

Instituto de Pesquisas Tecnológicas do Estado de São Paulo

Eder Mauricio Barbosa

**Aplicação de heurísticas spaçotemporais na análise de Registros
de Detalhes de Chamadas (CDR) para a construção de base de
dados de deslocamento urbano**

**São Paulo
2024**

Eder Mauricio Barbosa

**Aplicação de heurísticas espacotemporais na análise de
Registros de Detalhes de Chamadas (CDR) para a
construção de base de dados de deslocamento urbano**

2024

Eder Mauricio Barbosa

Aplicação de heurísticas espaço-temporais na análise de Registros de Detalhes de Chamadas (CDR) para a construção de base de dados de deslocamento urbano

Dissertação de Mestrado apresentada ao Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, como parte dos requisitos para a obtenção do título de Mestre em Computação Aplicada.

Data da aprovação _____ / _____ / _____

Prof. Dr. Alessandro Santiago dos Santos
(Orientador)
Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT

Membros da Banca Examinadora:

Prof. Dr. Alessandro Santiago dos Santos (Orientador)
Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT

Profa. Dra. Kelly Rosa Braghetto (Membro)
Instituto de Matemática e Estatística - IME - USP

Prof. Dr. Alfredo Goldman vel Lejbman (Membro)
Docente do Mestrado Profissional em Computação Aplicada

Eder Mauricio Barbosa

Aplicação de heurísticas espacotemporais na análise de Registros de Detalhes de Chamadas (CDR) para a construção de base de dados de deslocamento urbano

Dissertação de Mestrado apresentado ao Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, como parte dos requisitos para a obtenção do título de Mestre em Computação Aplicada.

Área de Concentração: Computação Aplicada.

Linha de pesquisa: Redes e Segurança de Computadores em Sistemas Ciberfísicos

Orientador: Prof. Dr. Alessandro Santiago dos Santos

São Paulo
Dez./2024

Ficha Catalográfica
Elaborada pela GITEB – Gerência de Gestão da Informação Tecnológica e Bibliográfica
do IPT - Instituto de Pesquisas Tecnológicas do Estado de São Paulo

B238a **Barbosa, Eder Mauricio**

Aplicação de heurísticas espaço-temporais na análise de registros de detalhes de chamadas (CDR) para a construção de base de dados de deslocamento urbano. /
Eder Mauricio Barbosa. São Paulo, 2024.
85p.

Dissertação (Mestrado em Computação) - Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Área de concentração: Redes de Computadores.

Orientador: Prof. Dr. Alessandro Santiago dos Santos

1. Registros de detalhes de chamadas - CDR 2. Mobilidade urbana 3. Cidade inteligente 4. Tese I. Santos, Alessandro Santiago dos, orient. II. IPT. Unidade de Negócios em Ensino Tecnológico III. Título

2025-12

CDU 004.65(043)

DEDICATÓRIA

Dedico esta dissertação a todos que, por meio da tecnologia e da colaboração em sociedade, impulsionam avanços que transformam vidas, solucionam desafios reais, promovem educação e constroem um futuro mais justo, inclusivo e inovador.

AGRADECIMENTOS

Primeiramente, agradeço a Deus, que me concedeu força, sabedoria e perseverança para enfrentar os desafios ao longo desta jornada no Mestrado Profissional do IPT. Sem Sua benção e graça, nada disso seria possível.

Expresso minha mais sincera gratidão à minha família: à minha esposa, Kárina Cristina G. Barbosa, às minhas filhas, Diana e Daniela G. Barbosa, e aos meus pais, Otalibes e Maria Barbosa. Vocês estiveram ao meu lado em todos os momentos, oferecendo suporte incondicional, palavras de incentivo e paciência durante as inúmeras horas dedicadas às aulas e ao desenvolvimento deste trabalho. Sou eternamente grato pelo amor e pelo apoio de vocês, que foram fundamentais para que eu alcançasse meus objetivos.

Manifesto também minha profunda gratidão ao Professor Doutor Alessandro Santiago dos Santos, que desempenhou um papel essencial como meu orientador. Sua dedicação, paciência e habilidade em apresentar alternativas e formas de abordar problemas foram fundamentais para o meu crescimento acadêmico e profissional.

Agradeço especialmente aos Professores Doutores Kelly Rosa Braghetto e Alfredo Goldman, que gentilmente aceitaram participar deste processo e enriqueceram este trabalho com suas valiosas contribuições. Suas análises e sugestões foram essenciais para a qualidade e relevância deste trabalho.

Registro ainda minha gratidão aos meus colegas de curso, que, com entusiasmo e companheirismo, compartilharam comigo esta jornada acadêmica. Seus incentivos constantes, a troca de ideias e a vivência de momentos únicos de aprendizado e evolução foram essenciais para tornar essa caminhada mais rica e significativa.

Por fim, estendo meus agradecimentos a todos os professores do IPT que ministraram as disciplinas do programa de Mestrado em Computação Aplicada. Suas aulas, repletas de conteúdos atualizados e ministradas com notável domínio, ampliaram significativamente meus conhecimentos e foram determinantes para maximizar os resultados deste trabalho.

RESUMO

A infraestrutura de telefonia celular gera um grande volume de Registros de Detalhes de Chamadas (do inglês, *Call Detail Records* CDR), comumente utilizados para tarifação e análises relacionadas à infraestrutura. Entretanto, esses registros podem ser processados de forma a expandir sua aplicação para além das funções tradicionais do setor de telecomunicações, possibilitando a análise de dados de CDR com foco no desenvolvimento de novos serviços e soluções inteligentes para ambientes urbanos. Com o acelerado crescimento da urbanização e seus efeitos no planejamento urbano, a análise dos dados de telefonia móvel são uma alternativa para identificar padrões de deslocamento populacional nas cidades. Dessa forma, a análise espacotemporal dos CDRs surge como uma abordagem eficaz para compreender os padrões de mobilidade urbana e para a construção de bases de dados de deslocamento em meio digital. O estudo de caso relacionado ao monitoramento do isolamento social durante a pandemia da COVID, evidenciou o potencial da análise de CDR para aplicações em políticas públicas e gestão de crises. Neste cenário, a extração deste estudo para possibilitar outros serviços baseados no processamento de CDRs, apresentavam uma série de desafios técnicos e científicos para serem superados para a ampliação desta abordagem em serviços inteligentes para cidades. Este trabalho demonstra métodos com a aplicação de heurísticas espacotemporais para a construção de bases de dados voltadas ao mapeamento da mobilidade urbana e ao suporte ao planejamento de cidades inteligentes. Sendo, que as heurísticas espaciais são aplicadas para o agrupamento de dados por regiões percorridas pelo caminho do usuário em sua trajetória de mobilidade, classificando estes caminhos com base nas áreas cobertas por Estações Rádio Base (ERB), enquanto as heurísticas temporais são aplicadas na análise do tempo de permanência na ERB, apontadas pelos padrões de conexões ao longo do tempo, identificando locais de origem e destino no trajeto do usuário. Essas técnicas permitem inferir padrões de mobilidade, que podem ser consumidos por serviços para gestão de cidades, incluindo a preservação da privacidade dos usuários por meio de anonimização. A dissertação demonstrou que, ao utilizar CDR de dados e técnicas analíticas, é possível inferir trajetos detalhados e compreender as dinâmicas de deslocamento, oferecendo uma visão sobre a mobilidade urbana, com uma segregação dos dados em *sandboxes*, com volume reduzido, mantendo as características essenciais para as análises de deslocamento. A validação foi evidenciada em diferentes experimentos de mobilidade, como deslocamentos via automóvel, bicicleta e a pé, sendo que os resultados das bases de dados de caminhos e de origem-destino (O-D) criadas, correspondiam aos padrões de deslocamento executados pelo usuário. As conclusões apontam que a aplicação de heurísticas espacotemporais gerando bases de caminhos e de Origem-Destino de forma digital, podem oferecer novos mecanismos de análise.

para o entendimento da mobilidade urbana, e para a gestão das cidades, possibilitando maior eficiência na alocação de recursos e na tomada de decisões estratégicas. Entretanto, cada cidade possui características territoriais e populacionais distintas, que podem incorporar desafios relacionados à representatividade dos dados e à precisão da localização, o que cabe uma análise de aplicabilidade ou não, para cada cidade.

Palavras-chave: registro detalhado de chamadas; cidade inteligente; cidades inteligentes; rede inteligente; estação rádio base

ABSTRACT

Application of Spatiotemporal Heuristics in the Analysis of Call Detail Records (CDR) for the Construction of Urban Mobility Databases

The cellular telephone infrastructure generates a large volume of Call Detail Records (CDR), commonly used for billing and infrastructure-related analysis. However, these records can be processed to expand their application beyond the traditional functions of the telecommunications sector, enabling the analysis of CDR data with a focus on developing new services and intelligent solutions for urban environments. With the rapid growth of urbanization and its effects on urban planning, the analysis of mobile phone data is an alternative to identify patterns of population movement in cities. Thus, the spatiotemporal analysis of CDRs emerges as an effective approach to understand urban mobility patterns and to build digital displacement databases. The case study related to monitoring social isolation during the COVID pandemic highlighted the potential of CDR analysis for applications in public policies and crisis management. In this scenario, extending this study to enable other services based on CDR processing presented a series of technical and scientific challenges to overcome for the expansion of this approach into intelligent services for cities. This work demonstrates methods for applying spatiotemporal heuristics to construct databases aimed at mapping urban mobility and supporting the planning of smart cities. Spatial heuristics are applied to cluster data by regions traversed by the user's path of mobility, classifying these paths based on the areas covered by Base Transceiver Stations (BTS), while temporal heuristics are applied to analyze the dwell time in the BTS, indicated by connection patterns over time, identifying origin and destination locations in the user's trajectory. These techniques allow inferring mobility patterns, which can be used by city management services, including preserving user privacy through anonymization. The dissertation demonstrated that by using CDR data and analytical techniques, it is possible to infer detailed routes and understand the dynamics of movement, offering a view on urban mobility with data segregation in sandboxes, with reduced volume, maintaining essential characteristics for displacement analysis. Validation was evidenced in different mobility experiments, such as travel by car, bicycle, and on foot, with the results of the path and origin-destination (O-D) databases created corresponding to the patterns of user movement. The conclusions indicate that the application of spatiotemporal heuristics generating digital path and Origin-Destination databases can offer new analysis mechanisms for understanding urban mobility and city management, enabling greater efficiency in resource allocation and strategic decision-making. However, each city has distinct territorial and population characteristics, which may incorporate challenges related to data representativeness and location accuracy, requiring an analysis of applicability for each city.

Keywords: Call Detail Record; Smart City; Smart Cities; Smart Network; Base Transceiver Station

Lista de Figuras

Figura 1	Arquitetura Básica de Rede de Celular	21
Figura 2	Fluxo de como o BSS e MSC Funcionam	22
Figura 3	Arquitetura Básica de Rede de Celular	23
Figura 4	Exemplo de CDR Bruto	28
Figura 5	CDR Enriquecido com outras Fontes	29
Figura 6	Sequência de Fases para a RS	31
Figura 7	Busca de Artigos nas Bases	33
Figura 8	Mapeamento Sistemático	34
Figura 9	Procedimento para Criar e Consolidar as Informações do Índice de Isolamento Social	37
Figura 10	Arquitetura de Referência e Passos Utilizados no Projeto COVID	38
Figura 11	Arquitetura para Plataforma Analítica de CDR	39
Figura 12	Sistema de Monitoramento Inteligente do Governo de São Paulo	43
Figura 13	Abordagem em Camadas	48
Figura 14	Segmentação Espacial dos CDRS	50
Figura 15	Segmentação Temporal dos CDRS	52
Figura 16	Fluxo e Passos das Bases	54
Figura 17	Fluxo de Coleta e Processamento dos CDR	62
Figura 18	Amostra de CDR Bruto	63
Figura 19	Experimento com Dados de Deslocamento via Automóvel.	64
Figura 20	Experimento com Dados de Deslocamento via Bicicleta.	65
Figura 21	Experimento com Dados de Deslocamento em uma Corrida.	66

Lista de Tabelas

Tabela 1	Estrutura de um Registro CDR	27
Tabela 2	Bases eletrônicas utilizadas na RS	32
Tabela 3	Formato dos Arquivos com Dados de Coleta por Antena	41
Tabela 4	Lições, Aprendizados e Novas Perspectivas.	44
Tabela 6	Primeira Parte da Tabela com Colunas e Campos dos CDR Analisados	53
Tabela 7	Segunda Parte da Tabela com Coluna Campos Adicionais dos CDR Analisados	53
Tabela 8	Tabela da Base de Caminhos	56
Tabela 9	Tabela com Campos dos CDR Analisados	57
Tabela 10	Tabela de Origem Destino	58
Tabela 11	Resumo do Ambiente de Sandbox Criado para o Processamento dos Dados	62
Tabela 12	Amostra da Base de Caminhos.	64
Tabela 13	Registros de Deslocamentos do Experimento	67

Lista de Abreviaturas e Siglas

- ERB *Estação Rádio Base*
CDR *Registros de Detalhamento de Chamadas*
SMS *Short Message Service*
MMS *Multimedia Message Service*
TIC *Tecnologias da Informação e Comunicação*
BSS *Broadcast Satellite Services*
GSM *Global System for Mobile Communications*
O-D *Origem e Destino*
site *Local ou Sítio*

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Contexto e Motivação	16
1.2 Objetivos, Questões e Proposições da Pesquisa	18
1.3 Contribuições	19
2 FUNDAMENTAÇÃO TEÓRICA	20
2.1 Ecossistema de infraestrutura de telecomunicações	20
2.1.1 Estação de Rádio Base	21
2.1.2 IP Multimedia Subsystem - IMS	22
2.1.3 Divisão Espacial e Arquiteturas Celulares	24
2.2 Registros de Dados de Chamadas CDR: Definição e Aplicações	24
2.2.1 RFC 2924 — Estrutura e Padronização de Atributos de CDRs	25
2.2.2 RFC 5114 – Importância para o Contexto de CDRs	26
2.2.3 Definições de CDR por Fabricantes	26
2.2.4 Exemplos e Modelos de CDR	27
3 REVISÃO SISTEMÁTICA E UM CASO DE REFERÊNCIA NO BRASIL	30
3.1 Revisão Sistemática da Literatura	30
3.1.1 Planejamento da Revisão Sistemática	31
3.1.2 Condução	32
3.1.3 Consolidação da Revisão	33
3.2 Estudo de caso: Índices de Isolamento Social para Enfrentar a COVID no Brasil	36
3.2.1 Desafios e Precisões na Análise de Conjuntos de Dados	38
3.2.2 Arquitetura para Análise de Dados de CDR	39
3.2.3 De que Maneira os CDRs são Preparados para Distinguir Dados de Origem e Destino	40
3.2.4 Mapa de Origem x Destino — Caso de Uso de Dados Amostrais de CDRs	41
3.2.5 Painel Informativo Utilizado no Estudo de Caso	43
3.2.6 Análise Crítica	44
3.3 Consolidação da Revisão e do Estudo de Caso	44
4 ABORDAGEM EM CAMADAS COM HEURÍSTICAS ESPAÇOTEMPORAIS PARA CRIAÇÃO DE BASES DE DESLOCAMENTO	47
4.1 Modelo Conceitual	47
4.1.1 Segmentação Espacial sobre os CDRs	49
4.1.2 Segmentação Temporal para Construção das Bases O-D	51

4.2	Métodos e Implementação	52
4.2.1	Implementação da Heurística Espacial (Base de Caminhos)	55
4.2.2	Implementação da Heurística Temporal (Base Origem-Destino)	57
5	EXPERIMENTO	61
5.1	Ambientes de operação e teste.	61
5.1.1	Ambiente operacional de análise	61
5.2	Execução do Experimento	62
5.2.1	Experimento com Carro de Aplicativo de Transporte	63
5.2.2	Experimento com Bicicleta	65
5.2.3	Experimento com Mobilidade a Pé	66
5.3	Disponibilidade dos Dados	67
5.4	Discussão dos Resultados	67
5.5	Relações e Associação dos Resultados com as Questões de Pesquisa	68
5.6	Contribuição Futura	71
6	CONCLUSÃO	72
	REFERÊNCIAS	76
	ANEXO A – ARTIGO DE MOBILIDADE IEEE	81
	ANEXO B – ARTIGO DO CASO DE USO IET	84

1 INTRODUÇÃO

As mais recentes evidências fornecidas pelo Censo de 2022, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), revelam uma taxa de crescimento populacional anual de 0,52% para o Brasil, situando a população atual em aproximadamente 203 milhões, 80 mil e 756 de habitantes o que apresenta diversos desafios para o planejamento urbano e o desenvolvimento sustentável do país (IBGE, 2022).

No mundo, pesquisas realizadas no ano de 2020 já indicavam que mais da metade da população mundial, equivalendo a mais de 4,3 bilhões de pessoas, se estabeleceria em ambientes urbanos, e espera-se que, até 2045, a população cresça em 1,5 vezes, atingindo a cifra de mais de 6 bilhões de habitantes em áreas urbanas (AMMAPA *et al.*, 2022). Este aumento demográfico ressalta a urgência de se repensar e fortalecer os sistemas de infraestrutura urbana, especialmente no que tange ao setor de transportes.

Segundo, Liu *et al.* (2020) a urbanização rápida, impulsionada tanto pelo crescimento populacional quanto pela migração para áreas urbanas, tem estimulado significativamente o desenvolvimento e a expansão das infraestruturas de transporte. No entanto, este crescimento nem sempre é acompanhado por planejamento e investimento adequados, resultando em impactos adversos consideráveis na mobilidade urbana.

No contexto atual, caracterizado por uma urbanização crescente e complexa, técnicas avançadas de “big data” emergem como soluções fundamentais para enfrentar os desafios urbanos. Uma dessas técnicas envolve a integração e análise de vastas quantidades de dados de telefonia móvel, incluindo informações de localização de residências e locais de trabalho dos cidadãos. Estes dados são de valor inestimável para o planejamento de transportes e o desenvolvimento urbano sustentável, uma vez que compreendem os padrões de mobilidade diários essenciais na vida das pessoas (BARBOSA *et al.*, 2023). A aplicação prática dessa técnica é realizada através da análise de registros de detalhes de chamadas, do inglês *Call Detail Records* (CDR), que fornecem percepções sobre a mobilidade de curto prazo das populações urbanas.

O CDR pode ser descrito como registros de informações sobre todas as comunicações realizadas pelo usuário de telefonia, seja o uso de voz, SMS (Short Message Service), MMS (Multimedia Message Service) ou Internet. Os dados brutos, contém dados da antena e dados da comunicação, tais como: data/hora de início e fim da chamada, id (identificador) telefone de origem, id de telefone de destino, além de diversas outras informações da aplicação como marcações de equipamentos da infraestrutura

da operadora de telefonia(IETF, 2024a).

Estes registros, quando combinados com dados geoespaciais, podem, possibilitar a identificação precisa de origem e destino da população em movimento e respectivamente a determinação das residências e locais de trabalho dos indivíduos (LUO *et al.*, 2020).

Para entender os impactos da pandemia da COVID nos deslocamentos das pessoas, conforme Metropolitano de São Paulo - Metrô (2024), a maior pesquisa de mobilidade urbana do país, realizada a cada 10 anos, foi antecipada pelo Metrô, tendo sido iniciada em 1º de agosto de 2023, com o objetivo de mapear todas as formas e motivos das viagens das pessoas na Região Metropolitana de São Paulo e em seus acessos. As linhas de metrô de São Paulo, por exemplo, são projetadas a partir das simulações de demanda de passageiros feitas com o uso dos dados fornecidos pela OD. Os resultados da pesquisa são também utilizados em modelagens de estudos dos setores de urbanismo, logística, segurança, saúde, educação e outros. Com a metodologia sistematizada em São Paulo, semelhante às consagradas em Londres e Paris, hoje, grandes cidades brasileiras, capitais e regiões metropolitanas aplicaram as pesquisas origem e destino nos seus planos de mobilidade e logística.

Conforme Jakovljevic *et al.* (2021) uma base de origem e destino (OD) é essencial para o planejamento e a gestão da mobilidade urbana em cidades inteligentes. Ela permite identificar e entender os padrões de deslocamento da população, capturando o número de viagens entre diferentes áreas da cidade. Esse tipo de análise oferece dados fundamentais sobre a demanda de transporte, ajudando a otimizar o tráfego, a alocar recursos de transporte público de forma mais eficiente e a planejar novas infraestruturas. A partir das matrizes OD, é possível antecipar comportamentos de mobilidade, ajustar serviços urbanos e promover um desenvolvimento mais sustentável e adaptado às necessidades da população urbana.

A importância dessa base de dados é evidenciada pela prática histórica do Metrô de São Paulo, que realiza a pesquisa OD a cada 10 anos desde 1967. Foi a partir desses estudos que se elaborou o primeiro plano para a construção da rede básica de metrô da cidade, destacando como a análise de padrões de origem e destino é vital para o desenvolvimento e expansão da infraestrutura de transporte urbano.

Na era das cidades inteligentes, a mobilidade urbana se beneficia significativamente do avanço do “big data”. O vasto volume de dados gerados pelas redes de telefonia emerge como uma fonte para compreender e otimizar os deslocamentos urbanos. Estes dados, especialmente os CDR, oferecem uma visão detalhada dos padrões de movimento da população, essencial para o planejamento de transportes e a gestão

de cidades inteligentes.

Contudo, a utilização desses dados traz consigo desafios significativos, incluindo questões de privacidade, precisão de localização, análise e interpretação. Um dos desafios críticos enfrentados é o erro inerente ao uso do posicionamento baseado em estações-base de telefonia móvel (ERB). Este problema é agravado em áreas onde a densidade de estações-base é baixa, levando a distâncias significativas entre as estações e, por consequência, a uma representação imprecisa das localizações dos indivíduos. Outro desafio importante é a ausência de um método eficaz para extrapolar os dados obtidos dos usuários de telefones móveis para a população da cidade na totalidade. De acordo com Luo *et al.* (2020), este é um passo importante em estudos de amostragem para assegurar a representatividade e a aplicabilidade dos resultados a um contexto mais amplo.

Apesar dessas barreiras, o potencial para transformar a mobilidade urbana é inegável. O “big data” de telefonia pode ser empregado para identificar, analisar e melhorar os padrões de mobilidade urbana, contribuindo assim para o desenvolvimento de soluções de transporte mais eficientes e sustentáveis.

Ao abordar tanto os desafios técnicos quanto as oportunidades, esta introdução lança as bases para um estudo sobre a aplicação de “big data” de telefonia na melhoria dos sistemas de mobilidade urbana, visando contribuir para a otimização dos deslocamentos na vida cotidiana das cidades inteligentes.

Esta abordagem busca equilibrar a discussão técnica com a aplicabilidade prática, estabelecendo um ponto de partida para investigações futuras no campo da mobilidade baseada em “big data” de telefonia.

1.1 Contexto e Motivação

A construção e transição para cidades mais inteligentes representam um anseio crescente da sociedade, impulsionado pela evolução significativa rumo a um modelo mais sustentável. Esta evolução influencia diretamente a gestão e administração das cidades contemporâneas. Nesse contexto, várias fontes de informação advindas das infraestruturas de telecomunicações surgem como alternativas modernas para incorporar cenários de “big data”, apoiando a gestão de cidades inteligentes.

Neste panorama, os CDR provenientes das redes de telefonia móvel emergem como uma fonte promissora de “big data”, capaz de revelar padrões complexos de mobilidade urbana. Estes registros, que incluem informações sobre chamadas, mensagens curtas (SMS) e transferências de dados, constituem um ambiente rico em dados, apresentando uma oportunidade única para análise e gestão de dados urbanos.

Uma média de 1 bilhão de registros por minuto é processada em âmbito nacional por uma operadora de telefonia brasileira, considerando a totalidade de sua base de clientes. No total de registros estimados para 108 milhões de usuários, aproximadamente 800 milhões de registros de CDR foram identificados e já convertidos do formato binário para CSV. A base original de CDR é alimentada com aproximadamente 70 gigabytes de informações por minuto em âmbito nacional, resultando em um volume total de 2,88 petabytes de dados por mês.

O processamento desse volume exige a utilização de servidores especializados em processamento distribuído, tornando-se essencial para a operação. O processo de conversão de arquivos binários de CDRs para CSV envolve a volumetria de 800 milhões de registros, sendo processada por meio de um algoritmo desenvolvido em Java, executado em um cluster Hadoop Cloudera versão 7.1. A infraestrutura configurada conta com 32 vmcores de 3.8 Ghz e 140 terabytes de memória RAM. O processamento é efetuado em uma fila dedicada, utilizando operações paralelizáveis.

Todo o volume e complexidade desta infraestrutura apresentada deixa evidente os desafios de processamento destes dados para aplicação de serviços de telecomunicações e também sua ampliação para outros escopos de cidades inteligentes.

No contexto do enfrentamento à pandemia da COVID em São Paulo, uma experiência inovadora de projeto de dados foi realizada, utilizando informações provenientes de CDR para analisar o padrão de mobilidade da população de forma agregada. Este projeto conforme Santos, Teixeira *et al.* (2021) representou uma colaboração entre operadoras de telecomunicações, especialistas em dados e o governo de São Paulo, visando otimizar as medidas de contenção do vírus por meio de uma compreensão mais precisa dos fluxos de movimento das pessoas pelo estado de São Paulo.

Em um mundo cada vez mais urbanizado, as cidades como São Paulo enfrentam desafios sem precedentes relacionados à mobilidade urbana. O aumento contínuo da população urbana exerce pressão sobre os sistemas de transporte existentes, resultando em congestionamentos, poluição atmosférica e uma diminuição na qualidade de vida dos habitantes em centros urbanos. Esses problemas não apenas afetam o bem-estar dos cidadãos, mas também representam obstáculos significativos para o desenvolvimento sustentável das cidades. A busca por soluções eficazes para esses desafios tem levado à concepção de cidades inteligentes, onde a tecnologia aliada a dados é utilizado para otimizar recursos, melhorar serviços públicos e promover a sustentabilidade. Nesse cenário, o “big data” emerge como uma ferramenta poderosa para entender e solucionar complexidades urbanas, especialmente no que diz respeito à mobilidade. Os CDR, gerados pelas operadoras de telefonia móvel, representam uma fonte rica e subutilizada de “big data”, capaz de oferecer percepções profundas

sobre os padrões de movimento da população. A análise desses dados tem o potencial de transformar o planejamento urbano e a gestão do transporte, permitindo uma compreensão mais precisa das necessidades de mobilidade dos cidadãos e possibilitando a implementação de soluções mais adaptadas e eficientes.

Segundo Santos, Teixeira *et al.* (2021) a utilização de CDR no projeto de enfrentamento ao COVID nos permitiu acessar grandes volumes de dados sobre a localização e movimentos dos usuários de telefonia móvel, respeitando, entretanto, a privacidade individual através da agregação e anonimização dos dados. Essas informações foram importantes para identificar áreas de alta concentração e movimentação de pessoas, auxiliando o governo no planejamento e implementação de medidas restritivas focadas, como o fechamento de determinadas áreas comerciais ou a limitação do transporte público em horários de pico, visando reduzir a disseminação do vírus.

Dessa forma, a análise de padrões de mobilidade utilizando CDR e “big data” não apenas endereça questões imediatas de planejamento e gestão de transporte, mas também contribui para o objetivo mais nobre de criar cidades mais inteligentes, sustentáveis e inclusivas. Este trabalho visa, portanto, fornecer uma contribuição valiosa para o campo do planejamento urbano e da mobilidade, abrindo novas perspectivas para o uso de tecnologias em infraestrutura de telecomunicações na resolução dos desafios urbanos contemporâneos.

Vale destacar que a utilização de dados de CDR para esses fins não está isenta de limitações e desafios. A privacidade e a segurança dos dados são preocupações primárias, exigindo a anonimização e agregação cuidadosa dos dados para proteger a identidade dos indivíduos. Além disso, a dependência de dados gerados por usuários de telefonia móvel pode introduzir vieses, uma vez que não representa perfeitamente toda a população ou todos os padrões de mobilidade.

