

MARCELO DE ALMEIDA BUOSI

MOBILIDADE INTELIGENTE:

**PROPOSTA DE FRAMEWORK DE BIG DATA ANALYTICS PARA ANÁLISE DE
DADOS DE MOBILIDADE URBANA EM UMA SMART CITY**

Dissertação apresentada ao Curso de Pós-graduação em Administração, da Universidade Estadual de Santa Catarina, como requisito parcial para a obtenção do grau de Mestre em Administração.

Orientador: Prof. Júlio da Silva Dias

**FLORIANÓPOLIS
2018**

B944m Buosi, Marcelo de Almeida

Mobilidade inteligente: proposta de framework de big data analytics para análise de dados de mobilidade urbana em uma smart city / Marcelo de Almeida Buosi. - 2018.

152 p. il.; 29 cm

Orientador: Júlio da Silva Dias

Bibliografia: p. 135-145

Dissertação (Mestrado) - Universidade do Estado de Santa Catarina, Centro de Ciências da Administração e Socioeconômicas, Programa de Pós-graduação em Administração, Florianópolis, 2018.

1. Planejamento - Cidades e vilas. 2. Planejamento urbano. 3. Big data. I. Dias, Júlio da Silva. II. Universidade do Estado de Santa Catarina. Programa de Pós-Graduação em Administração. III. Título.

CDD: 307.12 - 20. ed.

Ficha catalográfica elaborada pela Bibliotecária Alice de A. B. Vazquez CRB14/865
Biblioteca Central da UDESC

MARCELO DE ALMEIDA BUOSI

**MOBILIDADE INTELIGENTE: PROPOSTA DE FRAMEWORK DE BIG DATA
ANALYTICS PARA ANÁLISE DE DADOS DE MOBILIDADE URBANA EM UMA
SMART CITY**

Dissertação apresentada ao curso de Mestrado Profissional em Administração do Centro de Ciências Socioeconômicas da Universidade do Estado de Santa Catarina, como requisito parcial para obtenção do grau de Mestre em Administração.

Banca Examinadora

Orientador:

Prof. Dr. Júlio da Silva Dias
Universidade do Estado de Santa Catarina

Membro:

Prof. Dr. Carlos Roberto de Rolt
Universidade do Estado de Santa Catarina

Membro Externo:

Prof. Dr. Werner Kraus Júnior
Universidade Federal de Santa Catarina

Florianópolis, 31 de julho de 2018

AGRADECIMENTOS

A minha esposa, Janaína Ramos, por seu amor e apoio nos momentos e decisões mais difíceis e sua parceria nos momentos de alegria.

Aos meus pais, Maria Luiza de Almeida e Manoel Buosi, pelo respeito e apoio às minhas decisões.

A todos os colegas de sala, pela ajuda, união e amizade que mantivemos durante todo o curso e que certamente permanecerá.

Aos professores Christiane Kalb, Thiago Siqueira e Sérgio Boeira, pela amizade, paciência e ajuda no meu processo de reinserção ao mundo acadêmico e na preparação para o processo seletivo do curso.

Ao meu orientador Prof. Júlio da Silva Dias, pela sua orientação e amizade.

Ao Prof. Carlos de Rolt, pela organização do processo de trabalho, idealização do projeto do laboratório e amizade.

Aos professores Mário Moraes e Nério Amboni, nas orientações diretas sobre tópicos da dissertação.

Aos todos os professores da ESAG, que sempre estiveram à disposição para orientar nos assuntos mais diversos.

À CAPES, pelo financiamento da bolsa de estudos.

A Piero Zito, da UNIBO, pelo desenvolvimento técnico das ferramentas de análise de dados e pela amizade.

A Tatiane Simm, pela sua amizade e participação nas atividades do laboratório.

Ao Prof. Luca Foschini, da UNIBO, pelo conhecimento passado sobre o tema.

A Felipe Martins, pela elaboração das ilustrações nos tópicos de proposta.

A todos que colaboraram, direta e indiretamente, à elaboração do trabalho e em meu período de envolvimento com a ESAG/UDESC.

RESUMO

O presente trabalho tem por objetivo elaborar um *framework* para processos de análise de dados de mobilidade urbana em uma *smart city*, motivado pelo constante crescimento urbano e os impactos que isto vem causando na vida das pessoas no que diz respeito à qualidade de vida nas áreas urbanas. Nas estratégias para aprimorar o planejamento de trânsito, tanto para o transporte público quanto para o particular, a tecnologia atua como importante mecanismo de auxílio nas soluções orientadas a um trânsito inteligente e mobilidade sustentável. O conceito *smart city* surge como oportunidade para responder às necessidades da mobilidade inteligente e sustentável fazendo o uso de tecnologias inovadoras, como o *crowdsensing*, que explora a inteligência coletiva e organiza a colaboração de grandes grupos de pessoas e as soluções de *big data analytics*, utilizadas para extrair informações e conhecimento da massa de dados coletados através da aplicação de métodos analíticos, entre outras. Analisamos teorias e propostas de *framework* na literatura que pudessem auxiliar para a maior compreensão do assunto discutido, gerenciamos uma campanha de *crowdsensing* para coletar dados de geolocalização dos usuários envolvidos, submetê-los a rotinas de organização, processamento de dados e análise do conhecimento gerado, elaboramos um levantamento de requisitos, com base nas experiências da leitura dos estudos revisados e da campanha de *crowdsensing* realizada, para fornecer diretrizes para formulação do *framework*. O *framework* está estruturado em cinco categorias: coleta, organização e armazenamento, processamento, análise e publicação; e quatro abordagens de análise de mobilidade urbana. Ao final do trabalho, identificamos a importância de se analisar o desempenho do transporte público, da utilização de métodos de organização de dados, de engajar e comunicar claramente as pessoas participantes de uma campanha de *crowdsensing*, e a falta de organização sistêmica do fluxo de dados e de padronização dos processos de análise.

ABSTRACT

The present dissertation aims to elaborate a data analysis framework for urban mobility in a smart city, motivated by constant urban growth and its impacts over people's lives regarding life quality in the urban area. On traffic plan strategies, for both public and private transportation, technology acts as an important assistance mechanism for smart and sustainable mobility-oriented solutions. Smart city concept emerges as an opportunity to attend to smart and sustainable mobility needs using innovative technologies, such as crowdsensing, which exploits group intelligence and organizes collaboration amongst big groups of people, big data analytic solutions, used to extract information and knowledge from massive data collected through analytical methods application, among others. We analyzed theory and framework proposals that could help us understand deeply the topic discussed, we managed a crowdsensing campaign to collect geolocalization data from the users involved, subject them to data organization and process procedures, and analyze the knowledge arose and elaborated function requirements, based on the literature studied and the crowdsensing campaign, to supply guidelines for the framework formulation. The framework is structures in five categories: collect, organization and storage, processing, analysis and publication; and four urban mobility analysis approach. In the end of the study, we identified the importance of public transportation analysis, data organization methods, engagement and clear communication with crowdsensing participants, and the lack of a systematic organization of data flow and analysis process patterns.

LISTA DE ILUSTRAÇÕES

Lista de figuras

Figura 1: Fontes, abordagens analíticas e aplicações	62
Figura 2: Padrão de modelo de processos de negócio para ETL	65
Figura 3: Proposta de <i>framework</i> para dados em tempo real.....	67
Figura 4: Modelo de ambiente Big Data.....	68
Figura 5: Modo de deslocamento inferido pelo MMA	69
Figura 6: Arquitetura de sistema.....	70
Figura 7: Mapa de calor dinâmico de densidade.	73
Figura 8: Padrão de mobilidade de passageiros.....	75
Figura 9: Interfaces da análise de atrasos	76
Figura 10: Identificação de regiões com maior demanda de mobilidade	77
Figura 11: Regiões com maior densidade.....	78
Figura 12: Mapa de áreas com diferentes densidades.....	82
Figura 13: Arquitetura de três camadas isoladas	83
Figura 14: Exemplo de uma matriz O/D.....	84
Figura 15: <i>Clusters</i> de regiões e sua quantidade de deslocamentos	84
Figura 16: Exemplo de movimentações entre regiões	85
Figura 17: Fluxo de viagens entre regiões agregadas.....	86
Figura 18: Visão do Output Data Explorer.....	87
Figura 19: Representação do fluxo dos dados	98
Figura 20: Arquitetura para desenvolvimento de análise de dados.	99
Figura 21: Fluxo de informações do módulo de análise de dados PIR	102
Figura 22: Distribuição da quantidade de dados e trajetos W1	103
Figura 23: <i>Heatmap</i> de concentração de pessoas	104
Figura 24: Distribuição dos dados e pontos de interesse (continua)	104
Figura 25: Distribuição dos dados e pontos de interesse (conclusão)	105
Figura 26: Fluxo macro do processo do <i>framework</i>	112
Figura 27: Fluxo da etapa de organização e armazenamento	116
Figura 28: Representação da etapa de processamento.....	119
Figura 29: Representações das análises MCS	120
Figura 30: Representações das análises CPT.....	121
Figura 31: Representações das análises DTP	122

