

# MATRIZ ORIGEM DESTINO

---

**RELATÓRIO DE METODOLOGIA E RESULTADOS DA MATRIZ ORIGEM DESTINO DOS  
REAIS DESEJOS DE VIAGENS CONTEMPLADOS OU NÃO PELO TRANSPORTE AÉREO**





## **APOIO TÉCNICO NO PLANEJAMENTO DO SETOR DE AVIAÇÃO CIVIL**

### **FASE 3 – DESENVOLVIMENTO DE UMA MATRIZ ORIGEM/DESTINO DE PASSAGEIROS EM ÂMBITO NACIONAL PARA IDENTIFICAÇÃO DOS REAIS DESEJOS DE VIAGEM CONTEMPLADOS OU NÃO PELO TRANSPORTE AÉREO**

RELATÓRIO DE METODOLOGIA E RESULTADOS DA MATRIZ  
ORIGEM/DESTINO DOS REAIS DESEJOS DE VIAGENS CONTEMPLADOS OU  
NÃO PELO TRANSPORTE AÉREO

### HISTÓRICO DE VERSÕES

Data	Versão	Descrição	Autor
28/02/2020	1.0	Entrega da primeira versão do <i>Relatório de Metodologia e resultados da matriz Origem/Destino (OD) dos reais desejos de viagens contemplados ou não pelo transporte aéreo</i> , em referência ao produto previsto no Plano de Trabalho <i>Relatório de análise dos resultados e avaliação dos principais determinantes dos fluxos existentes de passageiros entre aeroportos e análise da estimativa do potencial de migração de passageiros para o transporte aéreo, por aeroporto</i> .	LabTrans/UFSC
06/05/2020	1.1	Atualização da metodologia e resultados da expansão da matriz Origem/Destino para o modo 0 (Não aéreo)	LabTrans/UFSC



# SUMÁRIO

<b>1. Aspectos metodológicos .....</b>	<b>7</b>
1.1. Estruturação da base e coleta de dados .....	10
1.1.1. Área de estudo e zoneamento.....	10
1.1.2. Escopo temporal .....	13
1.1.3. Big Data .....	13
1.1.4. Fornecimento da Base de Dados .....	15
1.1.5. Método de coleta de dados de telefonia móvel .....	21
1.1.6. Rastreamento do usuário – Identificação e categorização de viagem .....	30
1.2. Etapas do estudo .....	39
1.3. Processo de validação dos dados fornecidos .....	40
1.3.1. Bases de dados de comparação .....	40
1.3.2. Regras de validação .....	42
1.3.3. Análises parciais.....	44
1.3.4. Análise do extrato completo.....	50
1.4. Expansão da matriz .....	53
1.4.1. Descrição das variáveis utilizadas .....	57
1.4.2. Análise dos Componentes Principais .....	61
1.4.3. Estimativa do modelo .....	65
<b>2. Resultados gerais.....</b>	<b>69</b>
2.1. Deslocamento da população brasileira entre as regiões .....	72
2.2. Deslocamento da população brasileira entre as UTPs .....	75
2.3. Potencial demanda reprimida .....	86
2.4. Áreas de influência .....	88
<b>Considerações finais .....</b>	<b>99</b>
<b>Referências .....</b>	<b>101</b>
<b>Lista de figuras .....</b>	<b>105</b>
<b>Lista de gráfico .....</b>	<b>107</b>

<b>Lista de tabelas.....</b>	<b>109</b>
<b>Lista de siglas .....</b>	<b>111</b>
<b>Apêndices e anexos .....</b>	<b>113</b>
Apêndice 1– Usuários descartados por UF .....	115
Apêndice 2 – Os 50 principais pares de deslocamento para o Modo 0 e para o Modo 1.....	119

# INTRODUÇÃO

A Secretaria Nacional de Aviação Civil do Ministério da Infraestrutura (SAC/MInfra) está desenvolvendo, em cooperação com o Laboratório de Transportes e Logística da Universidade Federal de Santa Catarina (LabTrans/UFSC), estudos e pesquisas para apoio técnico no planejamento do setor de aviação civil brasileiro.

Em atendimento ao escopo definido para a Fase 3 do *Plano de Trabalho* apresentado pelo LabTrans/UFSC à SAC/MInfra, relativo à elaboração de uma matriz Origem/Destino (OD) de passageiros em âmbito nacional para identificação dos reais desejos de viagens contemplados ou não pelo transporte aéreo, este relatório visa apresentar:

- aspectos metodológicos de elaboração da matriz;
- análises dos resultados da matriz;
- avaliação dos principais determinantes dos fluxos existentes de passageiros entre aeroportos;
- e análise da estimativa do potencial de migração dos deslocamentos terrestres para o transporte aéreo, por aeroporto.

As análises foram obtidas a partir da matriz OD fornecida pela Seleção Pública nº 105/2018<sup>1</sup>, gerada através de dados Big Data de registros de telefonia móvel para análises de mobilidade da população.

No processo de validação da matriz pelo LabTrans/UFSC, foram utilizadas bases de dados de deslocamentos do modo de transporte aéreo – a pesquisa da Empresa de Planejamento e Logística S.A. (EPL, 2014), a Base de Dados Estatísticos da Agência Nacional de Aviação Civil (ANAC, 2019a) e seus Microdados de Tarifas Aéreas Comercializadas (ANAC, c2020) – e do modo terrestre – os dados operacionais da Agência Nacional de Transportes Terrestres (ANTT, [201-]).

As análises de validação permitiram auxiliar a Telefônica no ajuste dos parâmetros utilizados na extração de dados com vistas a reduzir limitações inerentes à tecnologia. Nesse processo, ao todo, houve oito versões da extração parcial (referentes ao mês de maio de 2017) que foram analisadas com o intuito de reduzir potenciais erros de classificação do modo de viagem, que geram falsos positivos, ou de alocação aos aeródromos de embarque e de desembarque. Ao final, esses tipos de erro representavam menos de 5% de toda a amostra.

---

<sup>1</sup> TELEFÔNICA BRASIL S.A., CNPJ 02.558.157/0001-62, contratada por meio da Seleção Pública nº 105/2018, Processo nº 331271729, da Fundação de Amparo à Pesquisa e Extensão Universitária (FAPEU).

A comparação da matriz com essas bases de dados permitiu, ainda, realizar análises a fim de identificar os fluxos (pares OD) não contemplados pela pesquisa da contratada, fazendo-se necessária uma expansão para complementação da matriz com os deslocamentos faltantes.

A expansão ocorreu através de modelagem econométrica, utilizando-se de dados socioeconômicos e dos próprios resultados auferidos na base da contratada. O resultado final foi a matriz OD dos reais desejos de viagens contemplados ou não pelo transporte aéreo.

Assim, o conteúdo do presente documento está dividido nos seguintes capítulos:

- Introdução.
- Aspectos metodológicos – apresenta as premissas e aspectos considerados em relação à área de estudo, obtenção dos dados de telefonia, tratamentos realizados e metodologia adotada para a expansão da matriz OD para os pares não contemplados.
- Resultados gerais – contempla a análise da matriz OD de passageiros do Brasil e suas especificidades por região e Unidades Territoriais de Planejamento (UTPs); e avaliação das áreas de influência dos aeroportos das capitais.
- Análises específicas – mostra os resultados apresentados após a expansão da matriz OD, como as potenciais rotas ainda não exploradas e deslocamentos que não possuem rota direta.
- Considerações finais.

# 1. ASPECTOS METODOLÓGICOS

Este capítulo apresenta os aspectos metodológicos da elaboração da matriz OD de passageiros em âmbito nacional, com a identificação dos municípios de reais origens e destinos dos usuários do transporte aéreo.

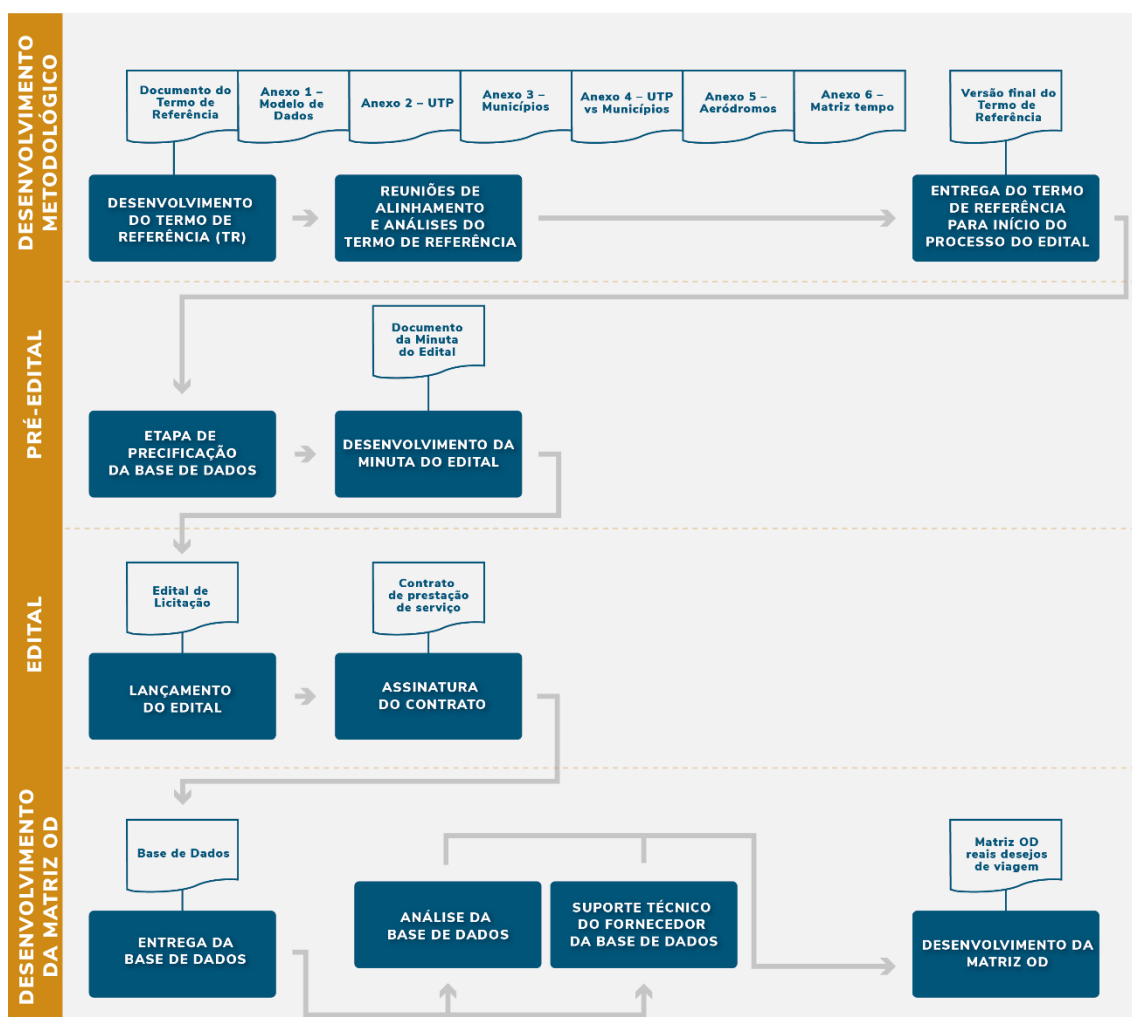
A matriz OD foi obtida a partir da base de dados fornecida pela empresa Telefônica, extraída através da plataforma Luca, que contempla as viagens realizadas entre as 780 UTPs definidas pela SAC/MInfra, através do modo aéreo ou não aéreo. A base foi desenvolvida a partir do fornecimento de dados agregados em volumes de deslocamentos com base nos registros dos usuários de telefonia móvel na rede de Estações Rádio Base (ERBs), popularmente conhecidas como antenas. Utilizando essas informações, foi possível identificar os deslocamentos e classificar o modo de transporte utilizado em aéreo ou não aéreo (ou terrestre), através de alguns critérios considerados. Quando classificado em aéreo, foram identificados, ainda, os municípios de origem e de destino dos passageiros, bem como os aeródromos de embarque e desembarque.

A seleção da empresa como fornecedora dos dados de telefonia a serem utilizados no estudo ocorreu através de seleção pública. Os critérios adotados estão descritos no *Termo de Referência* (TR), referente ao Anexo I da Seleção Pública nº 105/2018 (FAPEU, 2018). Salienta-se que os parâmetros para identificação da real origem e do real destino do usuário, e para classificação do modo de transporte em aéreo ou terrestre, foram apresentados também no *Relatório de Metodologia - Estimativa dos fluxos de passageiros do transporte aéreo*<sup>2</sup>. No entanto, a partir da primeira versão da extração parcial de dados pela Telefônica e das análises de validação realizadas pela equipe técnica do LabTrans/UFSC, verificou-se a necessidade de ajustes nos parâmetros utilizados, o que resultou em oito extratos parciais até se chegar ao máximo de refinamento, considerando-se assim, a base como válida para a extração completa e expansão dos dados a fim de contemplar deslocamentos faltantes não identificados pela contratada. As etapas realizadas para a elaboração da matriz OD são apresentadas na Figura 1.

---

<sup>2</sup> Produto entregue à SAC/MInfra em 14 de dezembro de 2018.

Figura 1 – Fluxograma das macroetapas de coleta de dados e elaboração da matriz OD

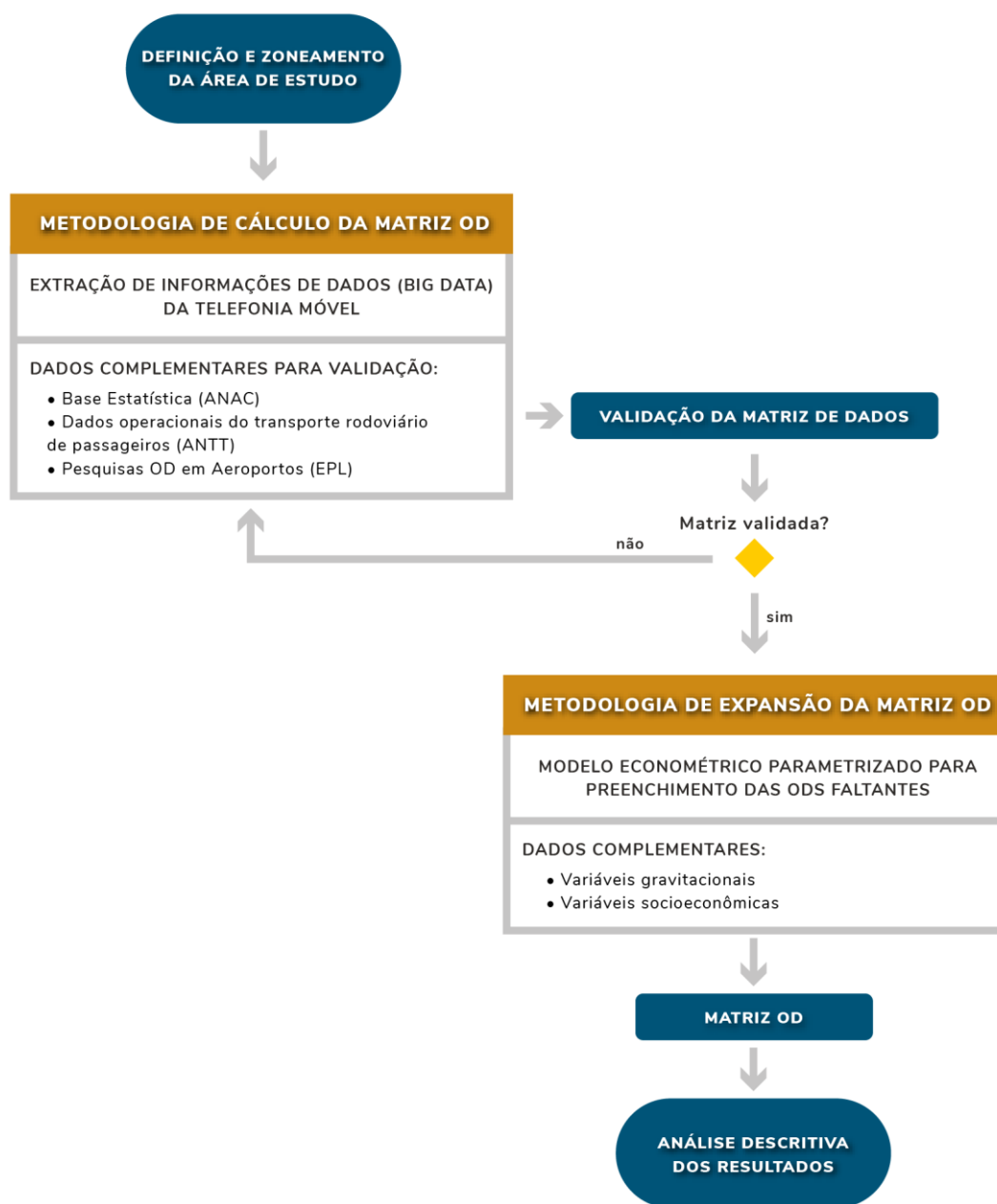


Elaboração: LabTrans/UFSC (2020)

A partir da base expandida foi possível realizar as análises referentes aos determinantes dos fluxos existentes de passageiros entre os aeroportos, identificação de possíveis rotas que atualmente não são contempladas pelo setor e o potencial de migração entre o modal terrestre e o modal aéreo.

A Figura 2 apresenta o fluxograma da metodologia utilizada para a elaboração da matriz OD da área de estudo utilizando-se os dados de registros de usuários de telefonia móvel para estimativa dos fluxos dos passageiros.

Figura 2 – Fluxograma da metodologia



Elaboração: LabTrans/UFSC (2020)

As sessões seguintes apresentam detalhadamente os aspectos metodológicos e as etapas decorridas na formulação da matriz OD.

## 1.1. Estruturação da base e coleta de dados

Neste item são apresentadas as definições iniciais do estudo, incluindo área de estudo e seu zoneamento, o período analisado, o conceito de Big Data e sua utilização em planejamento de transportes e logística, bem como as características da empresa contratada relacionadas à sua cobertura (abrangência geográfica) e participação de mercado. Além disso, são exibidos os métodos utilizados na coleta de dados, os critérios adotados na validação e classificação dos dados.

### 1.1.1. Área de estudo e zoneamento

Uma vez que o presente estudo consiste na coleta de dados da demanda atual por transporte aéreo, que servirá como base para um planejamento de nível estratégico do setor aéreo nacional, a área de estudo compreende o território nacional.

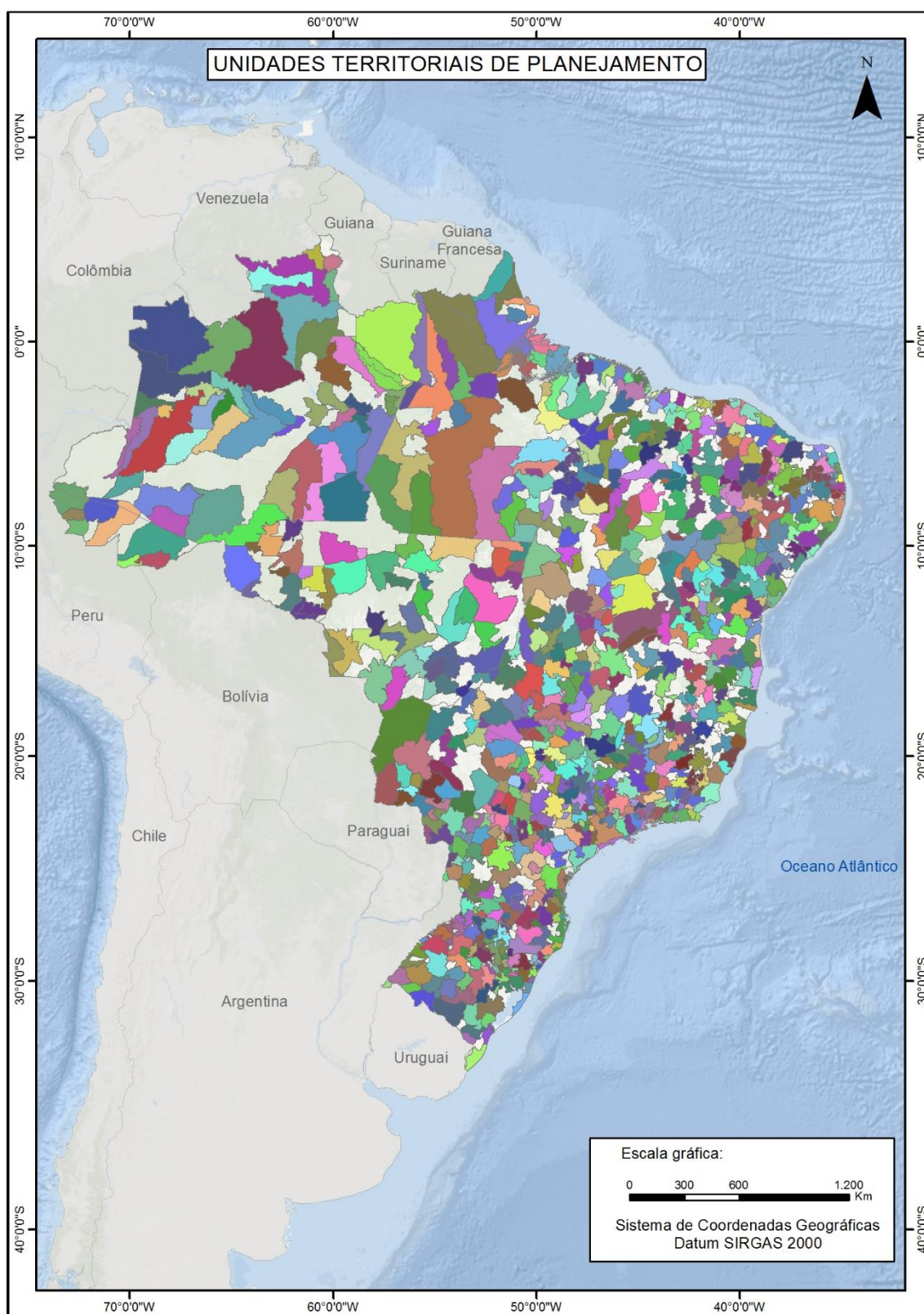
No que se refere ao zoneamento dessa área, na metodologia utilizada para as projeções de demanda para os aeroportos brasileiros (2017-2037), realizada pelo Ministério dos Transportes, Portos e Aviação Civil (MTPA) (BRASIL, 2017a, p. 16), atual Ministério da Infraestrutura (MInfra), foram definidas as UTPs diante da necessidade de:

[...] predefinir regiões onde se concentram a maior parte da demanda potencial e real pelo transporte aéreo, de modo que suas variáveis possam ser estudadas como explicativas no procedimento de modelagem, ou mesmo delimitar regiões onde já existem infraestruturas aeroportuárias instaladas, operantes ou não, que podem ser alvo de planejamento futuro e que representam regiões de concentração da população e, assim, do maior potencial de demanda pelo transporte aéreo. (BRASIL, 2017a, p.16).

As UTPs foram definidas não só para as projeções de demanda, mas também para o Plano Aeroviário Nacional (PAN) (BRASIL, 2018b). De acordo com o *Plano de Trabalho* do projeto de cooperação entre a SAC/MTPA, atual SAC/MInfra, e o LabTrans/UFSC, elas devem ser utilizadas também como as unidades territoriais de referência para a elaboração da matriz OD. As UTPs, de acordo com o Anexo 2 da Seleção Pública nº 105/2018, Processo nº 331271729, da FAPEU (2018), são apresentadas na Figura 3.



Figura 3 – UTPs



Fonte: FAPEU (2018). Elaboração: LabTrans/UFSC (2020)

A definição das UTPs teve o objetivo de representar regiões de concentração da população e, assim, do maior potencial de demanda pelo transporte aéreo. De acordo com a então SAC/MTPA (2017a), os resultados da pesquisa origem/destino realizada pela SAC com a EPL permitiu identificar que cerca de 84% da demanda dos aeroportos faz parte da população das aglomerações urbanas imediatamente próximas a eles.

Assim, a partir da hipótese de que a maior parte da demanda pelo transporte aéreo está presente nas aglomerações urbanas em que há infraestrutura aeroportuária instalada, a definição das UTPs partiu das informações do estudo do Instituto Brasileiro de Geografia e Estatística (IBGE) intitulado *Arranjos populacionais e concentrações urbanas* (IBGE, 2016).

Somadas a esse primeiro recorte, a partir dos arranjos populacionais e concentrações urbanas, foram criadas novas UTPs, a fim de contemplar:

- Municípios com aeródromos instalados segundo o cadastro de aeródromos públicos, sendo cerca de 640 à época em que foi consolidado o banco de dados (ANAC, [2016]), e aeródromos constantes nos Planos Aeroviários Estaduais (PAEs).
- Municípios indutores do turismo de classes A ou B, segundo o Programa de Regionalização do Turismo (BRASIL, 2017b).

Além disso, foram realizados os seguintes ajustes:

- Adequação de UTPs da Região Norte, de modo a considerar o transporte hidroviário, conforme base de dados cedida pela Agência Nacional de Transportes Aquaviários (ANTAQ) de 2013 referentes a uma pesquisa realizada sobre as ligações desse modo de transporte, a qual possibilitou a identificação de ligações pendulares.
- Ampliação da abrangência territorial de determinadas UTPs considerando-se na definição da área de captação dos aeródromos a presença de municípios em isócronas de 30, 60, 90 e 120 minutos de deslocamento rodoviário aos aeródromos.
- Ajustes para eliminar citações duplicadas de municípios em diferentes UTPs (sobreposições), tendo prevalência a influência de municípios-sede das UTPs cuja categoria de centralidade nas Regiões de Influência das Cidades (REGIC) é maior.
- Ajustes realizados em ambiente de Sistema de Informações Geográficas - SIG para eliminar as inconsistências nas formas das UTPs, a saber: municípios isolados de suas sedes e UTPs muito próximas, totalmente ou parcialmente inscritas em outras UTPs.

Desse processamento resultaram 780 UTPs, apresentadas nos anexos da Seleção Pública nº 105/2018, Processo nº 331271729, da FAPEU (2018). Elas serão objeto do estudo de deslocamento dos indivíduos devido a sua relevância em termos de aglomerações urbanas existentes no País, de aeródromos instalados, indução turística e participação do modo hidroviário no transporte de pessoas na Região Norte do País.

### 1.1.2. Escopo temporal

Em 2018, ano precedente à realização do presente estudo, houve a greve dos caminhoneiros que se iniciou no dia 21 de maio e teve uma duração de dez dias (UOL, 2018). Esse fato pode ter causado impactos significativos nos padrões de comportamento da população e também no setor aéreo, devido ao fato de que muitos aeroportos ficaram sem combustível para o abastecimento das aeronaves, prejudicando a normalidade dos deslocamentos. Desse modo, o ano de 2018 não foi considerado como representativo (BRASIL, 2018a).

Dessa forma, os dados coletados têm como período de referência o ano de 2017 integralmente, considerado o ano mais recente representativo.

### 1.1.3. Big Data

A literatura indica que um elemento fundamental no planejamento de transportes aéreo é a elaboração de uma matriz OD mais próxima do real, contemplando também os desejos de viagens. Para essa estimativa existem diversos métodos usualmente utilizados e um deles é o uso de Big Data de telefonia móvel. Nos itens a seguir são apresentados o conceito de Big Data, exemplos de seu uso para o planejamento de transportes e os aspectos de privacidade dos dados e de transparência de sua utilização.

#### 1.1.3.1. Conceito de Big Data

Para a criação de uma base de dados fez-se uso de material de Big Data<sup>3</sup> contendo as informações de telefonia móvel. De acordo com Chen e Zhang (2014), Big Data é uma coleção de volumosos conjuntos de dados com uma diversidade de tipos tão grande que torna difícil seu processamento, utilizando-se abordagens do estado da arte ou plataformas tradicionais de processamento de dados. O Big Data derivou do conceito de gerenciamento de dados 3-D formulado por Laney (2001), que consiste na exploração de dados nas dimensões volume, velocidade e variedade (3 Vs). Além disso, segundo Demchenko, Laat e Membrey (2014), existem as dimensões valor e veracidade (surgindo assim a definição de propriedades dos 5 Vs). O Big Data pode ser visto como um imenso volume de dados que apresentam veracidade quanto a sua origem e variedade quanto a sua estrutura e que surgem em torno dos negócios

---

<sup>3</sup> Mais informações do material no documento de Seleção Pública nº 105/2018, Processo nº 331271729, da FAPEU (2018).

diariamente, permitindo-se agregar valor (*insights* produzidos na análise) quando houver velocidade para obtenção dos resultados.

O Big Data fornece um melhor conhecimento de seus usuários e também a compreensão de seu comportamento e de suas preferências. Nos sistemas de transporte, os usuários são os indivíduos que realizam deslocamentos através dos modos de transporte de interesse. Informações de demanda provenientes de Big Data podem oferecer maior assertividade das ações estratégicas no planejamento da infraestrutura de transportes, otimizando investimentos. Para tanto, os dados devem ser tratados e analisados criteriosamente, conhecendo-se bem as questões a que devem responder.

Dados provenientes de Big Data são considerados dados de alto valor agregado aos negócios e organizações, uma vez que, através de um processamento eficiente e de inteligência analítica, os planejadores podem tomar decisões baseadas em evidências concretas. Chen e Zhang (2014) citam aplicações de Big Data em três principais domínios: comércio e administração, administração pública e campos de pesquisa científica.

### **1.1.3.2. O uso do Big Data no planejamento de transportes**

No contexto da coleta de dados no planejamento de transportes com o Big Data, avaliou-se a existência das diversas fontes de dados passíveis de serem utilizadas para a elaboração de matrizes OD. Uma fonte de obtenção de informações relevantes do deslocamento de pessoas que vem sendo utilizada é a análise da pegada digital, por meio de registros dos usuários de telefonia móvel na rede de telecomunicações, de registros de GPS, entre outros dados gerados durante seu deslocamento com informações de espaço e tempo. No que se refere à telefonia móvel, segundo a Agência Nacional de Telecomunicações (ANATEL, 2019), no Brasil, para cada 100 habitantes, há 109,5 acessos em operação na telefonia móvel. De acordo com García *et al.* (2016), são vantagens da utilização de dados de telefonia móvel para entender o deslocamento dos usuários em comparação às pesquisas tradicionais:

- Possibilidade de amostras grandes.
- Os dados são coletados passivamente, eliminando a interação com os entrevistados, o que frequentemente resulta em respostas incorretas ou imprecisas.
- Economia de custo, visto que os dados já são registrados pelas operadoras de telefonia móvel.

### 1.1.3.3. Privacidade dos dados e transparência quanto à utilização

Uma grande preocupação a respeito do uso de Big Data é a privacidade dos usuários. Ainda que estes possam se beneficiar quando a empresa utiliza dados de Big Data com o objetivo de melhorar seus produtos, processos e serviços, é necessária precaução para garantir a privacidade dos usuários. Nesse sentido, as empresas que coletam e armazenam dados de interesse podem fornecê-los não de forma individual, contendo informações de identificação pessoal (PII, do inglês – *Personally identifiable information*), mas agregados e anonimizados. A anonimização refere-se a um procedimento que faz com que os registros não possam ser associados, de forma direta ou indireta, a um indivíduo.

O conceito de privacidade e anonimização dos dados está alinhado com a Lei Geral de Proteção de Dados Pessoais (LGPD) – Lei nº 13.709/2018 (BRASIL, 2018e), que dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

### 1.1.4. Fornecimento da Base de Dados

A Telefônica (Telefónica, conforme seu nome comercial ou razão social) é uma empresa espanhola que se configura como uma das principais empresas globais integradas de telecomunicações. Ela é conhecida por seus clientes e por suas marcas comerciais, sendo as principais:

- Movistar, em diversos países da América Latina
- O<sub>2</sub>, na Alemanha, na Espanha e no Reino Unido
- Vivo, no Brasil desde 2003.

A empresa é composta, também, por marcas especializadas em negócios específicos. A unidade de análise de dados é chamada LUCA e fornece soluções a empresas, utilizando Big Data e inteligência artificial (AI, do inglês – *artificial intelligence*), aplicando algoritmos aos dados da Telefônica para detectar padrões.

Para a coleta de dados de telefonia móvel do presente estudo foi utilizada a plataforma LUCA Transit, uma ferramenta estatística com o suporte de uma infraestrutura de Big Data, a qual utiliza as informações operacionais da rede de telefonia móvel de forma anonimizada e agregada para a extrapolação do fluxo de pessoas pelo território (TELEFÔNICA, 2018a).

No setor de transportes, a LUCA auxilia empresas e a administração pública a otimizar a infraestrutura e o planejamento do sistema de transportes provendo maior conhecimento a respeito dos usuários, dos horários e das rotas. São alguns dos estudos realizados no setor (LUCA, [201-]b):

- Um estudo de análise da demanda do transporte na cidade espanhola de Zaragoza, com a construção de matrizes e a determinação de tendências demográficas.
- Em parceria com o World Bank e a Universidade de São Paulo (USP), um estudo de padrões de mobilidade da população residente de Paraisópolis, utilizando dados de passes do transporte público, dados de telefonia móvel e dados de uma lista de provedores de serviços de telecomunicações.
- Construção de matrizes OD para projetar planos de mobilidade urbana sustentável dos municípios brasileiros de Guarujá, Jandira, São Luis e Votuporanga.

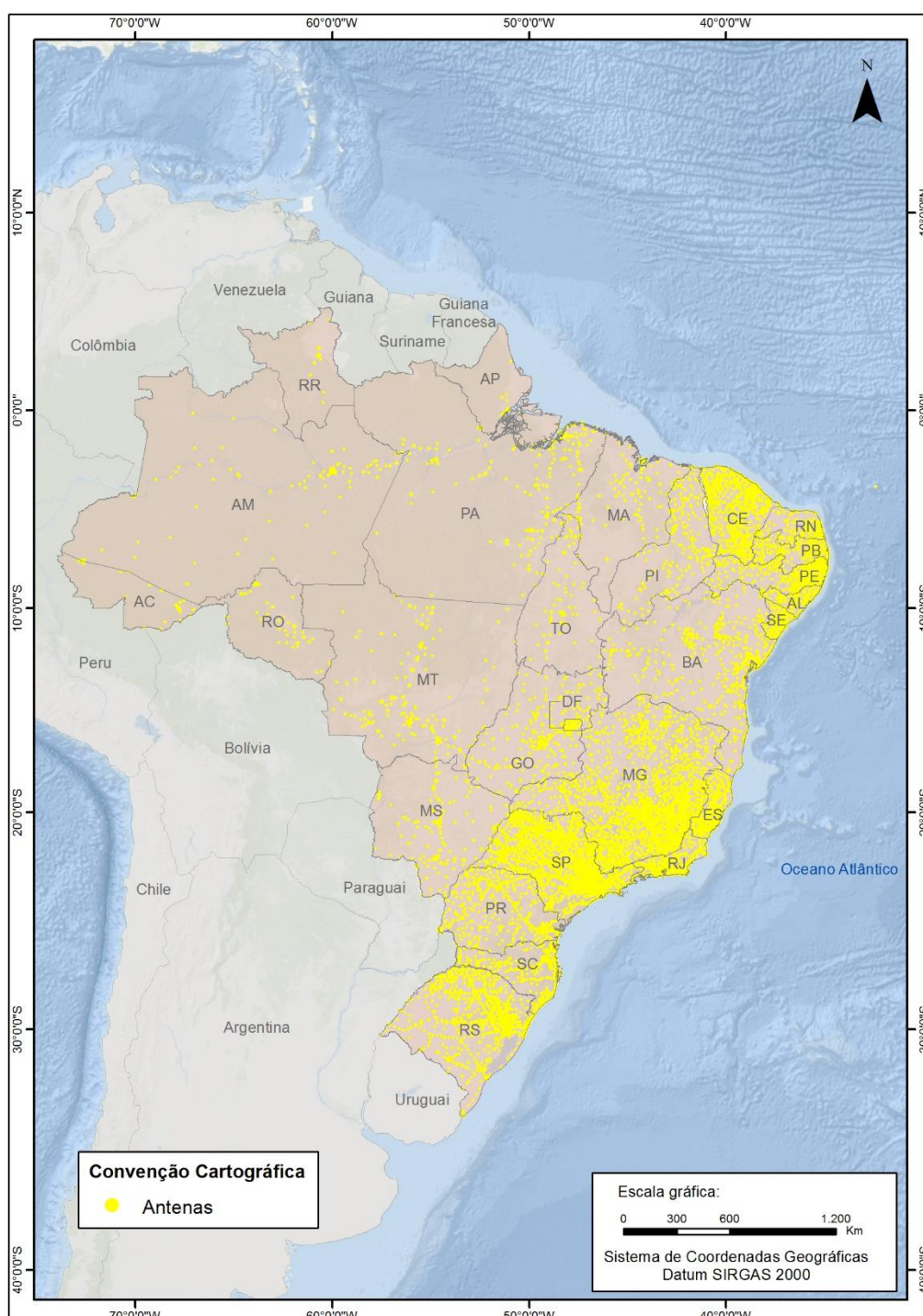
#### **1.1.4.1. Cobertura e participação de mercado da empresa contratada**

A amostra de dados provenientes da pesquisa referentes à cada UTP varia de acordo com aspectos técnicos, a saber: a cobertura de telefonia móvel e a participação de mercado da empresa.

No que se refere à cobertura, a empresa informou ter cobertura de telefonia móvel em 4.540 municípios, apresentados na Figura 4. Para o presente estudo foi considerando que um município possui cobertura móvel quando este possui uma torre de transmissão dentro de seu limite geográfico (TELEFÔNICA, 2018b). De acordo com a Anatel (2020), a Vivo (Telefônica no Brasil) tem cobertura do Serviço Móvel Pessoal (SMP) em 4.471 municípios, se considerada qualquer tecnologia (2G, 3G ou 4G); se 3G, há cobertura em 4.324 municípios. No ano de 2017, ela apresentava cobertura 3G em 3.996 municípios (ANATEL, 2020).



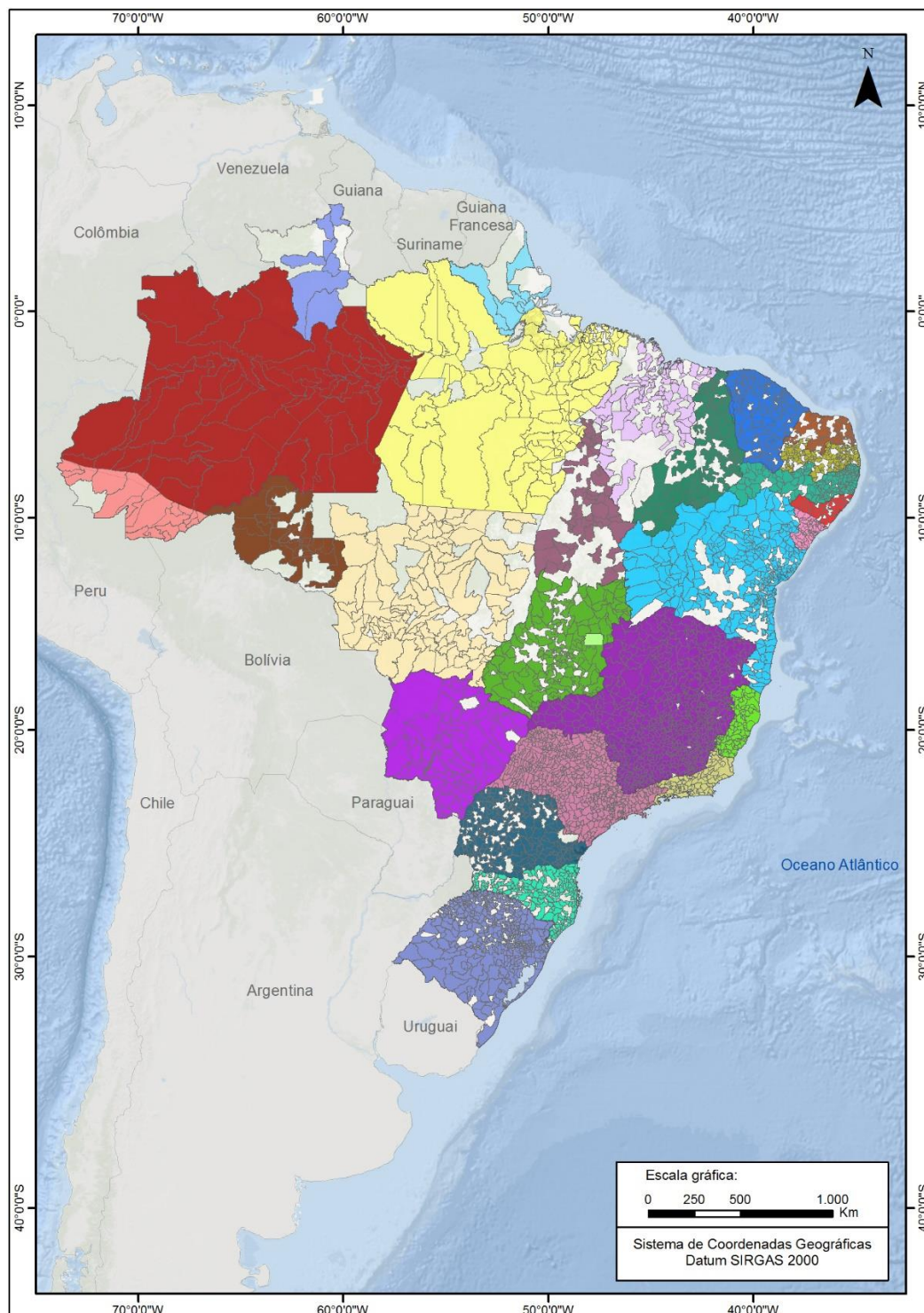
Figura 4 – Disposição geográfica das torres de telefonia móvel pesquisa – Brasil



Fonte: Anatel (2020). Elaboração: LabTrans/UFSC (2020)

18

## RELATÓRIO DE ANÁLISES





A Tabela 1, por sua vez, apresenta a participação de mercado por Unidade da Federação (UF). Pode-se observar que sua participação de mercado varia consideravelmente entre as UFs, inclusive entre aquelas pertencentes à mesma região geográfica.

