

GCS-H: A Geo-Structural Constructive Heuristic for the Uncapacitated r -Allocation p -Hub Median Problem

Carlos V. Dantas Araújo¹, Jailon W. B. Oliveira da Silva¹
Pablo L. Braga Soares¹

¹Laboratório de Pesquisa & Desenvolvimento do NEMO
Universidade Federal do Ceará – Campus Russas
Avenida Felipe Santiago – Nº 411 62.900-624, Russas/CE, Brasil

{carlosvinicio, williambrunocc}@alu.ufc.br, pablo.soares@ufc.br

Abstract. *This paper introduces GCS-H, a single-pass constructive heuristic for the Uncapacitated r -Allocation p -Hub Median Problem (UrApHMP). The approach uses Multidimensional Scaling (MDS) and K-Means clustering for geographic candidate diversification and PageRank for selection based on structural importance, followed by a greedy construction phase. Validated on 231 instances from the AP benchmark, GCS-H outperforms the GRASP metaheuristic (mean cost reduction of 2.47%) and achieves a mean gap of 2.35% from state-of-the-art Scatter Search solutions. The heuristic replaces iterative searches with structural analysis, offering a scalable solution for hub location problems.*

Resumo. *Este trabalho apresenta a GCS-H, uma heurística construtiva de passagem única para o Problema p -Hub Mediana com r -Alocação (UrApHMP). A abordagem utiliza Escalonamento Multidimensional (MDS) e K-Means para diversificação geográfica de candidatos e PageRank para seleção baseada em importância estrutural, seguida por uma construção gulosa. Validada em 231 instâncias do benchmark AP, a GCS-H supera a meta-heurística GRASP (redução de custo média de 2,47%) e alcança um gap médio de 2,35% em relação às soluções do estado da arte (Scatter Search). A heurística substitui buscas iterativas por análise estrutural, oferecendo uma solução escalável para problemas de localização de hubs.*

1. Introdução

A otimização de redes logísticas e de transporte é um pilar para a eficiência operacional e competitividade econômica em inúmeros setores. Dentre os problemas de otimização combinatória que modelam esses desafios, o Problema de Localização de Hubs se destaca por sua complexidade e aplicabilidade estratégica. Uma variante de grande interesse prático é o Problema p -Hub Mediana com r -Alocação Não Capacitado, do inglês *Uncapacitated r -Allocation p -Hub Median Problem* (UrApHMP), que busca determinar a localização de p centros de distribuição (hubs) e as regras de alocação de n pontos de demanda a esses hubs, visando minimizar o custo total da rede [Corberán et al. 2018].

A característica central do UrApHMP, proposta por [Yaman 2011], é a flexibilidade da r -alocação, que permite que cada nó se conecte a até r hubs. Este modelo

generaliza as versões clássicas de alocação única ($r = 1$) e múltipla ($r = p$), oferecendo um equilíbrio mais realista entre o custo da infraestrutura e a robustez da rede. Contudo, essa flexibilidade adiciona camadas de complexidade a um problema já classificado como NP-difícil [O’Kelly 1987], tornando a busca por soluções ótimas computacionalmente intratável para instâncias de tamanho realista.

A literatura de referência para o UrApHMP, como o trabalho de [Corberán et al. 2018], tradicionalmente emprega meta-heurísticas sofisticadas, como GRASP e Scatter Search (SS), para obter soluções de alta qualidade. Tais métodos, embora eficazes, são inerentemente iterativos: eles constroem e refinam soluções através de milhares ou milhões de passos. Essa natureza iterativa, embora poderosa, apresenta duas limitações críticas em contextos práticos: (1) **Custo Computacional Elevado**, que pode ser proibitivo para instâncias de larga escala ou aplicações que demandam agilidade; e (2) **Sensibilidade a Parâmetros**, exigindo um ajuste cuidadoso de múltiplos parâmetros para garantir um bom desempenho.