1.2 Objetivos, Questões e Proposições da Pesquisa

Esta dissertação apresenta como objetivo principal o desenvolvimento de estratégias para a construção de um conjunto de dados de deslocamento, aplicando heurísticas spaçotemporais na análise de Registros de Detalhes de Chamadas (CDR) provenientes de redes móveis para serviços de cidades inteligentes.

Como objetivos específicos pretende-se: (i) identificar quais são os principais usos de CDRs para prover serviços em cidades inteligentes; (ii) elaborar métodos para construir conjunto de dados por meio de processamento de CDRs, como subsídio para serviços inteligentes que se aproveitem de dados de deslocamento para suportar atividades de planejamento e operação de serviços em cidades inteligentes.

Para atingir os objetivos, foram elaboradas questões para orientar o desenvolvimento da pesquisa, expressas nas seguintes **Questões de Pesquisa**:

P1. Quais são os principais serviços de cidades inteligentes que poderiam utilizar dados de CDR?

P2. Como utilizar dados de CDR para criar uma base de caminhos e de Origem-Destino aplicável a serviços de cidades inteligentes?

1.3 Contribuições

- Para a Academia:

O desenvolvimento e a aplicação de heurísticas spaçotemporais na análise de dados de CDR contribuem para a criação de novos métodos de análise de mobilidade urbana, ampliando as ferramentas disponíveis para estudos em big data e ciência de dados aplicados à mobilidade.

Este estudo enriquece a literatura sobre cidades inteligentes, fornecendo uma base teórica e metodológica que pode ser replicada e aprimorada em pesquisas futuras para entender melhor os padrões de deslocamento urbano.

A proposta de utilizar dados de CDR como fonte primária para a construção de matrizes de origem e destino oferece à comunidade acadêmica uma alternativa inovadora aos métodos tradicionais de análise de dados de mobilidade.

- Para a Indústria:

As estratégias desenvolvidas podem ser aplicadas em empresas e órgãos de planejamento urbano, permitindo uma tomada de decisões mais informada e eficaz em relação à infraestrutura de transporte e mobilidade.

As operadoras de telefonia móvel podem explorar o uso de CDR como um recurso adicional de negócio, oferecendo visões de mobilidade para governos e empresas, criando novas oportunidades de serviços voltados para cidades inteligentes.

A base de dados criada com CDR facilita a implementação de soluções que melhoram a alocação de recursos e a eficiência dos serviços urbanos, contribuindo para a sustentabilidade e a qualidade de vida nas cidades inteligentes.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção proporciona uma análise detalhada da fundamentação teórica necessária e utilizada para a construção das estratégias e bases conceituais para a construção de conjuntos de dados de deslocamento, com o uso de “big data” e CDR, no contexto de tecnologias e infraestrutura de telecomunicações, e desenvolvimento de cidades inteligentes.

2.1 Ecossistema de infraestrutura de telecomunicações

Conjuntos anônimos de “big data” de telecomunicações, apoiados por uma infraestrutura robusta de telecomunicações, emergiram como uma fonte inovadora de dados para diversas análises em planejamento de transporte e mobilidade. Esta infraestrutura essencial não só facilita a coleta em larga escala de dados, refletindo a abrangência do mercado do operador móvel, mas também garante a confiabilidade e a rapidez necessárias para a obtenção de dados precisos e facilmente replicáveis. A importância dessa infraestrutura se destaca no tratamento de volumes massivos de informações, possibilitando a representação da posição aproximada da população ao longo do tempo conforme Vidovic *et al.* (2022). Entretanto, para extrair percepções significativas sobre o movimento populacional (viagens populacionais) a partir desses dados, são indispensáveis técnicas analíticas avançadas.

Segundo Zhang *et al.* (2018) a infraestrutura de telecomunicações desempenha um papel crucial nesse processo, oferecendo a capacidade técnica necessária para o manuseio e a análise de grandes volumes de dados, permitindo assim uma compreensão mais profunda dos padrões de mobilidade. Com base nas características desses movimentos, tais como tempo, distância aproximada, velocidade e duração, é possível inferir contextos adicionais como a finalidade da viagem e o modo de transporte utilizado. A detecção do modo de transporte, em particular, beneficia-se enormemente da robustez dessa infraestrutura, pois a identificação da utilização dos diferentes modos de transporte através desta análise emerge como uma solução promissora para obter uma repartição modal real e confiável. Isso contrasta com outros métodos que são complexos e dependem geralmente de pesquisas ou observações manuais, destacando assim a vital importância da infraestrutura de telecomunicações no avanço das capacidades de análise de mobilidade.

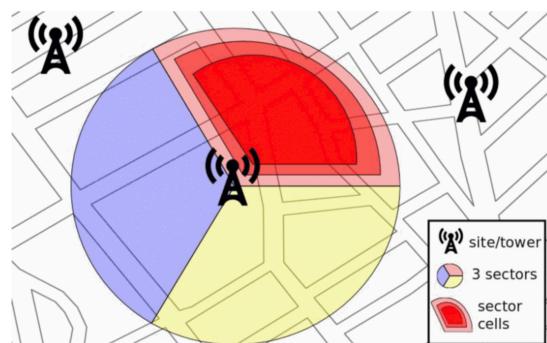
O ecossistema de infraestrutura de telecomunicações é essencial para a sociedade, por possibilitar acesso à informação e o funcionamento de uma variedade de serviços essenciais. Com a constante evolução da tecnologia, esse ecossistema con-

tinua a crescer e se adaptar para atender às demandas em constante mudança dos usuários finais e da indústria.

Um ecossistema de infraestrutura de telecomunicações é um conjunto complexo de tecnologias e serviços que possibilita a comunicação e a transferência de dados em uma rede de telecomunicações.

Segundo Elbakly e Youssef (2019) a rede celular é estruturada em uma série de sites, ou torres, distribuídos em locais físicos específicos. Cada uma dessas torres fornece cobertura a uma área circundante, subdividida em diversos setores. Um setor, então, é definido como uma segmentação do espaço ao redor da torre, delimitado por duas linhas que partem do centro da torre em ângulos precisos em relação ao norte, formando uma “fatia” do círculo. Embora teoricamente a cobertura de um setor possa se estender indefinidamente, na prática, a distância alcançada é limitada pela capacidade do sinal de viajar, ou seja, sua atenuação, determinando assim a área efetiva de cobertura de cada setor conforme Figura 1 .

Figura 1 – Arquitetura Básica de Rede de Celular



Fonte: Elbakly e Youssef (2019)

2.1.1 Estação de Rádio Base

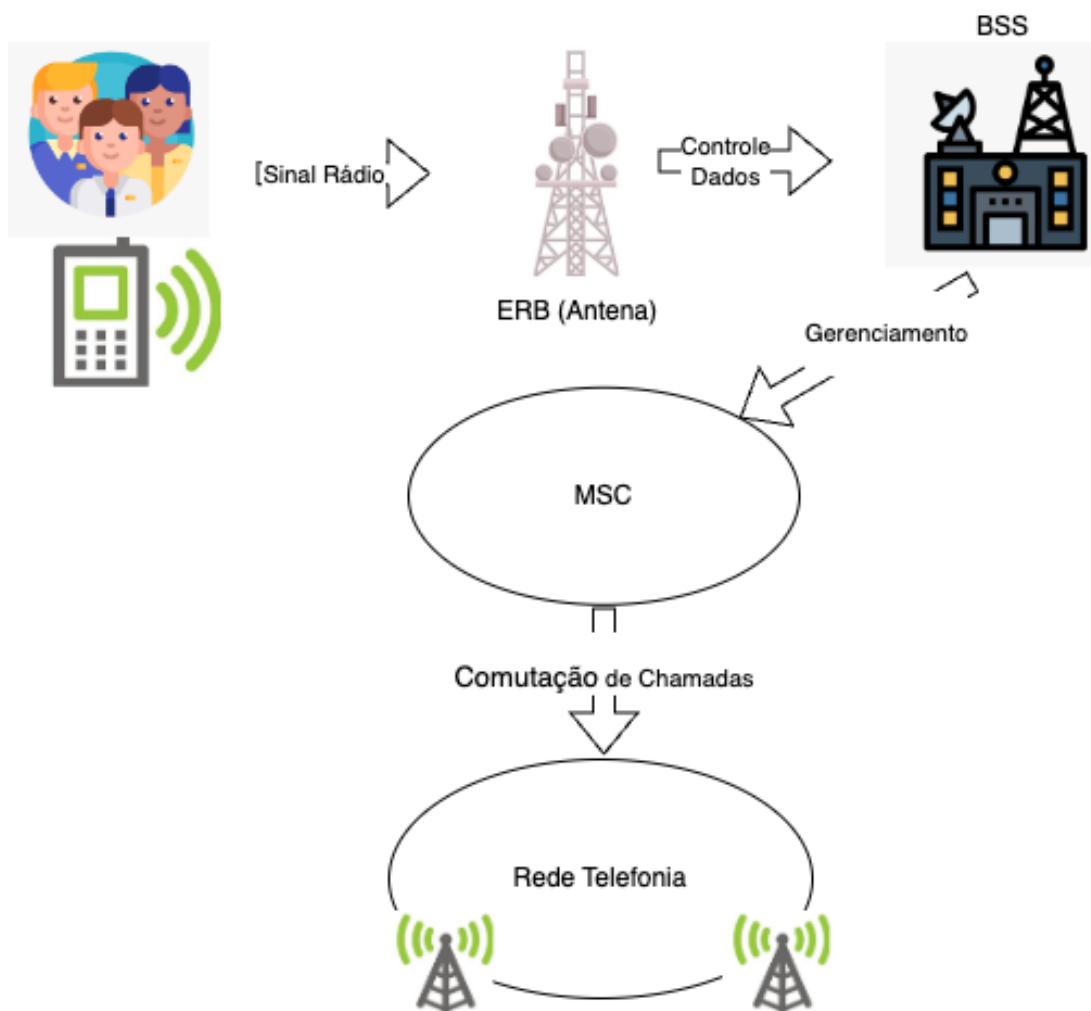
O Subsistema de Estação Base (BSS) e Central de Comutação Móvel (MSC) são elementos fundamentais da arquitetura do sistema (GSM), trabalhando de maneira integrada para garantir a comunicação móvel eficiente (RAHNEMA, 1993). O BSS é responsável por gerenciar os recursos de rádio e a conexão entre os dispositivos móveis e a infraestrutura da rede. Ele controla as Estações Rádio Base (ERBs), gerenciando operações como alocação de canais, controle de potência e passagem (handovers) locais, ou seja, transferências de conexão entre células sob sua supervisão.

O MSC atua como o núcleo de comutação da rede, sendo responsável pela interligação do BSS com outras partes da infraestrutura, como redes fixas e móveis. O

MSC coordena as chamadas de voz e dados, processa a mobilidade dos usuários e garante a entrega contínua de serviços, mesmo durante deslocamentos entre áreas controladas por diferentes BSS. Essa colaboração entre o BSS e o MSC é essencial para suportar a alta demanda de comunicação móvel, mantendo a qualidade e a eficiência da rede.

A Figura 2 ilustra a colaboração entre esses importantes componentes de uma infraestrutura celular.

Figura 2 – Fluxo de como o BSS e MSC Funcionam



Fonte: Adaptado de Goleniewski e Jarrett (2006)

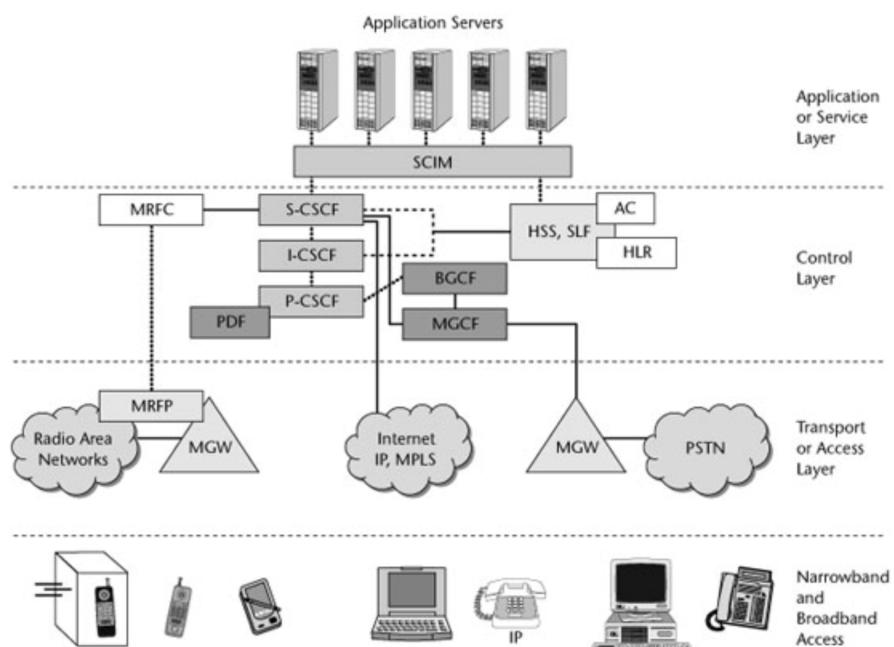
2.1.2 IP Multimedia Subsystem - IMS

Ao se enfocar a comunicação de dados pela rede celular, a arquitetura IMS (IP Multimedia Subsystem) é um dos elementos principais neste processo. A arquitetura IMS segundo Goleniewski e Jarrett (2006) define um modelo genérico para oferecer serviços de VoIP e multimídia em infraestruturas com ou sem fio. Essa arquitetura é

organizada em três camadas principais: camada de serviço (ou aplicação), camada de controle e camada de transporte (ou acesso) conforme Figura 3.

Um dos principais elementos da arquitetura, o Multimedia Resource Function Controller (MRFC), desempenha um papel fundamental ao controlar o processamento de recursos multimídia e gerar registros de detalhamento de chamadas (CDRs), cruciais para análise, cobrança e monitoramento de serviços.

Figura 3 – Arquitetura Básica de Rede de Celular



Fonte: Goleniewski e Jarrett (2006)

Na camada de controle, o MRFC interpreta as informações enviadas por servidores de aplicação e pelo principal controlador de chamadas da rede (S-CSCF), ajustando o comportamento do Multimedia Resource Function Processor (MRFP) para gerenciar fluxos de mídia. Além disso, o MRFP oferece funções como o mix de fluxos de mídia, geração de anúncios multimídia e processamento de streams.

Os CDRs gerados pelo MRFC permitem registrar eventos detalhados, como origem, destino, duração das chamadas e uso de recursos, sendo fundamentais para operadores realizarem cobrança, auditoria e análise de desempenho dos serviços. Esses registros são particularmente valiosos para redes IMS, devido à sua capacidade de suportar uma ampla gama de aplicações, como conferências multimídia, gerenciamento de presença e mensagens instantâneas.

A arquitetura IMS, conforme definida no padrão 3GPP Release 5 e em versões posteriores, evoluiu para incluir suporte a novas tecnologias, como interoperabilidade com redes WLAN e convergência entre redes fixas e móveis. Essa evolução destaca a

importância dos CDRs na adaptação dos serviços às necessidades de redes modernas, possibilitando a integração eficiente de diferentes protocolos e tecnologias, enquanto mantém a rastreabilidade e a qualidade do serviço.

2.1.3 Divisão Espacial e Arquiteturas Celulares

Conforme descrito por Goleniewski e Jarrett (2006), a telefonia móvel, existente desde a década de 1950, enfrentava inicialmente diversas limitações para atender o mercado em larga escala. Naquela época, as antenas de alta potência cobriam áreas de até 161 km (100 milhas) de diâmetro, mas havia poucas frequências disponíveis, geralmente reservadas para serviços essenciais, como polícia e emergências, ou para usuários que podiam pagar por equipamentos grandes e caros.

Na década de 1970, duas mudanças significativas revolucionaram o setor. Primeiro, o surgimento de torres transmissoras de baixa potência reduziu as áreas de cobertura para células menores, com cerca de 13 km (8 milhas) de diâmetro, chamadas de macrocélulas. Segundo, as agências reguladoras disponibilizaram mais frequências para comunicações móveis, aumentando a capacidade de atendimento.

Com o aumento da densidade de usuários, houve a necessidade de subdividir ainda mais as células, evoluindo para arquiteturas como microcélulas(1,5 km) e pico células (46 metros), descritas no livro (GOLENIEWSKI; JARRETT, 2006).

2.2 Registros de Dados de Chamadas CDR: Definição e Aplicações

Os registros de dados de chamadas CDR constituem uma fonte rica em informações, detalhando os padrões de comunicação - como, quando e com quem ou o que interagimos de acordo com Chin *et al.* (2019). Tradicionalmente, as operadoras de telefonia móvel utilizam esses registros primariamente para a finalidade de faturamento. No entanto, a aplicabilidade desses dados vai além, abrindo novos horizontes de pesquisa e inovação. Na jornada em direção ao desenvolvimento de cidades mais inteligentes e conectadas, esses registros oferecem uma oportunidade valiosa de exploração.

Os centros de controle de cidades inteligentes podem, ao tirar proveito da análise desses dados, otimizar significativamente suas operações. Isso não apenas melhora a eficiência dos serviços urbanos, mas também contribui para a criação de ambientes urbanos mais responsivos e adaptativos às necessidades de seus habitantes, de acordo com Barbosa *et al.* (2023). A integração e análise dos CDR, portanto, representam um passo crucial na evolução das cidades inteligentes, proporcionando percepções fundamentais para a otimização de recursos, planejamento de transporte, gestão de emergências e muito mais.

De acordo com Goleniewski e Jarrett (2006), a padronização em telecomunicações é fundamental para garantir a conexão entre sistemas e redes, possibilitar a portabilidade de equipamentos entre diferentes regiões e assegurar que produtos de diversos fabricantes sejam compatíveis. Essas iniciativas são essenciais para reduzir custos e ampliar o mercado global.

Diversas organizações globais têm papel central nesse processo, como a Internet Society (ISOC) e a União Internacional de Telecomunicações (ITU) sendo vinculada à Organização das Nações Unidas (ONU). Estas organizações definem alguns padrões como os RFCs (Request for Comments).

Os RFCs são documentos técnicos e organizacionais que descrevem métodos, comportamentos, pesquisas ou inovações aplicáveis à internet e sistemas interconectados. Eles são publicados pela Internet Engineering Task Force (IETF) para promover padrões abertos (IETF). Um exemplo relacionado ao CDR (Registros de Detalhamento de Chamadas) é o RFC 2924 conforme IETF (2024a) que aborda formatos de registro e atributos de contabilidade, além de discutir extensibilidade, formatos de transporte e vantagens de abordagens integradas para gerenciar eventos de uso de serviços. Isso inclui protocolos amplamente utilizados como RADIUS e DIAMETER.

2.2.1 RFC 2924 — Estrutura e Padronização de Atributos de CDRs

A RFC 2924 fornece diretrizes sobre quais atributos devem ser considerados em registros de contabilidade. Para análise espacotemporal, isso é relevante porque define claramente elementos como:

- Horário da chamada (para inferir deslocamentos).
- Localização da ERB (Estação Rádio Base).
- Duração da conexão.

O uso de formatos consistentes para os CDRs facilita a integração entre diferentes fontes de dados. Isso permite a agregação de CDRs de múltiplos provedores, aumentando a representatividade da base de dados de deslocamento urbano.

A RFC 2924 garante que CDRs coletados de diferentes redes ou operadoras possam ser interpretados uniformemente, essencial para análises em larga escala envolvendo várias fontes e define padrões que permitem calcular métricas como frequência e duração das chamadas em diferentes localidades, dados essenciais para inferir padrões de mobilidade urbana.

2.2.2 RFC 5114 – Importância para o Contexto de CDRs

Embora o RFC 5114 não mencione diretamente os CDRs conforme IETF (2024b), ele fornece diretrizes para proteger dados que transitam em redes, o que é fundamental para sistemas que gerenciam registros detalhados de chamadas.

A adoção de práticas de segurança descritas na RFC, como a implementação de grupos de Diffie-Hellman mais fortes, garante que os dados de CDR sejam transmitidos seguramente entre diferentes elementos da rede. O Diffie-Hellman é um protocolo de criptografia que permite a troca segura de chaves criptográficas entre duas partes em um canal de comunicação público.

A RFC 5114 define parâmetros de segurança, como grupos de Diffie-Hellman, que fortalecem o transporte seguro de dados sensíveis. Isso é crucial na análise de CDRs, pois esses registros contêm informações pessoais como localização e horário, que precisam ser protegidas contra acesso não autorizado.

Os métodos de segurança da RFC garantem que os CDRs possam ser transferidos entre elementos da rede (ex.: torres de celular e servidores de análise) sem risco de interceptação ou manipulação, essencial em ambientes de redes públicas ou expostas.

As RFC 5114 e RFC 2924 oferecem contribuições valiosas que podem apoiar a aplicação de heurísticas espaço-temporais na análise de CDRs para construção de bases de dados de deslocamento urbano, garantindo segurança, interoperabilidade e eficiência no processamento de dados.

2.2.3 Definições de CDR por Fabricantes

Cada fabricante de sistemas de telecomunicações adota um padrão específico para seus Registros de Detalhes de Chamadas (CDRs), adaptado às suas soluções e tecnologias.

Segundo Luo *et al.* (2020) esses padrões definem a estrutura, os campos e os métodos de codificação, atendendo às necessidades de monitoramento, faturamento e análise de redes móveis.

A Huawei, por exemplo, utiliza formatos binários proprietários como UCBILL, CDR e SOFTX, projetados para eficiência no armazenamento e transmissão de dados. Esses formatos incluem informações como identificadores do assinante (MSISDN, IMSI, IMEI), horários de início e término das chamadas, tipo de evento (voz, SMS, dados) e volume de dados transferidos. A Huawei também oferece o formato MINI, um CDR simplificado em texto, adequado para cenários onde o faturamento não é nec-

sário (HUAWEI, 2023a).

Essa diversidade de padrões destaca a importância de compreender as especificações técnicas de cada fabricante para integrar e analisar CDRs de forma eficaz.

2.2.4 Exemplos e Modelos de CDR

O tamanho de um Registro de Detalhes de Chamadas (CDR) da Huawei varia conforme o formato e os campos incluídos.

Por exemplo, em determinados formatos, a estrutura pode consistir em um cabeçalho fixo de 50 bytes, seguido por um cabeçalho repetitivo de 4 bytes para cada registro.

O tamanho total de um CDR é influenciado pelos campos adicionais que podem ser configurados. A Tabela 1 é de um CDR com tamanho de 59 bytes em sua estrutura.

Tabela 1 – Estrutura de um Registro CDR

Campo	Tamanho (bytes)	Descrição
MSISDN	11	Número de telefone do assinante (ASCII).
IMSI	15	Identificação única do assinante móvel.
IMEI	15	Identificação única do dispositivo móvel.
Start Time	4	Horário de início da chamada (UNIX timestamp).
End Time	4	Horário de término da chamada (UNIX timestamp).
Call Duration	4	Duração da chamada, em segundos.
Call Type	1	Tipo de evento (1=VOICE, 2=SMS, 3=DATA).
Data Volume	4	Volume de dados usados, em bytes.

Segue um exemplo de CDR Bruto do equipamento NetEngine AR6300 (HUAWEI, 2023b), apresentado na Figura 4 que consiste em dados básicos e compactos, utilizados principalmente para faturamento e monitoramento da rede.

A Figura 5 ilustra um CDR enriquecido, que inclui informações contextualizadas e detalhadas, ideal para análises avançadas e suporte a decisões estratégicas.

Figura 4 – Exemplo de CDR Bruto

000000000	31 31 36 39 36 34 34 39 30 38 38 33 31 30 32 36	1169644908831026
00000010	30 31 34 32 38 37 35 33 31 36 33 35 36 39 37 33	0142875316356973
00000020	37 39 34 30 39 34 34 31 39 39 60 61 a8 5a 60 61	7940944199`a.Z`a
00000030	b0 b5 00 00 08 5b 02 00 00 06 d7 31 31 36 37 31[....11671
00000040	33 32 39 38 35 36 33 31 30 32 36 30 38 38 39 32	3298563102608892
00000050	34 35 32 31 30 33 35 36 39 31 36 30 31 32 35 38	4521035691601258
00000060	32 32 31 39 31 64 c5 49 01 64 c5 49 be 00 00 00	22191d.I.d.I....
00000070	bd 03 00 00 19 ad 31 31 36 38 35 35 33 30 39 351168553095
00000080	34 33 31 30 32 36 30 31 34 31 36 39 31 39 35 36	4310260141691956
00000090	33 35 36 39 38 31 39 34 32 36 38 30 35 39 30 32	3569819426805902
000000a0	64 33 70 63 64 33 75 28 00 00 04 c5 03 00 00 15	d3pcd3u(.....
000000b0	79 31 31 33 37 37 35 33 30 38 37 36 33 31 30 32	y113775308763102
000000c0	36 30 35 30 32 38 35 31 37 31 38 33 35 36 39 33	6050285171835693
000000d0	37 32 32 30 33 36 33 36 39 30 35 62 e6 77 67 62	72203636905b.wgb
000000e0	e6 78 b2 00 00 01 4b 01 00 00 12 52 31 31 32 32	.x...K...R1122
000000f0	34 30 35 39 38 33 37 33 31 30 32 36 30 36 32 39	4059837310260629
00000100	34 38 35 31 32 32 33 35 36 39 37 36 33 38 37 32	4851223569763872
00000110	30 31 38 37 35 36 62 89 68 6a 62 89 6b db 00 00	018756b.hjb.k...
00000120	03 71 02 00 00 1e 1d 31 31 36 37 38 35 37 33 31	.q....116785731
00000130	37 36 33 31 30 32 36 30 34 33 31 33 39 33 34 33	7631026043139343
00000140	32 33 35 36 39 36 34 30 37 38 33 38 38 33 33 31	2356964078388331
00000150	30 64 85 42 82 64 85 43 46 00 00 00 c4 01 00 00	0d.B.d.CF.....
00000160	11 46 31 31 38 37 34 30 39 34 33 36 34 33 31 30	.F11874094364310
00000170	32 36 30 33 34 33 39 35 33 36 35 34 33 35 36 39	2603439536543569
00000180	36 39 30 38 36 37 34 34 35 33 31 36 63 cb 96 2c	690867445316c...
00000190	63 cb 9b 2c 00 00 05 00 01 00 00 15 03 31 31 37	c.....117
000001a0	38 30 35 32 35 39 35 36 33 31 30 32 36 30 32 34	8052595631026024
000001b0	36 36 33 37 32 32 38 33 35 36 39 36 36 35 34 30	6637228356966540
000001c0	36 32 37 36 39 39 33 64 42 07 f9 64 42 0b ca 00	6276993dB..dB...
000001d0	00 03 d1 01 00 00 14 ef 31 31 32 32 32 32 32 3811222228
000001e0	32 30 33 33 31 30 32 36 30 31 31 37 35 33 35 34	2033102601175354
000001f0	34 38 33 35 36 39 33 31 35 30 36 33 35 33 33 33	4835693150635333
00000200	39 35 61 38 d0 b1 61 38 d5 6e 00 00 04 bd 02 00	95a8..a8.n.....
00000210	00 02 a7 31 31 39 30 33 38 32 31 35 38 34 33 31	...1190382158431
00000220	30 32 36 30 32 37 39 36 30 34 33 32 34 33 35 36	0260279604324356
00000230	39 31 33 35 34 31 31 39 39 33 31 39 31 63 85 7e	9135411993191c.~
00000240	89 63 85 8a 7b 00 00 0b f2 01 00 00 25 c7 31 31	.c.{.....%.11
00000250	39 38 35 37 35 33 38 35 33 31 30 32 36 30 34	9857553853102604
00000260	33 32 32 36 30 31 37 30 33 35 36 39 33 35 32 38	3226017035693528
00000270	37 30 30 33 34 33 34 31 60 eb 60 fe 60 eb 6b 28	70034341`..`.k(
00000280	00 00 0a 2a 01 00 00 21 a4 31 31 38 31 32 36 31	...*....!1181261
00000290	33 36 30 32 33 31 30 32 36 30 38 32 39 36 37 34	3602310260829674
000002a0	30 30 31 33 35 36 39 35 36 36 36 34 32 31 31 39	0013569566642119
000002b0	39 32 36 60 29 c4 2b 60 29 c8 58 00 00 06 0d 01	926`).K`).X.....
000002c0	00 00 05 1e 31 31 34 39 30 30 38 33 34 35 31 33114900834513
000002d0	31 30 32 36 30 32 37 31 31 34 32 35 30 39 33 35	1026027114250935
000002e0	36 39 36 35 32 31 36 39 33 37 30 37 38 35 63 84	69652169370785c.
000002f0	96 80 63 84 9e 58 00 00 07 d8 02 00 00 1b ea 31	..c..X.....1
00000300	31 33 33 37 37 34 39 33 37 33 33 31 30 32 36 30	1337749373310260
00000310	36 35 30 33 36 38 36 32 38 33 35 36 39 34 38 34	6503686283569484
00000320	38 35 32 37 37 30 30 33 30 61 0b b1 62 61 0b b4	852770030a..ba..
00000330	4e 00 00 02 ec 01 00 00 14 8c 31 31 32 36 37 32	N.....112672
00000340	34 33 30 39 31 33 31 30 32 36 30 35 37 34 31 30	4309131026057410
00000350	30 31 36 32 33 35 36 39 34 33 37 31 32 33 35 33	0162356943712353
00000360	34 36 33 34 60 5c ab e0 60 5c ae 48 00 00 02 68	4634`\..\`.H..h
00000370	03 00 00 26 96 31 31 38 38 38 31 35 33 38 38 35	...&.11888153885
00000380	33 31 30 32 36 30 34 37 36 38 34 31 38 30 38 33	3102604768418083
00000390	35 36 39 31 35 38 33 33 38 32 31 34 37 35 37 65	569158338214757e
000003a0	6b 20 05 65 6b 2d 52 00 00 0d 4d 03 00 00 05 8e	k .ek-R...M.....
000003b0	31 31 35 39 39 33 34 30 30 33 38 33 31 30 32 36	1159934003831026
000003c0	30 31 36 33 33 37 38 30 30 33 35 36 39 34 32	0163378003356942
000003d0	37 31 39 39 36 36 37 32 37 38 65 06 62 98 65 06	7199667278e.b.e.
000003e0	6d 52 00 00 0a ba 01 00 00 0b 60 31 31 36 31 34	mR.....`11614
000003f0	30 38 31 36 31 30 33 31 30 32 36 30 36 33 35 32	0816103102606352
00000400	36 31 36 31 39 33 35 36 39 36 38 35 31 31 32 38	6161935696851128
00000410	36 32 36 37 36 62 c7 83 cd 62 c7 8f cb 00 00 0b	62676b...b.....
00000420	fe 01 00 00 08 28 31 31 33 35 39 35 31 38 33 36(1135951836
00000430	36 33 31 30 32 36 30 33 31 36 34 33 35 37 35 30	6310260316435750
00000440	33 35 36 39 31 35 36 35 33 39 31 36 34 31 38 34	3569156539164184
00000450	60 4f df 10 60 4f e0 32 00 00 01 22 01 00 00 23	`O..`O.2..."..#

Fonte: Elaborado pelo autor.