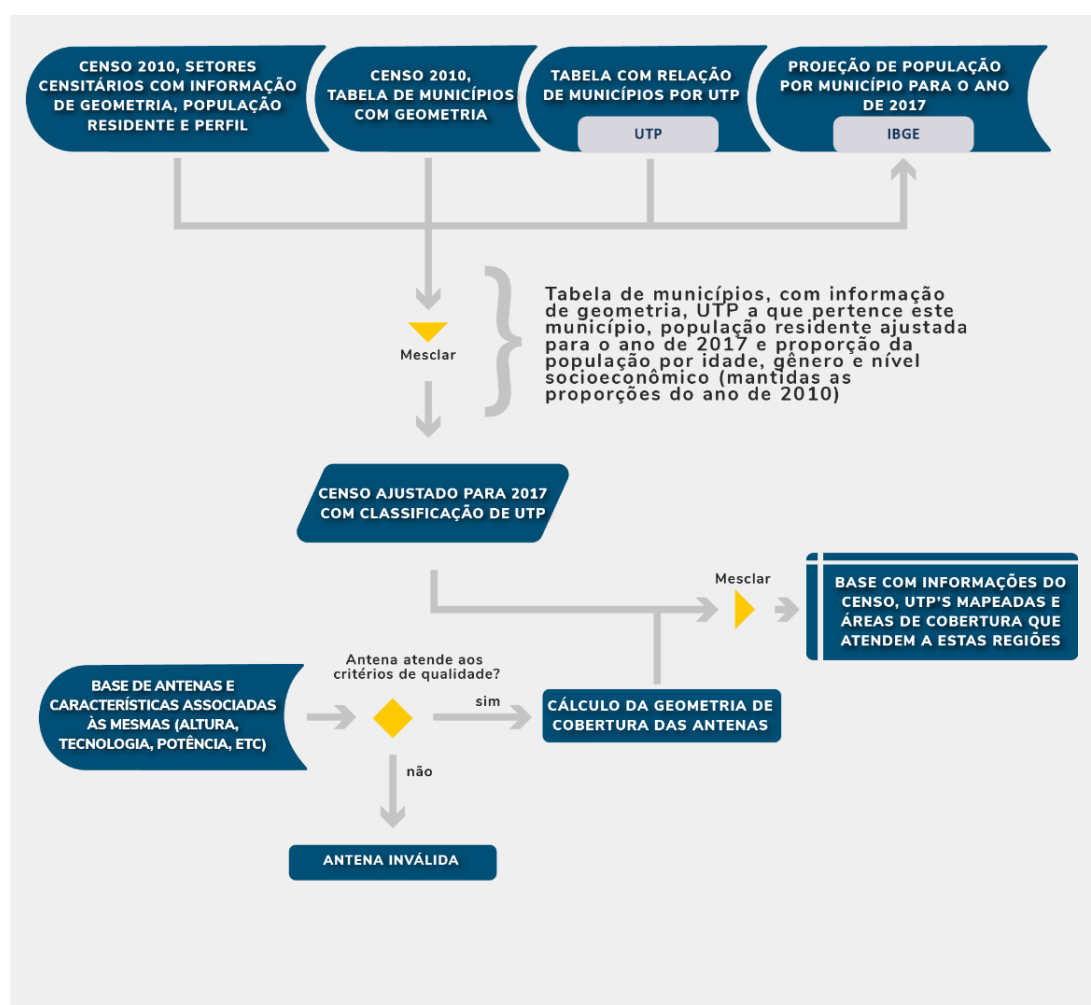
**Tabela 1 – Participação de mercado por UF (2019)**

Região	UF	Participação de mercado
Centro-Oeste	DF	31,2%
	GO	27,8%
	MS	48,8%
	MT	54,5%
Nordeste	AL	14,2%
	BA	30,6%
	CE	9,8%
	MA	20,4%
	PB	14,3%
	PE	11,3%
	PI	18,9%
	RN	7,1%
	SE	66,0%
Norte	AC	52,3%
	AM	52,9%
	AP	48,3%
	PA	37,8%
	RO	21,5%
	RR	59,0%
	TO	15,2%
Sudeste	ES	74,1%
	MG	42,5%
	RJ	28,5%
	SP	36,7%
Sul	PR	17,2%
	RS	47,3%
	SC	25,7%

Fonte: Teleco (2019). Elaboração: LabTrans/UFSC (2020)

Na Figura 6 é apresentado o diagrama de processo de formação da base de zoneamento da pesquisa, onde são agrupadas as informações das UTPs juntamente com outras variáveis, como população, perfil da população e dados de geometria. A partir disso é gerada uma base única de informações de cada UTP. Paralelamente, com base nas informações das antenas, são avaliadas quais atendem aos critérios de qualidade adotados, gerando-se uma base com informações da cobertura das antenas. Essas duas bases geradas são mescladas, formando a base de zoneamento para a pesquisa.

Figura 6 – Diagrama de processo: base de zoneamento



Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

Essa base é utilizada em etapas posteriores, na criação da Base de Estadias e no processo de extrapolação da quantidade de viagens em comparação ao nível censitário.

### 1.1.5. Método de coleta de dados de telefonia móvel

Conforme citado anteriormente, uma das preocupações em se utilizar Big Data refere-se à privacidade do usuário e à garantia da relevância estatística. Para isso, o processamento do uso dos serviços oferecidos pela empresa é feito através de uma tripla camada, que é composta por:

1. Anonimização: registros pessoais que são substituídos por identificadores encriptados na movimentação de cada dispositivo.
2. Agregação: dados sobre múltiplos dispositivos são agrupados no espaço e no tempo, fazendo com que se tenha a relevância estatística em quantidade de informação, e também para que não seja possível rastrear comportamentos individuais, ainda que anônimos.
3. Extrapolação: dados extrapolados para a base censitária do comportamento de toda a população.

Além disso, para o entendimento da origem e precisão dos dados é necessário o entendimento de alguns conceitos. O primeiro refere-se à rede de celular, que é uma rede sem fio, distribuída em áreas terrestres nomeadas de células, que possuem ao menos um transceptor de localização fixa, conhecido como um site celular ou ERB. Na rede celular cada célula utiliza um conjunto de frequências diferentes de células vizinhas para que sejam evitadas interferências e fornecida largura de banda garantida dentro de cada célula. Quando essas células estão juntas, elas fornecem cobertura de rádio em uma ampla área geográfica, fazendo com que seja possível uma grande quantidade de telefones celulares comunicarem-se entre si e com os transceptores fixos e telefones em qualquer lugar da rede, através das estações de base, mesmo que alguns telefones celulares estejam passando por mais de uma célula durante a transmissão (TELEFÔNICA, 2018a).

Quando um telefone móvel conecta-se a uma rede, o sinal é encaminhado por uma ERB. Uma ERB são os locais onde antenas e equipamentos eletrônicos são colocados, normalmente em lugares altos como torres e mastro, criando uma célula dentro de uma rede de celular. Cada célula é específica da tecnologia, que pode ser 2G, 3G e 4G, onde os eventos variam de acordo com a geografia de implantação. O alcance de uma célula depende do tipo de tecnologia, pois as células 2G tendem a ser capazes de cobrir áreas maiores que as células 3G e 4G. Para o contexto do estudo, o termo célula não se refere ao local onde os equipamentos telefônicos são colocados, mas à área coberta pelo equipamento. Cada célula possui características que estão disponíveis no catálogo de células (TELEFÔNICA, 2018a).

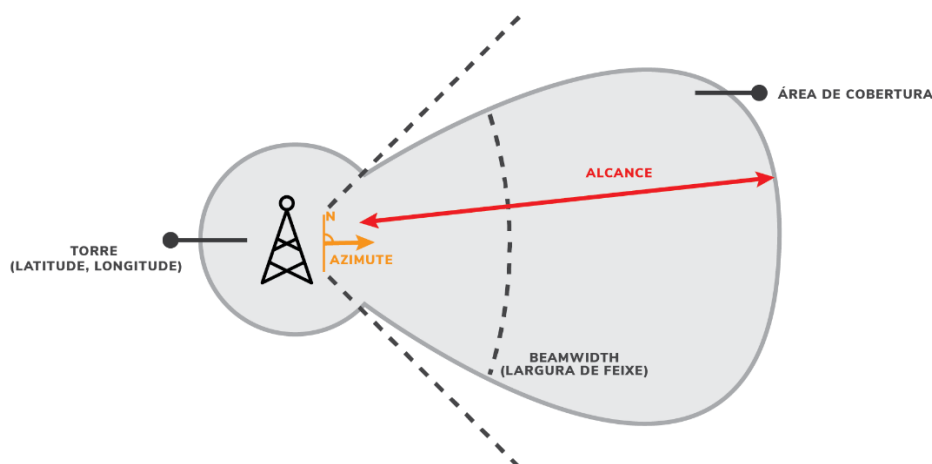
De acordo com a Telefônica (2018a), os telefones que estão conectados à rede móvel geram eventos de rede. A natureza desses eventos (tipo e frequência no tempo) depende de vários fatores, incluindo o comportamento do usuário, quando este faz uma chamada ou recebe um SMS (do inglês – Short Message Service), e também das características da rede móvel. Para a realização do estudo foram considerados como os eventos de rede ligar/receber uma ligação ou enviar/receber pacote de dados.

A seguir são apresentados os aspectos considerados para obter a origem e precisão dos dados, a seleção da amostra e tratamento dos dados e, por fim, a representação estatística da amostra e a extrapolação dos dados.

#### 1.1.5.1. Origem e precisão geográfica dos dados

Este tópico contém a descrição dos conceitos técnicos, que são parâmetros físicos para a cobertura das células de telefonia móvel e disponíveis no catálogo de células. A Figura 7 apresenta algumas dessas características de uma célula de telefonia móvel.

Figura 7 – Parâmetros de uma célula de telefone celular



Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

Onde, em suma:

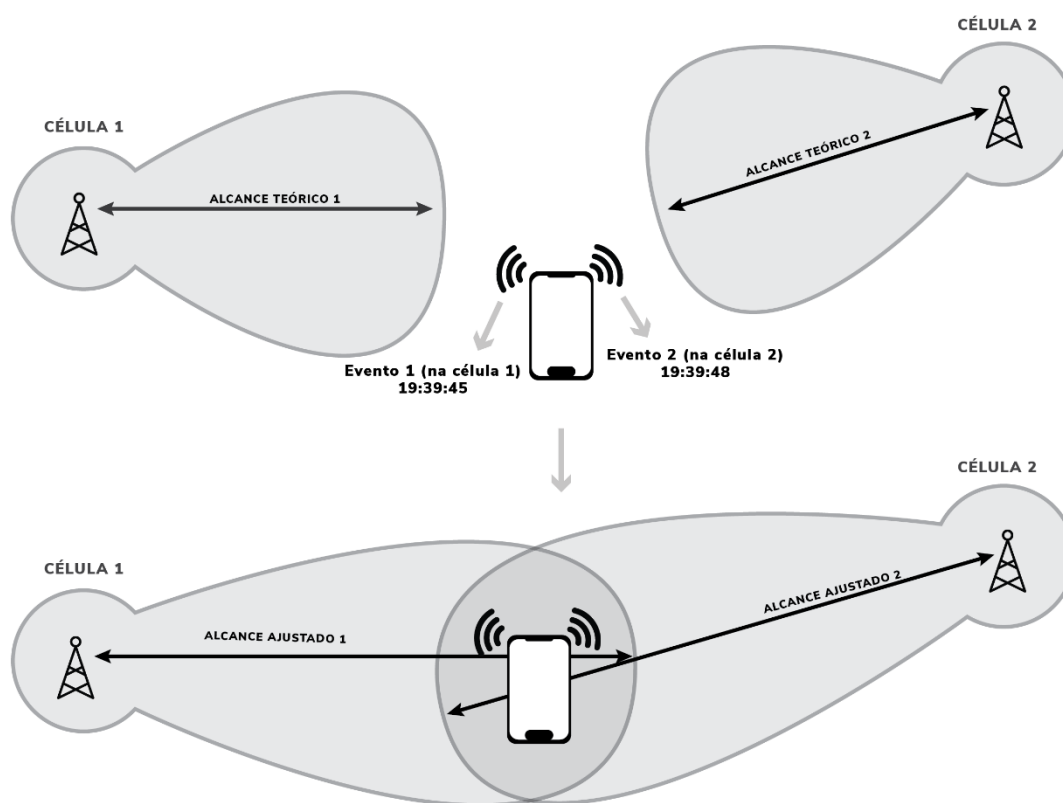
- Azimute: é a orientação da antena. 0 graus é orientação norte, 90 graus é a orientação leste etc.
- *Beamwidth*: é o ângulo de amplitude do feixe da célula. As mais comuns são as células de 65 graus de amplitude de feixe.
- Alcance: é a distância máxima em que a célula pode oferecer sinal na direção de seu azimute.
- Torre (latitude e longitude): é a localização exata do mastro da antena.

Com todos os parâmetros físicos apresentados na Figura 7 é possível obter, de forma muito aproximada, a área teórica de cobertura das células de telefonia móvel, fazendo com que os eventos sejam localizados na área de cobertura da célula em que foram produzidos. Salienta-se que a tecnologia da célula (4G, 3G e 2G) também afeta a precisão dos eventos (TELEFÔNICA, 2018a).

As células podem variar de alcance, cerca de 200 metros quando se tem tecnologia 4G, que é mais comum em áreas urbanas, ou até mesmo podem alcançar quilômetros. Células macro com baixa frequência podem chegar a ter um alcance de 12 quilômetros. Desta forma, a granularidade geográfica dos eventos de rede varia dependendo do intervalo de células em que ocorre e, conseqüentemente, se ocorre em ambientes urbanos ou rurais (TELEFÔNICA, 2018a).

Conforme a Telefônica (2018a), a cobertura real das células é afetada por agentes externos às características físicas da antena, como a saturação de usuários na célula ou o ambiente celular. A cobertura das células é ajustada com um recálculo do intervalo por um processo empírico que se baseia na observação dos eventos que de fato aconteceram. Por meio de um fenômeno chamado *flickering*, que corresponde à conexão de um determinado usuário em duas ou mais células em um curto espaço de tempo, onde esse usuário raramente consegue se mover, toda vez que este é observado pode-se afirmar que há uma sobreposição da cobertura das células envolvidas (TELEFÔNICA, 2018a). Isso pode ser representado pela Figura 8.

Figura 8 – Conceito de *flickering* como um instrumento para ajustar o alcance das células



Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

Considerando os eventos diários e as relações entre as células, o alcance de todas as células é ajustado. Depois de ser ajustado o intervalo, existe a segurança de que o usuário que produziu um evento esteja na área de cobertura ajustada, considerando-se a inserção desse processo empírico. Salienta-se que, de acordo com a Telefônica (2018a), a empresa já realizou estudos para medir a precisão geográfica dos eventos, analisando a diferença da distância entre o mastro da célula do evento com a localização do GPS produzida pelo mesmo usuário. Os resultados mostraram-se consistentes.

#### 1.1.5.2. Seleção da amostra e tratamento de dados

De acordo com a Telefônica (2018a), é necessário que sejam definidos os conceitos de usuários válidos e quem são esses usuários. Isso é necessário já que, durante a coleta de dados, alguns usuários deixam de ser clientes e outros passam a ser. Além disso, existem os clientes que fazem uso infrequente dos serviços, limitando, dessa maneira, as conclusões que se poderiam obter de seu comportamento. Sendo assim, para que a amostra não seja prejudicada, a remoção desses usuários do banco de dados mostra-se importante. Um usuário inválido é

definido como um usuário que não gera eventos suficientes para que sua mobilidade seja registrada de forma confiável. Um mesmo usuário pode ser válido em um determinado período de estudo, mas não ser válido para outro período. Logo, para a determinação de uma base válida de usuários, costuma-se observar uma redução na ordem de 10% a 15%<sup>4</sup>. Para que seja considerado como um usuário válido, usou-se o requisito de identificação de usuário de no mínimo sete dias do mês.

No final de cada fase de processamento das informações é realizada a limpeza dos dados, de modo que os erros possam ser detectados nos dados de entrada e nos resultados dos algoritmos de extração dos *insights*. Essa verificação de qualidade avalia o formato das informações, se os eventos possuem uma célula associada e se o volume de eventos está em um intervalo de três desvios padrões da distribuição de volumes dos últimos 30 dias. Ressalta-se que mesmo sem conhecimento da distribuição de probabilidade do volume de eventos, praticamente 100% dos eventos encontram-se em um intervalo de três desvios padrões em torno da média (WHEELER, 1995). Caso os dados não estejam de acordo com critérios anteriores, antes de descartá-los é realizada uma investigação para verificar a possibilidade de recuperação de tais informações.

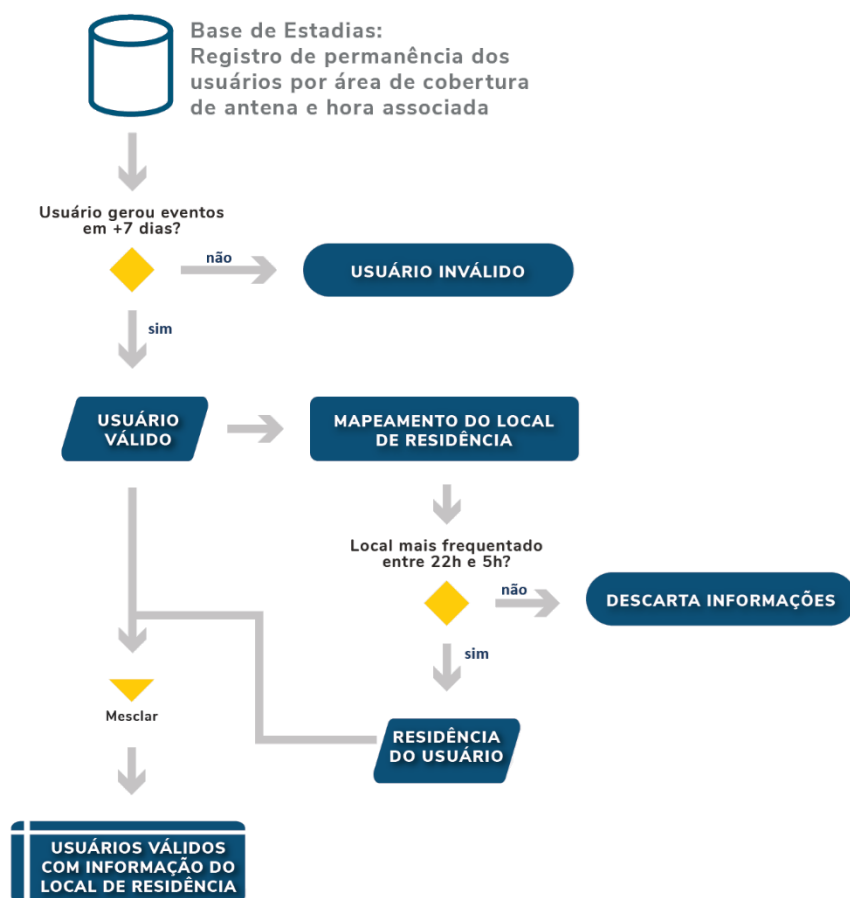
Após a realização da seleção da base de dados e limpeza da base, os eventos de rede são tratados para que se transformem em informações estruturadas e úteis para a análise de mobilidade. São elas: estadias, viagens e pontos intermediários. As estadias são locais em que os usuários permanecem por um tempo sem se movimentar. As viagens representam o deslocamento entre duas estadias. E os pontos intermediários são eventos observados entre a origem e o destino da viagem (TELEFÔNICA, 2018a).

Na Figura 9 é apresentado o diagrama de processos da criação da base de usuários, ilustrando as etapas e critérios citados anteriormente.

---

<sup>4</sup> A proporção de usuários descartados da amostra por UF está disponível no Apêndice 3.

Figura 9 – Diagrama de processo: base de usuários



Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

### 1.1.5.3. Representação estatística da amostra e extrapolação dos dados

Trabalhar com dados de telefonia móvel é algo vantajoso, uma vez que a amostra da população pode-se tornar muito maior do que a de uma pesquisa tradicional, aumentando a representatividade da amostra e também possibilitando a ampliação da quantidade de conclusões a serem obtidas. Para o cálculo da representatividade da amostra, deve-se considerar a margem de erro e o nível de confiança (TELEFÔNICA, 2018a).

Conforme Haag (2006, *apud* WEBER; PÉRSIGO, 2017) a margem de erro diz o quão perto a estatística da amostra<sup>5</sup> cai ou está em relação ao parâmetro da população. Ela existe porque

<sup>5</sup> Saliencia-se que isso acontece para amostras probabilísticas, ou seja, obtida por meio de algum sorteio não viciado.



os resultados que são obtidos em uma pesquisa partindo-se de uma amostra não são rigorosamente exatos ao universo/população onde foram extraídos. Ela diminui na medida em que a amostra é aumentada (GIL, 2007). Ademais, a margem de erro depende da unidade do parâmetro estimado. Como exemplo, caso a informação seja a média de tempo em segundos, a margem de erro será definida em segundos; se a amostra for de proporção (porcentagem) de ligações completadas, a margem de erro será definida em proporção absoluta ou percentual (BARBETA; REIS; BORNIA, 2010).

Segundo a Telefônica (2018a), a margem de erro é um índice que determina a estimativa máxima de erro dos resultados de uma pesquisa. Para o presente estudo, a margem de erro adotada refere-se à proporção de residentes no município. Assumindo um nível de confiança de 95%, se a margem de erro oferecida pela amostra for de 5%, isso significa que se a conclusão for de que 70% da amostra mora em uma determinada cidade, essa conclusão poderia ser tomada para a população com uma flutuação de 2,5% para mais ou para menos (com 95% de confiança).

Para Gil (2007), de acordo com a teoria das probabilidades, a distribuição das informações de média de tempo ou proporção de residentes coletadas a partir de amostras ajusta-se, geralmente, à curva normal gaussiana, em que as frequências associadas aos valores centrais são mais elevadas que as dos valores mais distantes dos centrais. Além disso, temos o nível de confiança, que é a certeza de que os dados que se pretende estimar em relação à população estão no intervalo estabelecido pela margem de erro. Pelo exposto, se a amostra possuir um nível de confiança de 95%, significa que a proporção da característica de interesse na população tem uma probabilidade de 95% de estar dentro da margem de erro especificada. Quando na seleção da amostra é considerado um desvio-padrão, trabalha-se com um nível de confiança de aproximadamente 68%, quando se tem dois desvios-padrão, trabalha-se com 95,5% de confiança, e quando se tem três desvios-padrão, tem-se um nível de confiança de 99,7% (GIL, 2007).

Conforme Triola (2005), a relação existente entre o tamanho da população, o tamanho da amostra, a margem de erro e o nível de confiança pode ser representada pela Equação:

$$n = \frac{N \cdot (Z_{(\alpha/2)})^2 \cdot p \cdot (1 - p)}{(N - 1) \cdot e^2 + (Z_{(\alpha/2)})^2 \cdot p \cdot (1 - p)}$$

Onde:

$N$ : tamanho da população

$n$ : tamanho da amostra

$e$ : margem máxima de erro permitida

$Z$ : desvio do valor médio para atingir o nível de confiança  $\alpha$  desejado (é dado pelas tabelas da distribuição normal, de modo que, se o nível de confiança desejado for de 99%, por exemplo, então  $Z = 2,57$ )

$p$ : proporção esperada para ser encontrada. Se não for conhecido, é usado 50%, o que representa um valor mais conservador.

Quando o tamanho da população é muito grande, maior que 100 mil indivíduos, a Equação pode ser simplificada da seguinte forma:

$$n = \frac{(Z_{(\alpha/2)})^2 \cdot p \cdot (1 - p)}{e^2}$$

Em relação à representatividade estatística da amostra, a Tabela 2 apresenta o nível de confiança e a margem de erro por UF.

**Tabela 2 – Representatividade estatística da amostra por UF – Fev/2019**

UF	População	Quantidade de celulares	Market Share Vivo	Nível de confiança	Margem de erro
AC	829.619	494.085	65,00%	99,00%	0,18%
AL	3.375.823	373.029	12,70%	99,00%	0,21%
AM	4.063.614	2.115.742	60,00%	99,00%	0,09%
AP	797.722	402.099	55,30%	99,00%	0,20%
BA	15.344.447	3.948.860	28,80%	99,00%	0,07%
CE	9.020.460	61.917	7,20%	99,00%	0,52%
DF	3.039.444	1.339.940	27,10%	99,00%	0,11%
ES	4.016.356	3.000.667	78,60%	99,00%	0,07%
GO	6.778.772	1.876.788	27,10%	99,00%	0,09%
MA	7.000.229	1.070.495	19,70%	99,00%	0,12%
MG	21.119.536	8.698.840	40,00%	99,00%	0,04%
MS	2.713.147	1.482.476	47,90%	99,00%	0,11%
MT	3.344.544	2.081.782	54,20%	99,00%	0,09%
PA	8.366.628	2.696.256	37,30%	99,00%	0,08%
PB	4.025.558	499.475	12,70%	99,00%	0,18%
PE	9.473.266	971.186	10,20%	99,00%	0,13%
PI	3.219.257	544.149	16,80%	99,00%	0,18%
PR	11.320.892	2.160.615	16,70%	99,00%	0,09%
RJ	16.718.956	5.368.366	27,80%	99,00%	0,06%
RN	3.507.003	180.935	5,50%	99,00%	0,30%
RO	1.805.788	359.212	20,20%	99,00%	0,22%

UF	População	Quantidade de celulares	Market Share Vivo	Nível de confiança	Margem de erro
RR	522.636	311.738	61,40%	99,00%	0,23%
RS	11.322.895	6.083.730	47,20%	99,00%	0,05%
SC	7.001.161	1.958.400	24,70%	99,00%	0,09%
SE	2.288.116	1.403.152	68,50%	99,00%	0,11%
SP	45.094.866	22.861.458	36,30%	99,00%	0,03%
TO	15.501.94	257.465	16,30%	99,00%	0,25%

Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

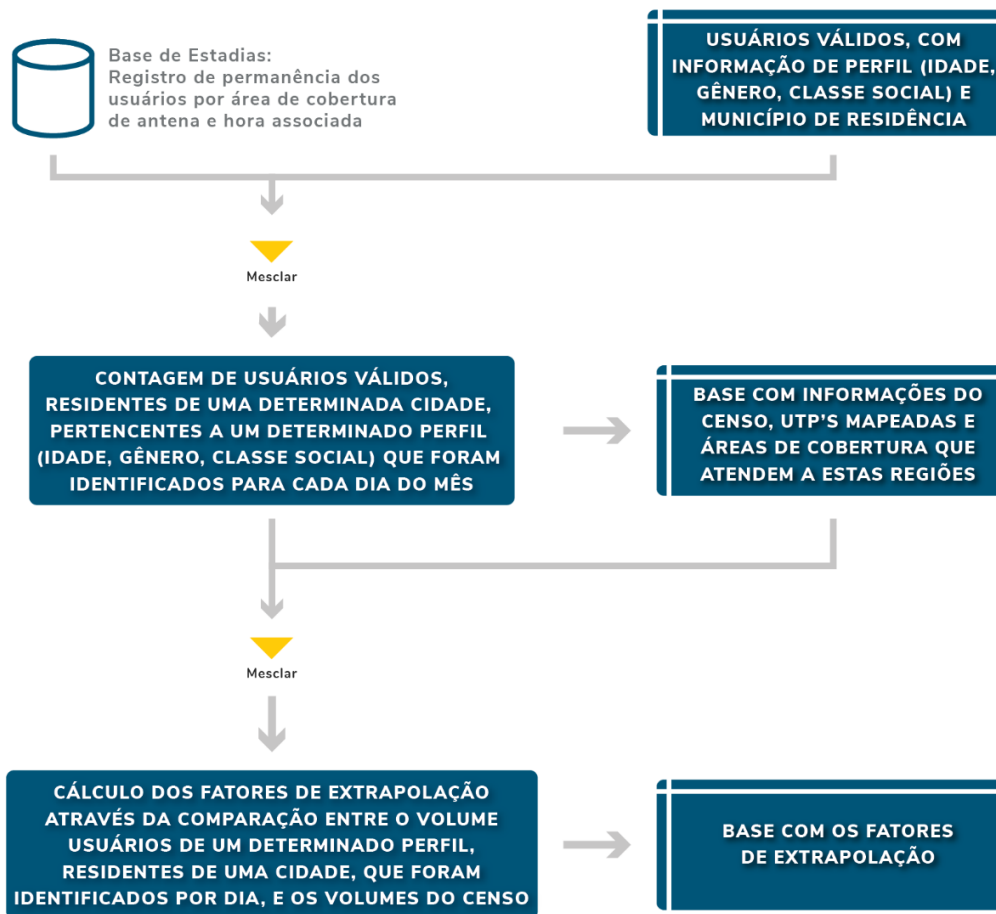
Conforme informado pela Telefônica (2018a), por questões regulatórias e respeito à privacidade dos seus usuários, a empresa possui algumas restrições quanto aos dados que pode compartilhar. Nesse sentido não serão compartilhados dados de deslocamento não extrapolados nem aqueles que representem volumes inferiores a dez viagens. Entretanto, existem informações públicas na esfera macro que permitem entender a relevância da amostra. Grande parte dessas informações pode ser encontrada no *site*: <http://www.teleco.com.br>, que resume uma série de estatísticas a respeito do mercado brasileiro de telecomunicações, como já citado na Tabela 1.

No que se refere à extrapolação, todos os dados do estudo foram extrapolados, ou seja, representam o comportamento de toda população, e não apenas dos usuários da contratada. Algumas etapas foram realizadas durante o processo de extrapolação. São elas:

1. A construção da população residente por zona (Telefônica)
2. A construção da população residente por zona (Censo)
3. A identificação de usuários por dia (usuários que deixam registro de permanência em determinado dia utilizado para identificação).

A Figura 10 mostra o processo de registro de dados e cálculo dos fatores de extrapolação para formação da base de extrapolação.

Figura 10 – Diagrama de processo da base com fatores de extrapolação



Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

### 1.1.6. Rastreamento do usuário – Identificação e categorização de viagem

Os usuários de telefonia móvel geram diariamente grande quantidade de registros na rede, através dos eventos de comunicação com as células descritos no item 1.1.5. Conforme a Telefônica (2018a), a compreensão íntima desses eventos possibilita entender contextualmente o movimento dos usuários no espaço e no tempo, o que permite registros que são interpretados como estadias e viagens. Para fins de elaboração da matriz pretendida no estudo, esses registros devem ser analisados de modo a identificar:

- O deslocamento entre UTPs
- A classificação do modo de transporte utilizado em aéreo ou terrestre

E, se o modo de transporte é classificado em aéreo:

- O município de origem real
- O município de destino real
- O aeródromo de embarque
- O aeródromo de desembarque.

Como abordado no item 1.1.4, a LUCA tem um portfólio relevante de *cases* nacionais<sup>6</sup> e internacionais de utilização da plataforma LUCA Transit para análises de padrões de mobilidade, especialmente com a utilização de transporte público. Dispondo dos dados de telefonia móvel e dos dados cadastrais dos usuários, tem-se a possibilidade de identificar o perfil demográfico dos usuários, seu local de residência e de trabalho, entre outras informações, apresentadas na Figura 11.

Figura 11 – Informações possíveis de serem levantadas pela LUCA



Fonte: LUCA ([201-]a). Elaboração: LabTrans/UFSC (2020)

No presente estudo, no entanto, devido a sua abrangência geográfica e às necessidades de classificação do modo de transporte em aéreo ou não aéreo e de identificação dos aeródromos de embarque e de desembarque dos usuários, ao longo do processo de validação do extrato parcial, foram identificados diversos desafios, principalmente em relação à correta classificação do modo de transporte.

A seguir são descritos os parâmetros adotados para essa classificação ao final do processo de validação do extrato parcial, bem como para obtenção dos dados para deslocamentos pelo modo aéreo.

<sup>6</sup> Na Espanha, uma vez que a empresa está localizada em Madrid.

### 1.1.6.1. Classificação do modo de transporte

De acordo com o TR, anexo ao documento de Seleção Pública nº 105/2018, Processo nº 331271729, da FAPEU:

A classificação do modo de transporte utilizado (aéreo ou não aéreo) será feita com base no tempo de deslocamento do usuário no par O/D entre as UTPs (Anexo 6). Deverá ser considerada uma velocidade de 200 km/h como valor de referência entre cada ligação O/D para a divisão entre aéreo e não aéreo. Quando na aferição do último local de um registro na origem e do primeiro local desse registro no destino verificar-se que, em razão do tempo e da distância de deslocamento, a velocidade resultar superior ao valor de referência da ligação O/D, aliado à proximidade a aeródromos em um dos extremos da viagem, o modo de deslocamento desse registro será classificado como aéreo (FAPEU, 2018, p. 4-5).

A Telefônica (2018a) indicou que, devido à natureza dos dados (de telefonia, gerados a partir de chamadas de voz e do consumo de dados), a inferência do modo de transporte utilizado só é possível com base no padrão de velocidade média das viagens, sendo calculada da seguinte maneira:

$$V_m \left( \frac{km}{h} \right) = \frac{C_{destino} - C_{origem} (km)}{T_{destino} - T_{origem} (h)}$$

Onde:

$V_m$  = Velocidade média do deslocamento (km/h)

$C_{destino}$  = Centróide da área de cobertura da célula em que houve a primeira detecção (registro) do usuário na UTP de destino

$C_{origem}$  = Centróide da área de cobertura da célula em que houve a última detecção (registro) do usuário na UTP de origem

$T_{destino}$  = Instante da última detecção (registro) do usuário na UTP de origem

$T_{origem}$  = Instante da primeira detecção (registro) do usuário na UTP de destino

No processo de validação do extrato parcial, que gerou oito versões (apresentadas no item 1.3.3), verificou-se a necessidade de uma terceira classificação do modo de transporte com o propósito de indicar um alerta para a necessidade de revalidação. Assim, os modos de transporte são classificados em:

- **Modo 0:** não aéreo (ou terrestre)
- **Modo 1:** aéreo
- **Modo 2:** modo de revalidação e realocação posterior.

Além disso, identificou-se a necessidade de definição de novos critérios visando à correta classificação do modo de transporte utilizado. A classificação nos três modos é realizada de acordo com critérios de distância e velocidade, além de checagem da identificação do viajante próximo a um aeródromo considerado na pesquisa, como apresenta a Tabela 3.

**Tabela 3 – Critérios de classificação do modo de transporte**

Distância	Velocidade	Proximidade de aeródromo	Classificação
> 500 km (viagem de longa distância)	≤ 200 km/h	-	<b>Modo 0</b>
	> 200 km/h	-	<b>Modo 1</b>
< 500 km (viagem de curta distância)	< 100 km/h	-	<b>Modo 0</b>
	> 100 km/h	Com identificação do usuário em um raio de 15 km do aeródromo de alocação da viagem	<b>Modo 1</b>
		Sem identificação do usuário em um raio de 15 km do aeródromo de alocação da viagem	<b>Modo 2</b>

Elaboração: LabTrans/UFSC (2020)

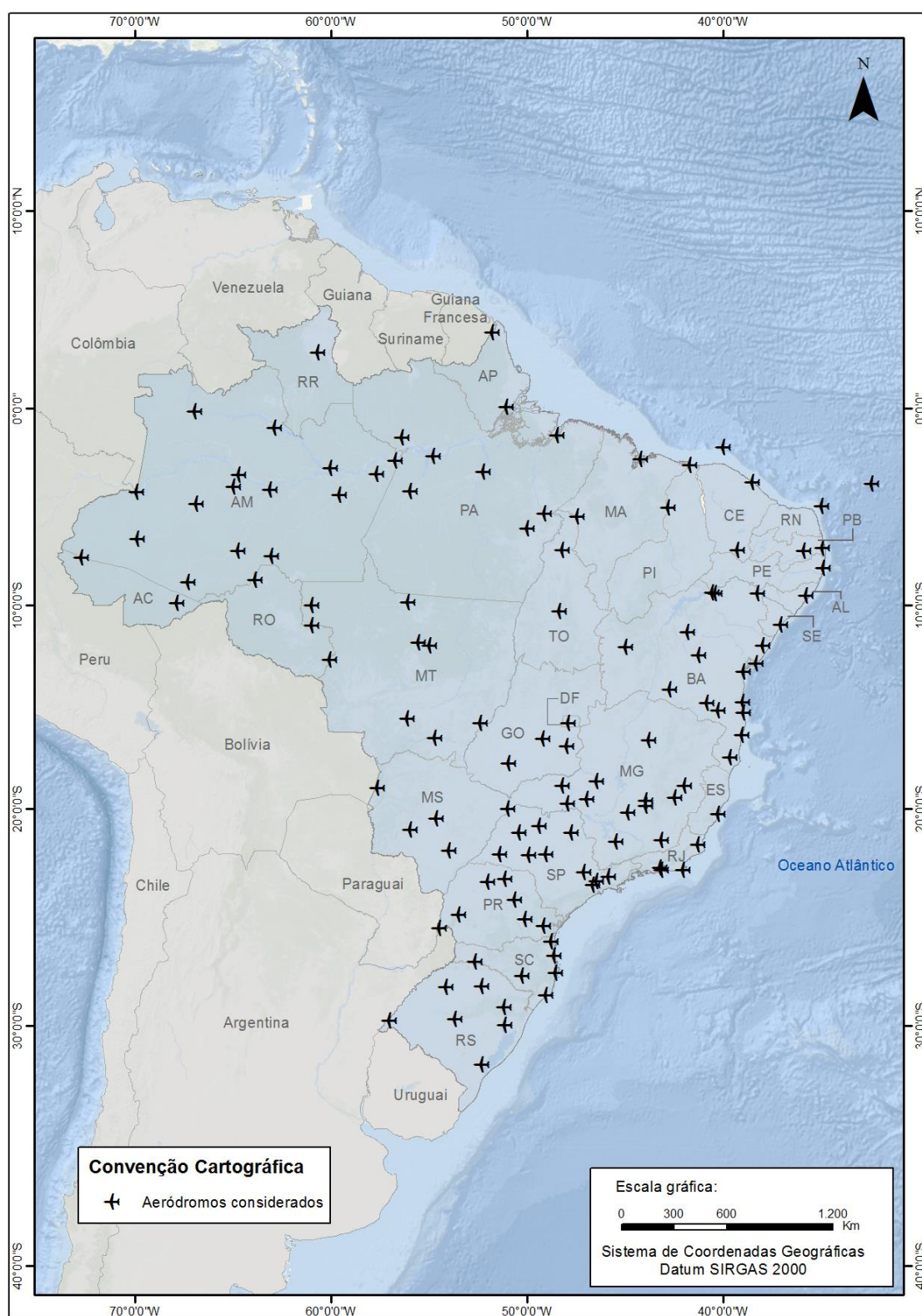
#### 1.1.6.2. Obtenção dos dados para deslocamentos aéreos

Conforme a Telefônica (2018a), para a classificação do deslocamento como sendo Modo 1, é definida a real origem como sendo o último município onde o usuário permaneceu por seis horas, desconsiderando-se o período utilizado para validação da viagem. Também é definido como real destino o primeiro município onde o usuário permaneceu por seis horas após o voo, desconsiderando-se as horas utilizadas para classificação do aeródromo como destino da viagem. Essa definição está de acordo com o definido no TR por Wang *et al.* (2013 *apud* FAPEU, 2018, p. 5) “Serão caracterizados como origem ou destino os locais (início ou final de uma viagem) em que o registro permanecer estacionado por período entre 4 a 6 (seis) horas”.

