



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2014/04.15.17.03-TDI

## **HEURÍSTICAS HÍBRIDAS PARA O PROBLEMA DE ALOCÇÃO DE BERÇOS PARA NAVIOS E PARA UM PROBLEMA DE AGRUPAMENTOS**

Rudinei Martins de Oliveira

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Luiz Antonio Nogueira Lorena, e Geraldo Regis Mauri, aprovada em 03 de abril de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G5QLMB>>

INPE  
São José dos Campos  
2014

## **PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

## **CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**

### **Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

### **Membros:**

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr<sup>a</sup> Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

### **BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

### **REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

### **EDITORAÇÃO ELETRÔNICA:**

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO  
**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

sid.inpe.br/mtc-m21b/2014/04.15.17.03-TDI

## **HEURÍSTICAS HÍBRIDAS PARA O PROBLEMA DE ALOCÇÃO DE BERÇOS PARA NAVIOS E PARA UM PROBLEMA DE AGRUPAMENTOS**

Rudinei Martins de Oliveira

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Luiz Antonio Nogueira Lorena, e Geraldo Regis Mauri, aprovada em 03 de abril de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G5QLMB>>

INPE  
São José dos Campos  
2014

Dados Internacionais de Catalogação na Publicação (CIP)

---

Oliveira, Rudinei Martins de.

Ol4h      Heurísticas híbridas para o problema de alocação de berços para navios e para um problema de agrupamentos / Rudinei Martins de Oliveira. – São José dos Campos : INPE, 2014.  
xxiv + 108 p. ; (sid.inpe.br/mtc-m21b/2014/04.15.17.03-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

Orientadores : Drs. Luiz Antonio Nogueira Lorena, e Geraldo Regis Mauri.

1. Otimização combinatória. 2. Clustering search. 3. Programação linear. 4. Heurísticas. 5. Geração de colunas. I.Título.

CDU 004.023:627.22

---

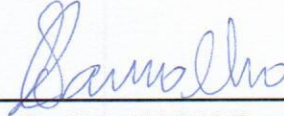


Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de **Doutor(a)** em  
**Computação Aplicada**

Dr. Solon Venâncio de Carvalho



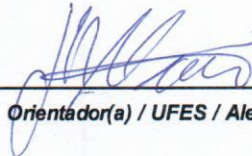
Presidente / INPE / SJC Campos - SP

Dr. Luiz Antonio Nogueira Lorena




Orientador(a) / INPE / São José dos Campos - SP

Dr. Geraldo Regis Mauri



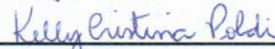
Orientador(a) / UFES / Alegre - ES

Dr. José Carlos Becceneri



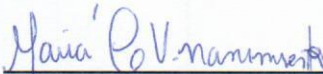
Membro da Banca / INPE / SJC Campos - SP

Dra. Kelly Cristina Poldi



Convidado(a) / UNIFESP / São José dos Campos - SP

Dra. Mariá Cristina Vasconcelos  
Nascimento



Convidado(a) / UNIFESP / São José dos Campos - SP

*Este trabalho foi aprovado por:*

( ) maioria simples

unanimidade

Aluno (a): **Rudinei Martins de Oliveira**

São José dos Campos, 03 de Abril de 2014



*A minha mãe Maria Rosa, a meu pai Sebastião, a minha  
tia Joana (in memoria) e as minhas irmãs Dalila e Júnia.*





## **AGRADECIMENTOS**

Agradeço a Deus pelo seu infinito amor e bondade.

Aos meus orientadores Profs. Drs. Luiz Antonio Nogueira Lorena e Geraldo Regis Mauri pelo incentivo e conhecimento compartilhado, pelos quais sou muito grato.

Ao Instituto Nacional de Pesquisas Espaciais (INPE) pela infraestrutura e a oportunidade de realizar meus estudos.

Aos professores da CAP.

Ao Marcos Antonio Pereira e Antonio Augusto Chaves pelo apoio técnico.

À CAPES pelo auxílio financeiro.

Aos meus amigos do INPE, pelo companheirismo.



## RESUMO

Este trabalho apresenta heurísticas híbridas para o Problemas de Alocação de Berços (PAB) e para o Problema de Agrupamentos. O PAB consiste em alocar navios a posições de atraque, de forma que seja utilizado o máximo de espaço do cais minimizado o tempo de serviço e é resolvido nos casos discreto e contínuo. O PAB pode ser considerado como um dos principais problemas de otimização em terminais marítimos, devido à crescente demanda de navios que transportam contêineres. Neste contexto, é proposta uma nova alternativa para resolvê-lo. Essa alternativa é baseada na aplicação do método *Clustering Search* (CS), utilizando o *Simulated Annealing* como gerador de soluções. O CS é um método iterativo que divide o espaço de busca em grupos e é composto por uma meta-heurística geradora de soluções, um processo de agrupamento e uma heurística de busca local. Em relação ao problema de agrupamentos, busca-se a separação de um conjunto de dados em grupos, tal que os membros de cada grupo sejam similares entre si. Os métodos de solução baseiam-se na aplicação de uma técnica de geração de colunas para resolver o problema  $p$ -medianas. São propostas cinco heurísticas: solução viável a partir do problema mestre, a solução de geração de colunas, uma heurística com *Path-Relinking* considerando as colunas iniciais do processo de geração de colunas, uma solução do problema mestre com *Path-Relinking* e o processo de geração de colunas com *Path-Relinking*. A qualidade dos resultados computacionais obtidos é comprovada com a comparação com métodos recentes da literatura.



# HYBRID HEURISTICS FOR THE BERTH ALLOCATION PROBLEM FOR SHIPS AND A CLUSTERING PROBLEM

## ABSTRACT

This paper presents hybrid heuristics for the Berth Allocation Problem (PAB) and a Clustering Problem. The BAP consists in allocating ships to positions of mooring using the maximum space of the quay and minimizing the service time of the ships and is solved for discrete and continuous cases. The BAP can be considered as a major optimization problem in marine terminals, due to increasing demand of ships carrying containers. In this context, we propose a new alternative to solve it. This alternative is based on the application of the method Clustering Search (CS), using the Simulated Annealing as the solutions generator. CS is an iterative method that divides the search space into groups and consists of a metaheuristic to generate solutions, a process of clustering and a local search. The issue Clustering Problem searches the separation of a set of objects into groups such that members of each group are similar. The hybrid heuristics proposed are based on the application of a column generation technique for solving the  $p$ -median problem. Five heuristics are examined: a feasible solution from the master problem, the column generation solution, a heuristic with Path-Relinking considering the initial columns of the column generation process, a solution of the master problem with Path-Relinking and the column generation with Path-Relinking. The quality of computational results is evidenced by comparison with recent methods of the literature.



## LISTA DE FIGURAS

	<u>Pág.</u>
2.1	Representação de um terminal marítimo. . . . . 6
2.2	Representação de um planejamento para os navios. . . . . 9
2.3	Divisão do cais em berços. . . . . 10
2.4	Representação do PAB como um Problema de Programação de Máquinas. 12
2.5	Representação gráfica do PAB. . . . . 13
2.6	Espaços de tempo em um berço de navios. . . . . 14
2.7	Representação das variáveis de tempo. . . . . 17
2.8	Divisão do cais para o caso do PAB-C. . . . . 21
2.9	Representação das variáveis de tempo e espaço. . . . . 21
2.10	Solução para o PAB-D (instância - i01 - FO = 1409). . . . . 22
2.11	Solução para o PAB-C (instância - i01 - FO= 1583). . . . . 23
2.12	Fluxograma do CS. . . . . 25
2.13	Exemplo de uma solução para o PAB. . . . . 26
2.14	Heurística de distribuição. . . . . 27
2.15	Heurística de programação para o PAB-D e PAB-C . . . . . 27
2.16	Heurística de Atualização. . . . . 28
2.17	Algoritmo <i>Simulated Annealing</i> utilizado no CS. . . . . 30
2.18	Movimento de re-ordenar navios. . . . . 31
2.19	Movimento de re-alocar navios. . . . . 31
2.20	Movimento de trocar navios. . . . . 32
2.21	Algoritmo EXECUTAR-CS. . . . . 33
2.22	<i>Path-relinking</i> utilizado na atualização dos centros de <i>cluster</i> . . . . . 34
2.23	Busca local. . . . . 35
3.1	Algoritmo GC. . . . . 48
3.2	Aplicação do PR. . . . . 50
3.3	Algoritmo HSI . . . . . 52
3.4	Gráfico das diferenças - City Block (HSI-PR) . . . . . 70
3.5	Gráfico das diferenças - City Block (HPL) . . . . . 71
3.6	Gráfico das diferenças - City Block (HPL-PR) . . . . . 71
3.7	Gráfico das diferenças - City Block (HGC) . . . . . 72
3.8	Gráfico das diferenças - City Block (HGC-PR) . . . . . 72
3.9	Gráfico das diferenças - Euclidiana (HSI-PR) . . . . . 73
3.10	Gráfico das diferenças - Euclidiana (HPL) . . . . . 73

3.11	Gráfico das diferenças - Euclidiana (HPL-PR)	74
3.12	Gráfico das diferenças - Euclidiana (HGC)	74
3.13	Gráfico das diferenças - Euclidiana (HGC-PR)	75
3.14	Gráfico das diferenças - Pearson (HSI-PR)	75
3.15	Gráfico das diferenças - Pearson (HPL)	76
3.16	Gráfico das diferenças - Pearson (HPL-PR)	76
3.17	Gráfico das diferenças - Pearson (HGC)	77
3.18	Gráfico das diferenças - Pearson (HGC-PR)	77
3.19	Gráfico das diferenças - Cosseno (HSI-PR)	78
3.20	Gráfico das diferenças - Cosseno (HPL)	78
3.21	Gráfico das diferenças - Cosseno (HPL-PR)	79
3.22	Gráfico das diferenças - Cosseno (HGC)	79
3.23	Gráfico das diferenças - Cosseno (HGC-PR)	80
B.1	i01 - FO = 1583	93
B.2	i02 - FO = 1315	94
B.3	i03 - FO = 1207	94
B.4	i04 - FO = 1380	95
B.5	i05 - FO = 1262	95
B.6	i06 - FO = 1413	96
B.7	i07 - FO = 1324	96
B.8	i08 - FO = 1381	97
B.9	i09 - FO = 1633	97
B.10	i10 - FO = 1333	98
B.11	i11 - FO = 1527	98
B.12	i12 - FO = 1505	99
B.13	i13 - FO = 1424	99
B.14	i14 - FO = 1273	100
B.15	i15 - FO = 1345	100
B.16	i16 - FO = 1491	101
B.17	i17 - FO = 1302	101
B.18	i18 - FO = 1518	102
B.19	i19 - FO = 1517	102
B.20	i20 - FO = 1406	103
B.21	i21 - FO = 1461	103
B.22	i22 - FO = 1437	104
B.23	i23 - FO = 1395	104
B.24	i24 - FO = 1351	105



B.25 i25 - FO = 1513 . . . . .	105
B.26 i26 - FO = 1448 . . . . .	106
B.27 i27 - FO = 1349 . . . . .	106
B.28 i28 - FO = 1461 . . . . .	107
B.29 i29 - FO = 1323 . . . . .	107
B.30 i30 - FO = 1487 . . . . .	108



## LISTA DE TABELAS

	<u>Pág.</u>
2.1 Comparação dos resultados obtidos (CS) para o PAB-D. . . . .	37
2.2 Comparação entre os resultados do BT, AM and CS para o PAB-C. . . .	40
2.3 Melhorias e comparação com o caso discreto. . . . .	41
3.1 Comparação do resultados de CRand para a distância City Block (cliques).	58
3.2 Comparação dos valores de função objetivo (FO) para o $p$ -medianas, o tempo da melhor solução (TM) e do tempo total de execução do algoritmo (TT) para a distância City Block. . . . .	58
3.3 Comparação do resultados de CRand para a distância Euclidiana (cliques).	59
3.4 Comparação dos valores de função objetivo (FO) para o $p$ -medianas, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a distância distância Euclidiana . . . . .	59
3.5 Comparação do resultados de CRand para a Correlação de Pearson (cliques). . . . .	60
3.6 Comparação dos valores de função objetivo (FO) para o $p$ -medianas, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a Correlação de Pearson. . . . .	60
3.7 Comparação do resultados de CRand para a correlação cosseno (cliques).	61
3.8 Comparação dos valores de função objetivo (FO) para o $p$ -medianas, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a Correlação Cosseno. . . . .	61
3.9 Comparação do resultados de CRand para a distância City Block ( $p$ -medianas). . . . .	63
3.10 Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e do tempo total de execução do algoritmo (T.T.) para a distância City Block ( $p$ -medianas). . . . .	63
3.11 Comparação do resultados de CRand para a distância Euclidiana ( $p$ -medianas). . . . .	64
3.12 Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a distância Euclidiana ( $p$ -medianas). . . . .	64
3.13 Comparação do resultados de CRand para a Correlação de Pearson ( $p$ -medianas). . . . .	65

3.14	Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a Correlação de Pearson ( $p$ -medianas).	65
3.15	Comparação do resultados de CRand para a correlação cosseno ( $p$ -medianas).	66
3.16	Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a correlação ( $p$ -medianas).	66
3.17	Diferenças para a distância <i>City Block</i> .	68
3.18	Diferenças para a distância Euclidiana.	68
3.19	Diferenças para a correlação de Pearson.	69
3.20	Diferenças para a correlação Cosseno.	69

## LISTA DE ABREVIATURAS E SIGLAS

CRAnd	–	<i>Correct (adjusted) rand index</i>
FO	–	Função objetivo
GC	–	Geração de colunas
HGC	–	Heurística de geração de colunas
HGC-PR	–	Heurística de geração de colunas com <i>path-relinking</i>
HPL	–	Heurística de programação linear
HPL-PR	–	Heurística de programação linear com <i>path-relinking</i>
PAB	–	Problema de alocação de berços
PAB-C	–	Problema de alocação de berços contínuo
PAB-D	–	Problema de alocação de berços discreto
PL	–	Programação linear
PM	–	Problema mestre
PR	–	<i>Path-relinking</i>
Rand	–	<i>Rand index</i>
SA	–	<i>Simulated annealing</i>



## LISTA DE SÍMBOLOS

$\alpha$	–	Razão de resfriamento
$\Delta$	–	Varição
$\mu, \rho$	–	Custos duais
$\omega$	–	Coefficiente de penalidade
$r$	–	Índice de ineficácia
$\lambda$	–	Limitante
$T$	–	Temperatura corrente
$T_0$	–	Temperatura inicial
$T_C$	–	Temperatura de congelamento





## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO . . . . .</b>	<b>1</b>
1.1 Objetivos . . . . .	2
1.2 Organização da tese . . . . .	3
<b>2 ALOCAÇÃO DE BERÇOS PARA NAVIOS . . . . .</b>	<b>5</b>
2.1 Problema de alocação de berços . . . . .	7
2.2 Classificação do PAB . . . . .	9
2.3 Revisão abreviada da literatura para alocação de berços . . . . .	11
2.3.1 Formulação matemática para o PAB-D . . . . .	16
2.3.2 Abordagem do PAB-C . . . . .	20
2.4 <i>Clustering Search</i> aplicado ao PAB . . . . .	23
2.5 Experimentos computacionais para o PAB . . . . .	36
2.6 Caso discreto do PAB . . . . .	36
2.7 Caso contínuo do PAB . . . . .	38
2.8 Considerações finais sobre o PAB . . . . .	42
<b>3 PROBLEMA DE AGRUPAMENTOS . . . . .</b>	<b>43</b>
3.1 Revisão abreviada da literatura para agrupamentos . . . . .	44
3.2 Geração de colunas para o problema de $p$ -medianas . . . . .	46
3.3 Heurísticas híbridas para o problema de agrupamentos . . . . .	49
3.4 Experimentos computacionais . . . . .	52
3.4.1 Dados utilizados . . . . .	52
3.4.2 Resultados obtidos para o problema de agrupamentos . . . . .	54
3.5 Conclusões sobre problemas de agrupamentos . . . . .	81
<b>4 CONCLUSÕES FINAIS E TRABALHOS FUTUROS . . . . .</b>	<b>83</b>
4.1 Trabalhos futuros . . . . .	84
<b>REFERÊNCIAS BIBLIOGRÁFICAS . . . . .</b>	<b>85</b>
<b>APÊNDICE A - TRABALHOS PUBLICADOS . . . . .</b>	<b>91</b>
A.1 Trabalhos publicados em periódicos . . . . .	91

A.2	Trabalhos publicados em periódicos e apresentado em evento internacional	92
A.3	Trabalhos apresentados em evento nacional . . . . .	92
	<b>APÊNDICE B - Soluções obtidas para o PAB-C . . . . .</b>	<b>93</b>

## 1 INTRODUÇÃO

Soluções de problemas de Otimização Combinatória (OC) possuem uma variedade de aplicações tais como automação, energia, petróleo, transportes, aviação, naval entre outras. Bons resultados foram obtidos, mas ainda muito precisa ser feito, como por exemplo, melhorias no planejamento urbano, logística de atendimento dos portos, comunicação celular, além de aprimorar o acesso aos meios de transportes, centros de saúde etc.

Para resolução de problemas de OC é necessário trabalhar a modelagem matemática, descrever por expressões matemáticas os fenômenos físicos, biológicos, químicos, econômicos e as leis que regem os problemas do dia a dia. Empresas utilizam técnicas de OC para minimizar despesas e maximizar lucros. Maximizar lucros envolve usar de forma eficiente os recursos disponíveis, a sequência de itens a serem produzidos e melhores rotas de distribuições.

Vários problemas podem ser classificados em OC: roteirização de veículos, problemas de cortes, problemas de empacotamento, problemas de manufatura, problema de rotulação cartográfica de pontos, problemas de atribuição, problema de alocação de berços, problema de agrupamentos, entre outros.

O Problema de Alocação de Berços (PAB) e o problema de agrupamentos são o foco desta tese. O PAB trabalha com a elaboração de um planejamento de berços, procurando determinar onde e quando um navio poderá atracar no cais (ELWANY et al., 2013). O objetivo é minimizar o tempo de serviço dos navios nos berços. Na maioria dos estudos encontrados, os berços usados para o atendimento dos navios são considerados como discretos, já o caso contínuo é uma novidade nas pesquisas atuais. Para o PAB discreto (PAB-D) as soluções obtidas neste trabalho são as ótimas para todas as instâncias do porto de *Gioia Tauro* (Itália). Neste trabalho além do PAB-D o PAB contínuo (PAB-C) também é estudado. No PAB-D, os berços atendem um navio de cada vez. O PAB-C considera o cais como uma linha contínua onde múltiplos navios podem compartilhar um mesmo berço. Assim, a quantidade de navios atendidos é maior, o que melhora o fluxo de atendimento dos navios no porto.

Quando se resolve problemas complexos como PAB, as restrições físicas, temporal e espacial, devem ser respeitadas na modelagem do problema. Em geral as restrições

estão relacionadas a profundidade da água, distância entre os guindastes. Em relação ao tempo, cada navio tem uma janela de tempo, com horário de início e fim do atendimento. Para o PAB-C é considerada a largura e comprimento de cada navio, desta forma, além destas restrições citadas há também aquelas relacionadas ao espaço disposto para os navios, de forma que as sobreposições devem ser evitadas.

O outro problema abordado neste trabalho foi o de problema de agrupamentos. O problema de agrupamentos se caracteriza na dificuldade de separar um conjunto de dados em grupos de objetos semelhantes, tal que, objetos de um mesmo grupo sejam diferentes de objetos de outros grupos (BERKHIN, 2006). Dependendo do procedimento utilizado para separar os dados haverá diferentes partições. O que se faz é definir um critério para determinar as distâncias entre todos os objetos em um conjunto de dados. Métricas como a distancia Euclidiana, *City Block*, Pearson e Cosseno usadas nesta tese, auxiliam a definir se um objeto é similar a outro.

Nos dias atuais, dentre as técnicas de resolução destes problemas cita-se os métodos heurísticos. Uma heurística é uma técnica que busca boas soluções a um custo computacional razoável sem poder garantir a viabilidade ou otimalidade da solução (REEVES, 1995). A junção de duas ou mais heurísticas com estratégias diferentes são denominadas heurísticas híbridas. Algumas heurísticas modificadas denominadas meta-heurísticas podem ser inspiradas em princípios físicos ou biológicos. Como exemplo, têm-se os métodos: *simulated annealing* (KIRKPATRICK et al., 1983), *busca tabu* (GLOVER; LAGUNA, 1997), *genetic algorithms* (GOLDBERG, 1989). Dentro deste contexto, no tópico seguinte são apresentados os objetivos propostos por esta tese.

## 1.1 Objetivos

O objetivo desta tese é resolver o problema de alocação de berços e o problema de agrupamentos. Para o PAB foi proposta uma aplicação do método híbrido conhecido como *Clustering Search* (CS), utilizando o *simulated annealing* como gerador de soluções. Este método é basicamente composto por uma meta-heurística (*simulated annealing*), um processo de agrupamento e uma heurística de busca local. A meta-heurística gera soluções, estas são armazenadas no grupo mais similar pelo processo de agrupamento e a busca local é executada quando o grupo se torna promissor. Grupo promissor é identificado quando a quantidade de soluções atingem certo limitante definido previamente.

Para solução do problema de agrupamentos é necessário definir os grupos que armarão as seleções dos objetos similares. Os grupos são obtidos pela técnica de geração de colunas para solução do problema de  $p$ -medianas. Soluções de  $p$ -medianas minimizam a soma das distâncias entre objetos e sua mediana mais próxima. Espera-se que as características entre os objetos que estão alocados a mesma mediana sejam similares.

Desta forma, propõem-se cinco heurísticas híbridas: uma solução viável a partir do problema mestre, a solução de geração de colunas, uma heurística com *path-relinking* considerando as colunas iniciais do processo de geração de colunas, uma solução do problema mestre com *path-relinking* e o processo de geração de colunas com *path-relinking*.

Os resultados computacionais das heurísticas híbridas propostas neste trabalho são comparados com métodos recentes da literatura, permitindo assim verificar a eficiência das abordagens apresentadas.

## 1.2 Organização da tese

O restante da tese está organizado como segue. O Capítulo 2 apresenta uma descrição detalhada sobre o problema de alocação de berços, sua classificação, uma revisão bibliográfica dos trabalhos sobre o PAB, a formulação matemática existente no contexto de alocação de berços e os resultados e considerações finais para o PAB.

O Capítulo 3 apresenta o problema de agrupamentos, sua revisão bibliográfica, a formulação de  $p$ -medianas utilizada na técnica geração de colunas, detalha as heurísticas híbridas propostas, os experimentos computacionais, dados, resultados e considerações finais.

O Capítulo 4 apresenta as conclusões finais, as contribuições e a possível continuação para o trabalho realizado.

O Apêndice A apresenta os trabalhos publicados e/ou apresentados em eventos científicos

Por fim, Apêndice B apresenta os gráficos das soluções obtidas pelo CS para PAB-C.



## 2 ALOCAÇÃO DE BERÇOS PARA NAVIOS

O transporte marítimo de mercadorias cresceu muito nos últimos anos em decorrência do aumento do comércio e do crescimento econômico internacional (HANSEN et al., 2008). O comércio internacional tem como pilar central o transporte marítimo. Aproximadamente 80% do comércio global de mercadorias é realizado através do mar (BUHRKAL et al., 2009). Em 2008, a frota de navios que transportam contêineres teve um aumento em sua capacidade de 17,3 milhões de toneladas, ou 11,9%, e passaram a representar 13,6% do total mundial. No início de 2009, a frota mercante mundial atingiu 1,19 milhões de toneladas, um crescimento de 6,7% em comparação à janeiro de 2008 e desde o início da década, a quantidade de contêineres transportados aumentou em 154% (UNCTAD, 2009). Em janeiro de 2010, a frota mercante mundial chegou a 1,276 milhões de toneladas aumentando em 84 milhões em relação ao ano anterior (UNCTAD, 2010).

Desse modo, devido ao crescimento do comércio internacional, o intenso fluxo de navios e contêineres nos portos, estes são forçados a investir pesadamente em meios otimizados para acomodar os navios, aprofundando e alargando canais e construindo novas instalações de atracação, tudo para que o tempo de atendimento do navio seja o menor possível. Assim, para que o navio complete seu ciclo no terminal, uma parte da responsabilidade diz respeito ao local de atracação, ou seja, ao cais, além da localização dos berços no terminal.

O cais é uma plataforma que se liga ao mar para facilitar o carregamento e o descarregamento de carga de navios. Já os berços são os locais onde os navios podem atracar. Esses são equipados com grandes guindastes, usados para carregar e descarregar contêineres, que são transferidos para um pátio por uma frota de veículos. No pátio esses contêineres ficam depositados até que lhes sejam determinado a próxima rota, seja ela por trens, por caminhões ou novamente por navios.

Uma representação de um terminal marítimo é apresentada na Figura 2.1, tal que os navios encontram-se em uma das três condições: chegando no porto, esperando para atracar ou em atendimento nos berços disponíveis. Quanto ao atendimento, uma vez iniciado, o navio ficará atracado no berço até que todo o processamento dos contêineres seja finalizado.

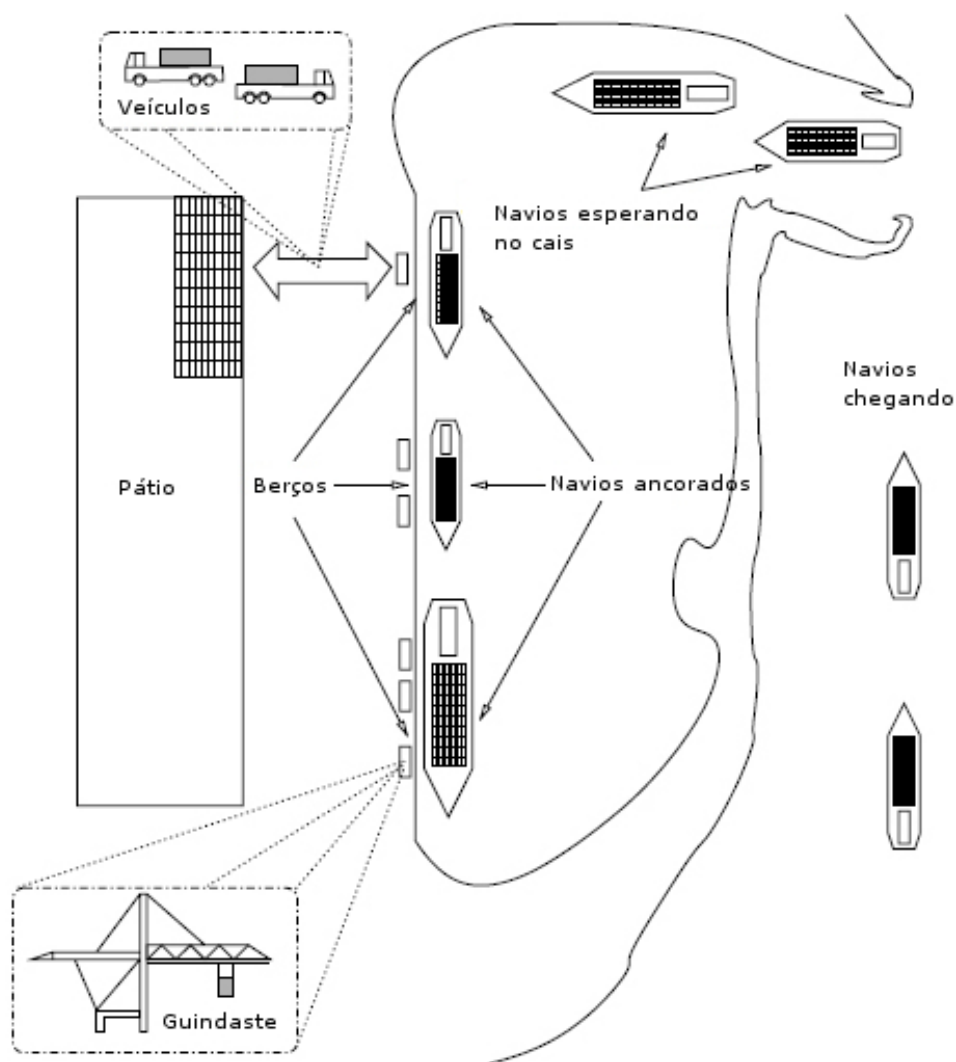


Figura 2.1 - Representação de um terminal marítimo.  
 Fonte: Adaptada de Cordeau e Laporte (2005).

Segundo Hansen et al. (2008) a importância da alocação de berços se torna mais visível com o aumento da frequência dos navios nos portos, o que por consequência gera o crescimento das filas, ocasionando prejuízos com o acréscimo no tempo de espera das embarcações. Isso poderia ser evitado se os navios fossem atendidos pelos berços com a rapidez necessária para o carregamento e descarregamento da carga. Nesse contexto, a acomodação dos navios nos berços é um fator primordial, visto que, os navios estão ficando cada vez maiores, graças aos avanços tecnológicos na sua construção. Vale ressaltar a necessidade de que juntamente com o desenvolvi-



mento comercial os portos consigam acompanhar tal crescimento se modernizando, produzindo mão de obra e tecnologia capaz de suprir o aumento da demanda do transporte de mercadorias.

Assim sendo, a necessidade de um bom planejamento logístico para acomodar os navios nos berços, além da busca por uma melhor distribuição do espaço, minimizando o tempo total de permanência dos navios no porto, com o objetivo de agilizar o atendimento dos navios, motivou o surgimento de um problema conhecido na literatura como Problema de Alocação de Berços (PAB).

## **2.1 Problema de alocação de berços**

O PAB pode ser classificado como um dos principais problemas de otimização em terminais marítimos. Isso por causa da crescente demanda de navios, além da preferência pelo uso de embarcações cada vez maiores, que reduzem o custo com o transporte e podem transportar milhares de toneladas de mercadorias. Com isto, estudos que possam contribuir com uma melhor distribuição do espaço no cais e uma eficiente vazão dos contêineres são necessários, principalmente em relação ao PAB que é o foco deste trabalho.

O Problema de Alocação de Berços consiste em alocar navios a posições de atraque, de forma que seja utilizado o máximo de espaço do cais minimizando o tempo de serviço. As decisões a serem tomadas dizem respeito à posição e ao instante em que o navio deverá atracar (IMAI et al., 2001).

Em relação à posição de atraque, há restrições relativas à profundidade da água e à máxima distância em relação ao local mais favorável ao longo do cais. Isso em relação aos contêineres estocados, que podem estar longe do berço disponível para o navio (CORDEAU; LAPORTE, 2005). Quanto ao tempo necessário para que o navio seja carregado ou descarregado, deve-se verificar a capacidade do berço onde ele será alocado. Em geral, o tempo de atendimento é determinado pela quantidade de guindastes disponíveis, além da distância entre o berço e a localização dos contêineres no pátio (MONACO; SAMMARRA, 2007). Assim, o tempo de serviço é o tempo de espera mais o tempo de atendimento para cada navio. Esses tempos podem ser fixos ou não. Segundo Bierwirth e Meisel (2009), o tempo de atendimento na maioria das vezes é determinístico, mas também podem ser:

- a) fixo, sendo conhecido de forma antecipada;
- b) dependente da posição de atraque dos navios;
- c) dependente da quantidade de guindastes a disposição dos navios, como comentado anteriormente;
- d) dependente da distribuição desses guindastes;
- e) dependente das três últimas combinações anteriores (itens (b), (c) e (d)).

Outrossim, o principal objetivo dessa programação em relação ao tempo é a minimização da espera junto com o tempo de atendimento, ou seja, o tempo de serviço que os navios gastam, buscando a reduzir o tempo que o navio fica no porto.

Dessa forma, uma solução para o PAB pode ser representada, segundo Bierwirth e Meisel (2009), como um diagrama de tempo-espço (Figura 2.2). Nessa representação, o eixo vertical (Cais) corresponde a um segmento de 600 m, o eixo horizontal (Tempo) está dividido em períodos de uma hora, totalizando 18 horas (eixo completo). O tempo de serviço gasto pelo navio é representado por um retângulo. A altura do retângulo corresponde ao comprimento do navio e a largura corresponde ao tempo de atendimento esperado. O vértice esquerdo abaixo do retângulo dá o instante de atracamento e a posição no cais. Por exemplo, o navio três tem tempo de atendimento igual a duas horas e sua posição no cais é a quatrocentos.

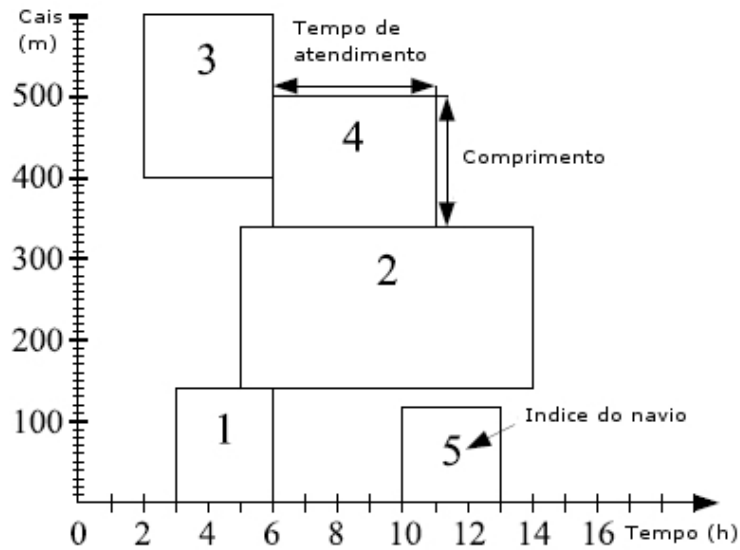


Figura 2.2 - Representação de um planejamento para os navios.  
 Fonte: Adaptada de Bierwirth e Meisel (2009).

## 2.2 Classificação do PAB

O PAB possui grande quantidade de restrições físicas, técnicas, entre outras. Isto faz com que seja possível modelá-lo de diferentes maneiras. Quanto aos aspectos espaciais dos berços, o PAB pode ser modelado como discreto, contínuo ou híbrido (IMAI et al., 2005). Além disso, se for levado em conta a chegada dos navios, o problema pode ser tratado como estático ou dinâmico (IMAI et al., 2001). O caso estático assume que todos os navios já estão no porto para o atendimento, o caso dinâmico permite aos navios chegarem a qualquer momento. A Figura 2.3 apresenta alguns tipos de divisões realizadas no cais e na sequência, são detalhados os diferentes tipos de modelagens enunciadas anteriormente.

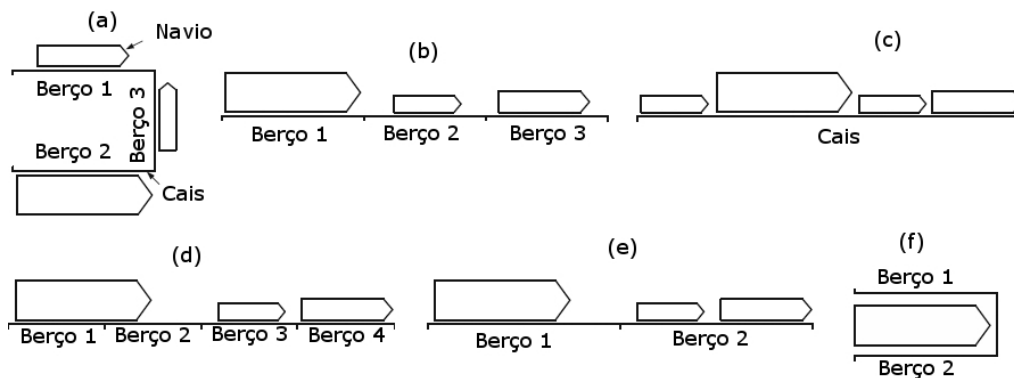


Figura 2.3 - Divisão do cais em berços.

Fonte: Adaptada de Bierwirth e Meisel (2009).

- PAB discreto (PAB-D): o cais é dividido em vários berços e somente um navio é atendido de cada vez em cada berço, independentemente do seu tamanho. O particionamento do cais pode ser como o apresentado na Figura 2.3 (a) ou para facilitar o planejamento pode ser fixo como na Figura 2.3 (b).
- PAB contínuo (PAB-C): não há nenhuma divisão do cais e, dessa forma, os navios podem atracar em qualquer posição (Figura 2.3 (c)). Ao contrário do caso discreto, os problemas envolvendo casos contínuos são mais difíceis de serem resolvidos devido a sua complexidade, apesar de o espaço no cais ser mais aproveitado.
- PAB híbrido: como no caso discreto, o cais é dividido em berços, entretanto, os navios grandes podem ocupar mais de uma posição (Figura 2.3 (d)), permitindo assim que navios pequenos compartilhem seu berço (Figura 2.3 (e)) e há também o caso em que um navio é servido por dois berços opostos, ou seja, o navio é atendido pelos dois lados (Figura 2.3 (f)).

Esta classificação considerando o caso híbrido do PAB foi apresentada inicialmente por Imai et al. (2005), entretanto esta é uma definição controversa, por exemplo Cordeau e Laporte (2005) considera este caso como um problema contínuo, o que na prática é uma realidade, porque não existe PAB contínuo segundo esta definição de Imai et al. (2005), pois os guindastes são fixo e não podem se deslocar para atender

os navios quando eles chegam. Desta forma, como neste trabalho foram usados os resultados de Cordeau e Laporte (2005) para comparação, o PAB será tratado apenas como discreto e contínuo.