Figura 5 – CDR Enriquecido com outras Fontes

```

310260322567992|2024-11-20T08:00:00Z|2024-11-20T08:05:00Z|-46.729118|-23.64075|SMS|831|310-260-6650|2540|Site Recife|PA-XXXXXXXXXX|SC-XXXXXXXXXX|SE-XXXXXXXXXX|XXXXXXXXXXXXXXXXXXXX
310260382254532|2024-11-20T08:07:00Z|2024-11-20T08:12:00Z|-46.649434|-23.55515|DATA|1812|310-260-8127|12674|Site Manaus|BA-XXXXXXXXXX|PE-XXXXXXXXXX|XXXXXXXXXXXX
310260598765847|2024-11-20T08:14:00Z|-46.669324|-23.550598|SMS|71|310-260-7000|3973|Site Florianópolis|MG-XXXXXXXXXX|MA-XXXXXXXXXX|XXXXXXXXXXXX
310260325499849|2024-11-20T08:19:00Z|-46.669324|-23.550598|SMS|71|310-260-7000|3973|Site Florianópolis|MG-XXXXXXXXXX|MA-XXXXXXXXXX|XXXXXXXXXXXX
310260277669364|2024-11-20T08:21:00Z|2024-11-20T08:26:00Z|-46.639716|-23.567387|SMS|1825|310-260-1162|3551|Site São Paulo|PI-XXXXXXXXXX|AP-XXXXXXXXXX|XXXXXXXXXXXX
310260277669364|2024-11-20T08:28:00Z|2024-11-20T08:33:00Z|-46.628951|-23.561567|VOICE|217|310-260-4971|1942|Site Aracaju|MS-XXXXXXXXXX|AP-XXXXXXXXXX|XXXXXXXXXXXX
18260993344371|2024-11-20T08:35:00Z|2024-11-20T08:40:00Z|-46.581561|-23.663244|SMS|1818|310-260-8777|1681|Site Rio de Janeiro|PI-XXXXXXXXXX|SE-XXXXXXXXXX|XXXXXXXXXXXX
310260728315998|2024-11-20T08:42:00Z|2024-11-20T08:47:00Z|-46.657485|-23.688349|DATA|722|310-260-4980|14088|Site Aracaju|RJ-XXXXXXXXXX|RS-XXXXXXXXXX|XXXXXXXXXXXX
310260343059818|2024-11-20T08:49:00Z|2024-11-20T08:54:00Z|-46.65893|-23.617527|SMS|752|310-260-5244|1671|Site Belém|ES-XXXXXXXXXX|HT-XXXXXXXXXX|XXXXXXXXXXXX
310260839627920|2024-11-20T08:56:00Z|2024-11-20T09:01:00Z|-46.573227|-23.692553|SMS|17|310-260-6517|1609|Site Belém|ES-XXXXXXXXXX|HT-XXXXXXXXXX|XXXXXXXXXXXX
310260839627920|2024-11-20T09:01:00Z|-46.573227|-23.692553|SMS|17|310-260-6517|1609|Site Belém|ES-XXXXXXXXXX|HT-XXXXXXXXXX|XXXXXXXXXXXX
310260357614720|2024-11-20T09:03:00Z|2024-11-20T08:54:00Z|-46.65893|-23.527985|VOICE|719|310-260-7359|461|Site Cuiabá|TO-XXXXXXXXXX|MA-XXXXXXXXXX|XXXXXXXXXXXX
310260482955337|2024-11-20T09:17:00Z|2024-11-20T09:22:00Z|-46.630701|-23.667942|VOICE|373|310-260-9263|1323|Site Florianópolis|SE-XXXXXXXXXX|SE-XXXXXXXXXX|XXXXXXXXXXXX
310260279939491|2024-11-20T09:18:00Z|2024-11-20T09:15:00Z|-46.652881|-23.610349|DATA|293|310-260-9336|3067|Site Porto Velho|RJ-XXXXXXXXXX|RN-XXXXXXXXXX|XXXXXXXXXXXX
310260279939491|2024-11-20T09:18:00Z|2024-11-20T09:15:00Z|-46.652881|-23.610349|DATA|293|310-260-9336|3067|Site Porto Velho|RJ-XXXXXXXXXX|RN-XXXXXXXXXX|XXXXXXXXXXXX
310260569191475|2024-11-20T09:31:00Z|2024-11-20T09:36:00Z|-46.563402|-23.584134|DATA|815|310-260-8409|1462|Site Fortaleza|SE-XXXXXXXXXX|ES-XXXXXXXXXX|ES-XXXXXXXXXX
31026021575449|2024-11-20T09:38:00Z|2024-11-20T09:43:00Z|-46.713431|-23.632489|VOICE|968|310-260-5542|2115|Site Rio Branco|TO-XXXXXXXXXX|MG-XXXXXXXXXX|XXXXXXXXXXXX
310260958013917|2024-11-20T09:45:00Z|2024-11-20T09:50:00Z|-46.561709|-23.69556|DATA|18041|310-260-2758|3171|Site Florianópolis|PA-XXXXXXXXXX|AM-XXXXXXXXXX|XXXXXXXXXXXX
310260101623249|2024-11-20T09:45:00Z|-46.561709|-23.69556|DATA|18041|310-260-2758|3171|Site Florianópolis|PA-XXXXXXXXXX|AM-XXXXXXXXXX|XXXXXXXXXXXX
310260101623249|2024-11-20T09:57:00Z|-46.564431|SMS|18054|310-260-5911|2424|Site Recife|RJ-XXXXXXXXXX|DF-XXXXXXXXXX|XXXXXXXXXXXX
310260718925332|2024-11-20T09:59:00Z|2024-11-20T10:04:00Z|-46.695112|-23.634835|DATA|47|310-260-4953|14753|Site Belém|RJ-XXXXXXXXXX|AM-XXXXXXXXXX|XXXXXXXXXXXX
310260718925332|2024-11-20T10:06:00Z|2024-11-20T10:11:00Z|-46.561405|-23.517095|VOICE|164|310-260-5577|1208|Site Porto Velho|RJ-XXXXXXXXXX|RS-XXXXXXXXXX|XXXXXXXXXXXX
310260221560368|2024-11-20T10:13:00Z|2024-11-20T10:18:00Z|-46.6179851|-23.608924|DATA|1026|310-260-4971|1425|Site Boa Vista|RJ-XXXXXXXXXX|HT-XXXXXXXXXX|XXXXXXXXXXXX
3102606408542035|2024-11-20T10:20:00Z|2024-11-20T10:25:00Z|-46.622079|-23.545451|VOICE|1088|310-260-7929|1851|Site Brasília|BA-XXXXXXXXXX|PE-XXXXXXXXXX|XXXXXXXXXXXX
310260620880173|2024-11-20T10:27:00Z|2024-11-20T10:32:00Z|-46.700002|-23.574963|DATA|948|310-260-3001|4753|Site Fortaleza|AM-XXXXXXXXXX|DF-XXXXXXXXXX|XXXXXXXXXXXX
310260880346958|2024-11-20T10:34:00Z|2024-11-20T10:39:00Z|-46.687356|-23.54269|DATA|1174|310-260-5777|127|Site Brasília|PI-XXXXXXXXXX|PR-XXXXXXXXXX|XXXXXXXXXXXX
310260829376281|2024-11-20T10:41:00Z|2024-11-20T10:46:00Z|-46.704832|-23.660976|VOICE|1210|310-260-7076|1484|Site Belo Horizonte|GO-XXXXXXXXXX|RO-XXXXXXXXXX|XXXXXXXXXXXX
310260935397696|2024-11-20T10:48:00Z|2024-11-20T10:53:00Z|-46.647102|-23.586135|DATA|235|310-260-6403|3768|Site Maceió|MS-XXXXXXXXXX|AP-XXXXXXXXXX|XXXXXXXXXXXX
310260381168067|2024-11-20T10:55:00Z|2024-11-20T11:00:00Z|-46.70281|-23.669893|VOICE|694|310-260-5345|2452|Site Curitiba|BA-XXXXXXXXXX|PR-XXXXXXXXXX|XXXXXXXXXXXX
310260414970639|2024-11-20T11:02:00Z|2024-11-20T11:07:00Z|-46.697601|-23.535476|SMS|861|310-260-7190|4731|Site Curitiba|ES-XXXXXXXXXX|BA-XXXXXXXXXX|XXXXXXXXXXXX
310260473746895|2024-11-20T11:09:00Z|2024-11-20T11:14:00Z|-46.690885|-23.54267|VOICE|251|310-260-5128|3066|Site Palmas|MS-XXXXXXXXXX|PA-XXXXXXXXXX|XXXXXXXXXXXX
310260597617793|2024-11-20T11:16:00Z|2024-11-20T11:21:00Z|-46.647682|-23.668444|VOICE|583|310-260-4794|1876|Site Belém|MS-XXXXXXXXXX|PI-XXXXXXXXXX|XXXXXXXXXXXX
310260700413692|2024-11-20T11:23:00Z|2024-11-20T11:28:00Z|-46.657856|-23.548087|VOICE|467|310-260-5985|1117|Site Goiânia|ES-XXXXXXXXXX|BA-XXXXXXXXXX|XXXXXXXXXXXX
31026027878953|2024-11-20T11:30:00Z|2024-11-20T11:35:00Z|-46.689233|-23.63062|SMS|1006|310-260-5252|3975|Site Macapá|RJ-XXXXXXXXXX|PE-XXXXXXXXXX|XXXXXXXXXXXX
310260455531112|2024-11-20T11:37:00Z|2024-11-20T11:42:00Z|-46.672342|-23.547341|VOICE|971|310-260-6879|3800|Site Natal|RJ-XXXXXXXXXX|AC-XXXXXXXXXX|XXXXXXXXXXXX
310260484770967|2024-11-20T11:44:00Z|2024-11-20T11:49:00Z|-46.666001|-23.508995|SMS|793|310-260-2808|614|Site Teresina|AC-XXXXXXXXXX|SC-XXXXXXXXXX|XXXXXXXXXXXX
31026027359912|2024-11-20T11:51:00Z|2024-11-20T11:56:00Z|-46.729809|-23.618529|VOICE|1002|310-260-5272|4145|Site Florianópolis|SE-XXXXXXXXXX|GO-XXXXXXXXXX|XXXXXXXXXXXX
310260889322365|2024-11-20T11:58:00Z|2024-11-20T12:03:00Z|-46.697125|-23.666937|SMS|1930|310-260-1293|3758|Site Belo Horizonte|RJ-XXXXXXXXXX|PB-XXXXXXXXXX|XXXXXXXXXXXX
310260287309812|2024-11-20T12:05:00Z|2024-11-20T12:10:00Z|-46.609663|-23.618173|DATA|199|310-260-6165|1593|Site Florianópolis|SP-XXXXXXXXXX|PI-XXXXXXXXXX|XXXXXXXXXXXX
310260630915428|2024-11-20T12:12:00Z|2024-11-20T12:17:00Z|-46.555515|-23.534795|VOICE|1898|310-260-9643|214|Site São Luís|AL-XXXXXXXXXX|AM-XXXXXXXXXX|XXXXXXXXXXXX
310260692766804|2024-11-20T12:19:00Z|2024-11-20T12:24:00Z|-46.670644|-23.697597|VOICE|250|310-260-8329|3818|Site Belo Horizonte|PI-XXXXXXXXXX|AC-XXXXXXXXXX|XXXXXXXXXXXX
310260617350758|2024-11-20T12:26:00Z|2024-11-20T12:31:00Z|-46.613708|-23.698497|SMS|793|310-260-7521|4696|Site Rio de Janeiro|MS-XXXXXXXXXX|PE-XXXXXXXXXX|XXXXXXXXXXXX
310260228750928|2024-11-20T12:33:00Z|2024-11-20T12:38:00Z|-46.671329|-23.538465|VOICE|921|310-260-6689|1171|Site Goiânia|RJ-XXXXXXXXXX|AL-XXXXXXXXXX|XXXXXXXXXXXX
310260588183886|2024-11-20T12:40:00Z|2024-11-20T12:45:00Z|-46.605456|-23.503847|VOICE|824|310-260-3001|482|Site Macapá|AP-XXXXXXXXXX|CE-XXXXXXXXXX|XXXXXXXXXXXX
310260797534581|2024-11-20T12:47:00Z|2024-11-20T12:52:00Z|-46.716468|-23.621958|DATA|157|310-260-2751|4266|Site Manaus|CE-XXXXXXXXXX|PE-XXXXXXXXXX|XXXXXXXXXXXX
3102606785558707|2024-11-20T12:54:00Z|-46.669983|-23.57832|VOICE|683|310-260-6982|3872|Site Macapá|ES-XXXXXXXXXX|ES-XXXXXXXXXX|XXXXXXXXXXXX
310260649602286|2024-11-20T13:01:00Z|2024-11-20T13:06:00Z|-46.627054|-23.630065|DATA|266|310-260-3688|4992|Site Natal|RS-XXXXXXXXXX|PB-XXXXXXXXXX|XXXXXXXXXXXX

```

Fonte: Elaborado pelo autor.

3 REVISÃO SISTEMÁTICA E UM CASO DE REFERÊNCIA NO BRASIL

Compreender e modelar corretamente a mobilidade urbana é uma questão crucial para o desenvolvimento de cidades inteligentes. A estimativa de viagens individuais a partir de dados de posicionamento de telefones celulares, ou seja, registos de detalhes de chamadas (CDR) pode naturalmente apoiar estudos urbanos e de transporte, bem como aplicações de marketing (MAMEI *et al.*, 2019).

3.1 Revisão Sistemática da Literatura

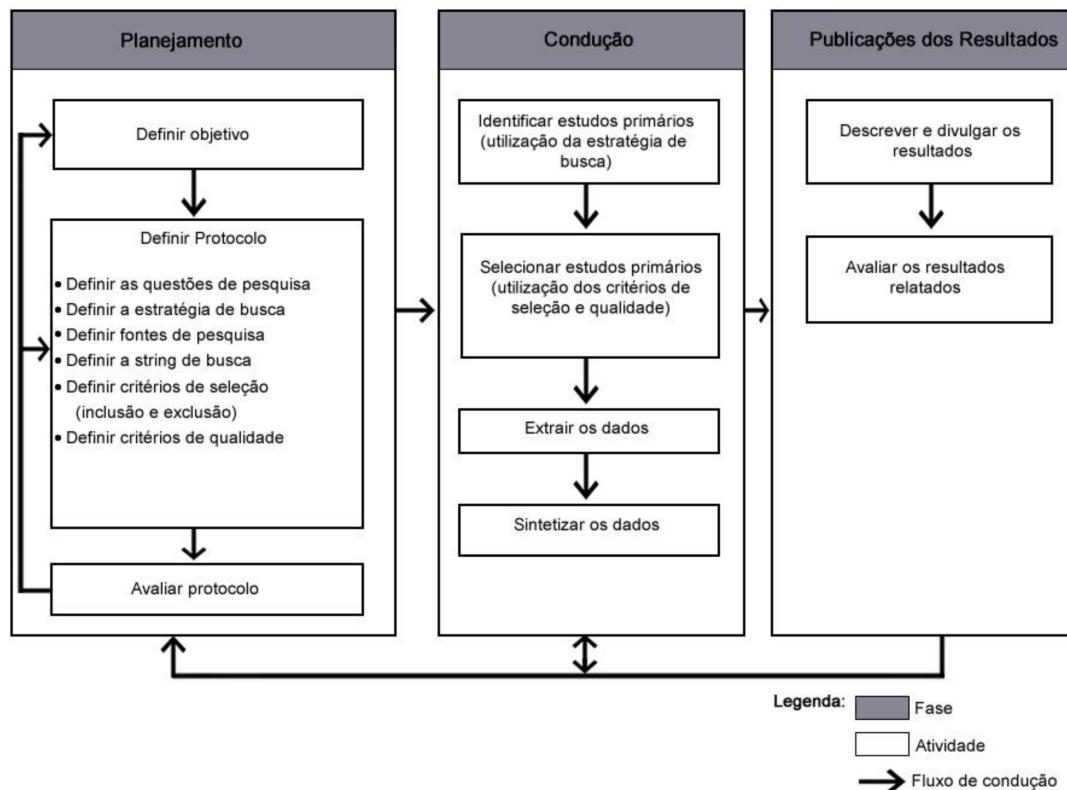
A revisão sistemática da literatura (RS) é uma abordagem rigorosa e estruturada para reunir, analisar e sintetizar pesquisas existentes em uma área específica. Segundo Felizardo *et al.* (2017) os protocolos de uma RS são fundamentais para garantir a qualidade e a transparência do processo de revisão. Esses protocolos incluem etapas como a formulação de perguntas de pesquisa, definição de estratégias de busca, critérios de inclusão e exclusão e métodos para extração e síntese de dados. A definição e refinamento desses protocolos, realizados iterativamente, ajudam a minimizar vieses, permitindo uma análise mais confiável dos estudos.

A RS é conduzida por meio de um processo com uma sequência de fases bem definidas, sendo elas:

- 1) Planejamento.
- 2) Condução.
- 3) Análise dos dados para publicação dos resultados.

A Figura 6 descreve cada uma destas fases.

Figura 6 – Sequência de Fases para a RS



Fonte: Felizardo *et al.* (2017)

3.1.1 Planejamento da Revisão Sistemática

Segundo Felizardo *et al.* (2017) o planejamento ou revisão sistemática é a etapa inicial em que são estabelecidas e apresentadas as estratégias e critérios que servirão de base para as fases subsequentes da revisão.

Como fontes de pesquisas foram estabelecidas as bases eletrônicas de consulta conforme apresentado na Tabela 2, sendo utilizada a string de consulta:

```
((("call detail record" OR "call detail records" OR "call details records" OR  
"calls details records") AND ("smart city" OR "smart cities" OR "smart finance"  
OR "smart grid" OR "smart health" OR "smart housing" OR "smart logistics"  
OR "smart mobility" OR "smart payment" OR "smart tourism" OR "smart envi-  
ronment")))
```

Base Eletrônica	Site
IEEEXplorer	https://ieeexplore.ieee.org/
Scopus	https://www.scopus.com/
ACM Digital Library	https://dl.acm.org/
Web of Science	https://www.webofscience.com/

Tabela 2 – Bases eletrônicas utilizadas na RS

Critérios de Inclusão:

I1. Estudos que representem modelos de trabalho, técnicas e processos para o uso de CDRs na infraestrutura de redes celulares aplicados em algum serviço para a gestão de Cidades;

I2. Estudos recentes publicados entre 2019 e 2022; foram consideradas bases de dados de acesso aberto com artigos revisados por pares.

Critérios de Exclusão:

E1. Estudos que utilizem outras tecnologias (GPS, Bluetooth, Wi-Fi, etc.) do ecossistema de ambiente celular, que não a análise de CDR, para fornecer serviços inteligentes em cidades.

3.1.2 Condução

Para a condução da revisão sistemática desta dissertação, foram utilizadas duas ferramentas de apoio: o Parsifal (COMPLEX, 2021) e o Microsoft Excel.

O Excel foi utilizado como suporte nas fases 1 e 2 descritas na seção 3.1 deste trabalho (“Estratégia para seleção dos trabalhos”), enquanto o Parsifal foi empregado para importar os metadados das pesquisas realizadas nas bases de dados.

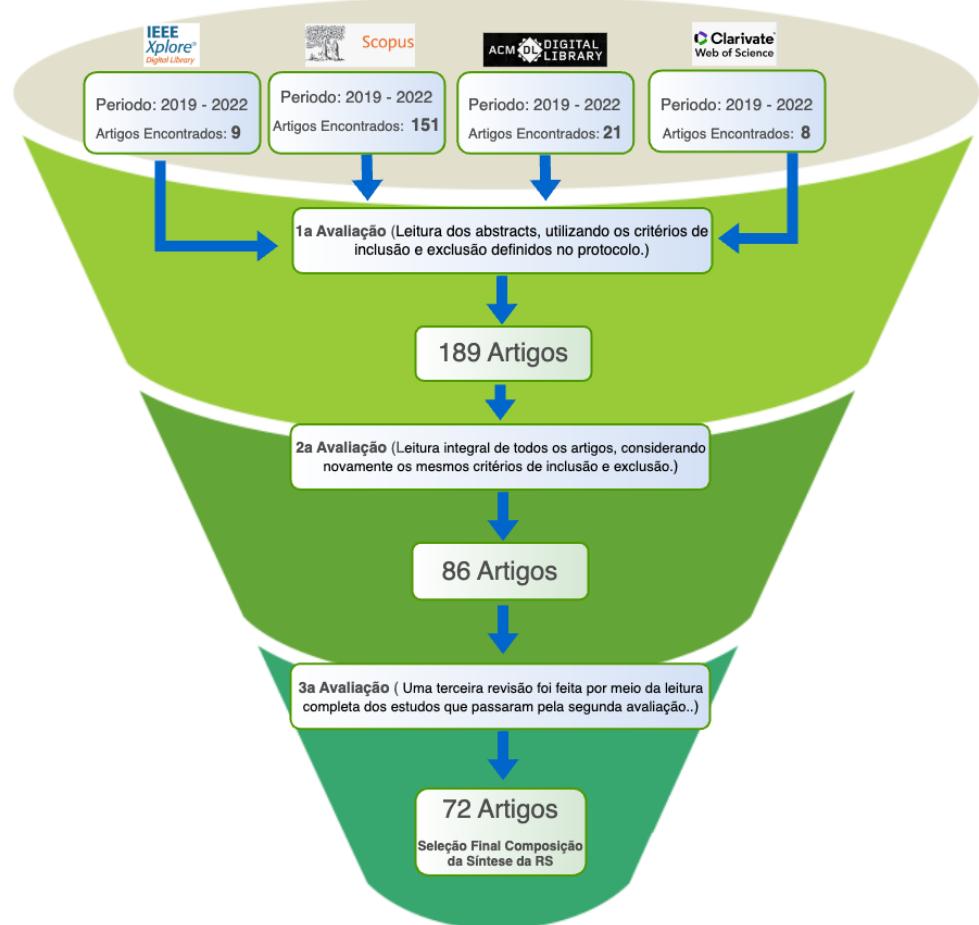
Com base nos critérios de inclusão e exclusão estabelecidos, foi realizada uma revisão sistemática que resultou inicialmente em 189 artigos. Desses, 86 artigos foram identificados como relevantes para responder às perguntas de pesquisa, pois atendiam ao critério de uso de dados de telefonia, especificamente CDR (Registros de Detalhamento de Chamadas), para a criação de uma base de caminhos e de origem-destino, destinada à aplicação em serviços de cidades inteligentes. Após a terceira revisão, com a leitura integral dos artigos e a aplicação rigorosa dos critérios de inclusão e exclusão, a lista final foi consolidada e está disponível no repositório público: ¹.

Na Figura7 é demonstrado o processo de busca e seleção de artigos na revi-

¹ https://github.com/edermau/RSL_Anexo_Mestrado

são sistemática, detalhando as etapas seguidas com base nos critérios de inclusão e exclusão definidos para a pesquisa.

Figura 7 – Busca de Artigos nas Bases



Fonte: Elaborado pelo autor.

Essa triagem permitiu focar nos principais serviços de cidades inteligentes que podem ser beneficiados pelo uso de CDR, respondendo questões sobre como dados de CDR podem ser empregados para criar matrizes de origem-destino e oferecer uma base sólida para o desenvolvimento de soluções urbanas baseadas em mobilidade.

3.1.3 Consolidação da Revisão

Esta revisão sistemática resultou em um artigo (BARBOSA *et al.*, 2023), que foi apresentado e publicado em 2023 nos anais do IEEE International Smart Cities Conference. Sendo que esta consolidação representa as oportunidades do uso de Bigdata baseados em CDR para fomentar Cidades Inteligentes.

A aspiração por cidades mais inteligentes reflete o desejo coletivo por um desenvolvimento urbano mais sustentável e eficiente. Este movimento em direção a um mo-

delo de cidade sustentável está se tornando cada vez mais prevalente, reformulando significativamente a gestão e administração urbanas. Neste contexto, a infraestrutura de Tecnologias da Informação e Comunicação (TIC) revela-se uma fonte vital de inovação, oferecendo novas maneiras de integrar cenários de "big data" para aprimorar a gestão das cidades inteligentes, como afirma MYEONG, KIM e AHN (2021).

Contudo, a implementação dessas inovações enfrenta obstáculos, notadamente no que tange à formulação e aplicação de políticas públicas adequadas. Estas políticas são cruciais para viabilizar a transição para cidades inteligentes, exigindo uma abordagem cuidadosa para superar os desafios existentes(BERNARDI; ET AL., 2020).

Consoante o estudo Barbosa *et al.* (2023) e conforme ilustrado na Figura 8, o mapeamento sistemático realizado destaca as principais oportunidades oferecidas pelos dados de infraestrutura de redes celulares no apoio aos centros de controle de cidades inteligentes.

