

# INSTITUTO TECNOLÓGICO DE AERONÁUTICA



João Marcos Sousa Miranda

Identificação de perfis de atrasos de voos no Aeroporto  
Internacional de Brasília

Trabalho de Graduação  
Ano 2015

Eng. Civil-Aeronáutica

JOÃO MARCOS SOUSA MIRANDA

**IDENTIFICAÇÃO DE PERFIS DE ATRASOS DE VOOS NO  
AEROPORTO INTERNACIONAL DE BRASÍLIA**

Orientador  
Prof. Rodrigo Arnaldo Scarpel (ITA)

**Divisão de Engenharia Civil-Aeronáutica**

SÃO JOSÉ DOS CAMPOS  
CENTRO TÉCNICO AEROESPACIAL  
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2015

## Dados Internacionais de Catalogação-na-Publicação (CIP)

### Divisão de Informação e Documentação

MIRANDA, João Marcos Sousa

Identificação de perfis de atrasos de voos no Aeroporto Internacional de Brasília

São José dos Campos, 2015.

44f.

ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Rodrigo Arnaldo Scarpel; coorientador: Alessandro Vinícius Marques de Oliveira. Publicado em 2015.

1.Controle de tráfego aéreo. 2.Atraso. 3.Analise de dados. 4. Aeroportos.5. Operações de linhas aéreas. 6.Transportes. I. Instituto Tecnológico de Aeronáutica II. Identificação de perfis de atrasos de voos no Aeroporto Internacional de Brasília

## REFERÊNCIA BIBLIOGRÁFICA

MIRANDA, João Marcos Sousa. **Identificação de perfis de atrasos de voos no Aeroporto Internacional de Brasília**. 2015. 44f. Trabalho de Conclusão de Curso. (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

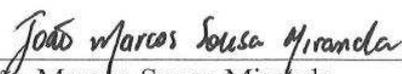
## CESSÃO DE DIREITOS

NOME DO AUTOR: João Marcos Sousa Miranda

TÍTULO DO TRABALHO: Identificação de perfis de atrasos de voos no Aeroporto Internacional de Brasília

TIPO DO TRABALHO/ANO: Graduação / 2015

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta monografia de graduação pode ser reproduzida sem a autorização do autor.



João Marcos Sousa Miranda

Rua Antonina do Norte, 351 – apto 402A São Gerardo

60325-610 – Fortaleza – CE

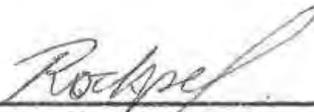
## ESTUDO DE ALOCAÇÃO DE HUBS PARA A DEMANDA INTERNACIONAL

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação



---

João Marcos Sousa Miranda  
Autor



---

Prof. Dr. Rodrigo A. Scarpel (ITA)  
Orientador



---

Prof. Eliseu Lucena Neto  
Coordenador do Curso de Engenharia Civil-Aeronáutica

São José dos Campos, 25 de novembro de 2015

## **AGRADECIMENTOS**

Agradeço a Deus por ter me guiado com seu amor sem medida por todos esses dias.

Agradeço à Virgem Maria, que com o seu ‘sim’ permitiu o anúncio do Evangelho.

Aos meus familiares, Maurício, Rita, Leandro e Isabella, que são minhas raízes e minhas asas. Sem vocês nada disso teria sido possível. Nunca poderei ser grato o suficiente por tanto bem que me fizeram.

Ao meu orientador, Prof. Rodrigo Scarpel, pela dedicação e paciência em me auxiliar durante a confecção deste trabalho, sempre contribuindo com sugestões valiosas.

Aos meus amigos do inesquecível H20B-111, que foram a família que fiz no ITA; e à Joana, minha namorada, que com amor sempre esperou que eu acabasse cada trabalhinho no ITA para que depois eu lhe desse merecida atenção. Obrigado a todos.

## RESUMO

Este trabalho tem como objetivo principal utilizar métodos de formação de agrupamentos em análise de dados para a identificação de perfis de atrasos em voos no Aeroporto Internacional de Brasília. Para isso, foram avaliadas todas as movimentações, tanto previstas como efetivamente realizadas, de voos no aeroporto durante o período que vai de janeiro de 2010 a maio de 2014. Todos os voos com diferença superior a 15 minutos entre o momento real e o previsto de sua movimentação na pista do aeroporto foram considerados como atrasados. Foram criados, através de técnicas hierárquicas de agrupamentos, grupos de dias com perfis semelhantes de movimentações previstas, de movimentações realizadas e de índices horários de atrasos, de modo a identificar possíveis padrões temporais que caracterizassem a formação desses agrupamentos. Enquanto tais padrões parecem emergir nos primeiros grupos citados, não há evidência clara que haja quaisquer padrões na divisão dos dias de acordo com seus perfis de atraso.

**Palavras-chave:** Controle de tráfego aéreo, Atraso, Análise de dados, Aeroportos, Operações de linhas aéreas, Transporte.

## **ABSTRACT**

This paper main goal is to identify flight delay patterns on Brasilia International Airport using group formation data analysis methods. All flights schedules, both predicted and realized, between January 2010 and May 2014 were used. Flights with 15 min difference from real to predicted time were considered as delayed. Using hierarchy grouping techniques. each day was grouped with others that shared similar predicted schedule, real schedule and delay profile so that delay creating patterns could be discovered. While such patterns seems to appear on the first groups mentioned, there is no evidence supporting any patterns on day division according to delay profiles.

**Palavras-chave:** Air traffic control, Delay, Data analysis, Airpoty, Airlines operations, Transportation.

## LISTA DE TABELAS

Tabela 1: Exemplos de informações recolhidas da base de dados VRA.....	19
Tabela 2: Movimentação de passageiros no Aeroporto Internacional de Brasília entre 2004 e 2014.....	19
20	
Tabela 3: Contagem do número de elementos de cada cluster do conjunto P .....	29
Tabela 4: Distribuição dos dias de semana nos elementos do conjunto P .....	30
Tabela 5: Distribuição do número de dias de cada ano nos elementos do conjunto P .....	31
Tabela 6: Contagem do número de elementos de cada grupo do conjunto R.....	33
Tabela 7: Distribuição dos dias de semana nos elementos do conjunto R .....	34
Tabela 8: Distribuição do número de dias de cada ano nos elementos do conjunto R .....	34
Tabela 9: Distribuição da quantidade de elementos dos grupos de movimentações previstas nos clusters movimentações realizadas .....	35
Tabela 10: Contagem do número de elementos de cada elemento do conjunto A.....	36
Tabela 11: Distribuição dos dias da semana nos elementos do conjunto A.....	37
Tabela 12: Distribuição do número de dias de cada ano nos elementos do conjunto A.....	37
Tabela 13: Distribuição da quantidade de elementos dos grupos de movimentações previstas nos grupos de atrasos .....	38
Tabela 14: Distribuição da quantidade de elementos dos grupos de movimentações realizadas nos grupos de atrasos .....	38
Tabela 15: Espaçamento médio para o pico matinal para cada um dos elementos do conjunto A.....	40
Tabela 16: Espaçamento médio para o pico noturno para cada um dos elementos do conjunto A.....	40

## LISTA DE FIGURAS

Figura 1: Evolução dos percentuais anuais de atrasos e cancelamentos de voos comerciais no Brasil .....	10
Figura 2: Evolução dos percentuais mensais para o ano de 2013 de atrasos e cancelamentos de voos comerciais no Brasil.....	11
Figura 3: Percentuais de atrasos por aeroporto de origem de voo .....	12
Figura 4: Exemplo ilustrativo de uma clusterização por critério de distância euclidiana .....	16
Figura 5: Participação dos 20 maiores aeroportos brasileiros na quantidade de decolagens para voos domésticos em 2013 .....	21
Figura 6: Vista aérea do Aeroporto Internacional de Brasília.....	22
Figura 7: Vista aérea da organização das pistas do Aeroporto Internacional de Brasília.....	23
Figura 8: Diagrama de árvore da divisão de clusters para movimentações previstas .....	29
Figura 9: Perfis de movimentação do acumulado horário para cada grupo do conjunto P...	30
Figura 10: Diagrama de árvore da divisão de clusters para as movimentações realizadas...	32
Figura 11: Perfis de movimentação do acumulado horário para cada grupo do conjunto R.	33
Figura 12: Perfis de movimentação do acumulado horário para cada grupo do conjunto A.	36

# SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>8</b>
1.1. Motivação.....	8
1.2. Justificativa .....	9
1.3. Situação brasileira .....	10
1.4. Objetivos.....	12
1.5. Organização do texto.....	13
<b>2. REFERENCIAL TEÓRICO</b> .....	<b>15</b>
2.1. Métodos de formação de agrupamentos.....	15
2.1.1. Princípios Gerais .....	15
2.1.2. Medidas de similaridade.....	16
2.1.2. Métodos hierárquicos.....	17
<b>3. MATERIAIS E MÉTODOS</b> .....	<b>18</b>
3.1. Base de dados .....	18
3.2. Caracterização do Aeroporto Internacional de Brasília.....	19
3.2.1. Considerações Gerais .....	19
3.2.2. Perspectivas .....	21
3.3. Método de análise dos dados .....	24
<b>4. RESULTADOS E DISCUSSÕES</b> .....	<b>28</b>
4.1. Movimentações previstas .....	28
4.2. Movimentações realizadas .....	31
4.3. Atrasos .....	35
<b>5. CONCLUSÃO</b> .....	<b>41</b>
<b>BIBLIOGRAFIA</b> .....	<b>42</b>

# 1. INTRODUÇÃO

## 1.1. Motivação

Sob as perspectivas e os interesses tanto de passageiros quanto de companhias aéreas e de gestores aeroportuários, atrasos em voos despontam como um dos principais contratempos dentro do mercado de aviação civil. A título de ilustração, de acordo com dados estatísticos do Departamento de Transportes dos Estados Unidos [1] (US Department of Transportation), aproximadamente 24% de todos os voos domésticos em território norte-americano aterrissaram com atraso superior a 15 minutos entre Janeiro de 2014 e Janeiro de 2015.

Os fatores responsáveis por gerar atrasos em voos são muitos e advêm de diferentes agentes. Condições como falta de capacidade de pista para o nível de demanda do aeroporto, problemas de segurança e de manutenção de aviões, controle precário de tráfego aéreo, condições climáticas desfavoráveis, tamanho reduzido de pátio para aeronaves, número de portões de embarque aquém do exigido e supersaturação de voos em horas picos estão entre as mais usuais causas de atrasos. Para Diana [2], atrasos são frutos de um trade-off entre a demanda por movimentações na pista e a capacidade do aeroporto.

Importa destacar que há um forte impacto econômico que surge em decorrência de atrasos e cancelamentos de voos. Naturalmente, atrasos aumentam os custos operacionais das linhas aéreas, que acabam por consumir mais de seus recursos com combustível, horas de trabalho de seus funcionários, manutenção de equipamentos e etapas do processo de apoio à aeronave em solo. Esse aumento de custos é repassado em parte aos passageiros e, de modo semelhante, também afeta a qualidade do serviço oferecido em voo, que é reduzido de modo a cortar custos operacionais a fim de manter os preços das tarifas em valores competitivos.