### 2.3 Revisão abreviada da literatura para alocação de berços

Os trabalhos iniciais sobre o PAB surgiram no final dos anos 80, quando Thurman (1989) propôs um modelo de otimização para o planejamento de berços para a estação naval Norfolk (EUA). A partir desse modelo, Brown et al. (1994) elaboraram um plano para minimizar os conflitos dos carregamentos nessa mesma estação. Os mesmos autores ainda apresentaram um planejamento de berços para submarinos em Brown et al. (1997). São raros os trabalhos relacionados ao PAB até meados dos anos 90. Entretanto, tais trabalhos vêm ganhando foco, principalmente na última década.

O trabalho proposto por Imai et al. (2001) abordou o PAB em sua forma estática e dinâmica. Os autores apresentam um método baseado na relaxação lagrangiana do problema original. Os testes foram realizados para 5, 7, e 10 berços, além de 25 e 50 navios. No mesmo ano, Nishimura et al. (2001) desenvolveram um algoritmo genético para solução do PAB, que é tratado como dinâmico e discreto. Foram usados a mesma quantidade de berços e navios de Imai et al. (2001). As instâncias utilizadas nos dois trabalhos foram geradas de forma aleatórias e não estão disponíveis.

No ano seguinte, Guan et al. (2002) formularam o PAB como um Problema de Programação de Máquinas. Nessa abordagem, um *job* exige um processo de vários processadores ao mesmo tempo. Passando para o problema de alocação, isso significa que os navios são atracados em um berço com múltiplos guindastes, visto que os navios são representados pelos *jobs* e os guindastes pelos processadores. Além disso, esse berço pode servir vários navios ao mesmo tempo. A quantidade de navios e berços utilizados nos testes foram 6 e 12 e as instâncias não foram divulgadas. Assim, uma heurística é proposta para o PAB com o objetivo de minimizar o tempo total de atendimento dos navios. A Figura 2.4 representa uma solução para o PAB e são considerados 12 processadores (guindastes) e 6 *jobs* (navios). No eixo horizontal do gráfico é dado o tempo que os navios poderão utilizar e no eixo vertical estão os processadores.

Imai et al. (2003) aprimoraram sua abordagem anterior (relaxação lagrangiana em

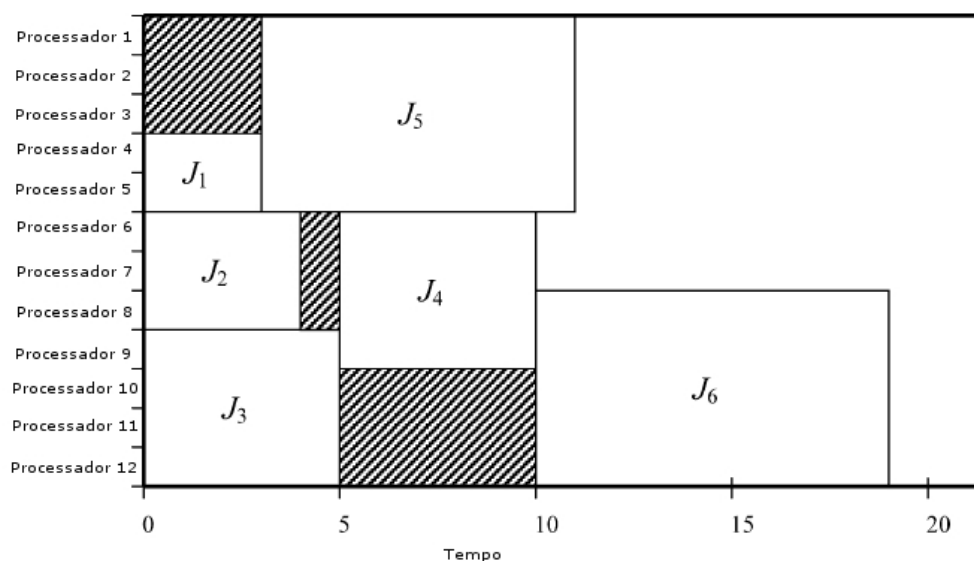


Figura 2.4 - Representação do PAB como um Problema de Programação de Máquinas.  
 Fonte: Adaptada de Guan et al. (2002).

(IMAI et al., 2001)) considerando diferentes prioridades de atendimento entre os navios. Ademais, os autores propuseram um algoritmo genético como método de solução. Os testes foram realizados para 25, 50, 75 e 150 navios e 5 berços. As instâncias não foram divulgadas.

O PAB é estudado como discreto e contínuo em Imai et al. (2005). Os autores demonstram sua semelhança com o problema de corte de estoques (PCE). A Figura 2.5 apresenta a modelagem geométrica utilizada, onde os navios são definidos como retângulos e os berços disponíveis são representados como uma caixa grande. No eixo horizontal os retângulos e a caixa correspondem ao comprimento físico do navio e do cais, respectivamente, enquanto o eixo vertical corresponde ao tempo gasto pelo navio e o tempo total disponível. Assim, o problema que surge é como os retângulos são empacotados.

Cordeau e Laporte (2005) propõem duas formulações e duas heurísticas baseadas na *busca tabu* para resolver o PAB. Os autores apresentam os testes realizados para o porto de Gioia Tauro (Itália). São definidos também três tipos de intervalos de tempos utilizados em um porto de navios. Desse modo, a Figura 2.6 apresenta o período de duração de cada intervalo, desde a chegada do navio no cais até sua saída. O tempo de espera é o tempo em que o navio fica parado no cais aguardando

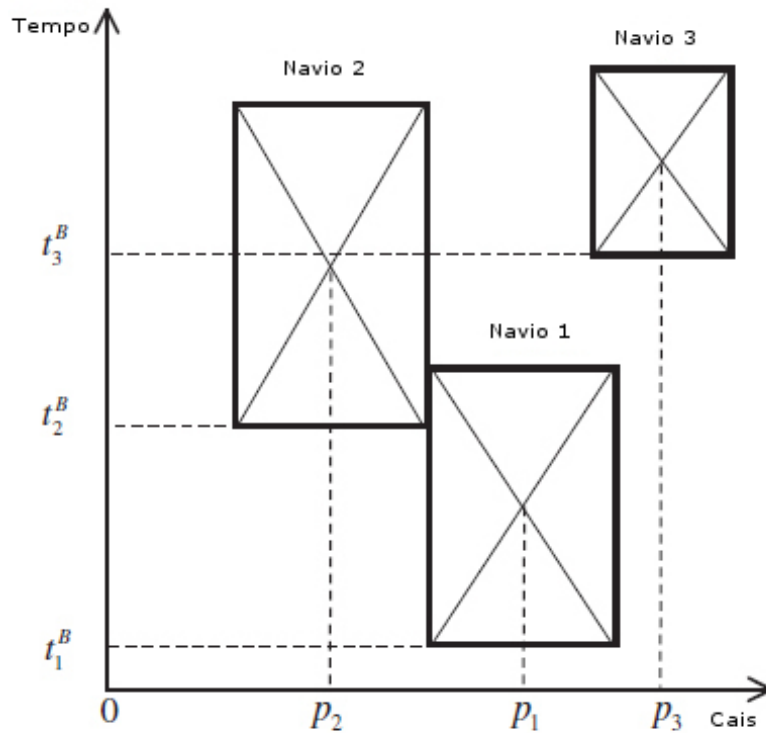


Figura 2.5 - Representação gráfica do PAB.  
 Fonte: Adaptada de Imai et al. (2005).

o seu atendimento. O tempo de atendimento é quando o navio está sendo servido no berço. Já o tempo de serviço é a soma do tempo de espera mais o tempo de atendimento. Nesse trabalho também aparece o conceito de janela de tempo, que é definida como o tempo disponível para o atendimento do navio. Os testes realizados abordaram o PAB discreto e contínuo.

Imai et al. (2008) consideram as restrições físicas do porto, representadas pela diversidade dos navios que chegam (comprimento). Os resultados são obtidos por meio de um algoritmo genético para apenas 5 berços.

Buhrkal et al. (2009) tratam o caso discreto do PAB, considerando uma representação baseada no problema de roteamento de veículos com múltiplas garagens e janelas de tempo, assim como apresentado por Cordeau e Laporte (2005). Além disso, os autores também utilizam um modelo de particionamento, que por sua vez encontra soluções ótimas para as instâncias propostas por Cordeau e Laporte (2005).

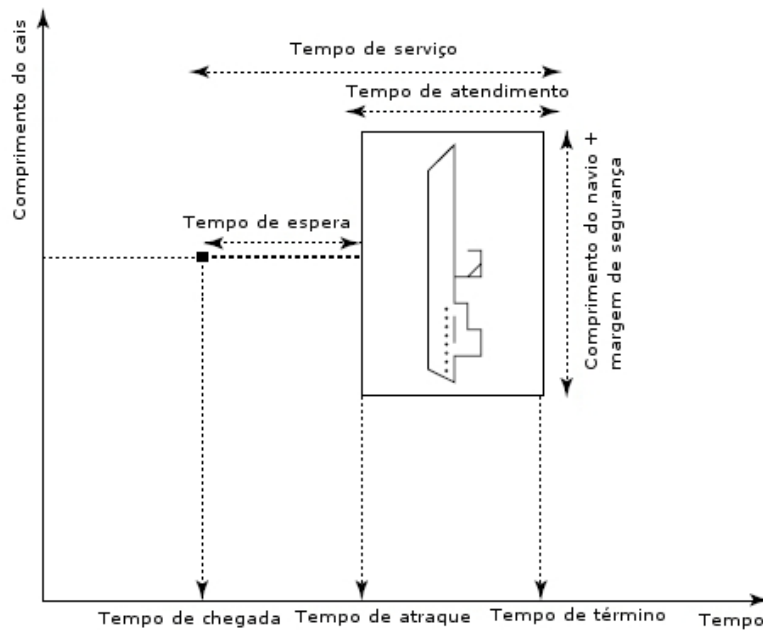


Figura 2.6 - Espaços de tempo em um berço de navios.  
 Fonte: Adaptada de Cordeau e Laporte (2005).

Meisel e Bierwirth (2009) estudam o PAB junto com o problema de atribuição de guindastes. Neste artigo, os dois métodos foram combinados, buscando minimizar o custo em utilizar os guindastes, além do tempo de atendimento dos navios. Enfim, o problema é tratado como discreto e para solução são utilizadas duas meta-heurísticas: a *squeaky wheel optimization* e a *busca tabu*.

Giallombardo et al. (2010) apresentam um modelo de programação quadrática e um de programação linear para representar o PAB. Além disso, os autores utilizam uma *busca tabu* e uma técnica de programação matemática para resolver as instâncias baseadas em dados reais. Cheong et al. (2010) apresentam uma aplicação do método *multiobjective evolutionary algorithm* para resolver o PAB. Xu et al. (2012) modelam o PAB como um problema de programação de máquinas paralelas, onde a distribuição dos navios nos berços consideram as restrições de profundidade e condições das marés.

A maioria dos trabalhos encontrados atualmente aborda o caso discreto do PAB, mas o caso contínuo vem ganhando ênfase nos últimos anos. Em seguida são apresentados alguns trabalhos onde o caso contínuo do PAB foi abordado. No artigo de



Lim (1998) o autor transformou o PAB-C em uma forma restrita do problema de empacotamento em duas dimensões e mostrou que o problema é NP-completo. Li et al. (1983) estudaram o problema de *multiple-job-on-one-processor pattern*, onde vários *jobs* podem ser processados por um único processador simultaneamente, desde que não exceda a capacidade do processador. Este problema é motivado pela operação de alocação de berços, que é alocar navios (*jobs*) para um berço (processador), onde os navios pequenos, podem dividir o berço com outras embarcações para carga ou descarga de mercadorias.

Especificamente no Brasil, vários trabalhos vêm se destacando no tratamento do PAB. Rios et al. (2003) elaboraram um modelo de decisão para o planejamento de capacidade dos terminais de contêineres. Silva (2007) faz uma análise das pesquisas científicas sobre os problemas existentes no sistema portuário, buscando verificar as diferentes formas aplicadas na resolução do PAB.

Mauri et al. (2008a) propõem uma abordagem baseada na aplicação do *simulated annealing* para resolução do caso discreto do PAB. Os autores tratam o problema como um problema de roteamento de veículos com múltiplas garagens e janelas de tempo. Os resultados computacionais superam os obtidos pelo CPLEX e pela *busca tabu* proposta por Cordeau e Laporte (2005). Por fim, Mauri et al. (2008b) tratam o PAB com um método híbrido chamado ATP/PL, que utiliza o algoritmo de treinamento populacional em conjunto com um modelo de programação linear por meio da técnica de geração de colunas. Os resultados obtidos superam os apresentados em Mauri et al. (2008a). Outra abordagem contínua do PAB foi proposto por Mauri et al. (2011), que usou um Algoritmo Memético (AM) para melhorar as soluções relatados em (CORDEAU; LAPORTE, 2005). O AM utiliza um algoritmo *Simulated Annealing* (SA) atuando como mecanismo de busca local, e o algoritmo SA também é aplicado de modo independente para resolver o PAB. Uma heurística de duas fases é também apresentada para calcular o tempo de atendimento e a posição de todos os navios durante a execução do AM e SA.

Barros et al. (2011) propõem um modelo de programação linear inteira para representar o PAB que considera o problema de alocação de berços em portos graneleiros com restrição de níveis de estoque. Neste caso o PAB considera as condições favoráveis de maré e estoque para atracar. Como o porto está associado a uma fábrica, as cargas transportadas estão em estoque, a fábrica produz a carga e estoca e os navios consomem este estoque. Assim, isto resulta em uma restrição para a atracação,

tendo em vista os navios que transportam certo tipo de carga não deve atracar se não tiver a quantidade que ele precisa disponível no estoque.

A partir desta revisão fica evidente que vários autores utilizaram métodos heurísticos para resolver o PAB, tanto para o caso discreto quanto para o caso contínuo, o caso contínuo é menos explorado, apenas alguns poucos trabalhos apresentam dados com instâncias que estão disponíveis para comparação. Já no caso discreto têm-se muitos trabalhos, mas a maioria fica limitada a problemas reais, cujas instâncias não estão disponíveis para comparação. De forma geral, a maioria dos trabalhos apresentados trabalham com problemas pequenos com poucos navios e berços, o que reforça a importância dos resultados obtidos por esta tese.

### 2.3.1 Formulação matemática para o PAB-D

Esta seção descreve a modelagem utilizada para o PAB-D descrita em Cordeau e Laporte (2005). O PAB é modelado como um Problema de Roteamento de Veículos com Garagens Múltiplas e Janelas de Tempo (PRVGMJT), onde os navios são vistos como clientes e os berços como garagens. Dessa forma, existem  $m$  veículos, um para cada garagem. Cada veículo inicia e termina sua rota em sua garagem. Esses navios são modelados como vértices em um multi-grafo. Cada garagem é dividida em um vértice de origem ( $o$ ) e um de destino ( $d$ ). No vértice inicial e final as janelas de tempo correspondem ao período de funcionamento dos berços.

O PRVGMJT é especificado como um multi-grafo  $G^k = (V^k, A^k), \forall k \in M$ , tal que  $V^k = N \cup \{o(k), d(k)\}$  e  $A^k \subseteq V^k \times V^k$ . Na Figura 2.7 são apresentados os intervalos de tempo utilizados por cada navio, além das diferentes variáveis utilizadas na formulação do PAB:

- $N$ : conjunto de navios,  $n = |N|$ ;
- $M$ : conjunto de berços,  $m = |M|$ ;
- $t_i^k$ : é o tempo de atendimento do navio  $i$  no berço  $k$ ;
- $a_i$ : horário de chegada do navio  $i$ ;
- $s^k$ : horário de abertura do berço  $k$ ;
- $e^k$ : horário de fechamento do berço  $k$ ;

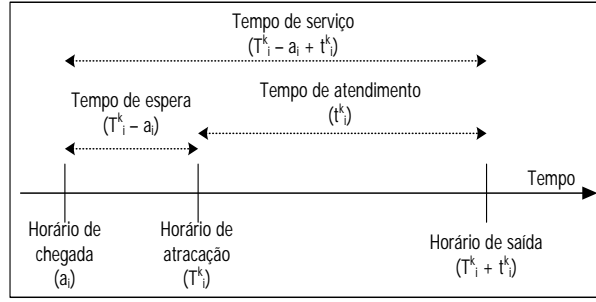


Figura 2.7 - Representação das variáveis de tempo.  
Fonte: Adaptada de Mauri et al. (2008a).

- $b_i$ : horário de término da janela de tempo para o navio  $i$ ;
- $v_i$ : custo do tempo de serviço do navio  $i$ ;
- $x_{ij}^k \in \{0, 1\}, \forall k \in M, \forall (i, j) \in A^k, x_{ij}^k = 1$  se o navio  $j$  é atendido pelo berço  $k$  após o navio  $i$ ;
- $T_i^k \forall k \in M, i \in N$  é o horário em que o navio  $i$  atracou no berço  $k$ ;
- $T_{o(k)}^k \forall k \in M$  é o horário em que o primeiro navio atracou no berço  $k$ ;
- $T_{d(k)}^k \forall k \in M$  é o horário em que o último navio saiu do berço  $k$ ;
- $M_{ij}^k = \max \{b_i + t_i^k - a_j, 0\}, \forall k \in M, \forall (i, j) \in N$ .

Segue a formulação matemática para o PAB proposta por Cordeau e Laporte (2005):

**Minimizar:**

$$\sum_{i \in N} \sum_{k \in M} v_i \left[ T_i^k - a_i + t_i^k \sum_{j \in N \cup \{d(k)\}} x_{ij}^k \right] \quad (2.1)$$

**Sujeito a:**

$$\sum_{k \in M} \sum_{j \in N \cup \{d(k)\}} x_{ij}^k = 1 \quad \forall i \in N \quad (2.2)$$

$$\sum_{j \in N \cup \{d(k)\}} x_{o(k)j}^k = 1 \quad \forall k \in M \quad (2.3)$$

$$\sum_{i \in N \cup \{o(k)\}} x_{id(k)}^k = 1 \quad \forall k \in M \quad (2.4)$$

$$\sum_{j \in N \cup \{d(k)\}} x_{ij}^k - \sum_{j \in N \cup \{o(k)\}} x_{ji}^k = 0 \quad \forall k \in M \forall i \in N \quad (2.5)$$

$$T_i^k + t_i^k - T_j^k \leq (1 - x_{ij}^k) M_{ij}^k \quad \forall k \in M, \forall (i, j) \in A^k \quad (2.6)$$

$$T_i^k \geq a_i \quad \forall k \in M, \forall i \in N \quad (2.7)$$

$$T_i^k + t_i^k - \sum_{j \in N \cup \{d(k)\}} x_{ij}^k \leq b_i \quad \forall k \in M, \forall i \in N \quad (2.8)$$

$$T_{o(k)}^k \geq s^k \quad \forall k \in M \quad (2.9)$$

$$T_{d(k)}^k \leq e^k \quad \forall k \in M \quad (2.10)$$

$$x_{ij}^k \in \{0, 1\} \quad \forall k \in M, \forall (i, j) \in A^k \quad (2.11)$$

A função objetivo 2.1 minimiza a soma dos tempos de serviço, ponderada por um custo associado ( $v_i$ ). As restrições (2.2) indicam que cada navio é atendido por apenas um único berço ( $k$ ). As restrições (2.3) e (2.4) asseguram que para cada berço ( $k$ ) um navio será o primeiro e outro será o último a ser atendido. As restrições (2.5) garantem a conservação do fluxo de atendimento para os navios restantes. As restrições (2.6) indicam a consistência do horário de atracação dos navios. As restrições (2.7) e (2.8) garantem que o horário de atracação seja maior que o tempo de chegada e o horário de saída do navio seja menor que seu tempo limite de atendimento (janela de tempo). As restrições (2.9) e (2.10) garantem o tempo de disponibilidade do berço. Por fim, as restrições (2.11) define o domínio das variáveis de decisão.

Mauri et al. (2008a) propõem a relaxação das restrições (2.7), (2.8), (2.9) e (2.10), de tal forma que, as restrições (2.7) e (2.8) são transferidas para o termo (2.13) da função objetivo e as restrições (2.9) e (2.10) são inseridas no termo (2.14). Além disso, coeficientes de penalizações ( $\omega = [\omega_0, \omega_1, \omega_2]$ ) são adicionados em cada termo da função. Dessa forma, tem-se a seguinte formulação:

**Minimizar:**

$$\omega_0 \sum_{i \in N} \sum_{k \in M} v_i (T_i^k - a_i + t_i^k \sum_{j \in NU\{d(k)\}} x_{ij}^k) + \quad (2.12)$$

$$\omega_1 \sum_{i \in N} \sum_{k \in M} (\max(0, a_i - T_i^k) + (\max(0, T_i^k + t_i^k \sum_{j \in NU\{d(k)\}} x_{ij}^k - b_i))) + \quad (2.13)$$

$$\omega_2 \sum_{k \in M} (\max(0, s^k - T_{o(k)}^k) + \max(0, T_{d(k)}^k + e^k)) \quad (2.14)$$

**Sujeito a:**

$$\sum_{k \in M} \sum_{j \in NU\{d(k)\}} x_{ij}^k = 1 \quad \forall i \in N \quad (2.15)$$

$$\sum_{j \in NU\{d(k)\}} x_{o(k)j}^k = 1 \quad \forall k \in M \quad (2.16)$$

$$\sum_{i \in NU\{o(k)\}} x_{id(k)}^k = 1 \quad \forall k \in M \quad (2.17)$$

$$\sum_{j \in NU\{d(k)\}} x_{ij}^k - \sum_{j \in NU\{o(k)\}} x_{ji}^k = 0 \quad \forall k \in M, \forall i \in N \quad (2.18)$$

$$T_i^k + t_i^k - T_j^k \leq (1 - x_{ij}^k) M_{ij}^k \quad \forall k \in M, \forall (i, j) \in A^k \quad (2.19)$$

$$x_{ij}^k \in \{0, 1\} \quad \forall k \in M, \forall (i, j) \in A^k \quad (2.20)$$

A partir dessa formulação, tem-se uma nova função objetivo (2.12, 2.13 e 2.14). Em cada termo dessa função há um fator de penalidade indicados pelos ômegas. O termo (2.12) mantém na função objetivo o tempo de serviço junto com um custo associado. No termo (2.13) as violações das janelas de tempo dos navios são minimizadas. Por último, o termo (2.14) minimiza as violações nas janelas de tempo dos berços.

Segundo Mauri et al. (2008a), com base nessa nova formulação do PAB e avaliando suas restrições, nota-se que apesar de o problema ter sido modificado para um problema menos árduo (Problema de Roteamento de Veículos com Garagens Múltiplas Sem Janelas de Tempo (PRVGMSJT)), ele poderá apresentar as mesmas soluções do problema original (com janelas de tempo). Por outro lado, o modelo também poderá apresentar soluções inviáveis, mas essas inviabilidades podem ser eliminadas por meio das penalizações inseridas no modelo.

### 2.3.2 Abordagem do PAB-C

Para a modelagem dos berços no caso do PAB-C foram utilizados os passos apresentados em Cordeau e Laporte (2005). Assim, o cais é dividido em comprimentos fixos de diferentes tamanhos, ou seja, em  $(m)$  segmentos (berços), onde cada segmento  $(k)$  tem duas partes, sendo o lado esquerdo  $(k - 1)$  e o lado direito  $(k + 1)$ , onde  $(k \neq 1)$  e  $(k \neq m)$ . A Figura 2.8 apresenta a divisão do cais para o caso do PAB-C. Nota-se que um navio grande poderá usar mais de um berço, podendo emprestar no máximo a metade do espaço de seus vizinhos, tanto do lado esquerdo quanto do lado direito e um navio pequeno poderá ceder seu espaço de sobra.

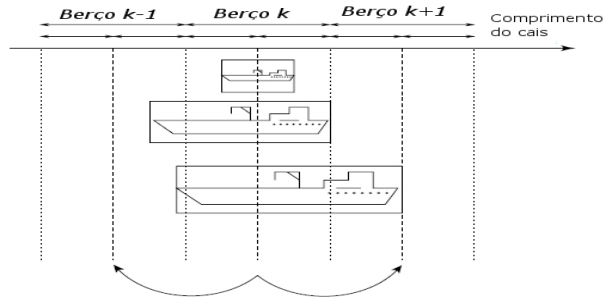


Figura 2.8 - Divisão do cais para o caso do PAB-C.  
 Fonte: Adaptada de Cordeau e Laporte (2005).

O PAB-C foi resolvido buscando minimizar a soma dos tempos de serviço, ou seja, da chegada até a saída do navio no porto. A Figura 2.9 ilustra os intervalos de tempo utilizados por cada navio (MAURI et al., 2011), tal que os navios são considerados retângulos em um plano cartesiano. O eixo horizontal representa o espaço físico do cais e o eixo vertical indica os diferentes tipos de tempo utilizados pelo navio no porto. A seguir são descritas as variáveis utilizadas na figura.

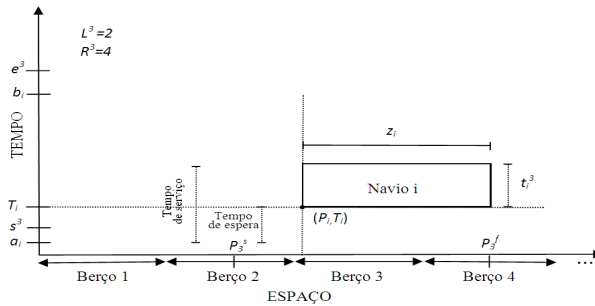


Figura 2.9 - Representação das variáveis de tempo e espaço.  
 Fonte: Adaptada de (MAURI et al., 2011).

- $z_i$ : comprimento do navio  $i$ ;
- $P_i$ : indica a posição para cada navio  $i$ ;
- $P_k^i$ : posição inicial para o berço  $k$ ;
- $P_k^f$ : posição final para o berço  $k$ ;

- $R^k$ : vizinho do lado direito do berço  $k$ ;
- $L^k$ : vizinho do lado esquerdo do berço  $k$ .

Outro fato importante a considerar no modelo de Cordeau e Laporte (2005) são as descontinuidades do cais, representadas nas Figuras 2.10 e 2.11 pelas linhas verdes com asterisco. Essas figuras apresentam as soluções do PAB-D e PAB-C para a primeira instância do conjunto de testes utilizando o algoritmo CS (veja Seção 2.4). As linhas cheias são as separações dos berços e as linhas pontilhadas indicam as divisões dos berços em dois lados (direito e esquerdo). Note que há berços que não se dividem (berços 1, 7, 8, 9 e 13). Os retângulos são os navios e as cores azul e vermelha são para facilitar a visualização. Para o caso discreto do PAB, note que as dimensões espaciais são ignoradas na Figura 2.10, fazendo com que haja sobreposições dos navios.

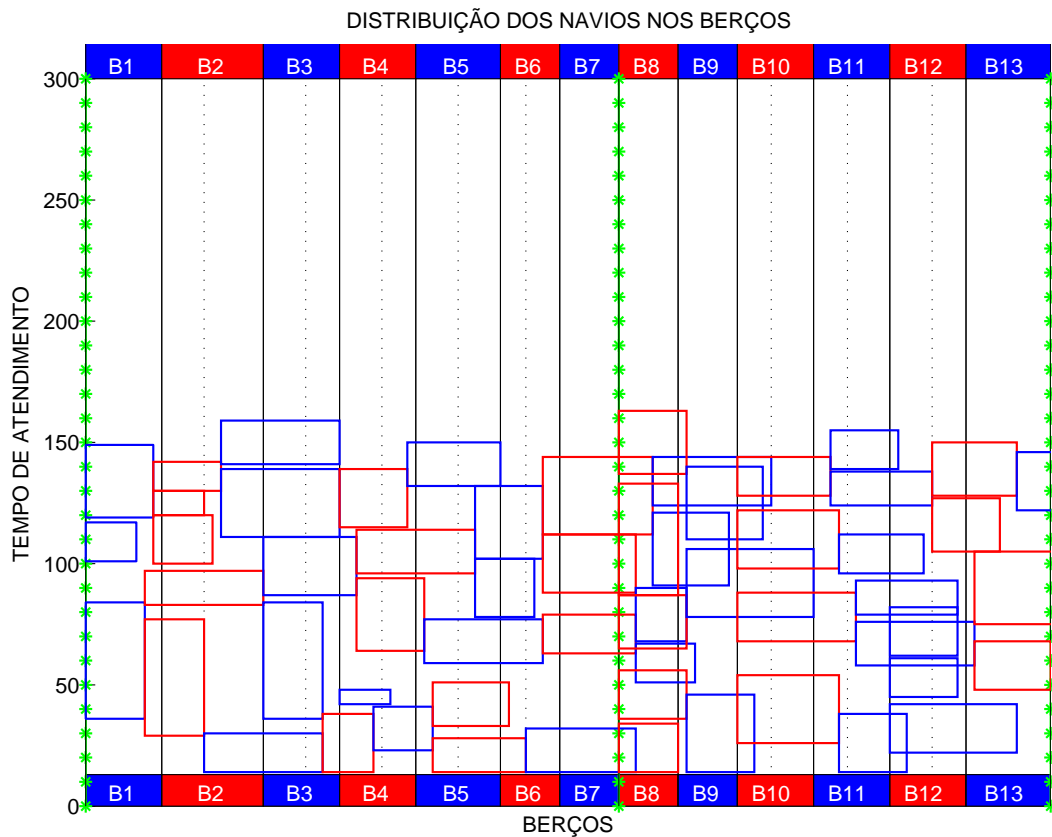


Figura 2.10 - Solução para o PAB-D (instância - i01 - FO = 1409).



Para o caso do PAB-C as sobreposições dos navios devem ser removidas. A Figura 2.11 apresenta a melhor solução conhecida para o PAB-C com valor de custo igual a 1583. Observe que não há sobreposição dos navios, pois essa é uma solução viável. Comparando as Figuras 2.10 e 2.11, o valor de custo para o PAB-C é maior que o valor de custo do PAB-D, o que é esperado, pois o caso discreto é uma relaxação do caso contínuo (CORDEAU; LAPORTE, 2005) .

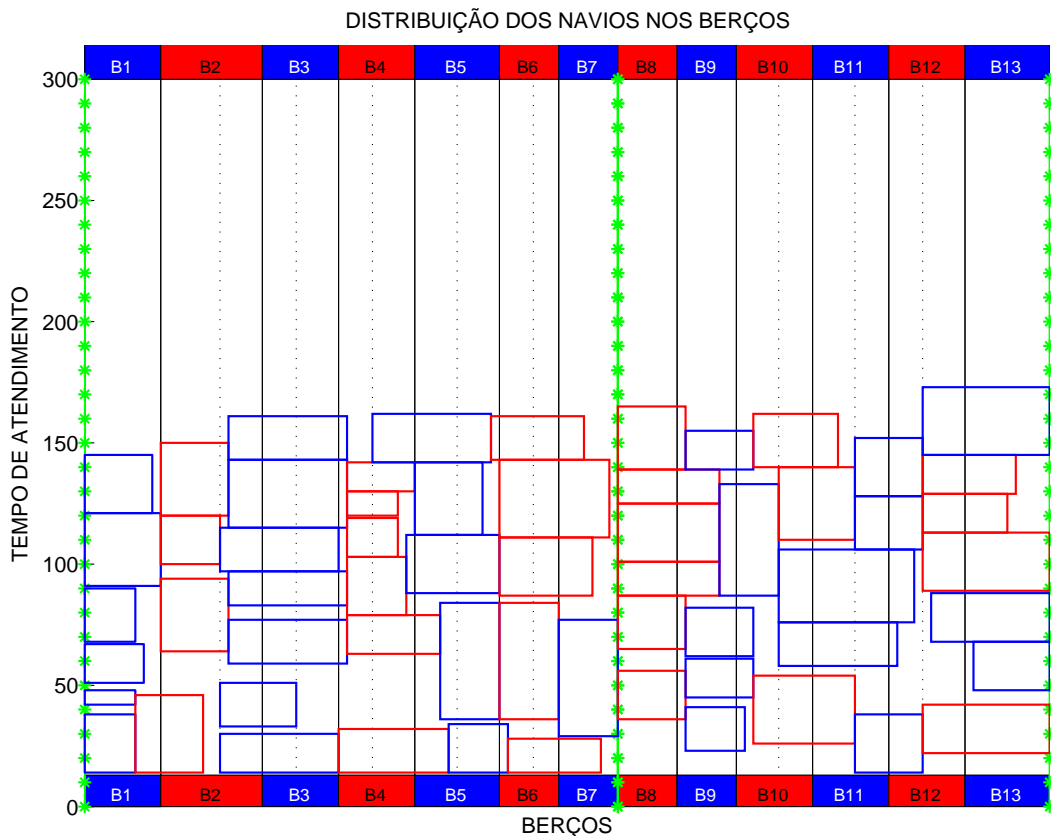


Figura 2.11 - Solução para o PAB-C (instância - i01 - FO= 1583).

## 2.4 *Clustering Search* aplicado ao PAB

Segundo Chaves (2009), o *Clustering Search* (CS) é um método iterativo que procura dividir o espaço de busca e localizar regiões promissoras por meio do enquadramento dessas em *clusters*. Um *cluster* pode ser definido por três atributos  $C = \{c, v, r\}$ . O centro  $c_i$  é uma solução que representa o *cluster*  $i$ , e identifica a sua localização dentro do espaço de busca. O volume  $v_i$  é a quantidade de soluções agrupadas no *cluster*  $i$ . Um *cluster* se torna promissor quando o volume atingir um certo limitante

$\lambda$ . O índice de ineficácia  $r_i$  é uma variável de controle para identificar se a busca local está ou não melhorando o centro do *cluster*  $i$ . O valor de  $r_i$  indica o número de vezes consecutivas que a busca local foi aplicada no *cluster*  $i$  e não melhorou a solução. Esse atributo evita que a busca local seja executada em regiões ruins ou regiões que já tenham sido suficientemente exploradas por mais de  $r_{max}$  vezes.

O CS é formado basicamente por três componentes principais: uma meta-heurística geradora de soluções, um processo de agrupamento e uma heurística de busca local. A cada iteração do CS, uma solução  $S$  é gerada pela meta-heurística e enviada para o processo de agrupamento. Essa solução é então agrupada no *cluster* mais similar  $C_j$  e o centro desse *cluster*  $c_i$  é atualizado com informações contidas na nova solução agrupada, fazendo com que o centro se desloque no espaço de busca.

Em seguida, é analisado o volume  $\nu_i$  do *cluster*  $e$ , caso esse volume atinja um limitante  $\lambda$  ( $\nu_i \geq \lambda$ ), percebe-se que algum padrão de solução está sendo predominantemente gerado pela meta-heurística. Portanto, esse *cluster* pode estar em uma região de busca promissora.

Por fim, é analisado o índice de ineficácia  $r_j$ , ou seja, caso a heurística de busca local não melhore a solução por  $r_{max}$  vezes consecutivas ( $r_j \geq r_{max}$ ), é aplicada uma perturbação aleatória no centro  $c_j$ , objetivando escapar de um possível ótimo local. Por outro lado, se  $r_j < r_{max}$ , a heurística de busca local é aplicada no centro  $c_j$  analisando a vizinhança do *cluster*. Encerrado esse processo, retorna-se para a meta-heurística que irá gerar uma nova solução. O critério de parada do CS é geralmente definido pela meta-heurística escolhida. A Figura 2.12 apresenta o fluxograma de execução do CS. Mais detalhes sobre o CS são apresentados em Oliveira (2004) e Chaves (2009).

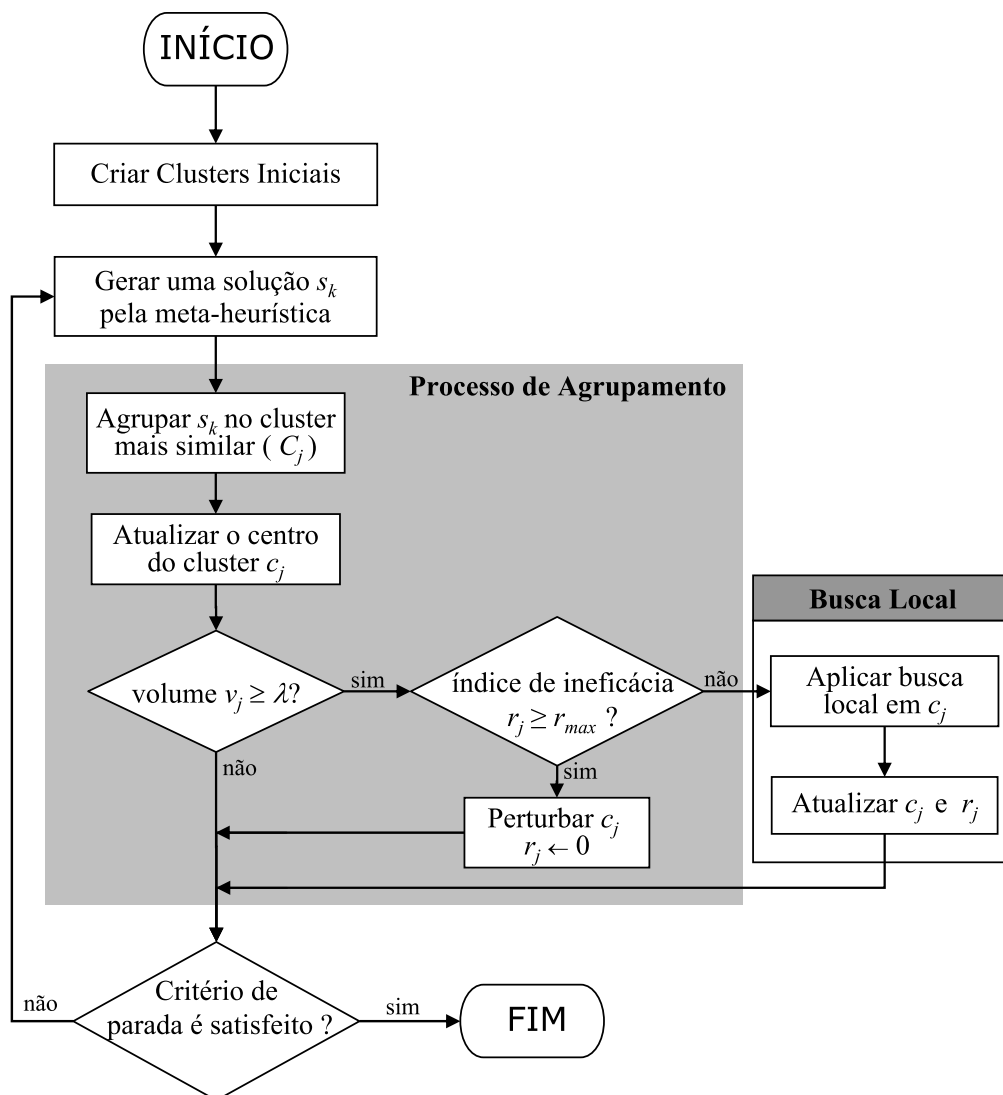


Figura 2.12 - Fluxograma do CS.