Figura 8 – Mapeamento Sistemático



Fonte: Barbosa *et al.* (2023)

Os CDRs são utilizadas para melhorar o planejamento de transporte, detectando modos de transporte detalhados e padrões de viagem. Além disso, têm potencial para otimizar perfis de ocupação em tempo real, essenciais para o planejamento energético, resposta a emergências e adaptação do fornecimento de transporte público(BARBOSA *et al.*, 2023)

Os Registros de Dados de Chamadas CDR encontram várias aplicações práticas nas cidades inteligentes, contribuindo significativamente para a otimização da

gestão urbana e o melhoramento da qualidade de vida dos seus habitantes (WANG, Y. et al., 2019). Abaixo, segue algumas dessas aplicações práticas:

- Análise de Mobilidade: CDRs são utilizados para entender padrões de mobilidade da população, ajudando a otimizar rotas de transporte público, reduzir congestionamentos e planejar infraestruturas de transporte, podem indicar áreas de alta atividade e crescimento populacional, orientando decisões sobre onde expandir serviços e infraestrutura conforme Rahimi-Golkhandan, Garvin e Wang (2021).
- Otimização de Serviços Públicos e Gestão de Saúde Pública: Segundo Maha-jan, Cantelmo e Antoniou (2021) a análise de padrões de deslocamento ajuda a melhorar a localização de serviços públicos como escolas, hospitais e parques.
- Monitoramento de Doenças: Durante surtos de doenças, CDRs podem ser utili-zados para rastrear movimentos de populações e entender a propagação de do-enças, contribuindo para estratégias de contenção conforme Carrillo et al. (2021).
- Monitoramento Ambiental: Informações derivadas de CDRs podem auxiliar no monitoramento de áreas urbanas sujeitas a problemas ambientais, como polui-ção do ar e ruído, permitindo a implementação de políticas mais eficazes para mitigação de acordo com Cumbane e Gidófalvi (2019).
- Melhoria da Conectividade: A análise de CDRs pode revelar áreas com defici-ênci a em serviços de telecomunicações, direcionando melhorias na cobertura e qualidade da conexão conforme DeAlmeida et al. (2021).
- Planejamento Urbano e Desenvolvimento de Infraestrutura:

O “big data” é utilizado para analisar tendências de crescimento populacional, otimizar a localização de serviços públicos e planejar o desenvolvimento de in-fraestruturas sustentáveis conforme Yeh, Lin e Huang (2022).

- Gestão de Transporte e Mobilidade Urbana: Através da análise de dados de mobilidade, as cidades podem melhorar sistemas de transporte público, reduzir congestionamentos e promover modos de transporte sustentáveis de acordo com Mamei et al. (2019).
- Eficiência Energética e Sustentabilidade: “big data” apoia o planejamento energé-tico urbano, otimizando o consumo de recursos e incentivando o uso de energias renováveis de acordo com Popova e Popovs (2022).
- Gestão de Emergências: Em emergências, CDRs podem ser analisados para identificar áreas com maior necessidade de recursos de socorro e para coordenar a resposta de emergência de forma mais eficaz como afirma Cumbane e Gidófalvi

(2019).

- Segurança Pública e Monitoramento: O monitoramento em tempo real através do “big data” aprimora a segurança pública, auxiliando na prevenção e resposta a atividades criminosas de acordo com Kozik *et al.* (2021).

Apesar de suas inúmeras vantagens, a implementação do “big data” em cidades inteligentes enfrenta desafios relacionados à privacidade dos dados, segurança cibernética e a necessidade de infraestruturas tecnológicas avançadas. Além disso, é fundamental abordar as considerações éticas, garantindo que a análise de dados seja conduzida de maneira responsável e transparente.

A integração do “big data” nas cidades inteligentes oferece uma oportunidade sem precedentes para melhorar a gestão urbana, promovendo ambientes mais eficientes, seguros e sustentáveis. À medida que avançamos, a adoção responsável e ética do “big data” será essencial para realizar seu potencial transformador.

3.2 Estudo de caso: Índices de Isolamento Social para Enfrentar a COVID no Brasil

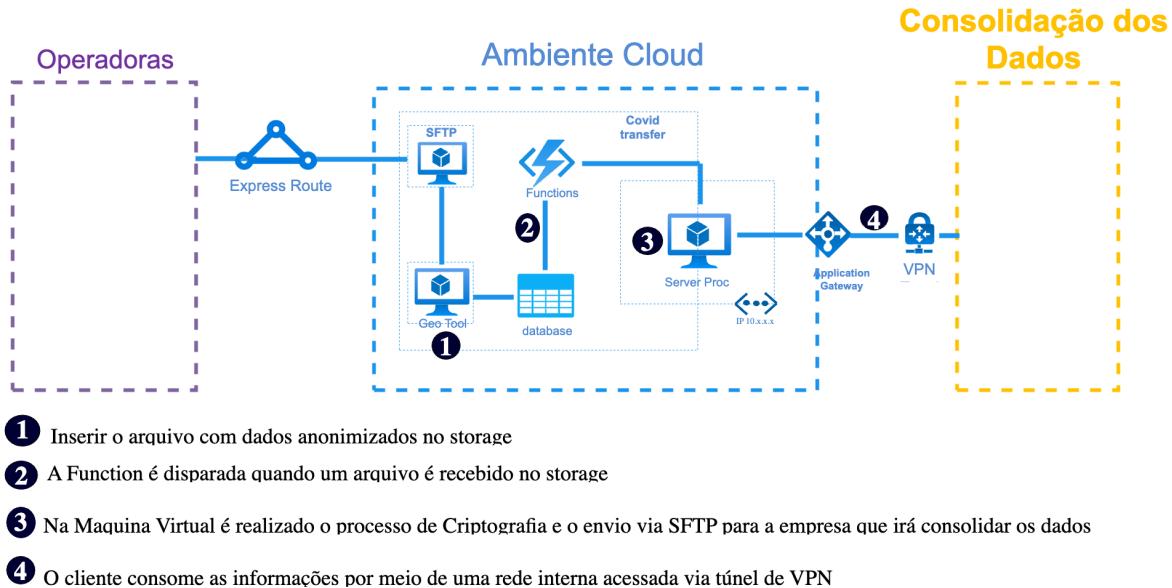
A pandemia de COVID desencadeou uma crise de saúde pública global sem precedentes, levando a uma necessidade urgente de monitorar e promover o isolamento social como uma medida chave para reduzir a transmissão do vírus. Neste contexto, a análise de dados de mobilidade provenientes de operadoras de telecomunicações emergiu como uma ferramenta crítica para avaliar a eficácia das políticas de isolamento.

O projeto COVID de Identificação do Índice de Isolamento Social na Cidade de São Paulo, conforme BRASIL (2020), teve como foco a identificação diária da porcentagem da população que estava respeitando as normas de distanciamento social durante a pandemia da COVID. Utilizou-se, como fonte de dados, as conexões de rede móvel, de forma agrupada e anonimizada, respeitando integralmente a Lei Geral de Proteção de Dados (LGPD). Este trabalho contou com a colaboração das quatro principais operadoras de telecomunicações do Brasil e destacou-se como um exemplo significativo de aplicação tecnológica voltada para o bem-estar público em tempos de crise sanitária.

Nesse projeto, atuei como líder técnico, sendo responsável pela construção da arquitetura para processar os dados de mobilidade baseados em CDR (Registros de Detalhamento de Chamadas), garantindo eficiência no tratamento dos dados e conformidade com os requisitos legais e técnicos.

Cada Operadora executou internamente o seguinte Procedimento para criar e consolidar as informações do índice de Isolamento Social, respeitando Anonimização e agrupamento dos dados conforme Figura 9.

Figura 9 – Procedimento para Criar e Consolidar as Informações do Índice de Isolamento Social



Fonte: Elaborado pelo autor.

Os dados finais foram consolidados por uma empresa neutra, e disponibilizados por meio de índices e mapas em uma plataforma Web para o Enfrentamento ao Covid pelo Governo do Estado de São Paulo.

O projeto COVID de integração de dados de quatro operadoras de telecomunicações distintas representa um desafio significativo em termos de interoperabilidade, governança de dados, e escalabilidade.

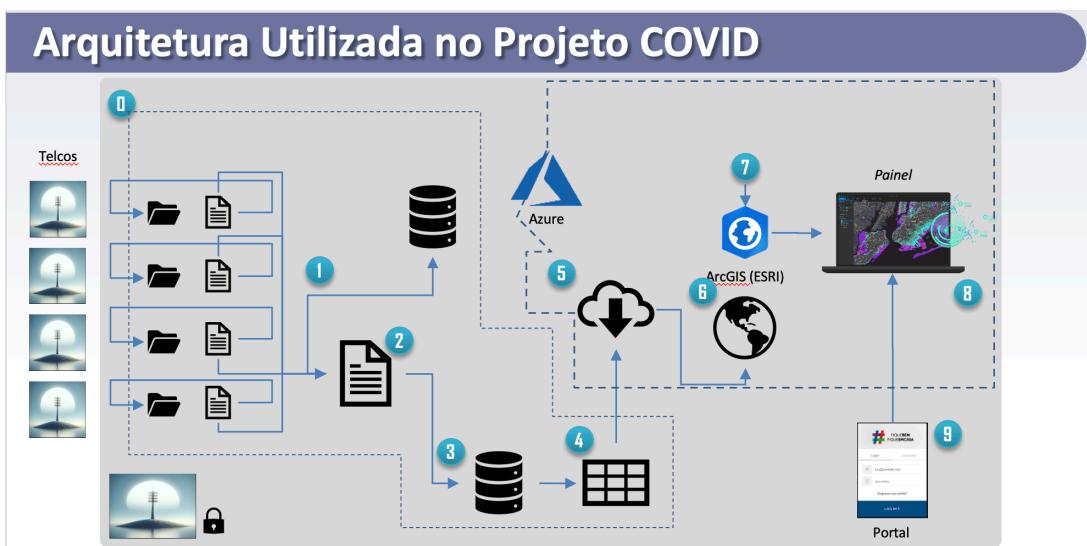
O objetivo central era criar uma arquitetura de dados unificada que pudesse processar, armazenar, e analisar grandes volumes de dados de CDR e outros tipos de dados gerados, mantendo ao mesmo tempo, altos padrões de segurança e privacidade.

A Figura 10 representa as dez etapas e arquitetura de referência.

- Primeira etapa: validações dos arquivos.
- Segunda etapa: armazenamento de dados gerais dos arquivos para rastreabilidade, anonimização e validações simples de conteúdo.
- Terceira etapa: consolidação dos arquivos e eliminação dos arquivos recebidos.
- Quarta etapa: criação de base de dados com o conteúdo dos arquivos para validações adicionais de qualidade.

- Quinta etapa: elaboração dos cálculos de percentuais em ambiente da Telecom escolhida para consolidar os dados para evitar acesso indevido às informações, tendo em vista a prontidão parcial das Prestadoras.
- Sexta etapa: acompanhamento e relatório do uso dos recursos em nuvem.
- Sétima etapa: administração do sistema de informação geográfica online.
- Oitava etapa: gestão da elaboração dos casos de uso, relatórios e fornecimento dos insumos.
- Nona etapa: elaboração do painel.
- Decima etapa: portal para acesso dos usuários.

Figura 10 – Arquitetura de Referência e Passos Utilizados no Projeto COVID



Fonte: Elaborado pelo autor.

3.2.1 Desafios e Precisões na Análise de Conjuntos de Dados

Segundo Jinzhong Wang *et al.* (2019) Registros de Detalhamento de Chamadas CDR são fundamentais para análises de mobilidade, embora apresentem desafios específicos. A principal questão é a variabilidade na frequência de registros, que impacta diretamente na confiabilidade dos estudos conduzidos. Os padrões de chamada demonstram uma anisotropia, com períodos de alta atividade seguidos por outros de calmaria, levantando o dilema de como estabelecer períodos de análise que sejam mais representativos. Adicionalmente, a questão da precisão na localização dos usuários se mostra premente. Em ambientes urbanos, a extensão da área de cobertura das antenas celulares pode oscilar significativamente, indo de poucas dezenas de metros até vários quilômetros, fazendo com que a localização geográfica inferida através das tor-

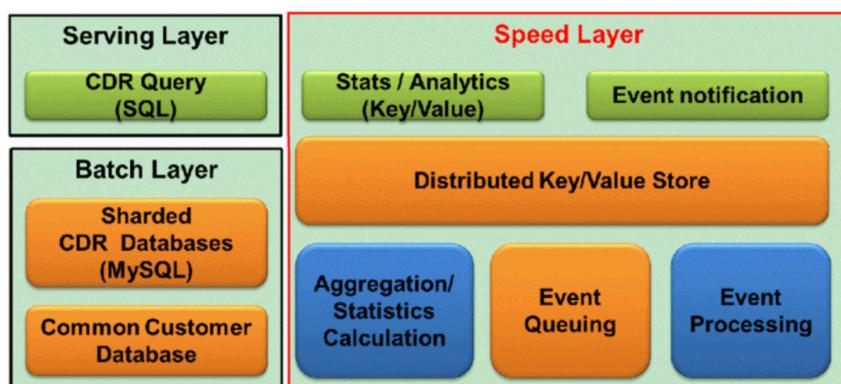
res celulares seja apenas uma aproximação da real posição do usuário. Tipicamente, o erro de localização pode variar entre 100 a 1000 metros, dependendo de uma série de fatores, incluindo a densidade de torres na área.

3.2.2 Arquitetura para Análise de Dados de CDR

No desenvolvimento do Projeto COVID, a equipe optou por empregar a arquitetura Lambda para a análise de dados. Esta escolha estratégica permitiu uma separação eficaz do processamento de dados em tempo real e do processamento de dados em lotes, facilitando assim uma análise mais abrangente e eficiente. Segundo Shih *et al.* (2014) a arquitetura Lambda é uma arquitetura de processamento de dados que visa satisfazer as necessidades de um sistema escalável, tolerante a falhas e robusto para atender a uma ampla gama de cargas de trabalho e casos de uso que requerem acesso de baixa latência. A arquitetura Lambda é composta por três camadas: camada “batch”, camada de serviço e camada de velocidade. Os dados recebidos são alimentados na camada de lote e na camada de velocidade. A camada de batch gerencia o conjunto de dados mestre e as visualizações de lote de pré-cálculo. A camada de serviço complementa a capacidade de consulta “ad-hoc”, que está faltando na camada em lote. É demorado disponibilizar novos dados na camada de serviço porque o índice e o pré-processamento levam tempo. A camada de velocidade compensa a alta latência de atualização da camada de serviço, mas processa apenas dados recentes.

A Figura 11 demonstra a arquitetura para plataforma analítica de CDR baseada em arquitetura Lambda utilizada no Projeto COVID. A camada de lote contém dois bancos de dados, um para CDR e outro para dados do cliente, ao passo que a camada de serviço atende a consultas de CDR.

Figura 11 – Arquitetura para Plataforma Analítica de CDR



Fonte: Shih *et al.* (2014)

3.2.3 De que Maneira os CDRs são Preparados para Distinguir Dados de Origem e Destino

A equipe de pesquisa do Projeto COVID desenvolveu uma técnica para mapear os locais de residência e trabalho dos indivíduos, utilizando uma combinação de Registros de Detalhamento de Chamadas CDR e informações provenientes de outros Sistemas Internos de cada Telecom. O cálculo era feito diariamente das 00:00 às 08:00; nesse horário, os usuários que conectavam na antena e permaneciam por um período caracterizavam como antena casa. A partir das 10:00, era analisado o total de tempo de permanência nessa antena, e isso determinava a taxa de isolamento social. Das 11:00 até às 22:00, eram calculadas a cada hora a taxa de isolamento, seguindo a mesma regra.

Observou-se também que ocorreram casos de usuários que conectavam em mais de uma antena; o critério para escolher a antena na qual o usuário permaneceu de fato era considerar o total de dados utilizados no período, e nessa situação, a antena que não tinha dados no mesmo período era descartada para esse usuário.

A taxa de isolamento era determinada na média do comportamento apresentado no intervalo das 22:00 às 10:00 e posteriormente esse dado era sumarizado por antena, bairro e município; após essa coleta, os dados eram enviados de forma anonimizada e sumarizados por município para entidade responsável em consolidar os dados das operadoras.

Seja $T_{conexão}(t)$ o tempo total de conexão de um usuário a uma antena em um determinado intervalo de tempo t , onde $t \in [0, 24]$ representa as horas do dia. A antena de residência é identificada pelo período das 00:00 às 08:00.

A taxa de isolamento social, τ_{iso} , é calculada como:

$$\tau_{iso} = \frac{1}{N} \sum_{i=10}^{22} \left(\frac{T_{conexão}(i)}{1000} \right), \quad (3.1)$$

onde N é o número total de horas consideradas no cálculo (das 10:00 às 22:00), e $T_{conexão}(i)$ é o tempo de conexão na antena de residência na i -ésima hora.

Um dado importante era manter um formato único conforme Tabela 3 dos arquivos de dados enviados para entidade centralizar as informações das operadoras, respeitando um layout idêntico entre as quatro operadoras referente a cada Parque de Antenas.

A Entidade, ao receber das operadoras os arquivos, realizava um agrupamento de município e bairro para enriquecer a visão e criar o índice de isolamento COVID do

Tabela 3 – Formato dos Arquivos com Dados de Coleta por Antena

Código Antena	Bairro	Município	Dia	Mês	Total Linhas	Intervalos
12345	Centro	Cidade XYZ	1	4	150	00:00-08:00
67890	Vila Nova	Cidade ABC	2	4	200	10:00-22:00
...

Fonte: Elaborado pelo autor.

Estado de São Paulo.

O método descrito em Luo *et al.* (2020), que identifica locais de residência e trabalho a partir de Registros de Detalhes de Chamadas (CDR), foi semelhante ao aplicado no caso de estudo do projeto COVID. Neste estudo, contudo, não se usaram dados da rede viária urbana. Uma diferença notável entre as duas abordagens é que, no projeto COVID, os dados de CDR foram compilados de várias operadoras e, em seguida, anonimizados.

Primeiro, os registros de CDR são combinados com dados de localização para mapear as trajetórias dos usuários. Em seguida, analisa-se a regularidade dessas trajetórias para identificar os locais de residência e trabalho, baseando-se no pressuposto de que as movimentações entre esses dois pontos são altamente regulares durante os dias úteis.

O método e os passos utilizados para determinar locais de residência e trabalho usando dados de (CDR) são detalhados da seguinte maneira:

Passo 1: Preparação e Agrupamento dos Dados: Os dados são agrupados por residente e por Origem-Destino (O-D), facilitando a análise subsequente.

Passo 2: Análise dos Grupos de Trajetórias: Identifica-se os dois grupos com o maior número de trajetórias para cada residente, focando nos horários acordados com as operadoras, correspondentes aos períodos de deslocamento para o trabalho e de retorno para casa.

Passo 3: Comparação e Identificação de Origem e Destino Comparação dos Grupos de Trajetórias: Se as origens e destinos de dois grupos são inversos entre si, esses são identificados como os locais de residência e trabalho. Esse passo baseia-se na regularidade dos padrões de deslocamento dos indivíduos durante os dias úteis.

3.2.4 Mapa de Origem x Destino — Caso de Uso de Dados Amostrais de CDRs.

Para calcular um Mapa de Origem e Destino no estudo de caso do projeto COVID foi utilizado uma estratégia similar a de Luo *et al.* (2020) que tem por base uma função matemática que calcula a origem e o destino de usuários durante dias úteis

utilizando (CDR), essa função foi simplificada e ajustada.

Definição dos Parâmetros: Definir os parâmetros de entrada da função, como o identificador do usuário, o “timestamp” da chamada, a localização da antena de telecomunicação, etc.

Determinação da Origem e do Destino:

- Origem: Determinar a localização da primeira chamada do usuário em dias úteis.
- Destino: Determinar a localização da última chamada do usuário em dias úteis.
- Filtragem de Dias Úteis: Aplicar critérios para filtrar os registros apenas para dias úteis.

Fórmula Matemática: Estruturar a fórmula que relaciona os parâmetros de entrada com as saídas desejadas (origem e destino).

Sejam:

- U_i o identificador único do usuário.
- T_{dia} o conjunto de “timestamp’s de chamadas durante os dias úteis.
- L_{CDR} as localizações correspondentes aos T_{dia} .

As funções para calcular a origem (O) e o destino (D) são definidas como:

$$O(U_i) = L_{CDR}(\min(T_{dia})) \quad (3.2)$$

$$D(U_i) = L_{CDR}(\max(T_{dia})) \quad (3.3)$$

Onde:

- $\min(T_{dia})$ e $\max(T_{dia})$ representam os horários da primeira e última chamada do usuário U_i em dias úteis, respectivamente.

A equação para calcular a distância média percorrida pode ser expressa como:

Considere a função de distância $D_{ist}(O, D)$, onde O e D representam as localizações de origem e destino. A distância média percorrida pelos usuários, D_{mdia} , é dada por:

$$D_{mdia} = \frac{1}{N} \sum_{i=1}^N D_{ist}(O(U_i), D(U_i)) \quad (3.4)$$

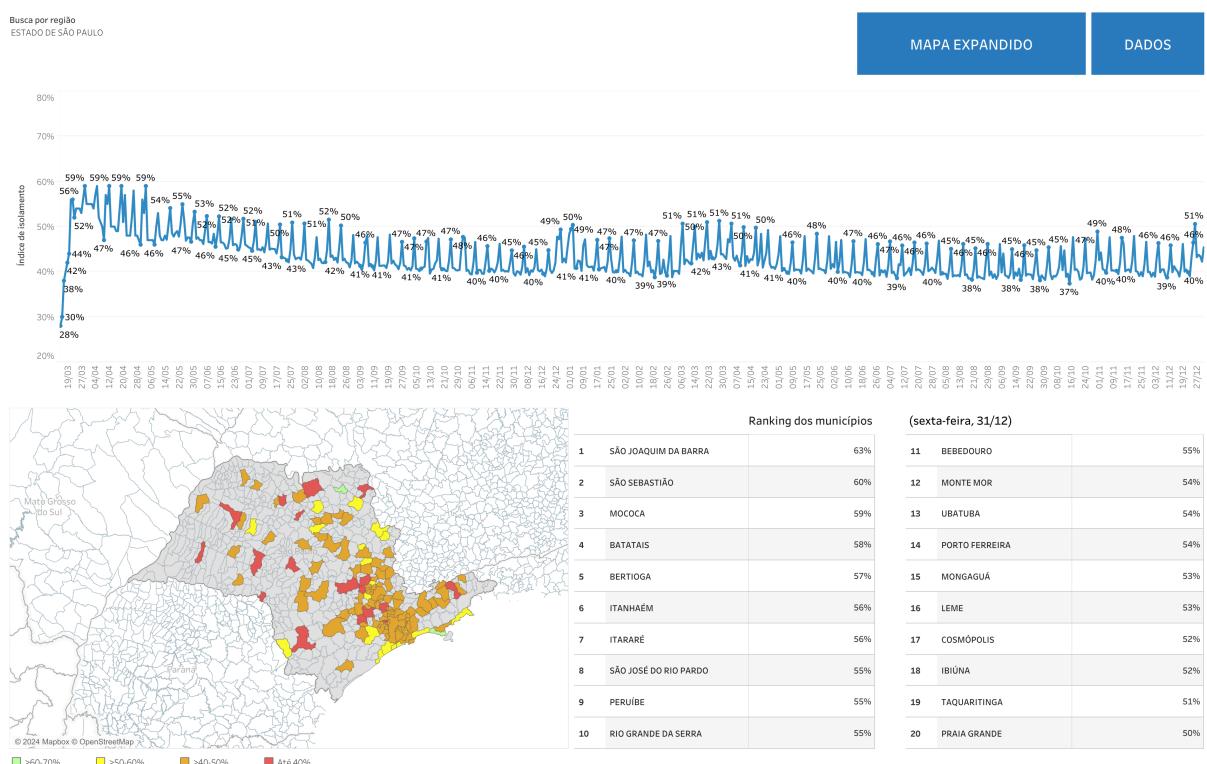
Onde:

- N é o número total de usuários.
- $D_{ist}(O(U_i), D(U_i))$ é a função que calcula a distância entre a origem O e o destino D do usuário i .

3.2.5 Painel Informativo Utilizado no Estudo de Caso

O Palácio dos Bandeirantes, sede do Governo do Estado de São Paulo e residência oficial do governador, desempenhou um papel importante na disseminação de informações durante a pandemia da COVID. Um Painel específico foi desenvolvido e frequentemente utilizado neste contexto conforme Figura 12 utilizando os dados de CDR do estudo de caso descrito, servindo como uma ferramenta vital para ilustrar e monitorar os dados e tendências relacionados ao coronavírus. Este painel informativo, que obteve ampla visibilidade ao ser exibido diversas vezes nas mídias de comunicação, tornou-se um recurso público essencial. Sua utilização efetiva no Palácio dos Bandeirantes evidencia a importância da transparência e do acesso à informação em tempos de crise, demonstrando o poder dos dados na gestão e no combate à pandemia.

Figura 12 – Sistema de Monitoramento Inteligente do Governo de São Paulo



O projeto foi realizado proativamente em prol de uma iniciativa social, que impactou positivamente toda população.

Através deste projeto, foi possível oferecer inteligência e insumos para os Governos tomarem as melhores medidas no combate à pandemia do COVID19.

3.2.6 Análise Crítica

O projeto promoveu uma importante ferramenta ao governo no enfrentamento da COVID para minimizar os impactos da pandemia, com uma gestão baseada em dados. A construção do processo resultou em lições, aprendizados, os quais podem ser vistos na Tabela 4

Tabela 4 – Lições, Aprendizados e Novas Perspectivas.

Pontos Positivos	Pontos Negativos
A colaboração em dados abre portas para a criação de novos serviços e produtos, impulsionando a inovação no setor. Por exemplo, a integração de dados de diferentes redes pode permitir o desenvolvimento de soluções IoT (Internet das Coisas) mais abrangentes ou melhorar a entrega de conteúdo digital.	Nem todas as operadoras tinham conhecimento em dados para compartilhar as informações no layout e tempo desejado.
Compartilhar dados relevantes pode facilitar a gestão coletiva da infraestrutura de telecomunicações, permitindo um planejamento mais eficaz da expansão da rede de telecomunicações.	Devido ao conflito de interesses e proteção de informações, foi necessário muito tempo para ter um contrato de proteção dos dados entre as operadoras, o que atrasou em muito o projeto.
Criar uma arquitetura de referência integrando dados na nuvem de forma anônima e com dados agrupados respeitando a LGPD foi complexo, porém inspirador para diversos outros projetos com essa característica.	Foi detectado muitos pontos nas cidades sem informações de mobilidade, o que atrapalhava em muito ter a visão clara de mobilidade em alguns pontos da cidade.
A criação de um algoritmo responsável pela geração dos CDR com as informações agrupadas e anônimas a serem compartilhadas entre as operadoras em um padrão único entre as operadoras participantes do projeto.	Muitas antenas se mostraram sobrecarregadas ou com falhas, o que atrasava a geração das informações consolidadas.
Integrar ferramentas de visualização com recursos de geolocalização demonstrando uma visão integrada para o centro de controle do Estado de São Paulo.	

3.3 Consolidação da Revisão e do Estudo de Caso

A revisão sistemática foi o instrumento utilizado para responder à questão de pesquisa: “Quais são os principais serviços de cidades inteligentes que poderiam utili-

zar dados de CDR?”.

Nesta jornada, o resultado da revisão gerou um trabalho publicado (BARBOSA *et al.*, 2023), com a análise de 72 artigos que apontam para a utilização de dados de CDR (Registros de Detalhamento de Chamadas) em serviços de cidades inteligentes, mapeando sua aplicabilidade em 9 eixos da norma ISO 37122. Esses eixos da ISO 37122 definem diretrizes para indicadores de cidades inteligentes, e o uso de CDR pode contribuir significativamente para monitorar e otimizar serviços urbanos.

Esses eixos evidenciam como o CDR é uma ferramenta valiosa para cidades inteligentes, proporcionando percepções essenciais para a melhoria de serviços e o planejamento urbano estratégico de cidades inteligentes.