Por sua vez, o aeródromo de embarque e o de desembarque são definidos como, respectivamente, o aeródromo mais próximo da antena em que houve a última detecção do usuário na UTP de origem e o aeródromo mais próximo da antena em que houve a primeira detecção do usuário na UTP de destino. Os aeródromos considerados estão restritos àqueles em que houve movimentação de passageiros da aviação comercial no ano de 2017, conforme o Anexo “005\_Aerodromo” à Seleção Pública nº 105/2018, dispostos na Figura 12.



Figura 12 – Aeródromos com registros de voo comercial em 2017

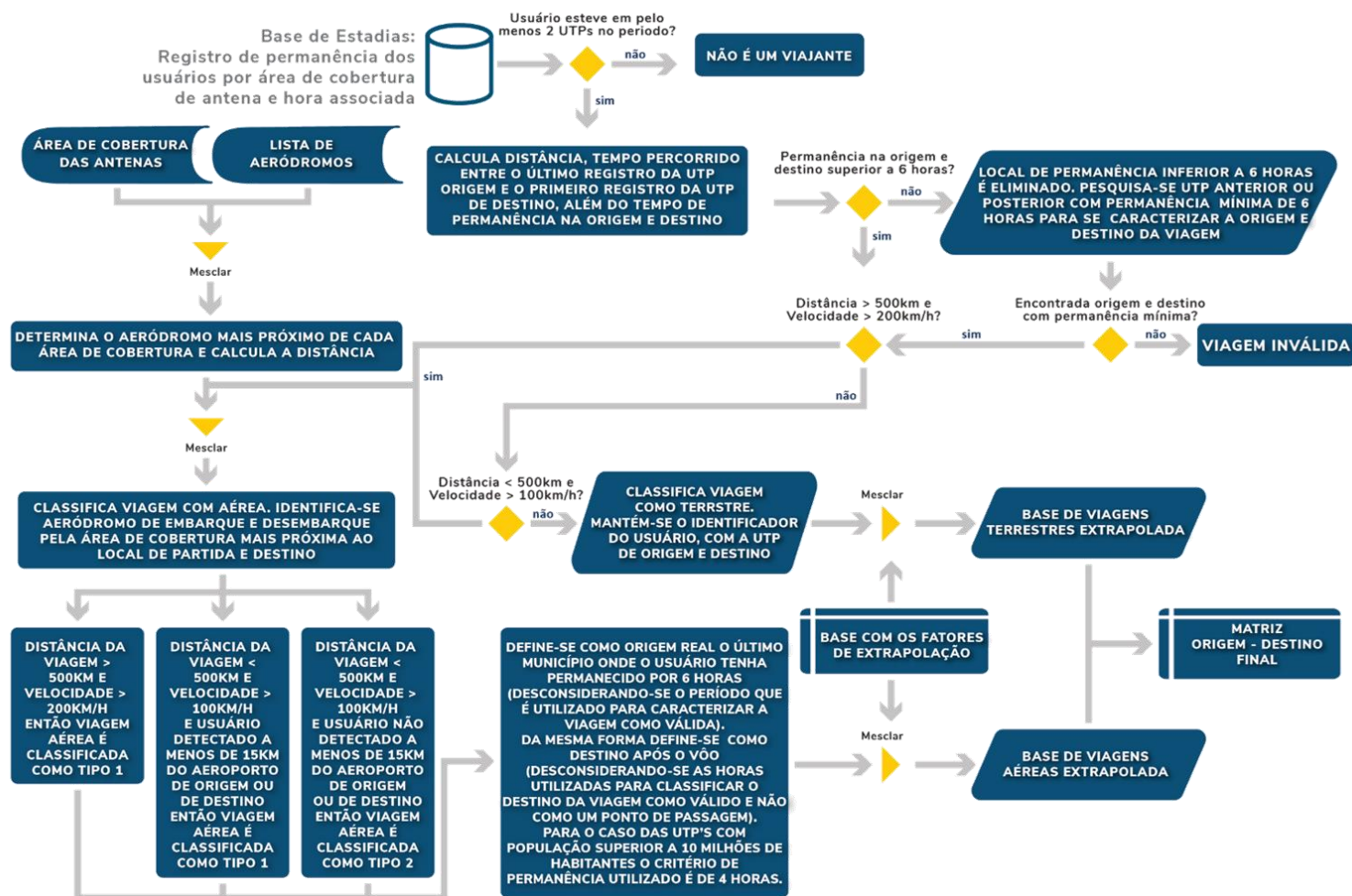


Fonte: FAPEU (2018). Elaboração: LabTrans/UFSC (2020)

A Figura 13 apresenta o diagrama de processo de determinação e caracterização das viagens.



Figura 13 – Diagrama de processo: base de viagens



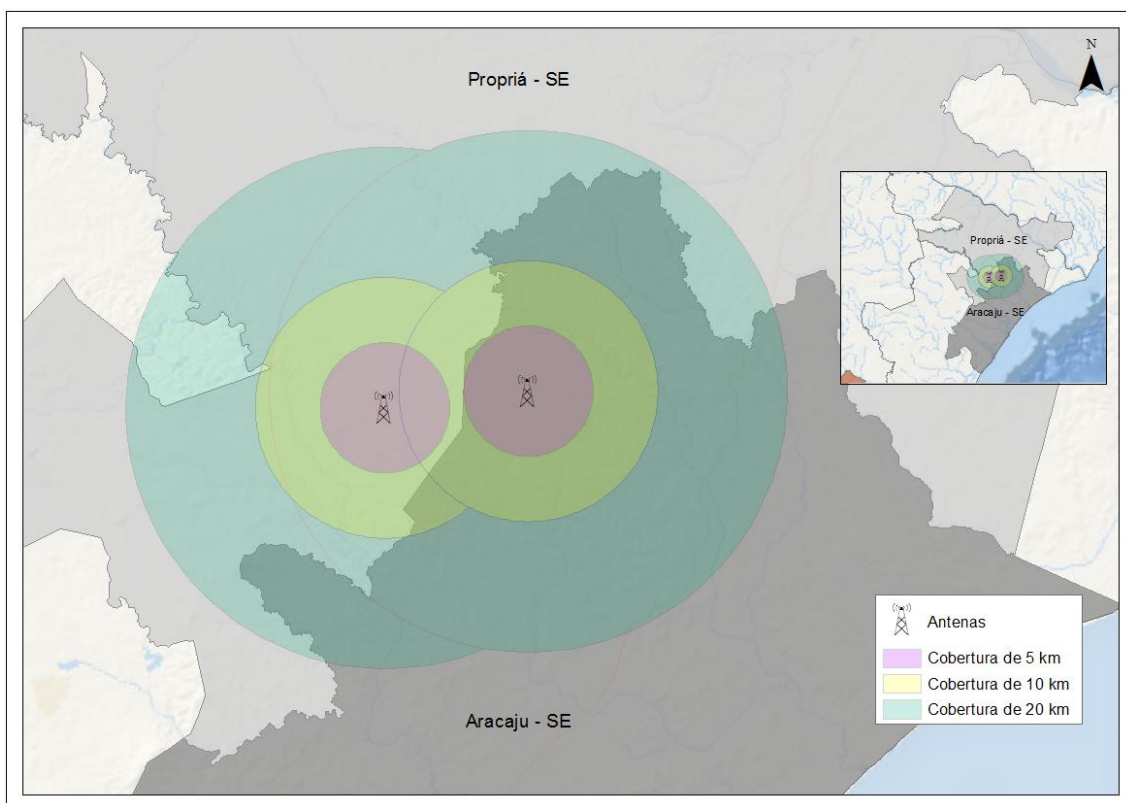
Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

### 1.1.6.3. Limitações metodológicas

Nas análises dos dados verificou-se que, quando há deslocamentos próximos aos limites de uma UTP, o usuário pode ser detectado por antenas instaladas em UTPs vizinhas e que tenham cobertura sobre a primeira. Tal detecção configura o deslocamento entre UTPs, conforme a metodologia aplicada. Em UTPs de menor área e/ou com alta densidade de torres, essa ocorrência pode ser comum.

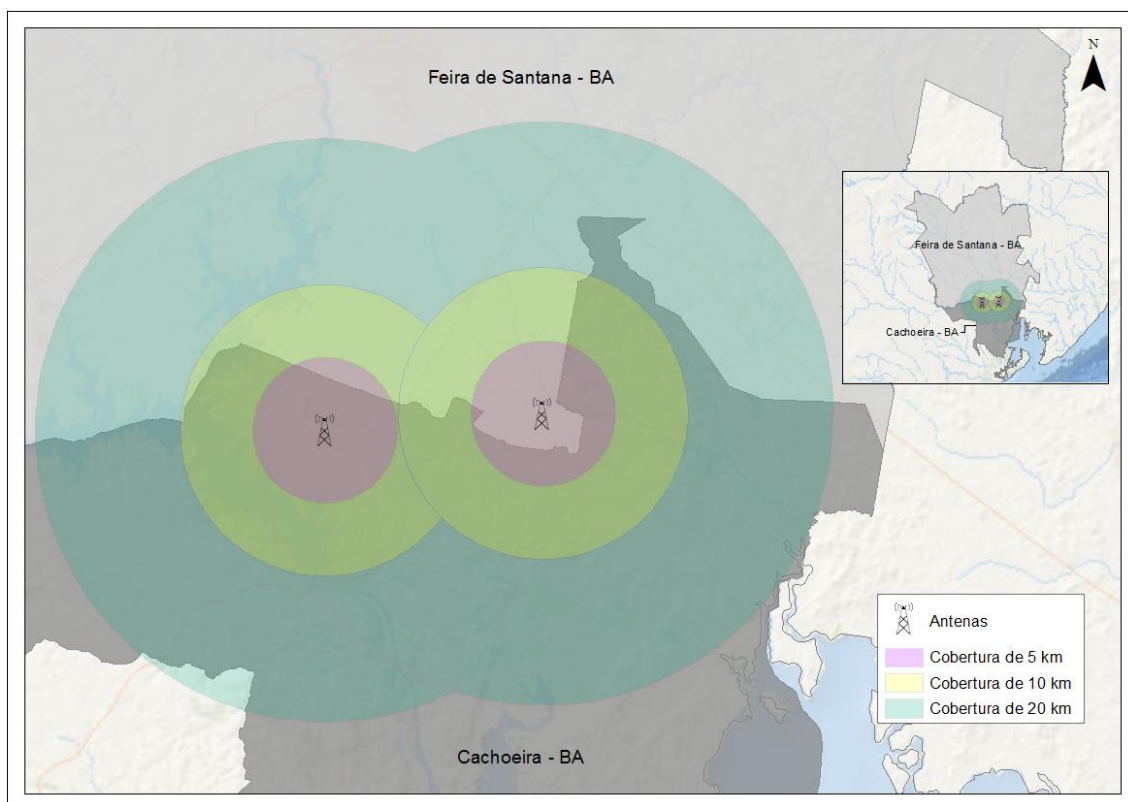
A Figura 14 e a Figura 15 apresentam exemplos para ilustrar esse caso. Neles, a localização das antenas é real e utilizou-se arquivo *shapefile* das antenas obtido com a contratada. As áreas de cobertura das antenas, no entanto, são fictícias e consideradas circulares. Conforme descreve a Anatel (2020, não paginado) “[...] a cobertura efetiva depende de fatores como relevo, construções, localização do usuário, tipo de ambiente (aberto ou dentro de construções e a depender da natureza dessas construções), altura do usuário em relação ao solo, etc.”

Figura 14 – Exemplo: UTP Propriá (SE) e Aracaju (SE)



Elaboração: LabTrans/UFSC (2020)

Figura 15 – Exemplo: UTP Feira de Santana (BA) e Cachoeira (BA)

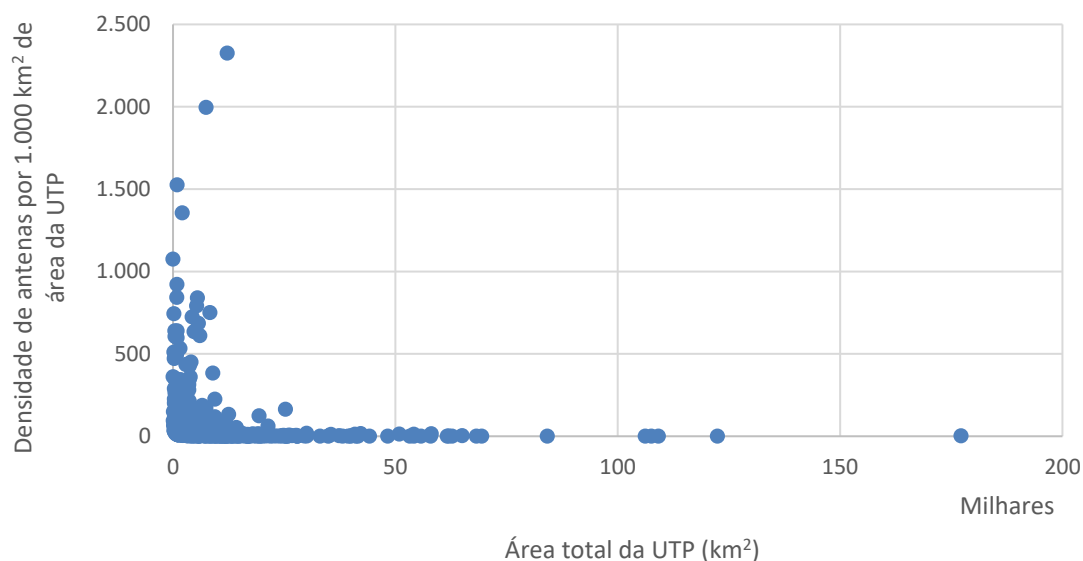


Elaboração: LabTrans/UFSC (2020)

A ocorrência descrita – e representada na Figura 14 e na Figura 15– é comum para UTPs de pequena área e/ou com elevada quantidade de antenas, especialmente quando combinadas essas características.

Nesse sentido, o Gráfico 1 apresenta a dispersão das UTPs em área (km<sup>2</sup>) e densidade de antenas, considerada a quantidade de antenas por 100 mil km<sup>2</sup>. Pode-se observar que boa parte delas apresenta área reduzida e uma elevada densidade de antenas.

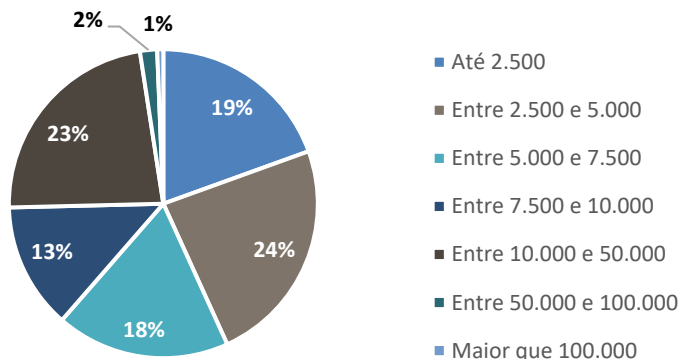
**Gráfico 1 – Dispersão das UTPs nas variáveis área (km<sup>2</sup>) e densidade de antenas**



Fonte: Dados fornecidos pela Telefônica (2019). Elaboração: LabTrans/UFSC (2020)

No que se refere às UTPs, como se observa no Gráfico 2, 43% delas apresentam área total até 5 mil km<sup>2</sup>.

**Gráfico 2 – Distribuição das UTPs por intervalos de área total (km<sup>2</sup>)**



Elaboração: LabTrans/UFSC (2020)

Além disso, a Telefônica (2018a) atenta para o fato de que, ainda que o deslocamento seja corretamente classificado no Modo 1 (aéreo), pode ocorrer um lapso de tempo entre a efetiva partida/chegada do usuário na UTP e sua detecção através de registros nas células. O deslocamento do usuário nesse lapso de tempo consiste em um dos elementos mais sensíveis no processo de definição dos aeródromos de embarque e de desembarque (TELEFÔNICA, 2018a).

Observou-se que esse fato, inerente à tecnologia utilizada na coleta de dados, poderia ter impacto na identificação de deslocamentos aéreos em pares OD específicos, a exemplo da

ponte aérea Rio-São Paulo, nos dois sentidos. Para contorná-lo, considerou-se reduzir o parâmetro de tempo estacionário, ou seja, o tempo em que um usuário permanece em uma UTP antes de realizar um deslocamento, para definição da origem e do destino reais para quatro horas, o limite inferior recomendado por Wang *et al.* (2013). Esse critério mostrou-se válido para esse par OD específico.

## 1.2. Etapas do estudo

As etapas do estudo, já apresentadas na Figura 1, consistem em:

- Desenvolvimento do tema e da metodologia de pesquisa
- Desenvolvimento do TR e execução do edital de contratação da base de dados
- Validação da base de dados e materiais correlatos
- Expansão e elaboração da matriz OD
- Atualização de aspectos metodológicos evoluídos na pesquisa
- Apresentação dos resultados alcançados através do *Relatório de Metodologia e resultados da matriz OD dos reais desejos de viagens contemplados ou não pelo transporte aéreo*.

No que se refere ao desenvolvimento da matriz OD, a Tabela 4 apresenta as entregas realizadas. Observa-se que houve oito versões do extrato da base de dados (com dados referentes ao mês de maio, quando das versões parciais, para facilitar no trabalho de validação da base).

Tabela 4 – Entregas da base de dados e suas versões de melhoramentos – 2019

Data	Entrega
27/02/2019	Relatório de Metodologia – Versão 1
31/03/2019	Extrato parcial – Versão 1
25/04/2019	Extrato parcial – Versão 2
30/04/2019	Relatório de Metodologia – Versão 2
28/05/2019	Extrato parcial – Versão 3
18/06/2019	Extrato parcial – Versão 4
17/07/2019	Extrato parcial – Versão 5
30/07/2019	Extrato parcial – Versão 6
19/08/2019	Extrato parcial – Versão 7

Data	Entrega
16/09/2019	Base de Dados Completa Relatório de Metodologia V3 Relatório Descritivo do Método de Geração da Base de Dados V1 Tabelas de Dados Auxiliares V1 Metadados V1
11/10/2019	Relatório de Metodologia V4 Relatório Descritivo do Método de Geração da Base de Dados V2 Tabelas de Dados Auxiliares V2 Metadados V2
05/12/2019	Base de Dados Completa – Versão 8

Elaboração: LabTrans/UFSC (2020)

### 1.3. Processo de validação dos dados fornecidos

Os itens a seguir descrevem as bases utilizadas para validação dos dados fornecidos, as regras de validação, as análises dos extratos parciais (com dados referentes ao mês de maio de 2017) e do extrato completo.

#### 1.3.1. Bases de dados de comparação

Na validação da base de dados foram realizadas comparações com outras bases e registros de deslocamento da população pelo modo aéreo e pelo modo terrestre. Além dessas bases, foram utilizadas informações quanto à concentração e influência urbana das cidades brasileiras que fazem parte das UTPs analisadas. As bases de dados, registros e informações complementares utilizadas para comparação dos resultados estão apresentadas na Tabela 5.

**Tabela 5 – Bases de dados utilizadas no processo de validação dos dados da Telefônica**

Nome	Fonte	Modo de transporte	Descrição
Pesquisa de origem/destino do transporte aéreo de passageiros	Empresa de Planejamento e Logística S.A. (EPL)	Modo 1 -Aéreo	Pesquisa realizada em 2014 com coleta de dados por meio de entrevistas em 65 aeroportos, sendo 53,7% dos aeroportos operantes de voos regulares em 2012 e com participação de 99% no total de passageiros embarcados.
Base estatística	ANAC		Base de dados operacionais do transporte aéreo, que apresenta, para cada par OD e empresa aérea, a quantidade total de passageiros pagos e de passageiros grátis, entre outras informações.
Microdados de tarifas aéreas domésticas			Base de dados das tarifas comercializadas pelas empresas aéreas, que apresenta o - número de assentos vendidos no mês de referência para a chave composta por ano e mês de venda da passagem, empresa aérea, aeroportos de origem e destino e valor de tarifa.

Nome	Fonte	Modo de transporte	Descrição
Dados operacionais	ANTT	Modo 0 - Não aéreo	Base de dados de demanda do transporte rodoviário intermunicipal e inter-regional
Arranjos populacionais e concentrações urbanas	IBGE		Identifica e delimita as maiores aglomerações de população, contempla os deslocamentos pendulares realizados em diversas regiões brasileiras
Hierarquia urbana brasileira			Define a hierarquia dos centros urbanos brasileiros e delimita as regiões de influência a eles associados

Fonte: ANAC (c2020, 2019a), ANTT ([201-]) e EPL (2014). Elaboração: LabTrans/UFSC (2020)

Referente à pesquisa de origem/destino do transporte aéreo de passageiros (EPL, 2014), embora ela não contemple todos os aeroportos com voos regulares, os dados coletados na pesquisa incluem os municípios de origem e de destino real das viagens. Tal informação permite a comparação direta do volume de viagens por modo aéreo dessa base com os dados fornecidos.

No que tange à comparação com bases da ANAC, os dados fornecidos pelo *Big Data* de deslocamentos pelo modo aéreo referem-se a pares OD considerando os aeroportos de origem e de destino reais do passageiro, ao passo que, na base estatística da ANAC, esses deslocamentos podem estar fragmentados em etapas, devido a escalas e conexões em aeroportos intermediários. No entanto, esta é a base de dados operacionais do modo aéreo mais completa e confiável com divulgação de movimentações de passageiros e foi, portanto, contemplada na validação dos dados fornecidos em termos dos fluxos com maior volume de deslocamento, ordens de grandeza e sazonalidade.

A base dos microdados de tarifas aéreas, por sua vez, contém o metadado “assentos comercializados” por par OD considerando a origem inicial e o destino final constantes do bilhete, independente de escalas e conexões no trajeto, como informado pela ANAC em questionamento pelo LabTrans/UFSC. Dessa forma, podem-se inferir movimentações de passageiros comparáveis com os dados fornecidos. No entanto, deve-se atentar para o fato de que “O número de assentos vendidos registrados corresponde a cerca de 50% do movimento de passageiros pagos transportados.” (ANAC, 2019b, não paginado).

Para o transporte não aéreo, a base de dados operacionais (ANTT, [201-]) apresenta a demanda do serviço regular rodoviário e semiurbano em deslocamentos entre diferentes UFs. É possível, portanto, utilizar a base para buscar correspondências e comparar com a proporção do número de viagens do Modo 0 nos dados fornecidos.

Relativamente à base de *Arranjos Populacionais e Concentrações Urbanas* do Brasil (IBGE, 2016), é possível identificar localidades com grande número de pessoas realizando

deslocamentos pendulares, caso que pode ocorrer em deslocamentos entre UTPs específicas da base, possibilitando a validação dos dados nesses casos.

Por fim, utilizaram-se também dados da Hierarquia Urbana Brasileira (IBGE, [201-]), que divide as cidades brasileiras em sete grandes grupos. Verificou-se o número de viagens entre pares dentro do mesmo grupo hierárquico e entre os grupos, avaliando-se a taxa de viagens (número total de viagens/população total do par OD) buscando semelhanças entre pares de mesmo grupo e distância.

### 1.3.2. Regras de validação

No que se refere às regras de validação, foram levantados três testes para avaliar a base de deslocamentos classificados no Modo 1 (aéreo), com base em situações consideradas como potenciais erros de classificação do modo de viagem, que geram falsos positivos, ou de alocação aos aeródromos de embarque e de desembarque. Essas regras são apresentadas na Tabela 6.

**Tabela 6 – Regras de validação dos deslocamentos do Modo 1**

Regra	Descrição
1	Aeroporto de origem igual ao aeroporto de destino
2	Distância rodoviária <sup>7</sup> entre o município de origem e o município de destino inferior a 300 km <sup>8</sup>
3	Distância rodoviária entre o município de origem e o município de destino inferior à soma da distância entre o aeródromo de embarque e o de desembarque e das distâncias rodoviárias <sup>9</sup> entre o município de origem e o aeródromo de embarque e o aeródromo de desembarque e o município de destino

Elaboração: LabTrans/UFSC (2020)

<sup>7</sup> A distância rodoviária foi obtida por meio de consulta em uma matriz distância gerada em SIG com camada do modal rodoviário.

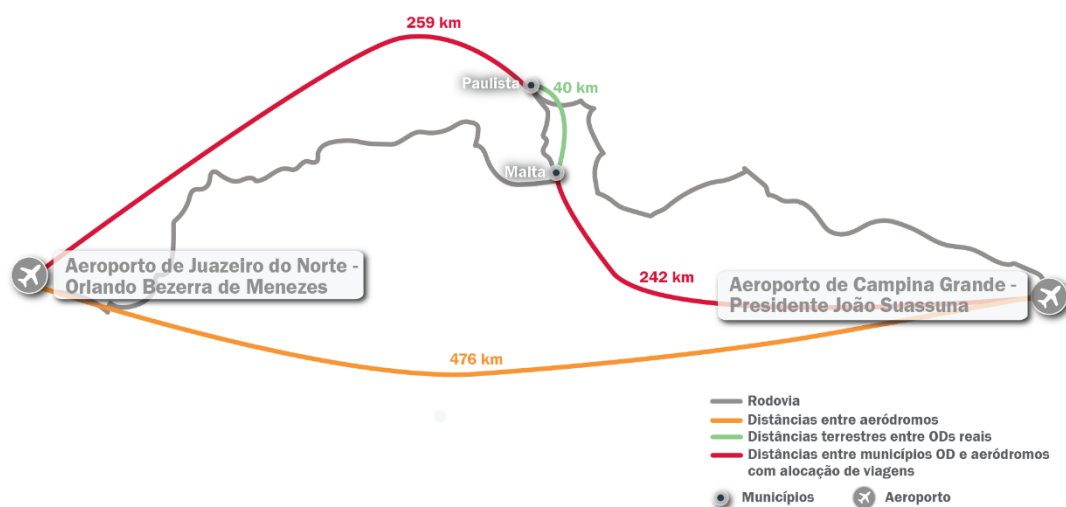
<sup>8</sup> O critério de 300km foi adotado a partir de análises das rotas atuais, em operação.

<sup>9</sup> Quando da inexistência de conexão rodoviária, utilizou-se a distância geográfica, considerando os centroides dos municípios, de acordo com o IBGE, e a localização dos aeroportos.



A Figura 16 representa um exemplo da confirmação da Regra 3.

Figura 16 – Exemplo representativo da Regra 3



Elaboração: LabTrans/UFSC (2020)

Ao longo do processo de validação dos extratos parciais, buscou-se verificar se os ajustes de parâmetros reduziram as quantidades de pares e de deslocamentos/viagens que atendem uma ou mais dessas regras. A Tabela 7 apresenta o processo de análise e a redução dessas ocorrências na evolução das versões da base de dados.

Tabela 7 – Histórico da análise das versões do extrato parcial pelas regras de validação dos deslocamentos do Modo 1

Regra	Descrição	Versão 1	Versão 2	Versão 3	Versão 4	Versão 5	Versão 6	Versão 7	Versão 8
1	Pares (%)	2%	2%	21%	0%	7,2%	7,6%	7%	1%
	Viagens (%)	1%	2%	11%	0%	2,6%	2,7%	2,7%	0,2%
2	Pares (%)	33%	34%	38%	23%	16%	17%	15%	3%
	Viagens (%)	16%	18%	19%	9%	4%	5%	5%	1%
3	Pares (%)	5%	5%	26%	6%	9%	9%	13,4%	-
	Viagens (%)	3%	3%	13%	3%	3%	3%	4,7%	-

Elaboração: LabTrans/UFSC (2020)

As análises parciais foram importantes para o refinamento da amostra e validação dos parâmetros, sejam eles critérios de modo de viagem, alocação de aeroportos, ou volume de

viagens levantadas nos pares OD identificados. Mais detalhes do refinamento da base de dados podem ser vistos no item Análises parciais.

### **1.3.3. Análises parciais**

As análises parciais referem-se às análises do extrato parcial, com dados de maio de 2017. Esse mês foi selecionado por não consistir em um mês atípico na sazonalidade de movimentação de passageiros observada no Brasil, como aqueles das temporadas de férias de inverno e de verão. Esse extrato foi gerado para fins de validação pelo LabTrans/UFSC antes da execução da consulta pela Telefônica dos dados referentes ao ano de 2017 completo.

Essas análises parciais foram realizadas durante o período de suporte técnico da Telefônica e, portanto, as necessidades de ajustes de parâmetros de consulta geraram novas versões do extrato parcial, totalizando oito versões, descritas a seguir.

#### **1.3.3.1. Versão 1**

Na primeira versão do extrato parcial, entregue em 31 de março de 2019, observou-se um número excessivo de viagens classificadas no Modo 0 (terrestre ou não aéreo). Isso ocorreu em virtude do parâmetro de tempo de permanência utilizado para identificação de deslocamento entre UTPs e definição da UTP de origem e de destino. O tempo de permanência/estacionário considerado foi de 30 minutos, não estando de acordo com a especificação do TR, igual a seis horas.

Além disso, identificou-se que a variável renda não havia sido utilizada na expansão dos dados e que esta seria relevante, especialmente para os deslocamentos classificados no Modo 1 (aéreo), conforme entendimento do LabTrans/UFSC e da SAC/MInfra.

#### **1.3.3.2. Versão 2**

Para a segunda versão, foi alterado o tempo de permanência para os deslocamentos classificados no Modo 0, de 30 minutos para seis horas, bem como para os classificados no Modo 1 (aéreo), alterado de quatro para seis horas, conforme definição do TR. Esses ajustes reduziram o número de viagens em 77%.

No entanto, nessa versão, entregue em 25 de abril de 2019, observou-se que, ao considerar os aeródromos do escopo inicial para definição dos aeródromos de embarque e

de desembarque para os deslocamentos classificados no Modo 1, que incluía aeródromos públicos e privados, inclusive helipontos, foram atribuídos volumes elevados de deslocamentos para fluxos entre aeródromos que não operaram voos comerciais em 2017, de acordo com os dados da base estatística da ANAC (2019a).

#### **1.3.3.3. Versão 3**

Foi sugerida, para a Versão 3, a mudança na seleção de aeroportos considerados para a definição como de embarque e de desembarque dos deslocamentos do Modo 1, sendo mantidos somente os aeródromos que apresentaram movimentação de passageiros da aviação comercial (regular e não regular) no ano de referência.

A partir das análises dessa versão, entregue em 28 de maio de 2019, constatou-se que classificar os deslocamentos no Modo 1, quando a velocidade média do deslocamento fosse superior a 200 km/h, seria inconsistente para viagens de curtas distâncias.

Com relação ao Modo 0, foram feitas análises levando-se em consideração a hierarquia urbana brasileira (IBGE, 2007) e os arranjos populacionais (IBGE, 2016), conforme apresentado na Seção 1.3.1, analisando-se a população das UTPs e o número de viagens entre estas, a fim de calcular a taxa de viagem/habitante entre os pares OD. Os resultados indicaram que o número de viagens dos pares da base da Telefônica (2018b) mostrava-se coerente.

#### **1.3.3.4. Versão 4 e Versão 5**

Com o objetivo da correta classificação do modo de transporte para deslocamentos de curtas distâncias, foram sugeridas duas novas extrações (Versão 4 e Versão 5) com alterações distintas nos parâmetros de consulta para avaliar a sensibilidade à cada uma.

Na Versão 4, o ajuste consistiu na incorporação de um critério: para os deslocamentos classificados no Modo 1, quando da identificação de registros intermediários entre a origem e o destino, apontando um deslocamento terrestre, os deslocamentos deveriam ser reclassificados para o Modo 0.

Para a Versão 5, foi incorporado o seguinte critério: para que viagens com 300 km de distância, consideradas de curta distância, fossem classificadas no Modo 1, um aeródromo deveria ser identificado em um raio de até 15 km do centroide da área de cobertura da célula na detecção do usuário na origem ou no destino. Caso não fosse identificado aeródromo nesse

raio, o deslocamento seria classificado para a análise da necessidade de revalidação através da nova classificação para o modo de transporte, o Modo 2, conforme descrito no item 1.1.6.1.

Em relação à avaliação do número absoluto de viagens por par OD, verificou-se que, até a Versão 5, entregue em 18 de junho de 2019, o principal fluxo de viagens (ou seja, com o maior volume de deslocamentos) ocorria entre os municípios de Borba (AM) e Pacaraima (RR). Era esperado que os principais fluxos ocorressem entre grandes metrópoles. Assim, solicitou-se que a empresa contratada, que dispõe dos dados brutos, buscasse analisar esse trecho a fim de averiguar a razão desse resultado.

### **1.3.3.5. Versão 6**

Para a Versão 6, foi utilizado o critério da Versão 5, que se mostrou mais consistente. No entanto, decidiu-se considerar como viagens de curta distância aquelas com distância inferior a 500 km e manter o raio para identificação de aeródromo próximo ao usuário na origem ou no destino de 15 km. Além disso, para classificação do modo de transporte das viagens de curta distância, a velocidade média de referência foi alterada de 200 km/h para 120 km/h.

A partir das alterações nas regras para a Versão 6, o principal fluxo de viagens passou a ser Rio de Janeiro (RJ) a São Paulo (SP), sendo esse resultado mais coerente com o esperado.

Analisando-se a Versão 6, no que concerne às áreas de influência dos municípios de origem, algumas não apresentavam viagens com origem nos municípios mais populosos de seu entorno. Por exemplo, para a área de influência do Aeroporto de Confins (MG), era esperado que o município com maior número de viagens fosse Belo Horizonte (MG), já que sua população é de 2,5 milhões de pessoas (IBGE, c2017a). Entretanto, na Versão 6, os dados mostravam que a origem com mais viagens era Confins (MG), onde a população residente é de 6,7 mil pessoas (IBGE, c2017b). Solicitou-se, então, reavaliação da base e refinamento desse aspecto nas regras de extração.

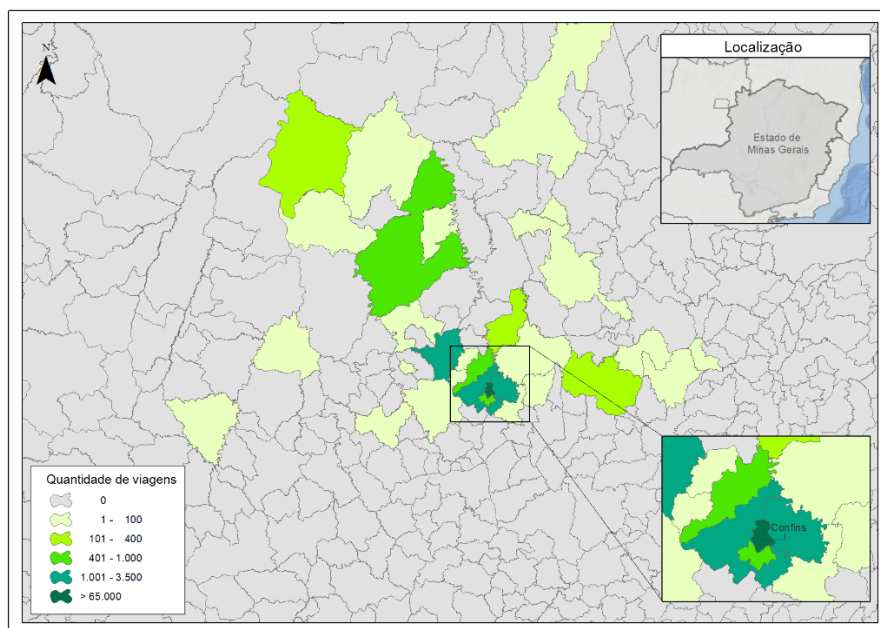
Quanto às viagens do Modo 0, foram utilizados os já mencionados dados do IBGE quanto aos arranjos populacionais e à hierarquia urbana. Além disso, foram incluídas novas análises com dados operacionais de transporte rodoviário interestadual (ANTT, [201-]). Conclui-se que, diante da amostra analisada, o Extrato V6 continuava coerente quanto às viagens não aéreas entre UTPs.

### 1.3.3.6. Versão 7

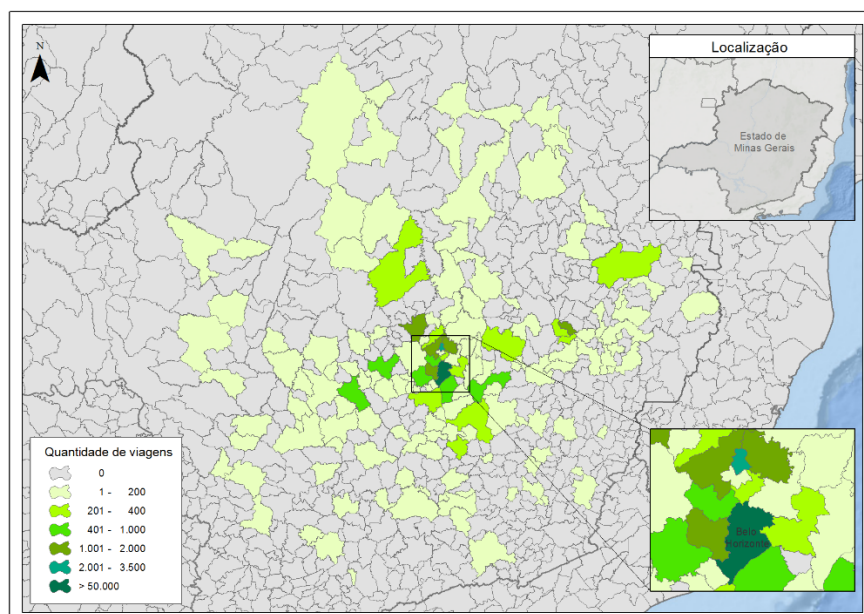
Após refinamentos nas regras de extração, como reordenação dos dados e busca de um período estacionário anterior ao local do aeroporto, o problema da identificação da área de influência do aeroporto foi corrigido para a Versão 7, conforme a Figura 17.

Figura 17 – Área de influência dos municípios de origem do Aeroporto SBCF no extrato 6 e 7

Extrato 6



Extrato 7



Elaboração: LabTrans/UFSC (2020)

Constatou-se que os resultados da Versão 7 alcançaram um refinamento considerado satisfatório e, levando-se em conta que os ajustes posteriores poderiam ser feitos durante o período de suporte, solicitou-se a extração completa da base de dados.

#### **1.3.3.7. Versão 8**

Por solicitação da SAC/MInfra, após a extração completa, foi solicitada uma nova versão, com os seguintes ajustes para melhor refinamento da base de dados:

- Para ODs entre UTPs com mais de 10 milhões de habitantes, a redução do tempo de permanência/estacionário, para definição da origem e do destino reais, de seis para quatro horas.
- Para as viagens de curta distância (inferiores a 500 km), a redução da velocidade de referência para classificação do modo de transporte de 120 km/h para 100 km/h.
- A expansão de dados utilizando um fator único para o grupo de municípios de mesma UTP.