Fonte: Adaptada de Chaves (2009).

Para que as soluções possam ser agrupadas em *clusters* similares, deve-se definir uma métrica para medir a distância entre duas soluções. Para isso, utiliza-se neste trabalho a distância de Hamming (HAMMING, 1950), que é definida como sendo o número de arcos diferentes entre as soluções. Assim, seja  $d_{Ham}(i, j)$  uma função de medida de distância entre as soluções  $i$  e  $j$ , onde  $d_{Ham}(i, j)$  é um número positivo que deverá aumentar de acordo com o aumento da distância entre as duas soluções. Considerando a Solução 1 apresentada na Figura 2.13 (e o ponto 0 como sendo o início e o fim de cada sequência de atendimento), pode-se definir os seguintes arcos:

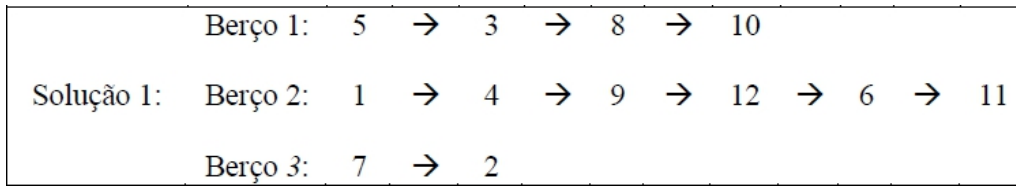


Figura 2.13 - Exemplo de uma solução para o PAB.

**Solução 1:**

(0,5), (5,3), (3,8), (8,10), (10,0), (0,1), (1,4), (4,9), (9,12), (12,6), (6,11), (11,0), (0,7), (7,2), (2,0).

Considerando agora uma Solução 2, tal que apenas foram invertidas três posições:

**Solução 2:**

(0,5), (5,8), (8,3), (3,10), (10,0), (0,1), (1,4), (4,9), (9,12), (12,6), (6,11), (11,0), (0,7), (7,2), (2,0).

A distância entre as soluções 1 e 2 é:  $d_{\text{Ham}}(\mathbf{S1}, \mathbf{S2}) = 3$  (arcos em azul).

Seguindo o fluxograma do CS (Figura 2.12), são criados então os *clusters* iniciais. Assim, para cada *cluster* é criada uma solução por meio das heurísticas de distribuição, programação e atualização (caso do PAB-C).

A heurística de distribuição (Figura 2.14) é responsável pela atribuição dos navios aos berços. Ela é iniciada com a criação de um conjunto de  $m$  berços, além de uma lista de  $n$  navios. Em seguida, os navios são ordenados nos berços de acordo com sua ordem de chegada ao porto. Caso o berço não possa atender o navio, outro berço é designado. Este procedimento assegura que cada navio será atribuído a um berço que deve ser capaz de atendê-lo, ou seja, o comprimento do cais deve ser suficiente para recebê-lo e os equipamentos do cais devem estar prontos para operar seu tipo de carga. Este procedimento não garante que o tempo de atracação ( $T_i$ ) e a posição ( $P_i$ ) para o navio  $i$  não apresentarão sobreposição das dimensões de tempo e espaço.

Após a aplicação da heurística de distribuição, é preciso definir o tempo de atracação e a posição para todos os navios. A heurística de programação (Figura 2.15) é responsável por determinar o horário de atendimento dos navios no porto e a posi-

ção (caso PAB-C), e nela também são calculados os termos das funções referentes as Equações (2.12), (2.13), (2.14). O tempo de atracação e a posição dos navios designados para um específico berço é definidos inicialmente de acordo com a Figura 2.15. Neste momento, os tempos de atracação são iguais aos tempos de chegada dos navios, se o berço está disponível. Inicialmente as posições de atracação são iguais a posição inicial do cais.

- |   |
|---|
| <ol style="list-style-type: none"> <li>1. <u>PARA</u> (cada berço pertencente a <math>c_j</math>) FAÇA</li> <li>2. <u>ENQUANTO</u> (melhorar a solução) FAÇA</li> <li>3. <u>PARA</u> (todos os navios de <math>c_j</math>);</li> <li>4. <u>INSERIR</u> (o navio <math>p</math> em todas as posições do berço);</li> <li>5. <u>ARMAZENAR</u> (os navios e as posições que resultam na melhor solução);</li> <li>6. <u>FIM-PARA</u>;</li> <li>7. <u>INSERIR</u> (os navios nas melhores posições armazenadas);</li> <li>8. <u>FIM-ENQUANTO</u>;</li> <li>9. <u>FIM-PARA</u>;</li> </ol> |
|---|

Figura 2.14 - Heurística de distribuição.

Fonte: Adaptada de Mauri (2008).

- |  |   |
|--|---|
| <ol style="list-style-type: none"> <li>1. <u>PARA</u> (cada berço <math>k, k = 1, 2, \dots, m</math>) FAÇA</li> <li>2. <u>PARA</u> (cada navio <math>i</math> atribuído a <math>k</math>) FAÇA</li> <li>3. <math display="block">T_i^k = \begin{cases} \max(a_i, s^k), &amp; i = 1 \\ \max(a_i, T_{i-1}^k + t_{i-1}^k), &amp; i &gt; 1 \end{cases}</math></li> <li>4. <u>FIM-PARA</u>;</li> <li>5. <u>FIM-PARA</u>;</li> <li>6. <u>CALCULAR</u> (a função objetivo para a solução atual);</li> </ol> | <ol style="list-style-type: none"> <li>1. <u>PARA</u> (cada berço <math>k, k = 1, 2, \dots, m</math>) FAÇA</li> <li>2. <u>PARA</u> (cada navio <math>i</math> atribuído a <math>k</math>) FAÇA</li> <li>3. <math display="block">T_i^k = \begin{cases} \max(a_i, s^k), &amp; i = 1 \\ \max(a_i, T_{i-1}^k + t_{i-1}^k), &amp; i &gt; 1 \end{cases}</math></li> <li>4. <math>P_i = P_k^i</math></li> <li>5. <u>FIM-PARA</u>;</li> <li>6. <u>FIM-PARA</u>;</li> <li>7. <u>CALCULAR</u> (a função objetivo para a solução atual);</li> </ol> |
|--|---|

Figura 2.15 - Heurística de programação para o PAB-D e PAB-C

Fonte: Adaptada de Mauri (2008).

A heurística de atualização é usada para o caso do PAB-C (Figura 2.16), nesta heurística a distribuição espacial dos navios nos berços são atualizadas e melhoradas. Este procedimento atualiza os horários de atracação e posições considerando uma ideia simples: se alguma sobreposição for detectada em algum navio, o seu tempo de atracação é atrasado até que a sobreposição seja eliminada. Se uma vaga não tem

vizinho do lado esquerdo, todos os navios que lhe são atribuídas tem à posição de atracação igual a posição de início da vaga, ou seja, todos os navios são alinhados à esquerda. Se uma vaga não tem vizinho do lado direito, todos os navios que lhe são atribuídas tem a posição de atracação igual a posição final do cais menos o comprimento do navio, ou seja, todos os navios são alinhados à direita. Finalmente, se uma vaga tem vizinhos de ambos os lados, tenta-se encaixar os navios entre eles.

1.	<u>ENTRADA:</u> berço $k$ ,
2.	<u>PARA</u> (cada navio $i$ atribuído a $k$ ) <u>FAÇA</u>
3.	<u>SE</u> $L^k=0$ então
4.	$P_i = P_k^s$
5.	<u>PARA</u> (cada navio $j$ atribuído a $k+1$ ) <u>FAÇA</u>
6.	<u>SE</u> $i$ sobrepõem $j$ <u>ENTÃO</u>
7.	$T_i = \max(T_i, T_j+t_j^{k+1})$
8.	<u>FIM-SE;</u>
9.	<u>FIM-PARA;</u>
10.	<u>SENÃO</u>
11.	<u>SE</u> $R^k=0$ então
12.	$P_i = P_k^f - z_i$
13.	<u>PARA</u> (cada navio $i$ atribuído a $k-1$ ) <u>FAÇA</u>
14.	<u>SE</u> $i$ sobrepõem $j$ <u>ENTÃO</u>
15.	$T_i = \max(T_i, T_j+t_j^{k-1})$
16.	<u>FIM-SE;</u>
17.	<u>FIM-PARA;</u>
18.	<u>SENÃO</u>
19.	$P_i = P_k^s$
20.	<u>ENQUANTO</u> $\exists j$ atribuído para $k-1$ sobrepondo $i$ <u>FAÇA</u>
21.	<u>SE</u> $(P_j + z_j \geq P_k^s)$ e $(P_j + z_j + z_i \leq P_k^f)$ <u>ENTÃO</u>
22.	$P_i = P_j + z_j$
23.	<u>SENÃO</u>
24.	$T_i = \max(T_i, T_j+t_j^{k-1})$
25.	$P_i = P_k^f - z_i$
26.	<u>FIM-SE;</u>
27.	<u>FIM-ENQUANTO;</u>
28.	<u>ENQUANTO</u> $\exists j$ atribuído para $k+1$ sobrepondo $i$ <u>FAÇA</u>
29.	$T_i = \max(T_i, T_j+t_j^{k-1})$
30.	<u>ENQUANTO</u> $\exists l$ atribuído para $k-1$ sobrepondo $i$ <u>FAÇA</u>
31.	$T_i = \max(T_i, T_l+t_l^{k-1})$
32.	<u>FIM-ENQUANTO;</u>
33.	<u>FIM-ENQUANTO;</u>
34.	<u>FIM-SE;</u>
35.	<u>FIM-SE;</u>
36.	<u>FIM-PARA;</u>

Figura 2.16 - Heurística de Atualização.

Fonte: Adaptada de Mauri et al. (2011).

A partir de então, o SA, baseado no proposto por Mauri et al. (2008a) é executado e, a cada temperatura, a solução corrente (não a melhor) é enviada ao CS. Os parâmetros do SA são:

- $\alpha$ : é a razão de resfriamento onde  $0 < \alpha < 1$ ;
- $SA_{max}$ : é o número de iterações para cada temperatura;
- $T$ : é a temperatura corrente;
- $T_0$ : é a temperatura inicial;
- $T_C$ : é a temperatura de congelamento;
- $S$ : é a solução corrente;
- $S'$ : é a solução vizinha da solução corrente;
- $S^*$ : é a melhor solução até o momento;

A Figura 2.17 apresenta um pseudocódigo do SA implementado. Pode-se notar que o CS é chamado na linha 20 desse algoritmo, ou seja, a cada temperatura.

1. ENTRADA ( $\alpha$ ,  $S_{\max}$ ,  $T_0$  e  $T_C$ ) FAÇA
2. GERAR (uma solução  $S$  por meio da heurística de distribuição);
3. AVALIAR (a solução  $S$  por meio da heurística de programação);
4.  $S^* \leftarrow S$ ; {Melhor solução obtida até então}
5.  $\text{IterT} \leftarrow 0$ ; {Número de iterações na temperatura  $T$ }
6.  $T \leftarrow T_0$ ; {Temperatura corrente}
7. ENQUANTO ( $T > T_C$ ) FAÇA
8. ENQUANTO ( $\text{IterT} < S_{\max}$ ) FAÇA
9.  $\text{IterT} \leftarrow \text{IterT} + 1$ ;
10. GERAR (um vizinho qualquer  $S'$  por meio de um dos mov. de troca);
11. APLICAR (a heurística de programação em todos os berços de  $S'$ );
12.  $\Delta \leftarrow f(S') - f(S)$ ;
13. SE ( $\Delta < 0$ )  $S \leftarrow S'$ ;
14. SE ( $f(S') < f(S^*)$ )  $S^* \leftarrow S'$ ; FIM-SE;
15. SENÃO
16. TOMAR ( $x \in [0,1]$ );
17. SE ( $x < e^{-\Delta/T}$ )  $S \leftarrow S'$ ; FIM-SE;
18. FIM-SE;
19. FIM-ENQUANTO;
20. **EXECUTAR-CS (solução corrente  $S$ )**;
21.  $T \leftarrow \alpha * T$ ;  $\text{IterT} \leftarrow 0$ ;
22. FIM-ENQUANTO;
23.  $S \leftarrow S^*$ ;
24. RETORNAR ( $S$ );

Figura 2.17 - Algoritmo *Simulated Annealing* utilizado no CS.

Como estrutura de vizinhança no SA (linha 10) foram utilizados três diferentes movimentos de troca: *Reordenar navios*, *Realocar navio* e *Trocar navios* (Figuras 2.18, 2.19 e 2.20). Assim, como na geração da solução inicial, esses movimentos garantem que cada navio seja atribuído apenas a berços que possam atendê-los. O movimento *reordenar navios* consiste em escolher um navio em qualquer um dos berços e trocá-lo de posição. Isso é apresentado na Figura 2.18, na qual o navio 2 na linha (a), uma vez tendo um lugar selecionado na linha (b), troca de posição com o navio 1 na linha (c). O movimento *realocar navios* (Figura 2.19) consiste em: (a) selecionar dois berços e em um dos berços escolher um navio; (b) acrescentar o navio selecionado no outro berço; e (c) os navios são atendidos na ordem de sua chegada. Por fim, no movimento *trocar navios*, seleciona-se dois berços e um navio



de cada berço (linha (a)) e troca-os de berços (linha (b)), mas a posição de chegada dos navios é mantida (linha (c)).

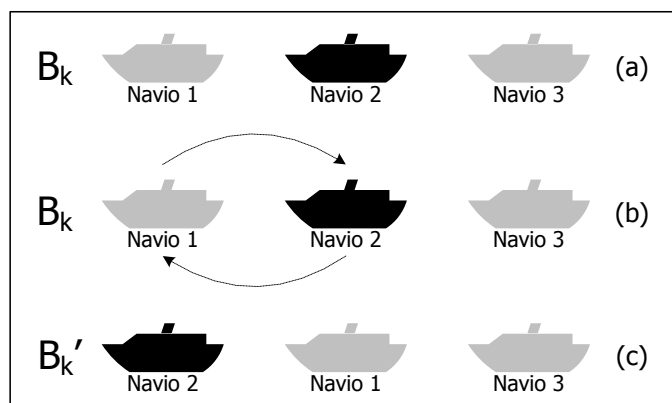


Figura 2.18 - Movimento de re-ordenar navios.  
Fonte: Adaptada de Mauri (2008).

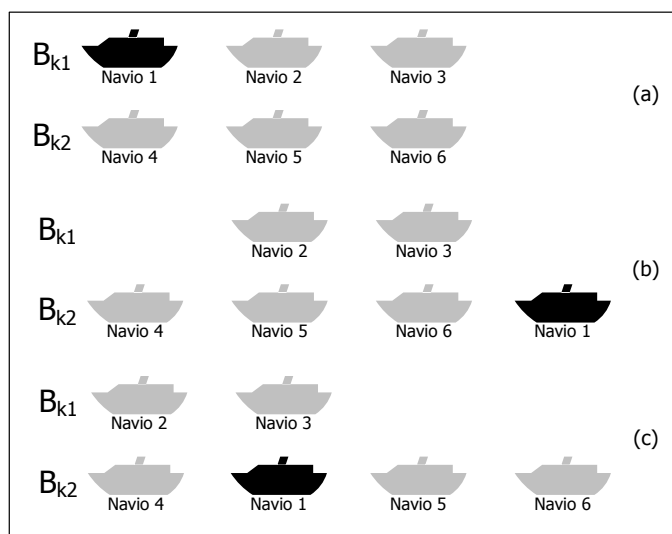


Figura 2.19 - Movimento de re-alocar navios.  
Fonte: Adaptada de Mauri (2008).

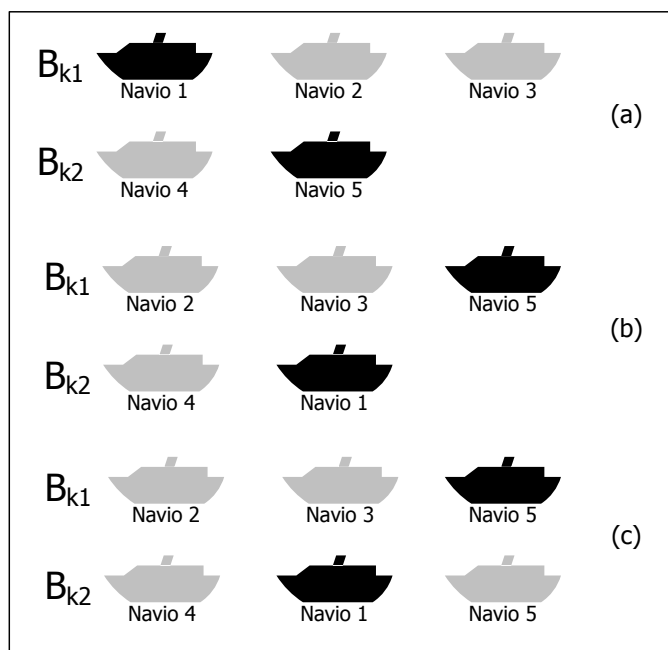


Figura 2.20 - Movimento de trocar navios.  
 Fonte: Adaptada de Mauri (2008).

Após a execução de cada um desses movimentos, a heurística de atualização é aplicada para eliminar as sobreposições e recalcular o valor da função objetivo da nova solução. Cada solução vizinha no SA (linha 10) é gerada por apenas um desses movimentos, sendo a sua escolha feita de forma aleatória e uniformemente distribuída, possibilitando assim uma boa diversidade entre as soluções intermediárias geradas, e consequentemente uma boa exploração do espaço de soluções. Mais detalhes desses movimentos são apresentados em Mauri et al. (2008a), Mauri et al. (2008b) e Mauri e Lorena (2009).

O CS é implementado de acordo com o apresentado em Chaves (2009). Vale destacar que antes mesmo da execução do SA, as soluções “centro” de *clusters* já foram criadas (como descrito anteriormente). Ao final da execução do SA, ou seja, do método CS-SA como um todo, a melhor solução encontrada é tomada como solução final para o problema. O algoritmo EXECUTAR-CS é apresentado na Figura 2.21.

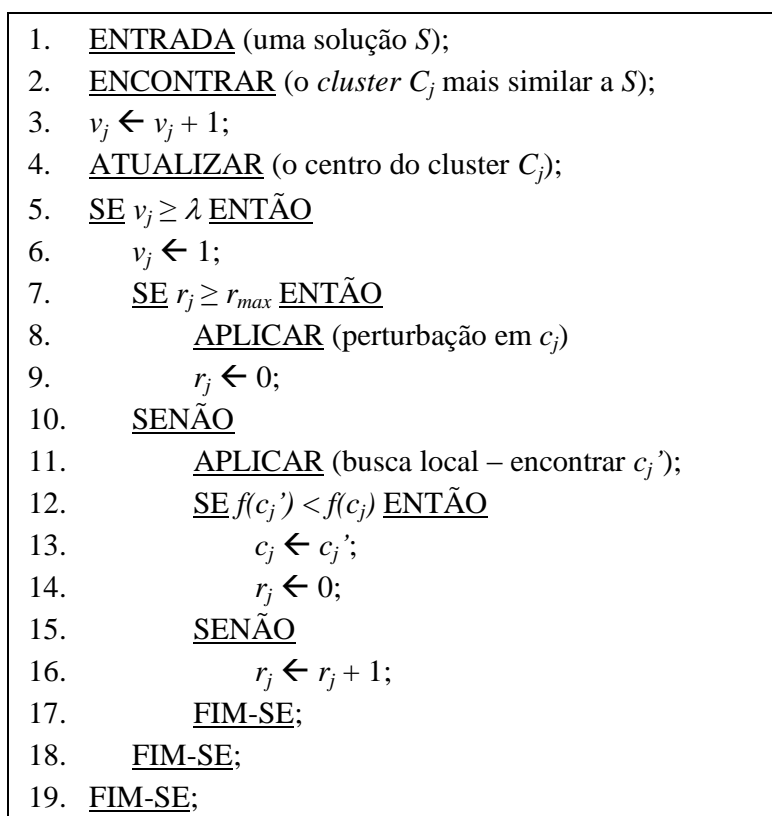


Figura 2.21 - Algoritmo EXECUTAR-CS.

Como mencionado anteriormente, a determinação do *cluster* mais similar (linha 2) é dada pela menor distância de *Hamming*. A atualização do centro do *cluster* (linha 4) é dada pela execução do *Path-Relinking* entre a solução dada  $S$  e a solução  $c_j$  centro do *cluster*  $C_j$  (Figura 2.22).

A ideia deste algoritmo é simples, e consiste em executar os movimentos necessários para “transformar” a solução  $S'$  (cópia de  $S$ ) na solução  $c_j$ . A partir destes movimentos, a melhor solução encontrada é tomada como novo centro do *cluster*  $C_j$ .

1. ENTRADA ( $S'$  e  $c_j$ )
2. PARA (cada navio  $i$ ,  $i = 1, \dots, n$ ) FAÇA
3.     SE (berço que atende o navio  $i$  em  $S' \neq$  berço que atende o navio  $i$  em  $c_j$ );
4.     REMOVER (o navio  $i$  de seu respectivo berço em  $S'$ );
5.     INSERIR (o navio  $i$  no berço em  $S'$  correspondente ao berço em  $c_j$ );
6.     SE ( $f(S') < f(c_j)$ ) ENTÃO
7.          $c_j \leftarrow S'$ ;
8.     SENÃO
9.         SE ( $f(S') = f(c_j)$ ) ENTÃO
10.         PARE;
11.     FIM-SE;
12. FIM-SE;
13. FIM-PARA;

Figura 2.22 - *Path-relinking* utilizado na atualização dos centros de *cluster*.

A perturbação apresentada na linha 8 do EXECUTAR-CS (Figura 2.21) é dada por uma simples aplicação do movimento trocar navios (Figura 2.20). Por fim, a busca local (linha 11 do EXECUTAR-CS) utilizada para intensificar a busca em *clusters* promissores é apresentada na Figura 2.23. É interessante destacar que a busca local é aplicada por berço, evitando assim um alto tempo de processamento.

1. PARA (cada berço pertencente a  $c_j$ ) FAÇA
2.     ENQUANTO (melhorar a solução) FAÇA
3.         PARA (todos os navios de  $c_j$ ) FAÇA
4.             INSERIR (o navio  $p$  em todas as posições do berço);
5.             ARMAZENAR (os navios e as posições que resultam na melhor solução);
6.         FIM-PARA;
7.             INSERIR (os navios nas melhores posições armazenadas);
8.         FIM-ENQUANTO;
9. FIM-PARA;

Figura 2.23 - Busca local.

## 2.5 Experimentos computacionais para o PAB

Para a realização dos experimentos computacionais para o caso do PAB-D e PAB-C foram utilizadas 30 instâncias distintas, cada uma com 60 navios e 13 berços. Essas instâncias são baseadas em dados do porto de Gioia Tauro (Itália), e foram geradas aleatoriamente por Cordeau e Laporte (2005). Para o caso do PAB-D os experimentos foram realizados em um *PC* com processador *AMD Athlon<sup>TM</sup> 64 3500* de 2.2 GHz e 1GB de memória RAM, mesma máquina utilizada pelo *Simulated Annealing* com *Re-annealing* - SA+RA (MAURI et al., 2008a) e pela geração de colunas - PTA/LP (MAURI et al., 2008b). Para o caso do PAB-C os experimentos foram realizados em um PC com processador intel core 2 duo de 1.66 GHz e 2 GB de memória RAM. A implementação foi desenvolvida na linguagem C++. Os parâmetros utilizados pelo CS, em todos os experimentos, foram:  $T_0 = 20000$ ,  $\alpha = 0.975$ ,  $T_c = 0.01$ ,  $SA_{max} = 1000$ ,  $\lambda = 7$ ,  $r_{max} = 3$ , e o número de *clusters* = 10. As penalizações utilizadas em ambos os casos foram  $\omega = [1, 10, 10]$ . Estes são os parâmetros que obtiveram as melhores soluções de um total de 5 testes para cada instância.

## 2.6 Caso discreto do PAB

Para o caso discreto do PAB o CS utilizou um tempo computacional médio de 12,79 segundos por instância. Além disso, o desvio entre a média das soluções obtidas em relação à melhor solução para cada instância foi de 0,04%.

A Tabela ?? apresenta uma comparação direta dos resultados obtidos com os apresentados nos trabalhos mais recentes encontrados na literatura. Nessa tabela, são apresentadas as melhores soluções obtidas pelo CS e o tempo médio (considerando os 5 testes para cada instância). As colunas FO (função objetivo) indicam os valores das soluções obtidas. Já as últimas 5 colunas indicam as melhoras nas soluções obtidas pelo CS proposto em relação aos demais métodos apresentados nessa mesma tabela. Os tempos computacionais apresentados estão em segundos.

O CS proposto encontrou as soluções ótimas (provadas por Buhrkal et al. (2009)) em todos os casos e, além disso, estas soluções foram obtidas em um tempo computacional expressivamente inferior aos demais métodos.

Tabela 2.1 - Comparação dos resultados obtidos (CS) para o PAB-D.

Inst.	CPLEX <sup>1</sup>		BT <sup>2</sup>		SA+RA <sup>3</sup>		PTA/LP <sup>4</sup>		GSPP <sup>5</sup>		CS		MELHORAS (%)					
	FO	Tempo	FO	Tempo	FO	Tempo	FO	Tempo	FO	Tempo	FO	Tempo	CPLEX	BT	SA+RA	PTA	LP	GSPP
i01	-	3600,00	1415	-	1409	53,12	1409	74,61	1409	17,92	1409	12,47	-	0,42	0,00	0,00	0,00	0,00
i02	2606	3600,00	1263	-	1261	58,94	1261	60,75	1261	15,77	1261	12,59	51,61	0,16	0,00	0,00	0,00	0,00
i03	2565	3600,00	1139	-	1129	54,03	1129	135,45	1129	13,54	1129	12,64	55,98	0,88	0,00	0,00	0,00	0,00
i04	4353	3600,00	1303	-	1302	67,33	1302	110,17	1302	14,48	1302	12,59	70,09	0,08	0,00	0,00	0,00	0,00
i05	2672	3600,00	1208	-	1207	55,38	1207	124,70	1207	17,21	1207	12,68	54,83	0,08	0,00	0,00	0,00	0,00
i06	-	3600,00	1262	-	1261	53,88	1261	78,34	1261	13,85	1261	12,56	-	0,08	0,00	0,00	0,00	0,00
i07	2887	3600,00	1279	-	1279	60,52	1279	114,20	1279	14,60	1279	12,63	55,70	0,00	0,00	0,00	0,00	0,00
i08	5177	3600,00	1299	-	1299	61,45	1299	57,06	1299	14,21	1299	12,57	74,91	0,00	0,00	0,00	0,00	0,00
i09	-	3600,00	1444	-	1444	57,91	1444	96,47	1444	16,51	1444	12,58	-	0,00	0,00	0,00	0,00	0,00
i10	-	3600,00	1213	-	1213	68,95	1213	99,41	1213	14,16	1213	12,61	-	0,00	0,00	0,00	0,00	0,00
i11	-	3600,00	1378	-	1368	76,77	1368	99,34	1368	14,13	1368	12,58	-	0,73	0,00	0,07	0,00	0,00
i12	3206	3600,00	1325	-	1325	62,84	1325	80,69	1325	15,60	1325	12,56	58,67	0,00	0,00	0,00	0,00	0,00
i13	-	3600,00	1360	-	1360	68,19	1360	89,94	1360	13,87	1360	12,61	-	0,00	0,00	0,00	0,00	0,00
i14	-	3600,00	1233	-	1233	75,06	1233	73,95	1233	15,60	1233	12,67	-	0,00	0,00	0,00	0,00	0,00
i15	4672	3600,00	1295	-	1295	54,55	1295	74,19	1295	13,52	1295	13,80	72,28	0,00	0,00	0,00	0,00	0,00
i16	4320	3600,00	1375	-	1364	63,91	1364	170,36	1364	13,68	1364	14,46	68,43	0,80	0,00	0,07	0,00	0,00
i17	-	3600,00	1283	-	1283	56,28	1283	46,58	1283	13,37	1283	13,73	-	0,00	0,00	0,00	0,00	0,00
i18	3681	3600,00	1346	-	1345	53,98	1345	84,02	1345	13,51	1345	12,72	63,46	0,07	0,00	0,00	0,00	0,00
i19	2400	3600,00	1370	-	1370	52,83	1370	123,19	1367	14,59	1367	13,39	43,04	0,22	0,22	0,00	0,00	0,00
i20	-	3600,00	1328	-	1328	53,38	1328	82,30	1328	16,64	1328	12,82	-	0,00	0,00	0,00	0,00	0,00
i21	-	3600,00	1346	-	1341	53,52	1341	108,08	1341	13,37	1341	12,68	-	0,37	0,00	0,00	0,00	0,00
i22	3489	3600,00	1332	-	1326	57,97	1326	105,38	1326	15,24	1326	12,62	61,99	0,45	0,00	0,00	0,00	0,00
i23	-	3600,00	1266	-	1266	53,75	1266	43,72	1266	13,65	1266	12,62	-	0,00	0,00	0,00	0,00	0,00
i24	4867	3600,00	1261	-	1260	54,09	1260	78,91	1260	15,58	1260	12,64	74,11	0,08	0,00	0,00	0,00	0,00
i25	1993	3600,00	1379	-	1377	53,56	1377	96,58	1376	15,80	1376	12,62	30,96	0,22	0,07	0,00	0,00	0,00
i26	2520	3600,00	1330	-	1318	57,34	1318	101,11	1318	15,38	1318	12,62	47,70	0,90	0,00	0,00	0,00	0,00
i27	3209	3600,00	1261	-	1261	69,98	1261	82,86	1261	15,52	1261	12,64	60,70	0,00	0,00	0,00	0,00	0,00
i28	-	3600,00	1365	-	1360	58,47	1360	52,91	1359	16,22	1359	12,71	-	0,44	0,07	0,00	0,00	0,00
i29	4809	3600,00	1282	-	1280	69,09	1280	203,36	1280	15,30	1280	12,62	73,38	0,16	0,00	0,00	0,00	0,00
i30	-	3600,00	1351	-	1344	70,67	1344	71,02	1344	16,52	1344	12,58	-	0,52	0,00	0,00	0,00	0,00
<b>Média</b>		3600,00		120,00		60,26		93,99		14,98		12,79						

<sup>3</sup>Mauri et al. (2008a)

<sup>5</sup>Buhrkal et al. (2009)

<sup>1</sup>Ilog (2006)

<sup>2</sup>Cordeau et al. (2005)

<sup>4</sup>Mauri et al. (2008b)

## 2.7 Caso contínuo do PAB

Esta seção apresenta os resultados computacionais do algoritmo CS para o caso do PAB-C em comparação com a *busca tabu* - BT (CORDEAU; LAPORTE, 2005) e Algoritmo Memético - AM (MAURI et al., 2011) relatado em trabalhos anteriores. O algoritmo CS foi executado 10 vezes para cada instância com critério de parada de 120 segundos, mesmo tempo utilizado pela BT e AM. Uma análise estatística baseada na média, desvio, diferenças e melhorias entre todos os métodos são apresentados nas próximas tabelas.

A Tabela 2.2 mostra uma comparação entre o BT, o AM e o CS proposto neste trabalho. As três primeiras colunas são as instâncias utilizadas, o valor ótimo para o PAB-D (OTM DISC) e a Melhor Solução Conhecida (MSC) entre os três métodos (BT, AM e CS). Os resultados de cada método são apresentados nas colunas restantes (de 4 a 17), tal que: MLR FO e Média FO são as melhores e a média para as soluções; Média MSC é o tempo médio para encontrar a melhor solução conhecida relatadas em Cordeau e Laporte (2005), e Média MLR é o tempo médio para encontrar a melhor solução (MLR FO). Os tempos de execução são expressos em segundos.

As colunas DESV e DESV MSC estão relacionadas aos desvios entre a média e as melhores soluções. DESV é calculado entre a média (Média FO) sobre a melhor solução (MLR FO) para cada método ( $DESV = 100 \times (Média\ FO - MLR\ FO) / MLR\ FO$ ) e DESV MSC é calculado entre a melhor solução (MLR FO) sobre a melhor solução conhecida por todos os métodos ( $DESV.MSC = 100 \times (MLRFO - MSC^*) / MSC^*$ ).

De forma geral, os valores das soluções na Tabela 2.2 para o PAB-C são maiores que os do PAB-D, o que é esperado, pois o caso discreto é uma relaxação do caso contínuo (CORDEAU; LAPORTE, 2005).

Quanto à comparação dos métodos, observa-se que em todas as instâncias o CS conseguiu os melhores resultados, além disso, comparando o CS com o AM observa-se que em média o CS foi mais rápido em encontrar as melhores soluções (Média MLR) e obteve os menores valores para os desvios com médias totais de  $DESV = 1,17\%$  e  $DESV.MSC = 0.00\%$ .

A Tabela 2.3 apresenta o cruzamento entre todos os métodos mostrando as melhorias obtidas em todas as instâncias. Pela tabela nota-se que tanto o AM quanto o CS



obtiveram melhorias em relação à BT, sendo que o CS obteve melhoras com média total de 1,7 comparado com o AM (CSxAM) e média total de 6,8 em relação ao BT (CSxBT). A Tabela 2.3 apresenta as diferenças entre o PAB-D e o PAB-C resolvido com a BT, o AM e o CS. O CS conseguiu os menores índices de diferença para todas as instâncias com média geral de 8%. É interessante notar que as soluções obtidas pelo algoritmo CS foram próximas dos valores do PAB-D, indicando bons limites superiores para o PAB-C.

Tabela 2.2 - Comparação entre os resultados do BT, AM and CS para o PAB-C.