Neste contexto, este trabalho explora uma via alternativa, fundamentada na aplicação de técnicas de aprendizado de máquina não supervisionado para guiar uma heurística construtiva. A hipótese central é que a estrutura topológica das matrizes de fluxo e distância contém informação latente suficiente para identificar um subconjunto de nós altamente promissores, eliminando a necessidade de uma busca iterativa extensiva. A contribuição deste trabalho é a Heurística por Seleção Geo-Clusterizada (GCS-H), um método de passagem única que se destaca pela velocidade e escalabilidade. A GCS-H integra análise geográfica, análise de redes e uma construção gulosa para gerar soluções de alta qualidade em segundos. Ao substituir a busca iterativa por uma análise estrutural direta, a abordagem proposta oferece uma alternativa eficiente.

2. Fundamentação Teórica

Esta seção estabelece os alicerces conceituais da abordagem proposta, detalhando em subseções distintas a formulação do problema de otimização, as métricas de teoria dos grafos e as técnicas de aprendizado de máquina utilizadas como base para o desenvolvimento da heurística.

2.1. O Problema p-Hub Mediana com r-Alocação (UrApHMP)

O Problema p-Hub Mediana com r-Alocação Não Capacitado (UrApHMP) é um problema de otimização combinatória definido sobre um conjunto de n nós, N . Para a modelagem, assume-se que o transporte é possível entre qualquer par de nós (i, j) , o que resulta em uma matriz de custos que representa um **grafo completo ponderado**, $G = (N, A)$, onde A é o conjunto de todas as arestas que conectam os pares de nós em N .

Para cada par de nós (i, j) , existe um fluxo de tráfego t_{ij} e um custo de transporte unitário d_{ij} . A matriz de distâncias D utilizada ao longo deste trabalho representa a matriz de caminhos mais curtos entre todos os pares de nós no grafo G . O objetivo central é selecionar um subconjunto de p nós para atuarem como centros de distribuição (*hubs*) e definir regras de alocação para os $n - p$ nós restantes, de forma a minimizar o custo total de transporte na rede.

A característica da r -alocação, proposta por [Yaman 2011], permite que cada nó não-hub seja conectado a até r hubs distintos ($1 \leq r \leq p$). O fluxo entre uma origem i e um destino j é roteado através de um par de hubs (k, l) , onde k pertence ao conjunto de hubs alocados a i (H_i) e l ao conjunto de hubs alocados a j (H_j). A função objetivo formal é:

$$\min \sum_{i \in N} \sum_{j \in N} t_{ij} \cdot C_{ij} \quad (1)$$

onde C_{ij} é o custo do caminho mais econômico para o fluxo entre i e j , dado por:

$$C_{ij} = \min_{k \in H_i, l \in H_j} (\chi d_{ik} + \alpha d_{kl} + \delta d_{lj}) \quad (2)$$

Nesta equação, os fatores de desconto χ , α e δ ponderam, respectivamente, os custos de coleta (origem ao hub), transferência (inter-hubs) e distribuição (hub ao destino). O UrApHMP é classificado como NP-difícil [Love et al. 1988], o que implica que a obtenção de soluções ótimas para instâncias de grande porte é computacionalmente intratável, motivando o uso de métodos heurísticos.

2.2. PageRank

O PageRank é um algoritmo de análise de links, originalmente desenvolvido por [Page et al. 1999], que atribui um escore de importância a cada nó em uma rede direcionada. Sua lógica é recursiva: um nó é considerado importante se recebe conexões de outros nós que, por sua vez, também são importantes. O algoritmo simula o comportamento de um “surfista aleatório” que navega pela rede, e a pontuação de um nó representa a probabilidade de o surfista se encontrar naquele nó. A fórmula é dada por:

$$PR(u) = \frac{1-d}{N} + d \sum_{v \in M(u)} \frac{PR(v)}{L(v)} \quad (3)$$

onde u e v são nós, $PR(u)$ é o escore de PageRank do nó u , $M(u)$ é o conjunto de nós que apontam para u , $L(v)$ é o número de links de saída do nó v , N é o número total de nós, e d é o fator de amortecimento (geralmente 0.85), que representa a probabilidade de o surfista seguir um link em vez de “saltar” para um nó aleatório.