Resumidamente, os ajustes realizados em cada uma das versões da extração parcial e completa da base de dados estão apresentados na Tabela 8.

**Tabela 8 – Resumo das alterações nos parâmetros para as oito versões do extrato parcial**

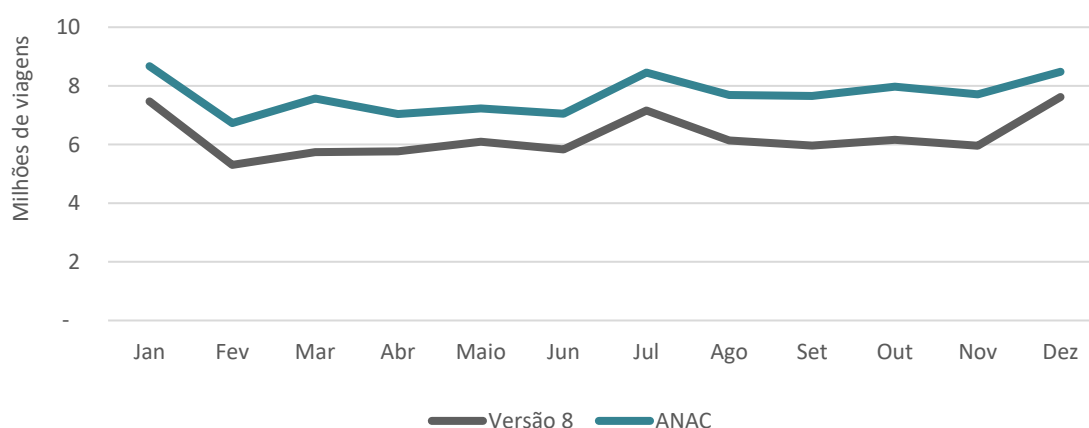
Parâmetro	Versão 1	Versão 2	Versão 3	Versão 4	Versão 5	Versão 6	Versão 7	Versão 8
Tempo estacionário para definição de UTP de origem e de UTP de destino	30 minutos para o Modo 0 e seis horas para o Modo 1	Seis horas para todos os modos						Quatro horas para par OD com mais de 10 milhões de habitantes e seis horas para o restante
Escopo de aeródromos para definição dos aeródromos de embarque e de desembarque	Todos constantes na base de dados do edital		Somente aeródromos com movimentação de passageiros da aviação comercial (regular e não regular) no ano de 2017					
Velocidade média de referência para classificação do modo de transporte	200 km/h					120 km/h, se distância < 500 km; senão 200 km/h		100 km/h, se distância < 500 km; senão 200 km/h
Identificação de registro intermediário	Não			Sim	Não			
Reclassificação dos deslocamentos com mesmo aeroporto de embarque e de desembarque	Não			Sim	Não			
Análise de proximidade ao aeródromo (raio de 15 km)	Não				Sim, se distância < 300 km	Sim, se distância < 500 km		
Expansão	Fator de expansão por usuário, sem variável renda	Fator de expansão por usuário, com variável renda						Média do fator de expansão por UTP

Elaboração: LabTrans/UFSC (2020)

### 1.3.4. Análise do extrato completo

Com a extração completa da base de dados para o ano de 2017, entregue em 16 de setembro de 2019, também foi avaliada, para os pares classificados no Modo 1, a sazonalidade das movimentações ao longo de 2017 em comparação com a dos dados da base estatística da ANAC (2019a). Pode-se observar que a base de dados fornecida pela Telefônica (2018a) apresenta um comportamento semelhante ao dos dados da ANAC no que se refere às variações mensais nas movimentações, conforme ilustrado no Gráfico 3.

**Gráfico 3 – Comparativo de movimentações mensais de 2017 do extrato com os dados da base estatística da ANAC**



Fonte: ANAC (2019a). Elaboração: LabTrans/UFSC (2020)

Além da avaliação da sazonalidade dos deslocamentos agregados por mês, os dados da Telefônica foram comparados com a base da ANAC em relação às variações mensais dos embarques e desembarques por aeródromo. De maneira geral, os resultados mostraram compatibilidade entre as duas fontes. A correlação<sup>10</sup> das variações mensais ponderadas pelo número de viagens por aeroporto foi de 0,7 para ambos os sentidos, indicando uma forte correlação.

No que se refere aos deslocamentos classificados no Modo 0, a primeira análise consistiu em uma avaliação considerando-se taxas de viagem em relação à população (IBGE, 2019) e o conceito de Hierarquia Urbana do IBGE, que divide as regiões brasileiras em Grande Metrôpole Nacional (GMN), Metrôpoles Nacionais (MN), Metrôpoles (M), Capitais Regionais (CR), Centros Sub-Regionais (CSB), Centros de Zona (CZ) e Centros Locais (CL). Além das classes do IBGE, foram criadas outras seis classes: Macrometrôpole Paulista, Microrregião de Bragança

<sup>10</sup> Utilizando-se o coeficiente de correlação linear de Pearson.



Paulista, Microrregião de Sorocaba, Polo Industrial, Região Serrana Fluminense, Região Integrada de Desenvolvimento Econômico (RIDE). Por fim, foram comparados os fluxos de viagens do extrato com o número total de viagens em grupos, informados na Tabela 9.

**Tabela 9 – Grupos de análise**

Análise de viagens terrestres por grupo
Grupo geral – Todos os fluxos
Grupo 1 – Grande Metrópole Nacional x Metrópoles Nacionais/Metrópoles Nacionais x Metrópoles Nacionais
Grupo 2 – Grande Metrópole Nacional/Metrópoles Nacionais x Metrópoles/RIDE/Microrregiões/Macrometrópole Paulista/Região Serrana Fluminense
Grupo 3 – Metrópoles x Metrópoles
Grupo 4 – Grande Metrópole Nacional / Metrópoles Nacionais x Capitais Regionais
Grupo 5 – Metrópoles x Capitais Regionais/Polos Industriais/RIDE/Microrregiões/Macrometrópole Paulista/Região Serrana Fluminense
Grupo 6 – Capitais Regionais A x Capitais Regionais A
Grupo 7 – Capitais Regionais A/B/C x Capitais Regionais B/C /Microrregiões/Macrometrópole Paulista/Região Serrana Fluminense/Polos Industriais/RIDE
Grupo 8 – Grande Metrópole Nacional x UTPs NCA*
Grupo 9 – Metrópole Nacional (Rio de Janeiro) x UTPs NCA*
Grupo 10 – Metrópole Nacional (Brasília) x UTPs NCA*
Grupo 11 – Metrópoles x UTPs NCA*
Grupo 12 – Centros Sub-regionais x Centros Sub-regionais
Grupo 13 – Centros Sub-regionais x Centros de Zona
Grupo 14 – Centros Sub-regionais x Centros Locais
Grupo 15 – Centros de Zona x Centros de Zona
Grupo 16 – Centros de Zona x Centros Locais
Grupo 17 – Centros Locais x Centros Locais

\*NCA: não contempladas nos grupos anteriores, representados pelos Centros Sub-regionais, de Zona e Locais, além de localidades sem classe definida.

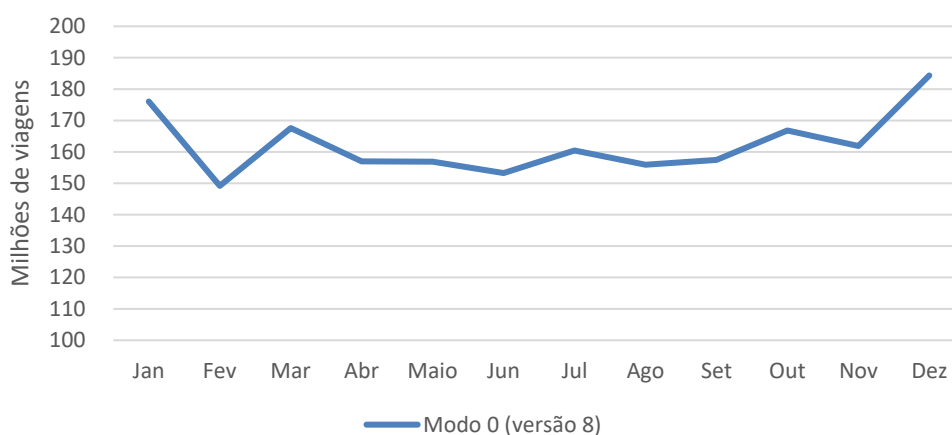
Elaboração: LabTrans/UFSC (2020)

As análises consistiram em identificar *outliers* pelas taxas de viagem/população do par e do número de viagens, pois, para hierarquias semelhantes, espera-se que estas taxas e o volume de viagens tenham ordens de grandeza semelhantes, diminuindo conforme as distâncias aumentam. Ao identificar *outliers*, os pares eram então separados para uma análise mais detalhada, buscando entender se existiam características que potencializassem seu número de viagens disperso (para mais ou para menos), como posição geográfica estratégica, turismo ou outras atividades. Dessa forma, aqueles pares não destacados podem ser caracterizados como válidos, fazendo parte da maioria dos casos, ou tendo um baixo nível de dispersão por seu volume ou taxa de viagens. Foi verificado que as maiores taxas se encontram em pares com

maior população e/ou menores distâncias e/ou em viagens envolvendo os grandes centros urbanos, identificando que, ante esta análise, a base encontra-se consistente.

Para verificar a consistência dos dados também foi realizado um processo comparativo entre os volumes totais de viagem para cada grupo ao longo dos meses. O objetivo foi verificar se a ordem de grandeza do número de viagens para cada mês apresenta compatibilidade, esperando, evidentemente, variações causadas (em sua maioria) por fatores de sazonalidade. Com relação a essa análise, as viagens apresentam-se bastante consistentes, destacando-se a sazonalidade já esperada, promovendo um maior número de viagens nos meses de dezembro e janeiro para a grande maioria dos grupos.

Gráfico 4 – Sazonalidade modo terrestre



Elaboração: LabTrans/UFSC (2020)

A segunda análise se refere aos deslocamentos pendulares comparando os pares registrados no Extrato Telefônica com a base de dados de arranjos populacionais do IBGE (2016). Essa comparação não se atentou propriamente ao volume de viagens, mas sim ao seu *ranking*.

Assim, os pares foram classificados por número de viagens e, então, separados. Depois, os pares foram classificados por número de deslocamentos pendulares e, por fim, foram comparados aos pares classificados por volume de viagens, buscando identificar correspondências entre os dois filtros e as classificações.

Deve-se citar que alguns pares dentro de arranjos populacionais presentes na base do IBGE encontram-se dentro da mesma UTP e, devido a isso, não apresentaram correspondência na base de dados, que só considera viagens entre UTPs distintas, conforme destacado anteriormente.

A terceira análise foi realizada de maneira análoga à anterior, só que considerando os dados operacionais da ANTT, de transporte rodoviário. Nesse caso, apenas deve-se destacar que os dados operacionais se referem basicamente ao transporte interestadual e semiurbano, ou seja, deslocamentos entre UTPs com características diferentes dessas não foram contemplados na análise, devido à inexistência de dados correspondentes. O objetivo foi comparar a classificação no número de viagens, mas, principalmente, a presença de pares da base da ANTT nos registros do extrato entregue. Grandes volumes de deslocamento na base da ANTT, quando entre UTPs distintas, ausentes no extrato, geram a necessidade de investigação, apontando um possível erro no extrato, assim como comportamentos com baixíssimo número de passageiros na base ANTT e forte presença no Extrato, o quais também podem indicar problemas.

Em suma, essas análises apontam para uma boa consistência nos dados classificados como Modo 0, aparentando refletir a realidade dos deslocamentos dentro do território brasileiro. Os pares com problemas apresentaram baixa frequência, podendo ser ajustados posteriormente ou descartados, viabilizando a continuidade e o aprofundamento das análises subsequentes.

## **1.4. Expansão da matriz**

De acordo com a Telefônica (2018a), ainda que municípios de economia e/ou população relevantes tenham cobertura de telefonia móvel, não foi realizado nenhum processo de expansão pela empresa para inferir os deslocamentos que tenham como origem ou destino regiões sem cobertura.

Assim, faz-se necessário realizar uma expansão da matriz a fim de estimar os dados de deslocamentos faltantes. Deve-se notar que esse procedimento de expansão diz respeito a uma etapa posterior à estimativa da matriz OD, aplicado, seguindo um método específico, ao modal terrestre e ao modal aéreo.

A matriz de viagens é expandida por duas razões. A primeira diz respeito à identificação de novos pares OD não considerados na matriz de viagens fornecida. A segunda, devido à baixa cobertura da Telefônica nas regiões de origem e/ou de destino da viagem.

Para identificação dos pares OD não contemplados pela extração, para o modal aéreo, foi realizada uma comparação com os Microdados de Tarifas Aéreas Comercializadas (ANAC, 2019b) e a base de dados de viagens aéreas (EPL, 2014). Já em relação ao modo terrestre, foi utilizada para comparação a Base de Dados Operacionais do Transporte Rodoviário de passageiros (ANTT, [201-]) e dos dados do Plano Aeroviário Nacional (BRASIL, 2018b).

A partir disso, para o modo 0, alguns critérios foram adotados para seleção desses pares, com o intuito de manter as características dos fluxos de transportes, apresentados pelo *big data*, nos resultados expandidos. Esses critérios foram elaborados a partir de análises feitas na base de dados da telefônica. O primeiro critério foi de selecionar pares com menos de 12 horas<sup>11</sup> de distância entre a origem e o destino. Considerando a Hierarquia Urbana do IBGE do município sede de cada UTP, o segundo critério utilizado foi de selecionar pares OD com tempo de deslocamento de 12 a 50 horas desde que as classificações sejam Metrópoles, Capital Regional A, B e C ou fluxos localizados na região Norte do país. Para os fluxos de viagem com mais de 50 horas de deslocamento, foram selecionados apenas os localizados na região Norte.

No que se refere aos pares com baixa cobertura, para o modo aéreo, foi utilizado como critério para a seleção a inexistência de cobertura no município de origem e/ou no município de destino. Em relação ao modal terrestre, considera-se baixa cobertura quando pelo menos 50% dos municípios da UTP de origem ou de destino não possuem cobertura em telefonia móvel.

Foram identificados, para expansão e agregação da Matriz, 12.356 pares OD do Modo 1 (aéreo), considerando-se as comparações entre a Base de Dados entregue e as bases supracitadas. Já em relação ao Modo 0, foram identificados, para expansão e agregação da Matriz, 32.638 pares OD.

Uma diferença entre a expansão da matriz OD dos modais aéreo e terrestre é que, no primeiro caso, os dados e resultados de número de viagens estão por município, enquanto que, no segundo caso, estão por UTP. Essa diferença de abrangência é uma condição do contrato com a fornecedora dos dados.

O modelo de expansão da matriz OD, tanto para o modal terrestre como aéreo, é expresso pela Equação (1). Nota-se que a estrutura matemática do modelo indica que é um modelo de corte transversal (do inglês – *cross section*), em que cada observação da matriz de dados refere-se a um par OD (modo aéreo ou terrestre, cada um constituindo uma matriz de dados). A estimativa dos parâmetros para a amostra total (do aéreo ou do terrestre), daqueles pares OD confirmados como válidos, permite que se obtenha uma projeção para qualquer par OD arbitrário (não constante da amostra). Esse procedimento é fundamentalmente a expansão da matriz de viagens para ODs não contempladas originalmente pela matriz da Telefônica.

---

<sup>11</sup> Essa definição levou em consideração a Lei 13.103/2015, da qual ficou deliberado que a jornada de trabalho de um motorista poderia chegar a, no máximo, 12 horas.

$$NV_{ij} = i\alpha_0 + \alpha_1 G_{ij} + \alpha_2 CP_{ij} + e_{ij} \quad (1)$$

Onde:

$NV_{ij}$ : é o número de viagens no modal aéreo ou terrestre da origem  $i$  ao destino  $j$ , com dimensão  $n \times 1$

$G_{ij}$ : é uma matriz de variáveis gravitacionais, que inclui PIB da origem  $i$  e do destino  $j$ , população de origem e de destino e a distância entre  $i$  e  $j$ , com dimensão  $n \times k_1$ , onde  $k_1$  é o número de variáveis gravitacionais

$CP_{ij}$ : é uma matriz de componentes principais dada a partir de um conjunto relativamente grande de variáveis socioeconômicas e logísticas que capturam a densidade econômica e a intensidade de tráfego nas áreas de influência da origem  $i$  e do destino  $j$ , com dimensão  $n \times k_2$ , onde  $k_2$  é o número de componentes principais

$i$ : é um vetor de 1

$\alpha_0$ : é um termo constante,  $\alpha_1$  e  $\alpha_2$  são vetores de parâmetros e

$e_{ij}$ : é um termo estocástico.

Além das variáveis convencionais de natureza gravitacional (expressas na matriz  $G$ ), considera-se um conjunto de outros determinantes – por exemplo, de natureza socioeconômica e de infraestrutura – do número de viagens entre a origem  $i$  e o destino  $j$ . Como essa amostra de determinantes representa um grande número de variáveis explicativas, a estimativa dessas variáveis isoladamente pode resultar em uma significativa perda de graus de liberdade do modelo. E também, muitas dessas variáveis explicativas devem medir impactos semelhantes e, portanto, são muito correlacionadas. Nesse sentido, procede-se a aplicação de Análise de Componentes Principais (ACP), que é uma ferramenta padrão do sistema de análise de dados e que tem como vantagem a concentração das informações do conjunto original de variáveis em componentes principais, reduzindo-se o número de variáveis (SABIN; FERRÃO; FURTADO, 2004). Cada componente principal é uma combinação linear de todas as variáveis originais e sua quantidade é, geralmente, indicada pelo número de critérios utilizados (ZOU; HASTIE; TIBSHIRANI, 2006).

A utilização de componentes principais permite a extração das informações mais relevantes de uma amostra de dados, as quais são expressas em um conjunto de variáveis ortogonais não correlacionadas, os Componentes Principais. Nesse sentido, a análise utiliza autovalores e autovetores da matriz de covariância (DJUKIC; VAN LINT; HOOGENDOORN, 2012). Na área de transportes, uma das aplicações mais usuais de ACP é para o dimensionamento na previsão de contagens de tráfego (SUN et al., 2009).

Em termos práticos, a análise de componentes principais foi aplicada a um conjunto de 34 variáveis independentes (além das três variáveis identificadas como variáveis gravitacionais). A abordagem de ACP estima, então, os componentes principais que são estatisticamente significantes. Se grande parte da variância das variáveis independentes puder ser atribuída ao primeiro ou aos primeiros componentes relevantes, então, esses componentes podem substituir as variáveis originais sem significativa perda de informação.

Em termos matemáticos, suponha que as variáveis originais (chamadas de  $X$ ) – de natureza socioeconômica, infraestrutura e outras – sejam expressas por um vetor aleatório e que a matriz de covariância, denominada por  $V$ , seja composta por autovalores (normalizados) e autovetores. Os  $m$  componentes principais, extraídos das variáveis originais, são obtidos como médias ponderadas, de acordo com a Equação (2).

$$cp_m = V_m X = V_1 X_1 + V_2 X_2 + \dots + V_{44} X_{44} \quad (2)$$

$$Cov(cp_m, cp_{m+q}) = v'_m V v_m = \sigma_m \text{ quando } q = 0 \text{ e } = 0 \text{ quando } q \neq 0$$

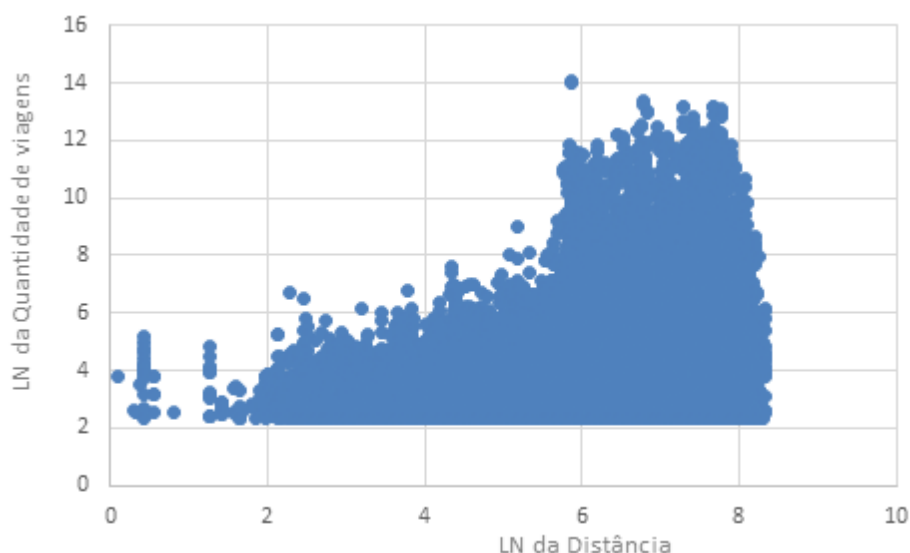
Assim, como descrito na equação acima, os componentes principais são não correlacionados entre si e são elencados em ordem decrescente conforme suas variâncias. Além disso, o número máximo de componentes corresponde ao número de variáveis independentes. O segundo passo é a estimativa da Equação (1), que estabelece os determinantes para gerar o número de viagens por par OD como função de variáveis gravitacionais e dos componentes principais. Foram considerados para inclusão na Equação (1) aqueles componentes principais cuja covariância dos componentes com as variáveis originais acumuladas alcançasse, ao menos, 95%.

### 1.4.1. Descrição das variáveis utilizadas

A partir da Equação (1), descrita na seção 1.4, as variáveis determinantes do número de viagens no modal aéreo ou terrestre são de natureza gravitacional, socioeconômica e de infraestrutura.

Mais especificamente para o componente gravitacional da equação, foram utilizadas as variáveis Produto Interno Bruto (PIB) e população, provenientes do IBGE referente ao ano de 2017, e a distância entre o município de origem e de destino, calculadas através do *software* de georreferenciamento Hórus Desktop<sup>12</sup>. De maneira ilustrativa, o Gráfico 5, Gráfico 6 e Gráfico 7 apresentam as relações entre o número de viagens extraído da base de dados da Telefônica com a distância entre os pares OD, população e PIB, para o Modo 1 (Aéreo). Ainda sobre a distância, observa-se que a relação desta variável com a quantidade de viagens é direta, o que é diferente do esperado pelo argumento tradicional do modelo gravitacional. Isso ocorre devido à característica do modal aéreo, que é mais utilizado quando a distância entre a origem e destino são maiores. Deve-se notar que a relação sugerida pelas figuras abaixo é bidimensional; enquanto que a verdadeira relação entre as variáveis é estimada em um contexto multidimensional – isto é, considerando-se a influência de diversas variáveis simultaneamente – e apresentada em seção posterior.

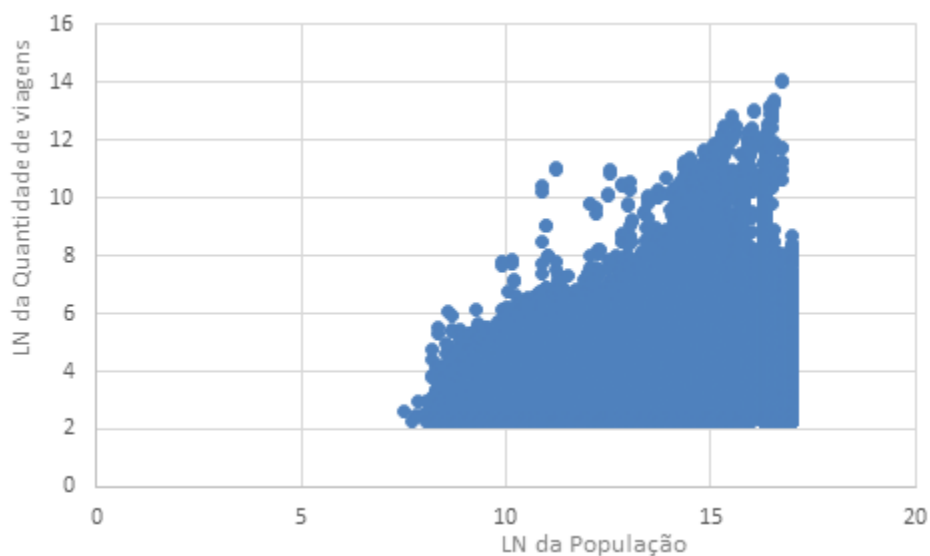
Gráfico 5– Relação entre quantidade de viagens X distância (variáveis em logaritmo natural) – Modo 1 (aéreo)



Fonte: Telefônica (2018a)). Elaboração: LabTrans/UFSC (2020)

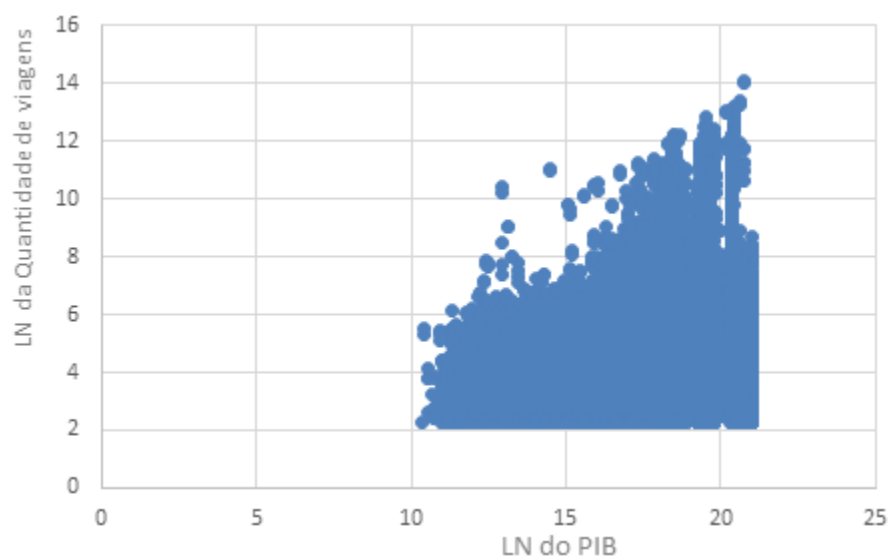
<sup>12</sup> O Hórus Desktop é um *software* desenvolvido pelo LabTrans/UFSC com camadas estruturadas em banco de dados, com revisões e atualizações recentes, e é voltado a aplicações para o transporte aéreo.

**Gráfico 6— Relação entre quantidade de viagens X população(variáveis em logaritmo natural) – Modo 1 (aéreo)**



Fonte: Telefônica (2018a) e IBGE (2019). Elaboração: LabTrans/UFSC (2020)

**Gráfico 7 – Relação entre quantidade de viagens X PIB (variáveis em logaritmo natural) – Modo 1 (aéreo)**

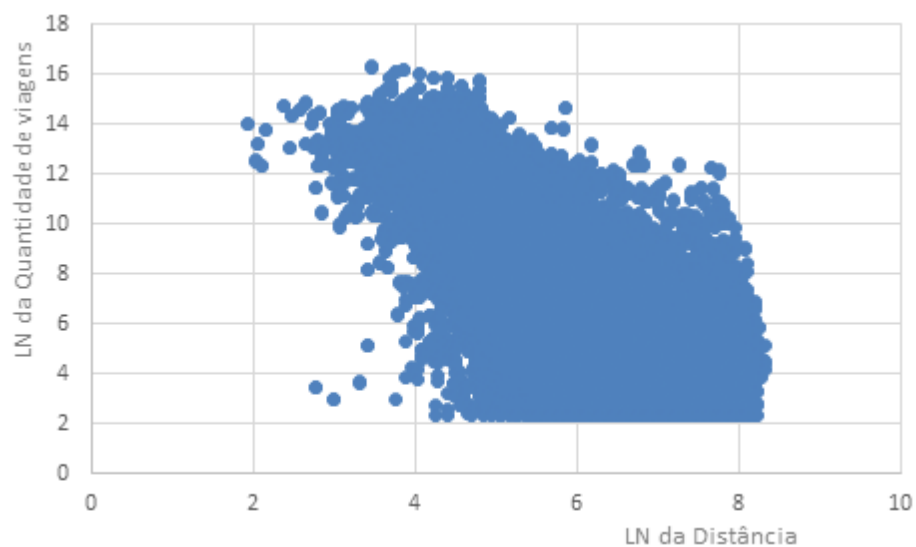


Fonte: Telefônica (2018a) e IBGE (2017). Elaboração: LabTrans/UFSC (2020)

Considerando-se essa mesma relação entre as variáveis gravitacionais com a quantidade de viagens para o Modo 0 (terrestre), pode-se observar que a relação com as viagens é inversa para a distância e direta para o PIB e população, em conformidade com o argumento gravitacional. O Gráfico 8, o Gráfico 9 e o Gráfico 10 ilustram essas relações.

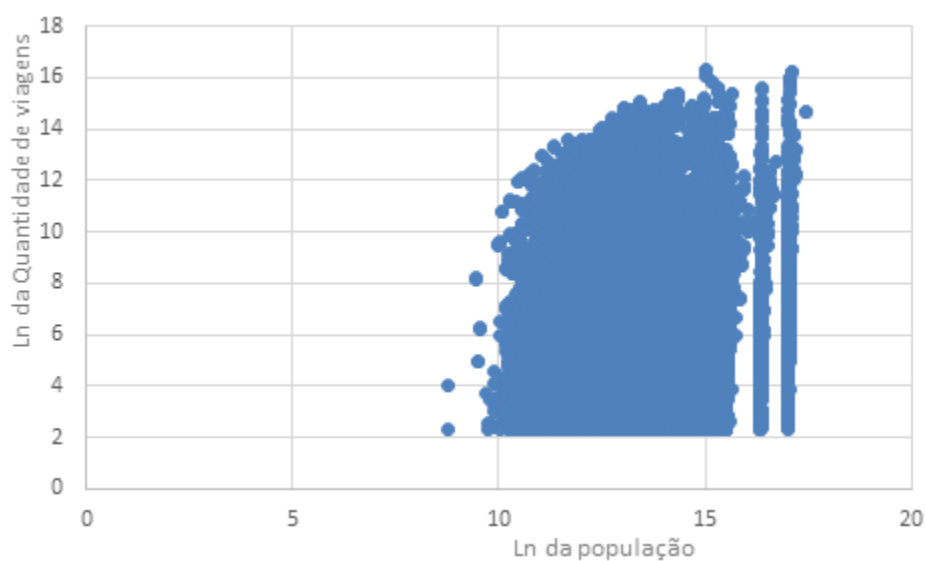


**Gráfico 8 – Relação entre quantidade de viagens X distância (variáveis em logaritmo natural) – Modo 0 (terrestre)**



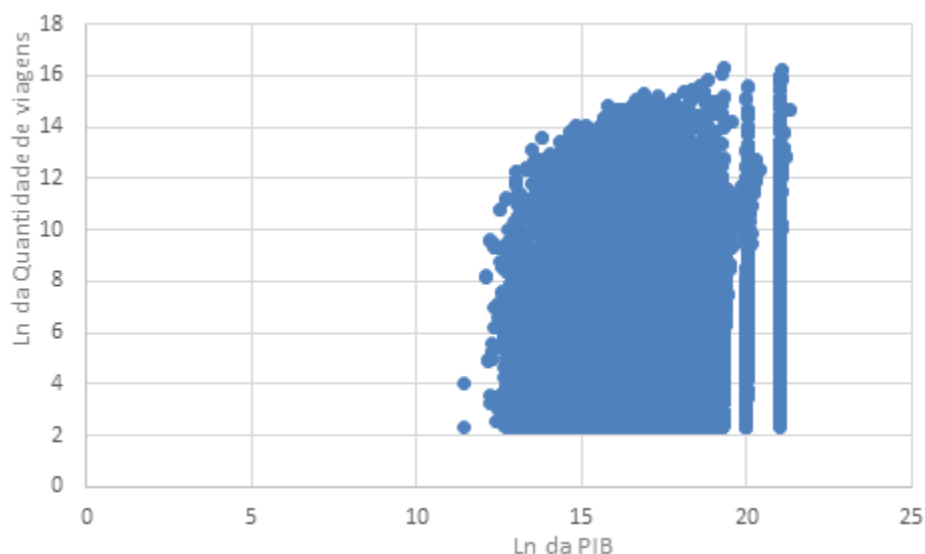
Fonte: Telefônica (2018a). Elaboração: LabTrans/UFSC (2020)

**Gráfico 9 – Relação entre quantidade de viagens X população (variáveis em logaritmo natural) – Modo 0 (terrestre)**



Fonte: Telefônica (2018a) e IBGE (2019). Elaboração: LabTrans/UFSC (2020)

**Gráfico 10 – Relação entre quantidade de viagens X PIB (variáveis em logaritmo natural) – Modo 0 (terrestre)**



Fonte: Telefônica (2018a) e IBGE (2017). Elaboração: LabTrans/UFSC (2020)

Em relação às variáveis socioeconômicas e de infraestrutura, foram, ao todo, utilizadas 34 variáveis com informações referentes aos municípios de origem e destino. Na Tabela 10 são apresentadas as variáveis consideradas.

**Tabela 10 – Variáveis utilizadas a para a estimativa do modelo**

Código	Município/ UTP	Variável	Fonte
v1	Origem	Total de Matrículas no Ensino Médio	Censo Escolar – 2017 (INEP, 2019b)
v2		Total de Matrículas em Educação Profissional	
v3		Total de Matrículas no Ensino Superior	Censo da Educação Superior – 2017 (INEP, 2019a)
v4		Número no IES	
v5		Categorização Turismo	Anuário Estatístico de Turismo 2018 – Ano Base 2017 (BRASIL, 2018c), Categorização dos municípios das regiões turísticas do mapa do turismo brasileiro (BRASIL, 2015), Estudo da Demanda Turística Internacional 2017 (BRASIL, 2018d)
v6		Quantidade de empregos no setor de Hospedagem	
v7		Demanda por turismo internacional	
v8		Demanda por turismo doméstico	
v9		Residentes a menos de 10 anos no município	Censo 2010 - (IBGE, 2010)
v10		Trabalha ou estuda em outro município	
v11		Residentes não naturais do município	
v12		Índice de Integração Nacional	Dados fornecidos pela SAC/MInfra (2020)
v13		Remuneração média	Relação Anual de Informações Sociais (RAIS) (BRASIL, [2019])
v14		Empregados no setor de Turismo	
v15		Empregados na Administração Pública	
v16		Empregados no setor de Saúde	
v17		Demais empregados	

Código	Município/ UTP	Variável	Fonte
v18	Destino	Total de Matrículas no Ensino Médio	Censo Escolar – 2017 (INEP, 2019b)
v19		Total de Matrículas em Educação Profissional	
v20		Total de Matrículas no Ensino Superior	Censo da Educação Superior – 2017 (INEP, 2019a)
v21		Número no IES	
v22		Categorização Turismo	Anuário Estatístico de Turismo 2018 – Ano Base 2017 (BRASIL, 2018c), Categorização dos municípios das regiões turísticas do mapa do turismo brasileiro (BRASIL, 2015), Estudo da Demanda Turística Internacional 2017 (BRASIL, 2018d)
v23		Quantidade de empregos no setor de Hospedagem	
v24		Demanda por turismo internacional	
v25		Demanda por turismo doméstico	
v26		Residentes a menos de 10 anos no município	Censo 2010 - (IBGE, 2010)
v27		Trabalha ou estuda em outro município	
v28		Residentes não naturais do município	
v29		Índice de Integração Nacional	Dados fornecidos pela SAC/MInfra (2020)
v30		Remuneração média	Relação Anual de Informações Sociais (RAIS) (BRASIL, [2019])
v31		Empregados no setor de Turismo	
v32		Empregados na Administração Pública	
v33		Empregados no setor de Saúde	
v34		Demais empregados	

Elaboração: LabTrans/UFSC (2020)

Salienta-se que, para as estimativas, as variáveis foram agrupadas por município para o modo aéreo e por UTP por modo terrestre, conforme estrutura dos dados disponibilizados pela Telefônica.

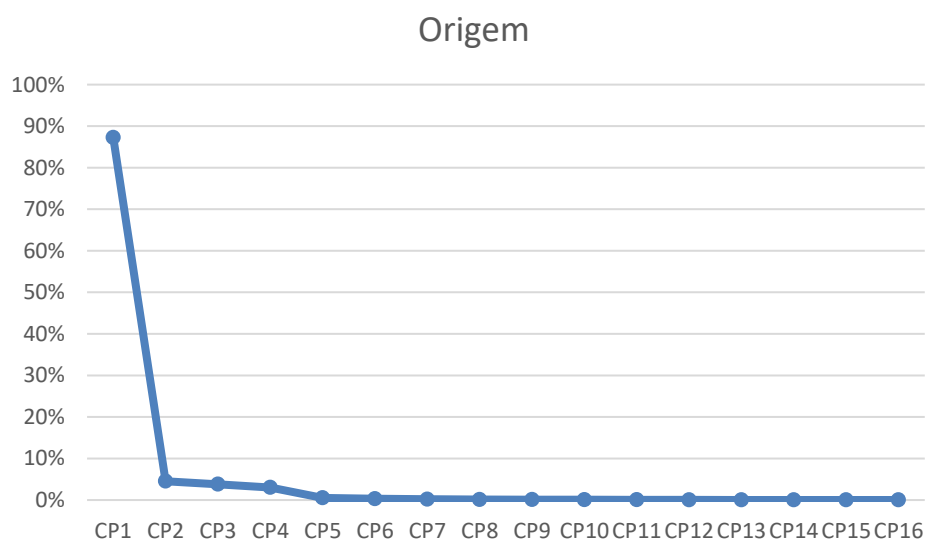
Conforme citado anteriormente, a utilização de um grande número de variáveis explicativas pode resultar em uma significativa perda de graus de liberdade, além de algumas medirem o mesmo impacto no número de viagens, ou seja, são muito correlacionadas. Dessa forma, foi aplicado o método de Análises dos Componentes Principais, conforme exposto na seção a seguir.

### 1.4.2. Análise dos Componentes Principais

Em relação à Análise dos Componentes Principais, a primeira etapa consiste em avaliar a correlação entre as variáveis. Foram elaboradas duas matrizes de correlação. Observou-se que as variáveis que continham informações da origem são fortemente correlacionadas, para o modo aéreo e para o modo terrestre. Da mesma forma, para as variáveis referentes ao destino, os resultados indicam forte correlação. Portanto foram realizadas duas análises dos componentes principais para cada modo.

De acordo com a abordagem dos componentes principais, para o modo aéreo, foram estimados 17 componentes para as variáveis que continham informações sobre os municípios de origem e o mesmo número para aquelas do destino. O Gráfico 11 indica que o primeiro componente principal corresponde a 87,3% da variância total dos dados de origem utilizados na análise, ao passo que os dois primeiros componentes principais juntos são equivalentes a 91,8%.

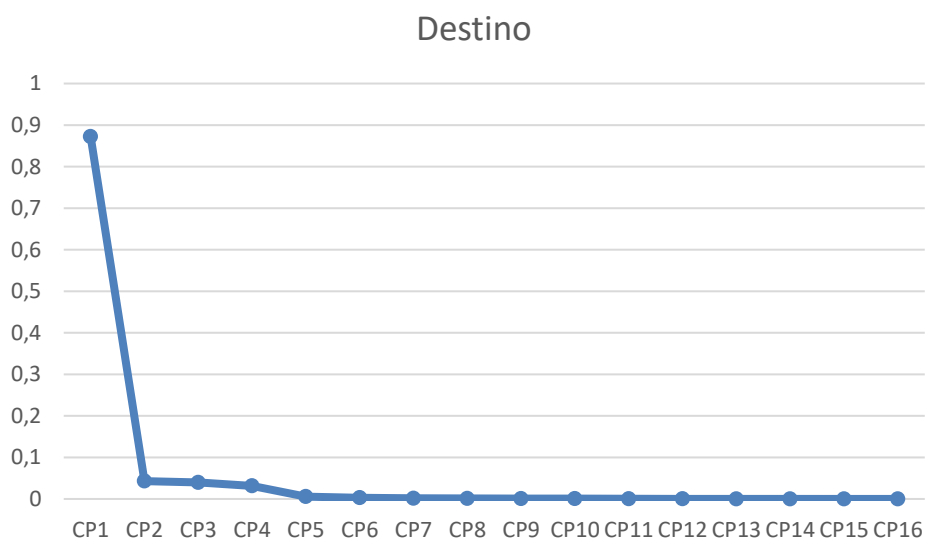
**Gráfico 11 – Participação dos componentes principais na variância total dos dados – Origem – Modo 1 (aéreo)**



Elaboração: LabTrans/UFSC (2020)

Em relação à análise realizada a partir dos dados com informações dos municípios de destino, o primeiro componente principal estimado corresponde a 87,2% e, em conjunto com o segundo componente, representa 91,5%. O Gráfico 12 ilustra as participações dos componentes estimados.