Inst.	OTM DISC	BT			AM						CS					
		MLR FO	DESV MSC	MSC	MLR FO	Média FO	Média MSC	Média MLR	DESV (%)	DESV MSC	MLR FO	Média FO	Média MSC	Média MLR	DESV (%)	DESV MSC
i01	1409	1706	7,77	1583	1613	1666,50	37,49	85,41	3,32	1,90	1583	1599,10	24,96	67,76	1,02	0,00
i02	1261	1355	3,04	1315	1326	1347,80	51,91	75,85	1,64	0,84	1315	1321,80	26,90	68,92	0,52	0,00
i03	1129	1286	6,55	1207	1234	1261,90	20,58	79,99	2,26	2,24	1207	1226,50	25,78	86,28	1,62	0,00
i04	1302	1440	4,35	1380	1392	1421,00	36,36	84,54	2,08	0,87	1380	1388,80	26,40	65,16	0,64	0,00
i05	1207	1352	7,13	1262	1285	1302,00	8,55	90,02	1,32	1,82	1262	1274,90	26,02	86,51	1,02	0,00
i06	1261	1565	10,76	1413	1461	1492,10	5,84	85,67	2,13	3,40	1413	1468,80	24,58	71,96	3,95	0,00
i07	1279	1389	4,91	1333	1425	1352,00	12,38	92,50	1,43	0,68	1324	1333,00	24,58	71,96	0,68	0,00
i08	1299	1519	9,99	1425	1466,40	1466,40	12,88	91,94	2,91	3,19	1381	1418,00	26,09	64,62	2,68	0,00
i09	1444	1633	1713	1633	1651	1698,20	40,06	89,57	2,86	1,10	1633	1665,80	25,19	80,43	2,01	0,00
i10	1213	1411	5,85	1333	1371	1393,00	38,93	72,64	1,60	2,85	1333	1347,70	35,25	80,21	1,10	0,00
i11	1368	1696	11,07	1527	1557	1602,10	5,70	72,93	2,90	1,96	1527	1551,20	23,99	76,27	1,58	0,00
i12	1325	1629	8,24	1505	1537	1565,80	7,64	88,01	1,87	2,13	1505	1521,00	24,74	81,50	1,06	0,00
i13	1360	1519	6,67	1424	1449	1482,50	23,66	83,02	2,31	1,76	1424	1442,50	25,61	52,14	1,30	0,00
i14	1233	1369	7,54	1273	1287	1306,90	6,57	80,19	1,55	1,10	1273	1284,60	24,90	59,46	0,91	0,00
i15	1295	1455	8,18	1345	1362	1394,20	14,44	75,85	2,36	1,26	1345	1356,60	24,96	67,75	0,86	0,00
i16	1364	1491	15,02	1491	1508	1581,10	4,28	87,78	4,85	1,14	1491	1512,50	23,79	72,49	1,44	0,00
i17	1283	1322	1,54	1302	1318	1335,30	68,71	83,41	1,31	1,23	1302	1305,30	27,96	59,88	0,25	0,00
i18	1345	1594	5,01	1518	1519	1552,10	21,68	89,47	2,18	0,07	1518	1531,30	25,88	85,80	0,88	0,00
i19	1367	1673	10,28	1517	1573	1628,60	26,00	84,78	3,53	3,69	1517	1546,80	24,24	92,88	1,96	0,00
i20	1328	1406	3,13	1406	1428	1469,30	67,47	91,01	2,89	1,56	1406	1414,90	27,21	87,27	0,63	0,00
i21	1341	1565	7,12	1461	1481	1510,40	7,64	80,47	1,99	1,37	1461	1464,40	24,22	76,96	0,23	0,00
i22	1326	1618	12,60	1437	1484	1521,00	12,57	86,19	2,49	3,27	1437	1449,10	24,34	72,13	0,84	0,00
i23	1266	1539	10,32	1395	1425	1456,10	3,56	84,86	2,18	2,15	1395	1413,60	24,45	89,60	1,33	0,00
i24	1260	1425	5,48	1351	1359	1383,40	14,40	87,65	1,80	0,59	1351	1362,00	26,10	69,03	0,81	0,00
i25	1376	1590	5,09	1513	1546	1604,50	38,04	79,12	3,78	2,18	1513	1545,10	26,32	67,80	2,12	0,00
i26	1318	1567	8,22	1448	1475	1520,10	22,25	80,51	3,06	1,86	1448	1458,40	24,40	67,92	0,72	0,00
i27	1261	1458	8,08	1349	1356	1382,30	6,59	70,23	1,94	0,52	1349	1354,70	25,22	89,33	0,42	0,00
i28	1359	1550	6,09	1461	1486	1553,90	53,80	81,15	4,57	1,71	1461	1471,50	25,62	63,48	0,72	0,00
i29	1280	1415	6,95	1323	1338	1360,80	6,76	74,49	1,70	1,13	1323	1329,70	25,17	75,00	0,51	0,00
i30	1344	1621	9,01	1487	1512	1555,60	7,43	84,63	2,88	1,68	1487	1504,10	25,29	66,10	1,15	0,00
<b>Média</b>	<b>1306,77</b>	<b>1516,87</b>	<b>7,36</b>	<b>1412,13</b>	<b>1436,37</b>	<b>1472,23</b>	<b>22,81</b>	<b>83,13</b>	<b>2,46</b>	<b>1,71</b>	<b>1412,13</b>	<b>1428,79</b>	<b>25,67</b>	<b>73,89</b>	<b>1,17</b>	<b>0,00</b>

Tabela 2.3 - Melhorias e comparação com o caso discreto.

Inst.	Melhorias			CONT $\times$ DISC		
	AMxBT	CSxAM	CSxBT	BT	AM	CS
i01	5,5	1,9	7,2	21,08	14,48	12,35
i02	2,1	0,8	3,0	7,45	5,15	4,28
i03	4,0	2,2	6,1	13,91	9,30	6,91
i04	3,3	0,9	4,2	10,60	6,91	5,99
i05	5,0	1,8	6,7	12,01	6,46	4,56
i06	6,6	3,4	9,7	24,11	15,86	12,05
i07	4,0	0,7	4,7	8,60	4,22	3,52
i08	6,2	3,2	9,1	16,94	9,70	6,31
i09	3,6	1,1	4,7	18,63	14,34	13,09
i10	2,8	2,9	5,5	16,32	13,03	9,89
i11	8,2	2,0	10,0	23,98	13,82	11,62
i12	5,6	2,1	7,6	22,94	16,00	13,58
i13	4,6	1,8	6,3	11,69	6,54	4,71
i14	6,0	1,1	7,0	11,03	4,38	3,24
i15	6,4	1,3	7,6	12,36	5,17	3,86
i16	12,1	1,1	13,1	25,73	10,56	9,31
i17	0,3	1,2	1,5	3,04	2,73	1,48
i18	4,7	0,1	4,8	18,51	12,94	12,86
i19	6,0	3,7	9,3	22,38	15,07	10,97
i20	1,5	1,6	3,0	9,19	7,53	5,87
i21	5,4	1,4	6,6	16,70	10,44	8,95
i22	8,3	3,3	11,2	22,02	11,92	8,37
i23	7,4	2,2	9,4	21,56	12,56	10,19
i24	4,6	0,6	5,2	13,10	7,86	7,22
i25	2,8	2,2	4,8	15,55	12,35	9,96
i26	5,9	1,9	7,6	18,89	11,91	9,86
i27	7,0	0,5	7,5	15,62	7,53	6,98
i28	4,1	1,7	5,7	14,05	9,35	7,51
i29	5,4	1,1	6,5	10,55	4,53	3,36
i30	6,7	1,7	8,3	20,61	12,50	10,64
Média	5,2	1,7	6,8	16,0	9,8	8,0

## 2.8 Considerações finais sobre o PAB

Soluções para o PAB contribuem no aprimoramento de uma logística na distribuição do espaço do cais minimizando o tempo total de serviço dos navios, evitando assim prejuízos para o porto com atrasos no atendimento das embarcações.

Para resolver o PAB, foi proposta uma aplicação do método híbrido CS utilizando o SA como gerador de soluções. O CS mostrou ser adequado e eficiente na localização de regiões promissoras por meio do enquadramento dessas em *clusters*. Dessa forma, percebe-se que o CS atuou como uma alternativa para acelerar a obtenção de boas soluções. Esse fato fica evidente quando os resultados obtidos são comparados diretamente com a BT (CORDEAU; LAPORTE, 2005) e o AM (MAURI et al., 2011).

De uma forma geral, os resultados obtidos demonstram que o CS foi capaz de gerar soluções de boa qualidade para todas as instâncias em tempos computacionais viáveis. Esses resultados ainda foram comparados com outras abordagens recentes encontradas na literatura e em todos os casos, as soluções apresentadas foram as mais favoráveis.

### 3 PROBLEMA DE AGRUPAMENTOS

Dado um conjunto de dados que apresenta atributos e representa algo do mundo real (dados de plantas, proteínas, câncer entre outros), o problema de agrupamentos consiste no processo de separar o conjunto de dados em grupos, tal que os membros de cada grupo sejam similares entre si. A dificuldade ocorre quando alguns atributos não são definidos de forma clara ou que geram interpretações equivocadas. A questão a ser resolvida é a maneira adequada de agrupar estes dados.

Para que os dados sejam agrupados é necessário identificar quão próximos ou distantes eles estão. Um dos caminhos seria a criação de uma matriz de distâncias e em seguida determinar a similaridade usando uma métrica. Se os dados estão próximos ou não, depende da escolha da métrica utilizada.

Os grupos criados são analisados por uma função custo que busca maximizar a similaridade dos dados dentro do mesmo grupo ou minimizar a similaridade em grupos distintos. A solução final do problema de agrupamentos (grupos criados) pode passar ainda por uma validação, analisada por um índice. As medidas mais utilizadas atualmente para medir similaridade entre conjuntos são o Índice Rand (RAND, 1971) e o seu derivado Índice Rand Corrigido (CRand) (HUBERT; ARABIE, 1985).

A dificuldade em resolver o problema acima vem de como identificar os dados próximos e desenvolver métodos que cumpram a função de agrupá-los. Essa não é uma tarefa fácil, devido a grande dificuldade em encontrar boas soluções, ou seja, classificar os dados de forma eficiente. Quanto a classificação dos dados eles podem ser supervisionados ou não (ABBASI; YOUNIS, 2007). Na análise supervisionada o método é treinado com padrões conhecidos dos dados para usá-los em novas amostras. Já na análise não supervisionada, o algoritmo busca encontrar estruturas nos dados que permitam a separação em grupos, sem o conhecimento prévio dos dados. Apesar dos métodos não supervisionados conseguirem classificações menos precisas que as soluções dos métodos supervisionados, eles são mais adequados quando nenhuma informação prévia dos grupos está disponível.

Além disso, as técnicas de agrupamentos podem ser divididas em hierárquico ou particional (não hierárquico) (BERKHIN, 2006). O hierárquico constrói uma árvore de grupos (dendrograma). Cada nó (dado) do grupo contém subgrupos, nós agregados

a grupos próximos são abrangidos pelo mesmo pai. O caso hierárquico é subdividido em aglomerativo e divisivo. No aglomerativo cada dado é um grupo e de modo recursivo unem-se dois ou mais grupos, até que se forme um único grupo com todos os dados. No divisivo o processo iterativo inicia com um grupo contendo todos os dados e de forma recursiva divide-se o conjunto até que cada grupo contenha apenas um elemento. No agrupamento particional escolhem-se os  $k$  dados como centros e agrupam-se os demais dados aos centros mais próximos. O critério de parada para ambos os casos também pode ser até uma quantidade de iterações do algoritmo.

Finalizando a classificação, os agrupamentos podem ser exclusivos ou não exclusivos. No caso exclusivo cada dado pertence a um único grupo, já no caso não exclusivo cada dado pode pertencer a vários grupos. O problema de agrupamentos analisado neste trabalho é considerado não supervisionado, particional e exclusivo.

Segundo Xu e Wunsch (2005) o problema de agrupamentos tem sido aplicado em uma ampla variedade de áreas de pesquisa, tais como, microbiologia, geologia, sensoriamento remoto, segmentação de imagens, inteligência artificial, aprendizagem de máquina, reconhecimento de padrões, mineração de dados espaciais, entre outros. Autores têm aplicado meta-heurísticas para sua resolução, tal como *busca tabu* (WANG et al., 2004), *Greedy Randomized Adaptive Search Procedure* - GRASP (NASCIMENTO et al., 2010), *Genetic Algorithms* - GA (MAULIK; BANDYOPADHYAY, 2000) e *Simulated Annealing* - SA (WANG et al., 2004).

### 3.1 Revisão abreviada da literatura para agrupamentos

O problema de agrupamentos tem sido muito estudado e pesquisadores utilizam-se de uma grande diversidade de métodos buscando resolvê-lo. Como exemplo, o trabalho de Rand (1971) propõe critérios que isolam aspectos do desempenho de um método, tais como, retorno, sensibilidade e estabilidade. Estes critérios dependem de uma medida de similaridade entre dois agrupamentos diferentes do mesmo conjunto de dados. A medida considera essencialmente o modo como cada par de pontos de dados é atribuído em cada conjunto.

Handl et al. (2005) mostram a grande quantidade de técnicas disponíveis para a validação dos resultados obtidos para o problema de agrupamento. O foco principal do trabalho é a análise de dados em pós-genômica. Os autores usam dados biológicos sintéticos e reais para demonstrar os benefícios, e também alguns riscos da validação.

O trabalho de Al-Sultan (1995) apresenta uma busca tabu para resolver o problema de agrupamentos, a função objetivo minimiza a distância entre os pontos pertencentes a um mesmo agrupamento e o seu centro. Mitra e Banka (2006) introduziram um método *multi objetivo biclustering* evolutivo com estratégias de busca local. Hong et al. (2008) propõem um algoritmo de seleção não supervisionado que combina métodos de agrupamentos de grupos. As soluções são obtidas por um método de agrupamentos, em seguida, a população é adaptada por um algoritmo de aprendizagem para encontrar o subconjunto característico que melhor se adapta a solução.

Chang et al. (2009) propõem um algoritmo de separação baseado em um algoritmo genético com rearranjo de genes para o problema *k-Means Clustering*, o qual busca remover degenerações para o propósito de uma busca mais eficiente. Um operador de *crossover* que explora similaridades entre os cromossomos em uma população também é apresentado.

Nascimento et al. (2010) apresentaram uma formulação matemática e um algoritmo *Greedy Randomized Adaptive Search Procedure* (GRASP) para resolver o problema de agrupamentos usando dados biológicos. Os resultados computacionais foram comparados com a aplicação do CPLEX (ILOG, 2009), *k-Means*, *k-Medians* e *Partitioning Around Medoids* (PAM). O índice CRand é usado para comparar os métodos. Serpa (2011) aplicou uma heurística de busca local iterativa combinada com a heurística *Clustering Search* para os mesmos dados de Nascimento et al. (2010) e obteve algumas melhorias nas soluções.

*Surveys* sobre o problema de agrupamentos, comentários e análises de alguns algoritmos podem ser encontrados em Nascimento e Carvalho (2011), que apresentam estudos sobre algoritmos de agrupamentos baseados em teoria espectral em grafos. Berkhin (2006) relata a variedade de algoritmos para aplicação em problemas de mineração de dados com dados reais. Em Xu e Wunsch (2005) é apresentada uma revisão das variedades de abordagens que aparecem na literatura sobre algoritmos de agrupamentos.

Como é possível observar nos trabalhos anteriormente citados, a abordagem de geração de colunas para problema de agrupamentos tem sido pouco explorada na literatura. Senne e Lorena (2001) e Senne et al. (2007) descrevem algumas relações entre a relaxação lagrangeana e *surrogate* juntamente com o processo de geração de colunas para problemas de programação linear (0-1). A relaxação lagrangeana/*surrogate* foi

combinada com a geração de colunas para acelerar e estabilizar os limites primal e dual obtidos utilizando a seleção do custo reduzido. Em Filho e Lorena (2010) uma heurística baseada em geração de colunas foi aplicada ao problema da formação de células de máquinas e partes. O trabalho apresenta um novo modelo para um problema de particionamento com uma restrição de cardinalidade.

### 3.2 Geração de colunas para o problema de $p$ -medianas

As heurísticas híbridas para resolver o problema de agrupamentos que serão descritas em 3.3, começam a partir de um conjunto de dados e sem qualquer informação dos padrões, constroem grupos que contenham características semelhantes entre seus objetos. Os grupos são obtidos pela técnica de Geração de Colunas (GC) para solução do problema de  $p$ -medianas. Soluções de  $p$ -medianas minimizam a soma das distâncias entre nós e sua facilidade (mediana) mais próxima e espera-se que as características entre os nós que estão alocados a mesma facilidade sejam semelhantes. Assim, uma possível solução viável para o problema de agrupamentos resultaria da separação de um grupo em  $p$  subgrupos distintos, na qual cada grupo contém uma facilidade e os nós alocados a ela.

Dada a matriz de distâncias  $[d_{ij}]_{n \times n}$ , de acordo com Senne et al. (2007), o problema de  $p$ -medianas pode ser modelado como o seguinte problema de particionamento de conjuntos:

$$\text{Minimizar: } \sum_{k=1}^m c_k y_k \quad (3.1)$$

$$\text{Sujeito a: } \sum_{k=1}^m A_k y_k = 1 \quad (3.2)$$

$$\sum_{k=1}^m y_k = p \quad (3.3)$$

$$y_k \in \{0, 1\} \quad (3.4)$$

o conjunto  $S = \{S_1, S_2, \dots, S_m\}$ , é formado por subconjuntos dos objetos  $N = \{1, \dots, n\}$ ;



$M = \{1, 2, \dots, m\}$  é o conjunto dos índices correspondentes as colunas; o subconjunto  $S_k$  corresponde a uma coluna  $A_k$  do conjunto de restrições 3.2;  $A_k = [a_i]_{n \times 1}$ , para  $k \in M$ ; com  $a_i = 1$  se  $i \in S_k$ , e  $a_i = 0$  caso contrário;  $c_k = \text{Min}_{i \in S_k} \left( \sum_{j \in S_k} d_{ij} \right)$ , para  $k \in M$ ;  $y_k$  são as variáveis de decisão, com  $y_k = 1$  se o subconjunto  $S_k$  é escolhido e  $y_k = 0$  caso contrário. Para cada conjunto  $S_k$ , a escolha da mediana é realizada pelo cálculo do custo ( $c_k$ ).

Como o número de colunas pode ser muito grande, o problema a ser resolvido é uma relaxação de programação linear de (3.1) - (3.4) conhecido como Problema Mestre (PM) e definido da seguinte forma:

$$\text{Minimizar:} \quad \sum_{k=1}^m c_k y_k \quad (3.5)$$

$$\text{Sujeito a:} \quad \sum_{k=1}^m A_k y_k \geq 1 \quad (3.6)$$

$$\sum_{k=1}^m y_k = p \quad (3.7)$$

$$y_k \in [0, 1] \quad (3.8)$$

Depois de definir um conjunto de colunas iniciais, o PM é resolvido e os seus custos duais finais ( $\mu_i$ , para  $i = 1, \dots, n$ ) e  $\rho$  são usados para gerar novas colunas ( $\beta_j = [\beta_{ij}]_{n \times 1}$ ), resolvendo o seguinte subproblema:

$$\text{Min}_{j \in N} \left[ \text{Min}_{\beta_{ij} \in \{0,1\}} \sum_{i=1}^n (d_{ij} - \mu_i) \beta_{ij} \right] \quad (3.9)$$

O problema (3.9) é resolvido considerando cada  $j \in N$  como uma mediana, e  $\beta_{ij} = 1$ , se  $(d_{ij} - \mu_i \leq 0)$  e  $\beta_{ij} = 0$ , se  $(d_{ij} - \mu_i) > 0$ . Para o novo conjunto  $S_j$  é definido o vértice  $i$  tal que  $\beta_{ij} = 1$  para o subproblema (3.9). Então a coluna  $\begin{bmatrix} \beta_j \\ 1 \end{bmatrix}$  é adicionada ao PM se o valor da solução do subproblema (3.9) é menor que  $\rho$ . Todas as colunas encontradas

satisfazendo a desigualdade (3.10) para  $j = 1, \dots, n$ , podem ser adicionadas ao conjunto de colunas, acelerando o processo de GC.

$$\left[ \text{Min}_{\beta_{ij} \in \{0,1\}} \sum_{i=1}^n (d_{ij} - \mu_i) \beta_{ij} \right] < \rho \quad (3.10)$$

O algoritmo de GC está resumido no fluxograma da Figura 3.1. Após a definição de um conjunto inicial de colunas, o software CPLEX (ILOG, 2009) é usado para resolver PM e obter os valores duais  $\mu_i$ ,  $i = 1, \dots, n$  e  $\rho$ . Eles são utilizados para gerar novas colunas  $\left( \begin{bmatrix} \beta_j \\ 1 \end{bmatrix} \right)$  por meio da solução do subproblema (3.9). Todas as colunas  $\left( \begin{bmatrix} \beta_j \\ 1 \end{bmatrix} \right)$ , que satisfazem a desigualdade (3.10) (para  $i = 1, \dots, n$ ) são colunas de entrada para o PM. Algumas colunas com alto custo reduzido são removidas para manutenção das boas soluções obtidas pelo PM. O processo iterativo termina se não forem encontradas novas colunas.

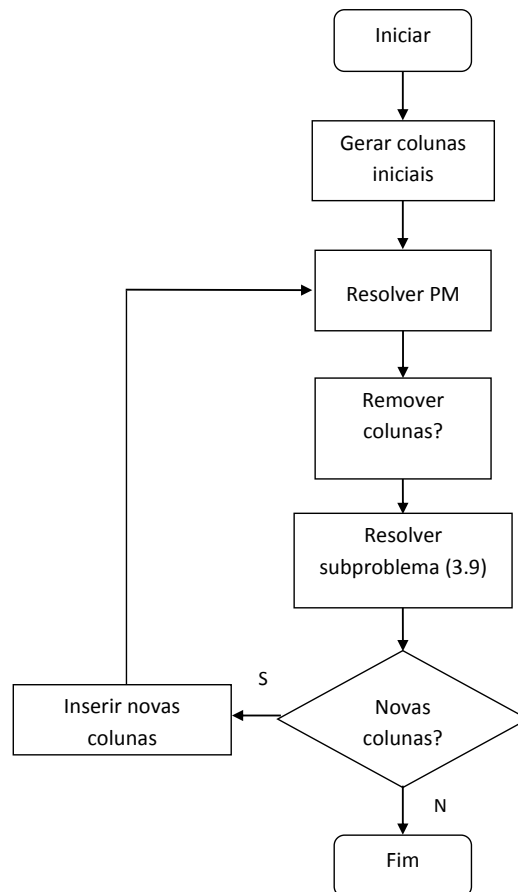


Figura 3.1 - Algoritmo GC.

### 3.3 Heurísticas híbridas para o problema de agrupamentos

As heurísticas híbridas examinadas neste trabalho podem ser classificadas como uma combinação de meta-heurísticas e GC. Duas possibilidades básicas são exploradas na literatura: aplicar meta-heurísticas no subproblema gerador de colunas ou diretamente no PM, para gerar colunas iniciais ou gerar colunas candidatas a participar do GC (MAURI; LORENA, 2007), (PIRKWIESER; RAIDL, 2010), (FILHO; LORENA, 2000), (MASSEN et al., 2013). As heurísticas híbridas propostas neste trabalho podem ser vistas como uma terceira opção na qual o processo de GC gera soluções do problema de agrupamentos que são melhoradas a seguir pelo método *Path Relinking* - PR (RESENDE; RIBEIRO, 2005).

Todas as heurísticas propostas geram soluções viáveis de  $p$ -medianas (agrupamentos) obtidas em várias fases do processo de GC. O custo ( $c_k$ ) do subconjunto  $S_k$  no modelo 3.5 - 3.8, será o valor obtido com a soma das distâncias de todas as arestas para os vértices nos agrupamentos (um clique), além do calculado em Senne et al. (2007), que seria a soma das distâncias dos pontos para a mediana mais próxima ( $p$ -medianas). Os agrupamentos finais obtidos pelas heurísticas são então avaliados pelo índice CRand, permitindo uma comparação justa com outros métodos utilizados na literatura. É importante observar que o índice CRand não foi usado nas funções objetivo das heurísticas propostas.

O valor de CRand varia entre  $[-1, 1]$  e quanto mais próximos de 1 as partições serão mais similares. Assim, sejam  $U$  e  $V$  duas partições, tal que  $N$  é o número de objetos de um conjunto de dados,  $n_i$  é o número de objetos no grupo  $i$  de  $U$ ,  $n_j$  o número de objetos no grupo  $j$  de  $V$ , os índices  $i$  e  $j$  variam de acordo com o número de grupos das partições. Assim, o índice CRand proposto por Hubert e Arabie (1985) é definido da seguinte forma:

$$CRand = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} / \binom{N}{2}}{\frac{1}{2} \left[ \sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2} \right] - \sum_i \binom{n_i}{2} \sum_j \binom{n_j}{2} / \binom{N}{2}} \quad (3.11)$$

O algoritmo PR é usado para intensificar e diversificar a busca em um grupo de soluções. Ele realiza movimentos exploratórios em uma vizinhança a partir de uma solução inicial buscando atingir uma solução guia. O objetivo é encontrar as melhores soluções no caminho que conectam estas soluções. Os movimentos gradualmente introduzem informações dos atributos da solução guia para a solução inicial (RESENDE; RIBEIRO, 2005).

A Figura 3.2 apresenta um exemplo do PR usado nesta tese. Supõe-se que a solução inicial (5, 10, 15) e a solução guia (6, 11, 16) são as medianas de cada solução. O PR calcula a diferença ( $\Delta$ ) entre as soluções (medianas), ou seja, o número de posições diferentes entre elas. A solução inicial e a solução guia diferem em três posições. A partir das próximas três soluções possíveis, tal que  $\Delta = 3$ , o PR escolhe a melhor solução de  $p$ -medianas e assim por diante até atingir  $\Delta = 0$ , na qual o processo é finalizado com a melhor solução no caminho.

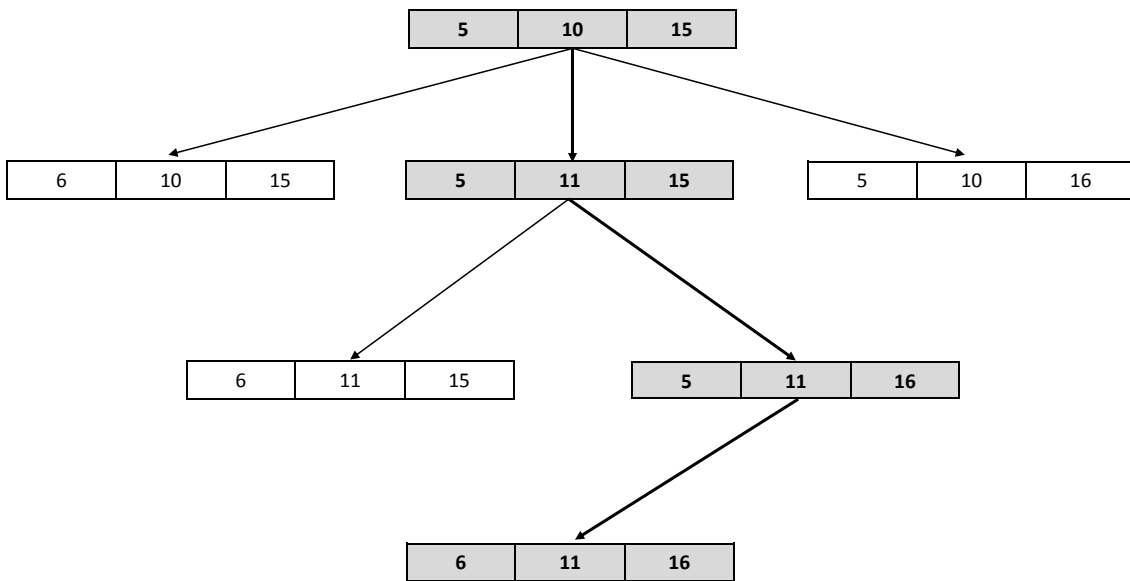


Figura 3.2 - Aplicação do PR.

As heurísticas propostas são:

- HPL: a solução viável do PM para o problema de  $p$ -medianas obtida depois de resolver o PM com as colunas iniciais;
- HGC: a solução viável do PM para o problema de  $p$ -medianas obtida após a resolução do PM final;
- HSI-PR: uma heurística com o PR considerando as colunas iniciais do processo de GC;
- HPL-PR: as soluções viáveis de  $p$ -medianas da HPL combinada com o PR;
- HGC-PR: as soluções viáveis de  $p$ -medianas da HGC combinado com o PR.

O algoritmo HPL considera a solução de programação linear do PM obtida pelo CPLEX usando as colunas iniciais encontradas pela Heurística de Solução Inicial (HSI) (Figura 3.3). Como essa solução não é inteira (em geral), as primeiras  $p$  colunas  $A_1, \dots, A_p$  são mantidas independente de seus valores reais ou inteiros  $y_1, \dots, y_p$ , e outras possíveis colunas fracionais são descartadas. A solução de agrupamento  $S_{sol} = \{S_1, \dots, S_p\}$  é então calculada pelo índice CRand.

A HGC é semelhante ao HPL, mas obtido com o PM na etapa final do processo de geração de colunas.

A HSI constrói um conjunto inicial de colunas para o PM. A HSI (Figura 3.3) inicia com a leitura dos dados e cálculo da matriz de distâncias entre os nós (ele usa as métricas descritas na Seção 3.4.1). Os grupos individuais são construídos ( $S_j$ ) ( $j = 1, \dots, p$ ), compostos pela mediana  $j$ , escolhida de forma aleatória, e os vértices mais próximos a mediana. Os grupos criados são convertidos em colunas de uns (se  $i \in S_j$ ) e zeros (se  $i \notin S_j$ ), e as colunas são adicionadas ao conjunto inicial de colunas de PM. Todo o processo é repetido até que o número de colunas ( $Num\_col$ ) atinja o número máximo de colunas geradas ( $Num\_Max\_Col$ ). A HSI-PR considera o PR aplicado a cada trinta soluções de  $p$ -medianas geradas no processo (uma solução de agrupamento com  $p$  medianas corresponde a  $p$  colunas  $A_1, \dots, A_p$ ) e finalmente, os melhores grupos gerados pelo PR ( $S_{sol}$ ) são avaliados pelo índice CRand.

**Solução Inicial:**

```

1 Ler dados();
2 Cálculo da matriz de distâncias;
3 Dado  $Num\_Max\_Col$  o número máximo de GC;
4  $Num\_col \leftarrow 0$ ;
5 Repita
6   Dado  $P = (n_1, \dots, n_p)$  um conjunto de vértices escolhidos de forma aleatória
7   Para ( $j=1, \dots, p$ ) faça
8      $S_j \leftarrow \{n_j\} \cup \left\{ q \in N - P \mid d_{qn_j} = \min_{t \in P} \{d_{qt}\} \right\}$ 
9      $c_j \leftarrow \min_{t \in S_j} \left\{ \sum_{i \in S_j} \{d_{it}\} \right\}$ 
10    Para ( $i = 1, \dots, n$ ) faça
11      Se ( $i \in S_j$ ), faça  $a_{ij} \leftarrow 1$ ;
12      Se  $i \notin S_j$ , faça  $a_{ij} \leftarrow 0$ ;
13      Adicione a coluna  $\begin{bmatrix} A_j \\ 1 \end{bmatrix}$  ao conjunto inicial de colunas;
14     $Num\_col \leftarrow Num\_col + 1$ ;
15 Enquanto ( $Num\_col < Num\_Max\_Col$ );
16 Fim

```

Figura 3.3 - Algoritmo HSI

Fonte: Adaptada de Pereira et al. (2007).

O algoritmo HPL-PR aplica o PR considerando a solução inicial como a solução final da HSI e como solução guia, a solução viável do final de execução da HPL. A solução de agrupamento  $S_{sol} = \{S_1, \dots, S_p\}$  é então avaliada pelo índice CRand.

A HGC-PR utiliza o PR em soluções intermediárias  $y_j$  (colunas  $\begin{bmatrix} y_j \\ 1 \end{bmatrix}$ ) do PM. As soluções do PM são consideradas as soluções iniciais e guia a cada trinta iterações do processo de GC. Todas as soluções de agrupamento geradas ( $S_{sol}$ ) são avaliadas pelo índice CRand.

### 3.4 Experimentos computacionais

Nesta seção são apresentados os dados utilizados nos experimentos e os testes computacionais para os métodos propostos.

#### 3.4.1 Dados utilizados

Os dados são no total de 8 instâncias, sendo eles: Íris, Yeast, Breast, BreastA, BreastB, Proteínas, DLBCLA e DLBCLB. Os dados de proteínas foram obtidos em <http://ranger.uta.edu/~chqing/protein>. Os dados de Yeast, Breast e Íris foram obtidos no repositório UCI (ABBASI; YOUNIS, 2007). Os dados de BreastA, BreastB, DLBCLA e DLBCLB são

do repositório de dados do programa de câncer (<http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi>).

Os dados de Íris são compostos por 150 objetos (flores) separados em quatro atributos: larguras e comprimentos da sépala e da pétala e são divididas em 3 grupos: íris virgínica, íris setosa e íris versicolor.

Os dados de Yeast são composto por um conjunto de 1884 proteínas de levedura com 8 atributos relativos às características calculadas a partir das sequências de aminoácidos e são divididos em 10 grupos de acordo com a localização dos sítios de proteínas: citoplasmática ou citoesqueleto (463 objetos), nuclear (429 objetos), mitocondrial (244 objetos), proteínas de membrana sem sinal N-terminal (163 objetos), proteína de membrana de mórula sem sinal de divisão (51 objetos), proteína de membrana de mórula com sinal de divisão (44 objetos), extracelular (37 objetos), vascular (30 objetos), peroxissomal (20 objetos) e localizadas no lúmen do retículo endoplasmático (5 objetos).

Os dados de Breast são composto por células cancerígenas com 699 objetos e 9 atributos cada uma e são divididos em benigno (458 objetos) ou maligno (241 objetos).

Os dados de BreastA e BreastB são dois conjuntos com 98 e 48 objetos de tumor de mama com 1213 atributos, onde BreastA é dividido em 3 grupos com 11, 51 e 36 objetos e BreastB possui 2 grupos com 25 e 24 objetos, que levam em consideração o receptor estrogênico positivo e negativo.

Os dados de DLBCLA e DLBCLB correspondem a exemplos de linfomas de grandes células B e são compostos por 141 e 180 objetos. O DLBCLA apresenta 661 atributos separados por 3 grupos: fosforilação oxidativa (49 objetos), respostas das células B (50 objetos) e respostas de hospedeiros (42 objetos). O DLBCLB apresenta 180 objetos divididos nos mesmos 3 grupos do DLBCLA com respectivamente: 12, 51 e 87.

Os dados de Proteínas são composto por 698 objetos com 125 atributos, divididos em 4 grupos de 116, 226, 260 e 96 objetos.

Os algoritmos híbridos propostos consideram estes dados representados em forma de grafos completos ponderados. Os pesos das arestas (distâncias ou dissimilaridades) são calculados por quatro métricas: A distância euclidiana, a distância City Block, a correlação de Pearson e a correlação cosseno.

Definindo  $x_{ik}$  como o  $k$ -ésimo atributo do objeto  $i$  e  $n_a$  como o número de atributos de um conjunto de dados, as métricas utilizadas são definidas da seguinte forma:

- **Euclidiana:** é a distância geométrica entre os objetos.

$$d_{ij} = \sqrt{\sum_{k=1}^{n_a} (x_{ik} - x_{jk})^2} \quad (3.12)$$

- **City block ou Manhattan:** é a soma das diferenças absolutas entre os atributos de dois objetos.

$$d_{ij} = \sum_{k=1}^{n_a} |x_{ik} - x_{jk}| \quad (3.13)$$

- **Correlação de Pearson:** mede o grau de correlação entre dois objetos produzindo um valor entre  $[-1, 1]$  (direção negativa ou positiva) no caso de valor zero não existe dependência linear entre os objetos.

$$c_{ij} = \frac{n_a \sum x_{ik}x_{jk} - \sum x_{ik} \sum x_{jk}}{\sqrt{n_a \sum x_{ik}^2 - (\sum x_{ik})^2} \sqrt{n_a \sum x_{jk}^2 - (\sum x_{jk})^2}} \quad (3.14)$$

A dissimilaridade entre dois objetos  $i$  e  $j$  pela correlação de Pearson é da forma:  $d_{ij} = 1 - |c_{ij}|$ .

- **Correlação cosseno:** é a correlação geométrica definida pelo ângulo entre dois objetos.

$$D_{ij} = \frac{\sum_{k=1}^{n_a} x_{ik}x_{jk}}{\sum_{k=1}^{n_a} x_{ik}^2 \sum_{k=1}^{n_a} x_{jk}^2} \quad (3.15)$$

O valor  $D_{ij}$  está entre  $[-1, 1]$ . Se  $D_{ij} = 1$  significa que o ângulo entre dois objetos é  $0^\circ$ . Se  $D_{ij} = -1$  significa que o ângulo entre os objetos é  $180^\circ$ . Assim, a dissimilaridade entre dois objetos  $i$  e  $j$  pela correlação cosseno é:  $d_{ij} = 1 - |D_{ij}|$ .

### 3.4.2 Resultados obtidos para o problema de agrupamentos

Os experimentos computacionais foram executados em um PC AMD Athlon 64 bits, processador dual core com 2.5 GHz e 4 GB de memória RAM. Os parâmetros utilizados nos cinco métodos propostos foram definidos de forma empírica. Em seguida são apresentados os que obtiveram os melhores resultados.

- número de colunas iniciais geradas com a HSI = 500,



- número máximo número de iterações = 500 (ou até que o algoritmo atinja um máximo de 10.000 colunas geradas),
- o algoritmo GC foi executado 100 vezes para cada conjunto de dados,
- o número de medianas variaram de 2 a 30. Durante os experimentos verificou-se que os resultados pioraram com valores acima de 10 medianas. Assim, os resultados finais foram realizados para até 10 medianas, e os melhores resultados são apresentados nas tabelas seguintes.

As tabelas de 3.1 a 3.16, mostram os resultados computacionais para as heurísticas híbridas propostas neste trabalho: HPL, HGC, HSI-PR, HPL-PR e HGC-PR. Nas tabelas de 3.1 a 3.8 o cálculo do custo é a somatória de todos os pontos dos agrupamentos (uma clique). Nas tabelas 3.9 a 3.16 o cálculo do custo é a somatória de todos os pontos a mediana mais próxima ( $p$ -medianas). Elas são comparadas com *GRASP*, *k-Means*, *k-Medians*, *PAM* (NASCIMENTO et al., 2010) e o *Iterated Local Search Algorithm* (ILS-CS) de (SERPA, 2011). As comparações com Nascimento et al. (2010) ficaram somente com os índices CRand, e os valores da função objetivo (FO) e tempos de processamento foram comparados apenas entre os métodos propostos neste trabalho.

As colunas nas tabelas são:

- **Melhor**: representa os melhores índices de CRand obtidos para todos os métodos para um tipo de dados específico, indicado na coluna **dados**. As melhores soluções são destacadas em negrito, e o símbolo (\*) indica uma melhoria dos resultados para essa instância.
- A coluna **M**: indica o número de agrupamentos (medianas da solução de  $p$ -medianas) na melhor solução obtida.
- O tempo para obter a melhor solução é indicada na coluna **TM**, enquanto o tempo total de execução de cada heurística é mostrado na coluna **TT**. A **Média** foi a média aritmética entre os valores das colunas.
- Os Métodos HGC e HGC-PR não conseguiram rodar para os dados de Yeast, Breast e proteínas. O *solver* CPLEX não conseguiu suportar a grande quantidade de colunas que iam sendo acrescentadas nas iterações do processo de geração de colunas.

