, Xie Beni, SSE	Dunn Index	Dunn Index
COSINE	Dunn Index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

**Tabela 12.** Superioridade do método DTRS por descriptor

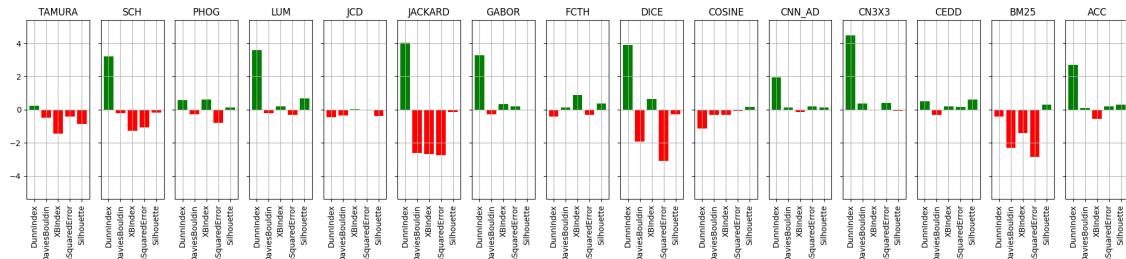
A Figura 17 mostra o ganho relativo para o método DTRS frente aos demais índices de validação de agrupamento nas métricas *Precision* e *Cluster Recall* em todos os descriptores do estudo.



(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 17. 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).**

### 3.2.3. Ganhos alcançados - 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 dos resultados de eficácia em valores máximos, para as faixas de ranking descritas na Tabela 11.

Descriptor	P@n	CR@n	F1@n
visual	LUM	20, 40	
	JCD	$\geq 30$	50
	GABOR	5	10
	FCTH	30	50
	CEDD	50	50
	ACC	$\geq 30$	50

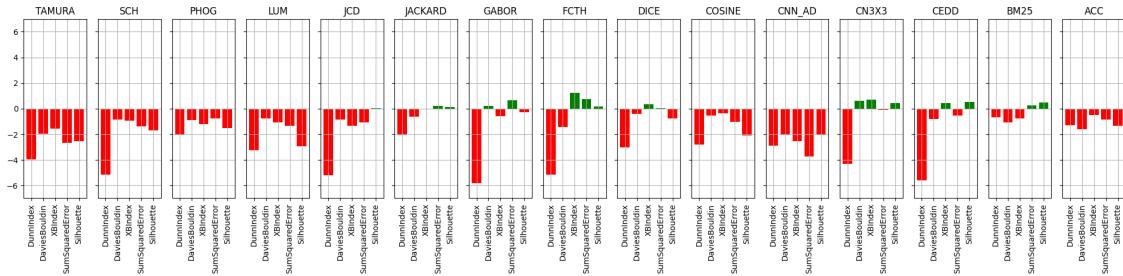
**Tabela 13. Descritores nos quais o método DTRS foi superior (valores brutos) a todos os outros métodos de validação. Resultados por faixa de ranking e sem aplicação do teste de significância estatística. Baseado nas observações dos resultados do experimento 03 (material complementar).**

Os resultados com a aplicação do teste estatístico do experimento 03 (algoritmo K-Medoids, usando o método auxiliar IASMIN) demonstraram superioridade estatística do DTRS nos casos específicos de acordo com a Tabela 14.

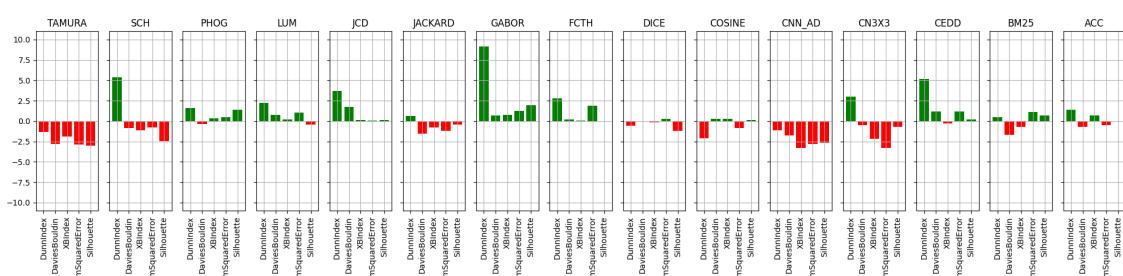
DESCRITOR	Precision	Cluster Recall	F1-Measure
PHOG	Dunn Index		
JCD	Dunn Index		
GABOR	Dunn Index	Dunn Index, SSE	
FCTH		SSE	
CN3X3	Dunn Index		
CEDD	Dunn Index	Dunn Index	