Figura 32: Representações da análise PU	122
Figura 33: Plataformas para disponibilização de resultados e interação.....	124
Figura 34: Fluxo geral do <i>framework</i>	126

Lista de gráficos

Gráfico 1: Taxa de urbanização do Brasil	28
Gráfico 2: Evolução do crescimento populacional brasileiro	29
Gráfico 3: Comparativo da variação do percentual de crescimento.....	30
Gráfico 4: Comparativo da evolução das viagens por modo	31
Gráfico 5: Evolução nos equipamentos usados na mobilidade	32
Gráfico 6: Evolução dos custos por tipo de transporte (bilhões de R\$/ano)	34
Gráfico 7: Evolução dos custos por tipo de custo (bilhões de R\$/ano).....	34
Gráfico 8: Quantidade de detecções identificadas a cada três horas.....	72
Gráfico 9: Comparativo de densidades de quatro <i>hotspots</i> diferentes	73
Gráfico 10: Atividade de telefone celular em Erlang.....	80
Gráfico 11: Atividades de celulares normalizados.....	81
Gráfico 12: Quantidade de viagens entre duas regiões	87

LISTA DE TABELAS

Tabela 1: Quantidade por tipo de veículo no país	30
Tabela 2: Comparativo do crescimento da população e de veículos	30
Tabela 3: Evolução das viagens por modo	31
Tabela 4: Evolução na quantidade dos veículos	32
Tabela 5: Distância percorrida pelas pessoas (bilhões de km/ano)	33
Tabela 6: Consumo de tempo por modo (bilhões de horas/ano)	33
Tabela 7: População por ano.....	35

LISTA DE QUADROS

Quadro 1: Exemplo de identificação de peso por número de avaliações	78
Quadro 2: Resumo da característica metodológica das etapas	94
Quadro 3: Características da etapa de coleta	114

SUMÁRIO

1	RESUMO EXECUTIVO	19
2	INTRODUÇÃO	21
2.1	Descrição da situação problema e oportunidade identificada.....	21
2.2	Frase da situação-problema.....	24
2.3	Objetivos.....	24
2.3.1	Objetivos específicos	24
2.3.2	Objetivo geral.....	25
2.4	Justificativa do estudo.....	25
2.5	Contribuição do trabalho	26
3	O CONTEXTO E A REALIDADE INVESTIGADA	27
3.1	Crescimento URBANO E populacional	27
3.2	A mobilidade urbana.....	36
3.3	Mobilidade urbana como política e planejamento	39
4	DIAGNÓSTICO DA OPORTUNIDADE IDENTIFICADA	43
4.1	<i>Smart city</i> e a mobilidade urbana inteligente	43
4.2	Tecnologias inovadoras da <i>smart city</i>	47
4.2.1	<i>Big Data</i>	47
4.2.2	<i>Big data analytics</i>	50
4.2.3	<i>IoT e Cloud Computing</i>	52
4.2.4	<i>Crowdsensing</i>	53
4.3	Estudos correlatos de <i>Smart city</i> /Mobilidade urbana	56
5	MODELOS DE ANÁLISE DE DADOS DE MOBILIDADE URBANA	61
5.1	Etapas envolvidas no processo de análise.....	61
5.2	Arquiteturas de análise de dados	67
6	PROCEDIMENTOS METODOLÓGICOS	89

6.1	Caracterização da pesquisa.....	89
6.2	Delineamento e métodos de pesquisa.....	91
6.3	Categorias de análise e dimensões	94
6.4	Possíveis limitações da pesquisa	95
7	FRAMEWORK: UMA PROPOSTA	97
7.1	Descrição pré-experimental da campanha de <i>crowdsensing</i>	97
7.1.1	ParticipACT Brasil, a plataforma de <i>smart city</i> utilizada.....	97
7.1.2	Métodos de processamento de dados e análise.....	99
7.1.3	Resultados experimentais	102
7.1.4	Questionário aplicado	105
7.2	Levantamento de requisitos.....	107
7.3	Apresentação do <i>framework</i> :.....	110
7.3.1	Coleta.....	113
7.3.2	Organização e Armazenamento.....	114
7.3.3	Processamento	116
7.3.4	Análise.....	119
7.3.5	Publicação.....	122
7.3.6	<i>Framework</i>	125
7.4	Discussão dos resultados	127
8	CONCLUSÃO/RECOMENDAÇÕES	129
9	REFERÊNCIAS	135
10	ANEXOS	147
10.1	Anexo 1: <i>Banner</i> de <i>feedback</i> da campanha de <i>crowdsensing</i>	147
10.2	Anexo 2: Questionário aplicado aos participantes da campanha.....	148

1 RESUMO EXECUTIVO

O crescimento da urbanização se manifesta diretamente na vida das pessoas, principalmente no que diz respeito ao bem-estar social. Este cenário se verifica em diversas cidades, especialmente nas grandes metrópoles, com aumento de população e de veículos automotores. Este cenário caótico contribui para o aumento do tempo e distância dos deslocamentos na cidade, exigindo ações corretivas e adaptativas não sustentáveis e que, em sua maioria, não estão em harmonia com o meio ambiente ao qual estão sujeitas.

As opções modais de transporte existentes naturalmente são questionadas. Governos de diversos países discutem a crescente utilização do veículo particular e a eficiência dos transportes públicos, desenvolvendo iniciativas e implementando estratégias para aprimorar a mobilidade urbana. A tecnologia é considerada como um dos caminhos para auxiliar nesta questão, ou até mesmo como mudança de paradigma, sendo abordada na orientação para um tráfego inteligente, planejamento urbano e compartilhamento de serviços, com o viés para a mobilidade sustentável.

A mobilidade sustentável compreende acessibilidade e conveniência no deslocamento, com mínimo impacto no meio ambiente. Considera a produtividade, inclusão, segurança, justiça social e pensamento verde como atributos. Na junção com as tecnologias disponíveis, surge o pensamento *smart mobility*, que visa a geração e compartilhamento de conhecimento, aprimoramento dos meios de mobilidade e de sistemas de transporte, com design, sistema, infraestrutura e utilização inteligentes.

Para poder responder às necessidades dos habitantes de forma eficiente e sustentável, segundo a visão da *smart mobility*, o conceito de *smart city* aparece como uma oportunidade, fazendo o uso de tecnologias de informação e comunicação de forma adequada, sempre associado a iniciativas sustentáveis da cidade. Combinando os serviços de diferentes disciplinas, a *smart city* conta com inovações como *machine learning*, *big data analytics*, *crowdsensing*, *internet of things*, *cloud computing*, entre outros.

Destacamos o *crowdsensing*, que explora a inteligência coletiva e organiza a colaboração espontânea e conjunta de grandes grupos de pessoas. Esse conceito faz uso dos dispositivos móveis das pessoas, caracterizando uma nova classe de serviços pervasivos, descartando a necessidade de instalação de sensores pela cidade, aproximando os mundos virtuais e físicos. De maneira geral, é um processo que atribui tarefas para os usuários

participantes das campanhas e aguarda os resultados para posterior exploração dos dados, de forma a aprimorar o gerenciamento das tarefas e das campanhas.

Da necessidade de manipulação de dados em massa, com escalabilidade, heterogeneidade, rapidez e privacidade, as soluções de *big data analytics* surgem como solução. Contemplando as fases de geração de dados, aquisição, armazenamento e análise, essas tecnologias são utilizadas para extrair informações e conhecimento da massa de dados coletados, através da aplicação de métodos analíticos para alcançar resultados descritivos, preditivos, exploratório e prescritivos. Esta tecnologia baseia sua coleta, armazenamento e análise primordialmente nas tecnologias de *smart city* já mencionadas.

Assumindo a mobilidade urbana deficiente como situação problema e vislumbrando as tecnologias inovadoras como oportunidade, o trabalho busca propor um *framework* para processos de análise de dados de mobilidade urbana em uma *smart city*. Para tal objetivo, foi realizada uma revisão bibliográfica, com pesquisas de teorias e propostas de *framework* na literatura que pudessem auxiliar para a maior compreensão do assunto discutido; uma campanha de *crowdsensing* para coletar dados de localização dos usuários envolvidos, submetê-los a rotinas de organização e processamento de dados, e analisar o conhecimento gerado; e um levantamento de requisitos, com base nas experiências da leitura dos estudos revisados e da campanha de *crowdsensing* realizada, para fornecer diretrizes para elaboração do *framework*.