O impacto econômico devido a atrasos em voos costuma ser de ordem maior do que usualmente se imagina, principalmente quando avaliada em escala nacional. De acordo com [3], estima-se que as perdas econômicas derivadas de atrasos e cancelamentos de voos foram superiores a 31 bilhões de dólares nos Estados Unidos no ano de 2007. Essa perda é fruto não só dos prejuízos das companhias aéreas, mas

principalmente das perdas financeiras dos passageiros, que perdem tempo produtivo de trabalho devido a recorrentes atrasos dos quais são vítimas e consequente perdas de conexões ou cancelamentos de voos.

É imprescindível, portanto, uma análise da conjuntura de atrasos em voos e de suas razões, de modo que, futuramente, seja possível que se mitigue esses desfavoráveis impactos econômicos e que se aumente o bem-estar dos passageiros dentro do mercado de aviação civil.

## **1.2. Justificativa**

A literatura especializada abrange, em geral, múltiplos aspectos relacionados aos fenômenos de atrasos em voos. Consideram-se, por exemplo, as causas diretas e indiretas do problema, a perda econômica associada, a mudança dos hábitos de consumo dos passageiros e o impacto de legislações vigentes associadas ao funcionamento de aeroportos.

A determinação de fatores associados a atrasos em voo é um dos aspectos mais amplamente abordados. Abdel-Aty et al. [4] aponta a existência de padrões periódicos de atraso, sejam esses diários, semanais ou mensais, e que estão diretamente associados com o nível de demanda do aeroporto. Santos e Robin [5], trabalhando com estatísticas de voos europeus entre 2000 e 2004, concluem que as variáveis que melhor explicam os atrasos em aeroportos da Europa são concentração de mercado, coordenação de slots, tamanho de Hubs e o nível de presença de companhias aéreas que operam em Hubs; também conclui que os níveis de atraso são, em geral, maiores em Hubs e menores em aeroportos com maior concentração de voos na mão de poucas companhias aéreas, conforme medido pelo índice HHI (Herfindal-Hirschman Index). Diana [6] afirma, em um estudo de caso em alguns aeroportos norte-americanos, que aeroportos com maior concentração de mercado não propagam mais atrasos que os demais.

Sobre a reação dos passageiros em relação a atrasos costumeiros, Brito et al. [7] demonstra, baseado principalmente em estimativas econométricas, que atrasos frequentes em determinadas rotas reduz a demanda de passagens para esses voos, fato esse que acarreta um natural aumento do valor da tarifa aérea, conforme também apontado por Zor e Hansen [8]. Uma abordagem acerca das consequências de múltiplos atrasos no comportamento de consumo de passageiros é feita por Ferrer et al. [9], que

conclui que aqueles tendem a voar menos à medida que passam por um maior número de más experiências com voos atrasados.

### 1.3. Situação da aviação comercial brasileira

Em se tratando da frequência de atrasos de voos, é possível uma breve percepção do estado da aviação comercial brasileira ao se observar os dados apontados no Anuário do Transporte Aéreo produzido pela Agência Nacional de Aviação Civil (ANAC) [10]. O percentual de voos atrasados, de acordo com a metodologia do documento, é calculado a partir da razão entre a quantidade de etapas de voo atrasadas e a quantidade total das etapas efetivamente realizadas (descontam-se, portanto, os cancelamentos).

Um panorama geral dos percentuais de atraso e cancelamento entre os anos de 2004 e 2013, de acordo com dados do Anuário do Transporte Aéreo de 2013, pode ser visualizado na Figura 1. Observe taxas elevadas de atraso entre os anos de 2006 e 2008, com um pico de 30,4% de atrasos em voos para o ano de 2007. Esses valores estão associados ao período da conhecida crise do transporte aéreo no Brasil, que foi responsável por uma posterior reestruturação do setor.

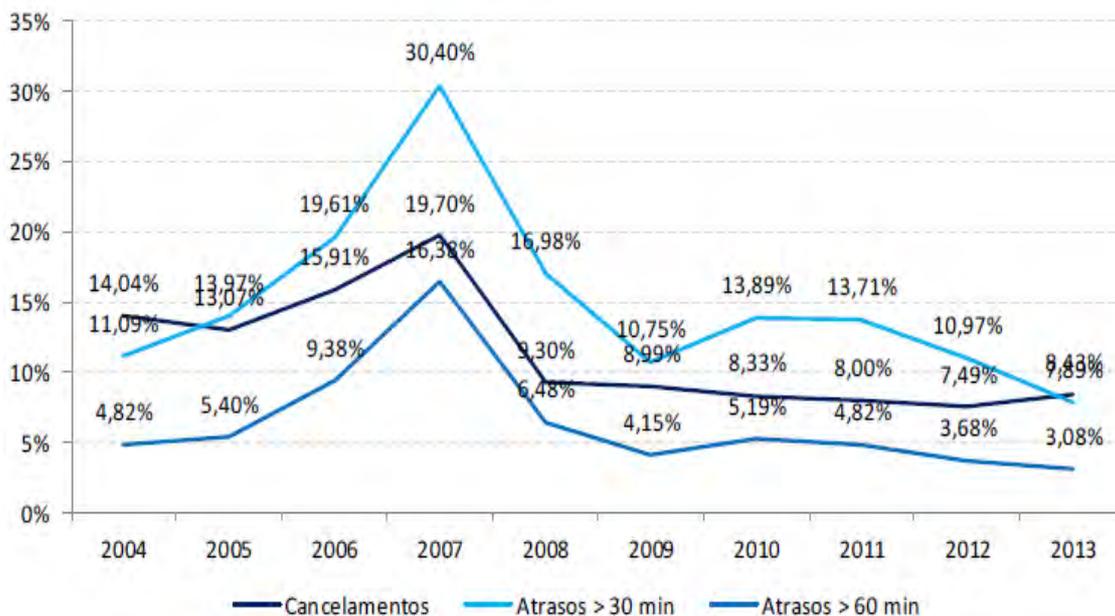


Figura 1: Evolução dos percentuais anuais de atrasos e cancelamentos de voos comerciais no Brasil

Observa-se também uma variação mensal nas taxas de atraso. Meses com uma demanda naturalmente maior por passagens aéreas, como nos casos das férias escolares de Julho e das festas de fim de ano em Dezembro, apontam valores mais elevados nas porcentagens de voos atrasados. Essas variações mensais para o ano de 2013, também retiradas do Anuário do Transporte Aéreo, podem ser melhor visualizadas através da Figura 2.



**Figura 2: Evolução dos percentuais mensais para o ano de 2013 de atrasos e cancelamentos de voos comerciais no Brasil**

Observe, ainda na Figura 2, que a taxa de cancelamento geralmente está associada à de atrasos, de modo que os picos de ambas usualmente coincidem. Esses pontos representam momentos de maior saturação da malha viária devido ao aumento sazonal da demanda, resultando em maiores índices de voos cancelados e atrasados em alguns aeroportos-chaves.

Por fim, é possível a obtenção da taxa de atraso, por aeroporto de origem de voo, discriminada para os aeroportos de maior circulação nacional e para os que apresentam índices mais elevados de atrasos. Esses dados, apontados na Figura 3, são retirados do Anuário do Transporte Aéreo de 2011, o mais recente em que se apresenta esse tipo de estatística.

É interessante notar que os aeroportos com maior demanda de passageiros, e consequentemente com maior número de movimentações de aeronaves, não coincidem

com os que apontam maiores incidências de atraso, que são em geral associados a aeródromos regionais. Esse fato se dá devido ao maior nível de planejamento alcançado pelos maiores aeroportos, o que causa de um controle maior sobre as operações de voo e reduz eventualidades negativas.

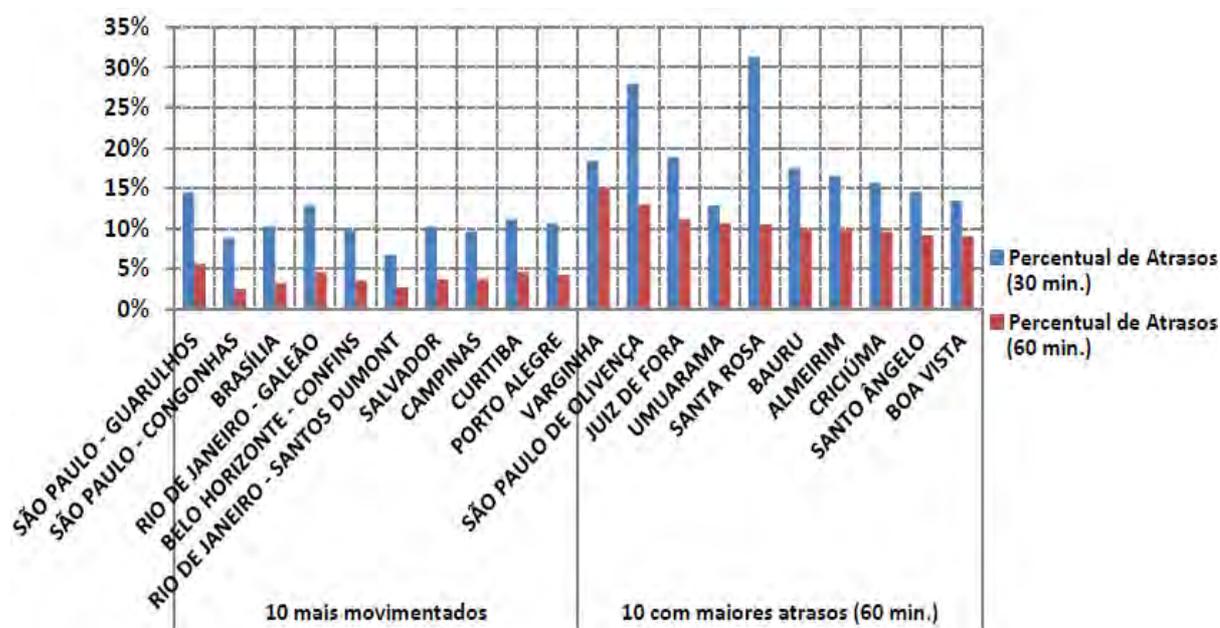


Figura 3: Percentuais de atrasos por aeroporto de origem de voo

## 1.4. Objetivos

Dentre os múltiplos aspectos relacionados às ocorrências de atrasos em voos, além de suas causas e desdobramentos, o presente trabalho limita-se à identificação de padrões de atraso em pousos e em decolagens de voos através de métodos de análise de dados, com enfoque especial em processos de formação de agrupamentos. Avalia-se especificamente, como caso amostral, a situação do Aeroporto Internacional de Brasília – Presidente Juscelino Kubitschek.

A escolha do Aeroporto Internacional de Brasília, que a partir de 2014 alcançou o posto de segundo maior do Brasil em número de movimentação de passageiros, deve-se à sua importância no cenário nacional e à sua localização estratégica no centro não só geográfico, mas também político do país. A completa caracterização do aeroporto e a análise de algumas estatísticas associadas ao seu volume de operações, bem como a

exposição de sua infraestrutura interna e suas perspectivas de crescimento, serão feitas na terceira seção deste trabalho.