2.3. Escalonamento Multidimensional (MDS)

O Escalonamento Multidimensional (MDS) é uma técnica de estatística multivariada utilizada para visualização e redução de dimensionalidade [Hastie et al. 2009]. Sua função é criar uma representação espacial de baixa dimensão para um conjunto de itens a partir de uma matriz que contém as dissimilaridades (distâncias) entre eles. O objetivo é que as distâncias na nova configuração reflitam da melhor forma possível as dissimilaridades originais. Para isso, o MDS posiciona os pontos x_1, \dots, x_n em um espaço de destino e os move iterativamente para minimizar uma função de perda, como a “Stress-1”:

$$\text{Stress}_D(x_1, \dots, x_n) = \left(\frac{\sum_{i < j} (d_{ij} - \|x_i - x_j\|)^2}{\sum_{i < j} d_{ij}^2} \right)^{1/2} \quad (4)$$

onde d_{ij} é a dissimilaridade original entre os itens i e j , x_i é o vetor de coordenadas do ponto i na nova configuração, $\|x_i - x_j\|$ é a distância euclidiana entre os pontos x_i e x_j , e o denominador atua como um fator de normalização.

2.4. Agrupamento K-Means (K-Means Clustering)

O K-Means é um algoritmo de aprendizado não supervisionado para problemas de agrupamento (*clustering*), proposto por [MacQueen 1967]. Seu objetivo é particionar um conjunto de n observações em K agrupamentos distintos, onde cada observação pertence ao agrupamento com a média (centroide) mais próxima. O algoritmo opera de forma iterativa através de duas etapas:

1. **Atribuição:** Cada ponto de dado é atribuído ao cluster cujo centroide é o mais próximo.
2. **Atualização:** O centroide de cada cluster é recalculado como a média de todos os pontos atribuídos a ele.

Este processo é repetido até a convergência, minimizando formalmente a soma dos quadrados intra-cluster, também conhecida como inércia:

$$J = \sum_{j=1}^K \sum_{x \in C_j} \|x - \mu_j\|^2 \quad (5)$$

onde J é a função objetivo (inércia), K é o número de clusters, C_j é o j -ésimo cluster, x é um ponto de dado pertencente a C_j , e μ_j é o vetor do centroide do cluster C_j .

3. Metodologia

Esta seção detalha o fluxo de trabalho proposto. Primeiramente, é descrita a Heurística por Seleção Geo-Clusterizada (GCS-H), o método central deste trabalho. Em seguida, são apresentados o conjunto de dados e os parâmetros utilizados para a validação experimental. Por fim, são fornecidos os detalhes de implementação e o ambiente computacional onde os experimentos foram conduzidos.

3.1. A Heurística GCS-H

A Heurística por Seleção Geo-Clusterizada (GCS-H) é uma abordagem projetada com foco em velocidade e análise estrutural, operando em uma única passagem sem o refinamento iterativo característico das meta-heurísticas. O método parte da premissa de que um conjunto de *hubs* eficaz deve equilibrar duas características, por vezes conflitantes: a **diversidade geográfica**, para uma cobertura eficiente da rede, e a **importância estrutural**, para o gerenciamento dos principais fluxos de tráfego. Para atender a esses requisitos, a GCS-H é composta por duas fases principais.

A primeira fase, de **Geração de Candidatos**, visa reduzir o espaço de busca, selecionando um subconjunto de nós promissores. Inicialmente, o Escalonamento Multidimensional (MDS) é aplicado sobre a matriz de distâncias para gerar um conjunto de coordenadas geográficas sintéticas (em \mathbb{R}^4) que preservam a topologia espacial da rede. Sobre estas coordenadas, o algoritmo K-Means particiona os n nós em k agrupamentos

geográficos, onde k é definido como $p \times 5$ para garantir uma supersegmentação deliberada (*oversampling*). Paralelamente, a matriz de fluxos é modelada como um grafo direcionado para que a centralidade PageRank de cada nó seja calculada. Ao final, eleger-se um “campeão regional” de cada um dos k agrupamentos: o nó com o maior escore de PageRank dentro de seu respectivo cluster é adicionado ao conjunto de candidatos C . Este processo resulta em um conjunto reduzido de candidatos que é, por construção, geograficamente diverso e estruturalmente relevante.