O objetivo foi alcançado com a apresentação detalhada de um *framework* que contempla cinco categorias: coleta, organização e armazenamento, processamento, análise e publicação. Cada categoria respeita abordagens de análise de objetivos e fontes diferentes, e é especificada pela descrição dos subprocessos e representações na forma de figuras e quadros. Ao final do trabalho, destacamos a importância de se analisar o desempenho do transporte público, da utilização de métodos de organização de dados, de engajar e comunicar claramente as pessoas participantes de uma campanha de *crowdsensing*, a falta de organização sistêmica do fluxo de dados e de padronização dos processos de análise.

2 INTRODUÇÃO

Apresentamos, neste capítulo, a mobilidade urbana como situação problema analisada e a tendência *smart city* como oportunidade identificada para auxiliar na solução dos problemas de mobilidade dentro das cidades. Após essa contextualização, descrevemos os objetivos do trabalho, juntamente com a justificativa e contribuições.

2.1 DESCRIÇÃO DA SITUAÇÃO PROBLEMA E OPORTUNIDADE IDENTIFICADA

Para que se mantenham seguras e agradáveis, as cidades precisam consumir uma quantidade de recursos não renováveis muito grande, despejando mais lixo do que a Terra pode absorver e tornando o processo de aproveitamento de recursos inseguro e vulnerável (MITCHELL, BORRONI-BIRD e BURNS, 2010). Apesar disso, esses problemas causados pelo aumento da urbanização também se manifestam diretamente nas vidas das pessoas. Acidentes de trânsito nos países da União Europeia totalizam, anualmente, aproximadamente 43000 casos, deixando 1.8 milhões de pessoas feridas, estimando um custo de 160 bilhões de euros, com 36% de mortes ocorrendo de acidentes (RUSSO e COMI, 2017).

No Brasil, principalmente nas grandes metrópoles, o cenário também é preocupante. De 2000 a 2012, a frota de veículos brasileiros passou de 19,9 para 40 milhões, num aumento de 9:1 para 5:1 na relação habitante por carro (MARTINE, OJIMA e FIORAVANTE, 2012). Esse cenário contribui para o aumento do deslocamento de casa até o trabalho, gerando um desgaste muito grande na população, principalmente das grandes metrópoles, reduzindo a produtividade e causando diminuição do bem-estar, tanto para o indivíduo quanto para a sociedade como um todo (STEFANELLI e PERO, 2015).

Os autores, no contexto apresentado, fazem as seguintes sugestões:

Há atualmente em curso um processo de deterioração das condições de mobilidade urbana, medido pelo aumento do tempo de deslocamento de casa ao trabalho, com impactos sobre o bem-estar dos trabalhadores, principalmente dos mais vulneráveis. Assim, é imperativo que os formuladores de políticas públicas reorientem suas prioridades no que diz respeito ao meio urbano e iniciem um vasto empreendimento de recuperação dos sistemas de transporte públicos metropolitanos, com foco na modernização, acessibilidade, ampliação, integração e sustentabilidade (STEFANELLI e PERO, 2015, p. 399).

Apesar da urgente necessidade da melhoria do transporte público coletivo, deve-se fazer um uso racional do veículo particular, adequando-o às reais condições de ocupação do espaço urbano e ao planejamento. Essa problemática deve ser enfrentada seriamente pelos governos, articulando-se o discurso de forma ampla, com todas as áreas envolvidas. Isso deve ser feito com certa urgência, “do contrário, a “janela de oportunidade” aberta pela pressão democrática da sociedade poderá ser perdida” (DE ÁVILA GOMIDE e GALINDO, 2013, p. 36).

Mas Pavone et al. (2012) entendem que o uso do automóvel particular é insustentável, pois “a demanda para rodovias e vagas de estacionamento irá aumentar dramaticamente, enquanto o espaço urbano disponível continua a diminuir” (PAVONE, SMITH, *et al.*, 2012, p. 839). Portanto, uma mudança de paradigma é necessária, onde a automação atuaria na mobilidade urbana com conceitos diferentes das atuais políticas em vigência. Os autores entendem que é importante assegurar a mesma qualidade e conforto do automóvel particular sem ser necessário aumento de infraestrutura urbana.

Os governantes de diversos países têm se preocupado com essas questões, desenvolvendo iniciativas para estudar a implementação de soluções. Em 2011, devido à alta influência do transporte urbano na vida das pessoas, a Comissão Europeia de Mobilidade Urbana incentivou cidades populosas a adotarem o SUMP (Sustainable Urban Mobility Plan), um plano estratégico desenvolvido para satisfazer as necessidades de mobilidade das pessoas e negócios nas cidades para uma melhor qualidade de vida, considerando princípios de integração, participação e avaliação (MAY, BOEHLER-BAEDEKER, *et al.*, 2017).

Com a deterioração da vida urbana nas cidades grandes e o elevado número de protestos contra o transporte público coletivo, o governo brasileiro anunciou, em 2013, o Plano Nacional de Mobilidade Urbana. Em um pacto com os governos estaduais e prefeituras das principais cidades do país, discutiu-se a melhoria da qualidade dos serviços de transporte coletivo, a maior participação e controle social dos serviços e redução das tarifas para os usuários (DE ÁVILA GOMIDE e GALINDO, 2013).

Conforme detalhado no PLAMUS (Plano de Mobilidade Urbana Sustentável), vários projetos foram estudados na Grande Florianópolis, visando identificar vários problemas de mobilidade urbana, nas esferas histórica, de atividade urbana, transporte coletivo, projetos, rede viária e infraestrutura, com o intuito de determinar as oportunidades que surgem de cada problema. Entre eles, temos a Concessão da Via de Contorno; Autopista Litoral Sul; Estudo de

Transporte Hidroviário, com duas rotas escolhidas; Proposta de Linha de Teleférico para Florianópolis, entre TICEN e UFSC; Estudos de aproveitamento da Ponte Hercílio Luz, com uso exclusivo da ponte para transporte coletivo e/ou transporte não motorizado; Corredores de BRT; Estudos de viabilidade para implantação de tecnologia sobre trilhos; Pesquisa de hábitos de turismo; Estudos sobre o novo Aeroporto Hercílio Luz; e Bacias Cicloviárias (PLAMUS, 2014).

Há diferentes abordagens da utilização das tecnologias para solucionar os problemas de mobilidade urbana, tais como tecnologias orientadas à técnica, buscando um tráfego inteligente, planejamento social e urbano, orientadas a veículo privado, apoiando-se no conceito do carro inteligente, e orientadas a serviço, onde há serviços aprimorados pela digitalização para atender a problemas de mobilidade e tráfego. O grande desenvolvimento da capacidade de coleta de dados em tempo real e histórico do tráfego aumentou muito a possibilidade de aprimorar a eficiência e sustentabilidade do transporte, ajudando tanto de forma individual quanto aos sistemas de serviço de tráfego, através de políticas de transporte, operacional e serviços digitais (HEISKALA, JOKINEN e TINNILÄ, 2016).

Políticos, pesquisadores e planejadores estão trabalhando para implementar estratégias e ações para promover crescimento econômico, inclusão social e maior qualidade de vida, principalmente devido aos desafios trazidos pela mudança climática e rápido crescimento urbano. Neste contexto, o paradigma da *smart city* se manifesta, cujo conceito aborda uma cidade que “é apta a responder às necessidades dos seus habitantes numa forma mais eficiente e sustentável, principalmente usando tecnologias da informação e comunicação (TIC) de forma apropriada” (PAPA, CARAGIULO e RUSSO, 2017, p. 409). Apesar de ter crescido muito nos últimos dez anos, este conceito atraiu mais ao setor empresarial do que ao científico e precisa de uma estrutura teórica mais consistente (FISTOLA, RAIMONDO e LA ROCCA, 2017).

A *smart city* combina serviços de diferentes disciplinas de forma interconectada para oferecer aumento na eficiência do transporte e da mobilidade urbana. Isso foi possível graças ao desenvolvimento das tecnologias de *machine learning*, *big data analytics*, manufatura de *hardware* e TIC (Tecnologia da Informação e Comunicação). Principalmente o Big Data, que pode impulsionar a revolução digital na sociedade cada vez mais voltada para o conhecimento (MANIAK, IQBAL e DOCTOR, 2018). Juntamente com o IoT (Internet of Things), o Big Data e o *big data analytics* podem transformar o sistema de mobilidade urbana buscando grande flexibilidade e melhor integração entre os modos de transporte existentes, também permitindo

soluções de mobilidade inteligentes e sustentáveis, como conceitos de compartilhamento de veículos, veículos elétricos, condução autônoma, entre outros (SCHATZINGER e LIM, 2015).