**Gráfico 12 – Participação dos componentes principais na variância total dos dados – Destino – Modo 1 (aéreo)**

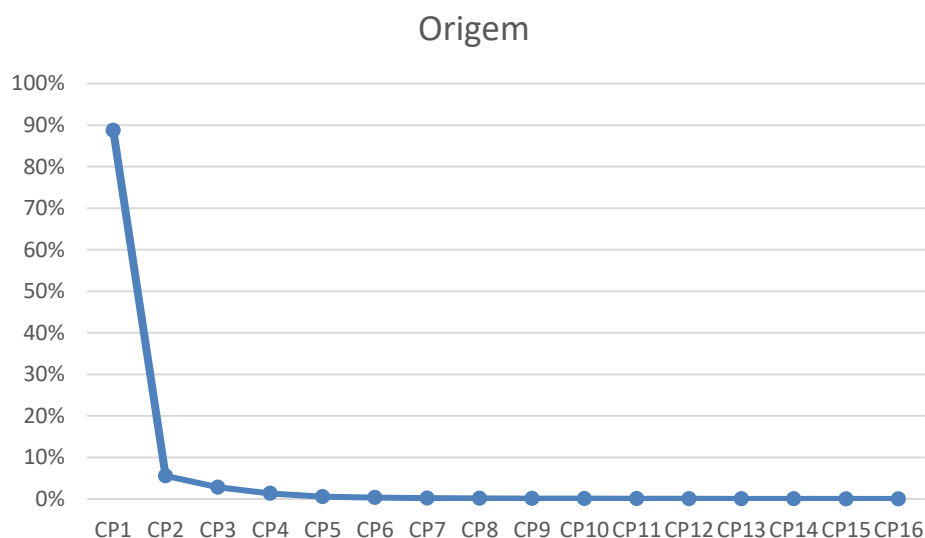


Elaboração: LabTrans/UFSC (2020)

Dessa forma, para a estimativa do modelo de expansão para o Modo 1 (aéreo), foram utilizados os três componentes principais da origem e os três componentes principais do destino, considerando-se que, no total, respondem, em ambos os casos, por pelo menos 95% da variância total dos dados considerados.

No que se refere à abordagem de Análise dos Componentes Principais para as variáveis com informações das origens do Modo 0 (terrestre), o primeiro componente principal estimado equivale a 88,9% e, em conjunto com o segundo componente, corresponde a 94,3%. O Gráfico 13 ilustra as participações dos componentes estimados.

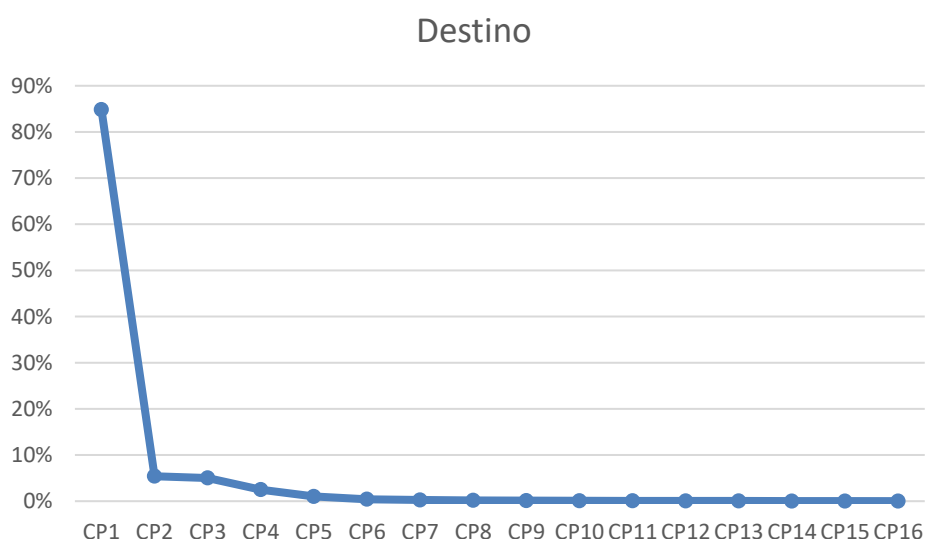
**Gráfico 13 – Participação dos componentes principais na variância total dos dados – Origem – Modo 0 (terrestre)**



Elaboração: LabTrans/UFSC (2020)

No que concerne à análise para as variáveis referentes às informações dos destinos, os resultados indicam que o primeiro componente responde por 84,8% da variância total dos dados utilizados. O segundo componente equivale a 5,4% da variância. O Gráfico 14 ilustra a participação de cada componente.

**Gráfico 14 – Participação dos componentes principais na variância total dos dados – Destino – Modo 0 (terrestre)**



Elaboração: LabTrans/UFSC (2020)

Superada a etapa de estimativa dos componentes principais, o modelo foi estimado, conforme as etapas descritas a seguir.

### 1.4.3. Estimativa do modelo

O modelo utilizado é de dados seccionais (*cross section*) gravitacional de  $i$  para  $j$ , com as variáveis em logaritmo. Após a realização de um teste paramétrico de Wald<sup>13</sup> foi constatado um problema de heterocedasticidade na amostra (quando a variância dos resíduos não é constante). Após a correção desse problema, o modelo proposto se mostrou estatisticamente significativo. Outro problema refere-se a alta correlação detectada na amostra em estudo (multicolinearidade) entre a população e o PIB. Os testes realizados com uma ou outra variável indicaram a superioridade (em termos de qualidade da regressão) da especificação com a população.

A especificação do modelo estimado para o Modo 1 (aéreo) foi:

$$NV_{ij} = \alpha_0 + \alpha_1 Pop_{ij} + \alpha_2 Dist_{ij} + \alpha_3 PC1_i + \alpha_4 PC2_i + \alpha_5 PC3_i + \alpha_6 PC1_j + \alpha_7 PC2_j + e_{ij} \quad (4)$$

$NV_{ij}$ : é o número de viagens no modal aéreo da origem  $i$  ao destino  $j$

$Pop_{ij}$  é a população da origem  $i$  e somada à população do destino  $j$

$Dist_{ij}$ : é a distância entre a origem  $i$  e o destino  $j$

$PC1_i, PC2_i$  e  $PC3_i$ : são os componentes principais referentes às informações da origem  $i$

$PC1_j$  e  $PC2_j$  são os componentes principais referentes às informações ao destino  $j$

$\alpha_0$ : é um termo constante,  $\alpha_1$  a  $\alpha_7$  são os parâmetros estimados

$e_{ij}$ : é um termo estocástico.

Os parâmetros estimados e as suas estatísticas calculadas estão disponíveis na Tabela 11.

<sup>13</sup> Teste usualmente encontrado na bibliografia especializada, que é utilizado para examinar as restrições impostas aos coeficientes da regressão (hipótese nula). Ele calcula a estatística de teste (WaldQui quadrado) que mede a eficiência das estimativas dos coeficientes da regressão original em satisfazer as restrições da hipótese nula. Avalia se o parâmetro é estatisticamente significativo. A estatística teste utilizada é obtida através da razão do coeficiente por seu respectivo erro padrão. Essa estatística teste tem distribuição normal, sendo seu valor comparado com valores tabulados de acordo com o nível de significância definido

**Tabela 11 – Coeficientes estimados para o Modo 1 (aéreo)**

	Coeficiente	Erro Padrão	Estatística t	p-valor	
(Intercept)	-1.48E+00	3.46E-02	-42.596	< 2e-16	***
log(Pop)	1.98E-01	6.06E-04	326.151	< 2e-16	***
Dist	7.99E-07	1.26E-08	63.32	< 2e-16	***
PC1Origem	-3.68E-05	6.57E-06	-5.6	2.15E-08	***
PC2Origem	-1.44E-04	1.41E-05	-10.161	< 2e-16	***
PC3Origem	-1.30E-04	1.80E-05	-7.207	5.73E-13	***
PC1Destino	-1.62E-07	2.65E-06	-0.061	0.95137	
PC2Destino	5.05E-05	9.30E-06	5.426	5.77E-08	***
R <sup>2</sup>	0.4585	R <sup>2</sup> ajustado	0.4585		

Elaboração: LabTrans/UFSC (2020)

Já para o Modo 0 (terrestre), o modelo estimado possui a seguinte especificação:

$$NV_{ij} = \alpha_0 + \alpha_1 Pop_{ij} + \alpha_2 Dist_{ij} + \alpha_3 PC1_i + \alpha_4 PC2_i + \alpha_5 PC1_j + \alpha_6 PC2_j + e_{ij} \quad (5)$$

$NV_{ij}$ : é o número de viagens no modal terrestre da origem  $i$  ao destino  $j$

$Pop_{ij}$  é a população da origem  $i$  e somada à população do destino  $j$

$Dist_{ij}$ : é a distância entre a origem  $i$  e o destino  $j$

$PC1_i$ , e  $PC3_i$ : são os componentes principais referentes às informações da origem  $i$

$PC1_j$  e  $PC2_j$  são os componentes principais referentes às informações ao destino  $j$

$\alpha_0$ : é um termo constante,  $\alpha_1$  a  $\alpha_7$  são os parâmetros estimados

$e_{ij}$ : é um termo estocástico.

O resultado da estimativa para o Modo 0 está disponível na Tabela 12.

**Tabela 12 – Coeficientes estimados para o Modo 0 (terrestre)**

	Coeficiente	Erro Padrão	Estatística t	p-valor	
(Intercept)	9.24E+00	3.24E-02	285.329	2.00E-16	***
log(Pop)	1.10E-01	1.11E-03	98.5	2.00E-16	***
Dist	-2.26E-04	1.46E-06	-155.009	2.00E-16	***
PC1Origem	-8.11E-03	2.04E-04	-39.707	2.00E-16	***
PC2Origem	1.83E-01	1.32E-03	138.211	2.00E-16	***
PC1Destino	-7.53E-03	2.06E-04	-36.561	2.00E-16	***
PC2Destino	2.35E-02	1.32E-03	17.773	2.00E-16	***



	Coeficiente	Erro Padrão	Estatística t	p-valor	
R <sup>2</sup>	0.9484	R <sup>2</sup> ajustado	0.9484		

Elaboração: LabTrans/UFSC (2020)

Após as estimativas, como forma de validar os resultados também foram comparados os valores estimados com a base de dados da Telefônica, considerando a hierarquia urbana do IBGE e o tempo de deslocamento entre a origem e o destino. Dessa forma, buscou-se comparar o resultado estimado com a média de viagens da base da telefônica para os pares que apresentavam as mesmas características, suavizando os resultados, quando necessário, utilizando *dummies* de escala do número de viagens.

Como resultado da expansão, são apresentados os seguintes números, de acordo com os seus respectivos modais:

- Modo 1 (aéreo): 12.452 pares e 2.011.525 viagens
- Modo 0 (terrestre): 19.348 pares e 214.838.359 viagens.

A expansão das viagens permitiu completar a base de dados com as informações de pares OD identificados que deveriam estar presentes, conforme explanado no documento. Para o Modo 1 (aéreo), essa expansão foi correspondente a um acréscimo de 5% de pares e de um aumento de 3% de viagens na matriz OD. Para o Modo 0 (terrestre), a expansão e agregação acrescentou 3% de pares e 0,03% de viagens na matriz OD. A modelagem utilizada para expansão se apresentou com resultados alinhados com o esperado, com conformidade estatística e valores condizentes com as demais bases consultadas. Os resultados foram agregados à matriz OD produzida pela Telefônica, de forma anual. Como a expansão se deu por município de origem e de destino, para a agregação do Modo 1 (aéreo) foi caracterizado como aeroporto de origem e destino, aquele localizado mais próximo ao município de origem ou destino da viagem, quando não o aeroporto do município. Para a caracterização da UTP, foi considerada de origem ou destino aquela onde o município de origem ou destino se encontra. No Modo 0 (terrestre) também foi considerada como a UTP de origem ou destino aquela em que o município de origem ou destino se encontra.



## 2. RESULTADOS GERAIS

A amostra de dados provenientes de registros de telefonia móvel trata-se da observação direta de mais de **219 milhões de viagens** e de mais de **26,5 milhões de usuários** únicos. Esses números evidenciam a vantagem do uso de *Big Data* apresentada no item 1.1.3 referente ao tamanho da amostra. Esses dados, compartilhados de forma anônima, agregados e extrapolados, geraram material primário para formação da base de dados para a Matriz Origem/Destino (OD).

A Matriz OD de reais desejos de viagens contemplados ou não pelo transporte aéreo apresenta os seguintes números gerais:

- **1.806.8423.489** viagens terrestres (modo 0 – não aéreo)
- **77.161.438** viagens aéreas
- **60.180** pares de UTPs com deslocamento terrestre, em ambos os sentidos
- **34.871** pares de UTPs com deslocamento aéreo, em ambos os sentidos
- **22.092** pares de UTPs com deslocamento terrestre e aéreo, em ambos os sentidos.

Do total de viagens observadas, **95,9%** são terrestres.

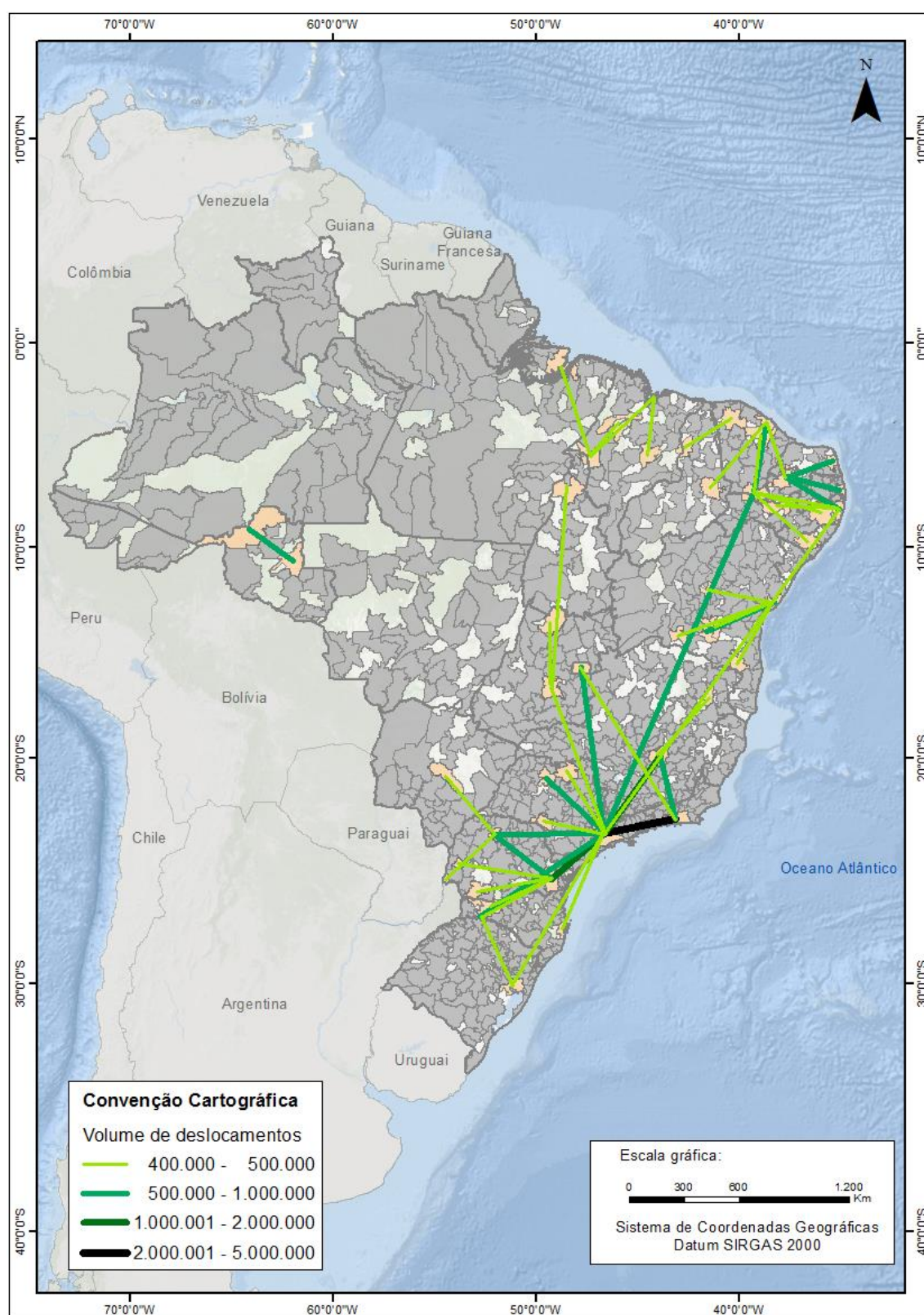
A Figura 18 e a Figura 19, por sua vez, apresentam os 50 pares OD com os maiores volumes de deslocamento, respectivamente, terrestres e aéreos, considerados os pares de Unidades Territoriais de Planejamento (UTPs) cujos municípios-sede estão distantes<sup>14</sup> de mais de 300 km. Optou-se neste momento por não apresentar os pares com volumes relevantes de viagens terrestres que correspondem aos conhecidos movimentos pendulares, caracterizados por deslocamentos diários entre grandes aglomerações urbanas e suas cidades satélites.

A identificação desses pares, com as distâncias em linha reta entre os municípios-sede e os respectivos volumes de viagens, é apresentada no Apêndice 2.

---

<sup>14</sup> Levando em consideração para cálculo a distância geográfica entre os dois pontos observados, utilizando a fórmula de Haversine, baseada na Lei dos Cossenos e que considera no modelo a curvatura da Terra, de aproximadamente 6.371 km.

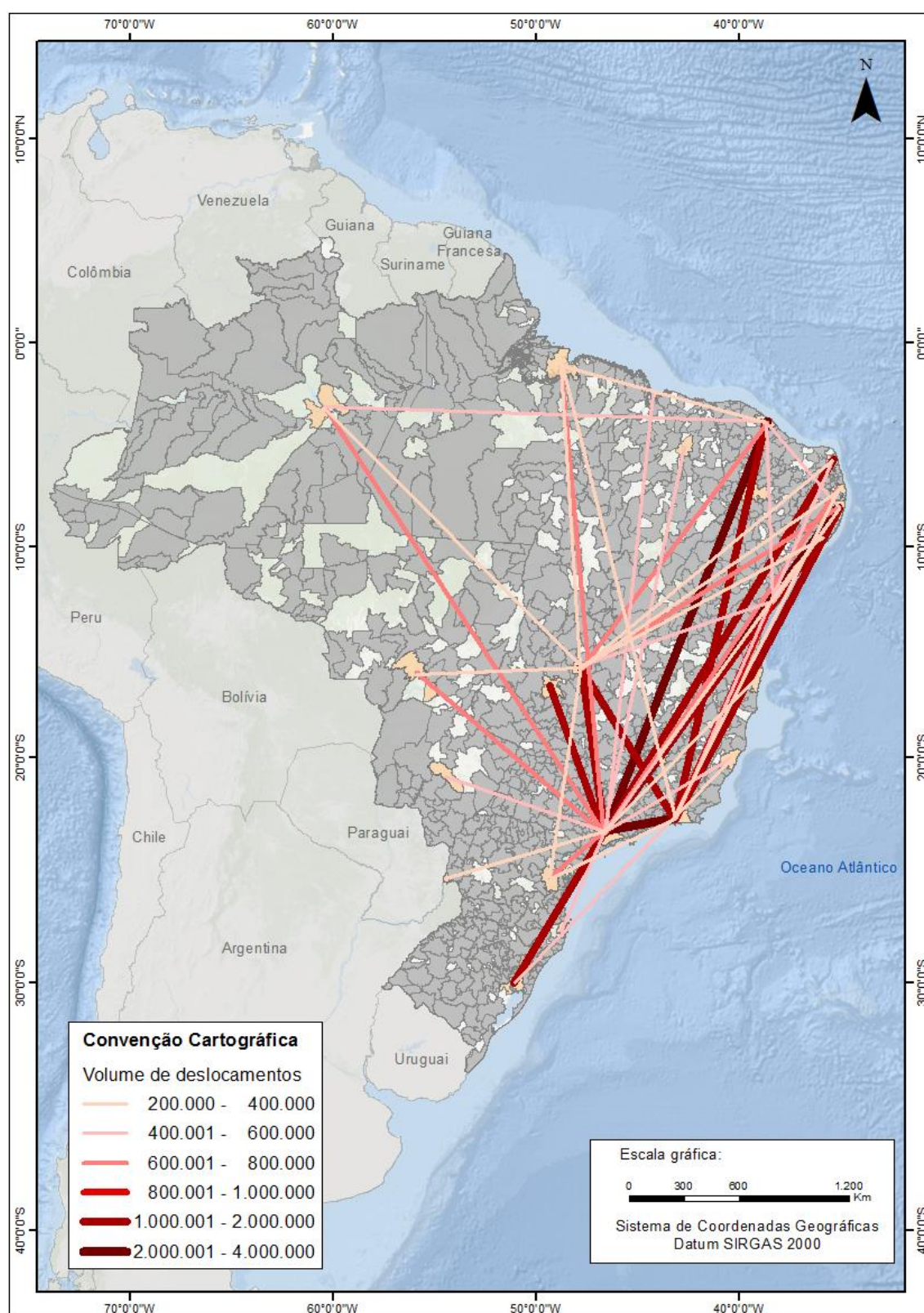
Figura 18 – Principais fluxos de deslocamento: Modo 0 (terrestre)



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



Figura 19 – Principais fluxos de deslocamento: Modo 1 (aéreo)



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

Os principais pares de deslocamento aéreo se caracterizam por serem de distâncias elevadas, cerca de 1.440 km, ao passo que a média das distâncias dos principais pares de deslocamento terrestre é em torno de 530 km.

Há 11 pares de viagens entre UTPs que estão entre as 50 principais para o Modo 0 (terrestre) e para o Modo 1 (aéreo). Vale destacar que esses pares correspondem a capitais brasileiras, como mostra a Tabela 13.

**Tabela 13 – Principais fluxos de deslocamento: Modo 0 (terrestre) e Modo 1 (aéreo)**

UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens		
		Modo 0 (terrestres)	Modo 1 (aéreas)	Ambos
São Paulo (SP) e Rio de Janeiro (RJ)	350	4.749.440	3.784.103	8.533.543
São Paulo (SP) e Recife (PE)	2.126	416.132	3.348.398	3.764.530
São Paulo (SP) e Fortaleza (CE)	2.367	364.348	2.841.560	3.205.908
São Paulo (SP) e Salvador (BA)	1.450	477.595	2.679.119	3.156.714
Brasília (DF) e São Paulo (SP)	875	750.195	1.995.054	2.745.249
São Paulo (SP) e Curitiba (PR)	343	1.957.299	623.716	2.581.015
São Paulo (SP) e Porto Alegre (RS)	854	468.530	1.272.796	1.741.326
Brasília (DF) e Rio de Janeiro (RJ)	927	465.038	1.273.385	1.738.423
São Paulo (SP) e Goiânia (GO)	816	480.824	1.005.846	1.486.670
São Paulo (SP) e Natal (RN)	2.322	113.507	1.272.112	1.385.619
Recife (PE) e Rio de Janeiro (RJ)	1.873	189.250	1.077.049	1.266.299

Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O par com maior volume de deslocamentos, tanto para o Modo 0 (terrestre) quanto para o Modo 1 (aéreo), envolve as UTPs São Paulo (SP) e Rio de Janeiro (RJ), entre as quais está a “ponte aérea” entre os aeroportos de Congonhas (SBSP) e Santos Dumont (SBRJ).

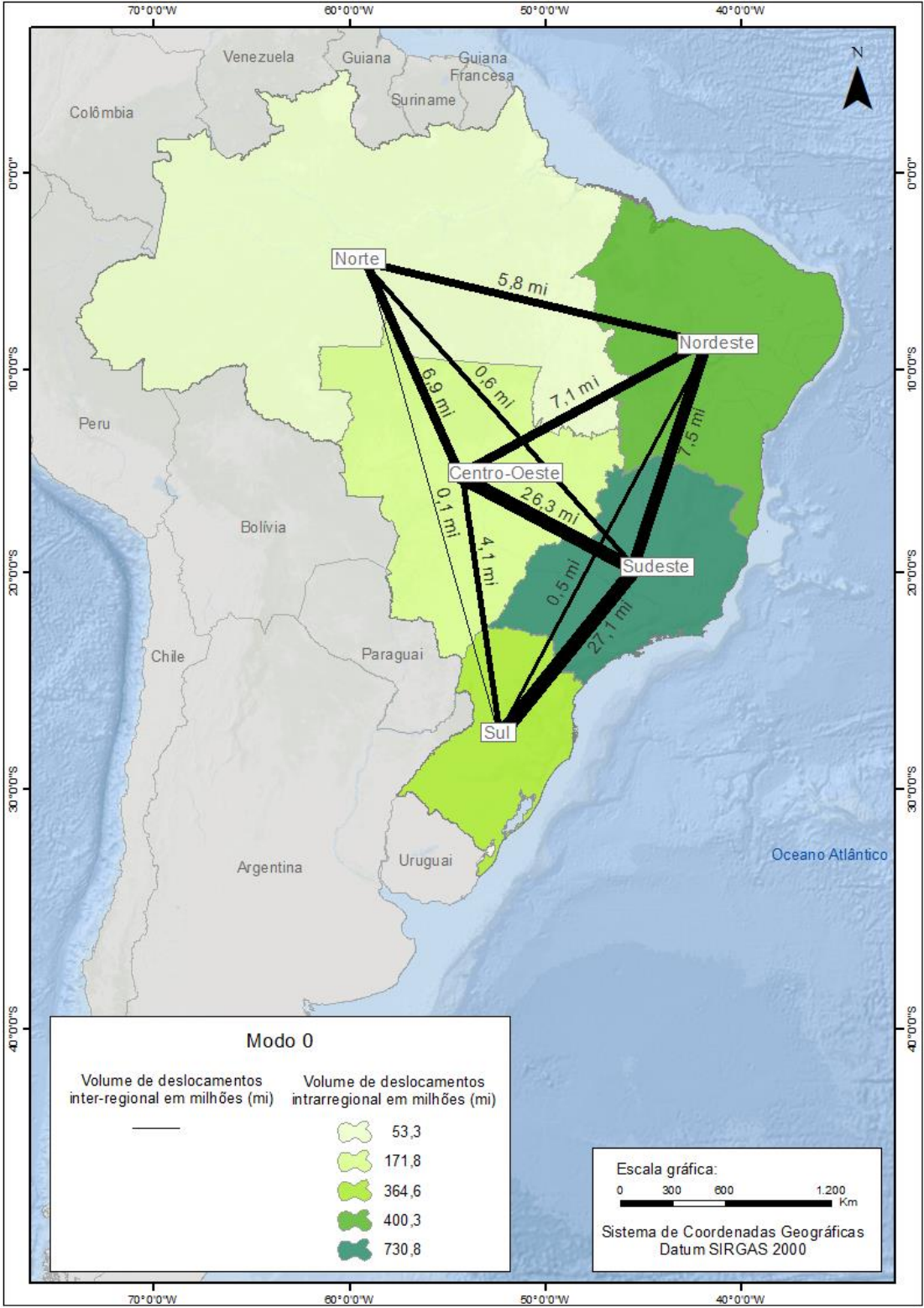
Observa-se que todos os pares envolvem UTPs localizadas na Região Sudeste. Com exceção do par Brasília (DF) e Rio de Janeiro (RJ), eles têm como origem ou destino a UTP São Paulo (SP), composta pela capital, pelos 39 municípios da Região Metropolitana de São Paulo (ou Grande São Paulo) (SÃO PAULO, 2011) e por mais 19 outros municípios, como Campinas e Jundiaí.

## 2.1. Deslocamento da população brasileira entre as regiões

A Figura 20 e a Figura 21 apresentam, referentemente aos deslocamentos observados para o Modo 0 (terrestre) e para o Modo 1 (aéreo), as quantidades de viagens entre as regiões

brasileiras e as quantidades de viagens internas, ou seja, entre Unidades Federativas (UFs) pertencentes à mesma região, independente do sentido.

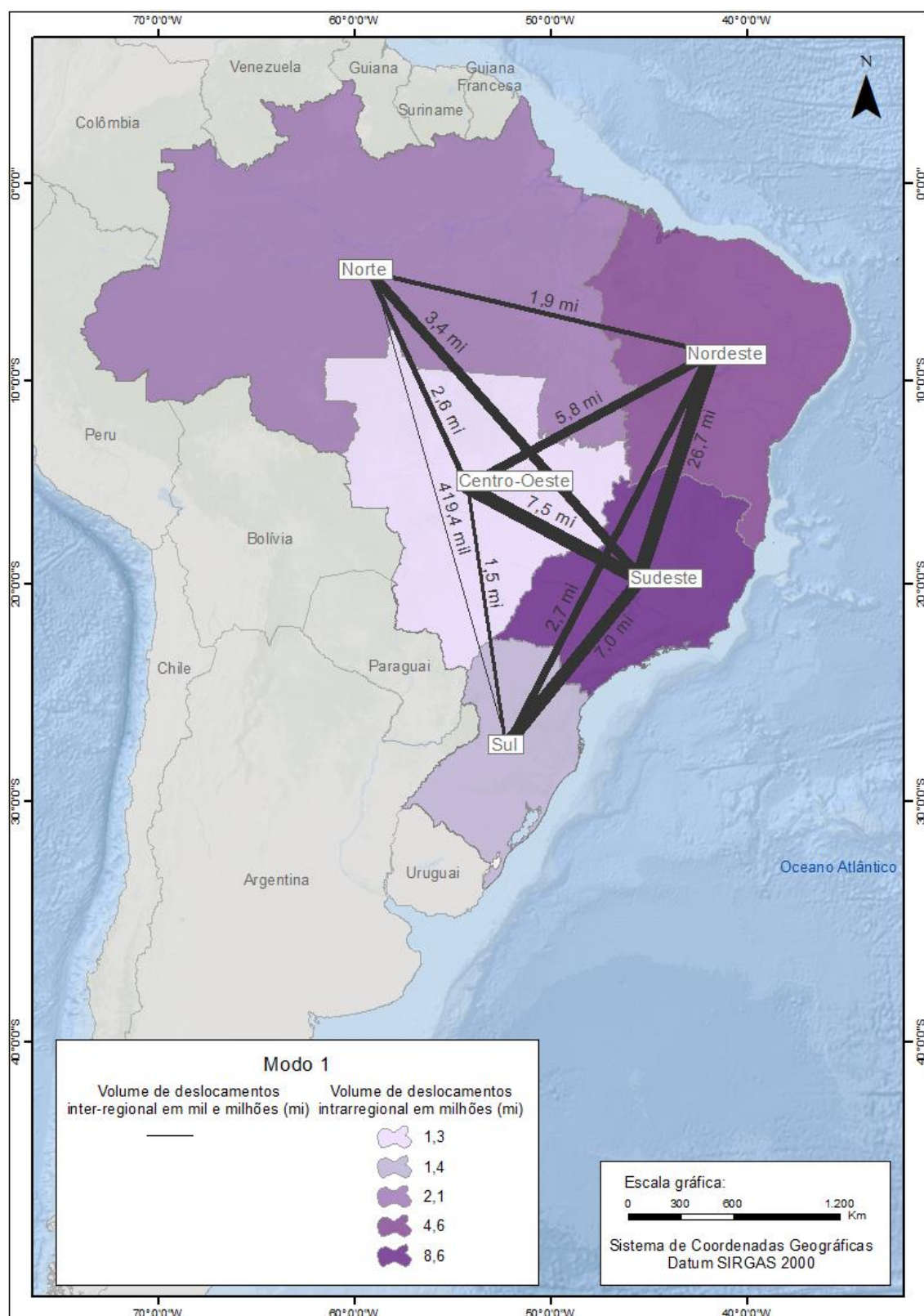
Figura 20 – Volume total de deslocamentos intra e inter-regionais: Modo 0 (terrestre) – Milhões de viagens



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



**Figura 21 – Volume total de deslocamentos intra e inter-regionais: Modo 1 (aéreo) –  
Milhões de viagens**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



Dentre os fluxos observados no Modo 0 (terrestre), 95,7% acontecem de maneira intrarregional, com destaque para a Região Sudeste, onde se concentram 42,2% do total de viagens terrestres, seguida da Região Nordeste (23,1%), e da Região Sul (21,0%). No que se refere às viagens intrarregionais no sudeste, vale destacar que 55,8% das viagens ocorrem dentro do estado de São Paulo, seguidas das viagens internas em Minas Gerais e no Rio de Janeiro, com respectivamente 21,3% e 10,4% de participação no total de viagens terrestres no sudeste. Analisando fluxos interestaduais, destacam-se os pares GO-DF e SP-MG, em ambos os sentidos, que representam, respectivamente, 9,8% e 5,9% do total das viagens terrestres.

Por sua vez, em relação aos deslocamentos aéreos, 76,7% das viagens realizadas ocorrem de maneira inter-regional. O maior volume ocorre entre as Regiões Sudeste e Nordeste, concentrando 34,4% do total de viagens aéreas. Destacam-se os volumes de viagens interestaduais entre SP-PE, SP-BA e SP-CE, com participação de, respectivamente, 16,3%, 15,5% e 14,8% entre as duas regiões, correspondente a cerca de 4 milhões de viagens. Destacam-se, ainda, os deslocamentos aéreos entre SP-RJ e entre SP-DF, que representam 5,2% e 2,9% do total de viagens aéreas observadas.

## **2.2. Deslocamento da população brasileira entre as UTPs**

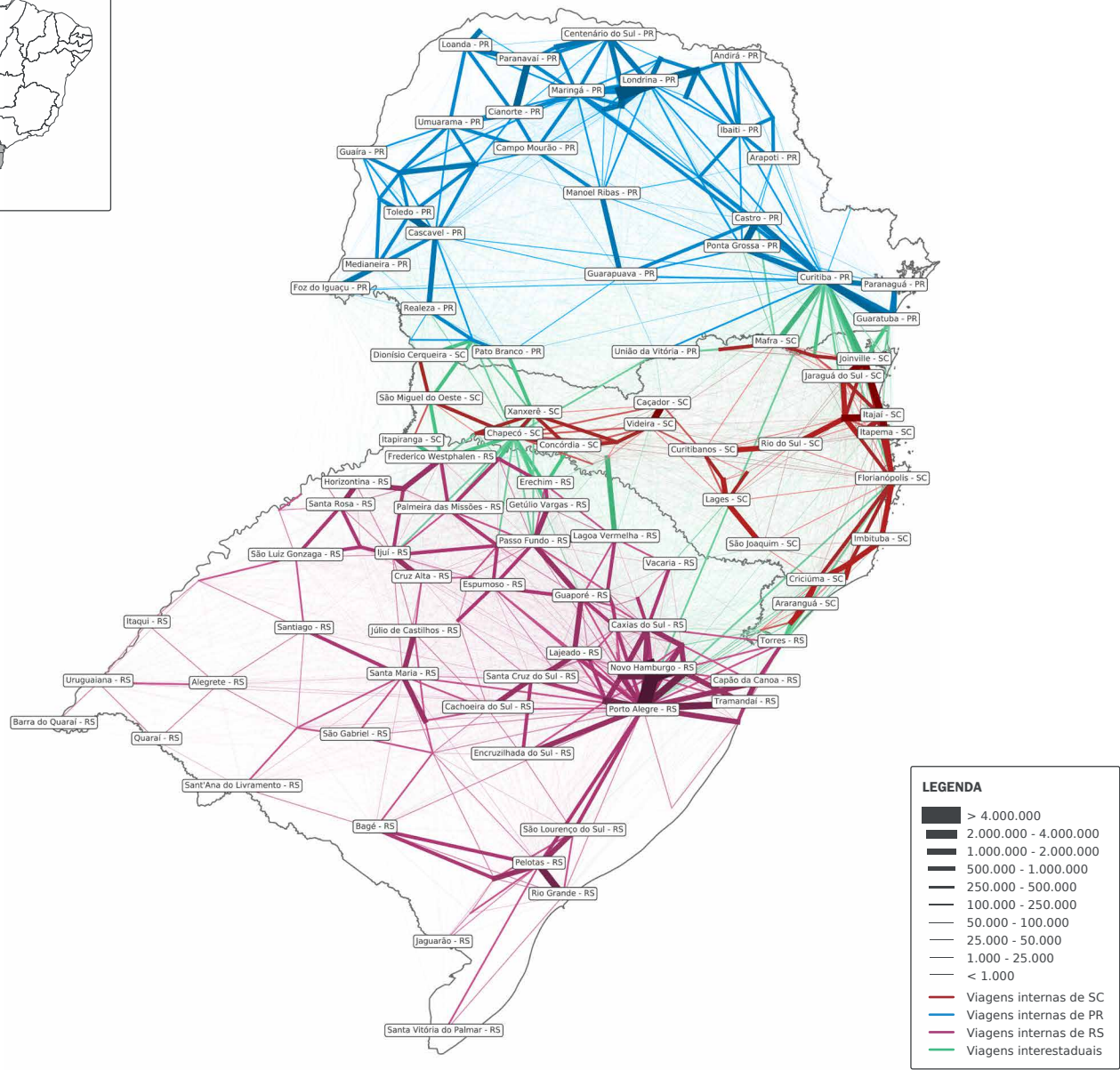
Na Figura 18 e na Figura 19 foram apresentados os principais fluxos de deslocamento entre as UTPs pelo Modo 0 (terrestre) e pelo Modo 1 (aéreo), respectivamente.

Neste item, considerando os volumes de viagens independentemente do modo de transporte, são identificadas as linhas de desejo de deslocamento da população brasileira entre as UTPs, por região.

A seguir, são apresentadas as linhas de desejo de deslocamento da população nas Regiões Sudeste, Sul, Nordeste, Centro-Oeste e Norte, respectivamente. Essas linhas apresentam uma escala de espessura baseada no volume de deslocamentos, independentemente do modo de transporte e sendo considerados ambos os sentidos do fluxo.