No entanto, cada artigo apresentava metodologia de uso de CDR própria, não havendo uma uniformidade nos métodos empregados, promovendo dificuldades de implementação de diversos serviços na mesma base de informações de CDR, sendo que nesta dissertação, procurou-se obter um modelo base para diversos serviços inteligentes.

O método PICO (acrônimo para População, Intervenção, Comparação e Resultado) é apresentado por Scannavino *et al.* (2017) como uma ferramenta amplamente utilizada para estruturar questões de pesquisa em revisões sistemáticas.

Utilizando os critérios PICO, podemos decompor as respostas da pergunta de pesquisa e evidenciar por meios de análise dos estudos encontrados, motivos para caracterizar o estudo desta dissertação:

População (P): População urbana cujos dados de mobilidade podem ser analisados por meio de registros de CDR para apoiar o desenvolvimento de cidades inteligentes.

Intervenção (I): Aplicação de heurísticas spaçotemporais na análise de CDR pode construir uma base de dados que represente deslocamentos urbanos.

Comparação (C): Avaliação da eficácia de heurísticas spaçotemporais aplicadas a CDR para a construção de dados de deslocamento urbano, em comparação com abordagens tradicionais de coleta de dados de mobilidade, pode proporcionar novas perspectivas para a análise desses dados.

Resultado (O): Criação de uma base de dados de deslocamento urbano pode permitir um planejamento mais eficiente dos serviços urbanos, com foco em padrões de origem e destino para cidades inteligentes.

Hipótese: A aplicação de heurísticas spaçotemporais na análise de dados de

CDR permite a construção de uma base de dados de deslocamento urbano, capaz de fornecer informações detalhados sobre padrões de mobilidade de origem e destino, facilitando o planejamento e a gestão de serviços em cidades inteligentes.

Por outro lado, aplicação disso é um processo complexo, no entanto, o estudo de caso da implementação de estatísticas de isolamento social durante a pandemia demonstrou mecanismos técnicos e funcionais da aplicação do processamento de CDRs para um serviço específico para prover apoio ao Governo no enfrentamento da pandemia. Inclusive, o trabalho publicado por Santos, Barbosa *et al.* (2024) incluiu a descrição do processo da implementação dos índices de isolamento social em São Paulo, sendo mais uma contribuição desta dissertação.

No entanto, a implementação realizada no estudo de caso da COVID era muito específica e não constituía uma base que poderia ser utilizada para outros serviços de cidades, o que motivou e inspirou esta dissertação com a criação de bases amplas que pudessem apoiar serviços em cidades inteligentes, com várias vantagens.

4 ABORDAGEM EM CAMADAS COM HEURÍSTICAS ESPAÇOTEMPORAIS PARA CRIAÇÃO DE BASES DE DESLOCAMENTO

Após imersão e os aprendizados adquiridos durante o projeto relacionado à COVID, a proposta deste trabalho é desenvolver uma estrutura de dados nova, desenhada especificamente para enfrentar e superar os desafios observados no projeto COVID. Esta abordagem envolve a construção do modelo conceitual e a implementação da abordagem em camadas, que amplie o escopo aplicado no projeto COVID para outros serviços de gestão de cidades.

4.1 Modelo Conceitual

Os usuários de serviços de telecomunicações podem usufruir de conectividade com a infraestrutura celular durante a realização de trajetos na cidade, e este uso provê registros CDR, no entanto, é necessário aplicar algumas heurísticas para inferir a origem e destinos dos usuários. As heurísticas podem ser estratégias para simplificar a resolução de problemas e a tomada de decisões em situações complexas ou com informações incompletas. Essas estratégias são como “atalhos” que nos permitem chegar a conclusões rápidas, mesmo que não sejam perfeitas. Esse trabalho adotou um modelo conceitual que utiliza heurísticas espaciais e temporais para inferir origem e destino de usuários com base em serviços de cidades inteligentes.

O modelo conceitual visa estabelecer ambientes com bases de dados fundamentais *sandboxes*, que são pré-processadas, representando e segmentando grupos de dados, que podem ser utilizados com diversos propósitos em serviços inteligentes. É proposto um fluxo de trabalho em quatro fases, empregando heurísticas espaciais e temporais, por exemplo, para a análise de caminhos e Origem-Destino(O-D) de usuários em um ambiente de simulação *sandbox* conforme Figura 13.

Figura 13 – Abordagem em Camadas



Fonte: Elaborado pelo autor.

Este trabalho concentra-se nas fases de um a três, estabelecendo a quarta fase como uma oportunidade para pesquisa e desenvolvimento futuro.

- Na primeira fase, é realizada a coleta e o tratamento inicial dos dados provenientes da base CDR bruta. Esses dados contêm registros detalhados de chamadas e conexões móveis, mas antes de qualquer processamento avançado, os dados foco (usuários, local ou período) são coletados e passam por um processo de anonimização e separação. Durante essa fase, todas as informações que possam identificar usuários são individualmente removidas ou mascaradas. Além disso, a base é filtrada para incluir apenas os dados relevantes, como identificadores anonimizados de dispositivos, dados censitários (relacionados às áreas geográficas) e os horários precisos de entrada e saída nas Estações Rádio Base (ERB). Esse processo garante tanto a privacidade quanto a conformidade com legislações de proteção de dados.
- Na segunda fase, os dados resultantes são organizados em uma "sandbox de caminhos". Essa nova base de dados é construída com foco na identificação de padrões espaciais nos trajetos registrados. Para isso, aplicam-se heurísticas, que permitem segmentar os dados em blocos coerentes de movimentos no espaço e no tempo. Essas segmentações são criadas para agrupar pontos de partida e chegada próximos, compondo uma estrutura que facilita a análise de trajetos.
- Na terceira fase, as informações da sandbox de caminhos são utilizadas para a construção de uma "sandbox de origens e destinos". Nesse estágio, os agrupamentos anteriores são refinados e integrados, formando uma base de dados

estruturada com segmentos claros de origem-destino. Esses segmentos detalham o fluxo entre diferentes localidades, permitindo uma análise mais robusta dos padrões de deslocamento. A partir dessa fase, os dados são preparados para integrações posteriores e análises em contextos de mobilidade urbana.

Na quarta fase, as informações sobre origem e destino, que foram identificadas na fase anterior, podem ser classificadas conforme as necessidades do serviço inteligente, ou mesmo combinadas com fontes de dados externas, seja para agrupar ou identificar padrões de mobilidade foco do estudo ou serviços de análise.

Como resultados obtêm-se 2 bases fundamentais:

- A primeira base de dados é dedicada à criação de um mapa de caminhos, uma ferramenta essencial para visualizar e compreender os percursos e movimentos num determinado espaço ou contexto.
- A segunda base de dados focará na descoberta de origens e destinos. Essa parte da estrutura será valiosa para entender de onde as pessoas estão vindo e para onde estão indo, uma informação chave para inúmeras aplicações, desde o planejamento urbano até a otimização de rotas em ambientes complexos.

4.1.1 Segmentação Espacial sobre os CDRs

A aplicação de heurísticas espaciais aos dados apresentados é ilustrada conforme a Figura 14 por uma jornada hipotética, em que um usuário realiza deslocamentos de um ponto inicial em um bairro na cidade 1 até um destino em outra cidade. A análise dos dados de CDR (Registros de Detalhamento de Chamadas) possibilita a criação de diferentes estratégias para mapear trajetos, como as abordagens baseadas em ERBs, bairros ou cidades.

As conexões registradas ao longo do percurso são detalhadas, possibilitando a identificação de diferentes formas de representação dos caminhos. Para o agrupamento de registros que segmentam os deslocamentos adotou-se 3 possibilidades heurísticas:

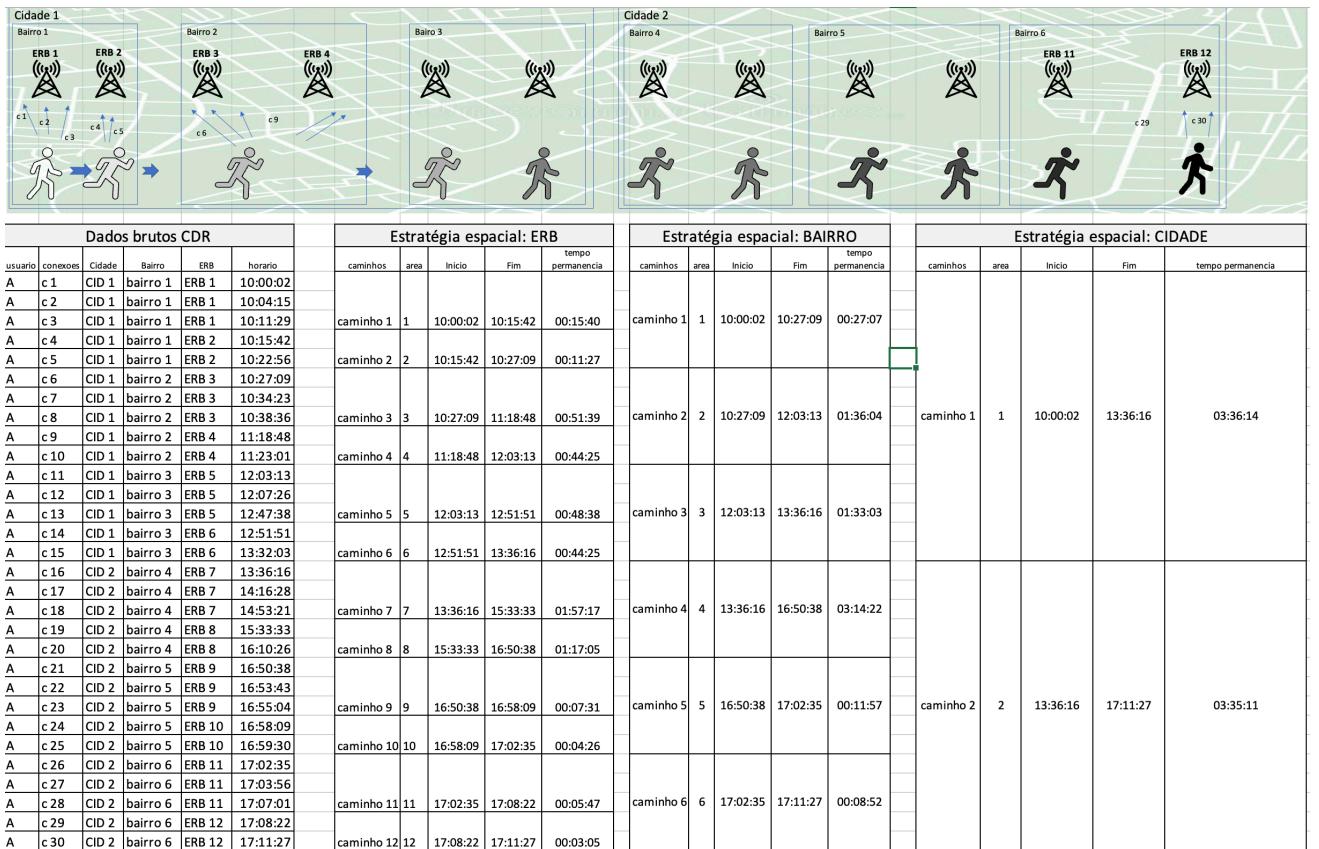
- Caminhos por ERBs: neste caso, adota-se como heurística espacial, que o caminho do usuário ocorre por regiões delimitadas pela ERBs a qual está conectado. Considerando a infraestrutura celular atual, esta é a menor granularidade espacial sem que utilize informações adicionais de outras fontes.
- Caminhos por Bairros: Ao agrupar as conexões em unidades espaciais maiores, os deslocamentos são categorizados com base nos bairros em que ocorrem. Esta heurística é útil para identificar padrões regionais de mobilidade, como rotas

casa-trabalho ou deslocamentos frequentes entre áreas adjacentes.

- Caminhos por Cidades: Esta abordagem, a mais ampla, segmenta os dados apenas em termos de cidades. Tal estratégia é adequada para análises que buscam entender deslocamentos intermunicipais ou fluxos entre grandes centros urbanos.

Essas heurísticas possibilitam a estimativa de tempos de permanência em cada região e estudos sobre padrões de mobilidade comportamentais com base em dados da infraestrutura de telecomunicações, identificando trajetos que conectam diferentes localidades entre Cidade 1 e Cidade 2.

Figura 14 – Segmentação Espacial dos CDRS



Fonte: Elaborado pelo autor.

Ainda na Figura 14 é possível observar o resultado do processamento dos dados brutos CDR, que utilizando diferentes heurísticas pode construir bases de caminhos com a estratégia de caminhos por ERB, caminhos por BAIRRO ou caminhos por CIDADE.

4.1.2 Segmentação Temporal para Construção das Bases O-D

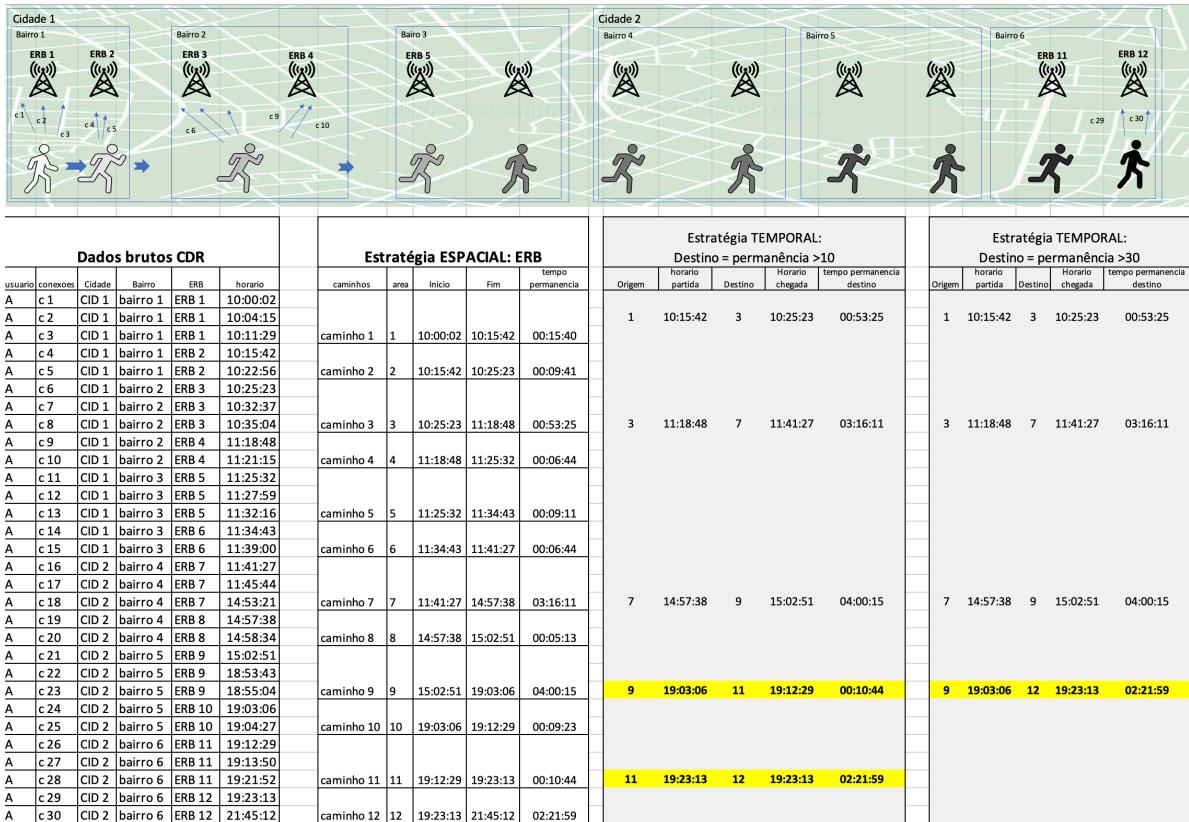
A construção das bases Origem-Destino é realizada utilizando uma heurística temporal para determinar os pontos de ORIGEM e DESTINO de cada trajetória. Na ausência de informações adicionais específicas sobre o comportamento dos usuários, essa abordagem permite aproximar a análise de mobilidade à realidade observada.

A heurística temporal é fundamentada no tempo de permanência em um local, considerando que a permanência prolongada pode indicar que o ponto é um DESTINO(JAKOVLJEVIC *et al.*, 2021). Embora diferentes tipos de serviços possam requerer valores distintos de tempo para caracterizar um destino, nesta análise foi implementado e adotado um parâmetro, TEMPO DE PERMANÊNCIA, para representar o mecanismo técnico que viabiliza a identificação de um DESTINO, gerando uma base O-D. Este conjunto de dado ainda pode ser a entrada no *sandbox* de serviços, que classificaria os tipos de Origem e destino, conforme a necessidade de análises dos serviços, como parte da fase 4 da Figura 13.

Para ilustrar a aplicação dessa heurística, a Figura 15 apresenta um caminho hipotético onde o usuário realiza deslocamentos iniciando no Bairro 1, na cidade 1, e visitando diversos destinos. Durante o percurso, são registradas conexões na infraestrutura celular, que servem para compor a base Origem-Destino com base em critérios temporais.

Na Figura 15 é apresentado o resultado do processamento conceitual da heurística temporal, que considera uma abordagem com a estratégia espacial de caminhos por ERBs, onde são aplicadas duas heurísticas temporais distintas: tempo de permanência superior a 10 minutos e tempo de permanência superior a 30 minutos. Essas abordagens permitem conceitualmente identificar pontos de origem e destino, considerando que conexões mais longas em determinados pontos sugerem localizações de maior relevância no trajeto do usuário. A Figura 15 apresenta o processamento das duas estratégias diferentes de agrupamento, destacando em amarelo as diferenças que refletem as escolhas por uma das heurísticas temporais.

Figura 15 – Segmentação Temporal dos CDRS



Fonte: Elaborado pelo autor.

4.2 Métodos e Implementação

A implementação do modelo conceitual exige uma série de procedimentos técnicos e científicos para a construção das bases.

Para a realização das análises e a construção dos “sandboxes” descritos, os registros detalhados de chamadas (CDR) foram disponibilizados em um formato mais acessível para manipulação e processamento. Inicialmente armazenados em sua forma compactada e binária, os CDR foram descompactados utilizando um software proprietário fornecido pela empresa de telefonia, que autorizou o uso e a manipulação desses dados exclusivamente para fins acadêmicos.

Após a descompactação, os CDR foram convertidos para o formato CSV e, em seguida, unificados em um único arquivo no formato XLSX Microsoft (2024) para facilitar as operações e análises no contexto do sandbox. Esse arquivo consolidado contém os principais campos necessários para a criação das bases de caminhos e de origem-destino. Entre os dados estruturados encontram-se identificadores anonimizados de dispositivos, informações temporais, dados espaciais (como a localização das ERBs)

e outros atributos essenciais para a análise de mobilidade.

O detalhamento completo dos campos recebidos e trabalhados está descrito nas Tabelas 6 e 7, onde são especificadas as características de cada campo.

Tabela 6 – Primeira Parte da Tabela com Colunas e Campos dos CDR Analisados

msisdn	mbou	dt_dia	cgi	hora_start	dt_start_time
Identificador único	Quantidade de dados utilizados	Dia do evento	Identificador da Antena	Hora de início	Timestamp no formato ano, mês, dia, hora, minuto, segundo

Fonte: Elaborado pelo autor.

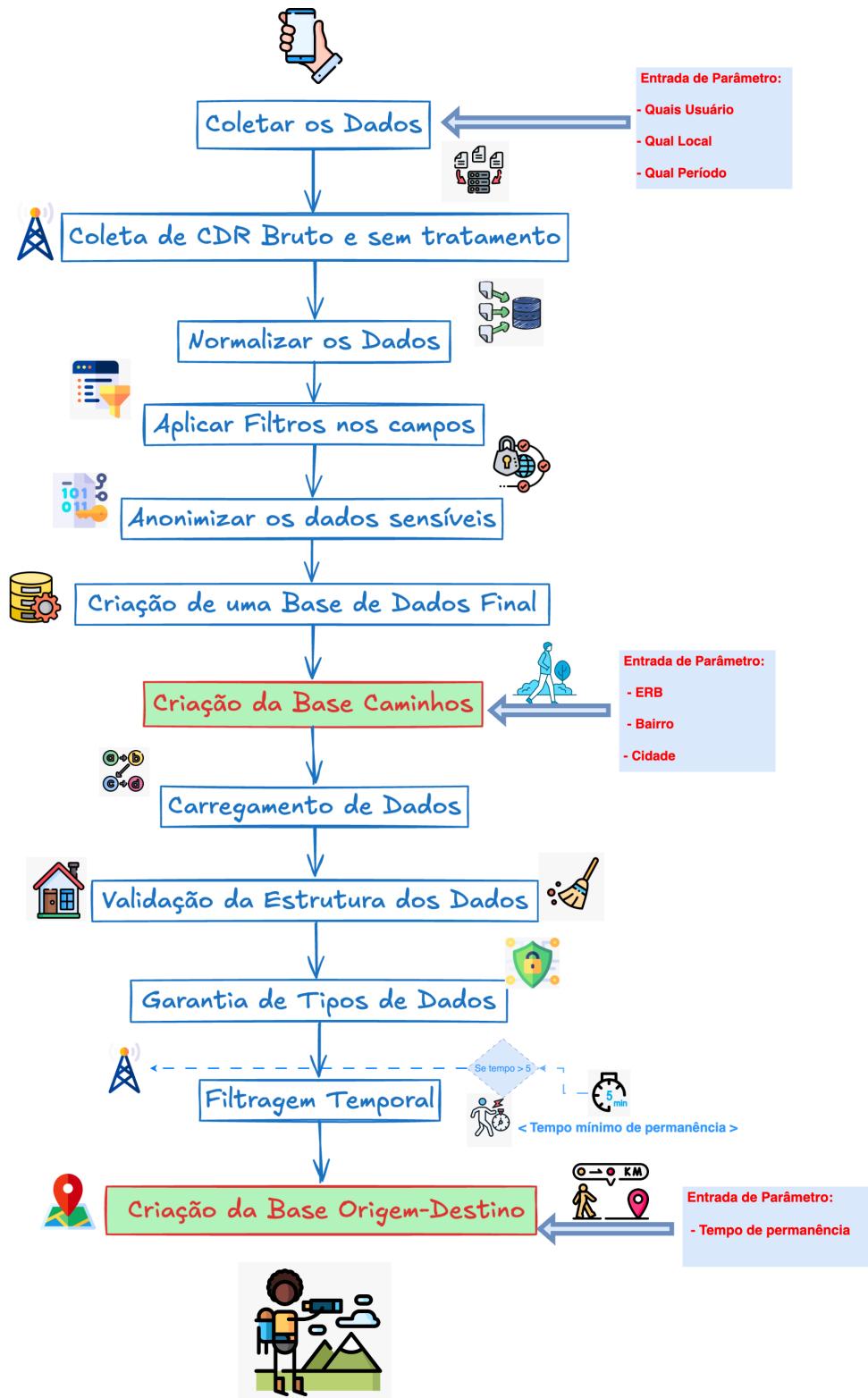
Tabela 7 – Segunda Parte da Tabela com Coluna Campos Adicionais dos CDR Analisados

ds_site	nome_site	long	lat	DT_HORA	Caminho
Código do site	Nome do site	Longitude	Latitude	Data e hora do registro	Condição específica

Fonte: Elaborado pelo autor.

Um fluxograma do procedimento completo é apresentado na Figura 16, proporcionando uma visão da sistemática das etapas realizadas:

Figura 16 – Fluxo e Passos das Bases



Fonte: Elaborado pelo autor.

- A etapa de coleta de CDR bruto e sem tratamento foi concebida para capturar os dados iniciais em seu estado bruto, sem qualquer pré-processamento. Essa coleta é parametrizável, permitindo que sejam definidos critérios específicos para

quais dados de CDR (Call Detail Records) serão coletados, independentemente do volume de informações.

- Na etapa de normalização dos dados, os registros coletados são ajustados e padronizados para garantir consistência e uniformidade. Esse processo também é parametrizável, permitindo que sejam aplicadas regras específicas para normalização conforme as necessidades do conjunto de informações.
- A etapa de aplicação de filtros nos campos foi desenvolvida para processar os dados com base em parâmetros definidos, como o conjunto de MSISDN (identificadores únicos de usuários) a ser analisado. Essa abordagem garante que apenas os dados relevantes sejam considerados no processamento, conferindo flexibilidade e precisão.
- Na fase de anonimização dos dados sensíveis, informações identificáveis, como CPF, MSISDN e outros dados que permitem a identificação direta ou indireta de indivíduos, são protegidos de maneira parametrizável, assegurando conformidade com regulamentos de privacidade e segurança. Os campos a serem anonimizados podem ser definidos conforme os requisitos da análise. No *sandbox* a anonimização foi aplicada no campo MSISDN.
- A criação de uma base de dados final organiza os dados tratados em um formato consolidado e estruturado no formato de arquivo XLSX. Essa etapa é parametrizável, permitindo que sejam definidos os critérios para inclusão e organização das informações, assegurando que a base seja ajustada às demandas do estudo.
- Na criação da base de caminhos, os dados consolidados são processados para identificar os fluxos de movimentação entre diferentes locais. O processo é parametrizável, possibilitando a definição de regras específicas para a detecção e registro de trajetos.
- A etapa de criação da base origem-destino estrutura os dados em um formato que detalha os deslocamentos entre os pontos de origem e destino. Essa criação é parametrizável, permitindo a configuração de filtros, como intervalos de tempo ou condições específicas, para delimitar os registros a serem considerados na base final.

4.2.1 Implementação da Heurística Espacial (Base de Caminhos)

O código-fonte, assim como descrição dos componentes, exemplos de utilização e métodos de coleta de dados estão disponíveis no repositório público:²

² <https://github.com/edermau/basedecaminhos>

De forma geral, a heurística espacial é iniciada com o mapeamento de conexões, cujas informações são processadas por meio do algoritmo responsável pela criação da base de caminhos.

A Coleta de dados dos CDR é realizada e armazenada em arquivo XLSX como entrada inicial do processo.

O algoritmo foi desenvolvido visando processar e anonimizar dados de mobilidade extraídos de um arquivo XLSX. Ele atende a necessidades de anonimização e análise, garantindo conformidade com exigências de privacidade, como a Lei Geral de Proteção de Dados (LGPD).

Os detalhes dos campos que serão apresentados nas operações abaixo podem ser consultados na Tabela 8.

Tabela 8 – Tabela da Base de Caminhos

msisdn	mbou	dt_dia	cgi	hora_start	dt_start_time
Identificador único	Quantidade de dados utilizados	Dia do evento	Identificador da Antena	Hora de início	Timestamp no formato ano, mês, dia, hora, minuto, segundo

next_dt_start_time	ds_site	next_ds_site	long	lat
Timestamp do próximo evento	Identificador do site atual	Identificador do próximo site	Longitude	Latitude

Fonte: Elaborado pelo autor.

Alguns blocos e informações relevantes sobre o algoritmo são apresentados a seguir:

- Dependências:

O algoritmo utiliza as bibliotecas pandas e openpyxl para manipulação de dados XLSX e hashlib para anonimização.

- Carregamento e Pré-processamento:

O algoritmo lê uma planilha específica do arquivo XLSX e converte a coluna de tempo para o formato datetime. Os dados são ordenados por tempo, e a diferença entre eventos consecutivos é calculada e armazenada.

- Anonimização:

O número de identificação do usuário (msisdn) é anonimizado usando a função SHA-256, garantindo privacidade e consistência nos dados anonimizados.

A função $anonymize_{msisdn}$ gera um *hash* criptográfico de 256 bits para proteger os identificadores de usuários, mantendo a integridade para análise repetida.