No entanto, os fatores que contribuem com o sucesso de modelos de negócio baseados no conceito *smart city*, tais como *crowdsensing*, ainda não foram aprofundados. Com relação à implementação desses modelos de serviço de mobilidade, Heiskala et al. (2016) sugerem que:

Para aprimoramentos sustentáveis emergirem de serviços de plataforma de múltiplas faces que usam dados de *crowdsensing* no campo do tráfego, o modelo de negócios do provedor de plataforma digital deve ser rentável no longo prazo. Ao mesmo tempo, o serviço deve ser naturalmente benéfico para os usuários para ser atraente de modo geral e compensar o fardo do *crowdsensing* participativo. Além disso, o sucesso das plataformas dupla ou multifacetada normalmente depende de como elas equilibram os custos e os benefícios de todas as faces da plataforma, enquanto fomentam positivamente efeitos da rede entre esses lados. Plataformas de *crowdsensing*, por sua vez, dependem do comprometimento ativo de usuários para fornecer dados e observações (HEISKALA, JOKINEN e TINNILÄ, 2016, p. 38-39).

2.2 FRASE DA SITUAÇÃO-PROBLEMA

O crescimento desmedido da urbanização gera problemas de mobilidade urbana que, devido a limitação e desorganização desses ambientes, causam impactos no meio ambiente e no bem-estar das pessoas.

2.3 OBJETIVOS

2.3.1 Objetivos específicos

- Pesquisar teorias e propostas de *frameworks* e arquiteturas sobre análise de dados de mobilidade urbana sob o pano de fundo da *smart city*;
- Executar e acompanhar uma campanha de *crowdsensing* para coleta de dados de mobilidade e efetuar processamento e análise dos dados;
- Aplicar questionário aos participantes da campanha visando analisar sua experiência e expectativas;
- Elaborar um levantamento de requisitos para processos de análise de dados de mobilidade urbana a partir da campanha executada e literatura estudada.

2.3.2 Objetivo geral

- Propor um *framework* para processos de análise de dados de mobilidade urbana em uma *smart city*.

2.4 JUSTIFICATIVA DO ESTUDO

Além dos problemas econômicos, como a perda de produtividade e o grande custo de investimento dispensado para manter e expandir a infraestrutura de tráfego, causados pelo crescimento urbano e populacional discutidos, ele também leva a sérios problemas sócio ambientais, como congestionamento, poluição, diminuição na qualidade de vida e bem-estar, morte, entre outros. Portanto, o mundo precisa de uma mobilidade urbana mais inteligente.

De acordo com a problematização elaborada anteriormente, pode-se identificar algumas carências apresentadas na situação atual da mobilidade urbana, como o aumento da urbanização, a diminuição da mobilidade nas cidades, os impactos negativos no meio ambiente, um sistema de trânsito não sustentável, diminuição da qualidade de vida nas cidades, diminuição do bem-estar social e a perda de eficiência nas cidades.

Como oportunidades, vemos o aumento do uso de *smartphones*, a maior participação do cidadão, o aprimoramento nos serviços de transporte coletivo, a disseminação de TIC, a disponibilidade de ferramentas de análise de dados, a arquitetura de Big Data e a tendência *smart city*. A estrutura do ParticipACT Brasil que será utilizada neste estudo também se apresenta como um elemento importante. Este projeto tem como objetivo desenvolver a ciência cidadã, com a utilização e disponibilização de dados úteis para o aprimoramento dos serviços na cidade, como dados abertos e sensoriamento a partir de habitantes e instituições públicas e privadas. Com isso, busca auxiliar gestores na solução de problemas urbano (GOMES, DANTAS, *et al.*, 2016).

O projeto ParticipACT Brasil foi desenvolvido pelo Laboratório de Tecnologias de Gestão (LabGES), um dos grupos da linha de pesquisa Gestão, Inovação e Empreendedorismo da ESAG/UDESC, e vislumbra “projetar, promover e desenvolver um sistema de gestão sócio técnica para formar gradual e progressivamente um *Big Data* para analisar problemas de uma

região urbana” (PARTICIPACT BRASIL). A proposta de elaboração de *framework* para análise de dados de mobilidade urbana se mostra aderente aos objetivos do laboratório e da linha de pesquisa.

2.5 CONTRIBUIÇÃO DO TRABALHO

- Fornecer fundamentos para elaboração de uma arquitetura de Big Data que vislumbre o estudo de problemas de mobilidade urbana;
- Auxiliar no planejamento de mobilidade urbana relativo a processos de análise de dados, tanto para transporte individual quanto para transporte coletivo;
- Sensibilizar quanto à importância da abordagem sustentável e de bem-estar social na discussão e estudos de mobilidade urbana e *smart city*;
- Incentivar a adoção do pensamento *smart city* e uso de tecnologias inovadoras para solucionar problemas urbanos, para que a cidade possa se adaptar às necessidades das pessoas e fornecer serviços de forma amigável para cada situação.

3 O CONTEXTO E A REALIDADE INVESTIGADA

Neste capítulo, discorreremos sobre a mobilidade urbana como a realidade investigada. Apresentamos brevemente como se deu o crescimento urbano no Brasil e apresentamos alguns números que evidenciam o crescimento populacional e de veículos no país. Após isso, contextualizamos a mobilidade urbana, apresentando alguns conceitos e como ela é compreendida pelo governo brasileiro.

3.1 CRESCIMENTO URBANO E POPULACIONAL

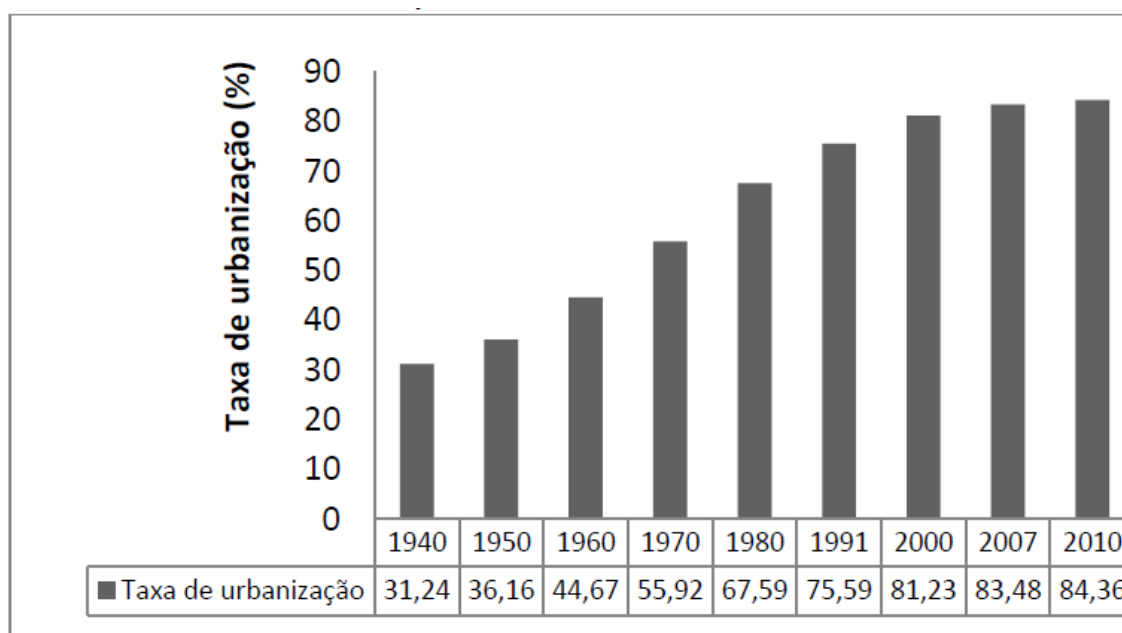
O Brasil caracteriza-se como um país urbano, com grande parte da população vivendo nas cidades. Em meados do século XIX, já era evidente um crescimento populacional desordenado, juntamente com a instauração do processo de Revolução Industrial e crescimento industrial (LODER, 1977). Esse processo de urbanização inicia-se a partir de 1930, com a intensificação da industrialização, com influência decisiva por parte do poder público. O surgimento e ascensão da burguesia e início do fluxo migratório intensificam a mudança na característica produtiva, passando de uma predominância rural para a concentração na industrialização e urbanização (WEISS, 2016).