É importante ressaltar que os padrões analisados de atrasos de voos no Aeroporto Internacional de Brasília, bem como em qualquer grande aeroporto, abrange uma série de variáveis das mais diversas naturezas. Essas podem estar, em geral, associadas ao volume horário de operações (em especial às horas-picos) e sua distribuição ao longo do dia, ao espaçamento médio entre movimentações consecutivas de pouso ou decolagem, a dias específicos da semana e mesmo à ocorrência de feriados ou datas comemorativas. A base de dados da qual são retiradas as informações referentes à ocorrência ou não de atrasos ou cancelamentos de voos para o Aeroporto Juscelino Kubitschek, assim como ao volume de suas operações, será apresentada na terceira seção deste trabalho.

A importância da identificação de padrões de atraso em quaisquer aeroportos reside no fato de que as autoridades responsáveis pela administração desses poderão ser munidas de um banco de informações que projete temporalmente as situações críticas de atrasos e de cancelamentos de voos que o aeroporto poderá vir a enfrentar. A partir dessa fonte de informações, os administradores do aeroporto podem estabelecer estratégias para contornar esses problemas e serão capazes de traçar planos de modo a responder de forma efetiva ao crescimento da demanda do aeroporto enquanto mantêm o nível de qualidade do controle das operações de voos. Como último fim, deseja-se estabelecer um método que possa ser aplicado como auxiliar na redução de índices de atrasos em voos e, conseqüentemente, mitigar a perda econômica associada a esse fenômeno e aumentar o nível de bem-estar dos passageiros do mercado de aviação civil.

Por fim, ressalta-se que a análise do fenômeno de atrasos em escala nacional, assim como para outros aeroportos diferentes do Aeroporto Internacional de Brasília, foge do escopo deste trabalho, muito embora tal estudo possa ser feito com a utilização de métodos semelhantes aos que aqui serão utilizados.

## **1.5. Organização do texto**

À parte esta introdução, o presente trabalho está dividido em outras quatro grandes seções. A segunda seção trata do referencial teórico necessário como pré-requisito para a compreensão deste estudo e aborda principalmente as bases estatísticas do processo de

clusterização. A seção seguinte aponta os materiais que foram utilizados como fonte de dados no processo, caracteriza a estrutura do aeroporto de Brasília e explica o método utilizado para a obtenção dos resultados. A quarta seção expõe os resultados obtidos no processo e discute os principais pontos observados. Por fim, a quinta e última seção traz consigo a conclusão do trabalho e fornece propostas para trabalhos futuros correlacionados.

## 2. REFERENCIAL TEÓRICO

### 2.1. Métodos de formação de agrupamentos

#### 2.1.1. Considerações gerais

Cabe a esta seção descrever, em suas linhas gerais, os métodos e objetivos da aplicação de um processo de agrupamento em uma base de dados, procedimento esse que será utilizado neste trabalho.

Em resumo, agrupar uma base de dados é dividi-la em conjuntos menores (grupos) de modo que elementos que apresentam maior similaridade entre si estejam no mesmo grupo e que elementos com maiores diferenças entre si sejam apresentados em grupos distintos.

Matematicamente, pode-se definir um problema de agrupamento como um tipo especial de partição de um conjunto. Particionar um conjunto  $A$  é dividi-lo em uma família de subconjuntos  $\{A_i; i \in I\}$ , tal que:

- (i)  $\bigcup_{i \in I} A_i = A$
- (ii)  $A_i \neq \emptyset \forall i \in I$
- (iii)  $A_i \cap A_j = \emptyset \forall i, j \in I \text{ e } i \neq j$

Em especial, um processo de agrupamento sobre um conjunto difere de uma partição qualquer sobre esse pelo fato, já mencionado, de que os elementos em cada um dos subconjuntos (grupos) apresentam um maior grau de similaridade entre si do que com os elementos dos demais subconjuntos.

Naturalmente, a partição de um conjunto por grupos apresenta algumas vantagens. De acordo com Webb [12], a análise de uma divisão em grupos resulta potencialmente na observação de estruturas internas na base de dados que podem ser usadas na geração de hipóteses sobre a mesma. Não à toa, métodos de agrupamento são utilizados em diversas áreas do conhecimento humano, tais como economia, psicologia, sociologia, medicina, entre outras.

### 2.1.2. Medidas de similaridade

É de fundamental importância que se estabeleça, em um processo de agrupamento, quais serão as variáveis analisadas de modo a identificar o grau de similaridade entre os elementos de um conjunto. Para que se ilustre a necessidade de se estabelecer um critério de comparação, imagine que, por exemplo, as cinquenta e duas cartas de um baralho podem ser divididas em grupos de acordo com seus naipes, cores ou mesmo valores numéricos, de modo que há, em cada uma das situações, partições perfeitas de acordo com as similaridades apontadas, muito embora totalmente distintas entre si. Para que se possa definir quais elementos devem estar presentes ou não em um mesmo cluster é necessário, portanto, que se utilize uma medida prévia de similaridade entre os elementos do conjunto.

A distância entre dois pontos em um conjunto é usualmente tratada como uma importante medida da similaridade entre eles, já que pontos próximos em um espaço  $p$ -dimensional tendem a apresentar maiores semelhanças entre seus atributos. Como medida de distância se utiliza com frequência a distância euclidiana, a fim de minimizar a distância intragrupos maximizá-la intergrupos.

A Figura 4 ilustra a divisão, feita partir de um critério de similaridade por meio de distância euclidiana entre pontos, de alguns elementos em três clusters em um espaço amostral genérico com a respeito a duas variáveis.

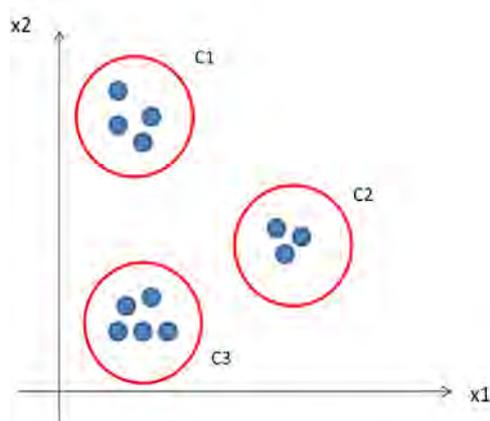


Figura 4: Exemplo ilustrativo de uma clusterização por critério de distância euclidiana

### 2.1.3 Métodos hierárquicos

Procedimentos que envolvem métodos hierárquicos são os mais comuns em processos de agrupamento de dados. Os elementos do conjunto base podem ser dispostos em um diagrama de árvore chamado dendograma, por meio do qual os grupos podem ser particionados ao se selecionar um determinado nível da árvore. Se dois grupos são selecionados em níveis distintos do dendograma, então um deles contém inteiramente o outro ou então não possuem elementos em comum. Um exemplo prático de divisão hierárquica provém da taxonomia biológica, em que reinos são divididos em filos; filos em subfilos; subfilos em classes; classes em ordens, e assim por diante, de modo que um grupo hierarquicamente inferior está sempre contido em um superior ou é disjunto desse.

A medida de similaridade entre dois grupos quaisquer deve ser previamente determinada para a formação de agrupamentos hierárquicos. Há diferentes métodos que fornecem parâmetros para que seja aferida essa medida, também chamada de distância entre grupos. Os métodos hierárquicos de formação de agrupamento diferem basicamente na escolha do modo no qual se calcula o grau de dissimilaridade entre grupos. Por exemplo, o método do centroide substitui cada grupo por um elemento médio que representa geometricamente o centro desse; pelo método do vizinho mais próximo, a distância intergrupos é calculada como o mínimo dentre as distâncias euclidianas de todos possíveis os pares de elementos entre os grupos, enquanto o método do vizinho mais distante usa essa máxima distância como parâmetro.

O método hierárquico de Ward para formação de agrupamentos será o utilizado neste trabalho. Esse método não calcula a distância intergrupos para a construção desses, mas, diferentemente, preocupa-se em maximizar o grau de homogeneidade dentro de cada grupo. A minimização da soma dos quadrados das distâncias dos pares de elementos dentro de cada grupo é utilizada como a medida de homogeneidade.

### 3. MATERIAIS E MÉTODOS

#### 3.1. Base de dados

Para que se fundamente um método de identificação de padrões de atrasos de voos, é imprescindível a posse de uma base de dados confiável que apresente as movimentações diárias nos aeroportos em questão. Ainda, é absolutamente necessário que essa base de dados indique eventuais cancelamentos de voos ou alterações de horários previstos para partidas e decolagens para que, por sua vez, possam ser comparados com os horários reais dessas movimentações a fim de que se verifique a eventualidade de atrasos.

Para tanto, o presente trabalho se vale da base de dados VRA (Voo Regular Ativo) [11] fornecida pela Agência Nacional de Aviação Civil (ANAC). A base VRA é construída a partir de informações de voos de empresas de transporte aéreo regular e dos dados fornecidos pela base de dados de Horário de Transporte (HOTRAN).

Para cada voo indicado na base de dados VRA, indica-se o aeroporto de origem e o de destino, a companhia aérea operante, a data e horário das movimentações previstas e das efetivamente realizadas, além de eventuais informações sobre cancelamentos de voos e suas respectivas razões e mesmo sobre a realização de voos não originalmente previstos.

A título de ilustração, a Tabela 1 apresenta algumas das informações relevantes que podem ser tomadas da base de dados VRA. Nesse caso, tomou-se à esmo um trecho da tabela dessa base de dados para alguns poucos voos realizados em Outubro de 2012.

É importante que se note, por meio da Tabela 1, que a base VRA fornece, para cada um dos voos relacionados, aeroporto de destino e de origem através da nomenclatura ICAO; note-se também que voos cancelados não apresentam, naturalmente, horários de partida ou chegada reais, mas apenas os respectivos horários previstos.

**Tabela 1: Exemplos de informações recolhidas da base de dados VRA**

<b>Sigla da empresa</b>	<b>Aeroporto Origem</b>	<b>Aeroporto Destino</b>	<b>Partida Prevista</b>	<b>Partida Real</b>	<b>Chegada Prevista</b>	<b>Chegada Real</b>	<b>Situação</b>
TAM	SBSR	SBSP	10/23/12 5:50	10/23/12 5:50	10/23/12 6:45	10/23/12 6:45	Realizado
DSM	SAEZ	SBGR	10/18/12 5:50	10/18/12 5:50	10/18/12 9:00	10/18/12 9:00	Realizado
TAM	SBSR	SBSP	10/29/12 5:50	10/29/12 5:50	10/29/12 6:45	10/29/12 6:45	Realizado
AZU	SBPA	SBGL	10/23/12 5:50		10/23/12 7:40		Cancelado
TIB	SBCG	SBLO	10/15/12 5:50	10/15/12 5:50	10/15/12 7:05	10/15/12 7:05	Realizado
GLO	SBMK	SBCF	10/10/12 5:50	10/10/12 5:50	10/10/12 6:38	10/10/12 6:38	Realizado
TUS	SBGR	SBEG	10/12/12 5:50	10/12/12 5:50	10/12/12 9:42	10/12/12 9:42	Realizado
GLO	SBMK	SBCF	10/1/12 5:50	10/1/12 5:50	10/1/12 6:38	10/1/12 6:38	Realizado

Como a proposta do presente trabalho se limita à análise dos voos comerciais referentes ao Aeroporto Internacional de Brasília, foi necessária uma limpeza da base de dados de modo que essa apontasse apenas os voos com aeroporto de origem ou aeroporto de destino relacionados com a sigla SBBR, nomenclatura ICAO para o Aeroporto Presidente Juscelino Kubitschek. Foram desconsiderados, para quaisquer efeitos, os demais voos apontados na base de dados VRA referentes aos demais aeroportos brasileiros ou internacionais.