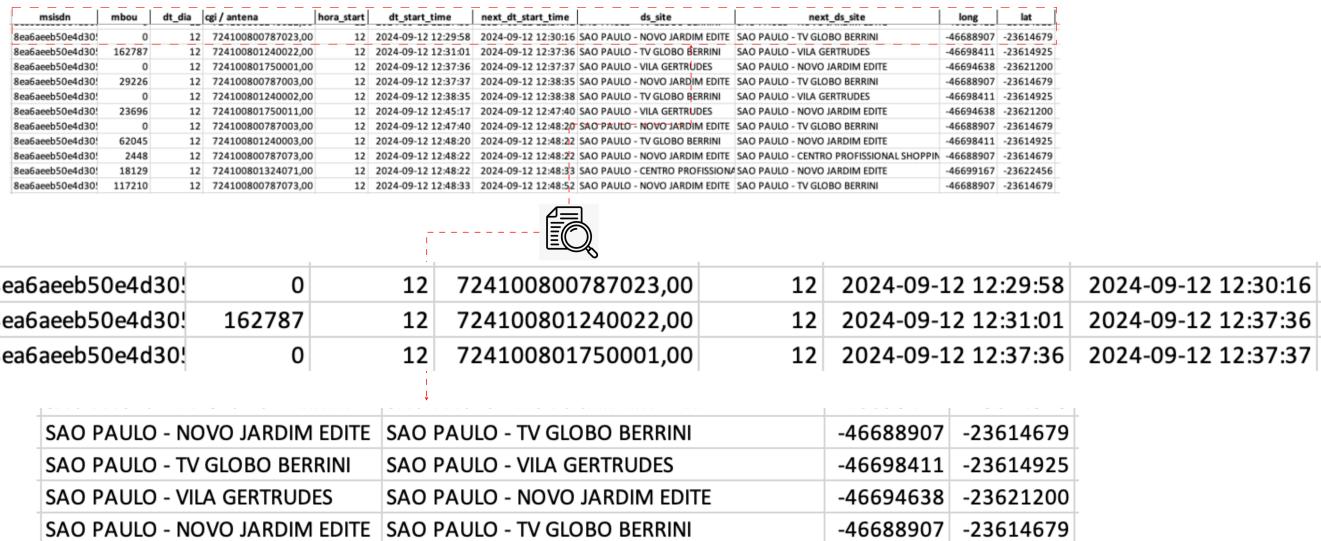
- Criação de uma Base de Dados Final:

A base final é organizada em colunas relevantes, como identificadores anonimizados, informações de tempo, localização e troca de sites, e salva em um novo arquivo XLSX para análise.

Esta base de caminhos identifica a origem ds_{site} e o destino $next_{ds_{site}}$ do usuário, removendo eventos consecutivos que ocorrem no mesmo local.

As práticas implementadas no algoritmo foram aplicadas para a geração de uma base de caminhos. O resultado do processamento dos dados de CDR foi consolidado em um arquivo XLSX, devidamente normalizado, conforme apresentado na Tabela 9.

Tabela 9 – Tabela com Campos dos CDR Analisados



msisdn	mbou	dt_dia	cgf / antena	hora_start	dt_start_time	next_dt_start_time	ds_site	next_ds_site	long	lat
8ea6aeeb50e4d30!	0	12	724100800787023,00	12	2024-09-12 12:29:58	2024-09-12 12:30:16	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907	-23614679,
8ea6aeeb50e4d30!	162787	12	724100801240022,00	12	2024-09-12 12:31:01	2024-09-12 12:37:36	SAO PAULO - TV GLOBO BERRINI	SAO PAULO - VILA GERTRUDES	-46698411	-23614925
8ea6aeeb50e4d30!	0	12	724100801750001,00	12	2024-09-12 12:37:36	2024-09-12 12:37:37	SAO PAULO - VILA GERTRUDES	SAO PAULO - NOVO JARDIM EDITE	-46694638	-23621200
8ea6aeeb50e4d30!	29226	12	724100800787003,00	12	2024-09-12 12:37:37	2024-09-12 12:38:35	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907	-23614679
8ea6aeeb50e4d30!	0	12	724100801240002,00	12	2024-09-12 12:38:35	2024-09-12 12:38:38	SAO PAULO - TV GLOBO BERRINI	SAO PAULO - VILA GERTRUDES	-46698411	-23614925
8ea6aeeb50e4d30!	23696	12	724100801750011,00	12	2024-09-12 12:45:17	2024-09-12 12:38:35	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - VILA GERTRUDES	-46694638	-23621200
8ea6aeeb50e4d30!	0	12	724100800787003,00	12	2024-09-12 12:45:17	2024-09-12 12:48:20	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - NOVO JARDIM EDITE	-46688907	-23614679
8ea6aeeb50e4d30!	62045	12	724100801240003,00	12	2024-09-12 12:48:20	2024-09-12 12:48:22	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - NOVO JARDIM EDITE	-46698411	-23614925
8ea6aeeb50e4d30!	2448	12	724100800787073,00	12	2024-09-12 12:48:22	2024-09-12 12:48:22	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - CENTRO PROFISSIONAL SHOPPIN	-46688907	-23614679
8ea6aeeb50e4d30!	18129	12	724100801324071,00	12	2024-09-12 12:48:22	2024-09-12 12:48:33	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - NOVO JARDIM EDITE	-46692167	-2362456
8ea6aeeb50e4d30!	117210	12	724100800787073,00	12	2024-09-12 12:48:33	2024-09-12 12:48:52	SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907	-23614679

SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907	-23614679
SAO PAULO - TV GLOBO BERRINI	SAO PAULO - VILA GERTRUDES	-46698411	-23614925
SAO PAULO - VILA GERTRUDES	SAO PAULO - NOVO JARDIM EDITE	-46694638	-23621200
SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907	-23614679

Fonte: Elaborado pelo autor.

4.2.2 Implementação da Heurística Temporal (Base Origem-Destino)

O código, incluindo documentação detalhada e exemplos, está disponível no repositório público:³

De forma geral, o algoritmo é utilizado para automatizar a criação de uma base de dados Origem-Destino (O-D) a partir de uma base de caminhos previamente gerada.

³https://github.com/edermau/base_od

Os dados são processados, os locais de origem e destino são identificados com base em heurísticas temporais, e as informações são organizadas em um formato apropriado para análises de mobilidade. Ressalta-se que a execução deste algoritmo deve ocorrer após a criação da base de caminhos, realizada pelo algoritmo descrito na seção 4.2.1

Os detalhes dos campos que serão apresentados nas operações abaixo podem ser consultados na Tabela 10.

Tabela 10 – Tabela de Origem Destino

msisdn	origem_site	origem_nome_site	origem_long	origem_lat	destino_site
Identificador único	Identificador do site de origem	Nome do site de origem	Longitude do site de origem	Latitude do site de origem	Identificador do site de destino

destino_nome_site	destino_long	destino_lat	dt_start_time	next_dt_start_time
Nome do site de destino	Longitude do site de destino	Latitude do site de destino	Timestamp do início do evento	Timestamp do próximo evento

Fonte: Elaborado pelo autor.

Etapas Principais do Algoritmo:

- Carregamento de Dados:

Os dados da base de caminhos são carregados a partir de um arquivo XLSX, permitindo a manipulação e o processamento no formato de *DataFrame* do pandas.

Um *DataFrame* segundo Team (2024b) é uma estrutura de dados bidimensional usada para armazenar e manipular dados tabulares, como em um arquivo XLSX ou uma tabela de banco de dados. Ele organiza os dados em linhas e colunas, facilitando operações como filtragem, agrupamento, ordenação e transformação.

Pandas segundo Team (2024a) é uma biblioteca do Python (FOUNDATION, 2024) amplamente utilizada para análise e manipulação de dados. Ela fornece ferramentas poderosas e flexíveis para trabalhar com estruturas como *DataFrames* e *Series* (dados unidimensionais, semelhantes a uma coluna). Com o Pandas, é possível realizar operações complexas de maneira simples e eficiente, tornando-o essencial para tarefas de ciência de dados e análise exploratória.

- Validação da Estrutura dos Dados: Uma validação inicial é realizada para garantir

que todas as colunas necessárias estejam presentes na base de caminhos.

- Garantia de Tipos de Dados: As colunas de tempo *dt_start_time* e *next_dt_start_time* são convertidas para o tipo *datetime* para permitir operações precisas de cálculo de tempo. Isso corrige possíveis inconsistências no formato de entrada.
- Cálculo da Diferença de Tempo: A diferença de tempo entre o início de um evento (*dt_start_time*) e o início do próximo evento (*next_dt_start_time*) é recalculada como um *Timedelta*. Esse valor é usado para aplicar critérios de filtragem baseados no tempo de permanência.
- Filtragem Temporal:

Apenas os registros com diferença de tempo superior ao tempo de permanência (parâmetro a ser definido, neste caso 5 minutos) são mantidos. Isso elimina eventos com conexão de curta duração, considerados irrelevantes para a análise de mobilidade.

- Criação da Base Origem-Destino:

Uma nova base é construída com as seguintes informações:

Origem: Informações da ERB de origem (*ds_site*, *nome_site*, *long*, *lat*).

Destino: Informações da ERB de destino (*next_ds_site*, *destino_long*, *destino_lat*).

Tempo: Colunas temporais, incluindo o tempo de início do evento e a diferença de tempo calculada.

- Tratamento de Dados Faltantes: Registros que possuem valores nulos em campos críticos (como coordenadas de destino) são removidos, garantindo a integridade dos dados.
- Exportação dos Resultados A base O-D final é exportada para um novo arquivo XLSX, facilitando análises posteriores. Uma mensagem de confirmação informa o usuário sobre o local de salvamento do arquivo.
- Código e Dependências:

O algoritmo utiliza a biblioteca pandas para manipulação de dados estruturados e openpyxl para leitura e gravação de arquivos XLSX.

- Detalhes dos Componentes:

Validação de Colunas: Verifica se todas as colunas essenciais estão presentes na base de dados. Isso evita erros em etapas posteriores.

- Filtragem Temporal: Mantém apenas os registros com tempo de permanência superior ao parâmetro definido, aplicando uma heurística temporal para identificar pontos relevantes no trajeto do usuário.
- Criação da Base O-D: Usa o método “shift()” para obter informações da ERB de destino e as associa ao registro atual da ERB de origem estruturando os dados em pares O-D, em Python esse método shift () é aplicado em um DataFrame do Pandas usado para deslocar os valores de uma coluna (ou linha) para frente ou para trás).

Exportação: Salva os resultados processados em um novo arquivo XLSX para facilitar a integração com outras ferramentas de análise.

Este algoritmo fornece uma ferramenta robusta para o estudo de padrões de mobilidade em dados de CDR, permitindo a construção de bases O-D detalhadas e protegendo a privacidade dos usuários por meio de anonimização prévia (realizada no algoritmo de base de caminhos).

5 EXPERIMENTO

Os dados de CDR utilizados para a construção da base de caminhos foram disponibilizados mediante autorização de uma empresa de telefonia. Essa contribuição enriqueceu o experimento, dada a alta qualidade dos CDRs empregados no estudo.

5.1 Ambientes de operação e teste.

5.1.1 Ambiente operacional de análise

Esta dissertação utilizou o ambiente de uma operadora de telefonia, como base de teste e validação do modelo apresentado.

O experimento foi realizado com a utilização de dados tratados de CDR, pré-selecionados, de um usuário específico durante 4 meses. Inicialmente, o conjunto de dados, normalizado no formato XLSX, foi composto por 16.384 registros.

A fim de não interferir no ambiente operacional é criado um *sandbox*, para processar os dados para os experimentos, o que está alinhado com o modelo apresentado na Figura 13.

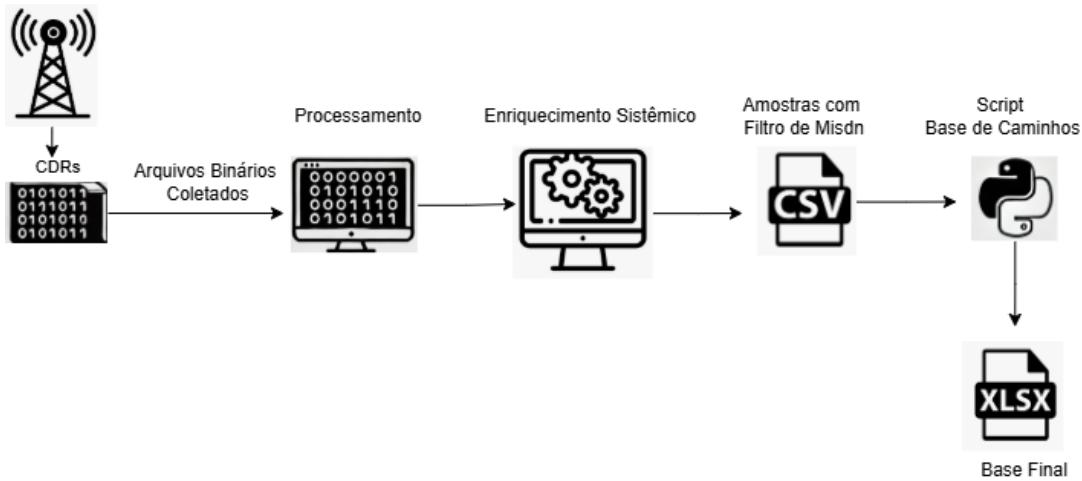
O *sandbox* de caminhos incorporou a infraestrutura e fluxo de processamento conforme a Figura 17, esse fluxo de processamento inicia-se com o recebimento de arquivos binários de CDR, que são coletados e processados. Em seguida, esses arquivos são convertidos para o formato CSV e enriquecidos com outras informações oriundas de ambientes sistêmicos. Após essa etapa, é realizado um filtro dessas amostras, selecionando um MSISDN específico para análise detalhada. Os dados filtrados são submetidos a um algoritmo de base de caminhos, que processa as informações e gera uma base final consolidada com todos os dados relevantes para a análise.

Antes de serem normalizados no Excel, os dados passaram por um fluxo de tratamento iniciado com 9.000 arquivos binários, cada um com tamanho médio de 56 kB. Esses arquivos brutos foram processados e enriquecidos adicionando dados como latitude e longitude. Após aplicar os filtros, os registros foram consolidados em um único arquivo CSV com um único MSISDN. A Tabela 11 apresenta o resumo do ambiente *sandbox* criado para o processamento.

Esses processamentos envolveram a aplicação do algoritmo⁴, gerando um conjunto de dados que foi reduzido para 2.139 registros. Posteriormente, foi aplicado o

⁴ <https://github.com/edermau/basedecaminhos>

Figura 17 – Fluxo de Coleta e Processamento dos CDR



Fonte: Elaborado pelo autor.

algoritmo⁵, utilizando a heurística temporal de 5 minutos(valor de teste), o que resultou em 1.972 registros correspondentes a quatro meses de informações associadas ao MSISDN.

Tabela 11 – Resumo do Ambiente de Sandbox Criado para o Processamento dos Dados

Parâmetro	Descrição
Servidor Utilizado	Chip Apple M1 Max, 10 Cores, 32 GB RAM
Sistema Operacional	Appe MacOS 14.7
Versão do Python	Python 3.12
Tempo Total de Processamento	2 segundos
Quantidade de Arquivos Binários Iniciais	9.000
Tamanho Médio dos Arquivos Binários	56 kB
Arquivos Gerados Após Enriquecimento	12.000 arquivos CSV
Quantidade Total de Registros Gerados	196.608.000 registros
Quantidade de Registros Após Consolidação	14.000 registros
Quantidade de ERBs Duplicados	1.904 registros
msisdn Consolidado	Dados de um único msisdn

5.2 Execução do Experimento

A proposição desse caso de uso utiliza dados móveis por usuário identificado pelo número de telefone em diferentes locais e horários. Os dados incluem informações sobre o volume de dados consumidos, a localização geográfica dos eventos (la-

⁵ https://github.com/edermau/base_od

titude e longitude), e o momento específico desses eventos.

A Figura 18, apresenta uma amostra de CDR em seu estado bruto, utilizada como ponto de partida para a elaboração da base de caminhos e a identificação do usuário esta devidamente anonimizada.

Figura 18 – Amostra de CDR Bruto

```

msisdn|mbo|cgl|dia|mes|hora|dt_start_time|nome_site|sig_estado|ibge_site|ds_mnco_site|lat|long
XXXXXXXXXXXXX|1511999|72406128600380|30|07|17|20240730171928|JARDIM GUANABARA|GO|5208707|GOI|CNIA|-16.622119|-49.223717
XXXXXXXXXXXXX|8|72406128600380|30|07|17|20240730171628|JARDIM GUANABARA|GO|5208707|GOI|CNIA|-16.622119|-49.223717
XXXXXXXXXXXXX|8|72406128600380|30|07|17|20240730171916|JARDIM GUANABARA|GO|5208707|GOI|CNIA|-16.622119|-49.223717
XXXXXXXXXXXXX|232|72406128600380|30|07|18|20240730081043|JARDIM GUANABARA|GO|5208707|GOI|CNIA|-16.622119|-49.223717
XXXXXXXXXXXXX|19643899|72406128600380|30|07|18|20240730083638|JARDIM GUANABARA|GO|5208707|GOI|CNIA|-16.622119|-49.223717
XXXXXXXXXXXXX|11682|7240612860018621|30|07|08|20240730080985|SETOR DA VAI|GO|5208707|GOI|CNIA|-16.638783|-49.229757
XXXXXXXXXXXXX|8|7240612860018621|30|07|08|20240730080985|SETOR DA VAI|GO|5208707|GOI|CNIA|-16.638783|-49.229757
XXXXXXXXXXXXX|232768|724061286013693|30|07|08|20240730089433|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|3389|724061286013693|30|07|08|20240730089433|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|18188413|724061286013693|30|07|08|20240730085850|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|3380|724061286013693|30|07|11|20240730114033|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|18689592|724061286013693|30|07|15|20240730154151|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|12176|724061286013693|30|07|18|20240730085842|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|3388|724061286013693|30|07|18|20240730150855|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|3236239|724061286013693|30|07|14|202407301845101|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|12228|724061286013693|30|07|15|20240730186255|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|5242923|724061286013693|30|07|15|20240730150865|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|6586696|724061286013693|30|07|18|20240730180580|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|7401871|724061286013693|30|07|18|20240730089580|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|8|724061286013693|30|07|18|20240730140932|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|232639|724061286013693|30|07|14|202407301845101|CAMPUS II UFG|GO|5208707|GOI|CNIA|-16.602211|-49.263061
XXXXXXXXXXXXX|1312|7240612860049011|30|07|12|20240730125127|URIAS MAGALHÉS|EES|GO|5208707|GOI|CNIA|-16.649811|-49.272711
XXXXXXXXXXXXX|425447|7240612860049021|30|07|12|20240730124653|URIAS MAGALHÉS|EES|GO|5208707|GOI|CNIA|-16.649811|-49.272711
XXXXXXXXXXXXX|6|7240612860049051|30|07|12|20240730125036|URIAS MAGALHÉS|EES|GO|5208707|GOI|CNIA|-16.649811|-49.272711
XXXXXXXXXXXXX|2388|7240612860049051|30|07|12|20240730125056|URIAS MAGALHÉS|EES|GO|5208707|GOI|CNIA|-16.649811|-49.272711
XXXXXXXXXXXXX|1465659|7240612860068601|30|07|17|20240730170845|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|7240612860068601|30|07|17|20240730124023|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|495528|7240612860068601|30|07|17|20240730144522|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|7240612860068601|30|07|17|20240730125129|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|7240612860068601|30|07|17|20240730144522|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|7240612860068601|30|07|18|20240730080582|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|107920|72406128600686021|30|07|18|20240730085747|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|72406128600686023|30|07|18|20240730084967|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|72406128600686023|30|07|18|20240730144846|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|103880|72406128600686023|30|07|18|20240730144846|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|72406128600686023|30|07|18|20240730144949|GOI|CNIA|2|GO|5208707|GOI|CNIA|-16.619139|-49.251444
XXXXXXXXXXXXX|8|724061286017016723|30|07|18|20240730081952|GOIANIA SETOR DE ABASTECIMENTO|GO|5208707|GOI|CNIA|-16.624439|-49.206500
XXXXXXXXXXXXX|192|724061286017016723|30|07|18|20240730081952|GOIANIA SETOR DE ABASTECIMENTO|GO|5208707|GOI|CNIA|-16.624439|-49.206500
XXXXXXXXXXXXX|2238704|7240612860170167503|30|07|18|202407300808957|GOIANIA SETOR DE ABASTECIMENTO|GO|5208707|GOI|CNIA|-16.624439|-49.206500
XXXXXXXXXXXXX|389484|7240612860170227921|30|07|18|202407300805259|SEFAZ DOI|GO|5208707|GOI|CNIA|-16.652738|-49.243970
XXXXXXXXXXXXX|8|724061282277021|30|07|12|20240730124712|SEFAZ DOI|GO|5208707|GOI|CNIA|-16.652738|-49.243970
XXXXXXXXXXXXX|8|724061282277021|30|07|12|20240730151487|GOIANIA PROGRESSO|GO|5208707|GOI|CNIA|-16.645658|-49.294688
XXXXXXXXXXXXX|8|724061282277021|30|07|12|20240730151487|GOIANIA PROGRESSO|GO|5208707|GOI|CNIA|-16.645658|-49.294688
XXXXXXXXXXXXX|8|72406128228253|30|07|17|20240730171543|BERNARDO SAAYO|GO|5208707|GOI|CNIA|-16.605939|-49.201531
XXXXXXXXXXXXX|8|72406128228253|30|07|17|20240730171543|BERNARDO SAAYO|GO|5208707|GOI|CNIA|-16.605939|-49.201531
XXXXXXXXXXXXX|3252|7240612824053001|30|07|17|20240730171904|GOIANIA SETOR SANTA GENOVEZA|GO|5208707|GOI|CNIA|-16.638333|-49.235472
XXXXXXXXXXXXX|395587|7240612824060001|30|07|17|20240730171613|GOIANIA DOIS BOULEVARD|GO|5208707|GOI|CNIA|-16.629500|-49.245500
XXXXXXXXXXXXX|882491|7240612824060051|30|07|17|20240730124944|GOIANIA DOIS BOULEVARD|GO|5208707|GOI|CNIA|-16.629500|-49.245500
XXXXXXXXXXXXX|7943814|72406128463011|30|07|16|20240730162000|MANSOES GOIANIAS|GO|5208707|GOI|CNIA|-16.623444|-49.282722

```

Fonte: Elaborado pelo autor

Este conjunto de dados é originalmente usado para analisar consumo de dados, tarifação ou a cobertura e qualidade de serviços de rede em diferentes áreas. No nosso caso, este conjunto será utilizado como uma nova linha de aplicação destes dados para outros fins, aplicados em serviços para gestão de cidades.

Para validar todo o processo foram realizados experimentos com diferentes modos de transporte, utilizando os parâmetros de utilização de heurísticas espaciais de caminho por FRBs e heurística temporal com tempo de permanência de 5 minutos

5.2.1 Experimento com Carro de Aplicativo de Transporte

A validação das bases de caminhos e de origem-destino foi realizada utilizando como referência um trajeto efetuado no dia 30 de julho de 2024, durante o período noturno. Nesse deslocamento, o usuário utilizou um aplicativo de transporte para percorrer 7,73 quilômetros. Os registros de CDR capturados demonstraram tanto o caminho percorrido quanto os pontos de origem e destino do usuário, que coincidiram

perfeitamente com o trajeto registrado pelo aplicativo.

A Tabela 12 ilustra a base de caminhos derivada de CDR, que foi devidamente normalizada e teve seus dados sensíveis anonimizados. Após este processo de tratamento, possibilita-se utilizar as informações para a elaboração de gráficos e gerar percepções.

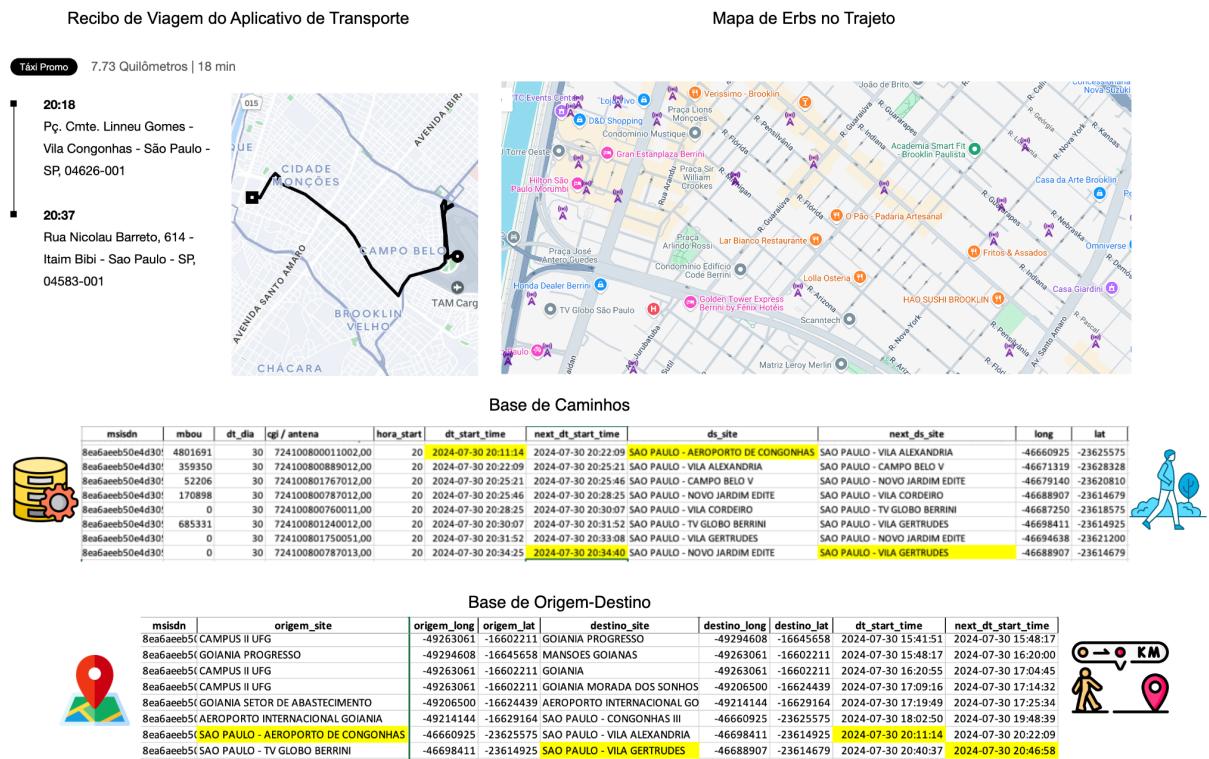
Tabela 12 – Amostra da Base de Caminhos.

msisdn	mbou	dt_dia	cgi	hora_start	dt_start_time	next_dt_start_time	time_diff	ds_site	next_ds_site	nome_site	long	lat
8ea6aeeb5c	3376	1	724100800787003,00	0	2024-07-01 00:05:08	2024-07-01 00:18:03		0 SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907 -23614679		
8ea6aeeb5c	128392	1	724100801240022,00	0	2024-07-01 00:18:03	2024-07-01 00:49:19		0 SAO PAULO - TV GLOBO BERRINI	SAO PAULO - NOVO JARDIM EDITE	-46698411 -23614925		
8ea6aeeb5c	0	1	724100800787013,00	0	2024-07-01 00:49:19	2024-07-01 00:50:27		0 SAO PAULO - NOVO JARDIM EDITE	SAO PAULO - TV GLOBO BERRINI	-46688907 -23614679		
8ea6aeeb5c	18178	1	724100801240022,00	0	2024-07-01 00:50:27	2024-07-01 01:20:12		0 SAO PAULO - TV GLOBO BERRINI	SAO PAULO - NOVO JARDIM EDITE	-46698411 -23614925		

Fonte: Elaborado pelo autor.

Na base de caminhos conforme Figura 19, é possível identificar o ponto de partida no “São Paulo — Aeroporto de Congonhas” às 20:11, percorrendo outras ERBs em diferentes bairros, e a sequência dos demais pontos até o horário final de 20:37 em “São Paulo — Vila Gertrudes”. Da mesma forma, a base de origem-destino reflete a origem do trajeto no Aeroporto de Congonhas e o destino na Vila Gertrudes, evidenciando a consistência dos dados capturados.