No entanto, além do fluxo migratório do ambiente rural para o urbano, o aumento da população se dá também por acontecimentos político-sociais, como a interiorização do processo de urbanização – este acelerado por mudanças no plano político econômico – e a sua hierarquização a nível nacional, influenciando a característica de configuração das cidades. Como resultado desse processo, o crescimento desproporcional da população acarretou em problemas de abastecimento alimentício para a população não produtiva, rede de estradas, infraestrutura, entre outros (LODER, 1977).

Até meados de 1920, o país apresentava uma taxa de urbanização de 10,7%, tendo esse indicador aumentado para 31,24% vinte anos depois (WEISS, 2016). O Gráfico 1 retrata a evolução da taxa de urbanização desde então, estimulada pelo cenário descrito. Em 2010, o país já apresenta uma taxa de urbanização de 84,36%.

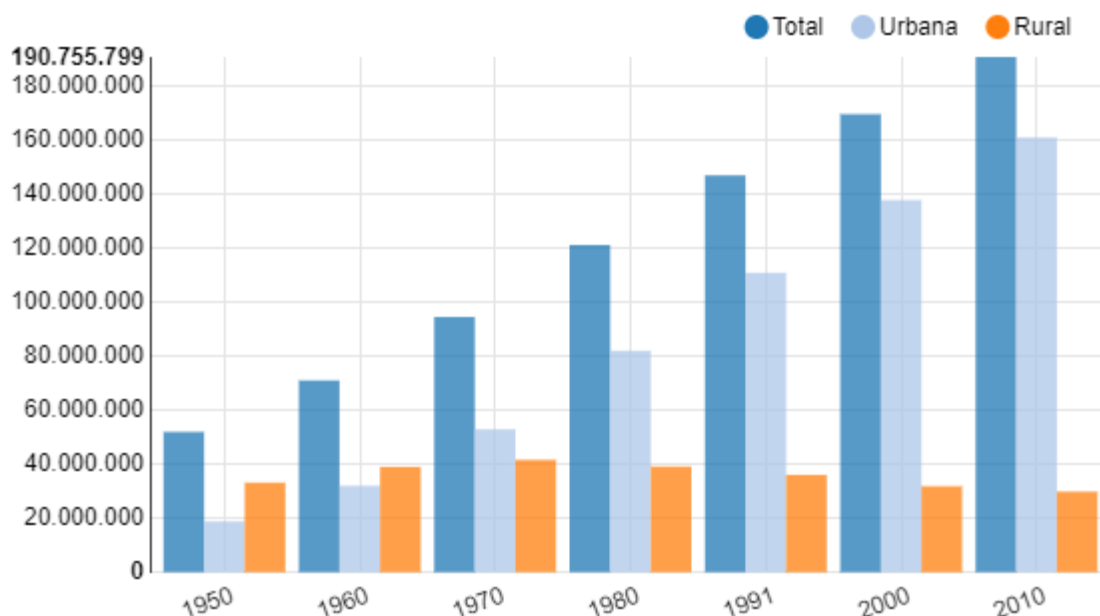
A onda de industrialização levou a problemas sociais, como alta taxa de desemprego nas cidades, aumentando a quantidade de pessoas que se aglomeravam em subabitações e moradias precárias nos centros urbanos. A maior parte dos trabalhadores buscava viver nas regiões periféricas, principalmente pela maior atratividade de aquisição, e pessoas de classe social mais privilegiada fugiam da concentração de pessoas e violência dos grandes centros, buscando viver em áreas afastadas dos centros industriais que permitissem um padrão de vida mais apropriado. Há então o surgimento de bairros e condomínios fechados, exigindo do município investimentos em estrutura viária para que se possa acessar os centros industriais via automóveis, atrelando, portanto, a mobilidade urbana ao uso de veículos particulares (WEISS, 2016).

Gráfico 1: Taxa de urbanização do Brasil



Fonte: (WEISS, 2016, p. 50).

Segundo o IBGE, cujo último censo foi realizado em 2010, o Brasil quase quadruplicou a sua população em 60 anos. Em 1950, o país contava com aproximadamente 50 milhões de habitantes, apresentando em 2010 cerca de 190 milhões, como podemos verificar no Gráfico 2. Esse crescimento é acompanhado do crescimento da população urbana, que aumentou de menos de 20 milhões para cerca de 160 milhões no mesmo período. O Gráfico 2 também nos permite verificar que a variação da população rural não ocorre na mesma proporção, mesmo apresentando queda a partir de 1970, baixando de aproximadamente 40 milhões para quase 30 milhões de habitantes.

Gráfico 2: Evolução do crescimento populacional brasileiro

Fonte: Elaborado a partir dos dados do IBGE (2010).

Apesar da queda na taxa de natalidade, o IBGE (2010) estima que a população seguirá crescendo, apresentando um total de 209.186.802 habitantes em 2018.

O país possui uma frota de aproximadamente 91 milhões de veículos, segundo último censo do IBGE (2010). Isso resulta em quase 1 veículo para cada 2 habitantes. E como podemos verificar na Tabela 1, a quantidade de veículos particulares como automóveis e motocicletas é a mais significativa entre todos os veículos, com cerca de 80% do total, resultando em quase 1 a cada 3 habitantes.

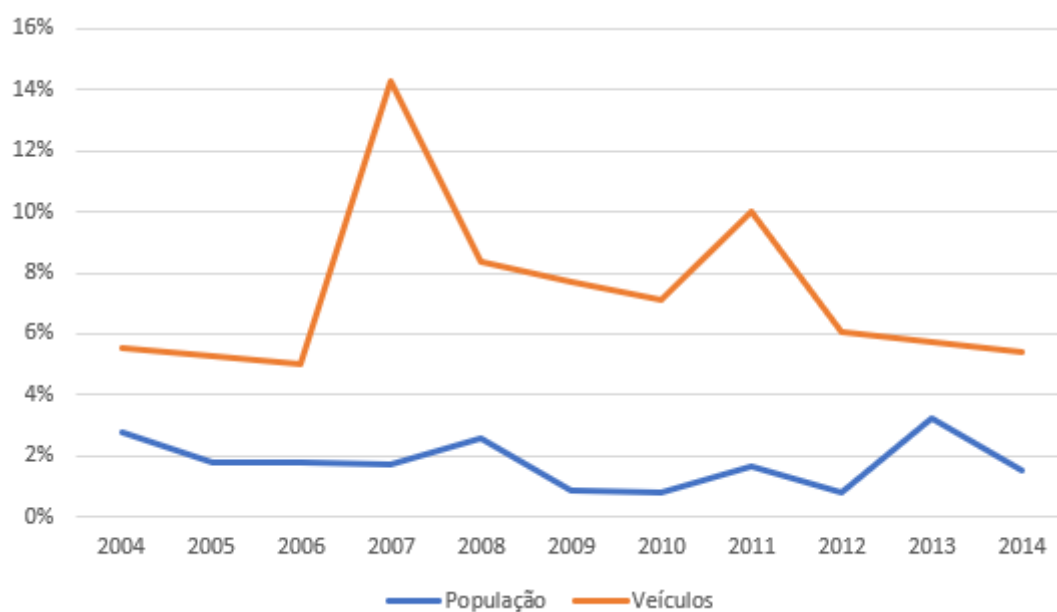
Ao analisarmos comparativamente a taxa de aumento de veículos com a da população no país pelo Gráfico 3, notamos uma desproporção. Segundo a ANTP (2016), a população aumentou em mais de 21% de 2003 a 2014. Já os veículos apresentaram um aumento de mais de 115% nos municípios que integram o sistema de informações da ANTP¹ (Tabela 2).

¹ Os municípios que integram o sistema de informações da ANTP são 438 municípios, com 60.000 ou mais habitantes em 2003, obtidos por meio de questionário enviado pela ANTP e preenchidos pelos responsáveis do transporte e trânsito municipais e metropolitanos (ANTP, 2016).

Tabela 1: Quantidade por tipo de veículo no país

Tipo de veículo	Quantidade
Automóveis	51.296.981
Caminhões	2.684.227
Caminhões-trator	606.679
Caminhonetes	6.880.333
Caminhonetas	3.053.759
Micro-ônibus	383.325
Motocicletas	20.942.633
Motonetas	3.990.558
Ônibus	601.522
Tratores	30.896
Utilitários	707.152

Fonte: (IBGE, 2010).

Gráfico 3: Comparativo da variação do percentual de crescimento

Fonte: gerado a partir dos dados da ANTP, (2016)

Tabela 2: Comparativo do crescimento da população e de veículos

Informação	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
População (milhões)	108	111	113	115	117	120	121	122	124	125	129	131
Veículos (milhões)	18	19	20	21	24	26	28	30	33	35	37	39

Fonte: (ANTP, 2016).