Os valores de CRand nas Tabelas 3.1, 3.3 3.5 e 3.7, mostram que em geral, todos os cinco métodos propostos encontraram bons índices de CRand para todas as amostras analisadas

e, na maioria dos casos, foram as melhores soluções quando comparadas com os resultados de *GRASP*, *k*-Means, *k*-Medians, *PAM* and *ILS-CS*.

Para a distância *City Block* (Tabela 3.1): HSI-PR conseguiu uma melhor solução para os dados de proteínas com 4 agrupamentos, a HPL obteve as melhores soluções para os dados de Íris com 3 agrupamentos, Yeast com 5 agrupamentos e DLBCLB com 2 agrupamentos. A HPL-PR conseguiu as melhores soluções para os dados de Íris com 3 agrupamentos, BreastA com 3 agrupamentos e DLBCLB com 2 agrupamentos. A HGC obteve a melhor solução para DLBCLB com 2 agrupamentos. O método GC-PR conseguiu as melhores soluções para BreastA com 4 agrupamentos e DLBCLB com 2 agrupamentos.

Considerando a distância Euclidiana (Tabela 3.3) a HSI-PR conseguiu a melhor solução para os dados de Íris com 3 agrupamentos. A HPL obteve as melhores soluções para os dados de Yeast com 5 agrupamentos e DLBCLA para 2 agrupamentos. A HPL-PR obteve as melhores soluções para os dados de BreastA com 3 agrupamentos, BreastB com 5 agrupamentos, Proteínas com 3 agrupamentos, DLBCLA com 3 agrupamentos e DLBCLB com 2 agrupamentos. A HGC obteve as melhores soluções para Íris com 3 agrupamentos e DLBCLB com 2 agrupamentos. A HGC-PR conseguiu a melhor solução para o dado de DLBCLB com 2 agrupamentos.

Para a correlação de Pearson (Tabela 3.5) a HSI-PR conseguiu boas soluções, onde nos dados de Yeast, BreastB e DLBCLA superaram as soluções dos métodos de (NASCIMENTO et al., 2010) e Serpa (2011). A HPL conseguiu as melhores soluções para Íris com 3 agrupamentos e Yeast com 5 agrupamentos. A HPL-PR obteve 3 soluções entre as melhores sendo elas para os dados de Íris com 3 agrupamentos, BreastB com 4 agrupamentos e Proteínas com 5 agrupamentos. A HGC obteve 2 soluções entre as melhores, para os dados de Íris com 3 agrupamentos e DLBCLA com 3 agrupamentos. A HGC-PR conseguiu a melhor solução para o dado de Íris com 3 agrupamentos.

Finalmente para a correlação Cosseno (Tabela 3.7) o desempenho dos métodos propostos não foram tão bons quanto os anteriores. Algumas das soluções são próximas dos métodos considerados para comparação. A HSI-PR conseguiu uma melhor solução para DLBCLA com 4 agrupamentos, a HPL também conseguiu uma melhor solução para o dado de Yeast com 8 agrupamentos e HGC com uma melhor solução para o dados de BreastB com 4 agrupamentos.

As Tabelas 3.2, 3.4, 3.6 e 3.8 comparam o valor da função objetivo (FO), o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para as distâncias *City Block*, Euclidiana e Correlações de Pearson e Cosseno. Os valores mínimos de FO para a HPL, HPL-PR, HGC e HGC-PR são muito próximos, com exceção da HSI-PR, que

apresentou valores mais elevados de FO que os outros métodos. Isto confirma o efeito de geração de colunas para as soluções de  $p$ -medianas. Em relação ao tempo de processamento (TM e TT), HPL e HPL-PR foram os mais eficientes, seguido pela HSI-PR.

Tabela 3.1 - Comparação do resultados de CRand para a distância City Block (cliques).

Data	Melhor		GRASP		$k$ -means		$k$ -medians		PAM		ILS-CS		HSI-PR		HPL-PR		HGC		HGC-PR		
	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand
Iris	<b>0.922</b>	3	0.818	3	0.717	3	0.772	3	0.818	3	0.818	3	0.904	3	<b>0.922</b>	3	<b>0.922</b>	3	0.904	3	0.904
Yeast	<b>0.205</b>	7	0.150	7	0.181	6	0.152	7	0.157	4	0.157	4	0.183	5	<b>0.205</b>	5	0.193	-	-	-	-
Breast	<b>0.877</b>	2	0.870	2	0.770	2	0.807	2	<b>0.877</b>	2	0.657	2	0.691	2	0.706	2	-	-	-	-	-
BreastA	<b>0.748</b>	2	0.682	2	0.583	2	0.560	4	0.560	-	0.740	2	0.722	3	<b>0.748</b>	2	0.722	2	0.722	2	0.722
BreastB	<b>0.689</b>	7	0.159	3	0.328	3	0.284	2	0.187	-	0.440	5	0.508	6	<b>0.689</b>	4	0.570	4	0.570	4	0.590
Proteins	<b>0.352</b>	5	0.293	8	0.223	7	0.229	3	0.192	5	0.295	4	<b>0.352</b>	3	0.295	6	0.305	-	-	-	-
DLBCLA	<b>0.805</b>	3	0.800	3	<b>0.805</b>	3	0.784	3	0.406	-	0.659	3	0.742	3	0.795	3	0.781	3	0.781	3	0.674
DLBCLB	<b>0.702</b>	2	0.700	2	0.690	2	0.690	3	0.350	-	0.693	2	<b>0.702</b>	2	<b>0.702</b>	2	<b>0.702</b>	2	<b>0.702</b>	2	<b>0.702</b>
Média	0.649	-	0.559	-	0.537	-	0.428	-	0.537	-	0.578	-	0.598	-	0.631	2	0.736	2	0.718	2	0.718

Tabela 3.2 - Comparação dos valores de função objetivo (FO) para o  $p$ -medianas, o tempo da melhor solução (TM) e do tempo total de execução do algoritmo (TT) para a distância City Block.

Dados	HSL-PR		HPL		HPL-PR		HGC		HGC-PR						
	FO	B.T.	T.T.	FO	B.M.	T.T.	FO	B.T.	FO	B.T.					
Iris	36237.00	0.109	11.66	11093.60	0.046	27.39	11093.60	0.156	20.04	11070.20	42.49	7021.88	11059.80	66.61	7565.00
Yeast	387992.94	14.30	1470.24	262107.47	1.87	979.43	249103.77	2.917	1225.20	-	-	-	-	-	-
Breast	5756264.00	3.07	317.41	2664376.00	0.437	390.54	2622860.00	0.796	384.72	-	-	-	-	-	-
BreastA	2599740.75	0.062	35.24	1690253.38	0.078	79.84	1401086.75	0.281	97.25	1688118.38	12.69	5150.32	1690253.38	42.32	5678.74
BreastB	493533.66	0.032	9.90	237496.08	0.015	113.08	213274.92	0.577	403.29	233409.05	2.044	908.28	227772.80	3.80	2455.42
Proteins	169824464.00	3.71	505.19	124025104.00	0.827	553.07	75825112.00	0.858	621.32	-	-	-	-	-	-
DLBCLA	1116767872.00	0.10	41.65	896423616.00	0.093	126.86	894320000.00	0.187	116.33	900440448.00	51.79	8623.82	902085952.00	78.20	12220.12
DLBCLB	10565468.00	0.17	67.11	8443896.00	0.234	192.00	8443896.00	0.437	176.23	8443896.00	167.98	28170.47	8443896.00	214.12	22506.37
Média	163303946.54	2.69	307.30	129219742.82	0.45	307.78	122885803.38	365.04	380.55	182163388.33	463.79	9974.95	182491786.80	81.01	10085.13

Tabela 3.3 - Comparação do resultados de CRand para a distância Euclidiana (cliques).

Dados	Melhor		GRASP		$k$ -means		$k$ -medians		PAM		ILS-CS		HSI-PR		HPL		HPL-PR		HGC		HGC-PR			
	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	
Iris	<b>*0.886</b>	3	0.756	3	0.730	3	0.744	3	0.730	3	0.756	3	<b>0.886</b>	3	<b>0.886</b>	3	<b>0.886</b>	3	<b>0.886</b>	3	<b>0.886</b>	3	0.868	-
Yeast	<b>*0.198</b>	9	0.150	7	0.170	6	0.173	8	0.143	9	0.150	5	0.173	5	<b>0.198</b>	4	0.195	-	-	-	-	-	-	-
Breast	<b>0.877</b>	2	<b>0.877</b>	2	0.803	2	0.782	2	0.828	2	<b>0.877</b>	2	0.746	2	0.757	2	0.772	-	-	-	-	-	-	-
BreastA	<b>*0.716</b>	2	0.682	2	0.654	2	0.654	2	0.543	-	-	2	0.706	3	0.571	3	<b>0.716</b>	2	0.560	3	0.610	3	0.610	
BreastB	<b>*0.687</b>	2	0.314	3	0.286	3	0.260	2	0.187	-	-	4	0.466	6	0.496	5	<b>0.687</b>	4	0.513	4	0.590	4	0.590	
Proteins	<b>*0.387</b>	4	0.322	7	0.322	7	0.313	6	0.250	4	0.322	5	0.353	3	0.350	3	<b>0.387</b>	-	-	-	-	-	-	
DLBCLA	<b>*0.597</b>	4	0.408	4	0.309	5	0.365	4	0.276	-	-	4	0.440	4	0.512	3	<b>0.597</b>	4	0.500	3	0.508	3	0.508	
DLBCLB	<b>*0.689</b>	4	0.481	2	0.420	3	0.424	3	0.391	-	-	2	0.595	2	<b>0.689</b>	2	<b>0.689</b>	2	<b>0.689</b>	2	<b>0.689</b>	2	<b>0.689</b>	
Média	0.630	-	0.499	-	0.462	-	0.464	-	0.419	-	0.526	-	0.546	-	0.553	-	0.616	-	0.630	-	0.630	-	0.653	

Tabela 3.4 - Comparação dos valores de função objetivo (FO) para o  $p$ -medianas, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a distância distância Euclidiana

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.
Iris	11637.66	0.12	11.52	6516.14	0.031	23.07	6593.08	0.140	26.44	6659.27	45.61	6437.42	6624.85	64.67	6153.99
Yeast	158027.47	16.67	1758.30	129986.63	0.936	1000.30	219591.53	4.134	1101.72	-	-	-	-	-	-
Breast	1645948.12	3.057	309.11	1274902.50	1.076	360.97	1270723.62	1.622	425.58	-	-	-	-	-	-
BreastA	165095.91	0.063	27.14	66732.62	0.078	76.90	108311.51	0.141	73.21	105037.94	33.89	5960.14	65112.63	6.64	3458.40
BreastB	12794.33	0.016	7.89	5485.38	0.016	249.28	8430.37	0.109	146.95	9313.03	2.18	857.78	9264.09	2.60	2379.34
Proteins	17324004.00	4.10	499.14	17436162.00	0.873	510.82	17655962.00	0.795	543.72	-	-	-	-	-	-
DLBCLA	59364408.00	0.109	33.88	61075896.00	0.171	99.38	81595160.00	0.234	102.47	58851824.00	14.06	4975.63	81810368.00	84.39	15988.05
DLBCLB	497995.78	0.171	51.18	444741.22	0.499	140.21	444741.22	0.422	157.29	444741.22	138.90	35149.89	444741.22	171.25	32922.54
Média	9897488.90	384.78	337.27	10055052.81	134.82	307.61	12663689.16	719.73	322.17	11883515.09	46.92	10676.17	16467222.15	65.91	12180.46

Tabela 3.5 - Comparação dos resultados de CRand para a Correlação de Pearson (cliques).

Dados	Melhor CRand	GRASP		$k$ -means		$k$ -medians		PAM		ILS-CS		HSI-PR		HPL		HPL-PR		HGC		HGC-PR			
		M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand	M.	C.Rand
Iris	<b>0.941</b>	3	0.886	3	0.886	3	<b>0.941</b>	3	0.886	3	0.886	3	0.904	3	0.922	3	0.922	3	0.922	3	0.922	3	0.922
Yeast	<b>*0.204</b>	9	0.131	8	0.135	8	0.133	7	0.145	7	0.133	5	0.195	5	<b>0.204</b>	5	0.186	-	-	-	-	-	-
Breast	<b>0.441</b>	3	0.284	2	<b>0.441</b>	2	0.368	2	0.289	3	0.288	3	0.145	4	0.142	3	0.248	-	-	-	-	-	-
BreastA	<b>0.705</b>	2	0.692	2	<b>0.705</b>	2	<b>0.705</b>	2	0.635	-	-	2	0.660	3	0.587	2	0.650	3	0.573	3	0.573	3	0.587
BreastB	<b>*0.681</b>	2	0.355	3	0.289	3	0.283	3	0.227	-	-	4	0.480	4	0.522	4	<b>0.681</b>	4	0.659	4	0.659	4	0.661
Proteins	<b>*0.353</b>	4	0.344	7	0.313	7	0.306	6	0.245	4	0.344	5	0.341	4	0.350	5	<b>0.353</b>	-	-	-	-	-	-
DLBCLA	<b>*0.791</b>	4	0.585	4	0.605	4	0.684	4	0.586	-	-	3	0.709	3	0.720	3	0.761	3	<b>0.791</b>	3	<b>0.791</b>	3	0.685
DLBCLB	<b>0.665</b>	2	0.527	3	<b>0.665</b>	3	0.561	3	0.545	-	-	3	0.230	5	0.176	3	0.183	3	0.106	3	0.106	3	0.117
Média	0.595	-	0.476	-	0.505	-	0.498	-	0.445	-	0.413	-	0.458	-	0.453	-	0.498	-	0.610	-	0.610	-	0.594

Tabela 3.6 - Comparação dos valores de função objetivo (FO) para o  $p$ -mediana, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a Correlação de Pearson.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	T.M.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.
Iris	175.84	0.109	11.71	39.43	0.031	16.22	41.68	0.124	22.48	38.36	4.58	3845.70	38.36	42.33	5884.14
Yeast	114907.46	16.20	1741.67	56309.22	2.668	1131.02	53614.71	3.369	1175.29	-	-	-	-	-	-
Breast	91488.14	2.589	275.37	53342.65	1.139	404.30	122048.76	1.607	436.55	-	-	-	-	-	-
BreastA	3496.24	0.032	42.52	2120.24	0.156	119.32	3484.47	0.265	113.88	2100.27	10.65	2325.05	2120.24	49.38	6.707.435
BreastB	245.06	0.016	11.77	141.64	0.031	88.26	175.46	0.437	131.53	138.68	3.08	1370.64	174.62	10.79	4009.91
Proteins	7922.43	4.166	594.69	5227.15	0.671	763.43	4139.86	1.014	641.44	-	-	-	-	-	-
DLBCLA	787.85	0.11	50.32	696.74	0.172	132.33	701.19	0.203	142.55	699.74	5264.31	7030.68	691.33	29.89	7658.01
DLBCLB	1216.46	0.156	80.13	4865.10	0.453	320.03	10262.70	0.609	291.65	8224.98	51.12	11718.60	10489.22	38.90	8336.00
Média	27529.93	846.45	351.02	15342.77	476.06	371.86	24308.60	748.95	369.42	2240.40	1066.74	5258.13	2702.75	34.25	1346664.61

Tabela 3.7 - Comparação do resultados de CRand para a correlação cosseno (cliques).

Dados	Melhor		GRASP		$k$ -means		$k$ -medians		PAM		ILS-CS		HSI-PR		HPL		HPL-PR		HGC		HGC-PR	
	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.
Iris	<b>0.941</b>	3	<b>0.941</b>	3	0.904	3	<b>0.941</b>	3	<b>0.941</b>	3	0.868	3	0.868	3	0.868	3	0.884	3	0.886	3	0.885	3
Yeast	<b>*0.213</b>	9	0.135	9	0.138	6	0.132	7	0.146	9	0.135	5	0.179	8	<b>0.213</b>	5	0.195	-	-	-	-	-
Breast	<b>0.332</b>	3	0.293	4	0.258	3	0.306	3	<b>0.332</b>	3	0.293	2	0.277	8	0.196	6	0.279	-	-	-	-	-
BreastA	<b>0.691</b>	2	0.686	2	<b>0.691</b>	2	<b>0.691</b>	2	0.664	-	-	2	0.638	2	0.665	2	0.638	2	0.638	2	0.638	2
BreastB	<b>*0.604</b>	2	0.314	2	0.269	3	0.264	4	0.239	-	-	5	0.430	6	0.473	5	0.440	4	<b>0.604</b>	3	0.444	3
Proteins	<b>0.349</b>	4	<b>0.349</b>	7	0.320	6	0.304	6	0.247	4	<b>0.349</b>	4	0.339	4	0.347	6	0.321	-	-	-	-	-
DLBCLA	<b>*0.685</b>	4	0.605	5	0.642	4	0.678	3	0.547	-	-	6	0.621	4	0.669	4	<b>0.685</b>	3	0.664	3	0.619	3
DLBCLB	<b>0.623</b>	4	0.502	3	0.501	3	<b>0.623</b>	5	0.385	-	-	6	0.159	4	0.123	4	0.185	4	0.111	2	0.111	2
Média	<b>0.554</b>	-	0.478	-	0.465	-	0.492	-	0.433	-	0.430	-	0.439	-	0.441	-	0.457	-	0.581	-	0.539	-

Tabela 3.8 - Comparação dos valores de função objetivo (FO) para o  $p$ -medianas, o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para a Correlação Cosseno.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.	FO	B.T.	T.T.
Iris	21106.71	0.919	92.57	6932.01	0.078	25.81	6973.61	0.093	27.11	6938.28	109.69	18528.16	6973.61	163.30	21094.45
Yeast	528368.06	39.31	3882.75	263647.44	2.184	1080.91	263821.19	3.837	1356.81	-	-	-	-	-	-
Breast	475494.41	4.13	421.43	151988.12	1.029	351.14	305381.91	1.435	348.53	-	-	-	-	-	-
BreastA	5855.21	0.046	33.89	4504.16	0.125	103.88	3774.94	0.249	94.83	4504.16	10.91	2039.08	4504.16	18.25	3755.11
BreastB	1369.81	0.047	11.62	507.93	0.016	161.78	649.91	0.421	127.40	517.93	2.387	1106.28	677.91	13.32	2231.81
Proteins	271311.09	4.493	601.33	149023.50	2.215	541.18	110243.62	1.451	856.71	-	-	-	-	-	-
DLBCLA	18632.00	0.437	69.41	6838.00	0.156	90.27	6022.00	0.326	99.57	6626.00	50.02	22678.49	6368.00	299.16	23221.41
DLBCLB	5477.40	0.234	71.32	7574.40	0.359	240.41	12316.43	0.856	290.81	10267.26	39.76	9796.28	13946.20	44.45	8436.51
Média	165951.83	567.26	648.04	73876.94	678.59	324.42	88647.95	840.61	400.22	5770.72	519.47	10829.65	6493.97	107.69	11747.85

Analisando-se os valores de CRand nas Tabelas 3.9, 3.11 3.13 e 3.15, onde os métodos propostos usam para o cálculo da função custo o  $p$ -medianas tradicional, observa-se que, de forma geral, os resultados não foram tão bons quanto os anteriores, mas alguns estão entre os melhores.

Para a distância City Block (Tabela 3.9) em geral os métodos conseguiram boas soluções se aproximando dos valores dos métodos em comparação. A HPL obteve as melhores soluções para os dados de BreastB com 5 agrupamentos e proteínas com 5 agrupamentos. O método HPL-PR conseguiu uma melhor solução para os dados de Yeast com 4 agrupamentos.

Considerando a distância Euclidiana (Tabela 3.11) a HPL obteve uma melhor solução para os dados de proteínas com 7 agrupamentos. O método HPL-PR surpreendeu obtendo 5 soluções entre as melhores, sendo elas para os dados de Íris com 3 agrupamentos, Yeast com 5 agrupamentos, BreastA com 6 agrupamentos, DLBCLA com 5 agrupamentos e DLBCLB com 3 agrupamentos.

Para a correlação de Pearson (Tabela 3.13), a HPL conseguiu uma melhor solução para Yeast com 7 agrupamentos. A HGC obteve uma solução entre as melhores, para os dados de BreastB com 4 agrupamentos.

Finalmente para a correlação Cosseno (Tabela 3.15), a HSI-PR conseguiu uma melhor solução para Yeast com 7 agrupamentos, a HPL-PR conseguiu uma melhor solução para o dado de BreastB com 4 agrupamentos e HGC-PR com uma melhor solução para o dados de BreastA com 3 agrupamentos.

As Tabelas 3.10, 3.12, 3.14 e 3.16 comparam o valor da função objetivo (FO), o tempo da melhor solução (TM) e o tempo total de execução do algoritmo (TT) para as distâncias *City Block*, Euclidiana e Correlações de Pearson e Cosseno. Assim como comentado anteriormente nos casos das tabelas com a função custo sendo uma clique, os valores mínimos de FO para a HPL, HPL-PR, HGC e HGC-PR são muito próximos, com exceção de HSI-PR, que apresentou valores mais elevados de FO que os outros métodos.



Tabela 3.9 - Comparação do resultados de CRand para a distância City Block ( $p$ -medianas).

Dados	Melhor CRand	GRASP		K-medias		K-medianas		PAM		ILS-CS		HSI-PR		HPL		HPL - PR		HGC		HGC - PR	
		M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand	M.	CRand
Iris	<b>0.818</b>	3	<b>0.818</b>	3	0.717	3	0.717	3	0.772	3	0.818	3	0.757	3	0.758	5	0.765	3	0.758	3	0.759
Yeast	<b>*0.262</b>	7	0.150	7	0.181	6	0.167	7	0.152	7	0.157	5	0.200	5	0.232	4	<b>0.262</b>	-	-	-	-
Breast	<b>0.877</b>	2	0.870	2	0.770	2	0.765	2	0.807	2	<b>0.877</b>	4	0.602	2	0.625	7	0.623	-	-	-	-
BreastA	<b>*0.770</b>	2	0.682	2	0.583	2	0.618	4	0.560	-	-	3	0.691	3	0.763	3	<b>0.770</b>	4	0.705	2	0.660
BreastB	<b>*0.507</b>	7	0.159	3	0.328	3	0.284	2	0.187	-	-	4	0.318	5	<b>0.507</b>	6	0.506	4	0.427	3	0.347
Proteínas	<b>*0.317</b>	5	0.293	8	0.223	7	0.229	3	0.192	5	0.295	4	0.292	5	<b>0.317</b>	5	0.300	-	-	-	-
DLBCLA	<b>0.805</b>	3	0.800	3	<b>0.805</b>	3	0.784	3	0.406	-	-	3	0.689	3	0.722	4	0.741	3	0.767	4	0.740
DLBCLB	<b>0.700</b>	2	<b>0.700</b>	2	0.690	2	0.690	3	0.350	-	-	3	0.487	3	0.672	3	0.582	3	0.680	3	0.608

Tabela 3.10 - Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e do tempo total de execução do algoritmo (T.T.) para a distância City Block ( $p$ -medianas).

Dados	FO	HSI-PR		HPL		HPL - PR		HGC		HGC - PR			
		T. M.	T. T.	T. M.	T. T.	T. M.	T. T.	T. M.	T. T.	T. M.	T. T.		
Iris	180.20	0.078	8.658	0.047	31.88	0.110	13.10	163.00	56.222	10446.06	113.60	103.131	15887.36
Yeast	598.41	6.940	742.99	2.340	1109.40	3.323	1139.49	-	-	-	-	-	-
Breast	5554.00	1.860	201.02	0.609	479.14	0.952	349.89	-	-	-	-	-	-
BreastA	30775.29	0.031	33.22	28586.53	0.125	115.09	29996.46	27515.52	3.713	1750.98	28968.27	83.445	8104.09
BreastB	20208.34	0.016	9.64	18887.47	0.015	54.21	18916.98	17937.39	0.078	237.94	16696.86	5.147	498.09
Proteínas	388491.81	1.904	324.79	367933.78	1.046	811.23	369516.38	-	-	-	-	-	-
DLBCLA	16597337.00	0.078	40.68	16442116.00	0.156	145.45	16135681.00	16510888.00	40.108	14054.16	15821051.02	4.975	60471.51
DLBCLB	83016.11	0.125	63.93	78858.80	0.562	223.20	79690.82	78510.31	54.768	56217.47	61466.38	17.143	1529.40

Tabela 3.11 - Comparação do resultados de CRand para a distância Euclidiana ( $p$ -medianas).

Dados	Melhor CRand		GRASP		K-medias		K-medias		PAM		ILS-CS		HSI-PR		HPL		HPL - PR		HGC		HGC - PR		
	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	
Iris	<b>*0.904</b>	3	0.756	3	0.730	3	0.744	3	0.730	3	0.756	3	0.786	3	0.773	3	<b>0.904</b>	3	0.758	3	0.760	3	0.760
Yeast	<b>*0.229</b>	9	0.150	7	0.170	6	0.173	8	0.143	3	0.150	4	0.203	4	0.219	5	<b>0.229</b>	5	-	-	-	-	-
Breast	<b>0.877</b>	2	<b>0.877</b>	2	0.803	2	0.782	2	0.828	2	0.877	2	0.682	2	0.752	3	0.705	3	0.705	-	-	-	-
BreastA	<b>0.682</b>	2	<b>0.682</b>	2	0.654	2	0.654	2	0.543	2	-	-	0.660	3	0.509	3	0.677	4	0.449	2	0.567	2	0.567
BreastB	<b>*0.506</b>	2	0.314	3	0.286	3	0.260	2	0.187	2	-	-	0.265	7	0.391	6	<b>0.506</b>	4	0.374	2	0.238	2	0.238
Proteínas	<b>*0.355</b>	4	0.322	7	0.322	7	0.313	6	0.250	4	0.322	5	0.316	7	<b>0.355</b>	5	0.334	5	-	-	-	-	-
DLBCLA	<b>*0.516</b>	4	0.408	4	0.309	5	0.365	4	0.276	4	-	-	0.383	4	0.483	5	<b>0.516</b>	4	0.380	3	0.395	3	0.395
DLBCLB	<b>*0.713</b>	4	0.481	2	0.420	3	0.424	3	0.391	-	-	-	0.562	3	0.662	3	<b>0.713</b>	3	0.646	2	0.545	2	0.545

Tabela 3.12 - Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a distância Euclidiana ( $p$ -medianas).

Dados	HSI-PR		HPL		HPL - PR		HGC		HGC - PR		
	FO	T. T.	FO	T. T.	FO	T. T.	FO	T. T.	FO	T. T.	
Iris	99.13	0.078	98.78	0.032	117.26	0.094	40.169	10946.58	69.78	96.735	10435.40
Yeast	322.26	7.877	821.73	2.324	289.33	2.980	942.33	-	-	-	-
Breast	3002.42	2.668	278.19	0.749	3028.57	0.874	364.40	-	-	-	-
BreastA	1655.64	0.062	28.25	1609.29	1989.12	0.266	100.94	1558.44	1300.08	32.807	2402.07
BreastB	839.71	0.016	8.14	734.49	785.82	0.094	115.36	737.66	573.32	6.271	588.32
Proteínas	57155.14	1.716	266.54	1.155	55099.99	1.138	564.14	-	-	-	-
DLBCLA	1417357.88	0.078	29.90	1399234.62	1460584.12	0.421	123.89	1377213.88	1433472.68	9.219	823.24
DLBCLB	4419.76	0.156	51.12	4193.70	4170.77	0.515	188.08	4138.71	3136.74	353.169	30464.29

Tabela 3.13 - Comparação do resultados de CRand para a Correlação de Pearson ( $p$ -medianas).

Dados	Melhor CRand		GRASP		K-medias		K-medias		PAM		ILS-CS		HSI-PR		HPL		HPL - PR		HGC		HGC - PR		
	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	
Iris	<b>0.904</b>	3	0.886	3	0.886	3	0.886	3	0.886	3	0.886	3	0.904	3	0.904	3	0.886	3	0.886	3	0.886	3	0.886
Yeast	<b>*0.213</b>	9	0.131	8	0.135	8	0.133	7	0.145	7	0.133	7	0.190	7	<b>0.213</b>	7	0.202	7	0.202	7	-	-	-
Breast	<b>0.441</b>	3	0.284	2	<b>0.441</b>	2	0.368	2	0.289	3	0.288	3	0.049	7	0.160	5	0.154	5	0.154	5	-	-	-
BreastA	<b>0.705</b>	2	0.692	2	<b>0.705</b>	2	<b>0.705</b>	2	0.635	3	-	-	3	0.463	3	0.505	4	0.580	3	0.505	2	0.496	
BreastB	<b>*0.466</b>	2	0.355	3	0.289	3	0.283	3	0.227	-	-	-	3	0.298	4	0.437	6	0.436	4	<b>0.466</b>	3	0.403	
Proteínas	<b>0.344</b>	4	<b>0.344</b>	7	0.313	7	0.306	6	0.245	4	<b>0.344</b>	4	0.293	7	0.341	6	0.323	6	0.323	6	-	-	-
DLBCLA	<b>0.684</b>	4	0.585	4	0.605	4	<b>0.684</b>	4	0.586	-	-	-	5	0.615	3	0.668	6	0.592	3	0.662	4	0.642	
DLBCLB	<b>0.665</b>	2	0.527	3	<b>0.665</b>	3	0.561	3	0.545	3	-	-	4	0.188	5	0.175	4	0.220	4	0.171	2	0.159	

Tabela 3.14 - Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a Correlação de Pearson ( $p$ -medianas).

Dados	HSI-PR		HPL		HPL - PR		HGC		HGC - PR				
	FO	T. T.	FO	T. T.	FO	T. T.	FO	T. T.	FO	T. T.			
Iris	0.49	8.89	0.45	0.015	0.45	0.125	16.53	0.45	29.718	6261.38	0.453	49.903	5852.17
Yeast	109.16	6.239	81.96	2.527	84.40	4.244	1260.55	-	-	-	-	-	-
Breast	172.03	1.544	172.30	159.61	433.23	191.63	152.91	-	-	-	-	-	-
BreastA	59.95	0.031	39.46	55.85	118.15	58.42	0.218	134.13	55.40	8.658	3650.84	258.742	22255.11
BreastB	11.67	0.016	11.43	10.93	43.79	10.55	0.094	58.46	10.31	0.078	165.56	9.39	4.430
Proteínas	20.86	1.607	341.42	15.32	890.35	15.10	689.50	-	-	-	-	-	-
DLBCLA	10.49	0.063	47.45	10.19	135.58	9.30	170.96	9.87	180.85	30018.36	9.12	6.047	188890.14
DLBCLB	120.18	0.094	74.31	112.04	331.15	116.42	258.77	111.58	16.02	135795.84	115.12	20.435	56496.40

Tabela 3.15 - Comparação do resultados de CRand para a correlação cosseno ( $p$ -medianas).

Dados	Melhor CRand		GRASP		K-medias		K-medias		PAM		ILS-CS		HSI-PR		HPL		HPL-PR		HGC		HGC-PR		
	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	M.	
Iris	<b>0.941</b>	3	<b>0.941</b>	3	0.904	3	0.941	3	0.904	3	0.941	6	0.535	3	0.868	5	0.750	3	0.868	3	0.868	3	0.670
Yeast	<b>0.146</b>	9	0.135	9	0.138	6	0.132	7	<b>0.146</b>	7	0.135	4	<b>0.146</b>	4	0.141	5	0.148	-	-	-	-	-	-
Breast	<b>0.332</b>	3	0.293	4	0.258	3	0.306	3	<b>0.332</b>	3	0.293	4	0.185	5	0.265	3	0.254	-	-	-	-	-	-
BreastA	<b>*0.862</b>	2	0.686	2	0.691	2	0.691	2	0.664	2	0.664	-	0.728	3	0.841	3	0.773	3	<b>0.862</b>	2	<b>0.862</b>	2	0.759
BreastB	<b>*0.417</b>	2	0.314	2	0.269	3	0.264	4	0.239	4	0.239	-	0.196	5	0.344	2	<b>0.417</b>	4	0.284	4	0.284	2	0.195
Proteínas	<b>0.349</b>	4	<b>0.349</b>	7	0.320	6	0.304	6	0.247	4	<b>0.349</b>	4	0.241	7	0.323	4	0.290	-	-	-	-	-	-
DLBCLA	<b>0.678</b>	4	0.605	5	0.642	4	<b>0.678</b>	3	0.547	-	0.547	-	0.236	3	0.452	4	0.434	4	0.375	2	0.375	2	0.055
DLBCLB	<b>0.623</b>	4	0.502	3	0.501	3	<b>0.623</b>	5	0.385	-	0.385	-	0.228	4	0.153	3	0.177	4	0.107	2	0.107	2	0.149

Tabela 3.16 - Comparação do valor da função objetivo (FO), do tempo da melhor solução (T.M.) e tempo total de execução do algoritmo (T.T.) para a correlação ( $p$ -medianas).

Dados	HSI-PR		HPL		HPL-PR		HGC		HGC-PR				
	FO	T. T.	FO	T. T.	FO	T. T.	T. M.	T. T.	T. M.	T. T.			
Iris	141.42	18.00	143.84	0.094	141.96	1.544	397.56	140.94	3.308	940.59	97.18	3.182	1497.23
Yeast	413.51	15.319	1584.96	4.633	3915.56	396.08	5.508	26901.06	-	-	-	-	-
Breast	678.56	3.479	362.06	3.496	903.41	689.48	1.669	770.39	-	-	-	-	-
BreastA	93.81	0.031	32.23	0.064	100.21	94.75	0.172	92.97	91.74	0.546	299.97	91.77	86.471
BreastB	16.99	2.684	275.06	43.99	108.28	46.99	0.078	25.78	40.99	0.078	1751.78	16.99	2.684
Proteínas	694.00	2.620	397.27	691.00	1238.17	694.00	2.808	1242.75	-	-	-	-	-
DLBCLA	139.00	0.141	45.34	138.00	147.38	137.00	0.352	168.75	133.00	0.187	6557.29	139.00	64.756
DLBCLB	176.89	0.109	60.20	175.84	282.72	176.87	0.765	301.58	171.86	0.780	656.68	162.86	58.047

As Tabelas de diferenças de 3.17 a 3.20 apresentam a comparação dos valores de CRand para os métodos propostos neste trabalho para as duas formas de cálculo da função custo na solução do problema de agrupamentos: a somatória entre todos os pontos dos grupos (clique) e a somatória de todos os pontos a mediana mais próxima (p-mediana tradicional), além disso, são mostrados as diferenças entre eles. Analisando os resultados pode observar que para a maioria das soluções o resultado foi melhor quando se usou a somatória da clique.

Para a distância *city block* usando o somatório da clique para a função custo a HSI-PR foi melhor em seis instâncias (Íris, Breast, BreastA, BreastB, proteínas e DLBCLB), a HPL foi melhor em cinco instâncias (Íris, Breast, BreastB, DLBCLA e DLBCLB), a HPL-PR foi melhor em seis instâncias (Íris, Breast, BreastB, proteínas, DLBCLA e DLBCLB), a HGC foi melhor em cinco instâncias (Íris, BreastA, BreastB, DLBCLA e DLBCLB), a HGC-PR foi melhor em quatro instâncias (Íris, BreastA, BreastB e DLBCLB).

Para a distância euclidiana com clique a HSI-PR foi melhor em sete instâncias (Íris, Breast, BreastA, BreastB, proteínas, DLBCLA e DLBCLB), a HPL foi melhor em seis instâncias (Íris, Breast, BreastA, BreastB, DLBCLA e DLBCLB), a HPL-PR foi melhor em cinco instâncias (Breast, BreastA, BreastB, proteínas, e DLBCLA), a HGC foi melhor em cinco instâncias (Íris, BreastA, BreastB, DLBCLA e DLBCLB), a HGC-PR foi melhor em cinco instâncias (Íris, BreastA, BreastB, DLBCLA e DLBCLB).

Para a correlação de Pearson a com clique a HSI-PR foi melhor em sete instâncias (Yeast, Breast, BreastA, BreastB, proteínas, DLBCLA e DLBCLB), a HPL foi melhor em seis instâncias (Íris, BreastA, BreastB, Proteínas, DLBCLA e DLBCLB), a HPL-PR foi melhor em seis instâncias (Íris, Breast, BreastA, BreastB, proteínas e DLBCLA), a HGC foi melhor em quatro instâncias (Íris, BreastA, BreastB e DLBCLA), a HGC-PR foi melhor em quatro instâncias (Íris, BreastA, BreastB e DLBCLA).

Para a correlação Cosseno a com clique a HSI-PR foi melhor em seis instâncias (Íris, Yeast, Breast, BreastB, proteínas e DLBCLA), a HPL foi melhor em quatro instâncias (yeast, BreastB, Proteínas e DLBCLA), a HPL-PR foi melhor em sete instâncias (Íris, Yeast, Breast, BreastB, proteínas, DLBCLA e DLBCLB), a HGC foi melhor em quatro instâncias (Íris, BreastB, DLBCLA e DLBCLB), a HGC-PR foi melhor em três instâncias (Íris, BreastB e DLBCLA).