Foram avaliadas todas as movimentações previstas e realizadas no Aeroporto Internacional de Brasília entre Janeiro de 2010 e Junho de 2014, segundo apontadas pelo VRA.

## **3.2. Caracterização do Aeroporto Internacional de Brasília**

### **3.2.1. Considerações gerais**

O Aeroporto Juscelino Kubitschek, de nomenclatura SBBR (ICAO) e BSB (IATA), localiza-se na capital do país e alcançou, a partir de 2014, o posto de segundo maior aeroporto do Brasil em número de passageiros em movimentação. Para que se tenha uma ideia da ordem de grandeza associada à movimentação dentro do aeroporto, em Dezembro de 2014 houve a circulação de aproximadamente 1,69 milhões de

passageiros, seja em embarque, desembarque ou conexão, para um número de 15.909 voos.

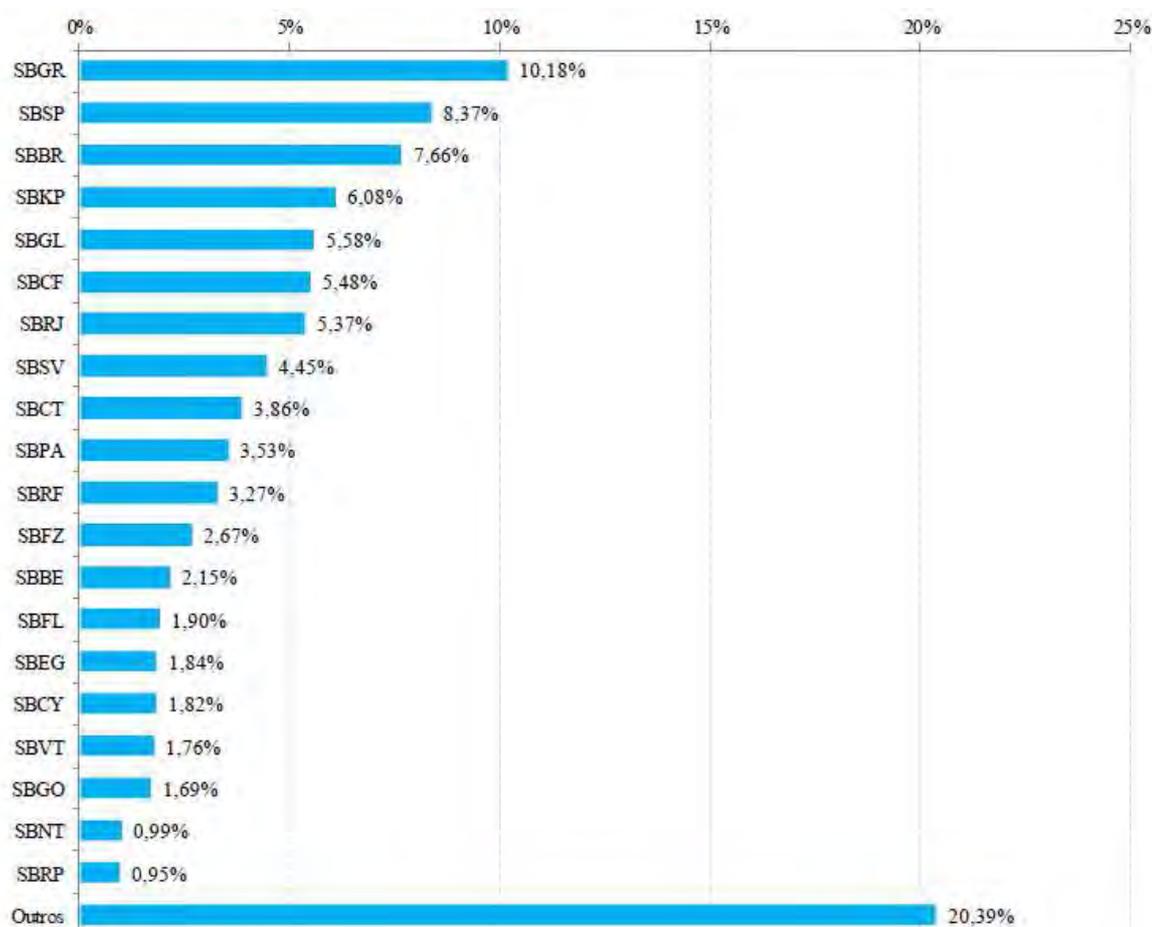
Por sua posição geográfica favorável, no centro do país, e por seu natural volume de movimentação, o Aeroporto Internacional de Brasília é largamente utilizado como ponto de conexão para algumas das principais cidades do Brasil, funcionando como uma dos principais hubs no cenário nacional. As principais companhias aéreas brasileiras, em especial a TAM, GOL, Azul e Avianca, operam com frequência no aeroporto, bem como grandes companhias internacionais, como a TAP, American Airlines e Air France.

Acompanhando uma tendência verificada em cenário nacional, o número de passageiros que passam pelo Aeroporto Juscelino Kubistchek, e conseqüentemente dependem de suas operações de voo e de sua infraestrutura, cresceu continuamente nos últimos anos. Uma evolução histórica do crescimento do aeroporto, em se tratando do número de passageiros em movimentação, é apontada na Tabela 2. Observe um crescimento acentuado, sempre medido em relação ao ano anterior, a partir do ano 2009. Naturalmente, um crescimento dessas proporções em um curto espaço de tempo traz consigo uma série de preocupações referentes à própria logística do aeroporto e ao controle sobre os horários dos voos.

**Tabela 2: Movimentação de passageiros no Aeroporto Internacional de Brasília entre 2004 e 2014**

<b>Ano</b>	<b>Número de passageiros em movimento</b>	<b>Taxa de crescimento em relação ao ano anterior</b>
2014	18.146.405	+10%
2013	16.489.987	+3,8%
2012	15.891.530	+3,2%
2011	15.398.737	+7,3%
2010	14.347.061	+15,8%
2009	12.213.825	+16,95%
2008	10.443.393	-6%
2007	11.119.872	+14,6%
2006	9.699.911	+2,8%
2005	9.426.569	-5%
2004	9.926.786	+45,1%

A relevância do Aeroporto Internacional de Brasília no cenário nacional pode ser devidamente considerada ao se avaliar os dados expostos na Figura 5, retirados do Anuário do Transporte Aéreo de 2013, organizado pela ANAC. Os dados apontam a participação dos 20 principais aeroportos brasileiros na quantidade de decolagens, restringidas ao mercado doméstico, no ano de 2013. O Aeroporto Juscelino Kubitschek é responsável por 7,66% de todas essas movimentações consideradas em escala nacional.



**Figura 5: Participação dos 20 maiores aeroportos brasileiros na quantidade de decolagens para voos domésticos em 2013**

### 3.2.2. Perspectivas

O Aeroporto Juscelino Kubitschek foi cedido, a partir de Dezembro de 2012, à administração privada por meio de um sistema de concessão para um período de vinte e cinco anos. O Consórcio Inframérica Aeroportos, uma das maiores operadoras de

aeroportos do mundo, ficou responsável por conduzir sua administração, papel que antes cabia à estatal INFRAERO.

A partir de sua concessão o aeroporto passou por uma série de modernizações e processos de ampliação estrutural a fim de adequar-se ao aumento progressivo da demanda. Foram construídas duas novas salas de embarque, o Píer Norte e o Píer Sul, que foram responsáveis por aumentar de 13 para 29 o número de pontes de embarque, além de um crescimento nos setores comerciais dentro do aeroporto, como lojas de conveniência e novas praças de alimentação. Ainda, a capacidade de pátio para estacionamento de aeronaves também foi ampliada para 70 posições, com consequente redução de 15% no uso de ônibus para transporte de passageiros para acesso às aeronaves em posições remotas.

A Figura 6 ilustra algumas das principais áreas do Aeroporto Internacional de Brasília, com exceção das pistas de pouso, decolagem e taxiways. É possível observar os dois píeres recém-construídos, que foram responsáveis pelo aumento do número de portões de embarque, como também o posicionamento das aeronaves no pátio para embarque ou desembarque.



**Figura 6: Vista aérea do Aeroporto Internacional de Brasília**

O grande destaque nas operações no Aeroporto Juscelino Kubitschek fica por conta de suas pistas de pouso e decolagem, assim como as taxiways. As pistas paralelas do aeródromo (de medidas 3300 m x 45 m e 3200 m x 45 m, com um espaçamento de 1,8 km entre elas) são as únicas do Brasil posicionadas a uma distância segura de modo a possibilitar ao aeroporto a realização de pousos e decolagens simultâneos. Com as recentes expansões e melhorias na infraestrutura do aeroporto, o Departamento de Controle do Espaço Aéreo (DECEA), órgão subordinado ao Comando da Aeronáutica que tem por responsabilidade o planejamento e a gerência do controle do espaço aéreo brasileiro, analisou e acabou por autorizar os procedimentos de operações simultâneas nas duas pistas. A previsão para que se inicie esse novo tipo de operação está para Novembro de 2015.

A Figura 7 representa uma vista aérea do Aeroporto Internacional de Brasília. Na imagem, é possível visualizar as duas grandes pistas, paralelas entre si, em que são realizadas as operações de pouso e decolagem. Note o espaçamento considerável entre elas, causa de que seja possível a realização de operações simultâneas em ambas as pistas.



**Figura 7: Vista aérea da organização das pistas do Aeroporto Internacional de Brasília**

Devido principalmente às alterações citadas, o Aeroporto Internacional de Brasília passará à primeira posição do país em termos de capacidade de pista para pousos e decolagens, com uma variação de 60 movimentos máximos por hora para um total de 80. Ressalta-se, a título de comparação, que o aeroporto também será o primeiro em toda a América do Sul a operar duas pistas simultaneamente, equiparando-se a outros grandes aeroportos do mundo.

### **3.3. Métodos de análise de dados**

A partir da base de dados fornecida pelo VRA, operou-se uma série de análises com o intuito de se gerar os resultados desejados para a formação de clusters e posterior avaliação. As etapas do processo estão sumarizadas a seguir.

#### 1) Limpeza dos dados

A base de dados traz consigo uma grande série de informações que não são úteis para o tipo de análise que é proposta neste trabalho. Retirou-se do banco, portanto, todos os dados desnecessários, tais quais o código de cada voo, a justificativa de eventuais cancelamentos, o tipo de linha ou a companhia aérea responsável por cada trajeto. Ainda, todos os voos que não tinham o Aeroporto Internacional de Brasília como aeroporto de origem ou destino foram excluídos da base de dados.