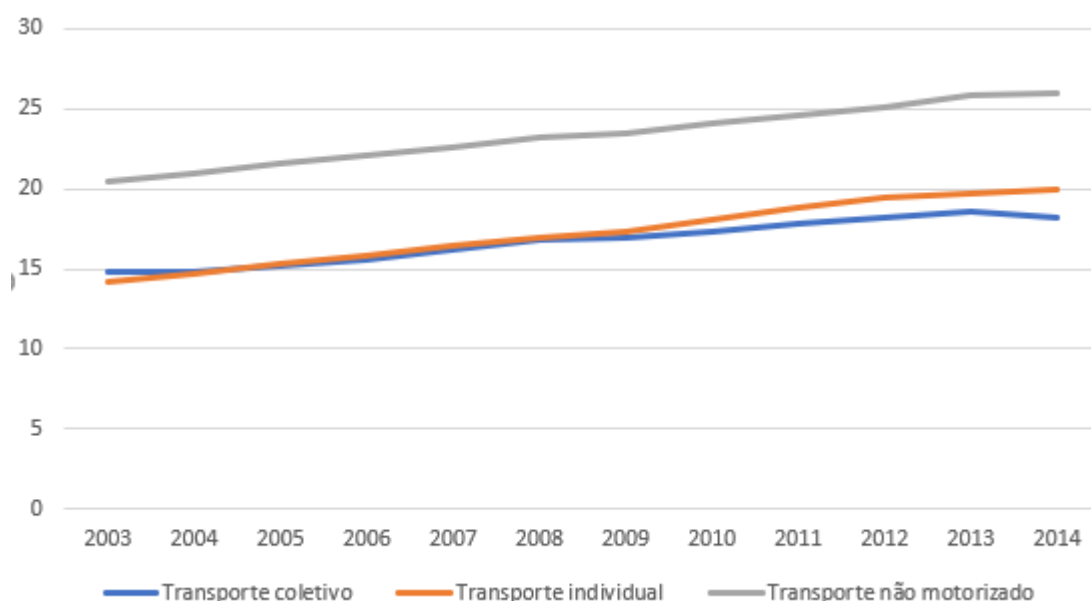
A quantidade de viagens que as pessoas estão fazendo também aumentou nos últimos anos. Importante notar que, apesar do aumento na quantidade em bilhões de viagens por ano, as viagens de transporte individual se tornam mais frequentes que as viagens de transporte público a partir de 2004, como podemos notar na Tabela 3 e Gráfico 4 em bilhões de viagens por ano. Isso evidencia a preferência por esse modo (ANTP, 2016).

Tabela 3: Evolução das viagens por modo

Sistema	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Transporte coletivo	0,46	0,44	0,45	0,45	0,46	0,48	0,47	0,47	0,48	0,49	0,48	0,46
Transporte individual	0,44	0,44	0,45	0,46	0,47	0,48	0,48	0,49	0,51	0,52	0,51	0,51
Transporte não motorizado	0,63	0,63	0,64	0,64	0,64	0,66	0,65	0,66	0,66	0,67	0,67	0,66

Fonte: (ANTP, 2016).

Gráfico 4: Comparativo da evolução das viagens por modo



Fonte: gerado a partir dos dados da ANTP, (2016).

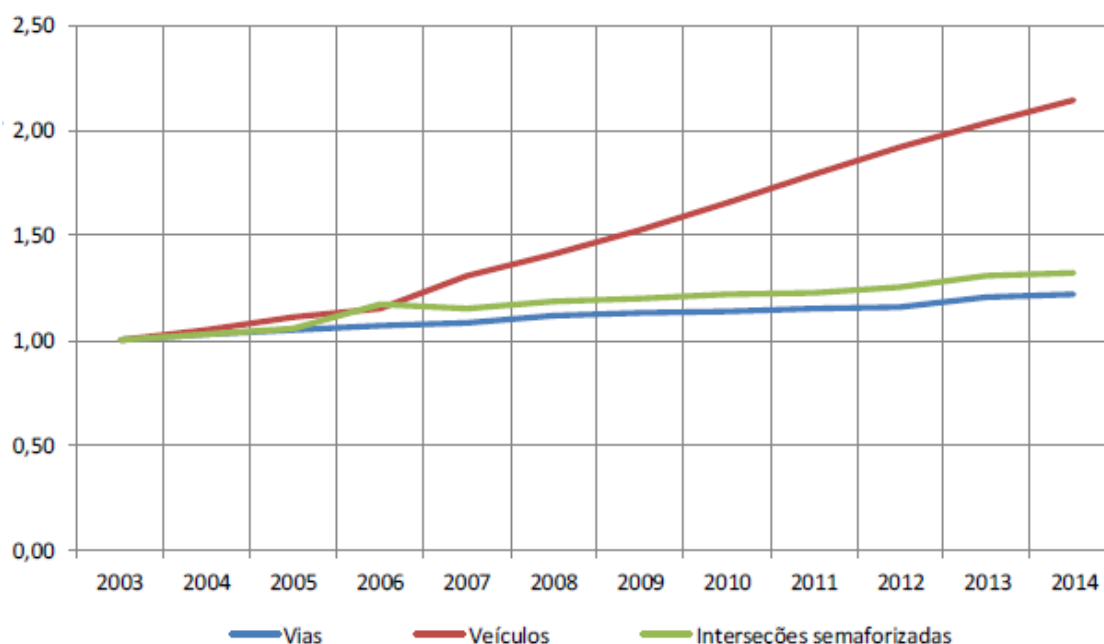
É esperado que a quantidade de veículos no país tenha aumentado, tendo em vista o aumento populacional já comentado. No entanto, esse crescimento foi desproporcional ao aumento de habitantes (Tabela 4). Na quantidade de veículos por mil habitantes, vemos um aumento de cerca de 77% entre 2003 e 2014 (ANTP, 2016).

Tabela 4: Evolução na quantidade dos veículos

Veículos	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Veículos (milhão)	18,1	19,1	20,2	20,9	23,7	25,6	27,7	30	32,5	34,8	37	38,9
Veículos/mil habitantes	168,4	172	179	182	203	214	228	245	263	279	286	298

Fonte: (ANTP, 2016).

Entretanto, o investimento feito nas condições de infraestrutura para suportar a quantidade de novos veículos nas ruas não foi proporcional ao seu crescimento. O Gráfico 5 nos apresenta o comparativo da evolução do índice de equipamentos utilizados na mobilidade urbana, considerando os valores de 2003 como 1,00 (ANTP, 2016).

Gráfico 5: Evolução nos equipamentos usados na mobilidade

Fonte: (ANTP, 2016).

Outros dados importantes também reforçam o cenário insatisfatório da mobilidade urbana no país, como o aumento da distância percorrida pelas pessoas e o aumento no consumo de tempo levado nas viagens realizadas (Tabelas 5 e 6). Independente do modal, notamos que tanto distância, com crescimento de aproximadamente 35% no total, quanto tempo gasto, cerca de 28%, aumentaram entre 2003 e 2014 (ANTP, 2016).

Tabela 5: Distância percorrida pelas pessoas (bilhões de km/ano)

Modo transporte	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Coletivo	187	192	199	208	217	226	230	236	241	248	256	250
Individual	113	116	121	125	130	134	137	143	147	152	154	155
Não motorizado	25	26	27	27	28	29	30	31	32	33	35	35
Total	325	334	347	360	375	390	397	410	420	432	444	440

Fonte: (ANTP, 2016).