**Tabela 14. Superioridade do método DTRS por descritor**

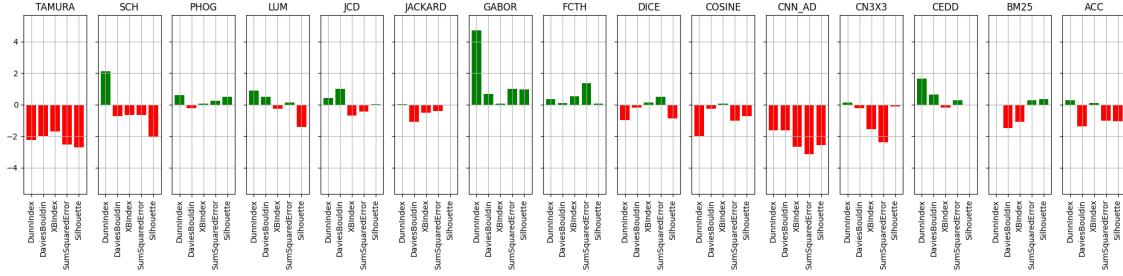
A Figura 18 mostra a representação gráfica do ganho médio relativo para o método DTRS frente aos demais índices de validação de agrupamento nas métricas *Precision* e *Cluster Recall* em todos os descritores do estudo.



(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 18. 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).**

### 3.3. Avaliação do tempo de execução

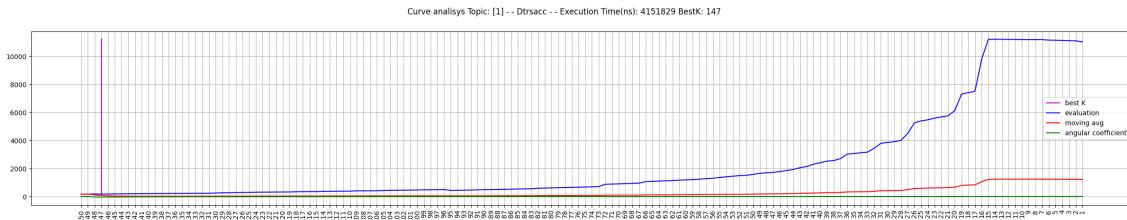
O objetivo principal desta etapa do estudo foi a redução do custo computacional associado ao processo de recuperação de imagens baseado em conteúdo. Normalmente, sistemas CBIR possuem um custo computacional elevado. A adoção de medidas para otimização desse custo, como a pré-filtragem ou o uso das matrizes de distância pré-calculadas, vistos no arcabouço apresentado no trabalho relacionado [Figueiredo and 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 2.4, 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  $t$ , 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 2.1.

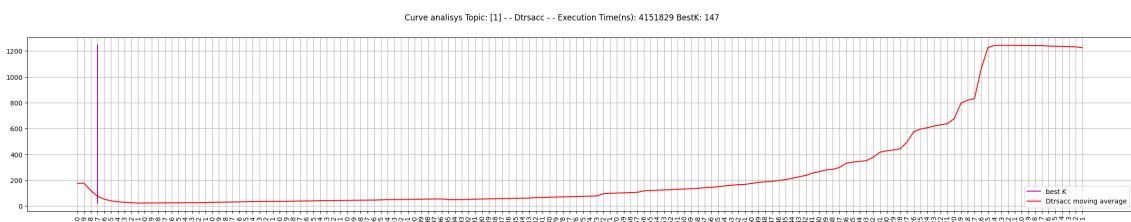
Nas Figuras 19 e 20 podemos ver o comportamento dos índices de validação do método DTRS para o descritor ACC em relação à *Query ID = 1 (angel\_of\_the\_north)*. No gráfico 19, temos a curva de risco gerada pelo método DTRS (azul) em conjunto com a média móvel (vermelho) e a curva do coeficiente angular da reta secante à média móvel (azul), curvas achataadas. O número de grupos ideal é representado pela reta vertical traçada no eixo das abscissas em  $k = 147$ .