Figura 19 – Experimento com Dados de Deslocamento via Automóvel.



Fonte: Elaborado pelo autor.

Nesta Figura 19, no canto superior esquerdo, está o comprovante de uma corrida de aplicativo, enquanto no canto superior direito há um mapa de ERBs da região do

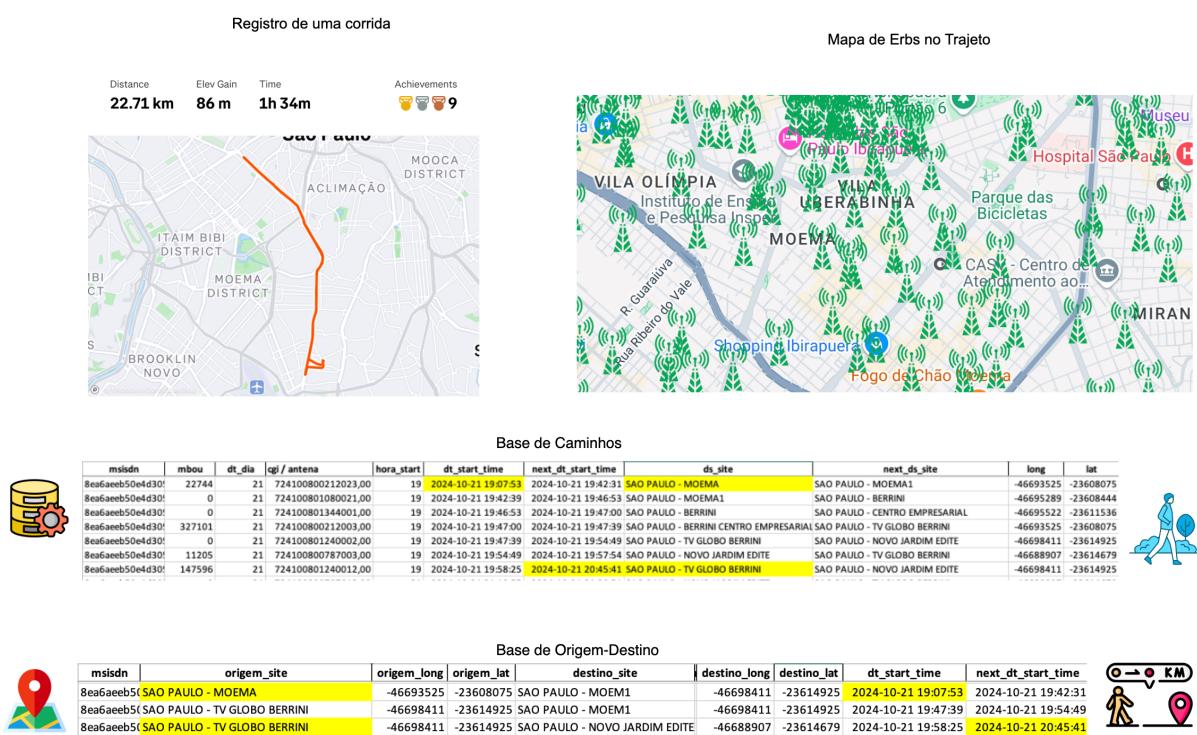
experimento. Na parte inferior da mesma figura, encontram-se os respectivos mapas de caminho e de origem-destino, com dados destacados em amarelo que correspondem às evidências apresentadas na parte superior da figura.

Esses dados reforçam a confiabilidade das bases utilizadas, que se mostraram capazes de mapear deslocamentos urbanos, alinhando-se ao trajeto real efetuado pelo usuário. Isso demonstra o potencial dessas bases para análises de mobilidade e planejamento urbano.

5.2.2 Experimento com Bicicleta

Uma segunda análise amostral conforme Figura 20 foi conduzida para validar as bases de caminhos e de origem-destino, tendo como referência um trajeto realizado no dia 21 de outubro de 2024, durante o período da noite, de bicicleta.

Figura 20 – Experimento com Dados de Deslocamento via Bicicleta.



Fonte: Elaborado pelo autor.

Na base de caminhos, destacam-se os pontos de origem e destino evidenciados pelos registros. O ponto inicial foi identificado em “São Paulo — Moema”, às 19:07:53, com a sequência de deslocamentos registrada até “São Paulo — Jardim Edite”, às 20:45:41. Esses dados capturaram com precisão as transições entre diferentes localidades, validando a rastreabilidade do trajeto.

Na base de origem-destino, observa-se uma correspondência consistente entre os pontos de origem (“São Paulo — Moema”) e destino (“São Paulo — Novo Jardim

Edite”), com informações complementares de longitude e latitude que reforçam a exatidão geográfica dos registros.

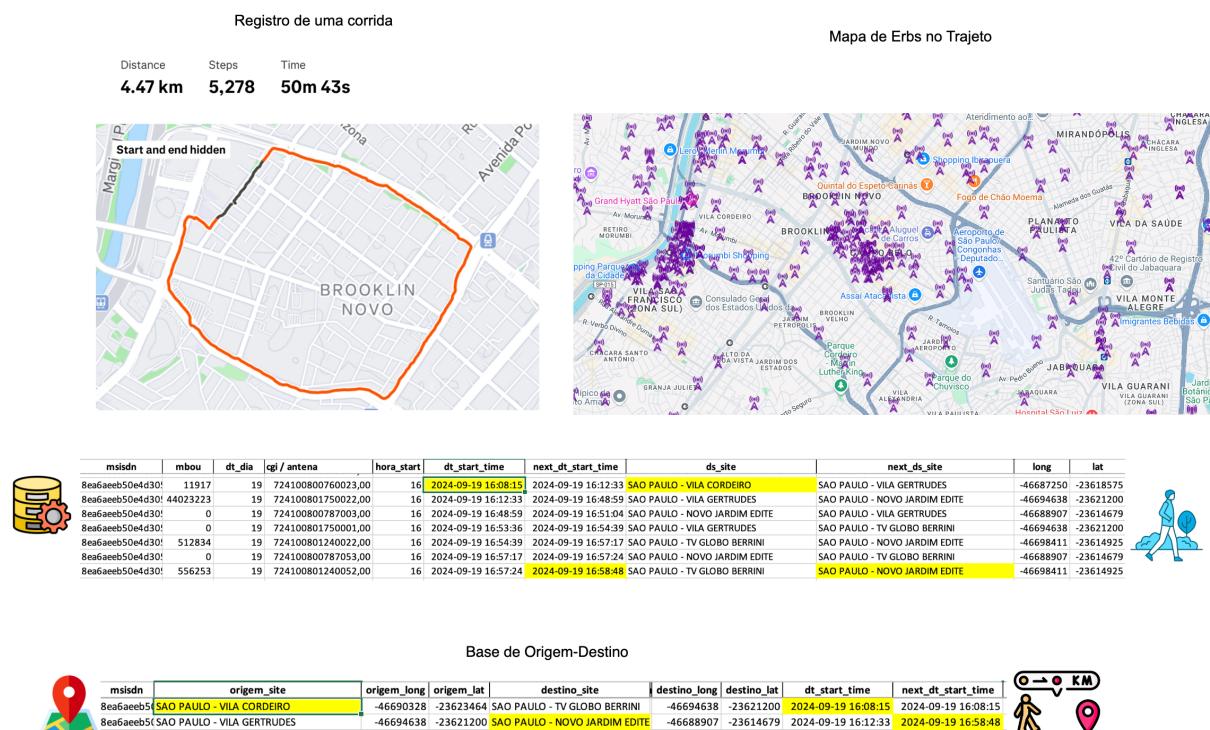
5.2.3 Experimento com Mobilidade a Pé

Uma terceira análise amostral foi conduzida para validar as bases de caminhos e de origem-destino, conforme ilustrado na Figura 21, utilizando como referência um trajeto realizado no dia 19 de setembro de 2024, durante o período da tarde. Nesse deslocamento foi percorrido 4,47 quilômetros a pé.

Na base de caminhos, destacam-se os pontos de origem e destino evidenciados pelos registros. O ponto inicial foi identificado em “São Paulo — Vila Cordeiro”, às 16:08:15, com a sequência de deslocamentos registrada até “São Paulo — Novo Jardim Edite”, às 16:58:48. Esses dados capturam com precisão as transições entre diferentes localidades, validando a rastreabilidade do trajeto.

Na base de origem-destino, observa-se uma correspondência consistente entre os pontos de origem (“São Paulo — Vila Cordeiro”) e destino (“São Paulo — Novo Jardim Edite”), com informações complementares de longitude e latitude que reforçam a exatidão geográfica dos registros.

Figura 21 – Experimento com Dados de Deslocamento em uma Corrida.



Fonte: Elaborado pelo autor.

A análise do tempo de deslocamento revelou os seguintes resultados:

- Total de registros: 1.971.

- Tempo médio de deslocamento: 27,54 minutos.

5.3 Disponibilidade dos Dados

Os dados adicionais coletados durante o experimento estão disponíveis em um repositório público e podem ser acessados no seguinte endereço: ⁶.

Esse repositório contém informações detalhadas sobre os trajetos registrados, permitindo a reprodução e análise dos resultados apresentados neste estudo.

Na Tabela 13 , estão registrados outros deslocamentos anotados durante o experimento, que podem ser consultados no repositório. Esses registros incluem caminhadas a pé, deslocamentos de carro por aplicativo e de bicicleta.

Tabela 13 – Registros de Deslocamentos do Experimento

Modalidade	Início	Término
Deslocamento via Automóvel	2024-07-02 07:44:55	2024-07-02 08:33:08
	2024-07-15 12:13:48	2024-07-15 12:40:49
	2024-07-17 14:37:19	2024-07-17 15:36:56
	2024-08-12 12:48:53	2024-08-12 13:48:38
Deslocamento via Bicicleta	2024-07-04 15:43:31	2024-07-04 16:01:17
	2024-07-16 11:46:07	2024-07-16 12:36:59
	2024-07-23 18:35:39	2024-07-23 19:31:31
	2024-08-14 13:54:26	2024-08-14 14:24:54
Deslocamento a Pé	2024-07-04 16:27:19	2024-07-04 16:35:31
	2024-07-21 10:05:41	2024-07-21 10:38:32
	2024-07-22 19:41:35	2024-07-22 20:38:59
	2024-09-01 19:55:10	2024-09-01 20:25:03

5.4 Discussão dos Resultados

O estudo conduzido por Jakovljevic *et al.* (2021) concentra-se na avaliação da precisão das matrizes de origem-destino (OD) construídas com base em dados de telecomunicações. A pesquisa adota uma abordagem meticolosa para entender os erros associados à localização de ERBs (Estações Rádio Base) e propõe métodos para corrigir esses desvios, visando aprimorar a representatividade espacial das matrizes OD.

Considerando estes aspectos, foram realizados experimentos controlados, para gerar evidências do trajeto percorrido por um usuário, confrontando com os resultados gerados no processamento nos *sandboxes* com a evidência apresentada. Nestes casos, todos os experimentos apresentaram na base de caminhos os locais (ERB) coerentes com a jornada do usuário, seja ela percorrida com um veículo de aplicativo de transporte, motocicleta ou a pé durante uma sessão de corrida do usuário.

⁶<https://edermau.github.io/experimento/>

Ainda na abordagem de, Jakovljevic *et al.* (2021) é relevante adotar intervalos temporais consistentes na análise de dados de mobilidade, principalmente para cenários onde a precisão espacial e a representatividade dos dados são críticas, como no planejamento de transportes e na gestão de cidades inteligentes. Embora o estudo não descreva explicitamente um critério, mencionando que para cada aplicação pode existir um valor mais adequado ao comportamento de mobilidade. Além disso, Mamei *et al.* (2019) sugere que, em situações de baixa granularidade de dados, é razoável desconsiderar ERBs que não contribuem significativamente para a identificação de trajetos principais, especialmente quando o objetivo é modelar rotas de alta prioridade, como deslocamentos casa-trabalho. Dessa forma, a exclusão de certas ERBs é justificada como uma simplificação necessária para tornar a análise viável e robusta dentro das limitações dos dados de CDR.

Nesta dissertação adotamos a heurística, que um ponto só será considerado um DESTINO, se ele permanecer em comunicação com a ERB por mais que 5 minutos. Assim, na geração da BASE O-D perceberíamos a variação em qualquer um dos modos de transporte experimentados (carro, motocicleta ou a pé), capturando os padrões regulares de deslocamento, garantindo maior representatividade das trajetórias urbanas e alinhando a análise aos princípios destacados em Jakovljevic *et al.* (2021). Considerando esta heurística, os resultados dos experimentos desta dissertação denotarão a correta marcação da Origem-Destino em todas as ERBs que o usuário permaneceu por mais que 5 minutos, excluindo aquelas ERBs que tiveram comunicação durante o trajeto entre a ORIGEM e o DESTINO, e não atenderam a esta heurística.

No entanto, vale salientar, este valor é um PARÂMETRO que deve ser ajustado para servir apropriadamente para um serviço inteligente que seria executado na etapa 4 do método proposto (Figura13), sendo que esta dissertação não incluiu atividades da etapa 4 (fora de escopo), sendo o tempo de permanência adotado nesta dissertação, apenas um valor empírico a ser adotado para confrontar os resultados obtidos.

5.5 Relações e Associação dos Resultados com as Questões de Pesquisa

Após a realização da revisão da literatura e das análises descritas no experimento, foram obtidos resultados significativos que permitem responder de maneira fundamentada às perguntas de pesquisa.

P1. Quais são os principais serviços de cidades inteligentes que poderiam utilizar dados de CDR? Na seção 3.1 foi apresentada uma revisão sistemática que respondia esta questão, e que culminou também na publicação (BARBOSA *et al.*, 2023), que foi apresentada e publicada em 2023 nos anais do *IEEE International Smart Cities Conference*. Sendo que esta consolidação representa as oportunidades do uso de “big

data” baseados em CDR para fomentar Cidades Inteligentes. O Anexo A apresenta um extrato desta publicação.

Ainda foi descrito o estudo de caso de serviço inteligente aplicado no Brasil, que utilizava este princípio e cujo resultados foram incorporados na publicação Santos, Barbosa *et al.* (2024), onde este autor participou da elaboração do trabalho. O anexo B apresenta um estrato desta publicação.

P2. Como utilizar dados de CDR para criar uma base de caminhos e de Origem-Destino aplicável a serviços de cidades inteligentes?

Como parte da pesquisa para responder esta questão, algumas referências foram fundamentais para esta contextualização, como, por exemplo, o caso de Luo *et al.* (2020), que propõem o uso de registros detalhados de chamadas (CDRs) combinados com dados de redes viárias da cidade para identificar localizações de residência e trabalho dos usuários. O método inclui etapas de pré-processamento de dados, fusão de dados e detecção de localização, e demonstrou resultados promissores em termos de precisão. Esse método mostra como os dados de CDR, integrados com informações de infraestrutura urbana, podem ser usados para rastrear trajetórias e criar uma base de dados de mobilidade para uso em serviços urbanos. Como se pretende realizar um modelo amplo para vários serviços, adotou-se como heurística espacial elementos da infraestrutura celular, no caso a ERB, como elemento que delimita áreas para a definição da região que faz parte do caminho percorrido, e estando contido na Base de Caminhos. No entanto, a heurística espacial a ser adotada pode ser também por BAIRRO ou CIDADE, dependendo da necessidade da criação de um *SANDBOX DE SERVIÇOS* (Etapa 4 do modelo apresentado na Figura13), que não está no escopo desta dissertação.

Jakovljevic *et al.* (2021) descrevem uma metodologia para avaliar a precisão das matrizes OD baseadas em grandes volumes de dados de telecomunicações. Esse estudo compara dados obtidos de CDR com metodologias tradicionais de coleta de dados, como censos e pesquisas de campo, demonstrando que as matrizes OD derivadas dos dados de telecomunicações podem ser eficazes para estimar a demanda de viagens e auxiliar no planejamento urbano. Essa metodologia reforça a aplicabilidade dos dados de CDR para criar uma base de caminhos e OD, especialmente ao validar a correlação dos dados de CDR com dados tradicionais. Nesta dissertação incluiu-se um parâmetro, TEMPO DE PERMANÊNCIA, como base de uma heurística temporal para definir destinos na base O-D, onde uma região era considerada DESTINO, se existisse comunicação do celular com a ERB, que evidenciasse que o usuário estava na região por mais tempo que o parâmetro definido.

Em comparação, a abordagem de Luo *et al.* (2020) foca mais na criação de trajetórias e identificação de locais específicos (residência e trabalho), enquanto o estudo de Jakovljevic *et al.* (2021) valida a precisão das matrizes OD usando CDR em relação a métodos tradicionais, o que é essencial para justificar o uso de CDR em contextos urbanos. Assim, a melhor forma de responder à pergunta é combinar o método de criação de trajetórias similares ao proposto por Luo *et al.* (2020) com a validação metodológica proposta por Jakovljevic *et al.* (2021) utilizando CDR para capturar dados de mobilidade e validar sua precisão em relação a métodos tradicionais, garantindo sua aplicabilidade para o desenvolvimento de serviços inteligentes urbanos. Nesta dissertação, realizaram-se experimentos com diferentes modais de transporte e validou-se com dados de outras fontes, para confrontar os resultados obtidos com os ocorridos, sendo que todos os experimentos e validações foram bem sucedidas.

Assim a proposição desta dissertação que inclui as etapas 1 a 3 do modelo proposto e apresentado na Figura 13 foi comprovada e propiciando bons mecanismos para proposições futuras de outros estudos que completem as demais etapas do modelo proposto, gerando serviços inteligentes para a gestão em cidades. Ainda, vários trabalhos de serviços para cidades inteligentes com CDRs, que foram apresentados na seção 3.1 poderiam se utilizar das base do modelo único proposto por esta dissertação, uma vez que cada um deles tinha um método próprio, muitas vezes não compatíveis entre si, dificultando a replicação em um ambiente real para múltiplos serviços.

Com relação aos serviços inteligentes que poderiam usufruir do modelo apresentado nesta dissertação, destaca-se exemplos e estudos que evidenciam o potencial desta abordagem, com é o caso de Liu *et al.* (2020), que exploram a estrutura comunitária urbana com base no fluxo populacional e na distribuição de serviços públicos. Esse estudo usa CDR para analisar as interações entre diferentes zonas geográficas, o que pode ajudar a entender a estrutura urbana e facilitar o planejamento de serviços de cidades inteligentes. A abordagem desse artigo enfatiza o uso de CDR para detectar comunidades urbanas e melhorar a gestão da distribuição de serviços, como transporte e infraestruturas de apoio. Outro exemplo é o trabalho de Luo *et al.* (2020), que utiliza o mesmo contexto para mapeamento de mobilidade urbana. Ambos os trabalhos exploram metodologias baseadas em CDR para entender a mobilidade urbana, mas com enfoques distintos e métodos diferentes.

Cabe salientar, que o uso de CDR's também traz limitações, que também foram apresentados por outros autores, por exemplo, Yankun Wang *et al.* (2019) e Luo *et al.* (2020) apresentam uma abordagem onde os dados de CDR são processados para identificar trajetórias individuais de mobilidade, onde a construção das trajetórias é realizada de maneira eficiente, mas os autores ressaltam que limitações, como a baixa

granularidade espacial dos dados de CDR, podem impactar a precisão em áreas que não são densamente urbanizadas e portanto com uma baixa cobertura de ERBs.

5.6 Contribuição Futura

Ainda como contribuição para trabalhos futuros, espera-se que seja disponibilizada uma amostra de usuários de uma empresa de telefonia, coletada ao longo de 1 mês, contendo bases de caminhos e OD, com os dados de MSISDN devidamente anonimizados. Para viabilizar essa iniciativa, foi iniciado um processo interno de autorização na empresa de telefonia, garantindo a conformidade legal e as aprovações necessárias para a disponibilização dos dados. A amostra será disponibilizada em um repositório público, e há a intenção de publicar um artigo para fomentar e apoiar novos estudos nesse âmbito, desde que as autorizações sejam concedidas.

6 CONCLUSÃO

A infraestrutura de telefonia celular produz um grande volume de CDRs, que são comumente utilizados como base para tarifação e análises relacionadas à infraestrutura. No entanto, esses registros podem ser processados para ampliar seu uso para além das aplicações tradicionais no setor de telecomunicações, viabilizando a análise de dados de CDR com foco no desenvolvimento de novos serviços e soluções inteligentes para cidades.

A presente dissertação evidenciou que a aplicação de heurísticas espaçotemporais permite a identificação de padrões individuais de deslocamento, possibilitando análises dos fluxos de movimento urbano. Esses fluxos podem servir como fontes essenciais para ampliar a compreensão do conglomerado urbano, tornando-se, assim, um instrumento para o desenvolvimento de cidades inteligentes. Dessa forma, a análise espaçotemporal dos CDRs surge como uma abordagem eficaz para compreender os padrões de mobilidade urbana e para a construção de bases de dados de deslocamento em meio digital.

Os resultados apresentados evidenciaram a relevância das heurísticas espaciais na segmentação de dados em diferentes níveis de granularidades, como ERBs, bairros e cidades. Conforme demonstrado na dissertação, essa abordagem possibilita a criação de bases de caminhos a partir das conexões registradas na infraestrutura celular. Adicionalmente, a aplicação de heurísticas temporais baseadas no tempo de permanência em uma determinada região possibilitou a inferência de pontos de Origem e Destino, oferecendo uma visão sobre os deslocamentos urbanos. Como resultado, foram geradas bases O-D, que geralmente são exploradas para estudos de identificação de rotas mais frequentes, áreas de alta concentração de deslocamentos, padrões de uso do espaço urbano, além de diversas outras aplicações no planejamento e na análise da mobilidade urbana.

Nessa dissertação foi apresentado um estudo de caso, sobre o uso de CRDs apoiando as decisões dos gestores públicos na condução da pandemia de COVID-19, sendo um motivador para a elaboração deste trabalho. Objetivando explorar novas possibilidades técnicas e científicas para o uso do CDR, esta dissertação investigou e abordou métodos e estratégias para expandir o uso, comprovando a viabilidade técnica e científica da aplicação dos resultados obtidos em empresas de telecomunicações, como uma possibilidade de novos serviços para a gestão de cidades.

Este trabalho deixa algumas contribuições relevantes, sejam elas técnicas ou

científicas, ou mesmo no contexto da promoção de estratégias para tornar as cidades mais inteligentes.

No âmbito das contribuições técnicas e científicas, a utilização prática de CDRs em um contexto amplo, como no estudo de caso voltado à geração de índices de isolamento social para o enfrentamento da COVID-19, revelou diversos desafios técnicos e científicos. Entre esses desafios, destaca-se a necessidade de anonimização para a preservação da privacidade dos dados, a agregação baseada na localização geográfica dos usuários e a representatividade da amostra populacional, além do elevado custo associado ao armazenamento e processamento dos grandes volumes de dados brutos dos CDRs para aplicações específicas. Apesar dessas limitações, a análise de CDRs pode representar um avanço significativo no campo da mobilidade urbana e das cidades inteligentes, especialmente devido à sua capacidade de utilizar grandes volumes de dados em tempo real.

A dissertação demonstrou que, ao utilizar CDR de dados e técnicas analíticas, é possível inferir trajetos detalhados e compreender as dinâmicas de deslocamento, oferecendo uma visão sobre a mobilidade urbana, com uma segregação dos dados em *sandbox*, com volume reduzido, mantendo as características essenciais para as análises de deslocamento. Possibilitando assim, um modelo de negócio que possa ser ampliado para outros serviços de cidades, promovendo inovação e novas oportunidades de negócios.

Adicionalmente, algumas contribuições secundárias foram identificadas durante o processo de coleta dos dados, por exemplo, uma falha identificada nos ambientes de monitoramento da empresa de telefonia que resultava em custos elevados de reprocessamento devido à ausência de métricas adequadas para a detecção de tais problemas. Essa descoberta levou à reestruturação do sistema de monitoramento, ajustando-o às boas práticas. Apesar da análise ter sido realizada com dados de um único CDR, uma amostra qualitativa foi gerada, desencadeando a criação de um projeto emergencial de âmbito nacional para implementar análises aleatórias nas unidades coletoras de CDR presentes na infraestrutura celular. Tal iniciativa contribuiu significativamente para a otimização das operações e redução de custos associados a falhas de coleta.

Ainda, em outro caso secundário, a análise detalhada dos trajetos urbanos revelou uma anomalia na funcionalidade de uma estação rádio base (ERB) localizada em uma torre alugada, que apresentava instabilidade constante em determinado trecho. A abertura de um chamado para a equipe de engenharia confirmou o mau funcionamento do equipamento, evidenciando o alto custo de validação para este tipo de análise pontual. A partir deste caso, foi estabelecida uma frente interna para a estudo de criação de um modelo de predição com variáveis internas das ERBs, visando demonstrar ao

nível executivo os benefícios financeiros e operacionais de evitar deslocamentos desnecessários. Essa iniciativa não apenas otimizou o uso de recursos, mas também promoveu melhorias na qualidade da rede e no atendimento aos usuários.

A análise extensiva dos CDRs destacou a falta de padronização entre fabricantes, conforme apontado pela literatura. Essa heterogeneidade foi reconhecida como um fator que dificulta o processamento eficiente, tanto em termos de tamanho quanto de compatibilidade com diferentes equipamentos. Assim, o estudo de vários artigos, que utilizavam a análise de CDRs, os quais foram base para implementar um modelo padrão e unificado, gera um modelo de referência para o processamento de grandes volumes de dados provenientes de CDRs, que podem impulsionar o desempenho de processamento em tempo real, com ambientes apartados em *sandbox*.

Este trabalho não apenas contribuiu para a resolução de problemas relevantes na empresa de telefonia, como também resultou na implementação de eficientes controles operacionais, além de abrir caminhos para pesquisas futuras no campo da análise e padronização de CDRs, ou mesmos exploração de serviços inteligentes como atividades de pesquisa futura que utilizem como base o modelo unificado.

No contexto de contribuições para Cidades Inteligentes, a aplicação prática das bases geradas, evidencia o potencial da abordagem utilizada nesta dissertação para transformar a gestão urbana. As bases de Origem-Destino, criadas com as heurísticas espaço-temporais, podem ser usadas diretamente no planejamento de transporte, na gestão de emergências e na alocação de recursos em cidades inteligentes. Esses dados são dinâmicos e permitem analisar o comportamento das cidades de forma mais flexível e com novas perspectivas, especialmente quando comparados aos métodos tradicionais, como as pesquisas de Origem-Destino feitas por órgãos governamentais, por exemplo, a pesquisa OD do metrô de São Paulo.

A integração de dados CDR para a construção de bases de caminhos e OD oferece novas possibilidades para o aprimoramento do planejamento urbano, para a otimização dos serviços públicos de transporte, alinhando-se aos objetivos de cidades mais inteligentes, sustentáveis e resilientes. A continuidade dessa pesquisa pode abrir novas oportunidades para ampliar o uso de CDR em diversos cenários urbanos, contribuindo para o desenvolvimento de cidades mais eficientes e responsivas às necessidades da população.

Embora os resultados tenham sido significativos, algumas limitações foram observadas, como foi o caso da exatidão dos dados de localização do usuário, que são afetadas pela variação na densidade das ERBs, que possuem uma cobertura limitada. Essa questão destaca a importância de pesquisas futuras para melhorar os métodos de

extrapolação e validação dos dados, utilizando fontes externas, para aumentar a exatidão do posicionamento dos fluxos de deslocamento, se houver este requisito como premissa dos estudos.