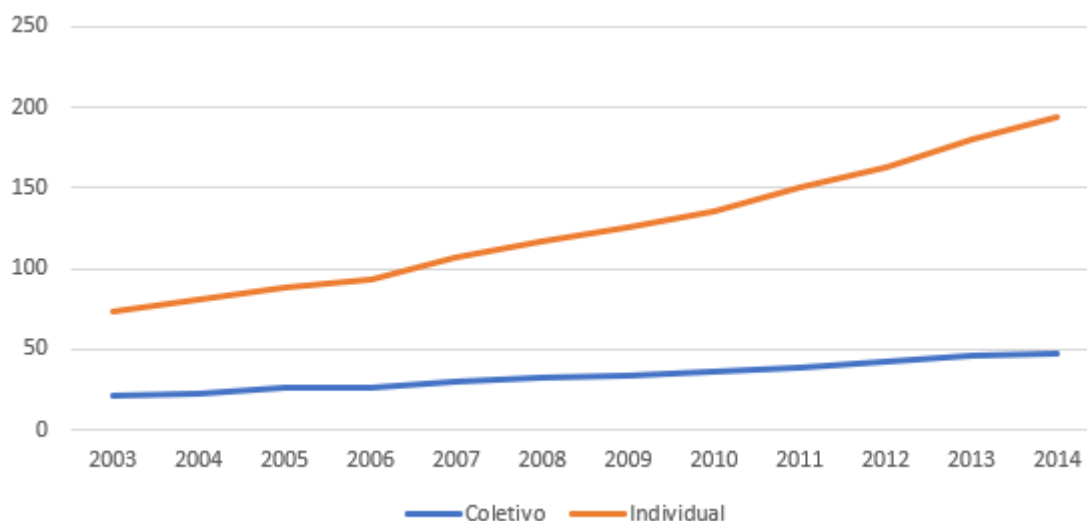
Tabela 6: Consumo de tempo por modo (bilhões de horas/ano)

Modo transporte	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Coletivo	8,7	8,8	9	9,3	9,7	10,1	10,2	10,4	10,7	10,9	11,3	11
Individual	3,7	3,8	4	4,1	4,2	4,4	4,5	4,6	4,8	4,9	5	5
Não motorizado	5,3	5,4	5,6	5,7	5,9	6	6,1	6,3	6,4	6,6	6,8	6,9
Total	17,7	18	18,6	19,1	19,8	20,5	20,8	21,4	21,8	22,4	23,1	22,8

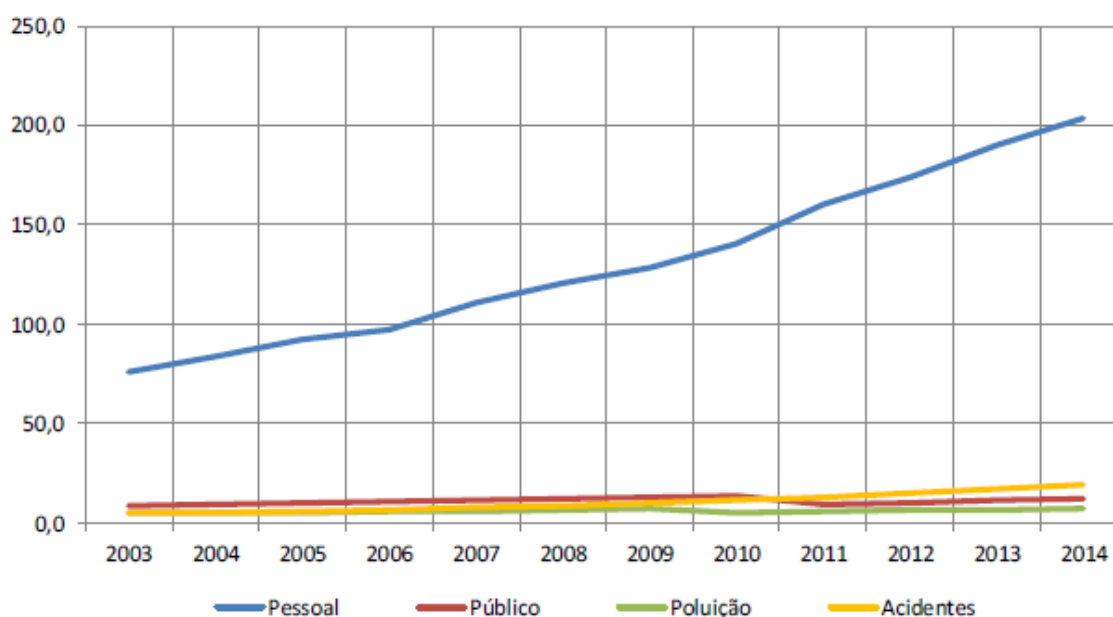
Fonte: (ANTP, 2016)

Os custos também aumentam juntamente com o incentivo ao crescimento dos equipamentos utilizados para a mobilidade. As categorias aqui consideradas pela ANTP nos cálculos são pessoal, público, poluição e acidentes, tanto para transporte individual quanto para coletivo². Novamente destacamos o aumento dos custos totais do transporte individual em relação ao transporte coletivo: apesar do aumento de cerca de 131% entre 2003 e 2014 nos custos do transporte coletivo, os custos do transporte individual aumentaram ainda mais, com mais de 163% no mesmo período (Gráfico 6). Para ambos os tipos de transporte, os custos pessoais são disparados os maiores e que mais aumentaram, contribuindo com 85% do custo total da mobilidade urbana (Gráfico 7) (ANTP, 2016).

² O custo pessoal do transporte individual são recursos gastos pelos usuários do transporte individual, enquanto do transporte coletivo são recursos gastos pelos usuários para utilização do sistema de transporte coletivo. Custo de poluição é baseado em custos por tonelada de cada poluente (CO, HC, NOx, MP, SOx e CO2), média de valores internacionais, adaptados à economia brasileira por meio da PPC – Paridade do Poder de Compra (Banco Mundial). Acidentes são custos de horas perdidas de trabalho, custos médios hospitalares, reparação de veículos e outros (ANTP, 2016).

Gráfico 6: Evolução dos custos por tipo de transporte (bilhões de R\$/ano)

Fonte: Gráfico gerado a partir dos dados da ANTP, (2016)

Gráfico 7: Evolução dos custos por tipo de custo (bilhões de R\$/ano)

Fonte: Gráfico gerado a partir dos dados da ANTP, 2016

Cidade utilizada como case de estudo dessa dissertação, Florianópolis é uma cidade que apresenta população de 421.240 habitantes, segundo o censo do IBGE (2010), e uma densidade demográfica de 627,24 hab/km². Sua taxa de crescimento populacional é sensivelmente mais alta que as taxas do estado de Santa Catarina e do Brasil. Como podemos verificar na Tabela 7, elaborada a partir do censo do IBGE de 2010, desde 1991 até 2010, o Brasil apresenta uma taxa

de crescimento de 30%, Santa Catarina em uma taxa de 38%, e Florianópolis conta com um crescimento de 65%, todos em números aproximados.

Tabela 7: População por ano

Ano	Florianópolis	Santa Catarina	Brasil
1991	255.390	4.541.994	146.825.475
1996	268.720	4.844.212	156.032.944
2000	342.315	5.356.360	169.799.170
2007	396.723	5.866.252	183.987.291
2010	421.240	6.248.436	190.755.799

Fonte: Elaborado a partir dos dados do IBGE Cidades. Disponível em <<https://cidades.ibge.gov.br/>>

Na década de 50, a população da região da Grande Florianópolis apresentava um crescimento ainda reduzido. É a partir da década de 70 que esse panorama começa a se alterar sensivelmente, quando ocorre o desenvolvimento do sistema viário, por meio do surgimento da BR-101, rodovia que possibilita a conexão das cidades circunvizinhas, inclusive às capitais mais próximas, como Curitiba e Porto Alegre (PLAMUS, 2014).

Contribuem para esse crescimento acima da taxa nacional e estadual a migração de áreas rurais e outros centros urbanos, independente da classe social. Assim como ocorre a nível nacional, a dispersão populacional em Florianópolis também segue a característica de afastamento residencial do grande centro, independente da classe social. As pessoas de classe mais baixa se estabelecem em áreas com pior infraestrutura e as de classe mais altas buscam locais com atributos paisagísticos, mas que contam com as vantagens de uma capital (BUENO, 2006).

Adicionado a essa característica populacional, Florianópolis ainda sofre com o aumento populacional no período de verão, considerado alta temporada na cidade: as pessoas se deslocam para a cidade nesses períodos para desfrutar dos seus atributos naturais, como o relevo acidentado e as belas e diversas praias (WEISS, 2016).

A frota de veículos na grande Florianópolis também conta com um aumento considerável: aumentou em cerca de 110% de 2003 a 2013. Isso é um reflexo também do grande aumento da frota no estado, de aproximadamente 123%. Contribuindo significativamente para o problema de mobilidade da região está o aumento da frota de automóveis particulares, contando com 91% de aumento, com destaque para municípios conturbados à capital, como

Palhoça e Biguaçu, com crescimento de mais de 180% e 152%, respectivamente (PLAMUS, 2014).