# REGIÃO SUL

## Fluxos totais - modo terrestre e aéreo



### Principais UTPs por número de viagens

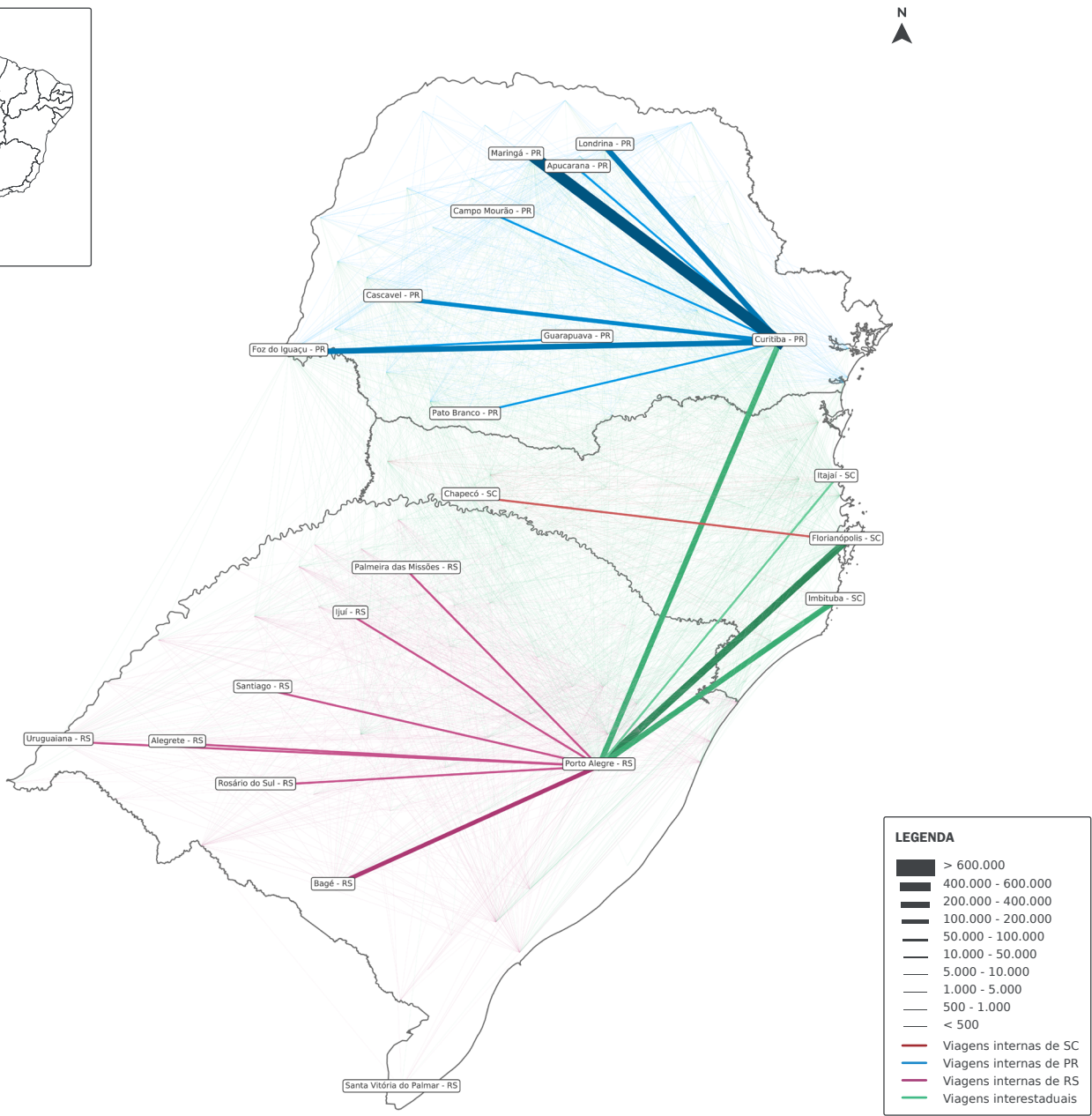
Nome	Viagens (ano)
Novo Hamburgo - RS	26.001.810
Curitiba - PR	24.694.691
Porto Alegre - RS	18.584.056
Blumenau - SC	11.257.991
Cascavel - PR	10.160.806
Londrina - PR	10.024.665
Florianópolis - SC	9.832.613
Chapecó - SC	9.666.081
Maringá - PR	9.439.966
Apucarana - PR	8.285.461

### Principais pares OD











Nome	Viagens (ano)	Viagens (semana)
Novo Hamburgo/RS - Porto Alegre/RS	15.675.839	301.458
Ivoti/RS - Novo Hamburgo/RS	5.073.442	97.566
Arapongas/PR - Londrina/PR	5.032.256	96.774
Novo Hamburgo/RS - Sapiranga/RS	4.155.890	79.921
Charqueadas/RS - Porto Alegre/RS	3.879.171	74.599
Ponta Grossa/PR - Castro/PR	3.766.558	72.434
Blumenau/SC - Itajaí/SC	3.640.023	70.000
Paranavaí/PR - Cianorte/PR	3.564.954	68.557
Apucarana/PR - Arapongas/PR	3.492.312	67.160
Cascavel/PR - Toledo/PR	3.295.410	63.373

# REGIÃO SUL

## Fluxos com distância geográfica acima de 300 km



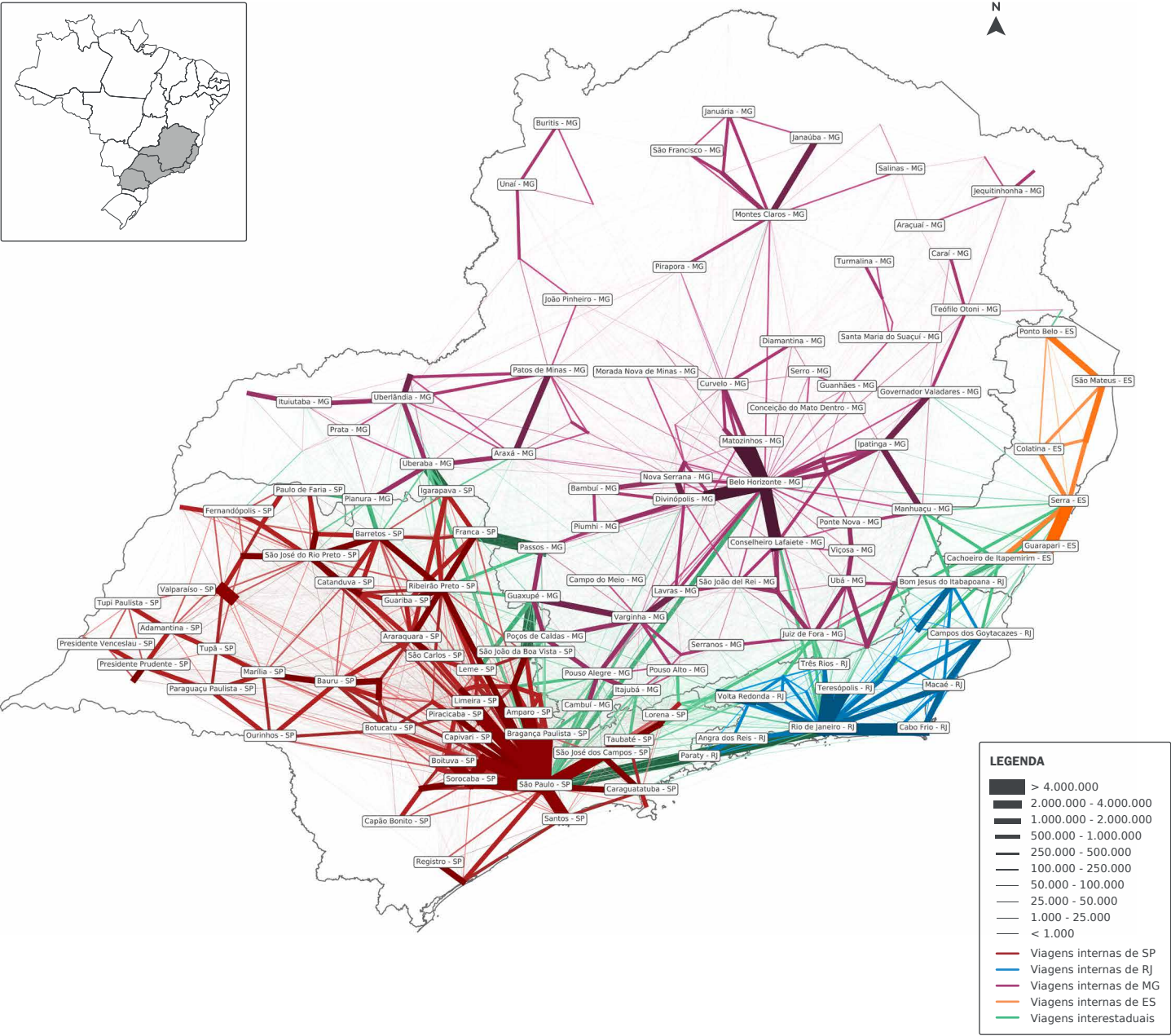
### Principais pares OD

Principais pares OD				Principais pares OD			
Nome		Viagens (ano)	Viagens (semana)	Nome		Viagens (ano)	Viagens (semana)
	Curitiba/PR - Maringá/PR	602.016	11.577		Imbituba/SC - Porto Alegre/RS	202.390	3.892
	Florianópolis/SC - Porto Alegre/RS	494.189	9.504		Cascavel/PR - Curitiba/PR	197.721	3.802
	Curitiba/PR - Porto Alegre/PR	358.270	6.890		Bagé/RS - Porto Alegre/RS	171.860	3.305
	Curitiba/PR - Londrina/PR	341.930	6.576		Itajaí/SC - Porto Alegre/RS	95.727	1.841
	Curitiba/PR - Foz do Iguaçu/PR	249.628	4.801		Porto Alegre/RS - Ijuí/RS	85.560	1.645



# REGIÃO SUDESTE

## Fluxos totais - modo terrestre e aéreo



### Principais UTPs por número de viagens

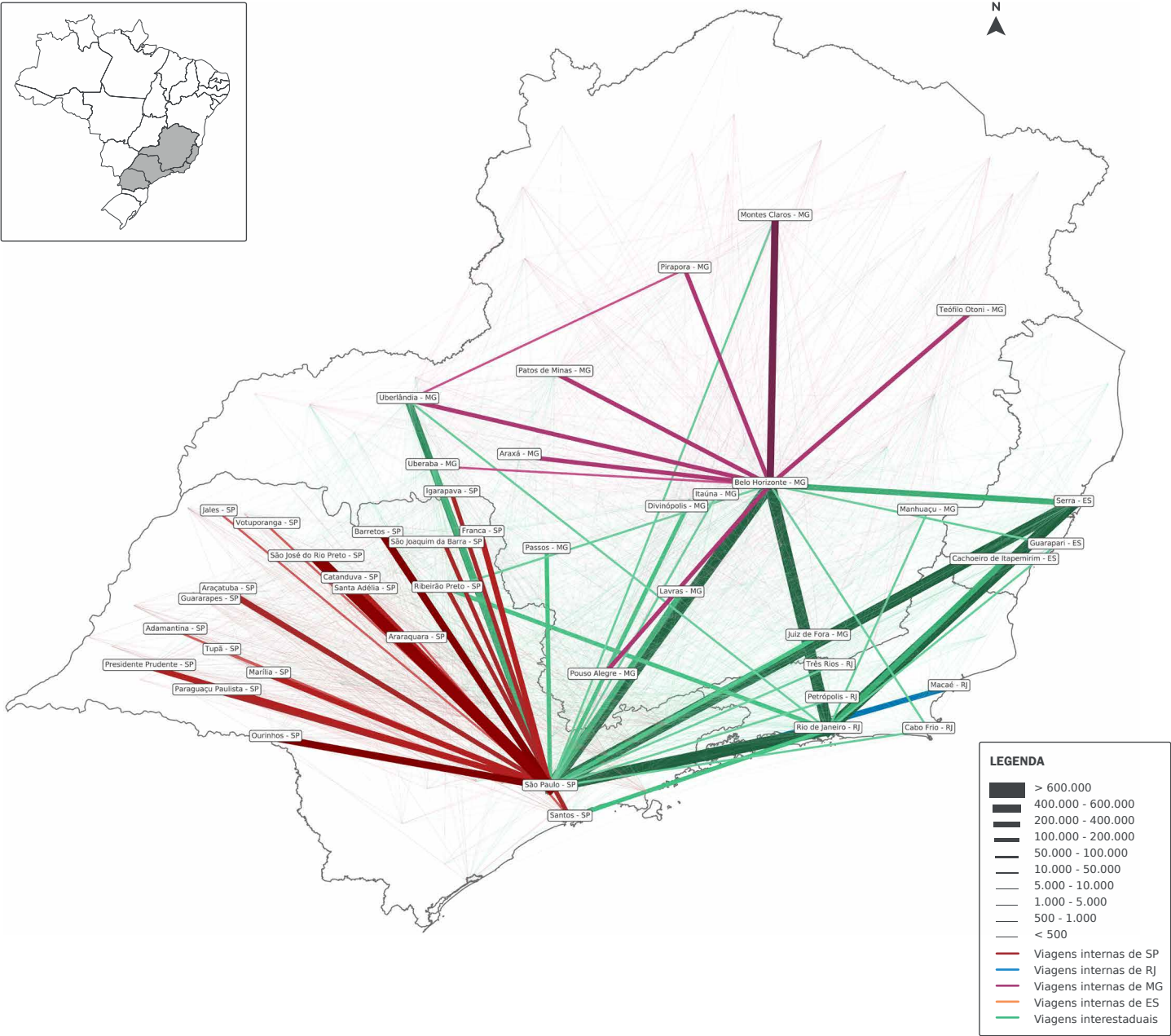
Nome	Viagens (ano)
Belo Horizonte - MG	43.385.716
São Paulo - SP	38.844.088
Rio de Janeiro - RJ	26.661.164
Santos - SP	26.446.878
São José dos Campos - SP	25.619.224
Americana - SP	23.510.555
Bragança Paulista - SP	18.565.799
Itu - SP	18.038.466
Araraquara SP	17.718.517
Barretos - SP	15.018.166

### Principais pares OD

Nome	Viagens (ano)	Viagens (semana)
Santos/SP - São Paulo/SP	21.986.609	422.819
São Paulo/SP - São Roque/SP	17.495.470	336.451
São José dos Campos/SP - São Paulo/SP	15.332.602	294.858
Bragança Paulista/SP - São Paulo/SP	14.908.639	286.705
Americana/SP - São Paulo/SP	13.514.712	259.898
Petrópolis/RJ - Rio de Janeiro/RJ	11.696.677	224.936
Itu/SP - São Paulo/SP	10.152.094	195.233
São José dos Campos/SP - Taubaté/SP	9.541.997	183.500
Rio de Janeiro/RJ - São Paulo/SP	8.533.543	164.107
Itatiba/SP - São Paulo/SP	7.573.164	145.638

# REGIÃO SUDESTE

## Fluxos com distância geográfica acima de 300 km

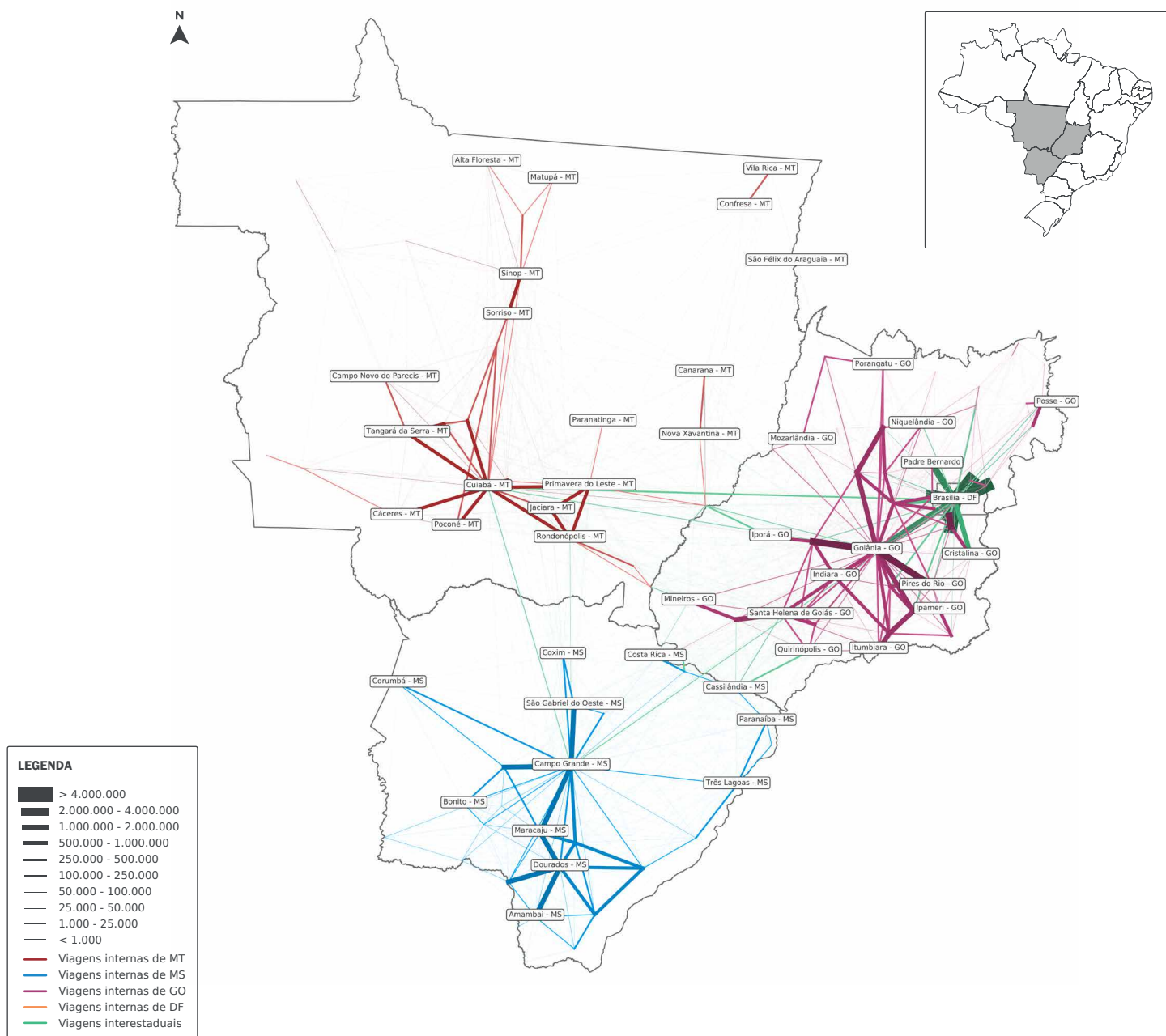


### Principais pares OD

Nome	Viagens (ano)	Viagens (semana)	Nome	Viagens (ano)	Viagens (semana)
Rio de Janeiro/RJ - São Paulo/SP	8.533.543	164.107	São Paulo/SP - Serra/ES	697.686	13.417
Belo Horizonte/MG - São Paulo/SP	1.753.166	33.715	São Paulo/SP - Uberlândia/MG	566.026	10.885
Belo Horizonte/MG - Rio de Janeiro/RJ	867.548	16.684	Barretos/SP - São Paulo/SP	493.237	9.485
Rio de Janeiro/RJ - Serra/ES	724.348	13.930	Ourinhos/SP - São Paulo/SP	451.866	8.690
São José do Rio Preto/SP - São Paulo/SP	718.553	13.818	Belo Horizonte/MG - Montes Claros/MG	433.525	8.337

# REGIÃO CENTRO-OESTE

## Fluxos totais - modo terrestre e aéreo



### Principais UTPs por número de viagens

Nome	Viagens (ano)
Brasília - DF	75.497.974
Goiânia - GO	22.345.467
Cuiabá - MT	9.218.668
Campo Grande - MS	8.904.343
Dourados - MS	7.502.413
Valparaíso de Goiás - GO	6.265.478
Itumbiara - GO	5.472.019
Caldas Novas - GO	4.395.165
Rio Verde - GO	3.894.735
Ceres - GO	3.880.317

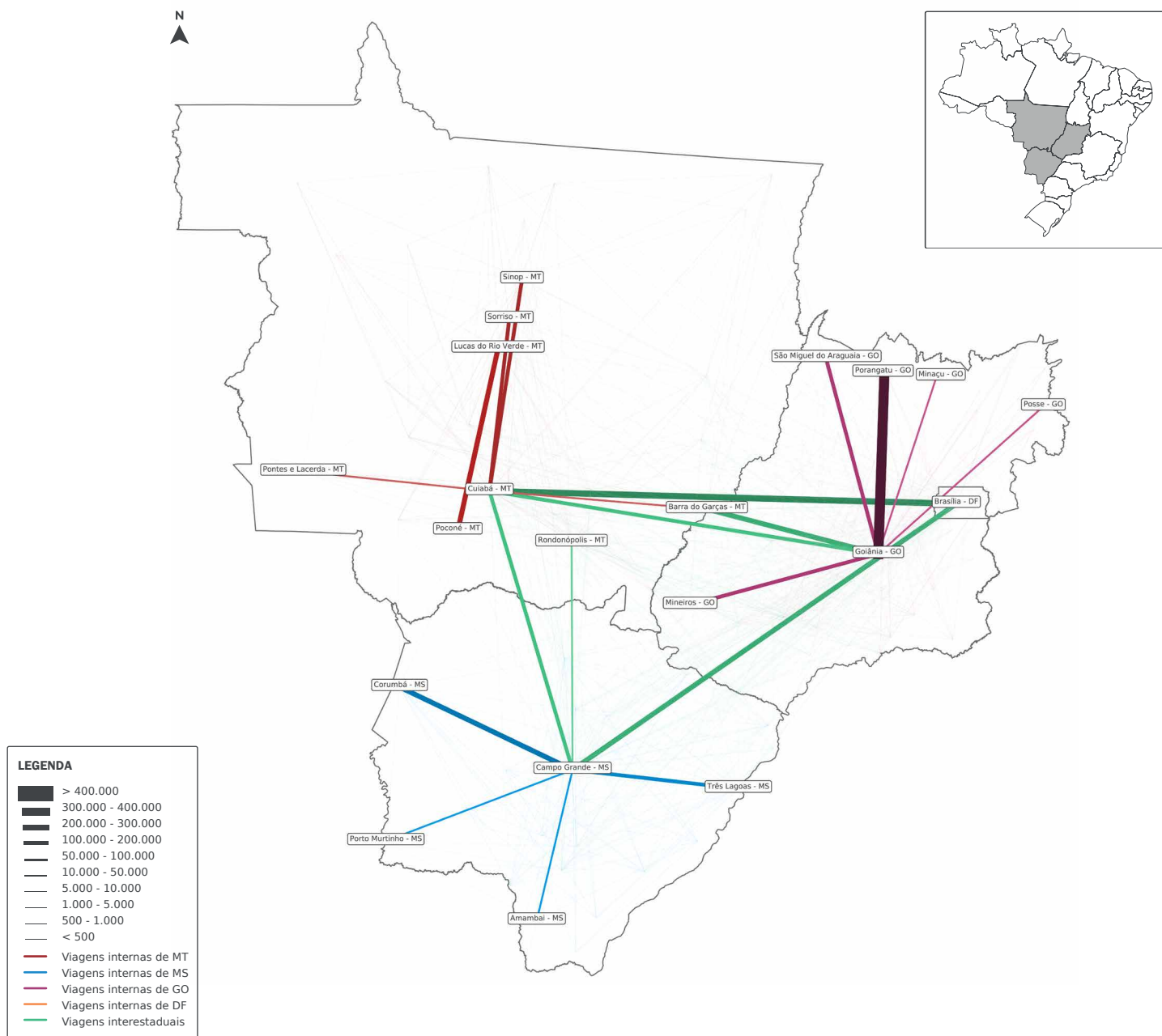
### Principais pares OD

Nome	Viagens (ano)	Viagens (semana)
Brasília/DF - Valparaíso de Goiás/GO	23.953.840	460.651
Brasília/DF - Água Lindas de Goiás/GO	19.442.836	373.901
Brasília/DF - Luziânia/GO	7.843.019	150.827
Brasília/DF - Planaltina/DF	6.188.634	119.012
Brasília/DF - Cidade Ocidental/GO	5.800.316	111.545
Brasília/DF - Formosa/GO	4.168.155	80.157
Brasília/DF - Santo Antônio do Descoberto/GO	3.723.418	71.604
Brasília/DF - Padre Bernardo/GO	3.668.056	70.540
Valparaíso de Goiás/GO - Luziânia/GO	3.377.355	64.949
Goiânia/GO - Brasília/DF	3.033.051	58.328













# REGIÃO CENTRO-OESTE

## Fluxos com distância geográfica acima de 300 km



### Principais pares OD

Principais pares OD				Principais pares OD			
Nome		Viagens (ano)	Viagens (semana)	Nome		Viagens (ano)	Viagens (semana)
	Goiânia/GO - Porangatu/GO	435.405	8.373		Campo Grande/MS - Brasília/DF	201.356	3.872
	Cuiabá/MT - Brasília/DF	364.567	7.011		Cuiabá/MT - Sinop/MT	179.310	3.448
	Lucas do Rio Verde/MT - Poconé/MT	273.830	5.266		Campo Grande/MS - Três Lagoas/MS	151.770	2.919
	Campo Grande/MS - Corumbá/MS	258.069	4.963		Campo Grande/MS -	143.642	2.762
	Barra do Garças/MT - Goiânia/GO	207.289	3.986		Cuiabá/MT - Goiânia/GO	131.005	2.519



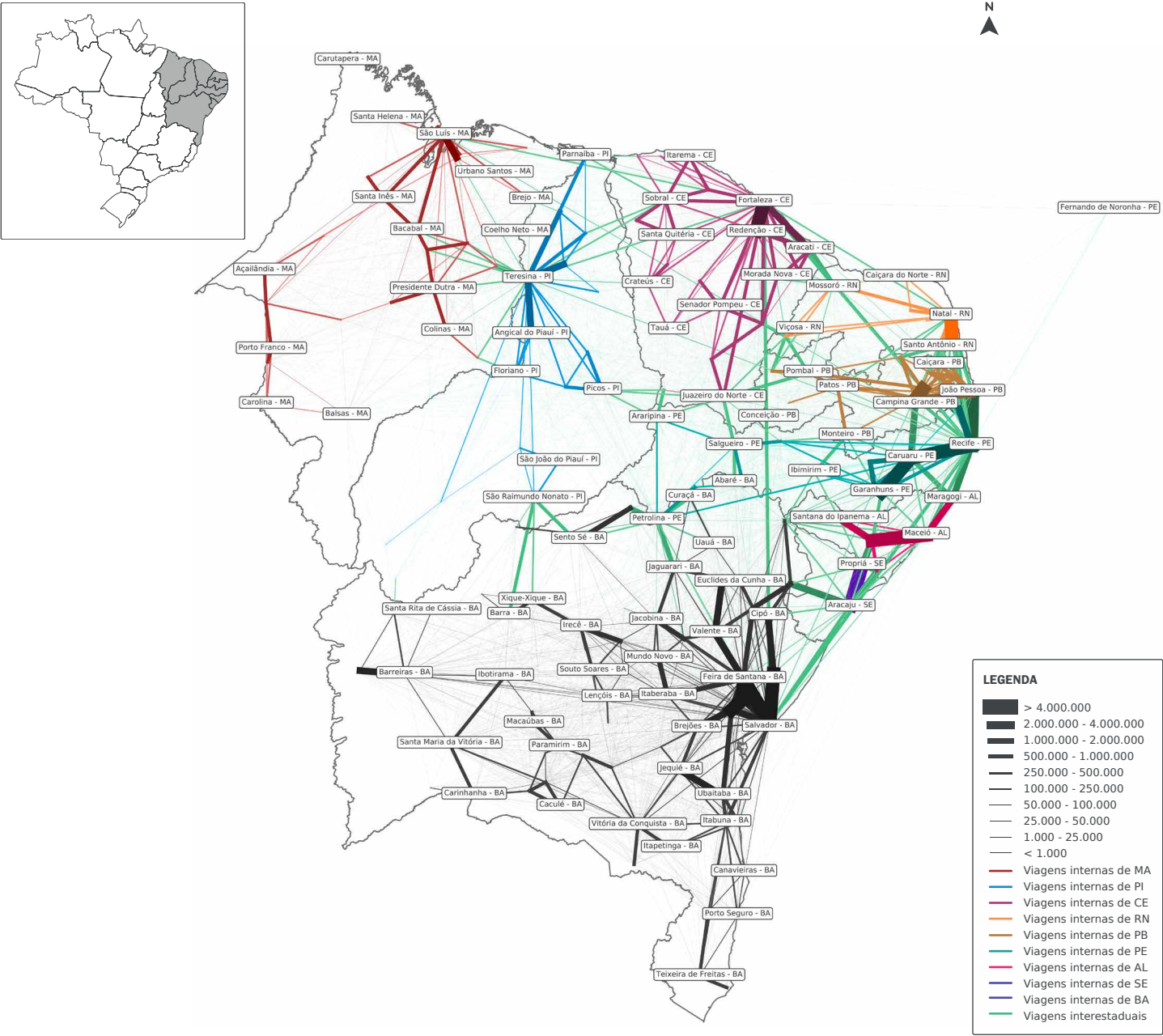
Rotas aéreas já existentes



Rotas aéreas potenciais

# REGIÃO NORDESTE

## Fluxos totais - modo terrestre e aéreo



### Principais UTPs por número de viagens

Nome	Viagens (ano)
Fortaleza - CE	30.202.151
Caruaru - PE	19.262.295
Feira de Santana - BA	18.848.447
Cachoeira - BA	18.415.583
Alagoinhas - BA	16.830.845
Aracaju - SE	15.120.811
Natal - RN	14.840.340
Arapiraca - AL	14.721.374
Campina Grande - PB	14.089.173
João Pessoa - PB	12.296.848

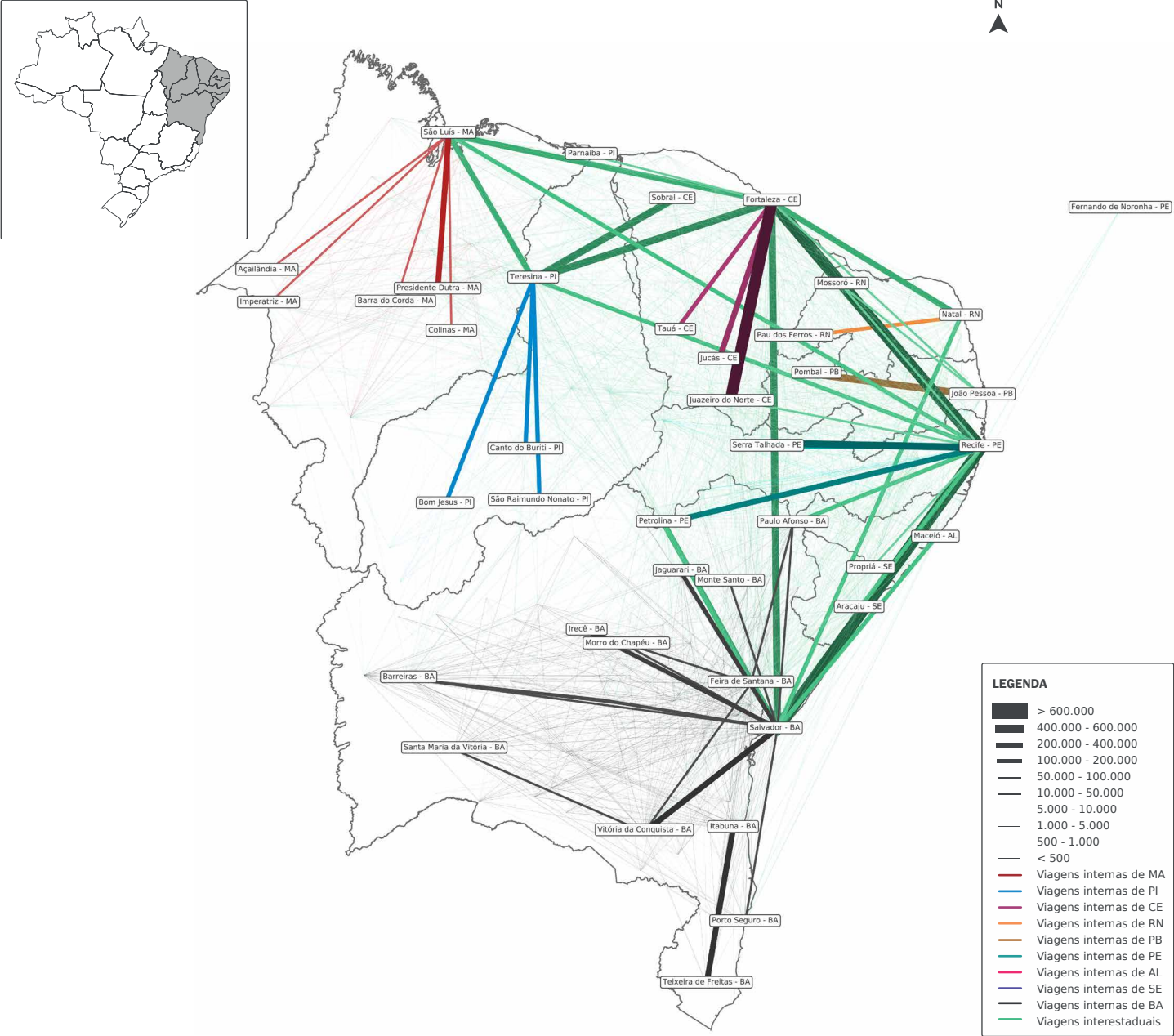
### Principais pares OD

Nome	Viagens (ano)	Viagens (semana)
Alagoinhas/BA - Salvador/BA	11.592.024	222.924
Fortaleza/CE - Redenção/CE	9.939.300	191.140
Caruaru/PE - Recife/PE	9.782.987	188.134
Cachoeira/BA - Feira de Santana/BA	8.994.158	172.965
Feira de Santana/BA - Salvador/BA	8.294.788	159.515
Natal/RN - São José do Mipibu/RN	7.810.311	150.198
Aracaju/E - Propriá/SE	6.352.738	122.168
João Pessoa/PB - Recife/PE	6.097.080	117.252
Arapiraca/AL - Maceió/AL	6.014.620	115.666
Itambé/PE - João Pessoa/PB	4.963.180	95.446













# REGIÃO NORDESTE

## Fluxos com distância geográfica acima de 300 km



### Principais pares OD

Principais pares OD				Principais pares OD			
Nome		Viagens (ano)	Viagens (semana)	Nome		Viagens (ano)	Viagens (semana)
	Fortaleza/CE - Recife/PE	875.599	16.838		Recife/PE - Serra Talhada/PE	479.994	9.231
	Recife/PE - Salvador/BA	710.817	13.670		Sobral/CE - Teresina/PI	465.305	8.948
	Fortaleza/CE - Juazeiro do Norte/CE	663.701	12.764		Fortaleza/CE - Teresina/PI	408.539	7.857
	Fortaleza/CE - Salvador/BA	556.562	10.703		Petrolina/PE - Recife/PE	395.924	7.614
	João Pessoa/PB - Pombal/PB	514.976	9.903		Natal/RN - Pau dos Ferros/RN	383.855	7.382



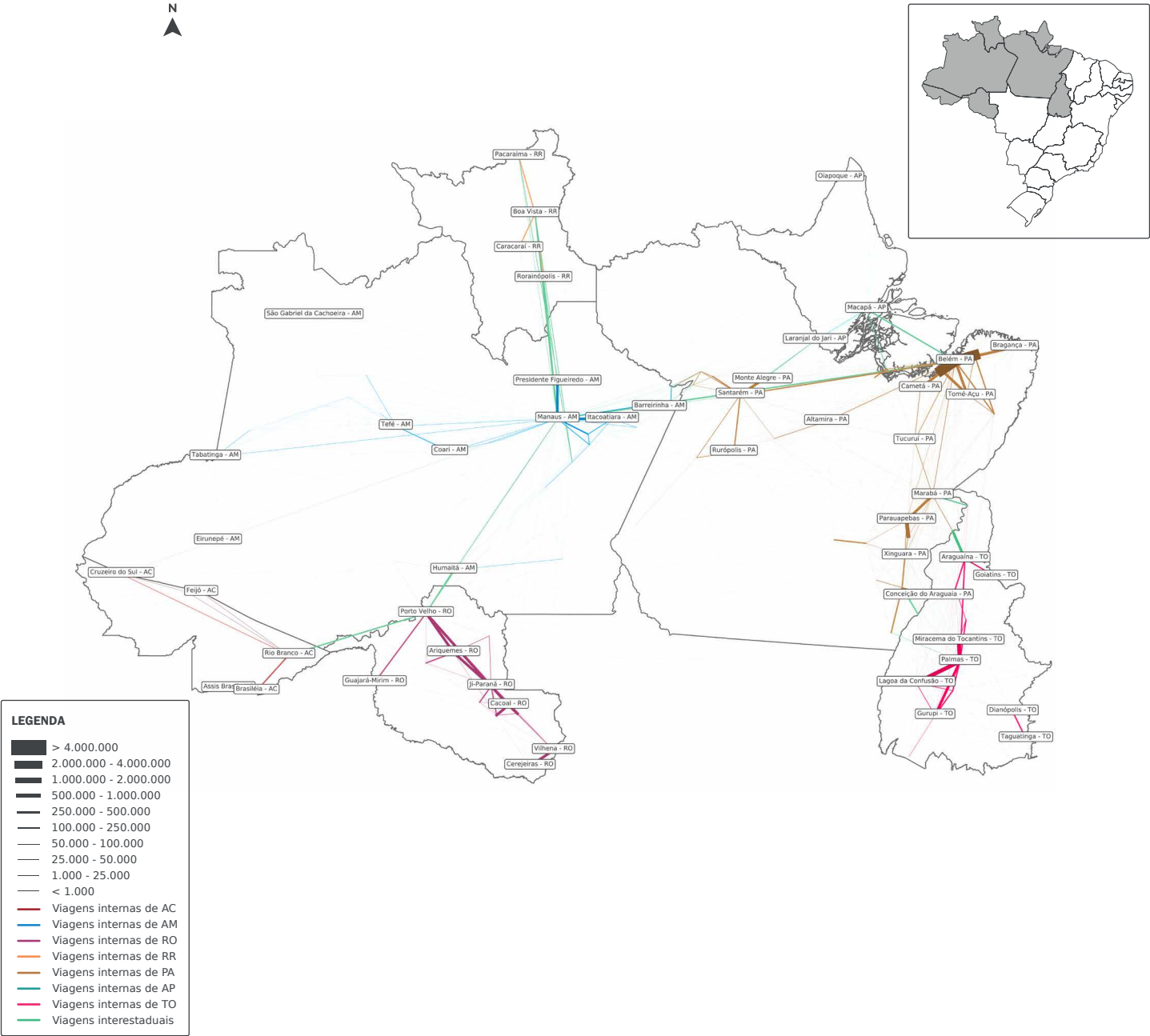
Rotas aéreas já existentes



Rotas aéreas potenciais

# REGIÃO NORTE

## Fluxos totais - modo terrestre e aéreo



### Principais UTPs por número de viagens

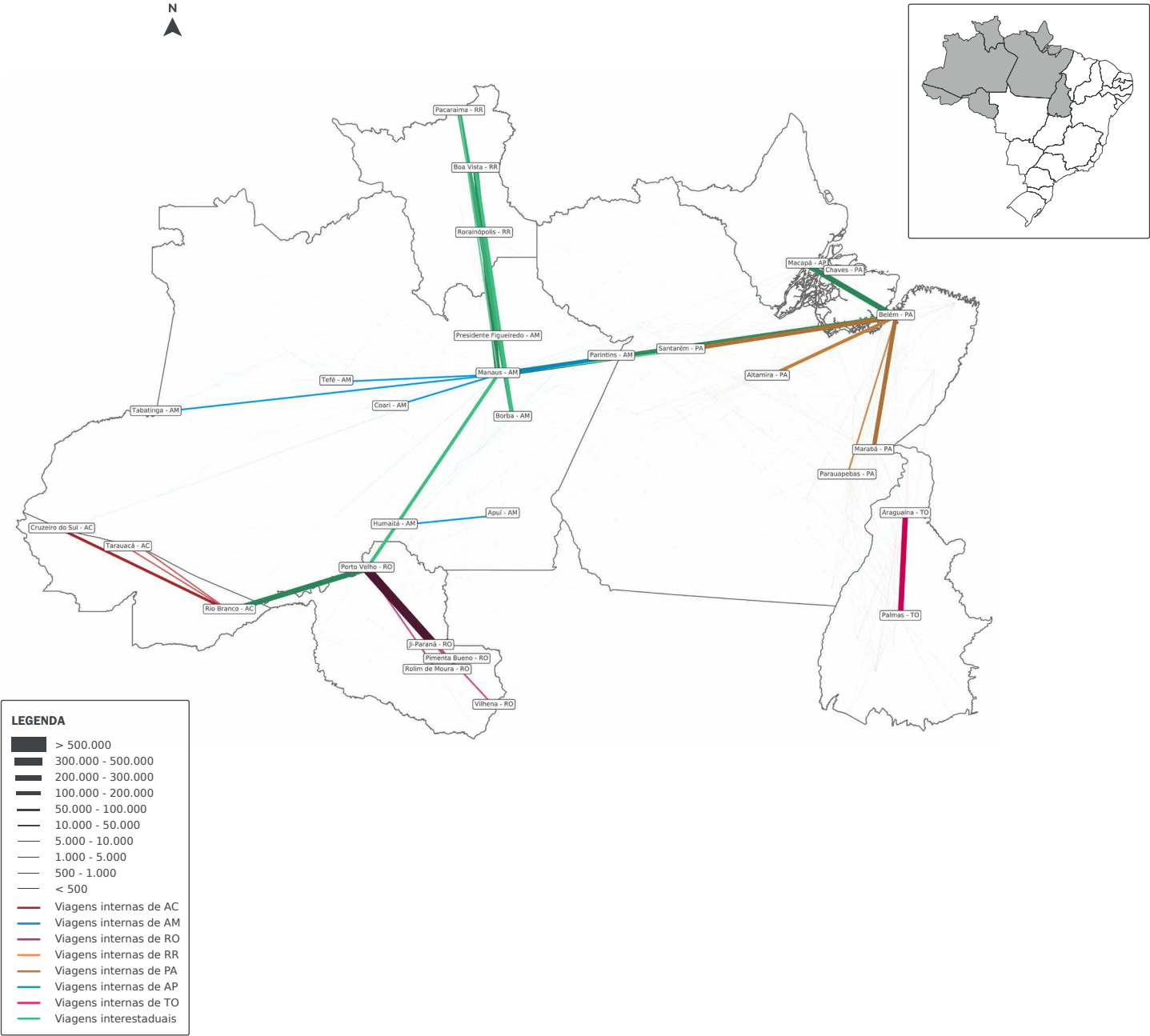
Nome	Viagens (ano)
Belém - PA	14.221.816
Abaetetuba - PA	6.695.783
Manaus - AM	6.123.007
Palmas - TO	4.051.447
Ji-Paraná - RO	3.709.244
Araguaína - TO	3.335.650
Porto Velho - RO	2.724.696
Marabá - PA	2.224.285
Santarém - PA	2.128.789
Boa Vista - RR	1.981.285

### Principais pares OD











Nome	Viagens (ano)	Viagens (semana)
Belém/PA - Castanhal/PA	6.998.495	134.586
Belém/PA - Abaetetuba/PA	5.673.647	109.109
Palmas/TO - Lagoa da Confusão/TO	1.268.008	24.385
Miracema do Tocantins/TO - Palmas/TO	1.196.533	23.010
Parauapebas/PA - Canaã dos Carajás/PA	1.075.749	20.687
Cacoal/RO - Pimenta Bueno/RO	971.364	18.680
Marabá/PA - Parauapebas/PA	895.175	17.215
Manaus/AM - Itacoatiara/AM	846.075	16.271
Manaus/AM - Presidente Figueiredo/AM	842.008	16.192
Ji-Paraná/RO - Pimenta Bueno/RO	762.076	14.655

# REGIÃO NORTE

## Fluxos com distância geográfica acima de 300 km



### Principais pares OD

Principais pares OD				Principais pares OD			
Nome		Viagens (ano)	Viagens (semana)	Nome		Viagens (ano)	Viagens (semana)
	Ji-Paraná/RO - Porto Velho/RO	509.349	9.795		Porto Velho/RO - Rio Branco/AC	327.024	6.289
	Araguaína/TO - Palmas/TO	362.142	6.964		Belém/PA - Manaus/AM	308.500	5.933
	Manaus/AM - Rorainópolis/RR	352.976	6.788		Belém/PA - Santarém/PA	277.334	5.333
	Belém/PA - Macapá/AP	336.392	6.469		Borba/AM - Pacaraima/RR	217.861	4.190
	Boa Vista/RR - Manaus/AM	335.272	6.448		Belém/PA - Marabá/PA	215.883	4.152



Rotas aéreas já existentes



Rotas aéreas potenciais

## 2.3. Potencial demanda reprimida

O item 2.2 expõe as linhas de desejo de deslocamento da população, apresentando os volumes de viagens independentemente do modo de transporte utilizado.