Tabela 3.17 - Diferenças para a distância City Block.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)
Iris	0.904	0.757	19.42	0.922	0.758	21.64	0.922	0.765	20.52	0.904	0.758	19.26	0.904	0.759	19.10
Yeast	0.183	0.200	-8.50	0.205	0.232	-11.64	0.193	0.262	-26.34	-	-	-	-	-	-
Breast	0.657	0.602	9.14	0.691	0.625	10.56	0.706	0.623	13.32	-	-	-	-	-	-
BreastA	0.740	0.691	7.09	0.722	0.763	-5.37	0.748	0.770	-2.86	0.722	0.705	2.41	0.722	0.660	9.39
BreastB	0.440	0.318	38.36	0.508	0.507	0.20	0.680	0.506	34.39	0.570	0.427	33.49	0.590	0.347	70.03
Proteínas	0.352	0.292	20.55	0.295	0.317	-6.94	0.305	0.300	1.67	-	-	-	-	-	-
DLBCLA	0.659	0.689	-4.35	0.742	0.722	2.77	0.795	0.741	7.29	0.781	0.767	1.83	0.674	0.740	-8.92
DLBCLB	0.693	0.487	42.30	0.702	0.672	4.46	0.702	0.582	20.62	0.702	0.680	3.24	0.702	0.608	15.46

Tabela 3.18 - Diferenças para a distância Euclidiana.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)	Clique Grand	P-mediana CRand	Melhorias (%)
Iris	0.886	0.786	12.72	0.850	0.773	9.96	0.886	0.904	-1.99	0.886	0.758	16.89	0.868	0.760	14.21
Yeast	0.173	0.203	-14.78	0.198	0.219	-9.59	0.195	0.229	-14.85	-	-	-	-	-	-
Breast	0.746	0.682	9.38	0.757	0.752	0.66	0.772	0.705	9.50	-	-	-	-	-	-
BreastA	0.706	0.660	6.97	0.571	0.509	12.18	0.716	0.677	5.76	0.560	0.449	24.72	0.610	0.567	7.58
BreastB	0.466	0.265	75.85	0.496	0.391	26.85	0.687	0.506	35.77	0.513	0.374	37.17	0.590	0.238	147.90
Proteínas	0.353	0.316	11.71	0.350	0.355	-1.41	0.387	0.334	15.87	-	-	-	-	-	-
DLBCLA	0.440	0.383	14.88	0.512	0.483	6.00	0.597	0.516	15.70	0.500	0.380	31.58	0.508	0.395	28.61
DLBCLB	0.595	0.562	5.87	0.689	0.662	4.08	0.689	0.713	-3.37	0.689	0.646	6.66	0.689	0.545	26.42

Tabela 3.19 - Diferenças para a correlação de Pearson.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %
Iris	0.904	0.904	0.00	0.922	0.904	1.99	0.922	0.886	4.06	0.922	0.886	4.06	0.922	0.886	4.06
Yeast	0.195	0.190	2.63	0.204	0.213	-4.23	0.186	0.202	-7.92	-	-	-	-	-	-
Breast	0.145	0.049	195.92	0.142	0.160	-11.25	0.248	0.154	61.04	-	-	-	-	-	-
BreastA	0.660	0.463	42.55	0.587	0.505	16.24	0.650	0.580	12.07	0.573	0.505	13.47	0.587	0.496	18.35
BreastB	0.480	0.298	61.07	0.522	0.437	19.45	0.681	0.436	56.19	0.659	0.466	41.42	0.661	0.403	64.02
Proteínas	0.341	0.293	16.38	0.350	0.341	2.64	0.353	0.323	9.29	-	-	-	-	-	-
DLBCLA	0.709	0.615	15.28	0.720	0.668	7.78	0.761	0.592	28.55	0.791	0.662	19.49	0.685	0.642	6.70
DLBCLB	0.230	0.188	22.34	0.176	0.175	0.57	0.183	0.220	-16.82	0.106	0.171	-38.01	0.117	0.159	-26.42

Tabela 3.20 - Diferenças para a correlação Cosseno.

Dados	HSI-PR			HPL			HPL-PR			HGC			HGC-PR		
	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %	Clique Grand	P-medianas CRand	Melhorias %
Iris	0.868	0.535	62.24	0.868	0.868	0.00	0.884	0.750	17.87	0.886	0.868	2.07	0.885	0.670	32.09
Yeast	0.179	0.146	22.60	0.213	0.141	51.06	0.195	0.148	31.76	-	-	-	-	-	-
Breast	0.277	0.185	49.73	0.196	0.265	-26.04	0.279	0.254	9.84	-	-	-	-	-	-
BreastA	0.638	0.728	-12.36	0.638	0.841	-24.14	0.665	0.773	-13.97	0.638	0.862	-25.99	0.638	0.759	-15.94
BreastB	0.430	0.196	119.39	0.473	0.344	37.50	0.440	0.417	5.52	0.604	0.284	112.68	0.444	0.195	127.69
Proteínas	0.339	0.241	40.66	0.347	0.323	7.43	0.321	0.290	10.69	-	-	-	-	-	-
DLBCLA	0.621	0.236	163.14	0.669	0.452	48.01	0.685	0.434	57.83	0.664	0.375	77.07	0.619	0.055	1.025.45
DLBCLB	0.159	0.228	-30.26	0.123	0.153	-19.61	0.185	0.177	4.52	0.111	0.107	3.74	0.111	0.149	-25.50

Os gráficos a seguir são referentes as Tabelas de 3.17 a 3.20 e apresentam a comparação entre os valores de CRand para os métodos propostos neste trabalho, usando clique e  $p$ -medianas no cálculo da função custo. Nos gráficos os eixos das abcissas são os dados (1-Íris, 2-Yeast, 3-Breast, 4-BreastA, 5-BreastB, 6-Proteínas, 7-DLBCLA e 8-DLBCLB) e os eixos das ordenadas são os valores de crand. Os valores zero para 2-Yeast, 3-Breast e 6-Proteínas, indicam que o algoritmo não obteve soluções para estes dados.

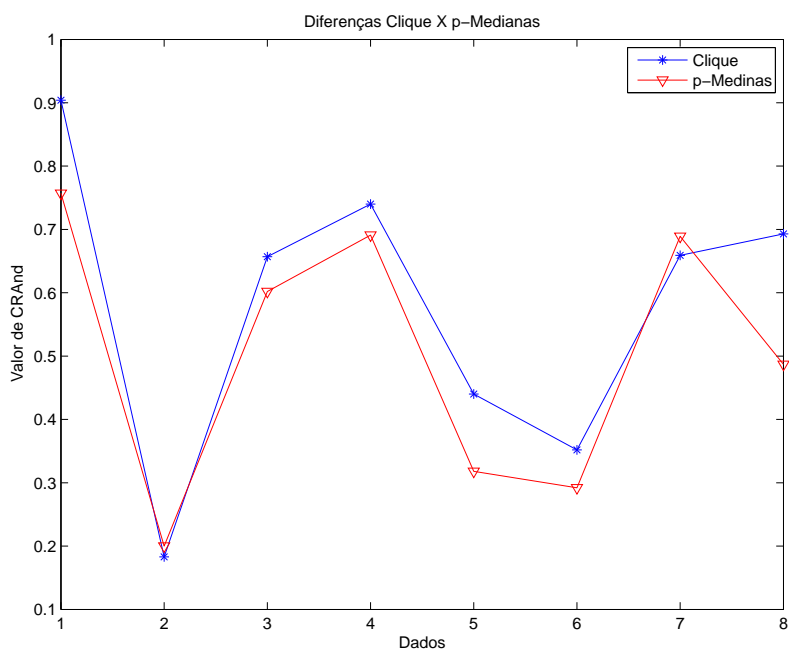


Figura 3.4 - Gráfico das diferenças - City Block (HSI-PR)



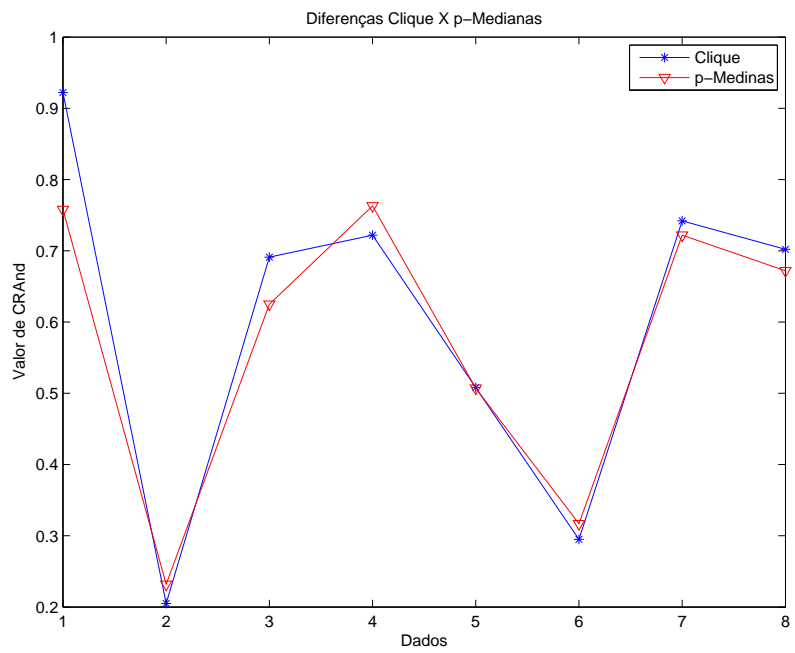


Figura 3.5 - Gráfico das diferenças - City Block (HPL)

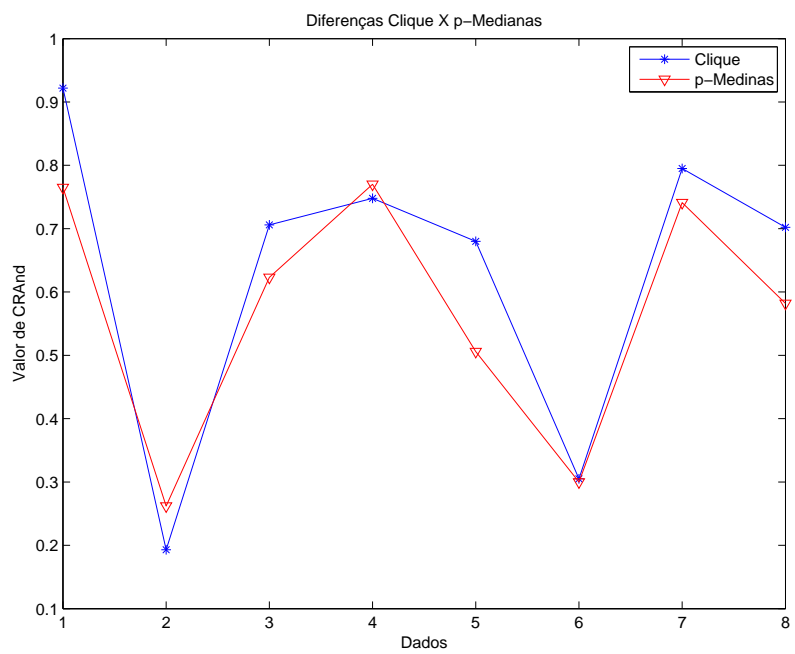


Figura 3.6 - Gráfico das diferenças - City Block (HPL-PR)

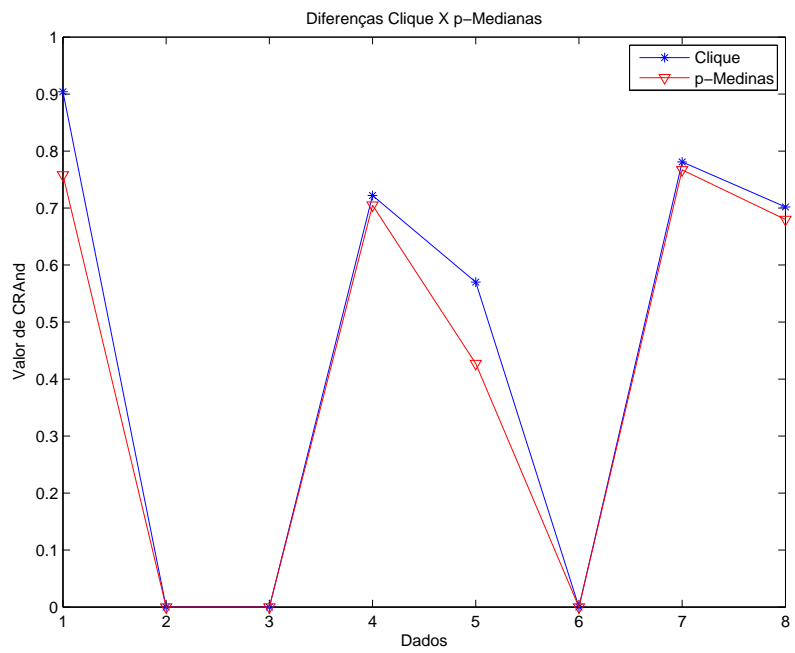


Figura 3.7 - Gráfico das diferenças - City Block (HGC)

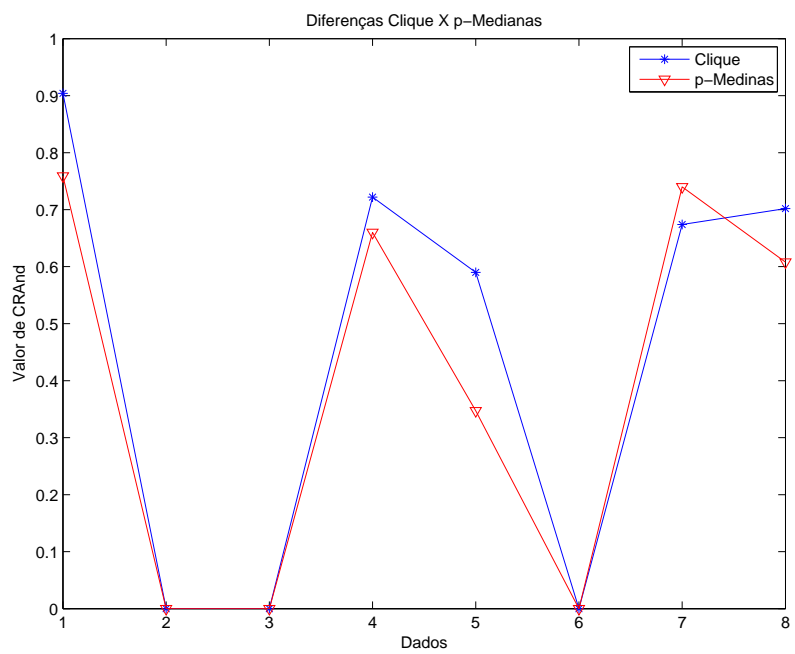


Figura 3.8 - Gráfico das diferenças - City Block (HGC-PR)

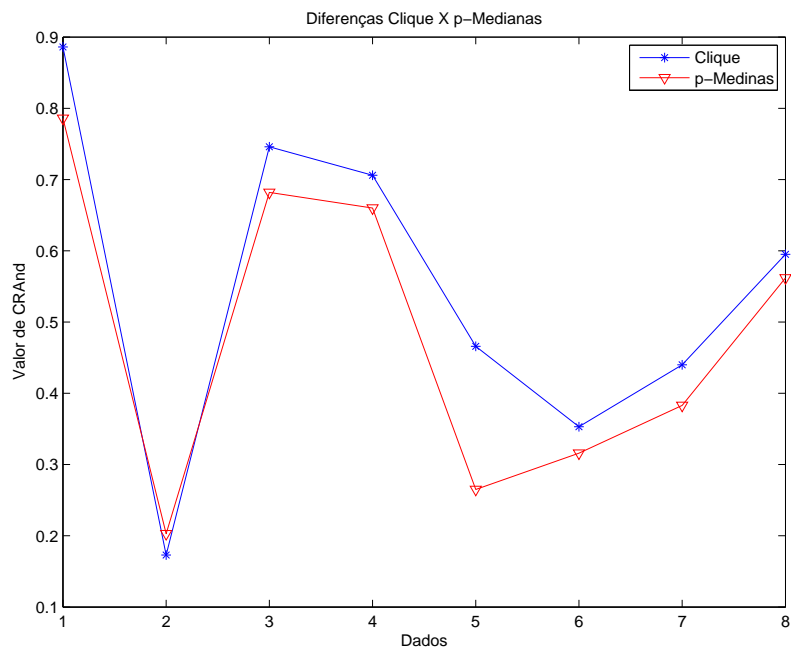


Figura 3.9 - Gráfico das diferenças - Euclidiana (HSI-PR)

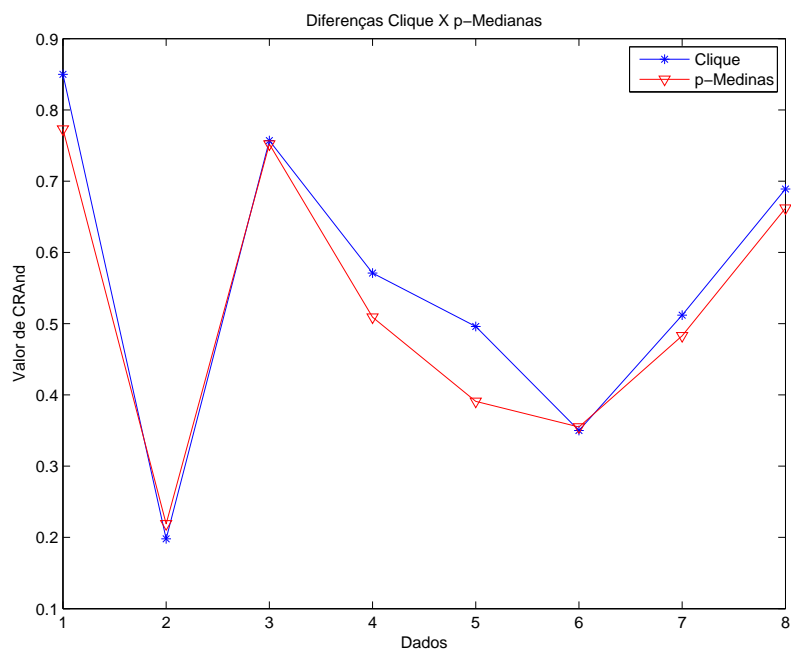


Figura 3.10 - Gráfico das diferenças - Euclidiana (HPL)

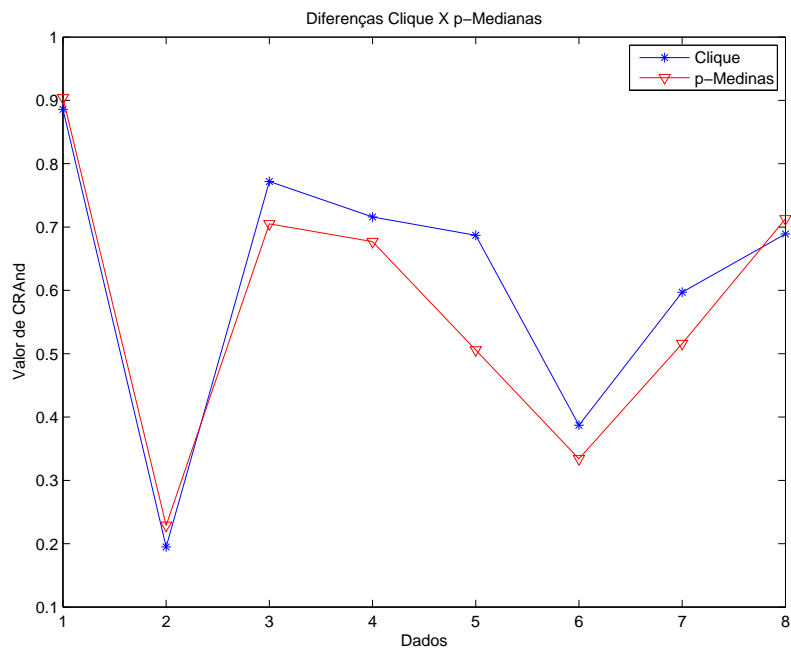


Figura 3.11 - Gráfico das diferenças - Euclidiana (HPL-PR)

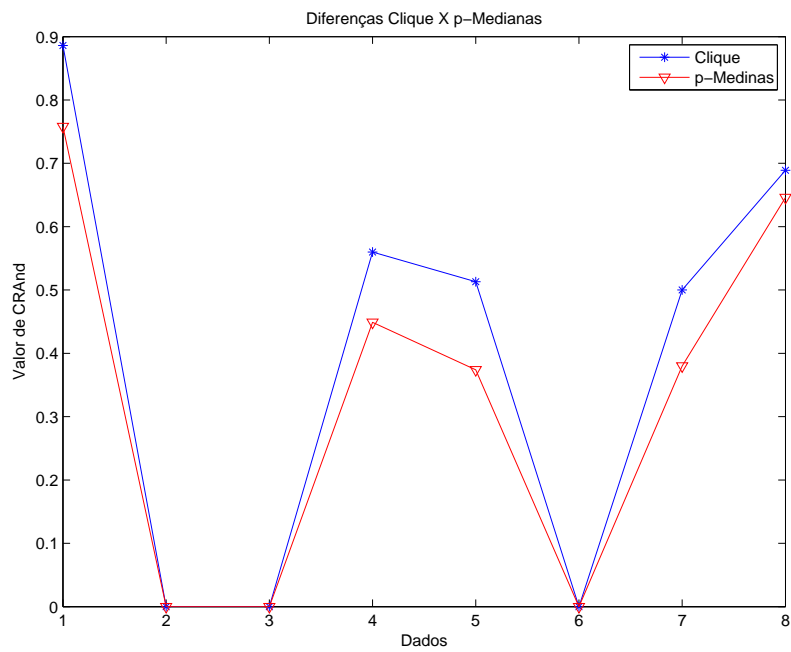


Figura 3.12 - Gráfico das diferenças - Euclidiana (HGC)

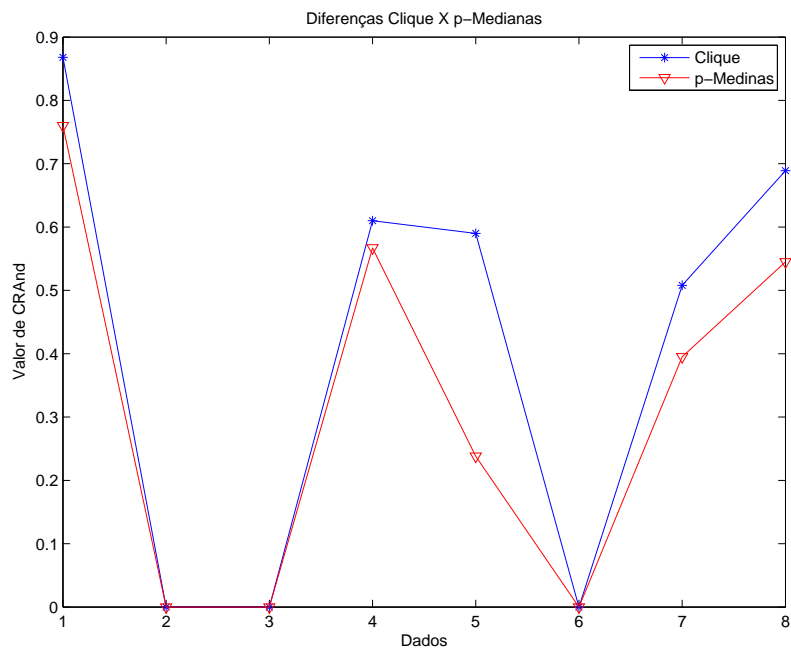


Figura 3.13 - Gráfico das diferenças - Euclidiana (HGC-PR)

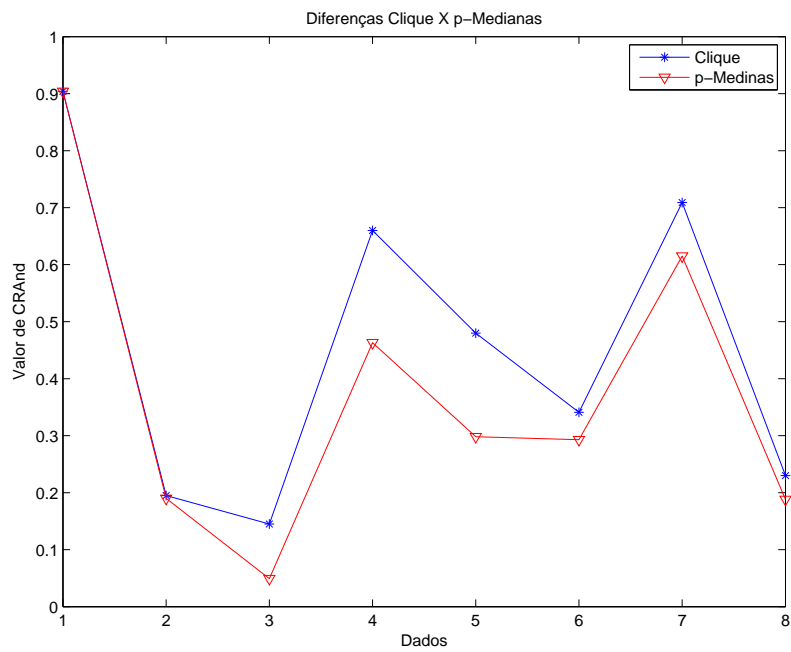


Figura 3.14 - Gráfico das diferenças - Pearson (HSI-PR)

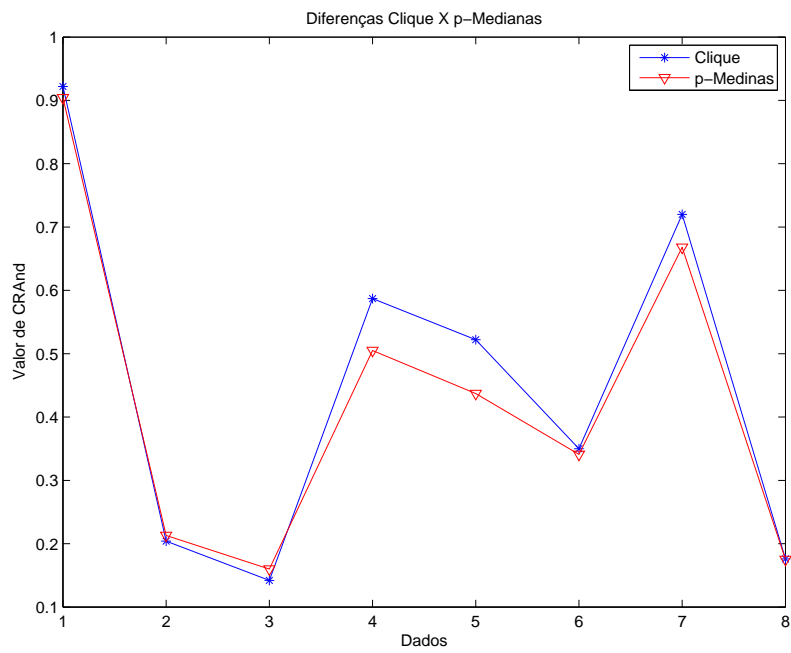


Figura 3.15 - Gráfico das diferenças - Pearson (HPL)

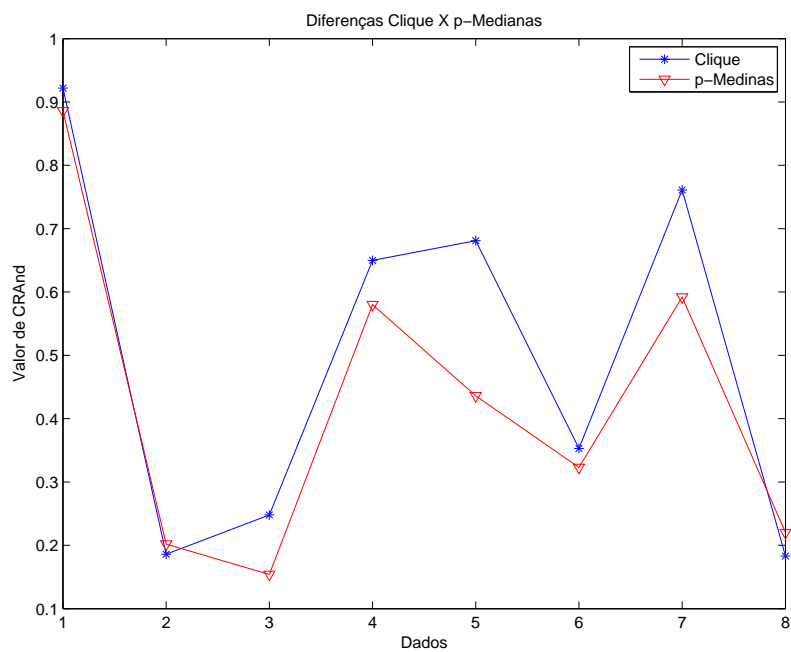


Figura 3.16 - Gráfico das diferenças - Pearson (HPL-PR)

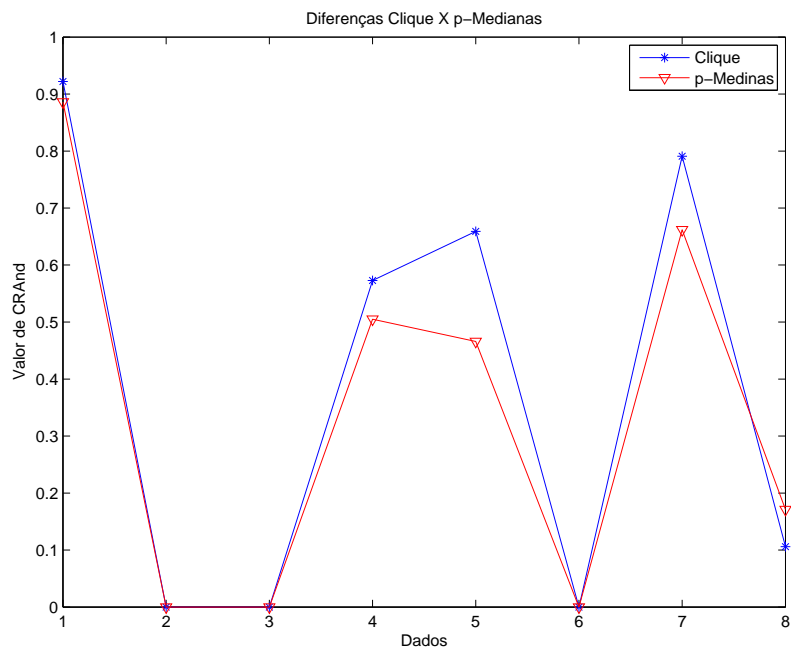


Figura 3.17 - Gráfico das diferenças - Pearson (HGC)

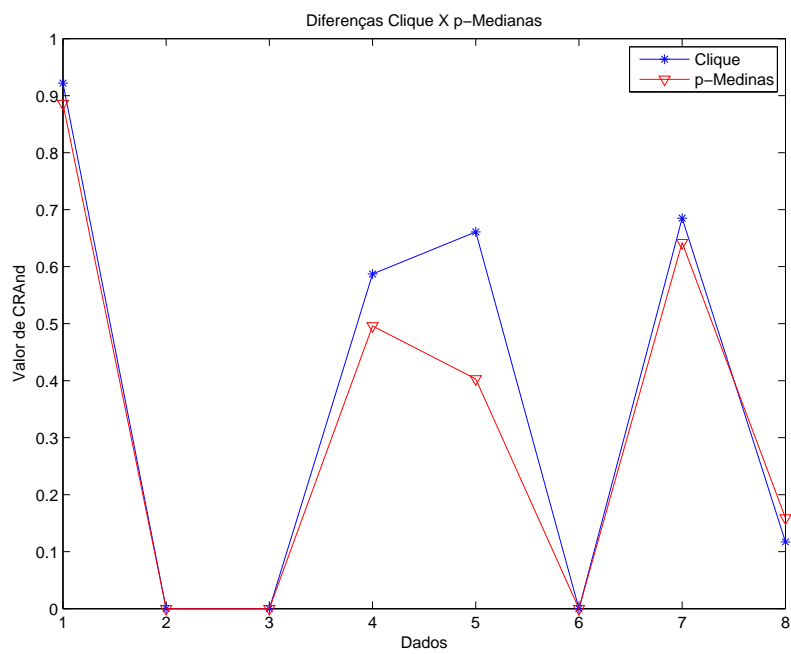


Figura 3.18 - Gráfico das diferenças - Pearson (HGC-PR)

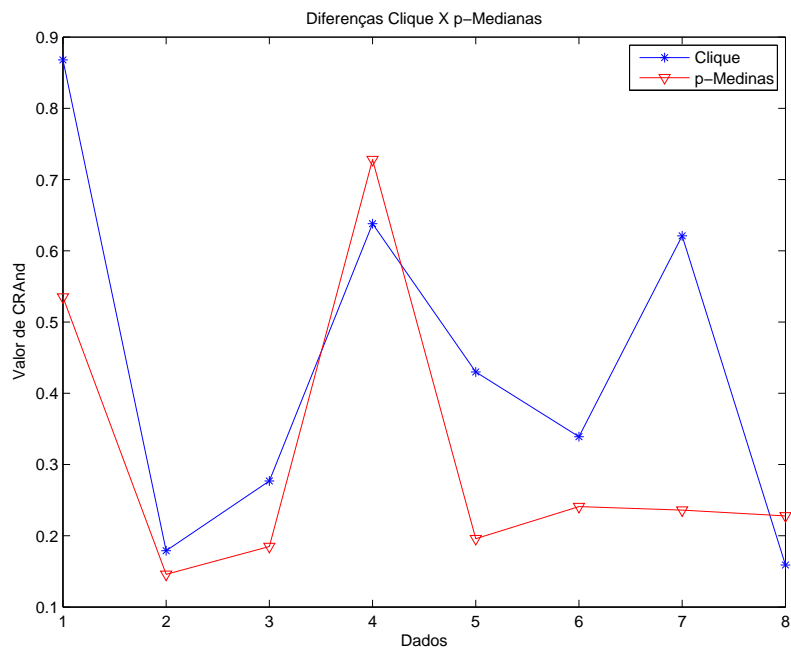


Figura 3.19 - Gráfico das diferenças - Cosseno (HSI-PR)

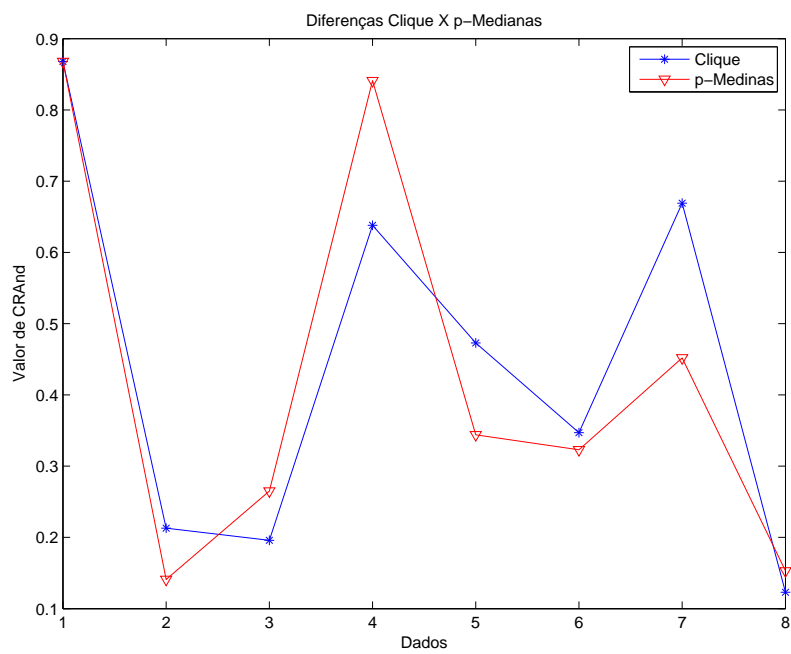


Figura 3.20 - Gráfico das diferenças - Cosseno (HPL)



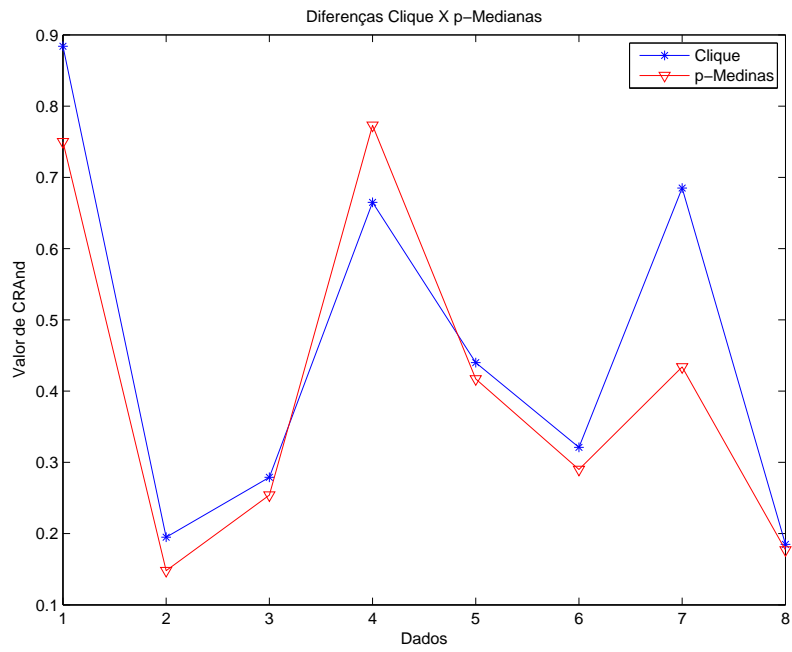


Figura 3.21 - Gráfico das diferenças - Cosseno (HPL-PR)

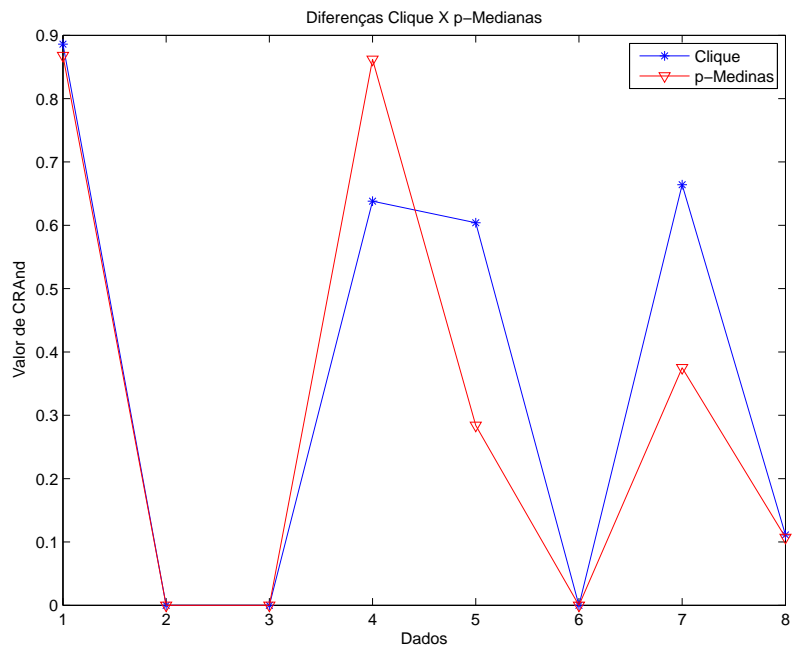


Figura 3.22 - Gráfico das diferenças - Cosseno (HGC)

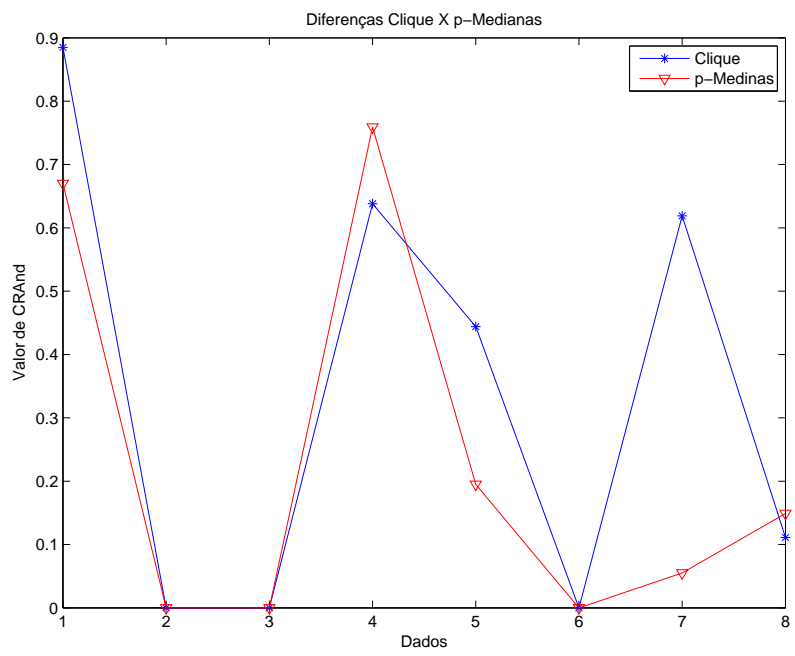


Figura 3.23 - Gráfico das diferenças - Cosseno (HGC-PR)

### 3.5 Conclusões sobre problemas de agrupamentos

Problemas de agrupamentos aparecem em vários contextos e aplicações reais. Esta tese apresenta uma contribuição para a solução de problemas de agrupamento utilizando heurísticas híbridas que combinam etapas de um processo de geração de colunas para resolver problemas de  $p$ -medianas.