A segunda fase, de **Construção da Solução**, aplica uma seleção gulosa para a frente (*greedy forward selection*) sobre o conjunto de candidatos C . Partindo de um conjunto de *hubs* vazio ($H = \emptyset$), o algoritmo itera p vezes. Em cada iteração, avalia-se o custo total da rede ao adicionar temporariamente cada candidato $c \in C \setminus H$ à solução parcial. O candidato que resulta na maior redução marginal de custo é permanentemente adicionado ao conjunto H . Esta estratégia, ao operar sobre um conjunto de candidatos pré-filtrado, constrói de forma incremental a solução final.

3.2. Conjunto de Dados e Parâmetros de Validação

Para a validação da abordagem, foi utilizado o conjunto de instâncias AP (Australian Post), introduzido por [Ernst and Krishnamoorthy 1996]. Este conjunto de dados é baseado em fluxos de correio e localizações reais do serviço postal australiano, sendo amplamente utilizado como *benchmark* na literatura de problemas de localização de *hubs*. Os experimentos foram conduzidos em uma coleção de 231 instâncias geradas a partir dos dados originais, com o número de nós (n) variando de 60 a 200. Para cada tamanho de instância, diferentes valores de p (número de *hubs*) e r (número de alocações) são considerados. Conforme a prática padrão para este conjunto de dados, os parâmetros de custo foram fixados em $\chi = 3$, $\alpha = 0.75$ e $\delta = 2$ para todas as execuções.

3.3. Implementação e Ambiente Experimental

A lógica da GCS-H é formalizada no Algoritmo 1. A implementação foi desenvolvida em linguagem Python, utilizando a biblioteca *Scikit-learn* para os algoritmos de MDS e K-Means, *NetworkX* para a modelagem do grafo e cálculo do PageRank, e *NumPy* para operações numéricas.

Algorithm 1 GCS-H: Heurística por Seleção Geo-Clusterizada

Require: Matriz de Fluxos T , Matriz de Distâncias D , p , r , Parâmetros de Custo

Ensure: Conjunto de hubs H_{final}

```
1:
2: {Fase 1: Geração de Candidatos}
3:  $n \leftarrow \text{dimensão}(D)$ ;  $k \leftarrow p \times 5$ 
4:  $\text{coords} \leftarrow \text{MDS}(D, \text{n\_components} = 4)$ 
5:  $\text{geo\_clusters} \leftarrow \text{KMeans}(\text{coords}, \text{n\_clusters} = k)$ 
6:  $G \leftarrow \text{GrafoDeFluxo}(T)$ 
7:  $\text{scores} \leftarrow \text{CalcularPageRank}(G)$ 
8:  $C \leftarrow \emptyset$  {Conjunto de candidatos}
9: for  $i \leftarrow 0$  to  $k - 1$  do
10:    $\text{nós\_cluster}_i \leftarrow \{j \mid \text{geo\_clusters}[j] = i\}$ 
11:   if  $\text{nós\_cluster}_i$  não está vazio then
12:      $\text{campeão}_i \leftarrow \text{argmax}_{j \in \text{nós\_cluster}_i} (\text{scores}[j])$ 
13:      $C \leftarrow C \cup \{\text{campeão}_i\}$ 
14:   end if
15: end for
16:
17: {Fase 2: Construção da Solução}
18:  $H_{final} \leftarrow \emptyset$ 
19: for  $i \leftarrow 1$  to  $p$  do
20:    $c_{melhor} \leftarrow \text{null}$ ;  $\text{custo}_{min} \leftarrow \infty$ 
21:   for all  $c \in C \setminus H_{final}$  do
22:      $H_{temp} \leftarrow H_{final} \cup \{c\}$ 
23:      $\text{custo}_{atual} \leftarrow \text{CalcularCustoObjetivo}(H_{temp}, D, T, r, \text{Parâmetros de Custo})$ 
24:     if  $\text{custo}_{atual} < \text{custo}_{min}$  then
25:        $\text{custo}_{min} \leftarrow \text{custo}_{atual}$ ;  $c_{melhor} \leftarrow c$ 
26:     end if
27:   end for
28:    $H_{final} \leftarrow H_{final} \cup \{c_{melhor}\}$ 
29: end for
30:
31: return  $H_{final}$ 
```