#### 2) Preparação dos dados

Os dados de cada voo foram separados de acordo com sua situação: cancelados ou realizados. Para os voos realizados previstos, a partir da diferença entre o horário da movimentação real e da movimentação prevista, foi possível o cálculo de eventuais atrasos bem como da grandeza desses. Todos os voos realizados com atrasos iguais ou superiores a quinze minutos foram considerados como atrasados, enquanto os demais foram considerados em situação regular.

Para a formação dos vetores diários de movimentação ou atrasos utilizados na formação dos grupos, foi necessário a discretização dos horários de movimentação em horários aproximados. Para isso, utilizou-se intervalos de 20 minutos a partir da

primeira hora do dia, de modo que o instante horário de cada movimentação fosse aproximado para o múltiplo de 20 mais próximo.

Assim, para cada dia foi estabelecido o número de voos previstos e realizados, bem como o número de atrasos, a cada intervalo de 20 minutos.

As amostras diárias, conforme mencionado nesta seção, vão de 01/01/2009 a 31/05/2014, resultando em um total de 1977 dias. Cada um desses dias, nesse processo de divisão, é considerado um vetor em que cada elemento representa ordenadamente o acumulado da última hora em número de movimentações previstas, acumuladas ou número de atrasos, iniciando à 1 da manhã e com incrementos de 20 em 20 minutos até o último minuto do dia. Para cada uma dessas três diferentes variáveis é possível a confecção de uma matriz de dissimilaridade, em que as linhas representam os dias considerados, enquanto as colunas apontam os horários arredondados de movimentação.

A título de exemplo, a matriz de dissimilaridade de movimentações previstas traz, em sua primeira linha e coluna, o número de movimentações previstas no Aeroporto Internacional de Brasília para a primeira hora do dia 01/01/2009; a segunda coluna da mesma linha o acumulado da última hora até às 1h20min da manhã; a terceira coluna até 1h40min; e assim sucessivamente, até a última coluna, com o acumulado da última hora do dia.

Em cada uma dessas três matrizes citadas, será feito o processo de formação de grupos conforme o método de Ward, por esse se preocupar principalmente com a coerência interna dentro de cada grupo, conforme mencionado na segunda seção deste trabalho.

### 3) Formação dos grupos

A participação dos dias em grupos de acordo com a similaridade de perfis de movimentação ou de atraso foi feita com o auxílio do software R versão 3.2.0. Esse programa possibilita a criação dos clusters, que nada mais são que essas partições, através de algum método à escolha do usuário. No caso, como já explicado, optou-se pelo método de Ward.

Foram feitos três agrupamentos distintos: perfis de movimentação diária prevista, de movimentação diária realizada e de índices horários de atrasos. Cada conjunto desses contém 1977 elementos, o número de dias considerados na análise. É

importante lembrar que um elemento em qualquer um desses grupos é um vetor ordenado em que cada elemento representa o número de movimentações acumuladas na última hora, em intervalos de 20 em 20 minutos, ou o número de atrasos, computados seguindo os mesmos critérios.

Como dado de saída, o software gera, por tratar-se de um método hierárquico, um dendograma. A escolha do número de grupos foi feita por um modo qualitativo a partir da caracterização do próprio dendograma, de um modo que a escolha parecesse natural com esse e que não se optasse por um número excessivo de grupos, de modo a tornar impossível a identificação de padrões de formação para cada um deles; nem que se escolhesse um número tão reduzido de grupos em que tais padrões não pudessem emergir.

Escolhido um número  $k$  de grupos, o software particiona os vetores de movimentação ou atrasos diários dentro de cada um dos conjuntos  $X_1, X_2, X_3, \dots, X_k$  de acordo com o grau de similaridade interna dos elementos em relação à variável escolhida na partição. Assim, por exemplo, dois dias estarão dentro do mesmo grupo gerado para os índices de atrasos se apresentarem níveis semelhantes dessa taxa ao longo do dia.

São montados assim três agrupamento distintos, uma para cada variável escolhida: partição de acordo com movimentações previstas, formando o conjunto  $P$ , de modo que cada grupo formado em respeito a essa variável é um subconjunto de  $P$ ; analogamente, de movimentações realizadas,  $R$ ; e também para os índices horários de atraso, formando o conjunto  $A$ .

Os grupos formados são analisados na seção seguinte deste trabalho.

#### 4) Análise dos resultados

Dadas as três diferentes partições citadas anteriormente, busca-se observar se há algum tipo de padrão temporal que determine se há algum padrão na formação dos conjuntos particionados. Assim, verifica-se se há predominância dos diferentes dias da semana dentro de algum dos grupos formados, bem como se há uma diferença nesses grupos com a evolução do tempo. Outra variável construída na busca desses padrões é o espaçamento médio previsto entre duas movimentações consecutivas. Para a criação dessa variável, divide-se o acumulado horário dessas movimentações por 60, de modo a

encontrar o intervalo médio, em minutos, de duas movimentações consecutivas de acordo com o previsto.

Os resultados alcançados são apontados na próxima seção deste trabalho, assim como as análises feitas para a identificação de possíveis padrões nas construções dos perfis de movimentação diária do aeroporto e de seus índices de atrasos.

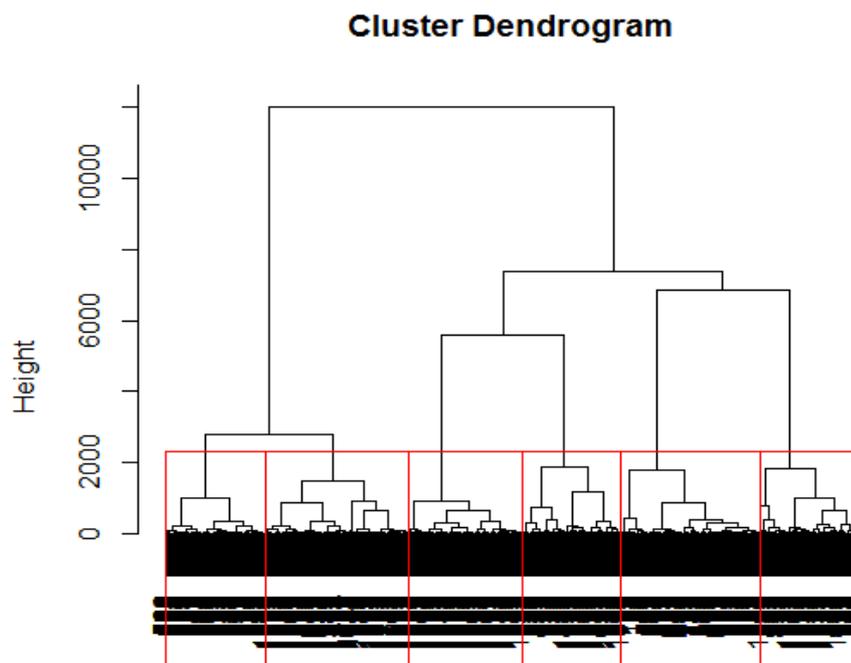
## 4. RESULTADOS E DISCUSSÕES

### 4.1. Movimentações previstas

A partir do acumulado horário da movimentação diária prevista é feita a primeira das divisões em grupos em busca de perfis de similaridade de movimentações nas pistas do Aeroporto Internacional de Brasília. Esses primeiros perfis representam a movimentação ideal, conforme planejado, de chegadas e partidas de aeronaves no aeródromo do aeroporto. Naturalmente, voos cancelados, mas inicialmente previstos, ainda são levados em conta, enquanto que voos realizados, mas fora da previsão regular, não são considerados.

A imagem do dendograma gerado pelo software, de onde foi tirada a divisão em grupos, está representada na Figura 8. Uma divisão em 6 grupos é escolhida a partir do próprio dendograma formado e apresenta um número pertinente para que se identifique possíveis padrões, uma vez que não é tão baixo a ponto de que não se formem perfis distintos nem tão alto a ponto de que essa identificação se torne impossibilitada. Observe que nessa divisão os elementos desses grupos estão delimitados e separados entre si pelos seis grandes retângulos inferiores. É importante recordar que o nível de similaridade cresce de cima para baixo, de modo que a medida do eixo y do diagrama representa a distância métrica entre os grupos calculado pelo método Ward.

Os diferentes grupos formados serão nomeados como  $P_1, P_2, P_3, P_4, P_5$  e  $P_6$  e são os elementos do conjunto  $P$ , que é definido como o conjunto dos grupos formados a partir do agrupamento dos dias considerados em respeito ao grau de similaridade do acumulado horário das movimentações previstas. Assim,  $P = (P_1, P_2, P_3, P_4, P_5, P_6)$ . Resta identificar a presença de padrões que marcam a divisão desses grupos.



**Figura 8: Diagrama de árvore da divisão de clusters para movimentações previstas**

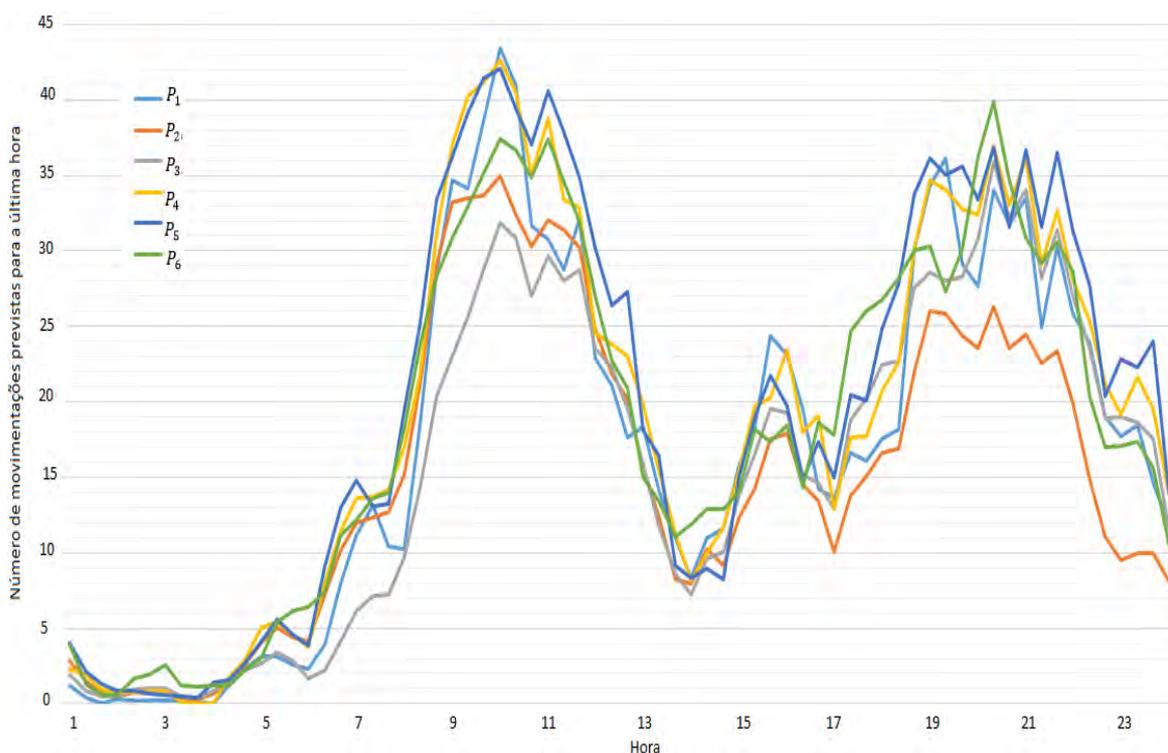
O número de elementos de cada um dos clusters é apontado na Tabela 3.