Ao se analisar o modo de transporte, os pares OD com volumes relevantes e participação significativa do Modo 1 (aéreo) indicam UTPs nas quais os aeroportos de embarque e de desembarque utilizados requerem especial preocupação quanto à manutenção e investimentos em infraestrutura, a fim de que estes possam atender à demanda de curto e longo prazo com níveis de serviço adequados.

Por sua vez, pares OD para os quais os volumes de deslocamento do Modo 0 (terrestre) são expressivamente superiores aos do Modo 1 (aéreo), quando as sedes das UTPs envolvidas são distantes de mais de 300 km<sup>15</sup>, podem indicar demanda reprimida, ou seja, desejos de viagem não contemplados pelo transporte aéreo. São motivos possíveis para tal:

1. Não há aeroportos próximos à origem e/ou ao destino, indicando localizações estratégicas para a implantação de aeródromos.
2. Há aeroportos próximos à origem e ao destino, no entanto, estes não apresentam infraestrutura compatível com as aeronaves que as empresas aéreas desejam operar ou não há interesse pelas empresas em operar voos entre eles.
3. Há aeroportos próximos à origem e ao destino e são operados voos entre eles, no entanto, é demonstrada uma preferência da população pelo modo de transporte rodoviário.

Nesse sentido, a Tabela 14 apresenta os deslocamentos para os quais o volume de viagens pelo Modo 0 (terrestre) supera o volume de viagens pelo Modo 1 (aéreo) em mais de 300 mil viagens. Na última coluna pode-se observar a participação relativa aproximada do Modo 0 (não aéreo ou terrestre).

**Tabela 14 – Pares de UTPs com viagens terrestres superando em mais de 300 mil as viagens aéreas**

UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens		Potencial demanda reprimida (viagens)	% Modo 0
		Modo 0 (terrestres)	Modo 1 (aéreas)		
São Paulo (SP) e Curitiba (PR)	343	1.957.299	623.716	<b>1.333.583</b>	75,8
São Paulo (SP) e Rio de Janeiro (RJ)	350	4.749.440	3.784.103	<b>965.337</b>	55,7
São Paulo (SP) e Maringá (PR)	548	669.173	154.215	<b>514.958</b>	81,3
Pombal (PB) e João Pessoa (PB)	326	511.213	3.763	<b>507.450</b>	99,3

<sup>15</sup> Caracterizando uma viagem de longa distância e considerada, para efeitos deste estudo, com viabilidade para operação de uma rota aérea. Sabe-se, no entanto, que essa viabilidade depende de inúmeros fatores, como: infraestrutura aeroportuária na origem e no destino compatível com a aeronave que a empresa aérea deseja utilizar, receita por assento quilômetro ofertado (RASK, do inglês – *revenue per available seat kilometer*), custo por assento quilômetro ofertado (CASK, do inglês – *cost per available seat kilometer*), entre outros.

UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens		Potencial demanda reprimida (viagens)	% Modo 0
		Modo 0 (terrestres)	Modo 1 (aéreas)		
Maringá (PR) e Curitiba (PR)	351	553.252	48.764	504.488	91,9
Rio de Janeiro (RJ) e Belo Horizonte (MG)	335	683.381	184.167	499.214	78,8
Porto Velho (RO) e Ji-Paraná (RO)	313	501.401	7.948	493.453	98,4
Serra Talhada (PE) e Recife (PE)	375	477.240	2.754	474.486	99,4
Teresina (PI) e Sobral (CE)	314	463.852	1.453	462.399	99,7
São Paulo (SP) e Barretos (SP)	393	476.438	16.799	459.639	96,6
São Paulo (SP) e Ourinhos (SP)	343	446.996	4.870	442.126	98,9
Porangatu (GO) e Goiânia (GO)	359	435.310	95	435.215	99,9
São Paulo (SP) e São José do Rio Preto (SP)	421	570.419	148.134	422.285	79,4
Juazeiro do Norte (CE) e Fortaleza (CE)	396	540.391	123.322	417.069	81,4
Pau dos Ferros (RN) e Natal (RN)	329	383.417	438	382.979	99,9
Teixeira de Freitas (BA) e Itabuna (BA)	310	373.371	321	373.050	99,9
Rorainópolis (RR) e Manaus (AM)	455	352.976	-	352.976	100,0
São Paulo (SP) e Marília (SP)	378	367.899	18.245	349.654	95,3
Montes Claros (MG) e Belo Horizonte (MG)	358	389.697	43.828	345.869	89,9
São Paulo (SP) e Belo Horizonte (MG)	487	1.047.138	706.028	341.110	59,7
Palmas (TO) e Araguaína (TO)	331	340.163	21.979	318.184	93,9
Manaus (AM) e Itapipoca (CE)	2269	329.140	13.116	316.024	96,2
Porto Alegre (RS) e Florianópolis (SC)	377	404.522	89.667	314.855	81,9
Sobral (CE) e Manaus (AM)	2185	316.494	10.640	305.854	96,7
Londrina (PR) e Curitiba (PR)	304	319.008	22.922	296.086	93,3
São Paulo (SP) e Santa Adélia (SP)	346	295.168	3.268	291.900	98,9
Poconé (MT) e Lucas do Rio Verde (MT)	363	273.830	-	273.830	100,0
São Paulo (SP) e Franca (SP)	348	275.681	6.670	269.011	97,6
Corumbá (MS) e Campo Grande (MS)	356	251.922	6.147	245.775	97,6
Jucás (CE) e Fortaleza (CE)	328	248.772	9.528	239.244	96,3
São Paulo (SP) e Lavras (MG)	305	235.474	1.332	234.142	99,4
Presidente Dutra (MA) e São Luís (MA)	307	225.011	914	224.097	99,6
Rio Branco (AC) e Porto Velho (RO)	457	274.016	53.008	221.008	83,8
Rio de Janeiro (RJ) e Cachoeiro de Itapemirim (ES)	314	221.135	3.663	217.472	98,4
Rorainópolis (RR) e Presidente Figueiredo (AM)	336	213.173	-	213.173	100,0
São Paulo (SP) e Catanduva (SP)	366	198.132	3.110	195.022	98,5
Porto Alegre (RS) e Imbituba (SC)	318	198.704	3.686	195.018	98,2
Tauá (CE) e Fortaleza (CE)	316	194.054	1.577	192.477	99,2
São Paulo (SP) e Paraguaçu Paulista (SP)	429	193.112	2.878	190.234	98,5
São Paulo (SP) e Passos (MG)	317	186.524	1.872	184.652	99,0
Uberaba (MG) e São Paulo (SP)	453	202.912	18.393	184.519	91,7
Goiânia (GO) e Barra do Garças (MT)	332	194.578	12.711	181.867	93,9
Vitória da Conquista (BA) e Salvador (BA)	326	208.414	26.783	181.631	88,6

UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens		Potencial demanda reprimida (viagens)	% Modo 0
		Modo 0 (terrestres)	Modo 1 (aéreas)		
Gurupi (TO) e Goiânia (GO)	550	179.782	5.202	174.580	97,2
Jaguarari (BA) e Salvador (BA)	358	174.161	2.398	171.763	98,6
Porto Alegre (RS) e Bagé (RS)	311	170.612	1.248	169.364	99,3
Pirapora (MG) e Belo Horizonte (MG)	307	169.583	669	168.914	99,6
São Paulo (SP) e Presidente Prudente (SP)	519	199.712	31.630	168.082	86,3
Irecê (BA) e Salvador (BA)	413	164.788	1.648	163.140	99,0
São Paulo (SP) e Curitiba (PR)	437	150.896	472	150.424	99,7
São Paulo (SP) e Rio de Janeiro (RJ)	343	1.957.299	623.716	1.333.583	75,8
São Paulo (SP) e Maringá (PR)	350	4.749.440	3.784.103	965.337	55,7
Pombal (PB) e João Pessoa (PB)	548	669.173	154.215	514.958	81,3
Maringá (PR) e Curitiba (PR)	326	511.213	3.763	507.450	99,3
Rio de Janeiro (RJ) e Belo Horizonte (MG)	351	553.252	48.764	504.488	91,9
Porto Velho (RO) e Ji-Paraná (RO)	335	683.381	184.167	499.214	78,8
Serra Talhada (PE) e Recife (PE)	313	501.401	7.948	493.453	98,4
Teresina (PI) e Sobral (CE)	375	477.240	2.754	474.486	99,4
São Paulo (SP) e Barretos (SP)	314	463.852	1.453	462.399	99,7
São Paulo (SP) e Ourinhos (SP)	393	476.438	16.799	459.639	96,6
Porangatu (GO) e Goiânia (GO)	343	446.996	4.870	442.126	98,9
São Paulo (SP) e São José do Rio Preto (SP)	359	435.310	95	435.215	99,9
Juazeiro do Norte (CE) e Fortaleza (CE)	421	570.419	148.134	422.285	79,4
Pau dos Ferros (RN) e Natal (RN)	396	540.391	123.322	417.069	81,4
Teixeira de Freitas (BA) e Itabuna (BA)	329	383.417	438	382.979	99,9
Rorainópolis (RR) e Manaus (AM)	310	373.371	321	373.050	99,9
São Paulo (SP) e Marília (SP)	455	352.976	-	352.976	100,0
Montes Claros (MG) e Belo Horizonte (MG)	378	367.899	18.245	349.654	95,3
São Paulo (SP) e Belo Horizonte (MG)	358	389.697	43828	345.869	89,9
Porangatu (GO) e Palmas (TO)	487	1.047.138	706028	341.110	59,7

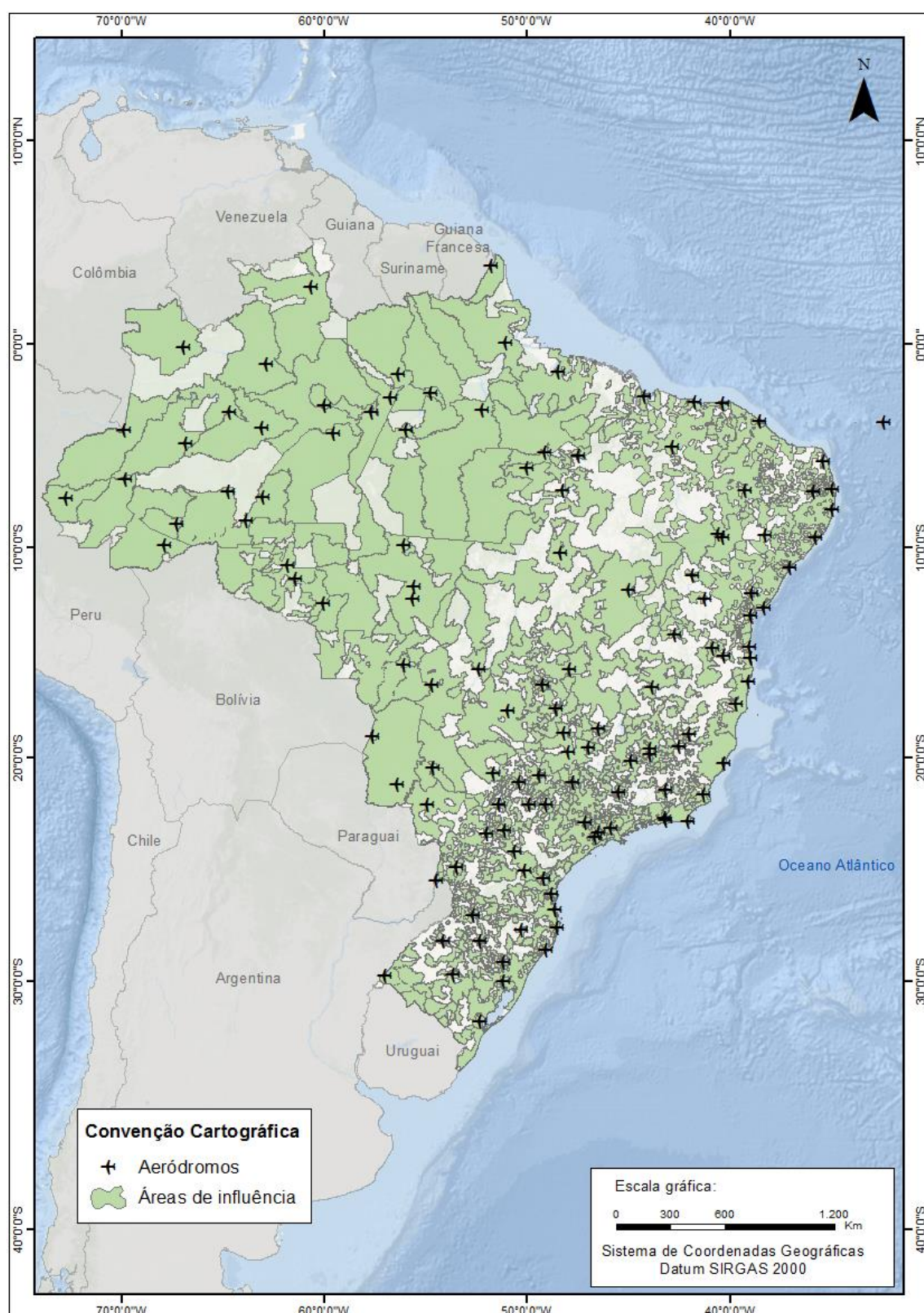
Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

## 2.4. Áreas de influência

Esta seção apresenta a delimitação da área de influência dos nove aeroportos brasileiros mais movimentados e de alguns aeroportos regionais. Para cada aeroporto, são analisados os municípios de origem dos passageiros embarcados e os volumes de deslocamentos gerados a partir deles. Analogamente são analisados os municípios de destino dos passageiros desembarcados e os volumes destinados a estes. Com isso, é possível delimitar as áreas de influência dos aeródromos através dos deslocamentos observados. A Figura 22 ilustra as áreas de influência dos aeroportos considerados em relação aos municípios com mais de 300 viagens por ano.



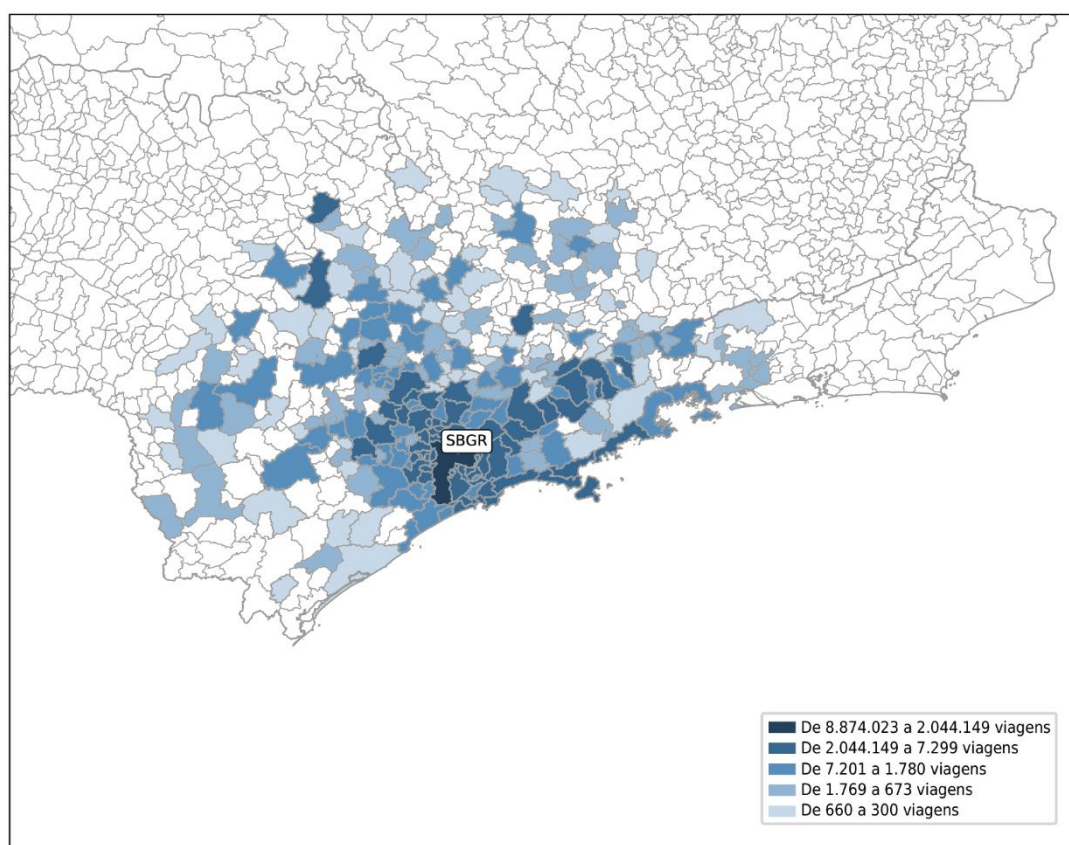
Figura 22 – Áreas de influência dos aeroportos considerados



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

De acordo com os dados disponíveis no Sistema Hórus, no ano de 2019 o Aeroporto de Guarulhos foi o que teve o maior fluxo de passageiros domésticos embarcados e desembarcados. Devido à sua localização dentro da região metropolitana de São Paulo e com acesso pelas rodovias Presidente Dutra e Ayrton Senna, sua área de influência se estende até os estados de Minas gerais e Rio de Janeiro - Considerando os municípios que tiveram mais de 300 viagens ao ano. A Figura 23 ilustra a área de influência do Aeroporto de Guarulhos.

**Figura 23 – Área de influência do Aeroporto de Guarulhos (SBGR) – Municípios com mais de 300 viagens/ano**

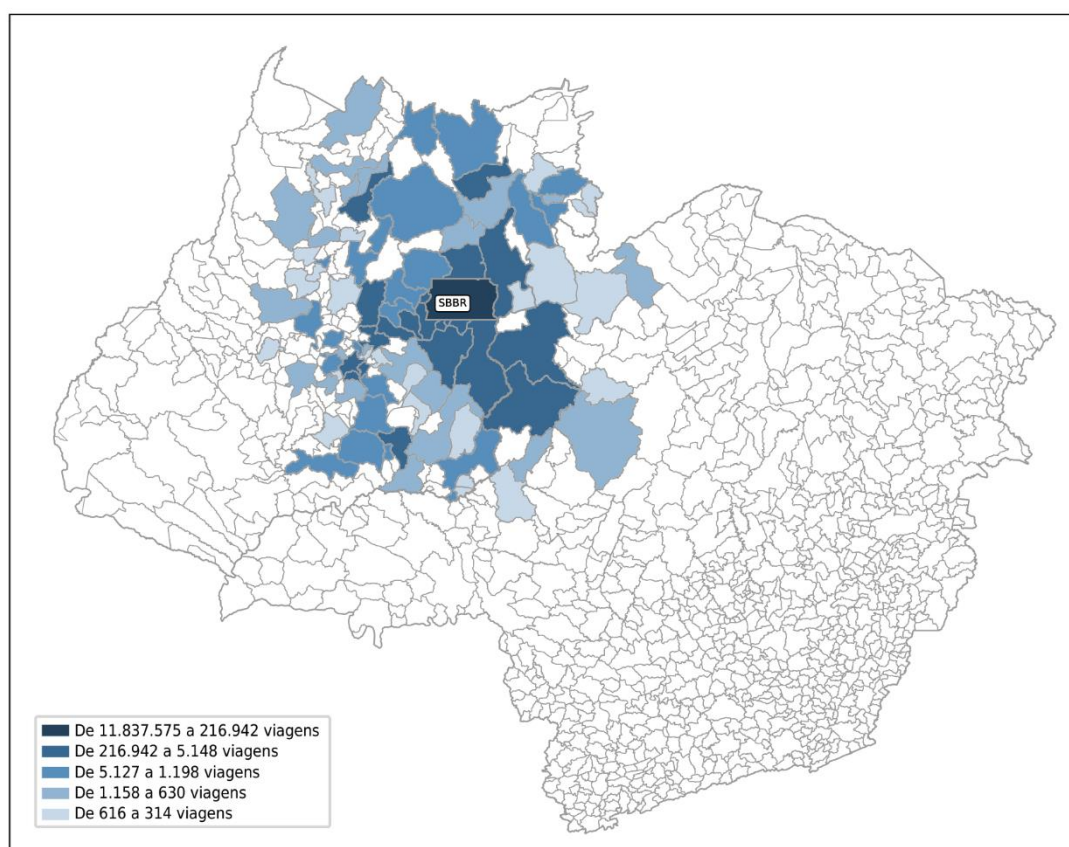


Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O Aeroporto de Brasília, em 2019, foi o terceiro com maior número de passageiros movimentados no transporte aéreo doméstico, de acordo com o Sistema Hórus (BRASIL, 2020). Vale destacar que o segundo município com maior número de viagens embarcadas ou desembarcadas no Aeroporto de Brasília foi Goiânia (GO), provavelmente pela maior oferta de voos nesse aeroporto. A Figura 24 retrata a sua área de influência.



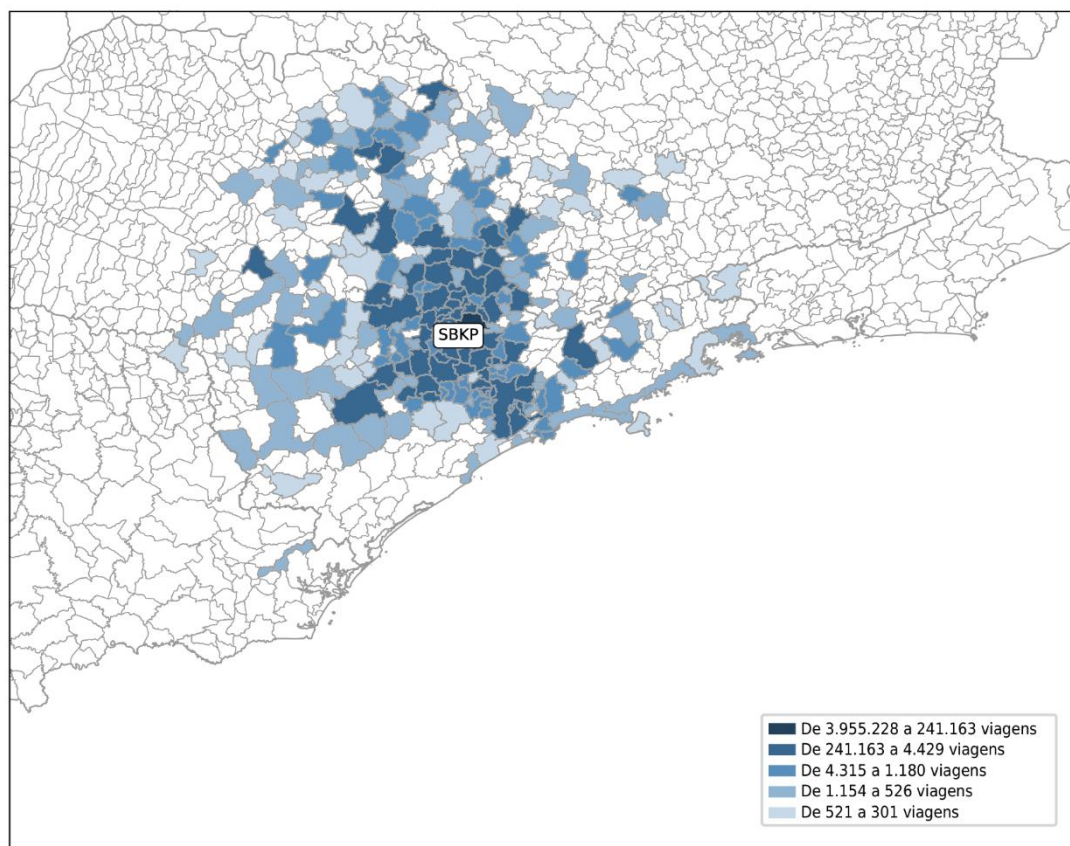
Figura 24 – Área de influência do Aeroporto de Brasília (SBBR) – Municípios com mais de 300 viagens/ano



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O aeroporto de Campinas é o quinto com maior movimentação de passageiros doméstico do País, principalmente por ser o *hub* da companhia Azul Linhas Aéreas (BRASIL, 2020). Devido à sua localização, parte de sua área de influência é sobreposta a do Aeroporto de Guarulhos, conforme dados obtidos com a Telefônica. A Figura 25 ilustra a sua área de influência.

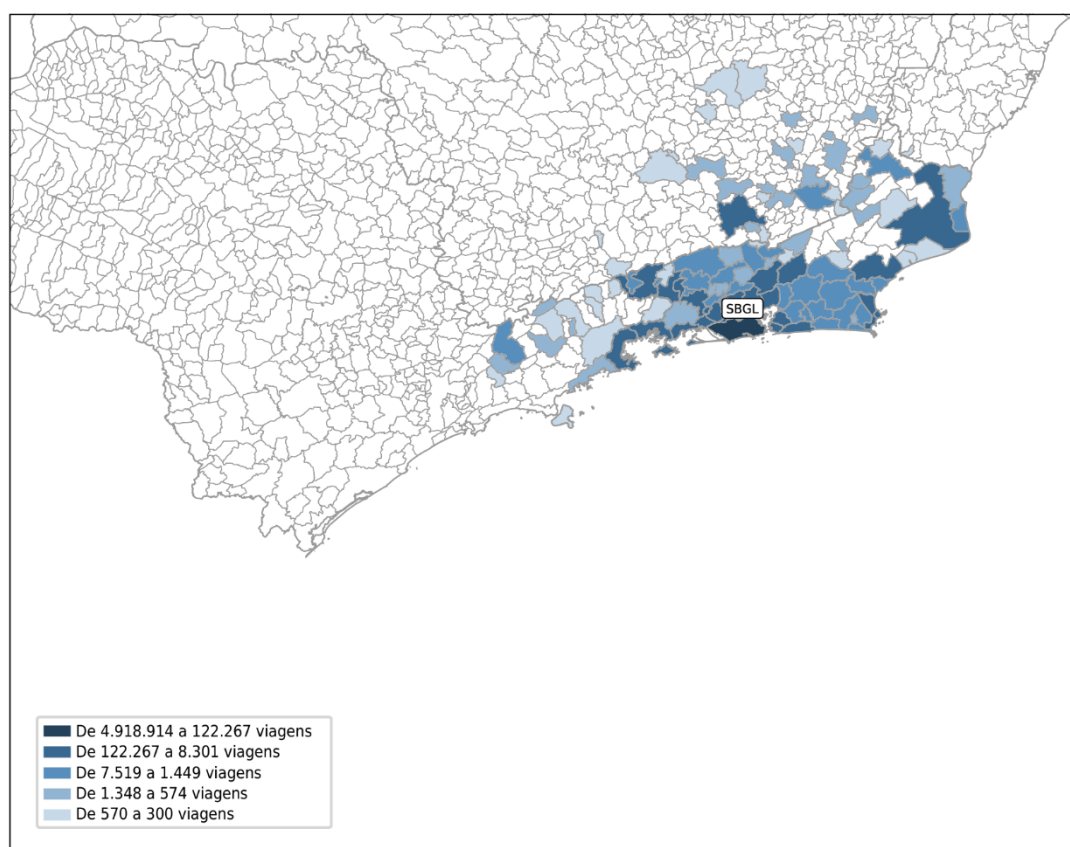
**Figura 25 – Área de influência do Aeroporto de Campinas (SBKP) –  
Municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

De acordo com os dados disponíveis no Sistema Hórus (BRASIL, 2020), o aeroporto do Galeão é o sexto em termos de volume movimentado de passageiros domésticos. Sua área de influência se estende do norte de São Paulo ao leste de Minas Gerais – considerando os municípios registrados com mais de 300 viagens por ano. A Figura 26 retrata a área de influência do aeroporto.

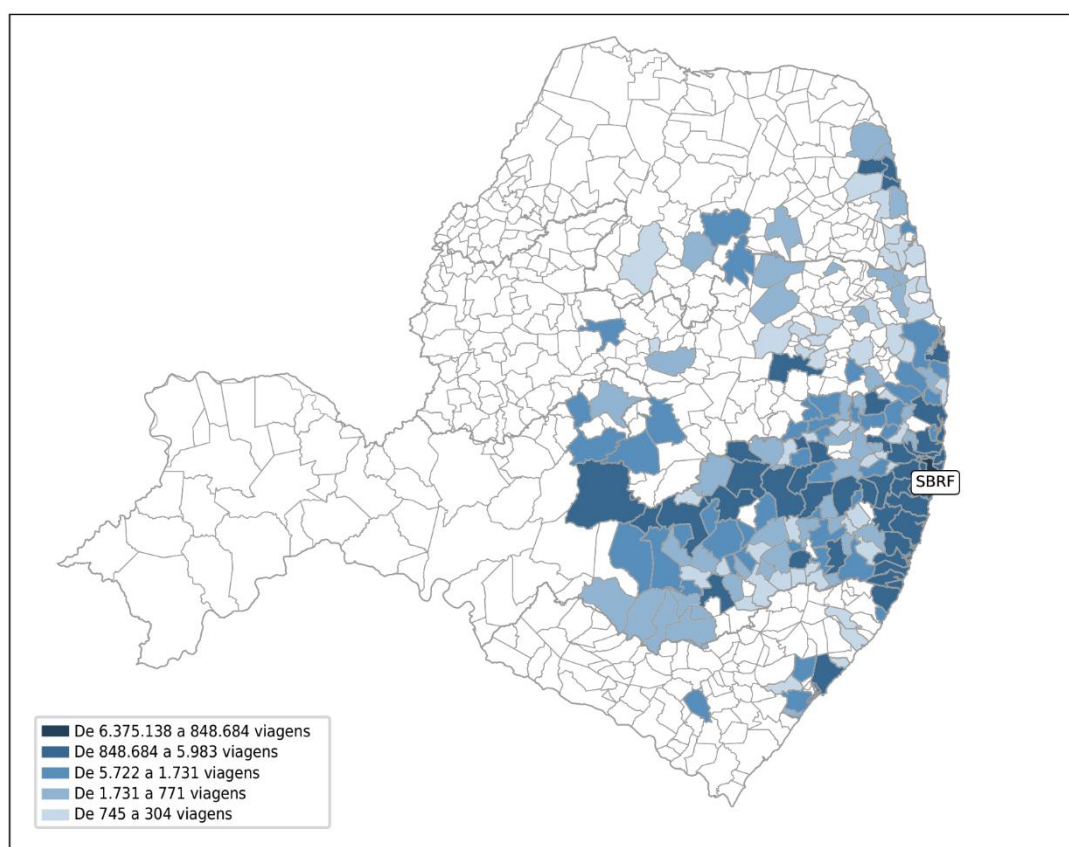
**Figura 26 – Área de influência do Aeroporto do Galeão (SBGL) – municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O oitavo aeroporto em termos de movimentação de passageiros doméstico e o maior do Nordeste do país é o Aeroporto de Recife (BRASIL, 2020). Atualmente, o aeroporto atua como *hub* regional da Azul Linhas Aéreas e da Latam. A Figura 27 retrata a sua área de influência.

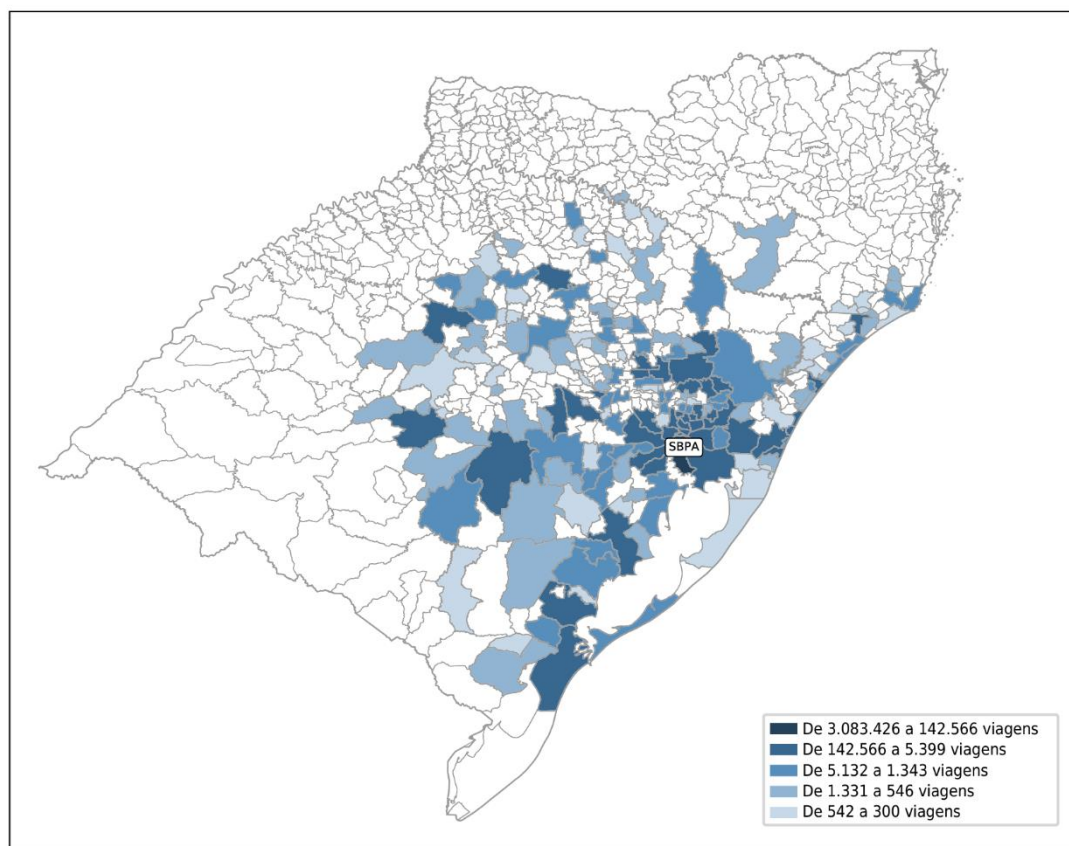
**Figura 27 – Área de influência do Aeroporto de Recife (SBRF) – municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O Aeroporto de Porto Alegre é o nono em número de passageiros domésticos transportados e o maior e mais movimentado do Rio Grande do Sul (BRASIL, 2020). Sua área de influência se estende desde o sul do estado de Santa Catarina, até o sul do Rio Grande do Sul, conforme ilustrado na Figura 28.

**Figura 28 – Área de influência do Aeroporto de Porto Alegre (SBPA) – municípios com mais de 300 viagens/ano**

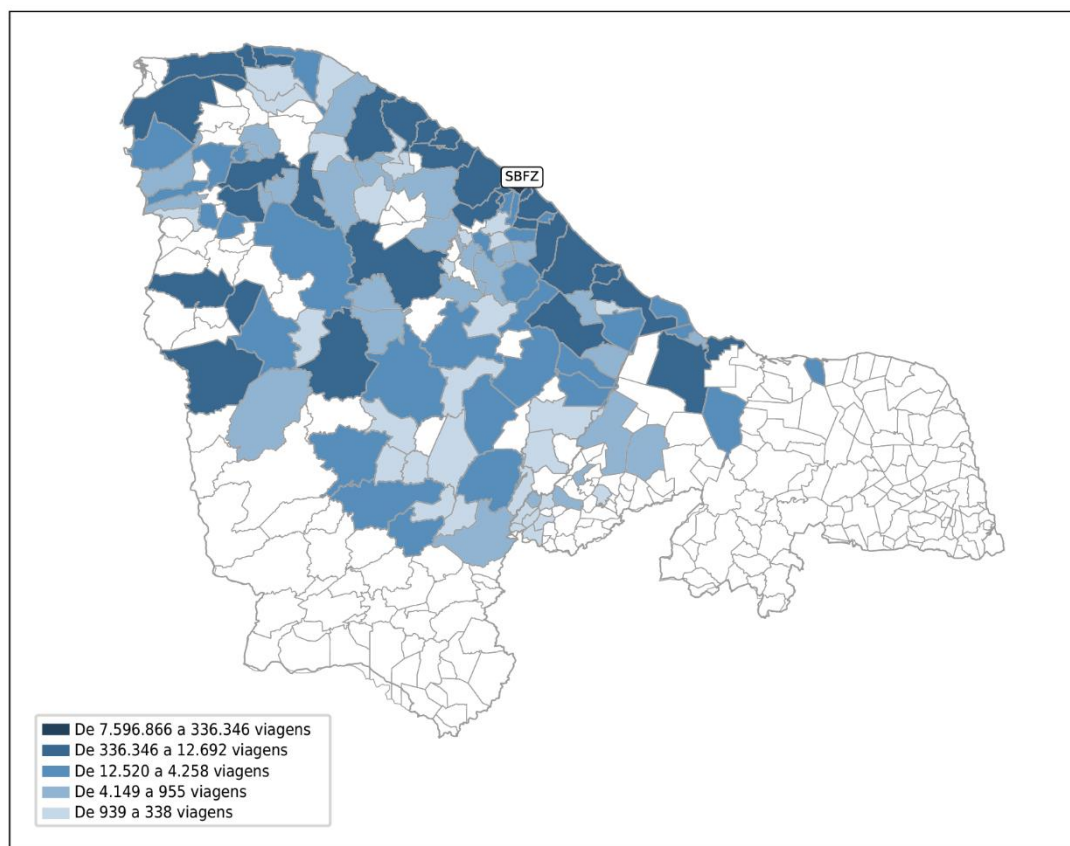


Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

O Aeroporto de Fortaleza é o terceiro aeroporto mais importante do nordeste (BRASIL, 2020). Atua como *hub* regional das operações da Gol Linhas Aéreas. A Figura 29 ilustra sua área de influência, considerando os aeroportos registrados com mais de 300 viagens por ano.