A Figura 20(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. A curva gerada pelo coeficiente angular da reta secante aos pontos de teste da média móvel pode ser vista na Figura 20(b). Aqui podemos identificar o ponto de inflexão negativo no qual a interrupção do

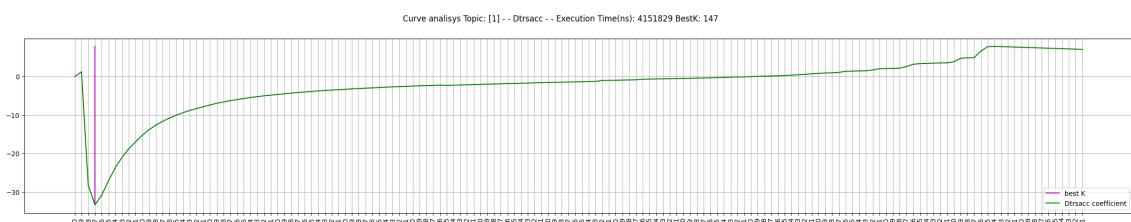
algoritmo foi alcançada. Neste ponto foram coletados o tempo de execução e o número  $k$  de grupos ideal anteriores ao ponto de inflexão. O tempo de execução ficou em 4151829 nanosegundos (0,04151829 segundos) para a amostra do tópico de consulta da Figura 20(b).



**Figura 19.** Experimento 02: curva gerada pelos resultados do método DTRS (curva de risco em azul) para o tópico de consulta *angel\_of\_the\_north*. Acha-tamento 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 logo nos primeiros esquemas de cluster fez com que o tempo de execução do sistema CBIR fosse o menor possível.

**Figura 20.** 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  $ID = 1$  (*angel\_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 2.4. 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:  $t_{total} = t_{k_1} + t_{k_2} + \dots + t_{bestK} + \dots + t_{k_f}$ , em nanosegundos, com  $k$  inicial, 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 15 e 16.

### 3.3.1. Tempo de execução - experimento 02

Assim, na Tabela 15 temos o tempo médio de execução, transformado para milisegundos, 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 15. 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.**

### 3.3.2. 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 16, 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 16. 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 16 e 15 (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.

#### 4. 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 (IASMIN), 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.

## 5. 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 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 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

- [Baeza-Yates et al. 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.
- [Bayes and Price 1763] Bayes, T. and Price, R. (1763). Philosophical transactions of the royal society of london. *Phil Trans R Soc*, 53:370–418.
- [Bezdek et al. 1997] Bezdek, J. C., Li, W., Attikiouzel, Y., and Windham, M. (1997). A geometric approach to cluster validity for normal mixtures. *Soft Computing*, 1:166–179.
- [Biasotti et al. 2013] Biasotti, S., Cerri, A., Giorgi, D., and Spagnuolo, M. (2013). Phog: Photometric and geometric functions for textured shape retrieval. In *Computer Graphics Forum*, volume 32, pages 13–22. Wiley Online Library.
- [Calumby et al. 2017] Calumby, R. T., Gonçalves, M. A., and da Silva Torres, R. (2017). Diversity-based interactive learning meets multimodality. *Neurocomputing*, 259:159–175. Multimodal Media Data Understanding and Analytics.
- [Chatzichristofis and Boutilis 2008a] Chatzichristofis, S. A. and Boutilis, Y. S. (2008a). Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In *International conference on computer vision systems*, pages 312–322. Springer.
- [Chatzichristofis and Boutilis 2008b] Chatzichristofis, S. A. and Boutilis, 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*, pages 191–196. IEEE.
- [Cinque et al. 2001] Cinque, L., Ciocca, G., Levialdi, S., Pellicano, A., and Schettini, R. (2001). Color-based image retrieval using spatial-chromatic histograms. *Image and Vision Computing*, 19(13):979–986.
- [Davies and Bouldin 1979] Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.
- [Dice 1945] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- [Dunn 1973] 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.
- [Figueiredo and Calumby 2022] Figueiredo, J. S. L. and Calumby, R. T. (2022). Unsupervised query-adaptive implicit subtopic discovery for diverse image retrieval based on intrinsic cluster quality. *Multimedia Tools and Applications*, pages 1–21.
- [Ionescu et al. 2014] Ionescu, B., Gînsca, A.-L., Boteanu, B., Popescu, A., Lupu, M., and Müller, H. (2014). Retrieving diverse social images at mediaeval 2014: Challenge, dataset and evaluation. *MediaEval*, 1263.