**Tabela 3: Contagem do número de elementos de cada cluster do conjunto P**

Grupo	Número de elementos
$P_1$	399
$P_2$	282
$P_3$	281
$P_4$	282
$P_5$	412
$P_6$	321
Total Geral	<b>1977</b>

É possível a construção do gráfico do número médio de movimentações horárias previstas ao longo do dia para cada um dos clusters. Tal gráfico é apresentado a seguir na Figura 9. Observe que as diferenças de comportamento são praticamente inexistentes em horas de baixa movimentação como a madrugada, enquanto que nas horas-pico da manhã e da noite (morning-peak e evening-peak) as diferenças são mais acentuadas.

Diferenças entre os grupos são facilmente visualizados no gráfico. Por exemplo, durante a manhã, o grupo  $P_3$  apresenta as mais baixas taxas de movimentação, enquanto apresenta maiores valores no pico noturno, ao passo que  $P_2$  apresenta comportamento contrário. De qualquer forma, as taxas de movimentação desses dois grupos são praticamente sempre inferiores às dos clusters  $P_4$  e  $P_5$ .



**Figura 9: Perfis de movimentação do acumulado horário para cada grupo do conjunto P**

Com a distribuição em dias da semana, conforme apreendido na Tabela 4, os primeiros padrões começam a emergir. Observe que o grupo  $P_3$  representa, praticamente em nível integral e exclusivo, os perfis de movimentação dos domingos, enquanto o grupo  $P_2$  representa o conjunto dos sábados. Os demais grupos apresentam relativa uniformidade em relação à distribuição dos dias úteis da semana, sem variações aparentemente representativas de valores, apontando certa uniformidade das movimentações previstas para esses dias.

**Tabela 4: Distribuição dos dias de semana nos elementos do conjunto P**

Dia da semana	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
<b>Domingo</b>	0,00%	0,00%	99,65%	0,00%	0,35%	0,00%
<b>Segunda</b>	28,37%	0,00%	0,00%	19,86%	28,72%	23,05%
<b>Terça</b>	28,37%	0,00%	0,00%	19,86%	29,08%	22,70%
<b>Quarta</b>	28,37%	0,00%	0,00%	19,86%	29,08%	22,70%
<b>Quinta</b>	28,27%	0,00%	0,00%	20,14%	28,98%	22,61%
<b>Sexta</b>	27,92%	0,35%	0,00%	20,14%	28,98%	22,61%
<b>Sábado</b>	0,00%	99,29%	0,00%	0,00%	0,71%	0,00%
<b>Total Geral</b>	<b>20,18%</b>	<b>14,26%</b>	<b>14,21%</b>	<b>14,26%</b>	<b>20,84%</b>	<b>16,24%</b>

A distribuição dos dias de cada ano nos elementos do conjunto  $P$  é apontado na Tabela 5 abaixo.

**Tabela 5: Distribuição do número de dias de cada ano nos elementos do conjunto  $P$**

Ano	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$
<b>2009</b>	71,23%	14,52%	14,25%	0,00%	0,00%	0,00%
<b>2010</b>	38,08%	14,25%	14,25%	33,42%	0,00%	0,00%
<b>2011</b>	0,00%	14,52%	14,25%	43,84%	27,40%	0,00%
<b>2012</b>	0,00%	14,21%	14,48%	0,00%	71,31%	0,00%
<b>2013</b>	0,00%	13,70%	13,97%	0,00%	13,97%	58,36%
<b>2014</b>	0,00%	14,57%	13,91%	0,00%	0,00%	71,52%

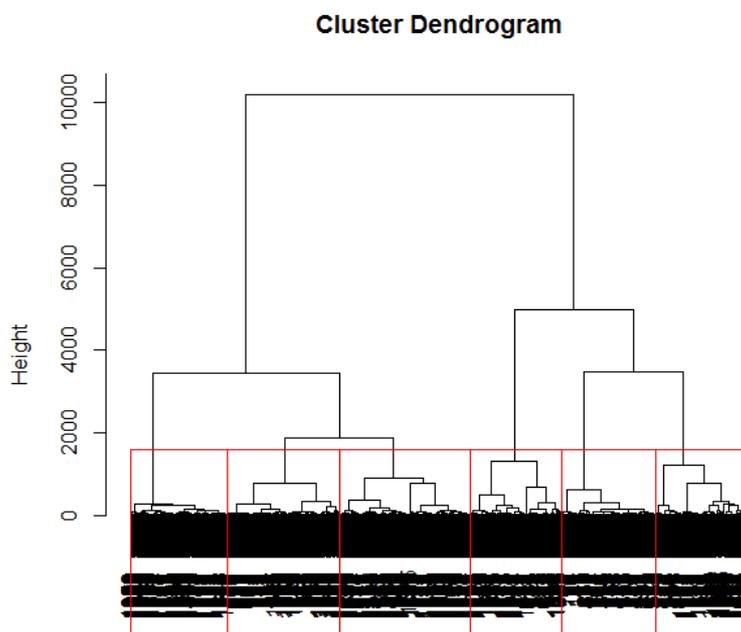
Um novo padrão de formação dos grupos é visível. Observe que os conjuntos  $P_2$  e  $P_3$  representam cada um, numericamente, um sétimo do número de dias de cada ano, o que é compatível com a representatividade de cada um deles no total de sábados e domingos do espaço temporal. Para os demais dias úteis, o modo de distribuição é temporal: o grupo  $P_1$  embarca os dias úteis de 2009 e parte de 2010,  $P_4$  está distribuído entre os dias úteis 2010 e 2011, enquanto  $P_5$  está associado principalmente a dias de 2011, 2013 e principalmente 2012 e, por fim,  $P_6$  abrange todos os dias úteis considerados de 2014 e a maior parte dos de 2013.

Essa divisão em dias da semana e em anos, associada ao gráfico representado na Figura 7, já encaminham uma boa ideia do que seja cada um dos elementos do conjunto  $P$ . Resta ainda que se analise uma divisão em grupos para as movimentações efetivamente realizadas e para as taxas de atraso, de modo que, relacionando entre si os dados encontrados, possa emergir perfis distintos em que possam ser identificados certos padrões de atraso em voos para o Aeroporto Internacional de Brasília.

## 4.2. Movimentações realizadas

É feita uma divisão em clusters para as movimentações efetivamente realizadas na pista do aeroporto análoga à que foi feita para as movimentações previstas. Nessas movimentações, são desconsiderados os voos cancelados e são adicionados os voos não previamente planejados. Para esses últimos voos não existe a possibilidade que se calcule o atraso por meio dos dados do VRA, uma vez que o atraso é calculado a partir da diferença entre o horário da movimentação prevista, que é inexistente nesses casos, e o horário da movimentação realizadas.

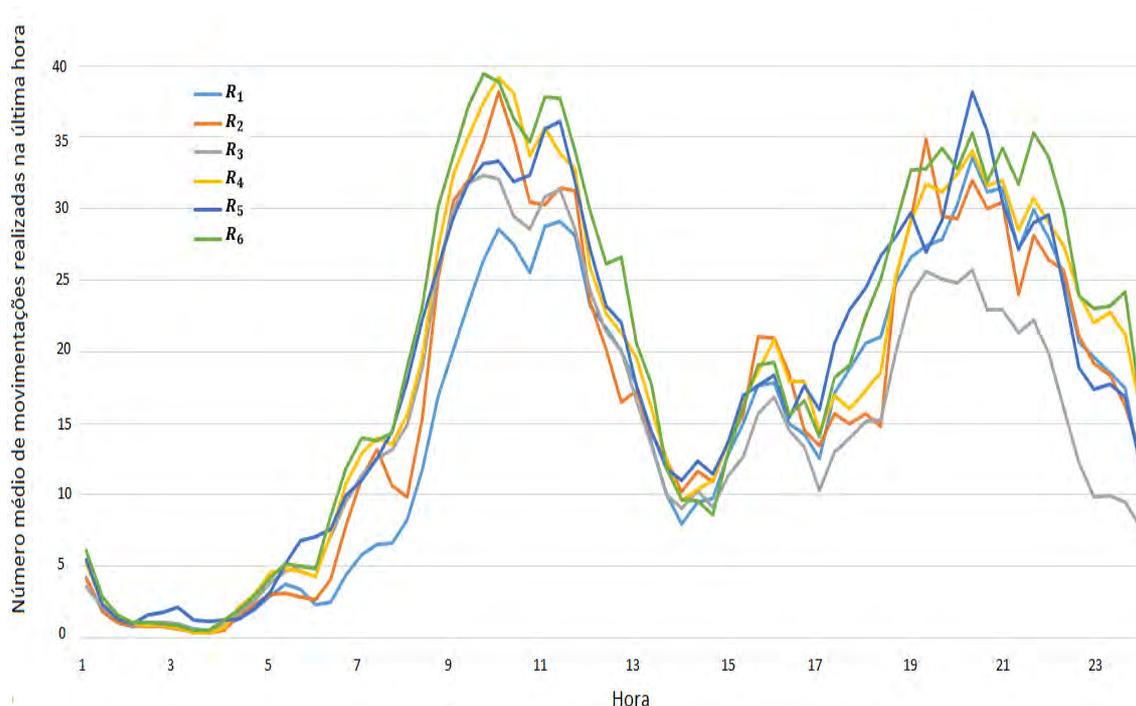
O diagrama de árvore encontrado para a divisão em clusters é representado na Figura 10.



**Figura 10: Diagrama de árvore da divisão de clusters para as movimentações realizadas**

Mais uma vez, uma divisão em 6 grupos parece natural, conforme indicado na Figura 10. Os grupos encontrados serão denominados  $R_1, R_2, R_3, R_4, R_5$  e  $R_6$  e são considerados como formadores do conjunto  $R$ , que representa o conjunto dos clusters do acumulado horário das movimentações diárias efetivamente realizadas. Assim, tem-se que  $R = (R_1, R_2, R_3, R_4, R_5, R_6)$ .

Os perfis horários de movimentação de cada um dos seis grupos é indicado no gráfico da Figura 11. Por meio dessa, é possível observar as diferenças de comportamento entre os elementos do conjunto  $R$ .



**Figura 11: Perfis de movimentação do acumulado horário para cada grupo do conjunto R**

A contagem do número de elementos em cada elemento do conjunto R é apontado na Tabela 6. A Tabela 7 indica a divisão por dia da semana entre esses grupos, enquanto a Tabela 8 estabelece a divisão dos dias de cada ano entre eles.

É possível observar alguns padrões de distribuição dos dias da semana por meio da Tabela 7. Observe que, na prática, o grupo  $R_1$  representa os domingos, enquanto o  $R_3$  engloba os sábados. Com o auxílio da Figura 11, nota-se, então, que as manhãs de domingo representam os menores índices de movimentação dentre todas as manhãs, enquanto as noites de sábado são as menos movimentadas entre todas as noites. Os dias úteis da semana não parecem, a princípio, organizarem-se em algum padrão dentro dos demais grupos.