Cinco diferentes abordagens foram propostas, considerando-se as soluções a partir das colunas iniciais, do problema mestre e da geração de colunas. As abordagens são combinadas com o PR e os resultados são avaliados pelo índice CRand. As heurísticas mostram-se adequadas e eficientes na localização e separação dos dados. Os resultados computacionais considerando as distâncias *City Block* e Euclidiana foram melhores do que os obtidos quando se considera as correlações de Pearson e cosseno, como demonstrado em (NASCI-MENTO et al., 2010).

Os resultados usando o somatório da clique para o cálculo da função custo foi melhor quando comparado com a  $p$ -mediana tradicional. Isso fica evidente nas colunas de diferenças nas Tabelas de 3.17 a 3.20 e nos gráficos das Figuras de 3.4.2 a 3.4.2.

O uso do PR melhora as soluções na maioria dos casos. A busca de um caminho intermediário entre as soluções de  $p$ -medianas iniciais e guia pode alcançar melhores soluções de acordo com índice CRand.

Espera-se que as heurísticas híbridas propostas possam ser úteis a outros problemas de agrupamentos em aplicações gerais ou mais específicos de classificação de dados.



## 4 CONCLUSÕES FINAIS E TRABALHOS FUTUROS

Este capítulo apresenta as conclusões finais para as heurísticas utilizadas para soluções do PAB, além do problema de agrupamentos. De forma geral, as heurísticas apresentadas foram satisfatórias cumprindo os objetivos propostos.

A modelagem utilizada para resolver o PAB foi uma versão relaxada de Mauri et al. (2008a) que foi elaborada para um problema de roteamento de veículos com garagens múltiplas e janelas de tempo em Cordeau e Laporte (2005). Para o PAB, foi aplicado o método híbrido *clustering search* com *simulated annealing* buscando minimizar o tempo de serviço dos navios no porto.

Foram utilizadas 30 instâncias distintas, cada uma com 60 navios e 13 berços. Essas instâncias são baseadas em dados do porto de Gioia Tauro (Itália). Para o PAB abordou-se os casos discreto e contínuo. Em ambos os casos, as soluções obtidas foram as melhores e com tempo computacional menor que os demais métodos em comparação. Além disso, no caso discreto as soluções obtidas foram as ótimas, provadas por Buhrkal et al. (2009). No caso contínuo apesar de não ser possível afirmar a otimalidade das soluções, os valores encontrados são um bom limitante inferior para as instâncias analisadas.

Para o problema de agrupamentos o objetivo foi separar um conjunto de dados (descritos em 3.4.1) em grupos, cujos objetos pertencentes ao mesmo grupo fossem similares. A configuração dos grupos foram obtidos pela solução do problema  $p$ -medianas pela técnica geração de colunas. Para este problema de  $p$ -medianas além de usar o cálculo do custo na função objetivo ( $c_k$  apresentado em 3.5) como mostrado na formulação tradicional de  $p$ -medianas em 3.5, analisou-se também quando  $c_k$  era definido a partir de uma clique. Assim, como é possível observar nas Tabelas de diferenças de 3.17 a 3.20 e nos gráficos das Figuras de 3.4.2 a 3.4.2, quando se usou a clique na função custo ( $c_k$ ) os resultados foram mais favoráveis.

Inicialmente para o problema de agrupamentos o objetivo era aplicar o algoritmo geração de colunas, mas no decorrer dos testes com o algoritmo, decidiu-se analisar as soluções intermediárias em cada fase desta técnica, o que deu origem as cinco heurísticas propostas: uma solução viável a partir do problema mestre, a solução de geração de colunas, uma heurística com *path-relinking* considerando as colunas iniciais do processo de geração de colunas, uma solução do problema mestre com *path-relinking* e o processo de geração de colunas com *path-relinking*.

De modo geral, os resultados obtidos para todos os métodos propostos, demonstraram que podem ser usados como uma nova ferramenta para soluções dos problemas abordados nesta

tese, cuja competitividade e até superioridade foi demonstrada claramente na maioria dos casos analisados.

Enfim, no Apêndice A são listados os trabalhos publicados desta tese.

#### **4.1 Trabalhos futuros**

Apesar dos bons resultados apresentados por este trabalho, os métodos utilizados podem ser melhorados. Assim, seguem algumas sugestões para trabalhos futuros.

Para solução do PAB e problema de agrupamento estudados nesta tese, propõem-se a junção das duas técnicas propostas, o *clustering search* com geração de colunas. O algoritmo de geração de colunas poderia substituir o *simulated annealing* na criação de soluções iniciais. Soluções estas que seriam encaminhadas ao processo de agrupamento do *clustering search*. Com isto, espera-se que haja uma melhoria na criação dos grupos promissores, facilitando a atuação do algoritmo de busca local.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABBASI, A. A.; YOUNIS, M. A survey on clustering algorithms for wireless sensor networks. **Comput. Commun.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 30, n. 14-15, p. 2826–2841, 2007. 43, 52
- AL-SULTAN, K. S. A tabu search approach to the clustering problem. **Pattern Recognition**, Elsevier, v. 28, n. 9, p. 1443–1451, 1995. 45
- BARROS, V. H.; COSTA, T. S.; OLIVEIRA, A. C. M.; LORENA, L. A. N. Model and heuristic for berth allocation in tidal bulk ports with stock level constraints. **Computers & Industrial Engineering**, v. 60, n. 4, p. 606–613, 2011. 15
- BERKHIN, P. A survey of clustering data mining techniques. **Grouping Multidimensional Data**, p. 25–71, 2006. 2, 43, 45
- BIERWIRTH, C.; MEISEL, F. A survey of berth allocation and quay crane scheduling problems in container terminals. **European Journal of Operational Research**, v. 202, n. 3, p. 615–627, 2009. 7, 8, 9, 10
- BROWN, G. G.; CORMICAN, K. J.; LAWPHONGPANICH, S.; WIDDIS, D. B. Optimizing submarine berthing with a persistence incentive. **Naval Research Logistics**, v. 44, p. 301–318, 1997. 11
- BROWN, G. G.; LAWPHONGPANICH, S.; THURMAN, K. P. Optimizing ship berthing. **Naval Research Logistics**, v. 41, p. 1–15, 1994. 11
- BUHRKAL, K.; ZUGLIAN, S.; ROPKE, S.; LARSEN, J.; LUSBY, R. Models for the discrete berth allocation problem: A computational comparison. **Transportation Research**, v. 47, n. 4, p. 461 – 473, 2009. 5, 13, 36, 83
- CHANG, D.-X.; ZHANG, X.-D.; ZHENG, C.-W. A genetic algorithm with gene rearrangement for k-means clustering. **Pattern Recognition**, Elsevier Science Inc., New York, NY, USA, v. 42, p. 1210–1222, July 2009. ISSN 0031-3203. 45
- CHAVES, A. A. **Metaheurísticas híbridas com busca por agrupamentos para problemas de otimização combinatória**. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2009. 23, 24, 25, 33
- CHEONG, C.; TAN, K.; LIU, D.; LIN, C. Multi-objective and prioritized berth allocation in container ports. **Springer**, v. 180, p. 63–103, 2010. 14

- CORDEAU, J.-F.; LAPORTE, G. Models and tabu search heuristics for the berth allocation problem. **Transportation Science**, v. 39, n. 4, p. 525–538, 2005. 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 22, 23, 36, 38, 42, 83
- ELWANY, M. H.; ALI, I.; ABOUELSEOUD, Y. A heuristics-based solution to the continuous berth allocation and crane assignment problem. **Alexandria Engineering Journal**, v. 52, n. 4, p. 671 – 677, 2013. 1
- FILHO, G. R.; LORENA, L. A. N. **Constructive Genetic Algorithm and Column Generation: an Application to Graph Coloring**. 2000. 49
- FILHO, G. R.; LORENA, L. A. N. Uma heurística de geração de colunas para o problema de formação de células de máquinas e partes. **Pesquisa Operacional para o Desenvolvimento**, v. 2, p. 188–202, 2010. 46
- GIALLOMBARDO, G.; MOCCIA, L.; SALANI, M.; VACCA, I. Modeling and solving the tactical berth allocation problem. **Transportation Research Part B**, v. 44, n. 2, p. 232–245, 2010. 14
- GLOVER, F.; LAGUNA, M. **Tabu Search**. Norwell, MA, USA: Kluwer Academic Publishers, 1997. 2
- GOLDBERG, D. E. **Genetic Algorithms in search, optimization and machine learning**. 1st. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989. 2
- GUAN, Y.; XIAO, W.-Q.; CHEUNG, R. K.; LI, C.-L. A multiprocessor task scheduling model for berth allocation: heuristic and worst-case analysis. **Operations Research Letters**, v. 30, p. 343–350, 2002. 11, 12
- HAMMING, R. W. Error detecting and error correcting codes. **Bell System Technical Journal**, v. 26, n. 2, p. 147–160, 1950. 25
- HANDL, J.; KNOWLES, J.; KELL, D. B. Computational cluster validation in post-genomic data analysis. **Bioinformatics**, Oxford University Press, Oxford, UK, v. 21, p. 3201–3212, August 2005. ISSN 1367-4803. 44
- HANSEN, P.; OGUZ, C.; MLADENOVIC, N. Variable neighborhood search for minimum cost berth allocation. **European Journal of Operational Research**, v. 191, n. 3, p. 636–649, 2008. 5, 6
- HONG, Y.; KWONG, S.; CHANG, Y.; REN, Q. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. **Pattern Recognition**, v. 41, n. 9, p. 2742–2756, 2008. 45



- HUBERT, L.; ARABIE, P. Comparing partitions. **Journal of Classification**, Springer New York, v. 2, n. 1, p. 193–218–218, dez. 1985. 43, 49
- ILOG. Ilog cplex 12.1. 2009. 45, 48
- IMAI, A.; CHEN, H. C.; NISHIMURA, E.; PAPADIMITRIOU, S. The simultaneous berth and quay crane allocation problem. **Transportation Research Part E**, v. 44, p. 900–920, 2008. 13
- IMAI, A.; NISHIMURA, E.; PAPADIMITRIOU, S. The dynamic berth allocation problem for a container port. **Transportation Research Part B**, v. 35, p. 401–417, 2001. 7, 9, 11, 12
- \_\_\_\_\_. Berth allocation with service priority. **Transportation Research**, v. 37B, p. 437–457, 2003. 11
- IMAI, A.; SUN, X.; NISHIMURA, E.; PAPADIMITRIOU, S. Berth allocation in a container port: using a continuous location space approach. **Transportation Research Part B**, v. 39, n. 3, p. 199–221, 2005. 9, 10, 12, 13
- KIRKPATRICK, S.; GELLAT, D. C.; VECCHI, M. P. Optimization by simulated annealing. **Science**, v. 220, p. 671–680, 1983. 2
- LI, C.; CAI, X.; LEE, C. Scheduling with multiple-job-on-one-processor pattern. **IIE Transactions**, v. 30, n. 5, p. 433–445, 1983. 15
- LIM, A. The berth scheduling problem. **Operations Research Letters**, v. 22, p. 105–110, 1998. 15
- MASSEN, F.; LÓPEZ-IBÁÑEZ, M.; STUTZLE, T.; DEVILLE, Y. **Experimental Analysis of Pheromone-Based Heuristic Column Generation Using irace**. [S.l.]: Springer Berlin Heidelberg, 2013. 92-106 p. (Lecture Notes in Computer Science, v. 7919). 49
- MAULIK, U.; BANDYOPADHYAY, S. Genetic algorithm-based clustering technique. **Pattern Recognition**, v. 33, n. 9, p. 1455–1465, set. 2000. 44
- MAURI, G.; LORENA, L. A. N. Uma nova abordagem para o problema dial-a-ride. v. 19, n. 1, p. 41–54, 2009. 33
- MAURI, G. R. **Novas abordagens para representação e obtenção de limitantes e soluções para alguns problemas de otimização combinatória**. 241 p. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2008. 27, 31, 32

MAURI, G. R.; ANDRADE, L. N. de; LORENA, L. A. N. A memetic algorithm for a continuous case of the berth allocation problem. In: . [S.l.]: SciTePress, 2011. p. 105–113. 15, 21, 28, 38, 42

MAURI, G. R.; LORENA, L. A. N. A new hybrid heuristic for driver scheduling. **International Journal of Hybrid Intelligent Systems**, v. 4, p. 39–47, 2007. 49

MAURI, G. R.; OLIVEIRA, A. C. M.; LORENA, L. A. N. hybrid column generation approach for the berth allocation problem. v. 4972, p. 110–122, 2008b. 15, 33, 36

MAURI, G. R.; OLIVEIRA, A. C. M. de; LORENA, L. A. N. Heurística baseada no simulated annealing aplicada ao problema de alocação de berços. **Gestão da Produção, Operações e Sistemas**, v. 3, n. 1, p. 113–127, 2008a. 15, 17, 19, 20, 29, 33, 36, 83

MEISEL, F.; BIERWIRTH, C. Heuristics for the integration of crane productivity in the berth allocation problem. **Transportation Research Part E**, v. 45, p. 196–209, 2009. 14

MITRA, S.; BANKA, H. Multi-objective evolutionary biclustering of gene expression data. **Pattern Recogn.**, Elsevier Science Inc., New York, NY, USA, v. 39, p. 2464–2477, December 2006. 45

MONACO, M. F.; SAMMARRA, M. The berth allocation problem: A strong formulation solved by a lagrangean approach. **Transportation Research Part B**, v. 41, n. 2, p. 265–280, 2007. 7

NASCIMENTO, M. C. V.; CARVALHO, A. C. P. L. F. D. Spectral methods for graph clustering - a survey. **European Journal Of Operational Research**, v. 211, n. 2, p. 221–231, 2011. 45

NASCIMENTO, M. C. V.; TOLEDO, F. M. B.; CARVALHO, A. C. P. L. F. de. Investigation of a new grasp-based clustering algorithm applied to biological data. **Comput. Oper. Res.**, Elsevier Science Ltd., v. 37, n. 8, p. 1381–1388, ago. 2010. ISSN 0305-0548. 44, 45, 55, 56, 81

NISHIMURA, E.; IMAI, A.; PAPADIMITRIOU, S. Berth allocation planning in the public berth system by genetic algorithms. **European Journal of Operational Research**, v. 131, p. 282–292, 2001. 11

OLIVEIRA, A. C. M. **Algoritmos evolutivos híbridos com detecção de regiões promissoras em espaços de busca contínuo e discreto**. Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2004. 24

- PEREIRA, M. A.; LORENA, L. A. N.; SENNE, E. L. F. A column generation approach for the maximal covering location problem. **International Transactions in Operational Research**, Blackwell Publishing Ltd, v. 14, n. 4, p. 349–364, 2007. ISSN 1475-3995. Disponível em:  
<<http://dx.doi.org/10.1111/j.1475-3995.2007.00593.x>>. 52
- PIRKWIESER, S.; RAIDL, G. **Multilevel Variable Neighborhood Search for Periodic Routing Problems**. [S.l.]: Springer Berlin Heidelberg, 2010. 226-238 p. (Lecture Notes in Computer Science, v. 6022). ISBN 978-3-642-12138-8. 49
- RAND, W. M. Objective criteria for the evaluation of clustering methods. **Journal of the American Statistical Association**, v. 66, n. 336, p. 846–850, 1971. 43, 44
- REEVES, C. **Modern Heuristic Techniques for Combinatorial Problems**. [S.l.]: McGraw-Hill Ryerson, Limited, 1995. (Advanced Topics in Computer Science Series). 2
- RESENDE, M. G. C.; RIBEIRO, C. C. Grasp with path-relinking: Recent advances and applications. In: **Metaheuristics: Progress as Real Problem Solvers**. [S.l.]: Springer, 2005. p. 29–63. 49
- RIOS, L. R.; MAÇADA, A. C. G.; BECKER, J. L. Modelo de decisão para o planejamento da capacidade nos terminais de containers. **in: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO (ENEGEP)**, Ouro Preto: ABEPRO/UFOP, n. 23, p. 1–8, 2003. 15
- SENNE, E. L. F.; LORENA, L. A. N. Stabilizing column generation using lagrangean/surrogate relaxation. **In EUROPEAN OPERATIONAL RESEARCH CONFERENCE (EURO 2001)**, Rotterdam, The Netherlands, v. 18, 2001. 45
- SENNE, E. L. F.; LORENA, L. A. N.; PEREIRA, M. A. A simple stabilizing method for column generation heuristics: an application to p-median location problems. **International Journal of Operations Research**, v. 4, p. 1–9, 2007. 45, 46, 49
- SERPA, D. R. **Abordagens heurísticas para problemas de agrupamentos**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, 2011. 45, 55, 56
- SILVA, V. M. D. Uma visão sobre o problema de alocação de berços. **Revista produção**, v. 7, n. 2, p. 85–98, 2007. 15
- THURMAN, K. P. **Optimal ship berthing plans**. 55 p. Masters of science in operations research — Naval Postgraduate School, Monterey, California - EUA, 1989. 11
- UNCTAD. **United nations conference on trade and development**. New York and Geneva: Report by the UNCTAD secretariat, 2009. Review of maritime transport. 5

\_\_\_\_\_. \_\_\_\_\_. 2010. Review of maritime transport. 5

WANG, X. Y.; WHITWELL, G.; GARIBALDI, J. M. The application of a simulated annealing fuzzy clustering algorithm for cancer diagnosis. In: **in IEEE International Conference on Intelligent Systems Design and Application**. Budapest: [s.n.], 2004. v. 4, p. 467–472. 44

XU, D.; LI, C.-L.; LEUNG, J. Y.-T. Berth allocation with time-dependent physical limitations on vessels. **European Journal of Operational Research**, v. 216, n. 1, p. 47–56, 2012. 14

XU, R.; WUNSCH, D. Survey of clustering algorithms. **Neural Networks, IEEE Transactions on**, v. 16, n. 3, p. 645–678, 2005. 44, 45

## APÊNDICE A - TRABALHOS PUBLICADOS

Nesta seção são listados os trabalhos publicados e/ou apresentados em evento científicos.

### A.1 Trabalhos publicados em periódicos

**Título:** *Hybrid heuristics based on column generation with path-relinking for clustering problems*

**Autores:** Rudinei Martins de Oliveira, Luiz Antonio Nogueira Lorena, Antonio Augusto Chaves e Geraldo Regis Mauri

**Periódico:** Expert Systems With Applications

**Editora:** Elsevier **ISSN:**

**doi:**

**Volume:**

**Número:**

**Páginas:**

**Ano:** 2014

**Título:** *Clustering Search for the Berth Allocation Problem*

**Autores:** Rudinei Martins de Oliveira, Geraldo Regis Mauri e Luiz Antonio Nogueira Lorena

**Periódico:** Expert Systems With Applications

**Editora:** Elsevier

**ISSN:** 0957-4174

**doi:** 10.1016/j.eswa.2011.11.072

**Volume:** 39

**Número:** 5

**Páginas:** 5499 - 5505

**Ano:** 2012

## A.2 Trabalhos publicados em periódicos e apresentado em evento internacional

**Título:** *Clustering search heuristic for solving a continuous Berth Allocation Problem*

**Autores:** Rudinei Martins de Oliveira, Geraldo Regis Mauri e Luiz Antonio Nogueira Lorena

**Local / data:** Malaga, Spain / 11-13 april 2012

**Evento:** *12<sup>th</sup> European conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP)*

**Periódico:** Lecture Notes in Computer Science

**Editora:** Springer

**ISSN:** 978-3-642-29123-4

**doi:** 10.1007/978-3-642-29124-1\_5

**Volume:** 7245

**Páginas:** 49-62

**Ano:** 2012

## A.3 Trabalhos apresentados em evento nacional

**Título:** *Clustering search* aplicado ao problema de alocação de berços

**Autores:** Rudinei Martins de Oliveira, Geraldo Regis Mauri e Luiz Antonio Nogueira Lorena

**Evento:** XLII Simpósio Brasileiro de Pesquisa Operacional (SBPO)

**Local/Data:** Bento Gonçalves - RS / 30 de agosto a 3 de setembro de 2010

## APÊNDICE B - Soluções obtidas para o PAB-C

Nesta seção são apresentados os gráficos das soluções da Tabela 2.2 obtidas pelo *Clustering Search* para PAB-C. Estas Figuras mostram a distribuição dos navios (retângulos, numerados de 1 a 60) nos berços. Desta forma, é possível observar que as soluções obtidas são viáveis, pois não há sobreposição dos retângulos (navios).