- [Kato 1992] Kato, T. (1992). Database architecture for content-based image retrieval. In Jamberdino, A. A. and Niblack, C. W., editors, *Image Storage and Retrieval Systems*, volume 1662, pages 112 – 123. International Society for Optics and Photonics, SPIE.
- [Kovács et al. 2005] Kovács, F., Legány, C., and Babos, A. (2005). Cluster validity measurement techniques. In *6th International symposium of hungarian researchers on computational intelligence*, volume 35.
- [Lewis et al. 2006] Lewis, J., Ossowski, S., Hicks, J., Errami, M., and Garner, H. R. (2006). Text similarity: an alternative way to search medline. *Bioinformatics*, 22(18):2298–2304.
- [Lux and Chatzichristofis 2008] Lux, M. and 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, page 1085–1088, New York, NY, USA. Association for Computing Machinery.
- [Manjunath and Ma 1996] Manjunath, B. S. and 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.
- [Müller et al. 2015] Müller, H., Ionescu, B., Gînsca, A. L., Boteanu, B., Popescu, A., and Lupu, M. (2015). Retrieving diverse social images at mediaeval 2015. In *Working Notes Proceedings of the MediaEval 2015 Workshop*. 14-15 September 2015.
- [Pawlak 1982] Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, 11:341–356.
- [Penatti et al. 2012] Penatti, O. A., Valle, E., and 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.
- [Rasyid and Andayani 2018] Rasyid, L. and 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, page 012082. IOP Publishing.
- [Rodriguez et al. 2019] Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. d. F., and Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1):1–34.
- [Rohm et al. 2018] Rohm, M., Ionescu, B., Gînsca, A. L., Santos, R. L. T., and 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, page 444–449, New York, NY, USA. Association for Computing Machinery.
- [Rojas-Thomas et al. 2017] Rojas-Thomas, J., Santos, M., and Mora, M. (2017). New internal index for clustering validation based on graphs. *Expert Systems with Applications*, 86:334–349.

- [Rousseeuw 1987] 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.
- [Shapiro and Wilk 1965] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- [Tamura et al. 1978] Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, man, and cybernetics*, 8(6):460–473.
- [Tyagi 2017] Tyagi, V. (2017). *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. Springer.
- [url ] url. Results from master’s thesis: Adaptive subquery discovery methods for diversified image search. <https://drive.google.com/drive/folders/1O7Oy7APXM1Es107Bc5dpg3NSocGbI7uw?usp=sharing>.
- [Van Leuken et al. 2009] Van Leuken, R. H., Garcia, L., Olivares, X., and van Zwol, R. (2009). Visual diversification of image search results. In *Proceedings of the 18th international conference on World wide web*, pages 341–350.
- [Woolson 2007] Woolson, R. F. (2007). Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3.
- [Xie and Beni 1991] Xie, X. L. and Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on pattern analysis and machine intelligence*, 13(8):841–847.
- [Yao 2007] Yao, Y. (2007). Decision-theoretic rough set models. In *Rough Sets and Knowledge Technology: Second International Conference, RSKT 2007, Toronto, Canada, May 14-16, 2007. Proceedings 2*, pages 1–12. Springer.
- [Yao et al. 1990] Yao, Y., Wong, S., and Lingras, P. (1990). A decision-theoretic rough set model. methodologies for intelligent systems 5, zw ras, m. zemankova and ml emrich.
- [Yu et al. 2014] Yu, H., Liu, Z., and 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 et al. 2010] Zagoris, K., Chatzichristofis, S. A., Papamarkos, N., and Boutalis, Y. S. (2010). Automatic image annotation and retrieval using the joint composite descriptor. In *2010 14th Panhellenic Conference on Informatics*, pages 143–147. IEEE.
- [Zhang et al. 2012] Zhang, S., Yang, M., Cour, T., Yu, K., and Metaxas, D. N. (2012). Query specific fusion for image retrieval. In *European conference on computer vision*, pages 660–673. Springer.
- [Zhao and Glotin 2009] Zhao, Z. and Glotin, H. (2009). Diversifying image retrieval with affinity-propagation clustering on visual manifolds. *IEEE MultiMedia*, 16(04):34–43.