**Tabela 6: Contagem do número de elementos de cada grupo do conjunto R**

Grupo	Número de elementos
$R_1$	300
$R_2$	300
$R_3$	290
$R_4$	361
$R_5$	308
$R_6$	418
Total Geral	<b>1977</b>

**Tabela 7: Distribuição dos dias de semana nos elementos do conjunto R**

Dia da semana	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$
<b>Domingo</b>	97,87%	0,00%	0,71%	0,71%	0,00%	0,71%
<b>Segunda</b>	2,13%	21,63%	0,71%	25,89%	21,63%	28,01%
<b>Terça</b>	1,77%	21,63%	0,00%	24,82%	22,34%	29,43%
<b>Quarta</b>	1,06%	22,34%	0,00%	26,24%	21,99%	28,37%
<b>Quinta</b>	0,35%	20,49%	0,35%	25,44%	23,32%	30,04%
<b>Sexta</b>	1,77%	20,14%	3,18%	24,38%	19,79%	30,74%
<b>Sábado</b>	1,41%	0,00%	97,53%	0,35%	0,00%	0,71%
Total Geral	<b>15,17%</b>	<b>15,17%</b>	<b>14,67%</b>	<b>18,26%</b>	<b>15,58%</b>	<b>21,14%</b>

**Tabela 8: Distribuição do número de dias de cada ano nos elementos do conjunto R**

Ano	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$
<b>2009</b>	15,34%	66,85%	16,16%	1,64%	0,00%	0,00%
<b>2010</b>	15,07%	15,34%	13,97%	55,34%	0,27%	0,00%
<b>2011</b>	13,70%	0,00%	15,07%	35,62%	0,27%	35,34%
<b>2012</b>	15,85%	0,00%	14,21%	4,37%	0,55%	65,03%
<b>2013</b>	15,34%	0,00%	13,97%	1,10%	55,89%	13,70%
<b>2014</b>	16,56%	0,00%	14,57%	1,99%	66,23%	0,66%

Algo semelhante ao encontrado na análise das movimentações previstas ocorre quanto à distribuição temporal dos dias de cada ano nos elementos de  $R$ . Enquanto os conjuntos  $R_1$  e  $R_3$  representam aproximadamente um sétimo do número de dias de cada ano, o que na prática representa o total de domingos e sábados, os demais dias úteis parecem ser distribuídos de acordo com o ano em que estão: os dias úteis de 2009 estão em  $R_2$ ; uma pequena parcela de 2010 em  $R_2$  e maior em  $R_4$ ; 2011 em  $R_4$  e  $R_6$ , e assim por diante, como indicado na Tabela 8.

Por fim, é útil que se estabeleça se há alguma relação entre os elementos do conjunto  $P$ , das movimentações previstas, e o conjunto  $R$ , das movimentações realizadas. A Tabela 9 aponta numericamente a distribuição dos dias de cada um dos clusters de  $P$  em relação aos de  $R$ .

**Tabela 9: Distribuição da quantidade de elementos dos grupos de movimentações previstas nos clusters movimentações realizadas**

Grupo	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	$R_6$
$P_1$	1,50%	74,19%	1,25%	23,06%	0,00%	0,00%
$P_2$	1,42%	0,00%	98,23%	0,35%	0,00%	0,00%
$P_3$	98,22%	0,00%	0,71%	0,71%	0,00%	0,36%
$P_4$	0,00%	1,42%	1,42%	84,75%	0,71%	11,70%
$P_5$	1,70%	0,00%	0,00%	4,85%	0,97%	92,48%
$P_6$	2,18%	0,00%	0,62%	2,18%	94,08%	0,93%

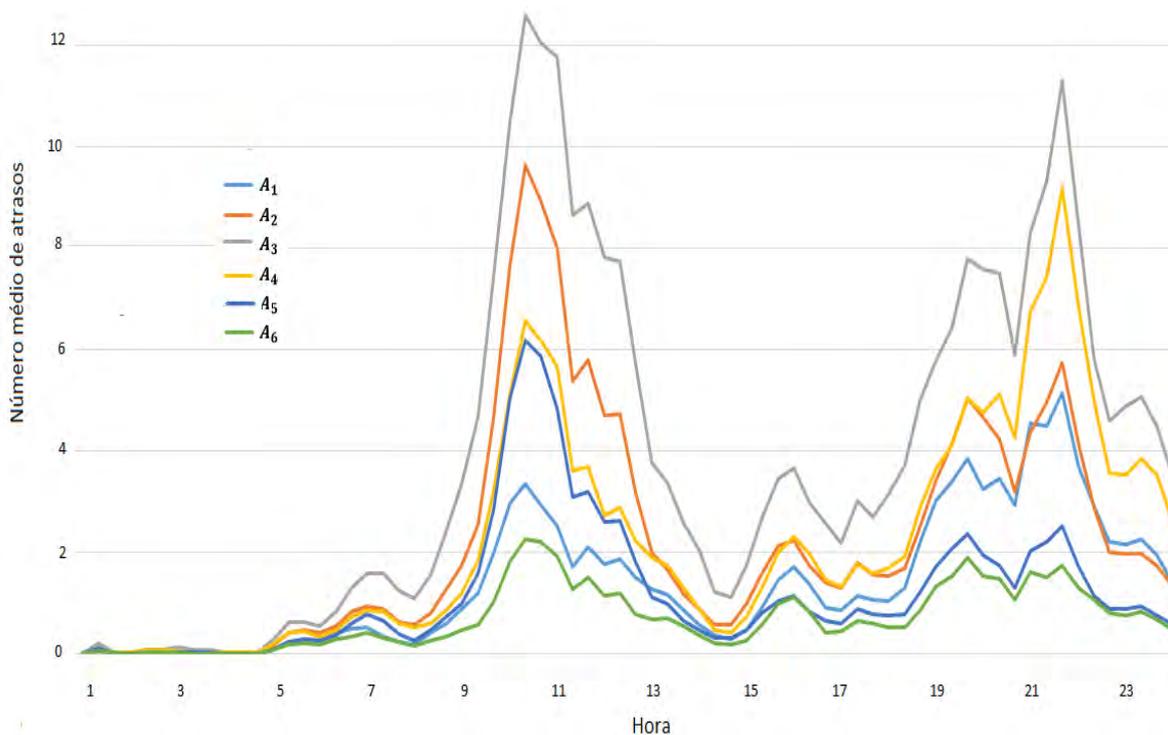
É possível notar, através da Tabela 9 uma forte correlação entre os elementos de  $P$  e de  $R$ , de modo que cada grupo de movimentações previstas pode ser associada predominantemente e um de movimentações realizadas.

### 4.3 Atrasos

Por fim, é feita a divisão dos dias em grupos de acordo com seus perfis de atraso. Como mencionado na seção anterior, são considerados voos atrasados aqueles com um intervalo de tempo maior ou igual a quinze minutos entre a movimentação prevista pelo aeroporto para determinado voo e a efetivamente realizada.

De modo análogo às divisões anteriores, foram formados seis grupos de acordo com a divisão apontada pelo software R. É construído, assim, o conjunto  $A$ , que representa a união dos 6 grupos de movimentações horárias acumuladas de atrasos no Aeroporto Internacional de Brasília. Desse modo,  $A = (A_1, A_2, A_3, A_4, A_5, A_6)$ .

A Figura 12 traz o gráfico com o perfil médio de cada um dos elementos do conjunto  $A$ . É possível observar padrões bem definidas para cada um dos grupos de atraso. Por exemplo, o grupo  $A_3$  apresenta invariavelmente as maiores taxas de atraso ao longo do dia, enquanto o  $A_6$ , as menores; o grupo  $A_2$  aponta taxas maiores de atraso no pico matinal em relação ao pico noturno, acontecendo o contrário com  $A_4$ .



**Figura 12: Perfis de movimentação do acumulado horário para cada grupo do conjunto A**

O número de elementos em cada grupo do conjunto  $A$  é contado na Tabela 10.

**Tabela 10: Contagem do número de elementos de cada elemento do conjunto A**

Grupo	Número de elementos
$A_1$	388
$A_2$	406
$A_3$	239
$A_4$	285
$A_5$	369
$A_6$	290
Total Geral	<b>1977</b>

A Figura 12 traz o gráfico com o perfil médio de cada um dos elementos do conjunto  $A$ . É possível observar padrões bem definidas para cada um dos grupos de atraso. Por exemplo, o grupo  $A_3$  apresenta invariavelmente as maiores taxas de atraso ao longo do dia, enquanto o  $A_6$ , as menores; o grupo  $A_2$  aponta taxas maiores de atraso no pico diurno em relação ao pico noturno, acontecendo o contrário com  $A_4$ .

**Tabela 11: Distribuição dos dias da semana nos elementos do conjunto A**

Dia da semana	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
<b>Domingo</b>	32,62%	10,64%	4,26%	18,44%	13,83%	20,21%
<b>Segunda</b>	15,25%	22,70%	13,12%	12,77%	20,21%	15,96%
<b>Terça</b>	18,79%	20,21%	15,25%	16,31%	15,96%	13,48%
<b>Quarta</b>	19,15%	19,50%	16,67%	19,86%	12,77%	12,06%
<b>Quinta</b>	15,90%	25,80%	14,84%	10,95%	24,03%	8,48%
<b>Sexta</b>	27,92%	21,55%	15,55%	16,96%	9,89%	8,13%
<b>Sábado</b>	7,77%	23,32%	4,95%	5,65%	33,92%	24,38%
<b>Total Geral</b>	<b>19,63%</b>	<b>20,54%</b>	<b>12,09%</b>	<b>14,42%</b>	<b>18,66%</b>	<b>14,67%</b>

Diferentemente dos casos anteriores, não há padrões facilmente identificáveis de distribuições de dias da semana temporais do conjunto  $A$ . Porém, algumas observações ainda podem ser feitas. Na Tabela 11, observa-se pouca participação dos finais de semana no grupo de maior atraso,  $A_3$ ; em contrapartida, esses dias apresentam participação considerável em  $A_6$ , conjunto com as menores taxas de atraso. Os sábados têm participação significativa em  $A_2$  e  $A_5$ , que possuem picos maiores de atraso pela manhã, enquanto os domingos em  $A_1$  e  $A_4$ , com picos maiores pela noite. As segundas-feiras têm maior participação em perfis de atrasos diurnos ( $A_2$  e  $A_5$ ) enquanto as sextas-feiras aparentam maiores atrasos pela noite, com relevante participação no grupo  $A_1$ .

Resta analisar as divisões diárias de cada ano dentro dos elementos do conjunto dos grupos de atraso. A Tabela 12 traz essas informações.