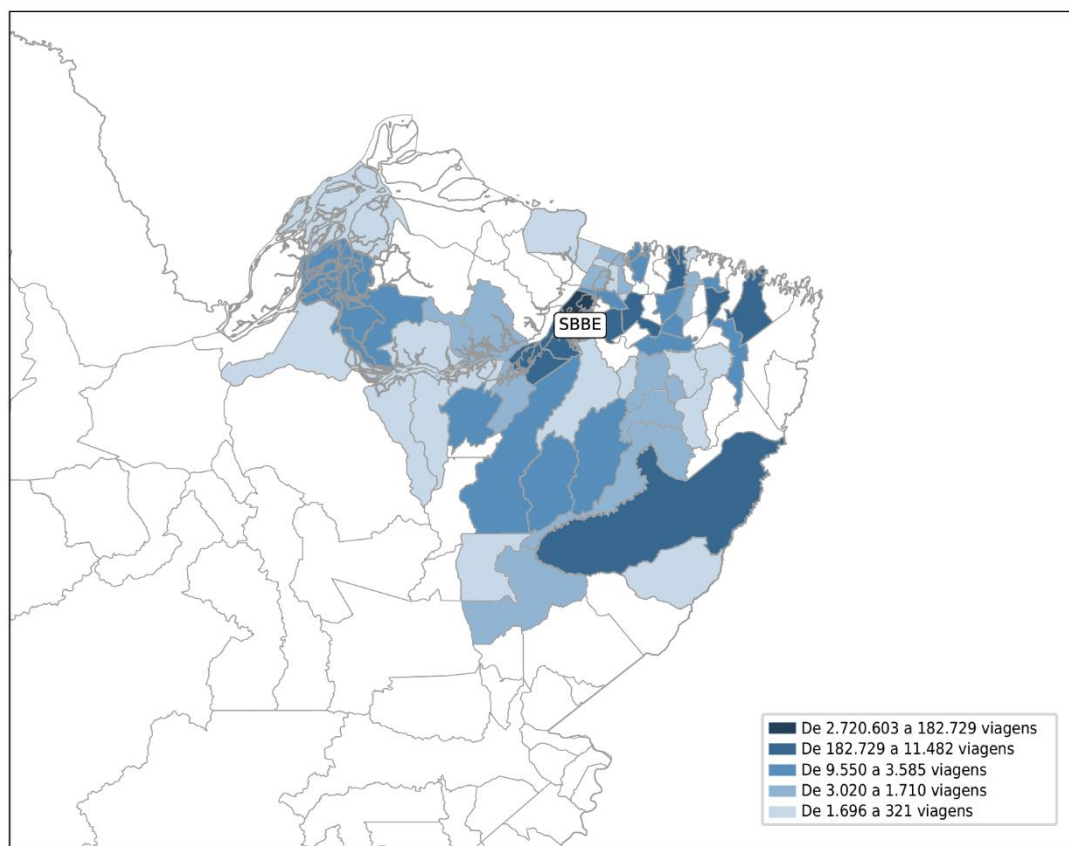
**Figura 29 – Área de influência do Aeroporto de Fortaleza (SBFZ) – municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

De acordo com os dados disponíveis no Sistema Hórus (2020), o Aeroporto de Belém é o aeroporto mais importante do norte do País. A Figura 30 retrata a sua área de influência.

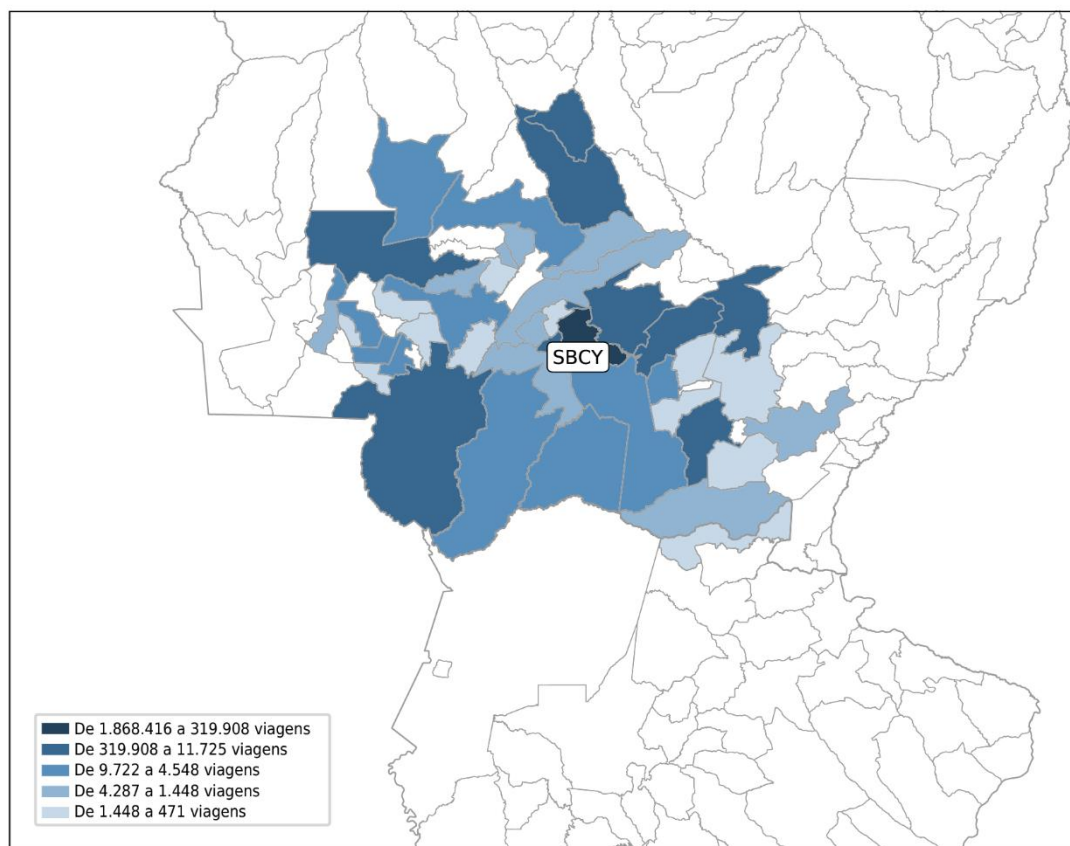
**Figura 30 – Área de influência do Aeroporto de Belém (SBBE) – municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

Localizado no município de Várzea Grande (MT), o Aeroporto de Cuiabá é o mais importante do estado e o terceiro da Região Centro-Oeste, segundo o Sistema Hórus (2020). A Figura 31 apresenta a área de influência do aeroporto, considerando os dados obtidos com a Telefônica.

**Figura 31 – Área de influência do Aeroporto de Cuiabá (SBCY) – municípios com mais de 300 viagens/ano**



Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



## CONSIDERAÇÕES FINAIS

Os resultados apresentados neste documento fornecem subsídio técnico para o planejamento no setor aéreo, em atendimento ao escopo definido para a Fase 3 do Plano de Trabalho apresentado pelo LabTrans/UFSC à SAC/MInfra, relativo à elaboração de uma matriz Origem/Destino (OD) de passageiros em âmbito nacional para identificação dos reais desejos de viagens contemplados ou não pelo transporte aéreo.

Do ponto de vista metodológico, inicialmente, foram avaliadas as premissas e critérios adotados pela vencedora na Seleção Pública nº 105/2018. Após constante interação da equipe técnica da empresa contratada com o pessoal técnico do LabTrans/UFSC e diversas análises realizadas, chegou-se ao máximo de aprimoramento possível da base de dados para o levantamento de matriz OD de passageiros em âmbito nacional para identificação dos reais desejos de viagens contemplados ou não pelo transporte aéreo. Em seguida, foram estimados o volume de viagens entre pares OD identificados que não possuíam cobertura de telefonia móvel pela contratada.

Em relação aos resultados gerais, foram avaliados os principais deslocamentos para o Modo 1 (aéreo) e para o Modo 0 (não aéreo ou terrestre), além de uma avaliação sobre rotas potenciais considerando os deslocamentos intrarregiões por estado, bem como a demanda reprimida. Além disso foram apresentadas as áreas de influência dos principais aeródromos brasileiros.

Destaca-se que os resultados obtidos a partir deste estudo podem servir de apoio a tomada de decisão em investimentos no setor, uma vez que a utilização de *big data* traz ganhos de valor agregado, podendo servir, ainda, como ponto de partida para estudos posteriores.



# REFERÊNCIAS

- AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (ANAC). **Dados Estatísticos**. [Brasília, DF], 2019a. Disponível em: <https://www.anac.gov.br/assuntos/dados-e-estatisticas/dados-estatisticos/dados-estatisticos>. Acesso em: 17 maio 2019.
- AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (ANAC). **Lista de aeródromos públicos registrados**. [Brasília, DF], [2016]. Disponível em: <http://www.anac.gov.br/assuntos/dados-e-estatisticas/aeroportos>. Acesso restrito.
- AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (ANAC). **Microdados de tarifas aéreas comercializadas**. [Brasília], 23 jul. 2019b. Disponível em: <https://www.anac.gov.br/assuntos/dados-e-estatisticas/microdados-de-tarifas-aereas-comercializadas>. Acesso em: 13 dez. 2019.
- AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (ANAC). **SAS - Superintendência de Acompanhamento dos Serviços Aéreos**. [Brasília, DF], c2020. Disponível em: <https://sas.anac.gov.br/sas/downloads/view/frmDownload.aspx>. Acesso em: 29 nov. 2019.
- AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (ANATEL). **Acessos de telefonia móvel no Brasil**. [Brasília, DF], 1 out. 2019. Disponível em: <http://www.anatel.gov.br/dados/destaque-1/283-brasil-tem-236-2-milhoes-de-linhas-moveis-em-janeiro-de-2018>. Acesso em: 12 fev. 2020.
- AGÊNCIA NACIONAL DE TELECOMUNICAÇÕES (ANATEL). **Telefonia Móvel - Municípios atendidos**. [Brasília, DF], 16 jan. 2020. Disponível em: <https://www.anatel.gov.br/setorregulado/component/content/article/115-universalizacao-e-ampliacao-do-acesso/telefonia-movel/423-telefonia-movel-municipios-atendidos>. Acesso em: 12 nov. 2019.
- AGÊNCIA NACIONAL DE TRANSPORTES TERRESTRES (ANTT). **Estatísticas e Estudos Rodoviários - Dados Operacionais**. Brasília, DF, [201-]. Disponível em: [http://www.antt.gov.br/passageiros/Dados\\_Operacionais](http://www.antt.gov.br/passageiros/Dados_Operacionais). Acesso em: 29 nov. 2019.
- BARBETTA, P. A.; REIS, M. M.; BORNIA, A. C. **Estatística para Cursos de Engenharia e Informática**. 3. ed. São Paulo: Atlas, 2010. 416 p.
- BRASIL. Ministério da Economia. **Relação Anual de Informações Sociais (RAIS)**. Brasília, DF, [2019]. Disponível em: <http://www.rais.gov.br/sitio/index.jsf>. Acesso em: 14 jun. 2019.
- BRASIL. Ministério da Fazenda. **Greve dos caminhoneiros impacta a economia em cerca de R\$ 15,9 bilhões**. [Brasília, DF], 14 jun. 2018a. Disponível em: <http://www.fazenda.gov.br/noticias/2018/junho/greve-dos-caminhoneiros-impacta-a-economia-em-cerca-de-r-15-9-bilhoes>. Acesso em: 5 dez. 2019.
- BRASIL. Ministério da Infraestrutura (MInfra). Secretaria Nacional de Aviação Civil (SAC). **Sistema Hórus**. Versão 9.5.6.269. Brasília, DF, 2019. Disponível em: <https://horus.labtrans.ufsc.br/gerencial/#Principal>. Acesso em: 24 abr. 2019.
- BRASIL. Ministério dos Transportes, Portos e Aviação Civil. Departamento de Planejamento e Gestão Portuária. **Plano Aeroviário Nacional 2018-2038: objetivos, estratégias e investimentos para desenvolvimento do transporte aéreo brasileiro**. Brasília, DF: MTPA, nov. 2018b. Disponível em:

[http://www.infraestrutura.gov.br/images/AVIACAO\\_CIVIL/PAN/PAN2018\\_ebook.pdf](http://www.infraestrutura.gov.br/images/AVIACAO_CIVIL/PAN/PAN2018_ebook.pdf). Acesso em: 4 dez. 2018.

BRASIL. Ministério dos Transportes, Portos e Aviação Civil. Departamento de Planejamento e Gestão Portuária. **Projeções de demanda para os aeroportos brasileiros 2017-2037: metodologia e resultados**. [Brasília, DF]: MTPA, set. 2017a. Disponível em: [http://www.transportes.gov.br/images/AEROPORTOS/Relatorio\\_Projecoes\\_2017\\_2037\\_Versao\\_final.pdf](http://www.transportes.gov.br/images/AEROPORTOS/Relatorio_Projecoes_2017_2037_Versao_final.pdf). Acesso em: 26 nov. 2018.

BRASIL. Ministério do Turismo. **Anuário Estatístico de Turismo 2018 - Ano Base 2017**. Brasília, DF, jul. 2018c. Disponível em: <http://www.dadosefatos.turismo.gov.br/2016-02-04-11-53-05.html>. Acesso em: 14 jun. 2019.

BRASIL. Ministério do Turismo. **Categorização dos municípios das regiões turísticas do mapa do turismo brasileiro**. Brasília, DF: Ministério do Turismo, 2015. Disponível em: [http://www.mapa.turismo.gov.br/mapa/downloads/pdf/categorizacao/Cartilha\\_da\\_Categorizacao.pdf](http://www.mapa.turismo.gov.br/mapa/downloads/pdf/categorizacao/Cartilha_da_Categorizacao.pdf). Acesso em: 14 jun. 2019.

BRASIL. Ministério do Turismo. **Estudo da Demanda Turística Internacional 2017**. Brasília, DF, maio 2018d. Disponível em: [dadosefatos.turismo.gov.br/2016-02-04-11-54-03/demanda-turistica-internacional.html](http://www.dadosefatos.turismo.gov.br/2016-02-04-11-54-03/demanda-turistica-internacional.html). Acesso em: 14 jun. 2019.

BRASIL. Ministério do Turismo. **Programa de Regionalização do Turismo**. [Brasília, DF], 16 mar. 2017b. Disponível em: <http://www.turismo.gov.br/aceso-a-informacao/63-aco-es-e-programas/4882-programa-de-regionalizacao-do-turismo.html>. Acesso em: 11 fev. 2020.

BRASIL. Presidência da República. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018e. Disponível em: <http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=515&pagina=59&data=15/08/2018>. Acesso em: 30 mar. 2019.

CHEN, C. L. P.; ZHANG, C. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. **Information Sciences**, Macau, v. 275, p. 314-347, Aug. 2014. Disponível em: [https://www.researchgate.net/publication/262305146\\_Data-intensive\\_applications\\_challenges\\_techniques\\_and\\_technologies\\_A\\_survey\\_on\\_Big\\_Data](https://www.researchgate.net/publication/262305146_Data-intensive_applications_challenges_techniques_and_technologies_A_survey_on_Big_Data). Acesso em: 23 nov. 2018.

DEMCHENKO, Y.; LAAT, C.; MEMBREY, P. Defining architecture components of the Big Data Ecosystem. In: INTERNATIONAL CONFERENCE ON COLLABORATION TECHNOLOGIES AND SYSTEMS (CTS), 2., 2014, Minneapolis. **Annals [...]**. Minneapolis: IEEE, 2014.

DJUKIC, T.; VAN LINT, J. W. C.; HOOGENDOORN, S. P. Application of Principal Component Analysis to Predict Dynamic Origin–Destination Matrices. **Transportation Research Record**, [s. l.], v. 2283, n. 1, p. 81-89, 2012.

EMPRESA DE PLANEJAMENTO E LOGÍSTICA S.A. (EPL). **Pesquisa de origem/destino do transporte aéreo de passageiros – relatório técnico**: pesquisa de origem/destino do transporte aéreo de passageiros. Brasília, DF: EPL, 2014. Disponível em: <https://www.epl.gov.br/pesquisa-de-origem-e-destino-nos-aeroportos>. Acesso em: 18 set. 2018.

FUNDAÇÃO DE AMPARO À PESQUISA E EXTENSÃO UNIVERSITÁRIA (FAPEU). **Anexo I – Termo de Referência:** Seleção Pública de Fornecedores nº. 105/2018 Processo nº. 331271729. [Florianópolis]: FAPEU, 2018. 18 p. [.pdf].

GARCÍA, P. *et al.* Big Data Analytics for a Passenger-Centric Air Traffic Management System: A Case Study of Door-to-Door Intermodal Passenger Journey Inferred from Mobile Phone Data. *In: SESAR INNOVATION DAYS*, 6., 2016, Netherlands. **Annals** [...]. Netherlands: Technical University of Delft, 2016.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2007. 220 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Arranjos populacionais e concentrações urbanas**. 2. ed. [Rio de Janeiro]: IBGE, 2016. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv99700.pdf>. Acesso em: 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Estimativas de população 2019**. Rio de Janeiro, 2019. Disponível em: <https://sidra.ibge.gov.br/Tabela/6579>. Acesso em: 14 jun. 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Produto Interno Bruto dos Municípios**. Rio de Janeiro, 2017. Disponível em: <https://sidra.ibge.gov.br/Tabela/5938>. Acesso em: 14 jun. 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Regiões de Influência das Cidades – REGIC**. O que é. Rio de Janeiro, [201-]. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/redes-geograficas/15798-regioes-de-influencia-das-cidades?=&t=o-que-e>. Acesso em: 6 fev. 2020.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Microdados**. Censo da Educação Superior - 2017. Brasília, DF: INEP, 2019a. Disponível em: <http://portal.inep.gov.br/web/guest/microdados#>. Acesso em: 14 jun. 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Microdados**. Censo Escolar - 2017. Brasília, DF: INEP, 2019b. Disponível em: <http://portal.inep.gov.br/web/guest/microdados#>. Acesso em: 14 jun. 2019.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. **META Group**, 6 Feb. 2001. Disponível em: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Acesso em: 23 nov. 2018.

LUCA. **Crowd analytics**. [Madrid], [201-]a. Disponível em: <https://luca-d3.com/products-services/business-insights/crowd-analytics>. Acesso em: 16 dez. 2019.

LUCA. **Success stories**. [Madrid], [201-]b. Disponível em: <https://luca-d3.com/artificial-intelligence-examples>. Acesso em: 16 dez. 2019.

SABIN, J. G.; FERRÃO, M. F.; FURTADO, J. C. Análise multivariada aplicada na identificação de fármacos antidepressivos. Parte II: Análise por componentes principais (PCA) e o método de classificação SIMCA. **Revista Brasileira de Ciências Farmacêuticas**, São Paulo, v. 40, n. 3, p. 387-396, 2004.

SÃO PAULO. Assembleia Legislativa do Estado de São Paulo. **Lei Complementar nº 1.139, de 16 de junho de 2011**. Reorganiza a Região Metropolitana da Grande São Paulo, cria o respectivo Conselho de Desenvolvimento e dá providências correlatas. São Paulo: ALESP, 16 jun. 2011. Disponível em:

<https://www.al.sp.gov.br/repositorio/legislacao/lei.complementar/2011/lei.complementar-1139-16.06.2011.html>. Acesso em: 14 fev. 2020.

SUN, Z. et al. Traffic flow forecast based on combining principal component analysis with support vector machine. **Journal of Highway and Transportation Research and Development**, Beijing, v. 26, n. 5, p. 127-131, 2009.

TELECO INTELIGÊNCIA EM TELECOMUNICAÇÕES (TELECO). **Estatísticas telefonia celular móvel**. [São José dos Campos], dez. 2019. Disponível em: . Acesso em: 7 fev. 2020.

TELEFÔNICA. **Relatório de Metodologia**: desenvolvimento de uma matriz origem/destino de passageiros em âmbito nacional para identificação dos desejos de viagem contemplados ou não pelo transporte aéreo, utilizando como base as unidades de planejamento territorial. [Madrid]: 2018a. 27 p. [.pdf].

TELEFÔNICA. **Relatório descritivo do método de geração da base de dados**: desenvolvimento de uma matriz origem/destino de passageiros em âmbito nacional para identificação dos desejos de viagem contemplados ou não pelo transporte aéreo, utilizando como base as unidades de planejamento territorial. [Madrid]: 2018b. 22 p. [.pdf].

TRIOLA, M. F. **Introdução à Estatística**. 9. ed. Rio de Janeiro: LTC, 2005.

UOL. **Greve dos caminhoneiros: a cronologia dos 10 dias que pararam o Brasil**. [São Paulo], 30 maio 2018. Disponível em: <https://economia.uol.com.br/noticias/bbc/2018/05/30/greve-dos-caminhoneiros-a-cronologia-dos-10-dias-que-pararam-o-brasil.htm>. Acesso em: 11 jan. 2020.

WANG, M. et al. Estimating Dynamic Origin-Destination Data and Travel Demand Using Cell Phone Network Data. **International Journal of Intelligent Transportation Systems Research**, New York, v. 11, p. 76-86, May 2013. Disponível em: [https://www.researchgate.net/publication/257797926\\_Estimating\\_Dynamic-Origin-Destination\\_Data\\_and\\_Travel\\_Demand\\_Using\\_Cell\\_Phone\\_Network\\_Data](https://www.researchgate.net/publication/257797926_Estimating_Dynamic-Origin-Destination_Data_and_Travel_Demand_Using_Cell_Phone_Network_Data). Acesso: 14 set. 2018.

WEBER, A. F.; PÉRSIGO, P. M. **Pesquisa de opinião pública**: princípios e exercícios. Santa Maria: FACOS-UFSM, 2017. *E-book*. Disponível em: <https://repositorio.ufsm.br/handle/1/13135>. Acesso em: 11 fev. 2020.

WHEELER, D. J. **Advanced Topics in Statistical Process Control**: the power of Shewhart's Charts. Knoxville, TN: SPC Press Inc., 1995.

ZOU, H; HASTIE, T.; TIBSHIRANI, R. Sparse principal component analysis. **Journal of computational and graphical statistics**, [s. l.], v. 15, n. 2, p. 265-286, 2006.

# LISTA DE FIGURAS

Figura 1 – Fluxograma das macroetapas de coleta de dados e elaboração da matriz OD .....	8
Figura 2 – Fluxograma da metodologia.....	9
Figura 3 – UTPs.....	11
Figura 4 – Disposição geográfica das torres de telefonia móvel pesquisa – Brasil.....	17
Figura 5 – Cobertura dos municípios com telefonia móvel da pesquisa – Brasil.....	18
Figura 6 – Diagrama de processo: base de zoneamento .....	20
Figura 7 – Parâmetros de uma célula de telefone celular .....	22
Figura 8 – Conceito de <i>flickering</i> como um instrumento para ajustar o alcance das células .....	24
Figura 9 – Diagrama de processo: base de usuários .....	26
Figura 10 – Diagrama de processo da base com fatores de extrapolação .....	30
Figura 11 – Informações possíveis de serem levantadas pela LUCA.....	31
Figura 12 – Aeródromos com registros de voo comercial em 2017 .....	34
Figura 13 – Diagrama de processo: base de viagens.....	35
Figura 14 – Exemplo: UTP Propriá (SE) e Aracaju (SE) .....	36
Figura 15 – Exemplo: UTP Feira de Santana (BA) e Cachoeira (BA) .....	37
Figura 16 – Exemplo representativo da Regra 3 .....	43
Figura 17 – Área de influência dos municípios de origem do Aeroporto SBCF no extrato 6 e 7 .....	47
Figura 18 – Principais fluxos de deslocamento: Modo 0 (terrestre) .....	70
Figura 19 – Principais fluxos de deslocamento: Modo 1 (aéreo).....	71
Figura 20 – Volume total de deslocamentos intra e inter-regionais: Modo 0 (terrestre) – Milhões de viagens.....	73
Figura 21 – Volume total de deslocamentos intra e inter-regionais: Modo 1 (aéreo) – Milhões de viagens.....	74
Figura 22 – Áreas de influência dos aeroportos considerados .....	89
Figura 23 – Área de influência do Aeroporto de Guarulhos (SBGR) – Municípios com mais de 300 viagens/ano.....	90



Figura 24 – Área de influência do Aeroporto de Brasília (SBBR) – Municípios com mais de 300 viagens/ano.....	91
Figura 25 – Área de influência do Aeroporto de Campinas (SBKP) – Municípios com mais de 300 viagens/ano.....	92
Figura 26 – Área de influência do Aeroporto do Galeão (SBGL) – municípios com mais de 300 viagens/ano.....	93
Figura 27 – Área de influência do Aeroporto de Recife (SBRF) – municípios com mais de 300 viagens/ano.....	94
Figura 28 – Área de influência do Aeroporto de Porto Alegre (SBPA) – municípios com mais de 300 viagens/ano.....	95
Figura 29 – Área de influência do Aeroporto de Fortaleza (SBFZ) – municípios com mais de 300 viagens/ano.....	96
Figura 30 – Área de influência do Aeroporto de Belém (SBBE) – municípios com mais de 300 viagens/ano.....	97
Figura 31 – Área de influência do Aeroporto de Cuiabá (SBCY) – municípios com mais de 300 viagens/ano.....	98

# LISTA DE GRÁFICO

Gráfico 1 – Dispersão das UTPs nas variáveis área (km <sup>2</sup> ) e densidade de antenas .....	38
Gráfico 2 – Distribuição das UTPs por intervalos de área total (km <sup>2</sup> ) .....	38
Gráfico 3 – Comparativo de movimentações mensais de 2017 do extrato com os dados da base estatística da ANAC .....	50
Gráfico 4 – Sazonalidade modo terrestre .....	52
Gráfico 5– Relação entre quantidade de viagens X distância (variáveis em logaritmo natural) – Modo 1 (aéreo) .....	57
Gráfico 6– Relação entre quantidade de viagens X população (variáveis em logaritmo natural) – Modo 1 (aéreo) .....	58
Gráfico 7 – Relação entre quantidade de viagens X PIB (variáveis em logaritmo natural) – Modo 1 (aéreo) .....	58
Gráfico 8 – Relação entre quantidade de viagens X distância (variáveis em logaritmo natural) – Modo 0 (terrestre) .....	59
Gráfico 9 – Relação entre quantidade de viagens X população (variáveis em logaritmo natural) – Modo 0 (terrestre) .....	59
Gráfico 10 – Relação entre quantidade de viagens X PIB (variáveis em logaritmo natural) – Modo 0 (terrestre) .....	60
Gráfico 11 – Participação dos componentes principais na variância total dos dados – Origem – Modo 1 (aéreo) .....	62
Gráfico 12 – Participação dos componentes principais na variância total dos dados – Destino – Modo 1 (aéreo) .....	63
Gráfico 13 – Participação dos componentes principais na variância total dos dados – Origem – Modo 0 (terrestre) .....	64
Gráfico 14 – Participação dos componentes principais na variância total dos dados – Destino – Modo 0 (terrestre) .....	64



# LISTA DE TABELAS

Tabela 1 – Participação de mercado por UF (2019) .....	19
Tabela 2 – Representatividade estatística da amostra por UF – Fev/2019 .....	28
Tabela 3 – Critérios de classificação do modo de transporte .....	33
Tabela 4 – Entregas da base de dados e suas versões de melhoramentos – 2019 .....	39
Tabela 5 – Bases de dados utilizadas no processo de validação dos dados da Telefônica .....	40
Tabela 6 – Regras de validação dos deslocamentos do Modo 1.....	42
Tabela 7 – Histórico da análise das versões do extrato parcial pelas regras de validação dos deslocamentos do Modo 1.....	43
Tabela 8 – Resumo das alterações nos parâmetros para as oito versões do extrato parcial .....	49
Tabela 9 – Grupos de análise .....	51
Tabela 10 – Variáveis utilizadas para a estimativa do modelo.....	60
Tabela 11 – Coeficientes estimados para o Modo 1 (aéreo) .....	66
Tabela 12 – Coeficientes estimados para o Modo 0 (terrestre) .....	66
Tabela 13 – Principais fluxos de deslocamento: Modo 0 (terrestre) e Modo 1 (aéreo) .....	72
Tabela 14 – Pares de UTPs com viagens terrestres superando em mais de 300 mil as viagens aéreas .....	86
Tabela 15 – Usuários descartados por UF .....	117
Tabela 16 – Os 50 principais fluxos de deslocamento: Modo 0 (terrestre) .....	121
Tabela 17 – Os 50 principais fluxos de deslocamento: Modo 1 (aéreo) .....	123



# LISTA DE SIGLAS

ACP	Análise de Componentes Principais
AI	Artificial intelligence
ANAC	Agência Nacional de Aviação Civil
Anatel	Agência Nacional de Telecomunicações
ANTAQ	Agência Nacional de Transportes Aquaviários
ANTT	Agência Nacional de Transportes Terrestres
CASK	Cost per Available Seat Kilometer
CL	Centros Locais
CR	Capitais Regionais
CSB	Centros Sub-Regionais
CZ	Centros de Zona
EPL	Empresa de Planejamento e Logística S.A.
ERB	Estação Rádio Base
FAPEU	Fundação de Amparo à Pesquisa e Extensão Universitária
GMN	Grande Metrópole Nacional
IBGE	Instituto Brasileiro de Geografia e Estatística
LabTrans	Laboratório de Transportes e Logística
LGPD	Lei Geral de Proteção de Dados Pessoais
M	Metrópoles
MIinfra	Ministério da Infraestrutura
MN	Metrópoles Nacionais
MPA	Ministério dos Transportes, Portos e Aviação Civil

OD	Origem/Destino
PAE	Plano Aeroviário Estadual
PAN	Plano Aeroviário Nacional
PIB	Produto Interno Bruto
PII	Personally identifiable information
RASK	Revenue per Available Seat Kilometer
REGIC	Regiões de Influência das Cidades
RIDE	Região Integrada de Desenvolvimento Econômico
SAC	Secretaria Nacional de Aviação Civil
SMP	Serviço Móvel Pessoal
SMS	Short Message Service
TR	Termo de Referência
UF	Unidade da Federação
UFSC	Universidade Federal de Santa Catarina
USP	Universidade de São Paulo
UTP	Unidade Territorial de Planejamento



## APÊNDICES E ANEXOS



## APÊNDICE 1– USUÁRIOS DESCARTADOS POR UF



**Tabela 15 – Usuários descartados por UF**

UF	% de usuários descartados
AC	8%
AL	20%
AM	6%
AP	7%
BA	17%
CE	24%
DF	14%
ES	8%
GO	16%
MA	15%
MG	12%
MS	10%
MT	10%
PA	11%
PB	17%
PE	22%
PI	18%
PR	22%
RJ	13%
RN	22%
RO	28%
RR	7%
RS	16%
SC	21%
SE	8%
SP	11%
TO	19%

Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



## APÊNDICE 2 – OS 50 PRINCIPAIS PARES DE DESLOCAMENTO PARA O MODO 0 E PARA O MODO 1





Tabela 16 – Os 50 principais fluxos de deslocamento: Modo 0 (terrestre)

#	UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens
1	São Paulo (SP) e Rio de Janeiro (RJ)	350	4.749.440
2	São Paulo (SP) e Curitiba (PR)	343	1.957.299
3	São Paulo (SP) e Belo Horizonte (MG)	487	1.047.138
4	São Paulo (SP) e Chapecó (SC)	724	779.945
5	Brasília (DF) e São Paulo (SP)	875	750.195
6	Rio de Janeiro (RJ) e Belo Horizonte (MG)	335	683.381
7	São Paulo (SP) e Maringá (PR)	548	669.157
8	São Paulo (SP) e Juazeiro do Norte (CE)	1.977	591.209
9	Recife (PE) e Pombal (PB)	352	587.557
10	São Paulo (SP) e São José do Rio Preto (SP)	421	570.419
11	Maringá (PR) e Curitiba (PR)	351	553.239
12	Juazeiro do Norte (CE) e Fortaleza (CE)	396	540.379
13	Pombal (PB) e João Pessoa (PB)	326	511.201
14	Rio de Contas (BA) e Salvador (BA)	366	507.829
15	Pombal (PB) e Natal (RN)	303	501.778
16	Porto Velho (RO) e Ji-Paraná (RO)	313	501.401
17	São Paulo (SP) e Florianópolis (SC)	489	499.061
18	Juazeiro do Norte (CE) e Caruaru (PE)	388	498.176
19	Xanxerê (SC) e Curitiba (PR)	351	486.703
20	Juazeiro do Norte (CE) e Arapiraca (AL)	407	484.782
21	São Paulo (SP) e Goiânia (GO)	816	480.824
22	Curitiba (PR) e Chapecó (SC)	382	478.775
23	São Paulo (SP) e Salvador (BA)	1.450	477.595
24	Serra Talhada (PE) e Recife (PE)	375	477.231
25	São Paulo (SP) e Barretos (SP)	393	476.438
26	Recife (PE) e Juazeiro do Norte (CE)	497	475.542
27	Toledo (PR) e Curitiba (PR)	457	472.567
28	São Paulo (SP) e Porto Alegre (RS)	854	468.530
29	Ituaçu (BA) e Salvador (BA)	318	467.174
30	São Luís (MA) e Imperatriz (MA)	485	465.100
31	Brasília (DF) e Rio de Janeiro (RJ)	927	465.038
32	Caraií (MG) e Belo Horizonte (MG)	385	464.132
33	Teresina (PI) e Sobral (CE)	314	463.838
34	Picos (PI) e Fortaleza (CE)	491	458.381
35	Pombal (PB) e Fortaleza (CE)	350	457.373
36	Guanambi (BA) e Salvador (BA)	483	455.266
37	Imperatriz (MA) e Belém (PA)	466	450.700

#	UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens
38	São Paulo (SP) e Ourinhos (SP)	343	446.996
39	Itapetinga (BA) e Salvador (BA)	312	443.767
40	Presidente Dutra (MA) e São Luís (MA)	307	443.552
41	Francisco Beltrão (PR) e Curitiba (PR)	386	442.238
42	Porto Alegre (RS) e Chapecó (SC)	352	441.308
43	Goiânia (GO) e Araguaína (TO)	1.060	436.236
44	Porangatu (GO) e Goiânia (GO)	359	435.298
45	Redenção (CE) e Juazeiro do Norte (CE)	339	433.812
46	Maringá (PR) e Campo Grande (MS)	430	432.729
47	Maringá (PR) e Foz do Iguaçu (PR)	357	432.457
48	Souto Soares (BA) e Salvador (BA)	358	418.578
49	Santa Inês (MA) e Imperatriz (MA)	312	417.138
50	São Paulo (SP) e Recife (PE)	2.126	416.132

Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)

Tabela 17 – Os 50 principais fluxos de deslocamento: Modo 1 (aéreo)

#	UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens
1	São Paulo (SP) e Rio de Janeiro (RJ)	350	3.935.471
2	São Paulo (SP) e Recife (PE)	2.126	3.736.761
3	São Paulo (SP) e Fortaleza (CE)	2.367	3.187.609
4	São Paulo (SP) e Salvador (BA)	1.450	2.958.058
5	Brasília (DF) e São Paulo (SP)	875	2.084.216
6	São Paulo (SP) e Porto Alegre (RS)	854	1.510.515
7	São Paulo (SP) e Natal (RN)	2.322	1.396.138
8	Brasília (DF) e Rio de Janeiro (RJ)	927	1.334.503
9	Rio de Janeiro (RJ) e Fortaleza (CE)	2.187	1.192.879
10	Rio de Janeiro (RJ) e Recife (PE)	1.873	1.165.186
11	São Paulo (SP) e Maceió (AL)	1.930	1.106.445
12	São Paulo (SP) e Goiânia (GO)	816	1.069.861
13	Salvador (BA) e Rio de Janeiro (RJ)	1.205	932.073
14	Brasília (DF) e Fortaleza (CE)	1.684	883.479
15	São Paulo (SP) e Cuiabá (MT)	1.334	835.975
16	Brasília (DF) e Recife (PE)	1.654	779.423
17	São Paulo (SP) e João Pessoa (PB)	2.212	747.191
18	São Paulo (SP) e Belo Horizonte (MG)	487	746.070
19	São Paulo (SP) e Belém (PA)	2.467	709.386
20	São Paulo (SP) e Curitiba (PR)	343	674.154
21	Rio de Janeiro (RJ) e Porto Alegre (RS)	1.125	656.268
22	São Paulo (SP) e Manaus (AM)	2.693	642.962
23	São Paulo (SP) e Florianópolis (SC)	489	632.835
24	Brasília (DF) e Salvador (BA)	1.058	622.404
25	São Paulo (SP) e São Luís (MA)	2.352	603.415
26	Serra (ES) e São Paulo (SP)	752	569.142
27	Recife (PE) e Fortaleza (CE)	633	558.573
28	Teresina (PI) e São Paulo (SP)	2.094	547.986
29	Manaus (AM) e Fortaleza (CE)	2.380	535.539
30	Salvador (BA) e Recife (PE)	677	531.188
31	São Paulo (SP) e Aracaju (SE)	1.732	522.784
32	Rio de Janeiro (RJ) e Natal (RN)	2.088	500.630
33	São Paulo (SP) e Campo Grande (MS)	898	495.941
34	Salvador (BA) e Fortaleza (CE)	1.033	458.744
35	Brasília (DF) e Natal (RN)	1.774	424.904
36	Rio de Janeiro (RJ) e Curitiba (PR)	676	399.682
37	Fortaleza (CE) e Belém (PA)	1.128	379.498
38	Recife (PE) e Belo Horizonte (MG)	1.640	378.064

#	UTPs envolvidas (ambos os sentidos)	Distância (em linha reta) entre municípios sede (km)	Quantidade de viagens
39	São Paulo (SP) e Foz do Iguaçu (PR)	840	361.027
40	Brasília (DF) e Belém (PA)	1.595	357.377
41	Serra (ES) e Rio de Janeiro (RJ)	430	354.027
42	Brasília (DF) e Manaus (AM)	1.935	348.707
43	Brasília (DF) e Maceió (AL)	1.487	343.485
44	Brasília (DF) e Cuiabá (MT)	877	331.868
45	Rio de Janeiro (RJ) e Belém (PA)	2.448	327.203
46	Brasília (DF) e Curitiba (PR)	1.081	323.483
47	São Paulo (SP) e Porto Seguro (BA)	1.114	322.536
48	Rio de Janeiro (RJ) e João Pessoa (PB)	1.964	320.525
49	Brasília (DF) e João Pessoa (PB)	1.712	310.316
50	São Paulo (SP) e Juazeiro do Norte (CE)	1.977	301.342

Fonte: Dados obtidos com a Telefônica. Elaboração: LabTrans/UFSC (2020)