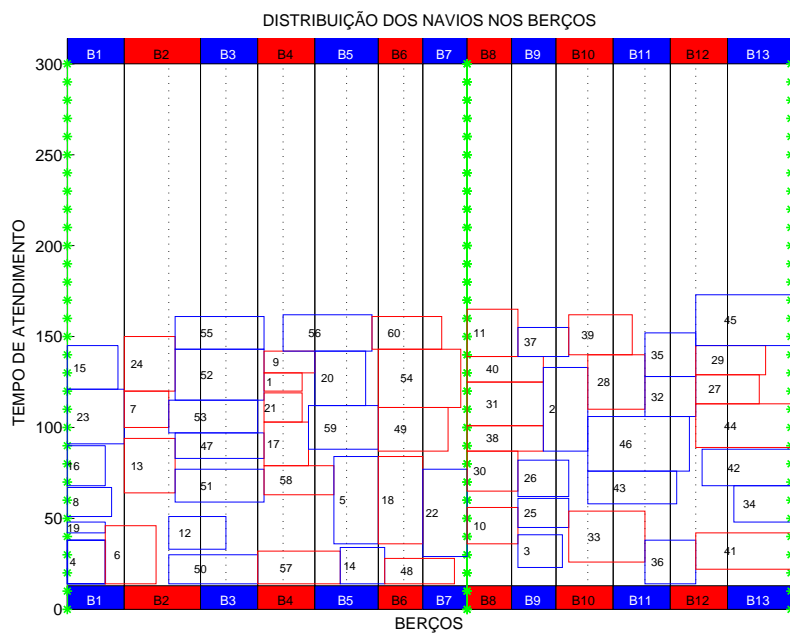


Figura B.1 - i01 - FO = 1583

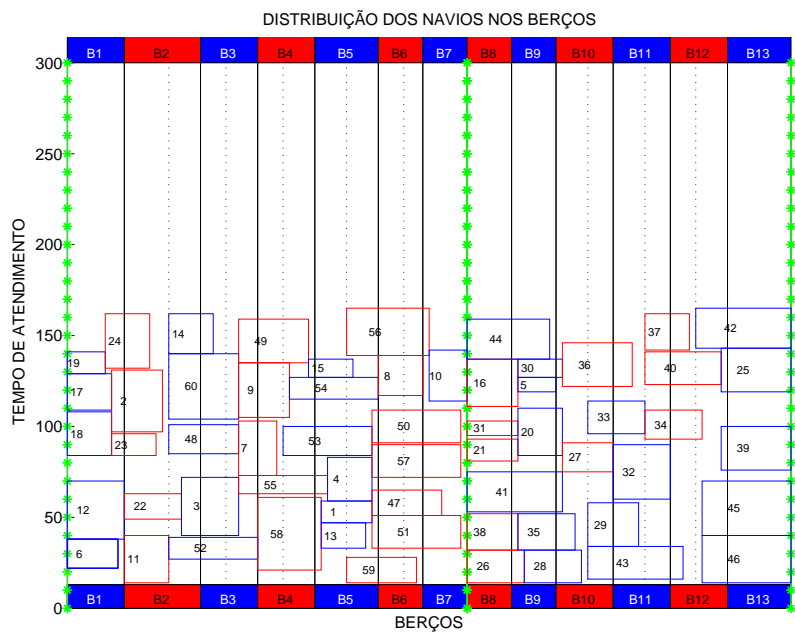


Figura B.2 - i02 - FO = 1315

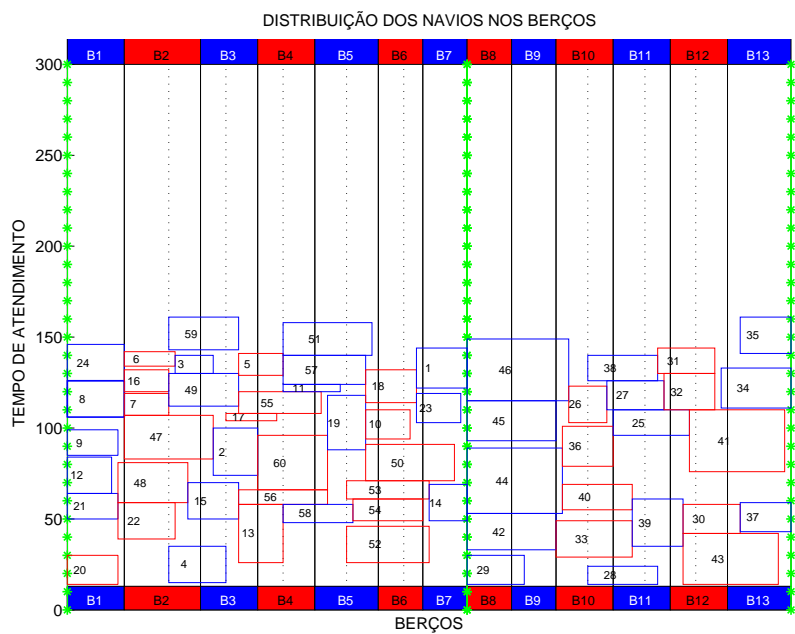


Figura B.3 - i03 - FO = 1207



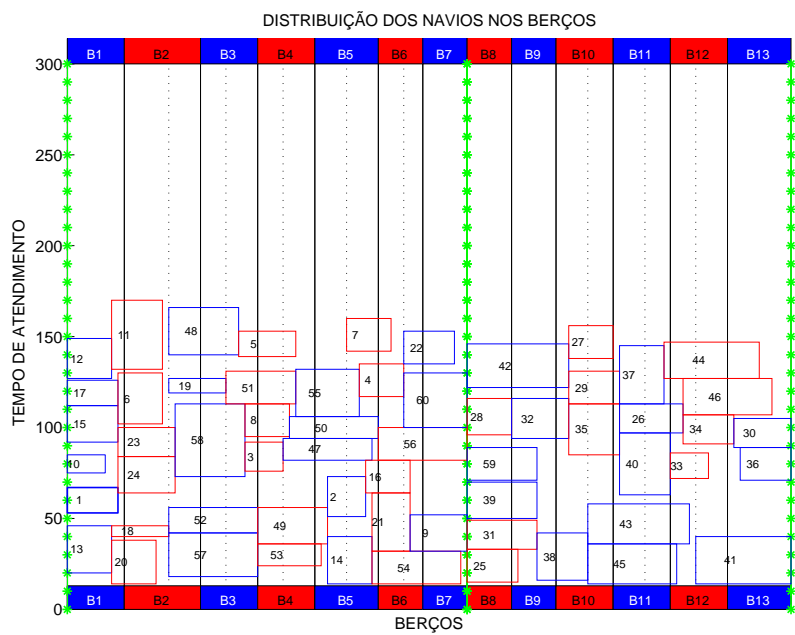


Figura B.4 - i04 - FO = 1380

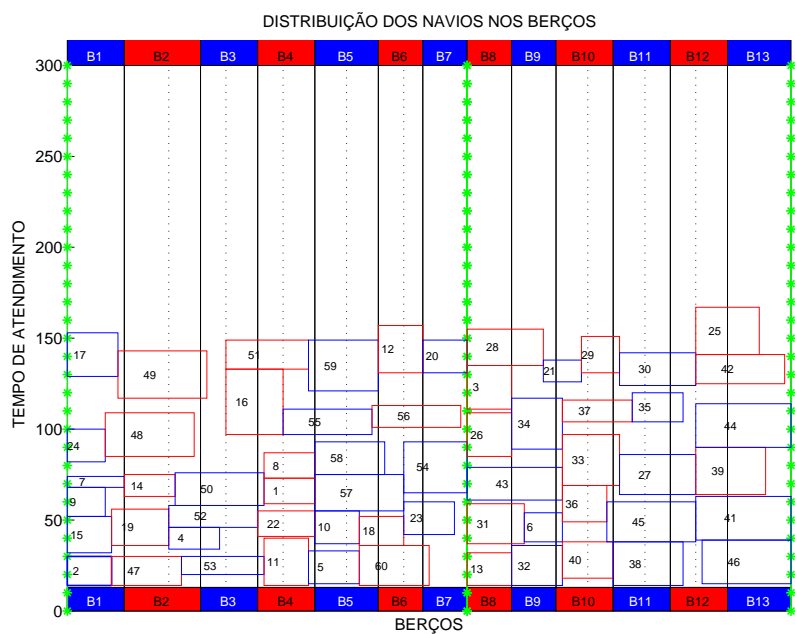


Figura B.5 - i05 - FO = 1262

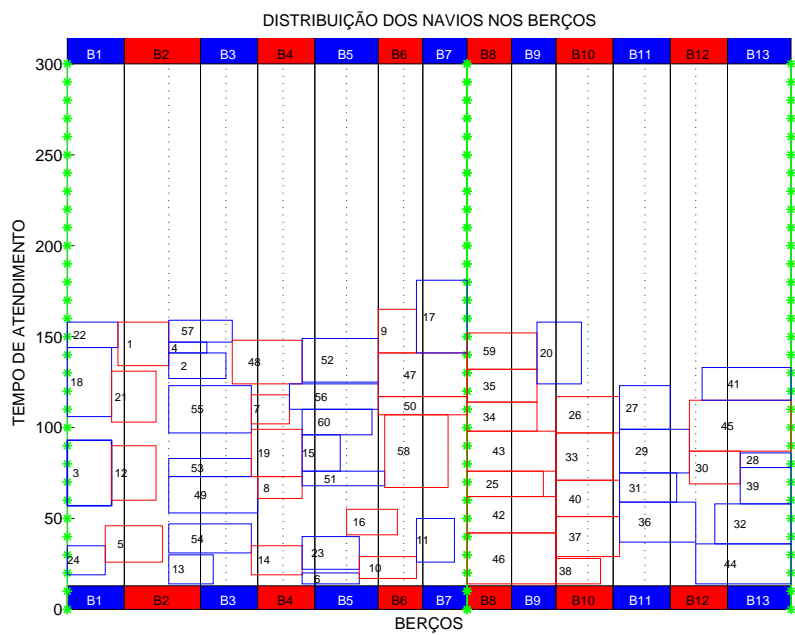


Figura B.6 - i06 - FO = 1413

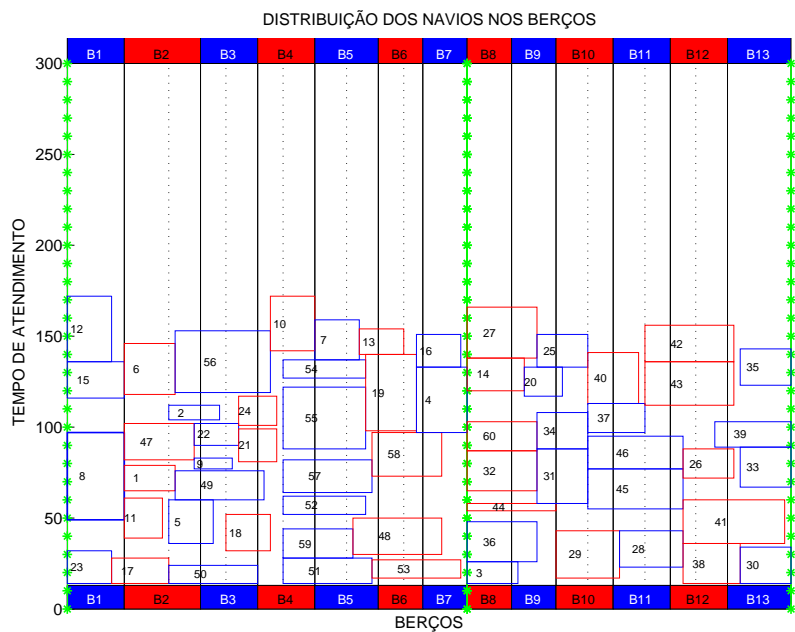


Figura B.7 - i07 - FO = 1324

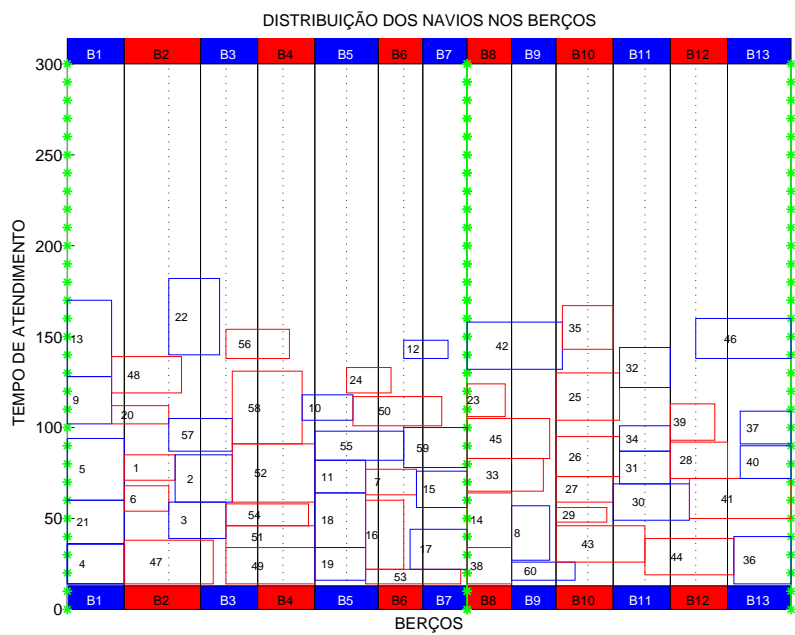


Figura B.8 - i08 - FO = 1381

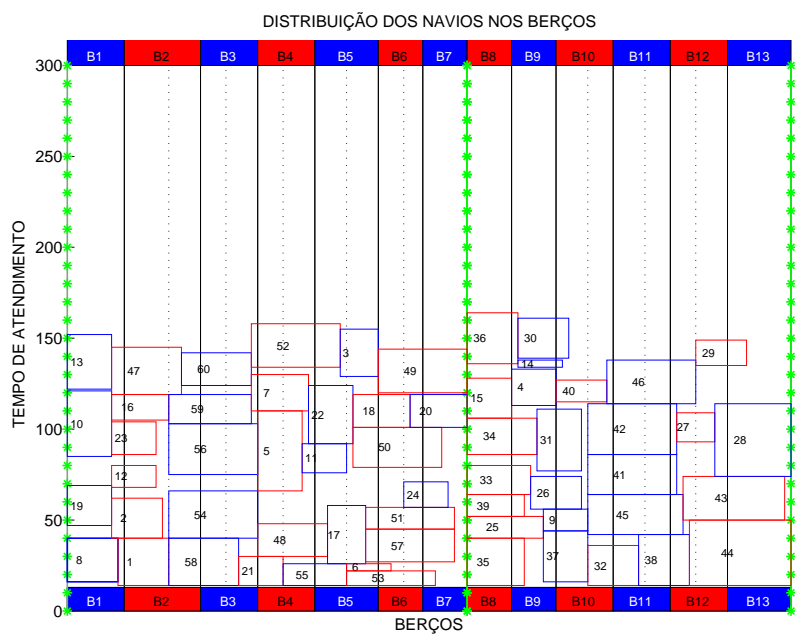


Figura B.9 - i09 - FO = 1633

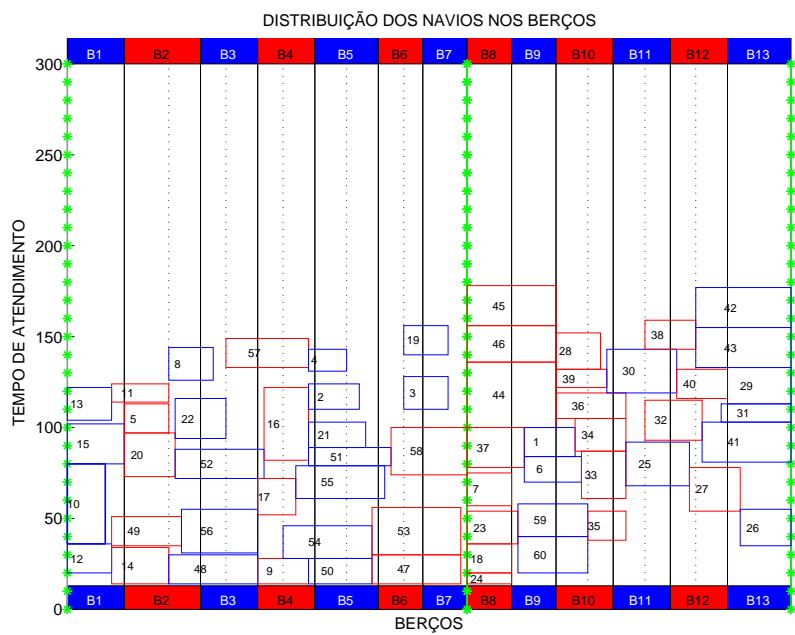


Figura B.10 - i10 - FO = 1333

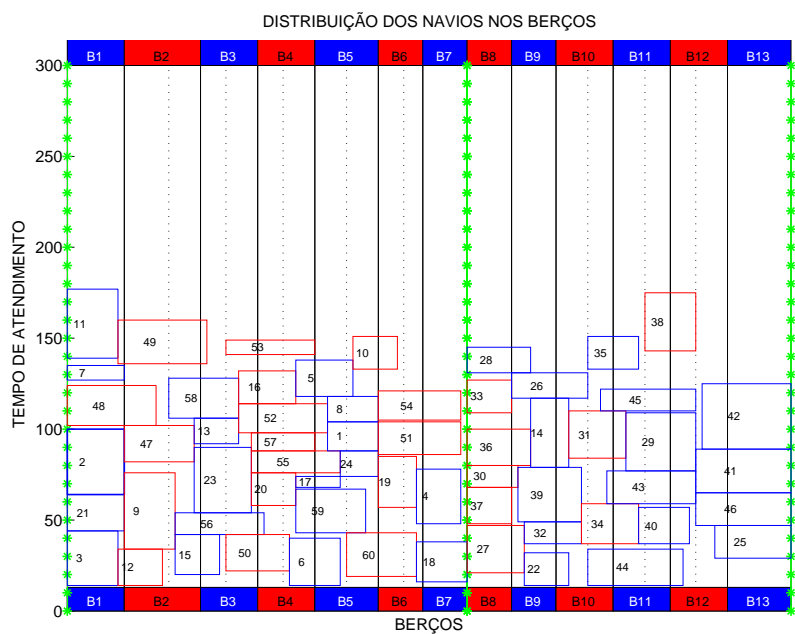


Figura B.11 - i11 - FO = 1527

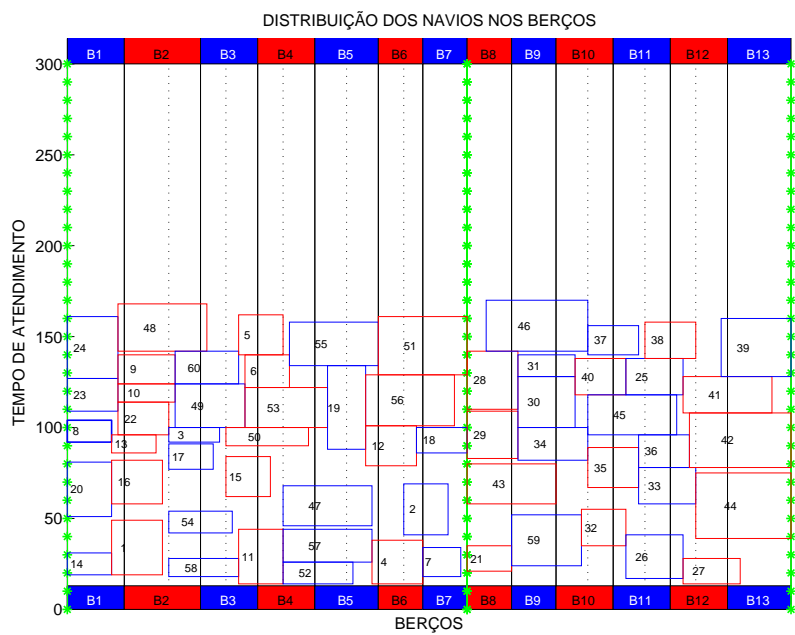


Figura B.12 - i12 - FO = 1505

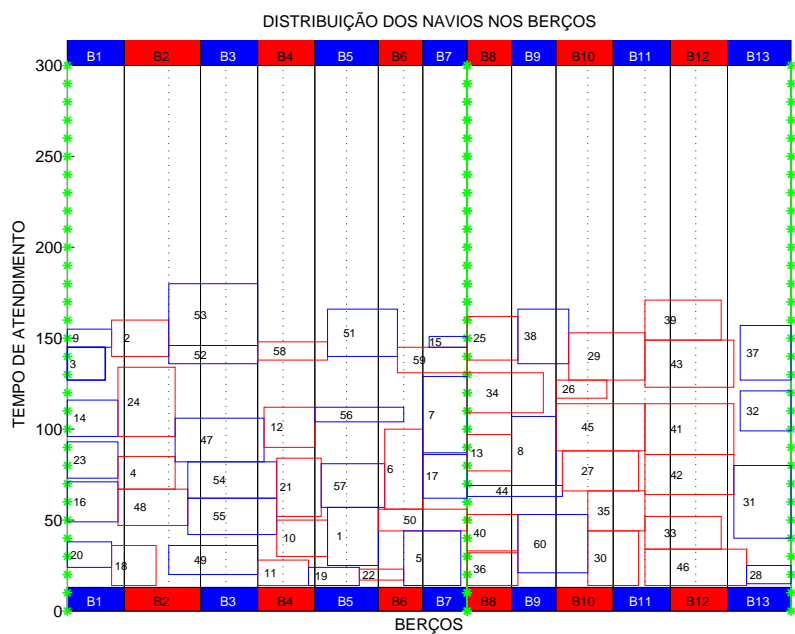


Figura B.13 - i13 - FO = 1424

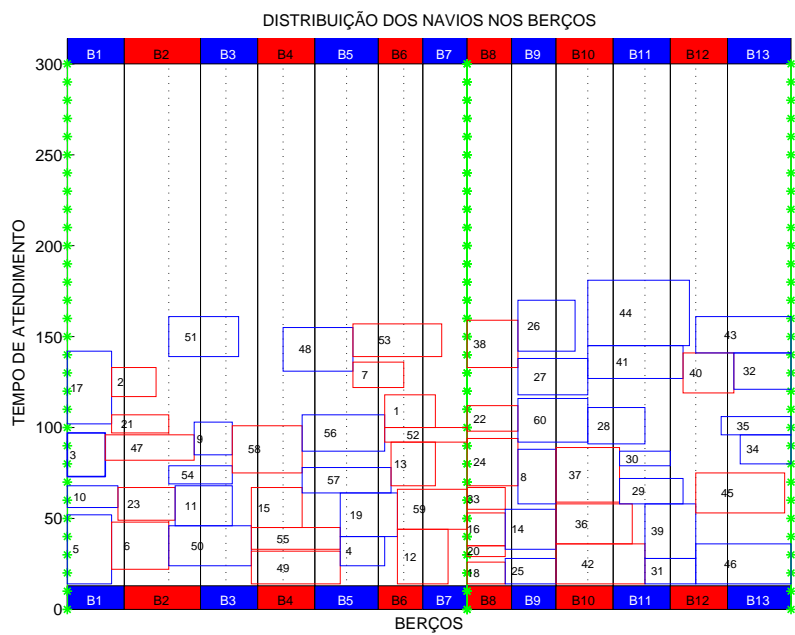


Figura B.14 - i14 - FO = 1273

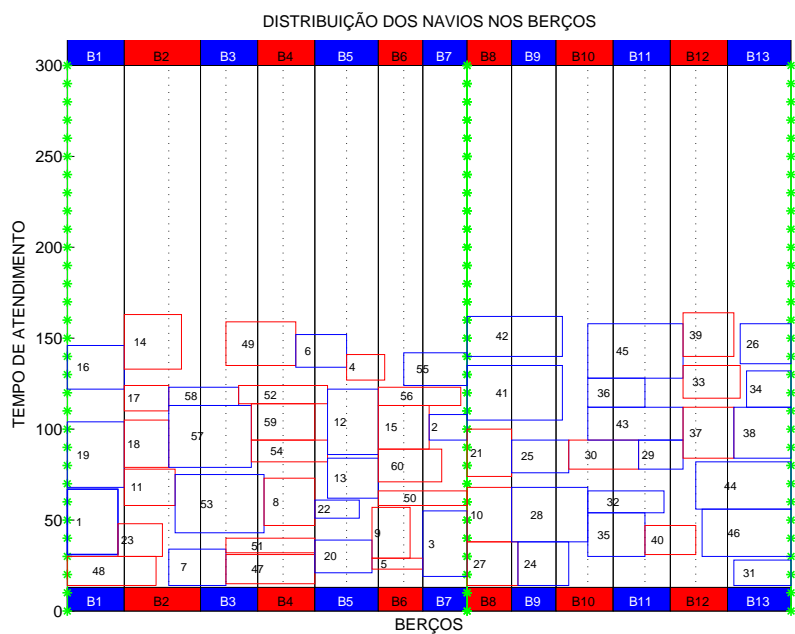


Figura B.15 - i15 - FO = 1345

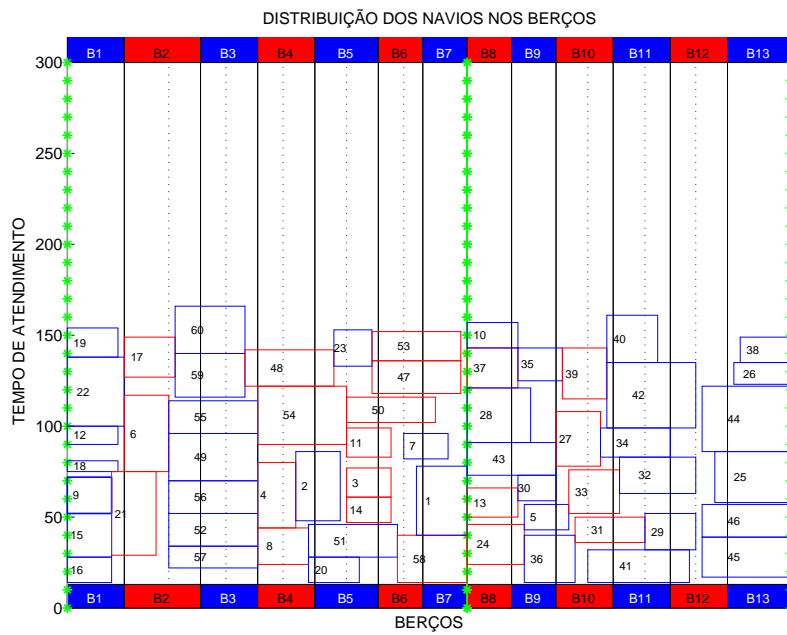


Figura B.16 - i16 - FO = 1491

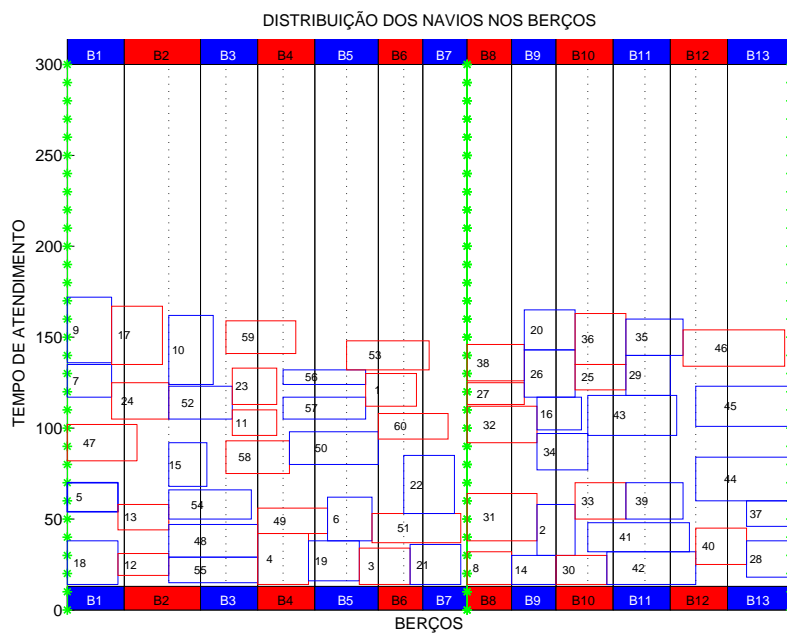


Figura B.17 - i17 - FO = 1302

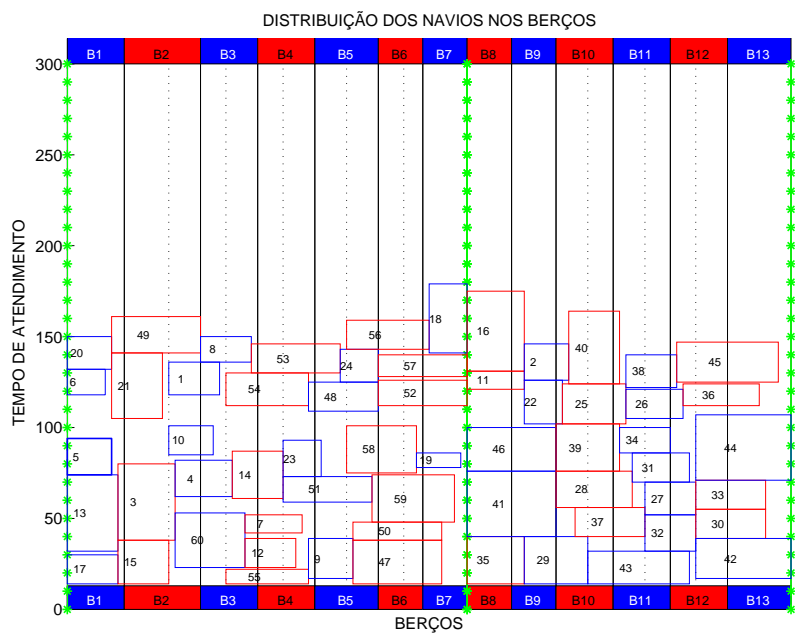


Figura B.18 - i18 - FO = 1518

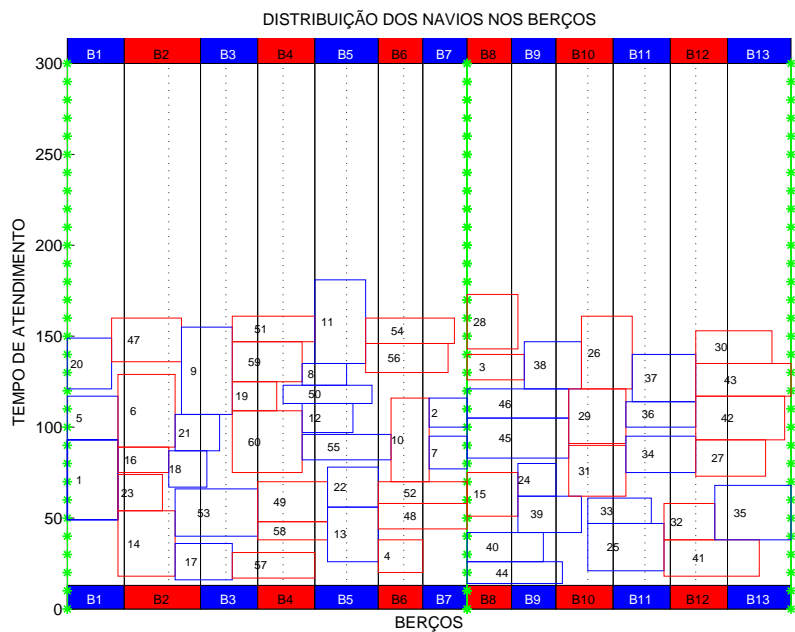


Figura B.19 - i19 - FO = 1517



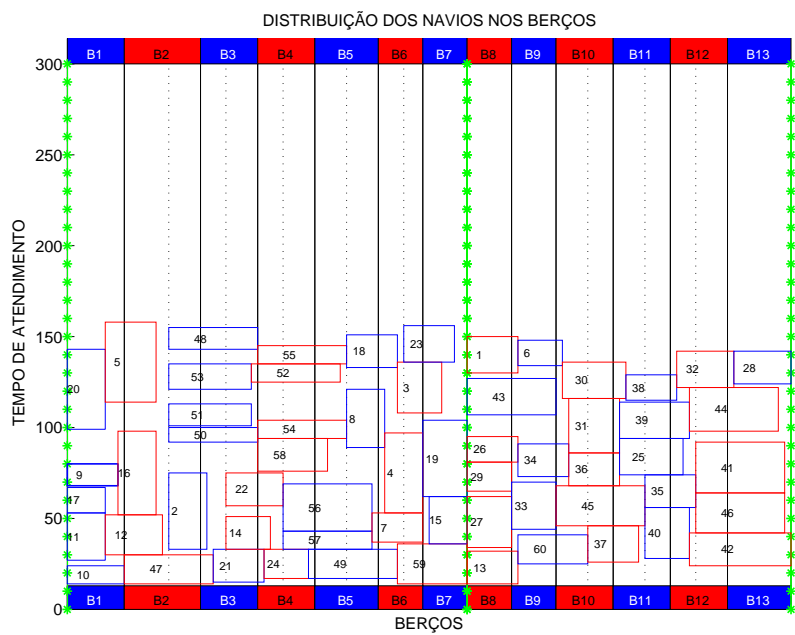


Figura B.20 - i20 - FO = 1406

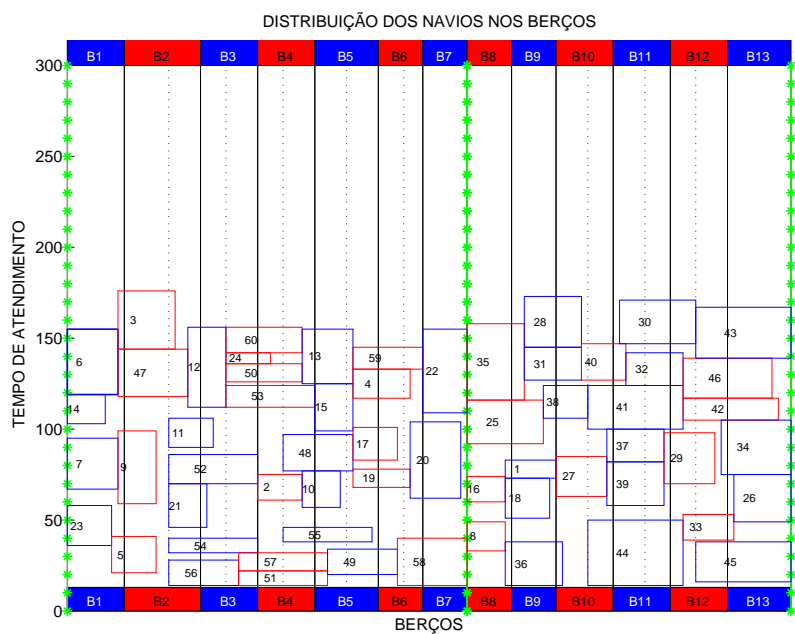


Figura B.21 - i21 - FO = 1461

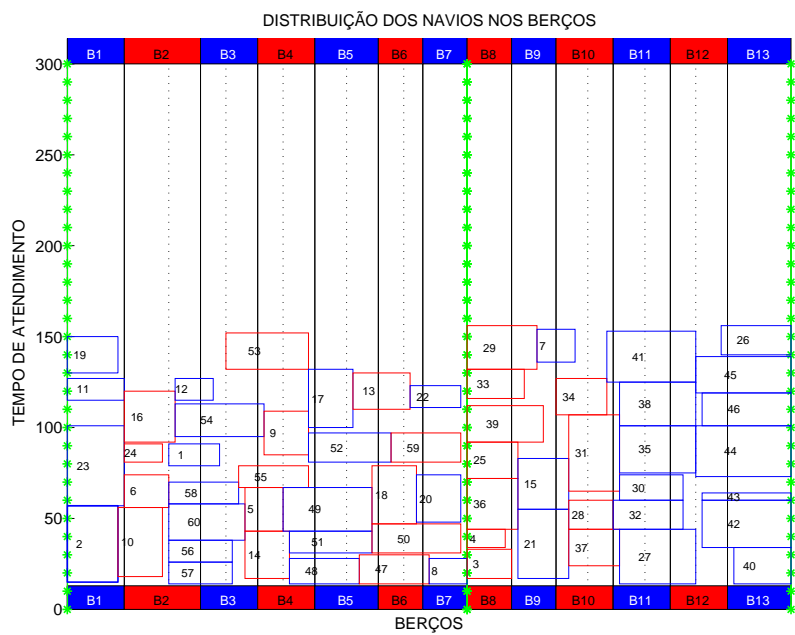


Figura B.22 - i22 - FO = 1437

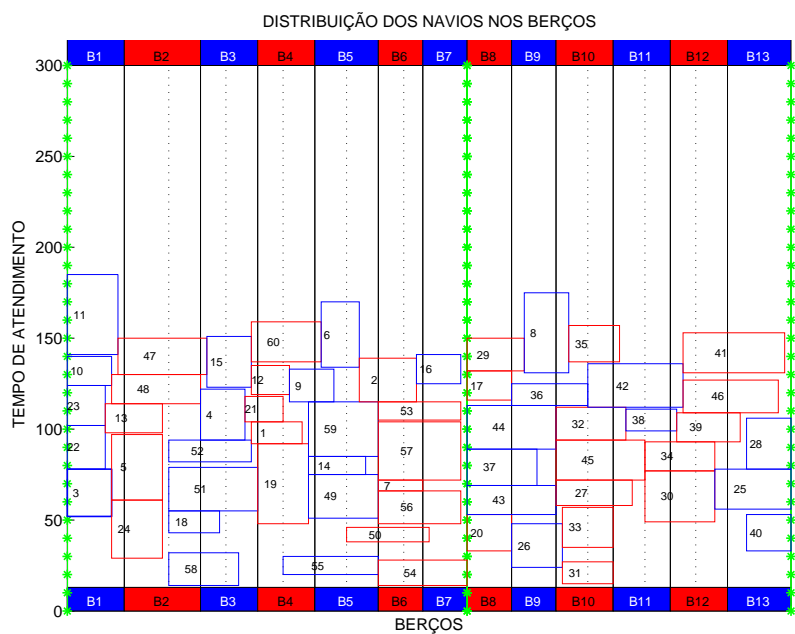


Figura B.23 - i23 - FO = 1395

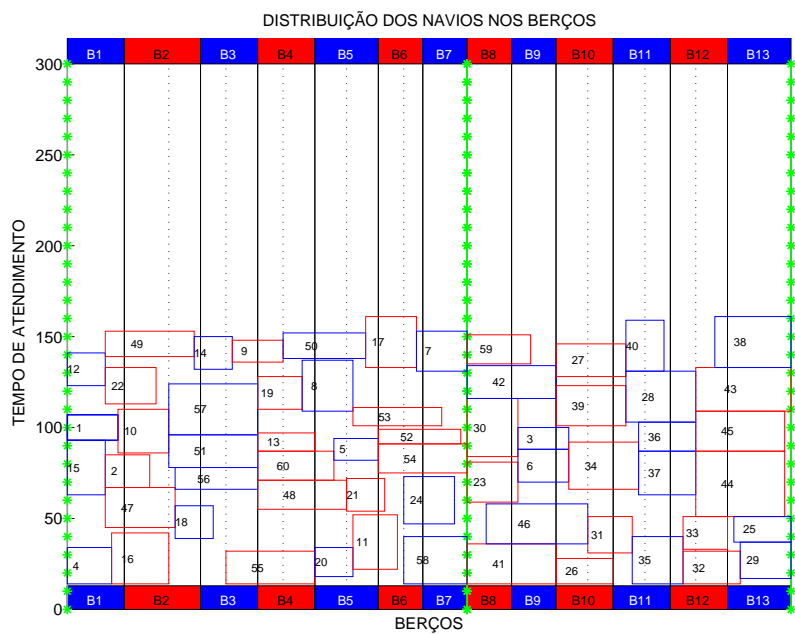


Figura B.24 - i24 - FO = 1351

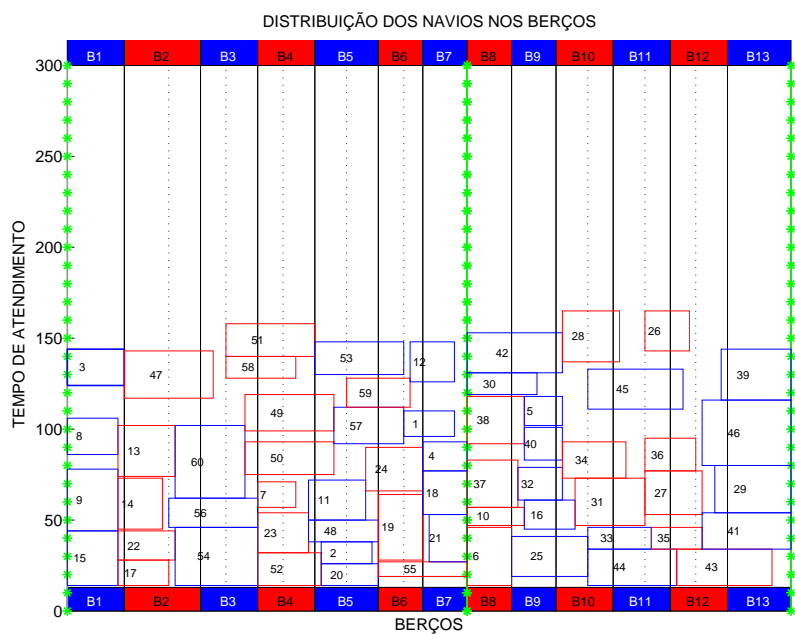


Figura B.25 - i25 - FO = 1513

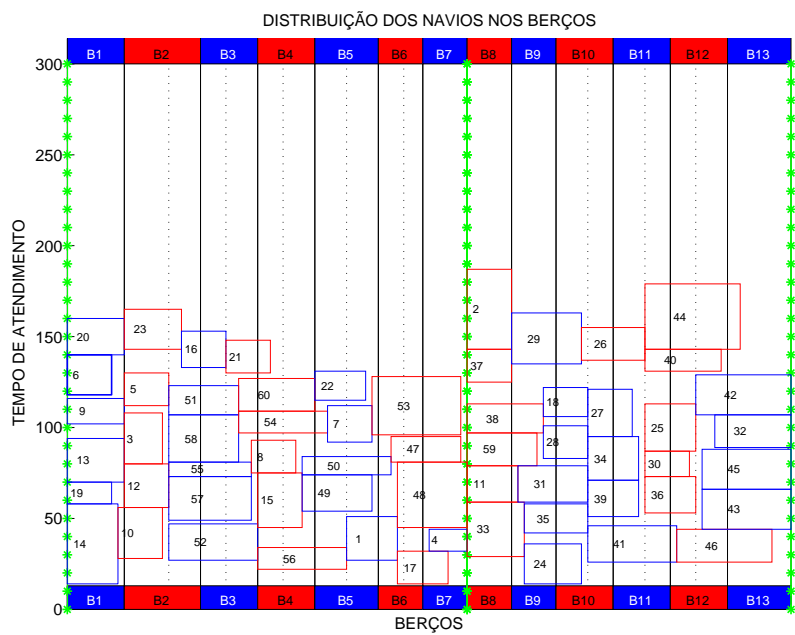


Figura B.26 - i26 - FO = 1448

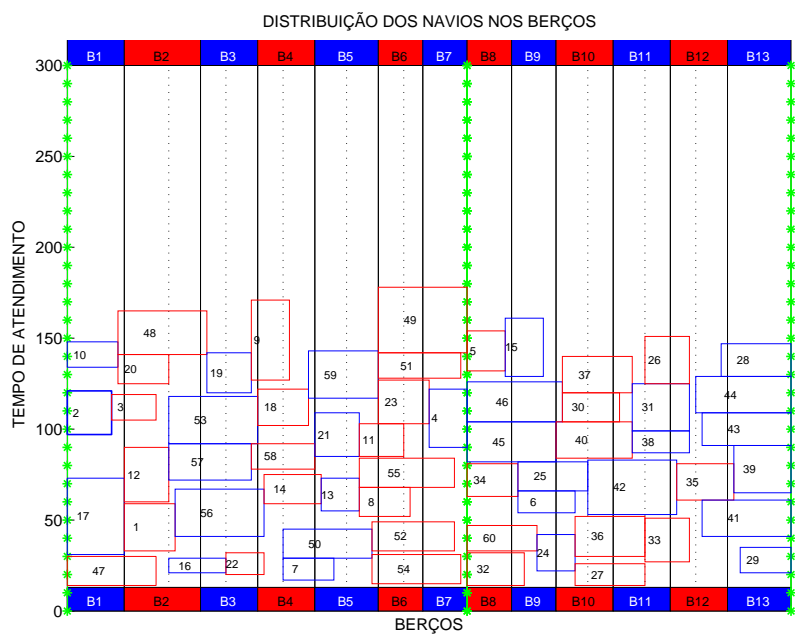


Figura B.27 - i27 - FO = 1349

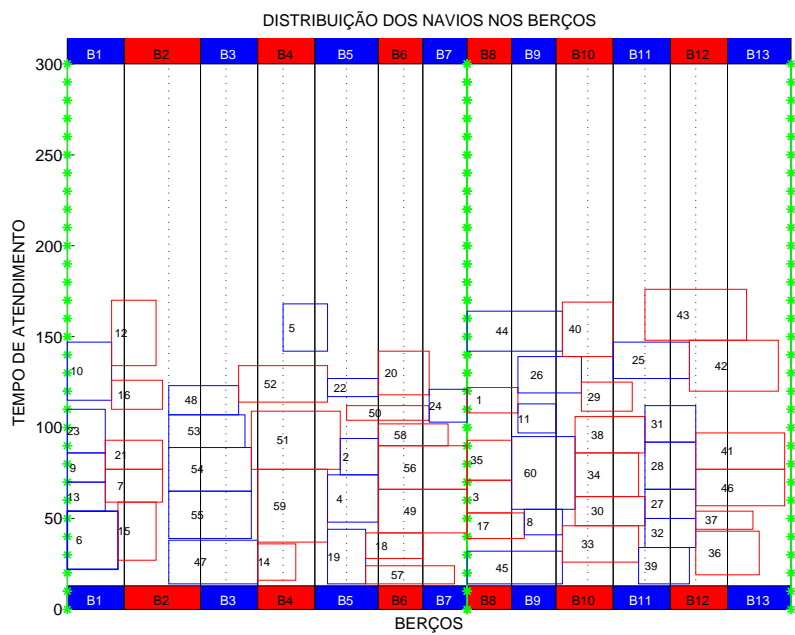


Figura B.28 - i28 - FO = 1461

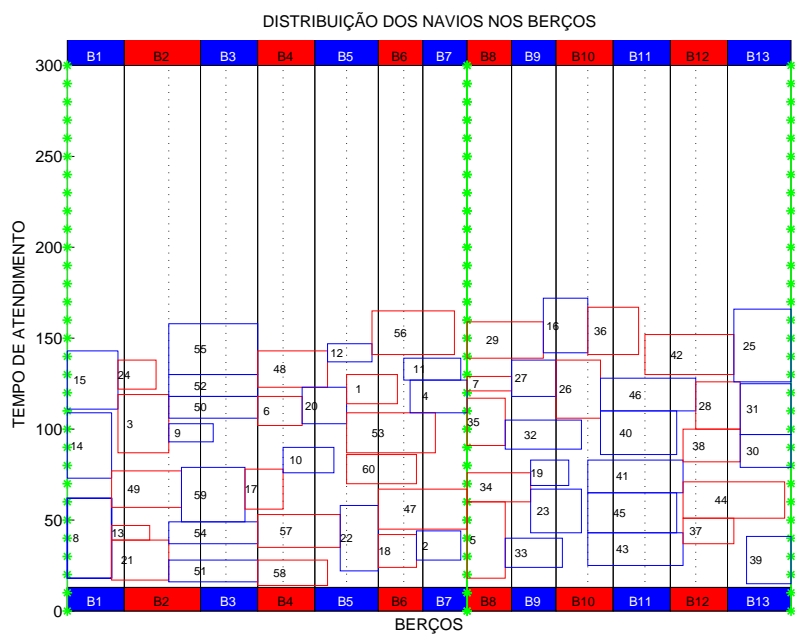


Figura B.29 - i29 - FO = 1323

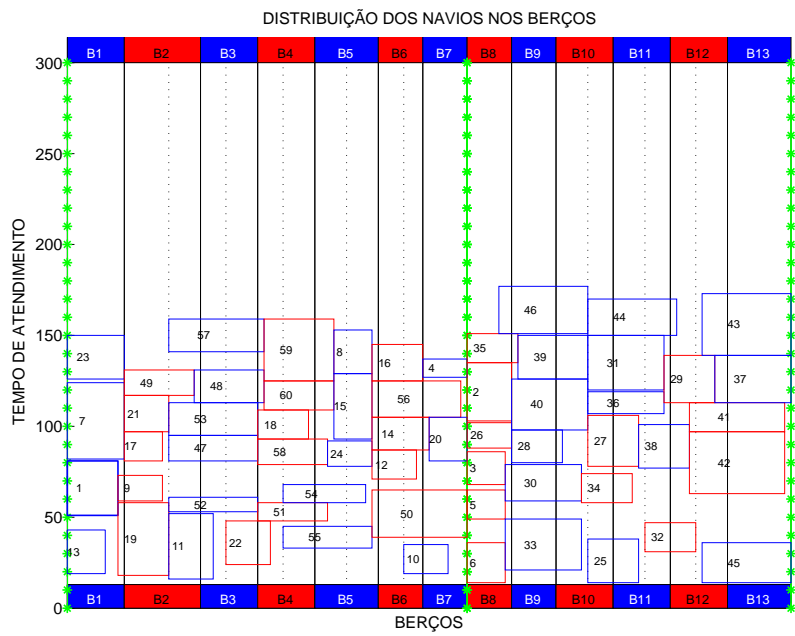


Figura B.30 - i30 - FO = 1487

## **PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE**

### **Teses e Dissertações (TDI)**

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

### **Manuais Técnicos (MAN)**

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

### **Notas Técnico-Científicas (NTC)**

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

### **Relatórios de Pesquisa (RPQ)**

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

### **Propostas e Relatórios de Projetos (PRP)**

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

### **Publicações Didáticas (PUD)**

Incluem apostilas, notas de aula e manuais didáticos.

### **Publicações Seriadas**

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

### **Programas de Computador (PDC)**

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

### **Pré-publicações (PRE)**

Todos os artigos publicados em periódicos, anais e como capítulos de livros.