**Tabela 12: Distribuição do número de dias de cada ano nos elementos do conjunto A**

Ano	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
<b>2009</b>	20,55%	24,11%	3,84%	8,22%	22,47%	20,82%
<b>2010</b>	21,64%	16,16%	9,86%	17,26%	16,44%	18,63%
<b>2011</b>	28,22%	7,40%	11,23%	21,64%	11,78%	19,73%
<b>2012</b>	19,40%	24,32%	14,48%	14,48%	17,21%	10,11%
<b>2013</b>	11,78%	30,96%	15,07%	10,96%	23,56%	7,67%
<b>2014</b>	11,26%	19,87%	26,49%	13,25%	23,18%	5,96%

É possível observar, por meio da Tabela 12, a participação estritamente crescente do número de dias de cada ano no grupo  $A_2$ , de maior atraso, ao passo que há uma diminuição gradual de participação em  $A_6$ , de menor atraso, com o passar do tempo. Outros padrões não são facilmente identificáveis como nos casos das divisões nos grupos de movimentações previstas e realizadas.

Por fim, para que se identifiquem possíveis padrões de atrasos, é importante que se estabeleça relações entre os elementos do conjunto  $A$  em relação aos de  $P$  e  $R$ . Tais relações são explicitadas na Tabela 13 e na Tabela 14.

**Tabela 13: Distribuição da quantidade de elementos dos grupos de movimentações previstas nos grupos de atrasos**

Grupo	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
$P_1$	20,55%	22,56%	5,51%	10,53%	22,31%	18,55%
$P_2$	7,80%	23,40%	4,96%	5,67%	33,33%	24,82%
$P_3$	32,74%	10,68%	4,27%	18,51%	13,88%	19,93%
$P_4$	25,18%	12,41%	16,67%	24,82%	12,77%	8,16%
$P_5$	20,87%	22,57%	17,96%	16,50%	10,19%	11,89%
$P_6$	10,90%	28,66%	21,81%	11,53%	21,50%	5,61%

**Tabela 14: Distribuição da quantidade de elementos dos grupos de movimentações realizadas nos grupos de atrasos**

Grupo	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
$R_1$	30,67%	10,67%	6,67%	17,33%	13,33%	21,33%
$R_2$	21,00%	25,00%	5,67%	10,67%	21,00%	16,67%
$R_3$	7,93%	22,76%	3,79%	5,17%	33,45%	26,90%
$R_4$	22,16%	12,47%	19,11%	20,22%	15,79%	10,25%
$R_5$	11,04%	29,55%	21,75%	11,04%	22,08%	4,55%
$R_6$	22,97%	23,21%	13,16%	18,90%	10,53%	11,24%

Não parece existir relações notáveis entre os conjuntos de movimentações previstas e o conjunto de movimentações realizadas com o conjunto de atrasos, de modo que não é possível inferir com razoável probabilidade de certeza que o perfil de movimentação prevista, ou mesmo realizada, determine um perfil de atraso esperado. É útil, portanto, que se busquem outros modos de se explicar quais os determinantes dos fatores de atraso.

Naturalmente, a capacidade de um aeroporto aparece como um dos mais importantes, senão o mais importante, fator de atraso em voos. Por isso, o espaçamento médio entre duas movimentações consecutivas é um forte candidato na explicação de maiores picos de atraso: quanto menor o intervalo de tempo entre dois pousos ou decolagens consecutivas, mais saturado está o aeroporto e mais atrasos são esperados. Para cada um dos grupos de atrasos foi calculado o espaçamento médio entre duas movimentações consecutivas tanto para o pico matinal (entre 9h e 12h) quando para o

pico noturno (entre 19h e 22h) de solicitação da pista. Os valores encontrados estão indicados na Tabela 15 e na Tabela 16. Em cada uma das tabelas, os grupos foram posicionados do de maiores índices de atraso para o horário, segundo a Figura 10, para os de menores índices, em ordem decrescente.

**Tabela 15: Espaçamento médio para o pico matinal para cada um dos elementos do conjunto A**

Grupo	Espaçamento médio entre 9h e 12h (em min)
$A_3$	1,701
$A_2$	1,766
$A_4$	1,767
$A_5$	1,809
$A_1$	1,820
$A_6$	1,848

**Tabela 16: Espaçamento médio para o pico noturno para cada um dos elementos do conjunto A**

Grupo	Espaçamento médio entre 19h e 22h (em min)
$A_3$	1.864
$A_4$	1.888
$A_2$	1.975
$A_1$	1.907
$A_5$	2.044
$A_6$	2.036

Os resultados apontados na Tabela 15 vão ao encontro do esperado: quanto maior o índice de atraso apresentado em um grupo do conjunto A, menor é o intervalo de tempo entre duas movimentações consecutivas no horário indicado. Tal dado levanta a hipótese de saturação de pista em determinados horários, o que acarreta inevitável atraso nas decolagens de voos.

A Tabela 16 apresenta algumas divergências quanto à hipótese inicial. Embora os dois grupos mais atrasados apontem, de fato, menores valores para o espaçamento médio no período noturno, os demais valores apresentam algumas inversões nessa ordem. Isso aponta que o intervalo de tempo médio entre duas movimentações

consecutivas, ainda que seja um fator relevante na determinação de atrasos, não aparece como única causa do fenômeno e, por si só, não é suficiente para explicar o fenômeno.

Outros fatores como determinantes de atraso ainda podem ser analisados de acordo com análises multivariáveis, porém tal análise, mais sofisticada, fugiria do escopo do presente trabalho.

## 5. CONCLUSÃO

A proposta deste trabalho consistiu na identificação de perfis de atrasos em voo no Aeroporto Internacional de Brasília. Para cumprir esse objetivo, utilizou-se técnicas hierárquicas de agrupamento de dados para a identificação de padrões nos perfis diários de movimentações na pista do aeroporto, tanto previstas como efetivamente realizadas, bem como para os índices horários de atraso, todos no período que vai de janeiro de 2009 a maio 2014.

Foram encontrados perfis horários de atrasos bem distintos entre os dias considerados, alguns com elevadas taxas de atraso na hora pico diurna e noturna, outros com índices relevantes apenas em um desses períodos, enquanto alguns apresentavam baixos índices ao longo de todo o dia.

As causas que determinam a disposição de cada dia dentro de um algum dos perfis de atraso não foram possíveis de serem determinadas com exatidão. Os índices de atraso não aparentam depender apenas dos perfis de movimentação previstas ou mesmo realizadas.

A hipótese de atrasos na hora pico estarem diretamente associadas ao espaçamento médio entre duas movimentações consecutivas no período apontou alguns resultados razoáveis, porém insuficientes para explicar o fenômeno. Uma análise mais profunda da relação entre essas variáveis ainda é necessária.

Há outras circunstâncias cujas análises são fundamentais para a avaliação proposta, como índice HHI, frequência de voos internacionais, fenômeno de HUB, condições climáticas, movimentações em períodos festivos ou de férias, dentre outros. Esse tipo de análise, com complexidade multivariável, é proposta para estudos futuros sobre o tema.

## BIBLIOGRAFIA

[1] **United States Department of Transportation.** Airline On-Time Statistics and Delay Causes, On-Time Arrival Performance National (Jan – May 2015). Disponível em:

<[http://www.transtats.bts.gov/OT\\_Delay/ot\\_delaycause1.asp?pn=1&periodfrom=24181%20&periodto=24185](http://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?pn=1&periodfrom=24181%20&periodto=24185)> 12 Ago 2015a.

[2] DIANA, T. Validating delay constructs: An application of confirmatory factor analysis. **Journal of Air Transport Management**, 35, 87-91, 2014.

[3] BALL, M., BARNHART, C., DRESNER, M., HANSEN, M., NEELS, K., ODONI, A., PETERSON, E., SHERRY, L., TRANI, A., ZOU, B., 2010. Total Delay Impact Study.

[4] ABDEL-ATY, M.; LEE, C; BAI, Y.; Li, X.; MICHALAK, M. Detecting periodic patterns of arrival delay. **Journal of Air Transport Management**, 13, 355-361, 2007.

[5] SANTOS, G.; ROBIN, M. Determinants of delays at European airports. **Transportation Research Part B**, 44, 392-403, 2010.

[6] DIANA, T. Do market-concentrated airports propagate more delays than less concentrated ones? A case study of selected U.S. airports. **Journal of Air Transport Management**, 15, 280-286, 2009.

[7] BRITTO, R.; DRESNER, M.; VOLTES, A. The impact of flight delays on passenger demand and social welfare. **Transportation Research Part E**, 48, 460-469, 2012.

[8] ZOU, B.; HANSEN, M. Flight delay impact on airfare and flight frequency: A comprehensive assessment. **Transportation Research Part E**, 69, 54-74, 2014.

- [9] FERRER, J. C.; R. e OLIVEIRA, P.; PARASURAMAN, A. The behavioral consequences of repeated flight delays. **Journal of Air Transport Management**, 20, 35-38, 2012.
- [10] **Anuário do Transporte Aéreo 2013**, volume único, 1. ed., Agência Nacional de Aviação Civil.
- [11] **Anuário do Transporte Aéreo 2011**, volume único, 1. ed., Agência Nacional de Aviação Civil.
- [12] WEBB, A. T. **Statistical Pattern Recognition**. 2. ed. New York: John Wiley & Sons, Inc, 2002.
- [13] DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2.ed. New York: Wiley-Interscience, 2000.
- [14] SHARMA, S. **Applied Multivariate Techniques**. New York: John Wiley & Sons, Inc, 1996.
- [15] **Agência Nacional de Aviação Civil**. Histórico de Voos. Disponível em: <<http://www2.anac.gov.br/vra/>> 15 Ago 2015a.

**FOLHA DE REGISTRO DO DOCUMENTO**

1. CLASSIFICAÇÃO/TIPO TC	2. DATA 19 de novembro de 2015	3. REGISTRO N° DCTA/ITA/TC-047/2015	4. N° DE PÁGINAS 44
5. TÍTULO E SUBTÍTULO: Identificação de perfis de atrasos de voos no aeroporto internacional de Brasília			
6. AUTOR(ES): <b>João Marcos Sousa Miranda</b>			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica – ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Aeroporto, Agrupamento, Atraso.			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Controle de tráfego aéreo; Atraso; Análise de dados; Aeroportos; Operações de linhas aéreas; Transporte.			
10. APRESENTAÇÃO: <p align="center"><b>X Nacional      Internacional</b></p> ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Rodrigo Arnaldo Scarpel; coorientador: Alessandro Vinícius Marques de Oliveira. Publicado em 2015.			
11. RESUMO: <p>Este trabalho tem como objetivo principal utilizar métodos de formação de agrupamentos em análise de dados para a identificação de perfis de atrasos em voos no Aeroporto Internacional de Brasília. Para isso, foram avaliadas todas as movimentações, tanto previstas como efetivamente realizadas, de voos no aeroporto durante o período que vai de janeiro de 2010 a maio de 2014. Todos os voos com diferença superior a 15 minutos entre o momento real e o previsto de sua movimentação na pista do aeroporto foram considerados como atrasados. Foram criados, através de técnicas hierárquicas de agrupamentos, grupos de dias com perfis semelhantes de movimentações previstas, de movimentações realizadas e de índices horários de atrasos, de modo a identificar possíveis padrões temporais que caracterizassem a formação desses agrupamentos. Enquanto tais padrões parecem emergir nos primeiros grupos citados, não há evidência clara que haja quaisquer padrões na divisão dos dias de acordo com seus perfis de atraso.</p>			
12. GRAU DE SIGILO: <p align="center"><b>(X) OSTENSIVO      ( ) RESERVADO      ( ) SECRETO</b></p>			