REFERÊNCIAS

- AMMAPA, Jirattikan *et al.* Using GIS-Based Spatial Analysis: Comparing Pattern of Urbanization and Transportation Networks. *In:* 2022 10th International Conference on Traffic and Logistic Engineering (ICTLE). [S. I.: s. n.], 2022. P. 17–21. DOI: 10.1109/ICTLE55577.2022.9902095.
- BARBOSA, Eder M. *et al.* Using Cellular Infrastructures Data to Foster the Transition Towards Smart Cities: a Systematic Mapping. *In:* 2023 IEEE International Smart Cities Conference (ISC2). [S. I.: s. n.], set. 2023. P. 1–7. DOI: 10.1109/ISC257844.2023.10293311.
- BERNARDI, ELY; ET AL. Brazilian scenarios for smart cities deployment from public policies perspectives. *In:* 2020 IEEE International Smart Cities Conference (ISC2). [S. I.: s. n.], 2020. P. 1–8. <https://doi.org/10.1109/ISC251055.2020.9239096>.
- BRASIL, CNN. **São Paulo faz parceria com operadoras de telefonia para monitorar quarentena.** [S. I.: s. n.], 2020. Disponível em: <https://www.cnnbrasil.com.br/nacional/sao-paulo-faz-parceria-com-operadoras-de-telefonia-para-monitorar-quarentena/>. Acesso em: 19 mar. 2024.
- CARRILLO, Maria Angelica *et al.* The use of mobile phones for the prevention and control of arboviral diseases: a scoping review. English. **BMC Public Health**, BioMed Central Ltd, v. 21, n. 1, 2021. ISSN 14712458. DOI: 10.1186/s12889-020-10126-4.
- CHIN, Kimberley *et al.* Inferring fine-grained transport modes from mobile phone cellular signaling data. English. **COMPUTERS ENVIRONMENT AND URBAN SYSTEMS**, ELSEVIER SCI LTD, THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND, v. 77, set. 2019. ISSN 0198-9715. DOI: 10.1016/j.comenvurbssys.2019.101348.
- CUMBANE, Silvino Pedro; GIDÓFALVI, Gyozo. Review of big data and processing frameworks for disaster response applications. English. **ISPRS International Journal of Geo-Information**, MDPI AG, v. 8, n. 9, 2019. ISSN 22209964. DOI: 10.3390/ijgi8090387.
- DEALMEIDA, Jonathan M. *et al.* Abnormal Behavior Detection Based on Traffic Pattern Categorization in Mobile Networks. English. **IEEE Transactions on Network and Service Management**, Institute of Electrical e Electronics Engineers Inc., v. 18, n. 4, p. 4213–4224, 2021. ISSN 19324537. DOI: 10.1109/TNSM.2021.3125019.
- ELBAKLY, Rizanne; YOUSSEF, Moustafa. Crescendo: An Infrastructure-free Ubiquitous Cellular Network-based Localization System. *In:* 2019 IEEE Wireless

Communications and Networking Conference (WCNC). [S. I.: s. n.], 2019. P. 1–6. DOI: 10.1109/WCNC.2019.8885420.

ESTADO DE SÃO PAULO, Governo do. **ADESÃO AO ISOLAMENTO SOCIAL EM SP**. [S. I.: s. n.], 2022. Disponível em: <https://www.saopaulo.sp.gov.br/coronavirus/isolamento/>. Acesso em: 22 mar. 2024.

FELIZARDO, Katia Romero *et al.* Defining Protocols of Systematic Literature Reviews in Software Engineering: A Survey. In: 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA). [S. I.: s. n.], 2017. P. 202–209. DOI: 10.1109/SEAA.2017.17.

FOUNDATION, Python Software. **Python - Official Website**. [S. I.: s. n.], 2024. Disponível em: <https://www.python.org/>. Acesso em: 11 dez. 2024.

GOLENIEWSKI, Lillian; JARRETT, Kitty Wilson. **Telecommunications Essentials: The Complete Global Source**. 2nd. Upper Saddle River, NJ: Addison-Wesley Professional, 2006. P. 514. Includes index. ISBN 0-321-42761-0.

HUAWEI. **Configuring Connection Data for the CDR Server**. Shenzhen, China, 2023. Disponível em: <https://support.huawei.com/enterprise/en/doc/EDOC1100112359/2ba508cb/configuring-connection-data-for-the-cdr-server>.

HUAWEI. **Hardware Installation and Maintenance Guide for NetEngine AR600, AR6100, AR6200, and AR6300 Series Routers**. Shenzhen, China, 2023. Disponível em: <https://support.huawei.com/enterprise/en/doc/EDOC1100316318>.

IBGE, Instituto Brasileiro de Geografia e Estatística. **Panorama do Censo 2022**. Rio de Janeiro: IBGE, 2022. Primeiros resultados. Disponível em: <https://censo2022.ibge.gov.br/panorama/indicadores.html>.

IETF. **RFC 2924 - Accounting Attributes and Record Formats**. [S. I.: s. n.], 2024. Disponível em: <https://datatracker.ietf.org/doc/rfc2924/>. Acesso em: 20 nov. 2024.

IETF. **RFC 5114 - Additional Diffie-Hellman Groups for Use with IETF Standards**. [S. I.: s. n.], 2024. Disponível em: <https://datatracker.ietf.org/doc/html/rfc5114>. Acesso em: 20 nov. 2024.

JAKOVLJEVIC, Marijan *et al.* Methodology for Assessment of Mobile Telecom Databased Origin-Destination Matrices Accuracy. English. In: M., Koricic *et al.* (Ed.). **2021 44th International Convention on Information, Communication and Electronic Technology, MIPRO 2021 - Proceedings**. [S. I.]: Institute of Electrical e Electronics Engineers Inc., 2021. P. 1081–1085. ISBN 978-953233101-1. DOI: 10.23919/MIPRO52101.2021.9596646.

KOZIK, Rafał *et al.* Proposition of innovative and scalable information system for call detail records analysis and visualisation. English. Edição: Herrero A. **Advances in Intelligent Systems and Computing**, Springer Science e Business Media Deutschland GmbH, 1267 AISC, p. 174–183, 2021. Cited by: 1; Conference name: 13th International Conference on Computational Intelligence in Security for Information Systems, CISIS 2020; Conference date: 16 September 2020 through 18 September 2020; Conference code: 244669. ISSN 21945357. DOI: 10.1007/978-3-030-57805-3_17.

LIU, Qinghe *et al.* Toward Identifying the Urban Community Structure from Population Flow and Public Services Distribution. English. In: X., Wu *et al.* (Ed.). **Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020**. [S. I.]: Institute of Electrical e Electronics Engineers Inc., 2020. P. 3133–3142. ISBN 978-172816251-5. DOI: 10.1109/BigData50022.2020.9378225.

LUO, Xusen *et al.* Research on Home and Work Locations Based on Mobile Phone Data. English. **Journal of Physics: Conference Series**, Institute of Physics Publishing, v. 1486, n. 5, 2020. ISSN 17426588. DOI: 10.1088/1742-6596/1486/5/052013.

MAHAJAN, Vishal; CANTELMO, Guido; ANTONIOU, Constantinos. Explaining demand patterns during COVID-19 using opportunistic data: a case study of the city of Munich. English. **European Transport Research Review**, Springer Science e Business Media Deutschland GmbH, v. 13, n. 1, 2021. ISSN 18670717. DOI: 10.1186/s12544-021-00485-3.

MAMEI, Marco *et al.* Evaluating origin–destination matrices obtained from CDR data. English. **Sensors (Switzerland)**, MDPI AG, v. 19, n. 20, 2019. ISSN 14248220. DOI: 10.3390/s19204470.

METROPOLITANO DE SÃO PAULO - METRÔ, Companhia do. **Metrô antecipa elaboração da maior pesquisa de mobilidade urbana para analisar os impactos da pandemia**. [S. I.: s. n.], 2024. Disponível em: <https://www.metro.sp.gov.br/2023/07/31/metro-antecipa-elaboracao-da-maior-pesquisa-de-mobilidade-urbana-para-analisar-os-impactos-da-pandemia/#:~:text=Com%20a%20metodologia%20sistematizada%20em,anos%20pelo%20Metr%C3%B4%20desde%201967..> Acesso em: 22 mar. 2024.

MICROSOFT. **Office Open XML File Formats - XLSX Specification**. [S. I.: s. n.], 2024. Disponível em: <https://docs.microsoft.com/en-us/office/open-xml/>. Acesso em: 11 dez. 2024.

MYEONG, SEUNGHWAN; KIM, YOUNHEE; AHN, MICHAEL J. Smart City Strategies—Technology Push or Culture Pull? A Case Study Exploration of Gimpo and Namyangju, South Korea. **Smart Cities**, Multidisciplinary Digital Publishing Institute, v. 4, n. 1, p. 41–53, 2021.

POPOVA, Yelena; POPOVS, Sergejs. Impact of Smart Economy on Smart Areas and Mediation Effect of National Economy. **SUSTAINABILITY**, v. 14, n. 5, 2022. DOI: 10.3390/su14052789.

RAHIMI-GOLKHANDAN, Armin; GARVIN, Michael J.; WANG, Qi. Assessing the Impact of Transportation Diversity on Postdisaster Intraurban Mobility. English. **Journal of Management in Engineering**, American Society of Civil Engineers (ASCE), v. 37, n. 1, 2021. ISSN 0742597X. DOI: 10.1061/(ASCE)ME.1943-5479.0000872.

RAHNEMA, M. Overview of the GSM system and protocol architecture. **IEEE Communications Magazine**, v. 31, n. 4, p. 92–100, 1993. DOI: 10.1109/35.210402.

SANTOS, Alessandro S.; BARBOSA, Eder *et al.* Smart resilience through IoT-enabled natural disaster management: A COVID-19 response in São Paulo state. **IET Smart Cities**, v. 6, n. 3, p. 211–224, 2024. DOI: <https://doi.org/10.1049/smcc2.12082>. eprint: <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/smcc2.12082>. Disponível em: <https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/smcc2.12082>.

SANTOS, Alessandro S.; TEIXEIRA, Igor C. *et al.* Challenges and Strategies for Information Systems in the Decision-Making Process to Face the COVID-19 Pandemic: The São Paulo Case. In: ROCHA, Álvaro *et al.* (Ed.). **Trends and Applications in Information Systems and Technologies**. Cham: Springer International Publishing, 2021. P. 630–640. ISBN 978-3-030-72660-7.

SCANNAVINO, Katia Romero Felizardo *et al.* **Revisão Sistemática da Literatura em Engenharia de Software: teoria e prática**. Rio de Janeiro, Brasil: Editora Ciência Moderna, 2017.

SHIH, Chia-Chun *et al.* Building a CDR analytics platform for real-time services. In: THE 16th Asia-Pacific Network Operations and Management Symposium. [S. I.: s. n.], 2014. P. 1–5. DOI: 10.1109/APNOMS.2014.6996538.

TEAM, Pandas Development. **Pandas - Python Data Analysis Library**. [S. I.: s. n.], 2024. Disponível em: <https://pandas.pydata.org/>. Acesso em: 11 dez. 2024.

TEAM, Pandas Development. **pandas.DataFrame - Pandas Documentation**. [S. I.: s. n.], 2024. Disponível em: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>. Acesso em: 11 dez. 2024.

VIDOVIC, Krešimir *et al.* Methodology for public transport mode detection using telecom big data sets: case study in Croatia. English. Edição: Petrovic M., Dovbischuk I. e Cunha A.L. **Transportation Research Procedia**, Elsevier B.V., v. 64, n. 100, p. 76–83, 2022. ISSN 23521457. DOI: 10.1016/j.trpro.2022.09.010.

WANG, Jinzhong *et al.* Urban Human Mobility: Data-Driven Modeling and Prediction. **SIGKDD Explor. NewsL.**, Association for Computing Machinery, New York, NY, USA, v. 21, n. 1, p. 1–19, 2019. ISSN 1931-0145. DOI: 10.1145/3331651.3331653.

Disponível em:

<https://doi-org.ez67.periodicos.capes.gov.br/10.1145/3331651.3331653>.

WANG, Yankun *et al.* A hybrid intelligent approach for constructing landslide displacement prediction intervals. **Applied Soft Computing**, v. 81, p. 105506, 2019. ISSN 1568-4946.

YEH, En-Hau; LIN, Phone; HUANG, Ming-Wey. ADPD: Anomaly Detection for Population Distribution in Geo-Space Using Mobile Networks Data. English. **IEEE Internet of Things Journal**, Institute of Electrical e Electronics Engineers Inc., v. 9, n. 22, p. 22774–22784, 2022. ISSN 23274662. DOI: 10.1109/JIOT.2022.3184072.

ZHANG, Desheng *et al.* Urban-Scale Human Mobility Modeling With Multi-Source Urban Network Data. **IEEE/ACM Transactions on Networking**, v. 26, n. 2, p. 671–684, 2018. ISSN 1558-2566. DOI: 10.1109/TNET.2018.2801598.

ANEXO A – ARTIGO DE MOBILIDADE IEEE

Using cellular infrastructures data to foster the transition towards smart cities: a systematic mapping

Eder M. Barbosa, Josias Lima, Alessandro Santos

Applied Computing dept.

Institute for Technological Research (IPT)

São Paulo, Brazil

{eder.barbosa,josias.lima}@ensino.ipt.br;alesan@ipt.br

Patricia Baptista

IN+ Center for Innovation, Technology and Policy Research, LARSyS

Instituto Superior Técnico, Universidade de Lisboa

Lisbon, Portugal

patricia.baptista@tecnico.ulisboa.pt

Abstract—Smartphones connected to the cellular network generate records about connection information (Call Details Record - CDR), creating considerable amounts of telecommunications data, which creates an opportunity to support city management, namely to foster the smart cities concept. This work performs the systematic mapping of research in this area and identifies the main opportunities with cellular infrastructure data to support smart city control centers. The methodological process encompasses a sequence of intricately executed stages, including formulating a research protocol, utilizing search engines, rigorous application of inclusion and exclusion criteria, meticulous selection of relevant studies, thorough data validation, and exhaustive synthesis of research outcomes. Thus, 72 papers represent use cases divided into smart services groups from ISO 37122. As a result, Mobility and Urban planning groups presented the most studies and uses, recently increasing the research in other axles, such as Economy, Environment and Climate Change, Health, Social Conditions, and Security. However, there are challenges to overcome to increase smart services to cities.

I. INTRODUCTION

Building and transitioning to smarter cities is a yearning of society. The significant evolution towards this more sustainable model tends to be more and more frequent and influences the way the city is currently managed and administered [1]. In this way, several sources of information from ICT infrastructures emerge with modern alternatives to incorporate big data scenarios to support the management of smart cities. However, there are still challenges to the implementation of this scenario, especially considering the public policies that must be established for this purpose [2].

There are several definitions to smart cities, and smart service in this concept. However, The ISO 37120 series [3] define indicators for city services and quality of life, which become a reference to establish criteria, and, in particular way,

Thanks for support from grant #2020/01275-7, #2017/50343-2, #2014/50937-1, #2015/24485-9 São Paulo Research Foundation (FAPESP), and CNPq proc. 465446/2014-0, CAPES Código de Financiamento 001. Thanks are also due to Project BE.Neutral – Agenda de Mobilidade para a neutralidade carbónica nas cidades, contract number 35, funded by the Resilience and Recovery Plan (PRR) through the European Union under the Next Generation EU

the ISO 37122 [4] establish indicators for evaluating smart cities.

The ISO 37122 standard establishes a common set of indicators and evaluation methodologies, enabling a more consistent and accurate comparison between cities.

ISO 37122 allows cities to compare themselves internationally, identifying their strengths and areas that need improvement in other cities. This is particularly relevant for cities seeking to attract investments, demonstrating their commitment to sustainable development and quality of life. In summary, ISO 37122 is essential for smart city services as it provides a common reference framework, sustainability guidance, performance monitoring, and urban development.

Governments should develop ICT strategies, enabling the improvement of service operations, economy, and conservation of the environment for future generations, reflecting directly on the management of cities. Therefore, intelligent mechanisms become fundamental for making more assertive decisions [5]. A good initiative is to consolidate ICT in a Control Center (CC), a centralized command communication environment to strategic decisions [6].

In an urban scenario, information from the city and citizens is uploaded to the Internet with a ubiquitous and pervasive approach provided by telecommunication infrastructures [7]. Registrations on calls, short messages (sms) and data transfers, named Call Details Records(CDR), build a big data environment that can become an opportunity for city data analysis.

This paper does a systematic mapping for understanding the use of CDR for smart services to cities, in order to identify new routes to improve the city management and decision-making process, where the information provided by smart services are based mainly in CDR, and could be integrated in smart city control centers.

II. CDR CHARACTERISTICS

Call data records contain an enormous amount of information on how, when, and with whom or what we communicate. Usually, mobile phone operators use CDRs for billing purposes, however, other uses are being investigated [8]. In the

pursue of smarter cities, this information is a great opportunity for research, and smart city control centers could improve their operations using analysis from this source.

Each transaction (voice, data, or SMS) from mobile phones with the telecommunication system generates "Call Details records". These records have aspects that must be considered, such as:

- CDR associate their location based on the position of the Radio Base Station (RBS) to which they connected, however the location of the mobile device is estimated within the radius of operation of the RBS, which can vary from region to region, based on an RBS distribution planning that uses demand criteria as the basis of this placement. Thus, the radius can vary from 100m (micro-cells) to kilometers (rural area or low demand region).
- The number of CDRs varies according to mobility and connectivity standards from smartphones, with operators considering an average of between 50 and 100 CDRs per day, thus this can generate billions of CDRs in big cities everyday.
- The voice and SMS generated CDR directly, when the user send a message or execute a phone call, however, the data flow could generate CDRs indirectly, due to background app inside smartphone. Furthermore, each type of CDR (sms, voice and data) is in a separated data base, with complementary information for each one, for example, in voice CDR has phone number from origin and destination call, however in data CDR this information does not exist.
- The aspects about information security over these CDR are complex, with a legal requirements involving GDPR roles, privacy and telecommunication regulations.

III. METHODS

This paper conducts a systematic mapping for providing an overview of the challenges and opportunities of using CDR for smart services. We adopted the approach from [9], following the following steps: (1) definition of research protocol; (2) the application of the search in scientific research bases; (3) the selection, prioritization, and refinement of search results; and (4) critical analysis.

The Methodology had three levels (Figure 1). Level 1 performs automatic search on scientific bases based on the string; Level 2 is a filtering process where we select papers that may have aspects that answer the research questions through the analysis of titles and abstracts; in Level 3, we use analytical reading to identify the use cases in smart service from ISO 37122, and how CDR was used by the solutions.

A. Protocol definition

1) Research questions: The main goal is to identify the main challenges of applying CDR in smart cities. However, these challenges are not presented directly, justifying questions that indirectly meet these challenges through the analysis of uses in smarter services.

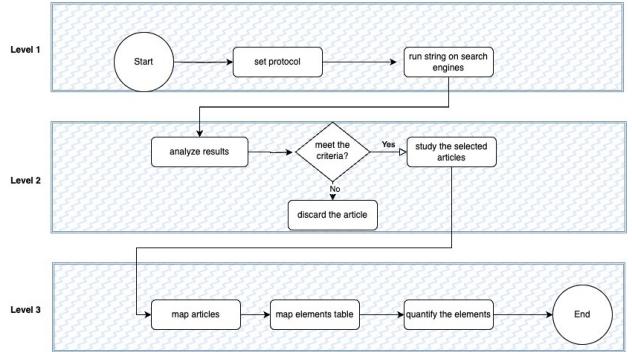


Fig. 1. Methodological flowchart. Adapted from: [9]

- **RQ1. Which ISO37122 smart services groups are using CDR to enable services?** This will allow to identify the scope of use of CDRs for city management;
- **RQ2. How smart services use CDR and their main objectives?** Some papers present specific benefits of CDR, and we will identify the most used strategies;
- **RQ3. What are the challenges of using CDR for managing smart cities through control centers?** Some papers mention the difficulties and challenges of achieving the expected results, which may be related to the use of CDR.

2) *Search engines and Strings:* IEEEXplorer; Scopus; ACM Digital Library; and Web of Science. The search expression was: (((“call detail record” OR “call detail records” OR “call details records” OR “calls details records”) AND (“smart city” OR “smart cities” OR “smart finance” OR “smart grid” OR “smart health” OR “smart housing” OR “smart logistics” OR “smart mobility” OR “smart payment” OR “smart tourism” OR “smart environment”))).

3) Criteria:

- Inclusion Criteria: studies that represent work models, techniques and processes for the use of CDRs in the infrastructure of cellular networks applied in some service for the management of Cities; recent studies published between 2019 and 2022; Open-access databases was considered with peer-reviewer papers.
- Exclusion Criteria: Studies that use other technologies (GPs, bluetooth, wifi, etc) from the cellular environment ecosystem, other than through CDR analysis, to provide smart services in cities.

B. Executing and refining searches

Table I presents the automatic execution of the search strings (Level 1), which undergo a second interaction (Level 2), which applies the inclusion and exclusion criteria. In the last iteration (Level 3), there is a need for manual reading and inspection models to assess whether the paper fits within the scope of answering the research questions (RQ1, RQ2, RQ3).

At the end of this whole process, we selected 72 papers relevant to our research.

ANEXO B – ARTIGO DO CASO DE USO IET

CASE STUDY

Smart resilience through IoT-enabled natural disaster management: A COVID-19 response in São Paulo state

Alessandro S. Santos¹  | Icaro Goncales¹  | Angelina Silva¹  | Rodrigo Neves¹  |
 Igor Teixeira¹ | Eder Barbosa¹  | Vagner Gava¹  | Olga Yoshida² 

¹Digital Technologies, Institute for Technological Research, São Paulo, Brazil

²Regulatory and Metrological Technologies, Institute for Technological Research, São Paulo, Brazil

Correspondence

Alessandro S. Santos.
Email: alesan@ipt.br

Present address

Alessandro S. Santos, Rua Prof. Almeida Prado, 532, Butantã, São Paulo, Brazil.

Funding information

National Council for Scientific and Technological Development, Grant/Award Number: 465446/2014-0; São Paulo State Research Support Foundation, Grant/Award Numbers: 14/50937-1, 15/24485-9, 17/50343-2, 20/01275-7

Abstract

Natural disaster management approach establishes stages of prevention, preparation, response, and recovery. With the Internet of Things (IoT), Bigdata, Business Intelligence, and other Information Communication Technologies, data can be gathered to support decisions in stages of the response to natural disaster events. In biological natural disasters, the ICTs can also support efforts to promote social distancing, public health, and economic monitoring to face the threads. São Paulo state used IoT in scenarios to face COVID-19, such as monitoring vehicular interurban mobility, social distancing, and economic activity. Frameworks, strategies, data views, and use cases are presented to support the decision-making process to face this biological natural disaster. The data-driven approach supports several purposes, including the communication of social distancing indices, economic recovery, the progression of contagion, and deaths. It also played a pivotal role in fostering transparency initiatives for society and supporting the crisis committee by facilitating situational analyses, and this approach became standard practice for pandemic response. Studies and innovative visualisation perspectives have produced positive outcomes, guiding the decision-making process through data analysis. Noteworthy use cases were interurban traffic fence monitoring; mapping of virus spreading; tracking the economic impact concerning recovery plans; and, evaluating the effectiveness of public policies.

KEY WORDS

city design, governance, planning & policy, IoT and mobile communications, smart cities applications, transport control

1 | INTRODUCTION

A modern model of understanding cities emerges within a concept of smart, sustainable, and resilient cities. Public policies should consistently address issues pertaining to quality of life, health, and well-being. City resilience, a crucial aspect of smart cities, enables sustainable adaptation to urban ecosystem changes and a swift return to normalcy following abnormal situations. Furthermore, the “Sustainable Development Goals 13” mentions that cities must strengthen their resilience and adaptive capacity to climate-related hazards and natural disasters in all countries [1].

Natural disasters have a direct impact on the resilience of cities, with a striking statistic revealing that 59% of cities with populations exceeding 500,000 in 2018 were highly vulnerable to at least one type of natural disaster [2]. This finding emphasises the criticality and challenges faced by public managers in their arduous task of providing citizens with a high quality of life.

The Centre for Research on the Epidemiology of Disasters (CRED), a partner agency of the World Health Organisation, has categorised natural disasters into Meteorological, Biological, Climatological, Geophysical, Hydrological, or Extraterrestrial events [3]. Effective natural disaster risk management

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *IET Smart Cities* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

follows a well-defined life cycle that encompasses best practices implemented by civil protection entities. Figure 1 illustrates the temporal relationship between the stages of a natural disaster management cycle: Prevention, preparedness, response, and recovery [4].

Civil Protection outlines several stages, including PREVENTION, PREPAREDNESS, RESPONSE, and RECOVERY, each with specific objectives in disaster management. PREVENTION encompasses actions aimed at minimising risks and mitigating the impact of a disaster. PREPAREDNESS focuses on maintaining a state of readiness and building capabilities to swiftly respond to a disaster. RESPONSE involves providing emergency services and public assistance during or immediately after a disaster. RECOVERY entails the restoration process, which includes efforts to address and reduce factors that contribute to disaster risks.

São Paulo stood as one of the initial Brazilian states grappling with the onset of the COVID-19 pandemic. A pivotal development in the state's response was the implementation of the Intelligent Monitoring System (SIMI). Its core mission was the consolidation and harmonisation of data pertaining to vehicular and human mobility, epidemiological conditions, and economic factors. The aim was to provide crucial support to the state government in its strategic decision-making processes.

Overcoming a set of substantial challenges, such as data integration, ensuring anonymity and privacy, and fostering multidisciplinary collaboration. These challenges were instrumental in yielding results that contributed to scenario analysis and offered improved insights for informed decision-making. The indicators generated by SIMI wielded a profound impact across São Paulo society, influencing the establishment of criteria for the flexibilisation or restriction of economic sectors and guiding strategic shifts in response to the pandemic.

The entire set of strategies in the creation, operation, and improvements of SIMI adopted an approach based on natural disaster risk management. In this case study, we intend to record the main strategies adopted, such as a memory mechanism and improvement of actions to face new biological natural disasters, which can be adopted in an analogous way for

other types of disasters that can use intelligent monitoring technological resources to support decision-making for emergency centres and combating disasters. Thus, we describe the approach, the technical architecture used in the monitoring ecosystem, and the samples of actions and analyses of response and recovery from the natural disaster.

2 | SAO PAULO STATE APPROACH

São Paulo, the largest state in Brazil, boasts a population of over 40 million people and has more than 80 municipalities with populations exceeding 100,000. Furthermore, it holds the distinction of having the largest GDP in Latin America, serving as a vital hub for Brazil's financial and economic activities. With its significant concentration of industries, companies, and financial institutions, São Paulo plays a crucial role in the national economy [5, 6].

Brazil holds the fifth position globally in terms of the highest number of registered COVID-19 cases [7, 8]. The first reported case of COVID-19 in Brazil originated in São Paulo (February 25th, 2020), which had the highest number of cases in the country. For example, when the WHO Director informed that COVID-19 was not a public health emergency of international concern (May 5th, 2023), São Paulo state had 6,589,278 cases and 179,991 deaths by COVID-19. There are several factors contributing to the fast spread of the virus in the state, including its high population density, a large fleet of vehicles, and a huge interurban traffic flow (with 988 million vehicles passing through toll plazas in 2022) [9].

Figure 2 presents four graphics with data sets: (i) daily number of cases and (ii) deaths reported [8], (iii) percentage difference in the number of people visiting the retail store, compared to the pre-pandemic baseline, based in google mobility report [10], and (iv) social isolation rate. The São Paulo quarantine started in 24th march, 2020 (milestone (a)), only after almost a year, the vaccination process began with the first dose of the COVID vaccine (milestone c - January 27th, 2021), which resulted in a subsequent decrease in mortality just after a few months. In the meantime, protective actions (social distancing, mask, and others) were the health prevention measures, while economically, a recovery plan began on June 3rd, 2020 (milestone b), which allowed the gradual opening of strategic sectors; however, there were moments of setback when the pandemic indicators worsened, which can be seen in Figure 2(iii) with peaks of reduced presence in retail and leisure locations when plans suggested closing these locations. Social distancing was monitored daily, peaking on weekends and holidays, with the highest rates being reached at the beginning of the pandemic (Figure 2(iii)).

Natural disasters can have devastating impacts on communities and societies. Disaster management requires processes, modern technology for data analysis, Internet of Things (IoT), and architectural approaches, which play a critical role in supporting decision-making and actions during these events.

The emergency required an agile implementation approach, which allowed gradual continuous improvement cycles, where



FIGURE 1 Risk management strand [4].