4. Resultados e Discussão

A avaliação experimental da Heurística por Seleção Geo-Clusterizada (GCS-H) foi conduzida no conjunto completo de 231 instâncias da família AP. O desempenho do método foi comparado com duas meta-heurísticas de referência da literatura para este problema: GRASP e Scatter Search (SS). Os valores de *benchmark* para estes métodos foram adquiridos a partir da replicação dos métodos propostos no trabalho de [Corberán et al. 2018], que detalha os melhores resultados conhecidos, do inglês *Best Known Values* (BKV) para estas instâncias. A análise a seguir foca na qualidade da solução (custo) e na eficiência computacional (tempo de execução).

4.1. Análise Comparativa com GRASP

A primeira análise experimental compara o desempenho da GCS-H com a meta-heurística GRASP, um método de referência na literatura para o UrApHMP. A avaliação foca na qualidade da solução (custo total) e na eficiência computacional. A coluna “Melhora (%)” nas tabelas subsequentes quantifica a redução de custo obtida pela GCS-H em relação ao GRASP. Um valor positivo indica um custo menor para a GCS-H.

A análise sobre o conjunto completo de 231 instâncias indica que a GCS-H obteve soluções com custo inferior em todos os cenários, com um tempo de execução ordens de magnitude menor. A melhoria média no custo da solução foi de 2.47%. Este resultado sugere que a fase de análise geo-estrutural da GCS-H identifica configurações de hubs de alta qualidade de forma mais eficaz que a abordagem semi-aleatória do GRASP, com um custo computacional significativamente reduzido.

A Tabela 1 apresenta os resultados para as instâncias de pequeno porte ($n \leq 75$). Observa-se que a GCS-H obteve soluções com custo inferior ao GRASP em todas as configurações apresentadas.

Tabela 1. GCS-H vs. GRASP: Instâncias de Pequeno Porte.

Inst.	p	r	GCS-H	GRASP	Melhora (%)
AP60	5	4	129.807.35	133.649.33	2.87
AP60	6	5	121.979.40	126.073.75	3.25
AP60	8	7	109.651.38	112.789.93	2.78
AP65	5	4	130.094.52	135.272.11	3.83
AP65	6	5	122.962.93	126.278.38	2.62
AP65	8	7	111.324.46	113.779.73	2.16
AP70	5	4	132.055.96	133.409.89	1.01
AP70	8	7	113.315.79	116.647.14	2.86
AP75	6	5	125.184.65	128.987.03	2.95
AP75	8	7	114.086.67	115.968.17	1.62

Como observado, a Tabela 1 mostra uma melhoria de custo média de 2.60% para este subconjunto. Em casos específicos, como na instância AP65 com $p = 5$ e $r = 4$, a redução de custo atinge 3,83%. Este desempenho indica que a estratégia de seleção de candidatos baseada em *PageRank* e clusterização geográfica tem sucesso na captura de características relevantes da estrutura do problema. A seguir, a Tabela 2 detalha os resultados para instâncias de médio porte, permitindo avaliar a manutenção do desempenho com o aumento da complexidade.

Tabela 2. GCS-H vs. GRASP: Instâncias de Médio Porte.

Inst.	p	r	GCS-H	GRASP	Melhora (%)
AP80	5	4	132.424.08	134.172.20	1.30
AP80	8	7	114.404.95	115.350.95	0.82
AP85	6	5	125.915.82	127.495.60	1.24
AP85	8	7	114.966.55	116.349.03	1.19
AP90	5	4	133.081.65	133.552.36	0.35
AP90	8	7	113.803.56	116.465.61	2.28
AP95	6	4	125.386.26	129.627.98	3.27
AP95	8	7	113.732.39	116.812.27	2.64

Para as instâncias de médio porte ($80 \leq n \leq 95$), a GCS-H mantém a superioridade em relação ao GRASP, com uma melhoria média de custo de 1.64%. Mesmo nos cenários com ganhos mais modestos, como para a instância AP90 ($p = 5$), a GCS-H não produz soluções de custo superior, e a diferença no tempo de execução permanece significativa. A avaliação da escalabilidade e eficácia da abordagem é estendida às instâncias de grande porte, detalhadas na Tabela 3.

Tabela 3. GCS-H vs. GRASP: Instâncias de Grande Porte.

Inst.	p	r	Custo GCS-H	Custo GRASP	Melhora (%)
AP100	5	4	133.494.74	136.484.19	2.19
AP100	8	7	114.296.13	117.488.46	2.72
AP150	5	4	134.102.43	138.798.99	3.38
AP150	8	7	117.709.98	120.116.12	2.00
AP200	5	4	136.146.23	139.663.15	2.52
AP200	8	7	118.152.46	119.636.32	1.24

Os resultados para as instâncias de grande porte ($n \geq 100$), apresentados na Tabela 3, confirmam a tendência observada. A melhoria de custo neste conjunto é consistente, com uma média de 2.34% e atingindo 3,38% para a instância AP150. O desempenho em problemas com um espaço de busca maior corrobora a hipótese de que a pré-seleção de candidatos por análise geo-estrutural é uma estratégia eficaz. A obtenção de soluções de menor custo com tempos de execução inferiores indica o potencial da GCS-H para a otimização de redes em larga escala.

4.2. Análise Comparativa com Scatter Search (SS)

Para avaliar a competitividade da GCS-H, suas soluções são comparadas com os melhores valores conhecidos (BKV) da literatura, obtidos pela meta-heurística *Scatter Search* (SS). A análise utiliza o *gap* percentual, que mede a distância da solução da GCS-H em relação ao BKV, onde um *gap* menor indica maior proximidade da melhor solução conhecida.

Sobre o conjunto completo de 231 instâncias, o *gap* médio observado foi de 2.35%. Este resultado indica que a heurística construtiva, que opera em uma única pas-

sagem, gera soluções com pequena diferença de custo em relação a métodos iterativos complexos. A relação entre a qualidade da solução e a eficiência computacional representa um dos principais atributos da abordagem proposta.

A Tabela 4 apresenta a comparação para as instâncias de pequeno porte, onde se observa a proximidade das soluções da GCS-H com os BKV.

Tabela 4. Competitividade (GCS-H vs. SS): Instâncias de Pequeno Porte.

Inst.	p	r	Custo GCS-H	BKV (SS)	GAP (%)
AP60	4	3	145.473.55	142.268.41	2.25
AP60	6	5	125.179.80	121.979.40	2.62
AP60	8	7	112.455.11	109.651.38	2.56
AP65	5	4	133.294.40	130.094.52	2.46
AP65	8	7	113.665.20	111.231.25	2.19
AP70	5	4	134.109.76	132.055.96	1.56
AP70	8	7	114.987.10	112.082.12	2.59
AP75	6	5	128.214.50	125.184.65	2.42
AP75	8	7	115.968.17	114.086.67	1.65

Para as instâncias de pequeno porte, o *gap* médio foi de 2.25%. Os valores variam entre 1.56% (AP70, $p=5$) e 2.62% (AP60, $p=6$), indicando um desempenho consistente da heurística. A análise prossegue com as instâncias de médio porte na Tabela 5, para verificar se esta competitividade se mantém com o aumento da dimensão do problema.

Tabela 5. Competitividade (GCS-H vs. SS): Instâncias de Médio Porte.

Inst.	p	r	Custo GCS-H	BKV (SS)	GAP (%)
AP80	6	5	128.450.15	125.332.81	2.49
AP80	8	7	115.950.25	114.404.95	1.35
AP85	5	4	136.193.30	133.081.65	2.34
AP85	8	7	116.849.11	114.966.55	1.64
AP90	5	4	135.881.95	133.081.65	2.10
AP90	8	7	116.465.61	113.803.56	2.34
AP95	6	4	128.686.16	125.386.26	2.63
AP95	8	7	116.312.20	113.732.39	2.27

Conforme a Tabela 5, a GCS-H mantém sua competitividade para as instâncias de médio porte. O *gap* médio para este subconjunto é de 2.14%, com o menor valor sendo 1.35% para a instância AP80 ($p = 8$). A capacidade de gerar soluções próximas ao BKV em cenários de maior complexidade é um indicador positivo. A avaliação final da escalabilidade da abordagem é realizada com as instâncias de grande porte, apresentadas na Tabela 6.

Tabela 6. Competitividade (GCS-H vs. SS): Instâncias de Grande Porte.

Inst.	p	r	Custo GCS-H	BKV (SS)	GAP (%)
AP100	5	4	136.484.19	133.494.74	2.24
AP100	8	7	117.188.46	114.296.13	2.53
AP150	5	4	137.450.20	134.102.43	2.50
AP150	8	7	120.116.12	117.709.98	2.04
AP200	5	4	139.263.15	136.146.23	2.29
AP200	8	7	119.636.32	118.152.46	1.26

Os resultados para as instâncias de grande porte ($n \geq 100$) reforçam os achados anteriores. O *gap* médio neste conjunto foi de 2.15%. Mesmo para as maiores instâncias (AP200), o *gap* permanece baixo, como o valor de 1.26% para a configuração com $p = 8$ e $r = 7$. Estes dados sugerem que a GCS-H é uma heurística eficaz, que oferece um balanço favorável entre a qualidade da solução e o tempo de execução. A capacidade de produzir soluções com um desvio médio de 2.35% do BKV em segundos posiciona a GCS-H como uma alternativa viável para aplicações que requerem agilidade.

4.3. Análise de Escalabilidade e Impacto dos Parâmetros

A Figura 1 apresenta duas análises complementares sobre o desempenho da GCS-H. O gráfico à esquerda ilustra a escalabilidade computacional da heurística em função do tamanho da instância (n). O tempo de execução, apresentado em escala logarítmica, exibe um crescimento polinomial controlado. Embora o cálculo inicial da matriz de distâncias tenha complexidade teórica de $O(n^3)$, o tempo total de execução na prática se mostra altamente competitivo, como demonstrado na comparação a seguir.

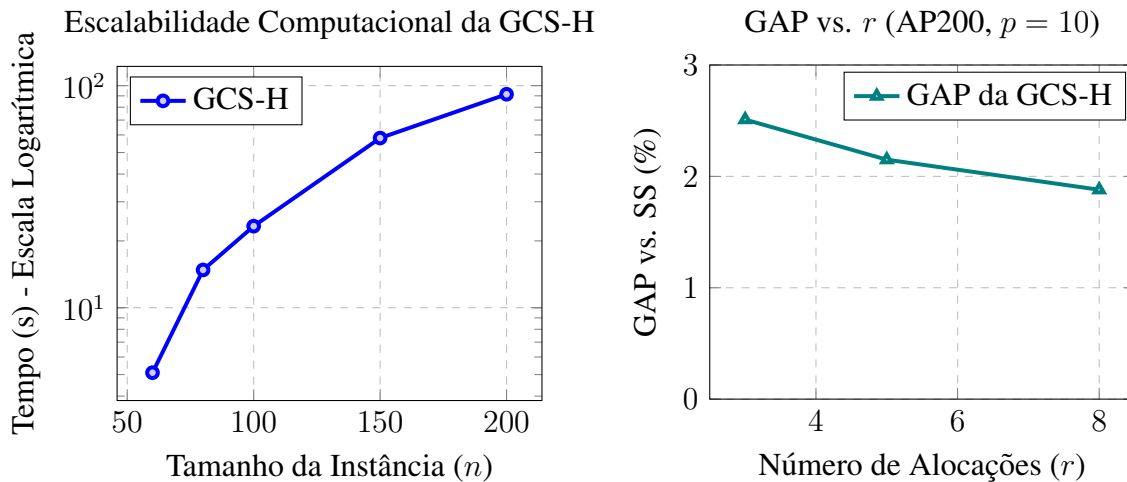


Figura 1. Análise de escalabilidade do tempo (esquerda) e impacto do parâmetro r na qualidade (direita).

O gráfico à direita analisa o impacto do parâmetro de alocação r na qualidade da solução, utilizando como exemplo a instância AP200 com $p = 10$. Os dados mostram uma tendência de melhora na solução à medida que o valor de r aumenta: o *gap* em

relação ao SS diminui de 2.51% para $r = 3$ para 1.88% para $r = 8$. Este comportamento sugere que a GCS-H utiliza de forma eficiente a flexibilidade da alocação múltipla para construir redes de menor custo, se aproximando mais das melhores soluções conhecidas em cenários com maior conectividade permitida.

5. Conclusão

Este trabalho propôs a Heurística por Seleção Geo-Clusterizada (GCS-H), uma abordagem não supervisionada e construtiva para o Problema p-Hub Mediana com r-Alocação. A metodologia demonstrou que, ao substituir buscas iterativas por uma análise geo-estrutural inteligente, é possível alcançar um equilíbrio notável entre a qualidade da solução e a eficiência computacional.

Os resultados obtidos em um conjunto de 231 instâncias da família AP foram conclusivos. A GCS-H provou ser consistentemente superior à meta-heurística GRASP, encontrando soluções de menor custo em tempos de execução drasticamente inferiores. Em comparação com o estado da arte representado pelo Scatter Search, a GCS-H produziu soluções altamente competitivas, com um desvio médio da melhor solução conhecida de apenas 2.35%. Essa pequena diferença na qualidade é compensada por uma vantagem de velocidade que a torna ordens de magnitude mais rápida, consolidando a GCS-H como uma ferramenta prática e poderosa.

Uma limitação deste estudo é a validação em um único benchmark (instâncias AP), embora seja um padrão na área. Trabalhos futuros devem validar a generalidade da GCS-H aplicando-a a outros conjuntos de dados de referência, como o CAB (Civil Aeronautics Board) ou instâncias de maior escala.

A principal contribuição deste trabalho é, portanto, um método prático e escalável que torna a otimização de redes de hubs mais acessível e ágil. Como trabalhos futuros, a hibridização da GCS-H se apresenta como a via mais promissora. A solução gerada pela GCS-H pode servir como uma semente de alta qualidade para meta-heurísticas tradicionais, potencialmente guiando-as para ótimos globais de forma mais rápida e unindo o melhor das abordagens de Machine Learning e Otimização Combinatória.

Referências

- Corberán, , Peiró, J., Martí, R., and Saldanha-da Gama, F. (2018). Heuristic solutions for a class of stochastic uncapacitated p-hub median problems.
- Ernst, A. T. and Krishnamoorthy, M. (1996). Efficient algorithms for the uncapacitated single allocation p-hub median problem. *Location science*, 4(3):139–154.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Love, R. F., Morris, J. G., and Wesolowsky, G. O. (1988). Facilities location: Models & methods. *Publications in operations research series*, 7.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14):281–297.
- O’Kelly, M. E. (1987). The p-hub median problem. *Transportation Science*, 21(3):183–195.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web.

Yaman, H. (2011). Allocation strategies in hub networks. *European Journal of Operational Research*, 211(3):442–451.