3.2 A MOBILIDADE URBANA

Além do crescimento populacional, a opção pelo transporte individual e a composição das cidades também influenciam para o problema da mobilidade urbana. A indústria, uma atividade tipicamente urbana, acaba se localizando, de maneira geral, em localidades metropolitanas, principalmente pela necessidade de parceria com estabelecimentos que fornecem componentes entre outras negociações. O mercado atacadista e outras áreas também buscam se fixar nesses locais, visando redução de transporte na redistribuição e redução de custos de comunicação. Logo, o transporte motorizado é importante na manutenção desses mercados, pois seus funcionários, de maneira geral, vivem em áreas residenciais, normalmente distantes das áreas industriais. O mesmo acontece com áreas centrais, onde se concentram o comércio e serviços. Pela distância das áreas residenciais, as pessoas necessitam de transporte motorizado para se deslocarem para essas áreas, tanto para consumo quanto para trabalhar (QUADROS JUNIOR, 2011).

Automóveis e motocicletas são relativamente acessíveis, tanto no Brasil quanto em outros países, principalmente de economia central. Em 1930 já havia indícios de problemas de congestionamento em Nova Iorque, o que mostra uma preocupação antiga com o tema, que não é exclusivo dos tempos atuais. Problema este causado pelo paradigma norte-americano, adotado também pelo Brasil, voltado para o incentivo do uso do transporte individual. Entendido como o melhor meio de transporte possível, este pensamento remodelou a estrutura das cidades para atender às suas necessidades e causou danos ao sistema social (QUADROS JUNIOR, 2011).

Mas o que é a mobilidade? A mobilidade é intrínseca ao ser humano e suas relações sociais. O indivíduo, para se mover, pode depender tanto de meios próprios, como os pés, quanto de equipamentos e da tecnologia, como através do uso de veículos motorizados. Também está condicionado ao espaço onde sua movimentação ocorre, como as cidades, e como isso acontece, sendo esta produção e produto dessas relações interpessoais. Portanto, a mobilidade é a relação que o indivíduo tem com o espaço e as disponibilidades que ele tem disponíveis (BOHUSCH, 2013).

Também está relacionada com a forma de ocupação do território. Dentro das relações interpessoais que ocorrem nas cidades, existe uma ação seletiva do sistema político hegemônico nas configurações de redes de transporte que protege a especulação imobiliária, fazendo com que a circulação de pessoas e bens ocorram de forma desigual. Ao mesmo tempo que promovem integração, incentivam a fragmentação, com o isolamento de alguns lugares, dando uma característica de relações de poder na organização do território para a mobilidade (PEREIRA, 2009).

A concepção rodoviarista de transporte no Brasil foi idealizada visando mais fluidez e velocidade para o transporte, influenciando a característica do espaço geográfico. Fato este que causou degradação do meio ambiente e criação de lugares impróprios para convivência, tornando o veículo, não o cidadão, como o principal elemento do sistema urbano. Logo, quando as liberdades individuais entram em conflito com as liberdades associadas ao automóvel, o problema de mobilidade está instaurado, tendo no congestionamento sua maior manifestação, tornando impossível a prática da liberdade individual nos espaços regionais. Esses espaços se tornam, portanto, campo de disputas territoriais dinâmicas entre os diferentes atores sociais da cidade (BOHUSCH, 2013).

O Código de Trânsito Brasileiro define o trânsito como “a utilização das vias por pessoas, veículos e animais, isolados ou em grupos, conduzidos ou não, para fins de circulação, parada, estacionamento e operação de carga ou descarga. (BRASIL, 1997, art. 1º)”. É importante fazer a distinção entre trânsito, que é a movimentação de veículos e pessoas, e tráfego, que compreende apenas a movimentação de veículos e está associada ao comércio. Portanto, a definição de trânsito é mais abrangente, pois tráfego visa o deslocamento de mercadorias e veículos da forma mais apropriada e ordenada (SILVA, 2001).

Já a mobilidade urbana pode ser compreendida como a capacidade do indivíduo de se locomover de acordo com a infraestrutura de meios de transporte disponível. É a facilidade de deslocamento de pessoas e bens no centro urbano, de viagens no espaço urbano, considerando não somente a ocorrência, mas as possibilidades que a cidade proporciona (ALVES e RAIA JUNIOR, 2009).

Ao fazermos a decomposição do sentido etimológico da expressão mobilidade urbana, temos *mobilis*, que significa “o que pode ser movido, deslocado”, *movere*, “deslocar, colocar em movimento” e *urbs*, que significa “cidade”. Portanto, entendemos mobilidade urbana como algo que se move pela cidade (LORA, 2012). No entanto, trazendo um sentido mais amplo para

o tema, devemos estendê-lo às relações dos indivíduos com o espaço, os objetos, os transportes, entre outros elementos da sociedade, como a organização que visa a otimização dos seus atores nas diferentes mobilidades (ASCHER, 1998).

Teles (2005) explica o conceito mais amplo do assunto:

[...] o conceito de mobilidade é referente ao deslocamento de pessoas, bens, e informação e relaciona locais de trabalho com acessos a bens de consumo, com as residências e espaços de entretenimento e lazer, com as compras e com as relações de amizades. É assim um conceito de integração com a noção de proximidade. Este conceito está associado às questões políticas, às questões ambientais e de sustentabilidade das cidades [...] (TELES, 2005, p. 38)

É importante o enfoque na visão mais ampla da mobilidade urbana, principalmente em uma abordagem política, para que se permita o acesso a todas as pessoas que disputam o espaço urbano. Visando uma democracia real, é essencial que se otimize a mobilidade e acessibilidade de todos os cidadãos. Se existir diferenciação horizontal, diversidade de funções e desigualdade de ofertas, a mobilidade das diferentes classes sociais será condenada, desfavorecendo a cidadania (TELES, 2005).

Os grandes centros urbanos brasileiros sofrem uma crise na história da mobilidade urbana no país. Podemos notar essa crise no crescente aparecimento de transporte coletivo clandestino, na desvinculação das políticas de uso do território e diminuição dos investimentos em transporte coletivo face ao favorecimento do transporte particular (MINISTÉRIO DAS CIDADES, 2004b).

A ampliação da infraestrutura viária favoreceu o uso do veículo particular nos grandes centros do país. No entanto, qualquer tipo de investimento nesse sentido permitirá uma redução do tempo de viagem num primeiro momento, mas atrairá outros usuários que normalmente não optariam por esse tipo de transporte, fazendo com que as vias voltem a ficar congestionadas, exigindo novos investimentos (NOBRE, 2010). Com a formação da cultura do automóvel, nota-se a perda de eficiência do transporte coletivo, com a segregação evidente da população que tem acesso ao transporte individual da minoria que depende do transporte coletivo, limitando a mobilidade desta última (LORA, 2012), apresentando o paradoxo do aumento do custo com transporte urbano quanto menor a renda familiar (MINISTÉRIO DAS CIDADES, 2004a).

O aumento das tarifas de transporte coletivo em níveis mais altos que o da inflação impedem que a população de baixa renda tenha acesso ao deslocamento por esse tipo de

transporte, confirmando a queda na demanda do transporte público. O incentivo ao uso do transporte individual pelo interesse público e a dependência do transporte público pelo mercado financeiro e recursos do governo se tornam um obstáculo para o progresso do transporte coletivo (MINISTÉRIO DAS CIDADES, 2004a).

Esse modelo político voltado para o transporte individual traz uma série de problemas, como congestionamento, redução da velocidade de deslocamento de ônibus, diminuição do uso do transporte público, aumento da emissão de poluentes e ruído, fragmentação das comunidades, aumento nos acidentes de trânsito e necessidade de investimento no sistema viário (LORA, 2012).

3.3 MOBILIDADE URBANA COMO POLÍTICA E PLANEJAMENTO

Sob um novo paradigma de mobilidade urbana no país, o Brasil incorpora um capítulo sobre política urbana na Constituição Federal de 1988. O Conselho das Cidades apresenta as seguintes diretrizes na elaboração dos Planos de Mobilidade:

Art. 8º Nos casos previstos pelo art. 41, § 2º do Estatuto da Cidade, o plano de transporte urbano integrado, ora denominado de Plano Diretor de Transporte e da Mobilidade, deverá contemplar os seguintes princípios e diretrizes gerais: I. garantir a diversidade das modalidades de transporte, respeitando as características das cidades, priorizando o transporte coletivo, que é estruturante, sobre o individual, os modos não-motorizados e valorizando o pedestre; II. garantir que a gestão da Mobilidade Urbana ocorra de modo integrado com o Plano Diretor Municipal; III. respeitar às especificidades locais e regionais; 4 IV - garantir o controle da expansão urbana, a universalização do acesso à cidade, a melhoria da qualidade ambiental, e o controle dos impactos no sistema de mobilidade gerados pela ordenação do uso do solo (CONSELHO DAS CIDADES, 2005, p. 89).

Conta com o Plano Diretor de Transporte e Mobilidade, definido pelo Ministério das Cidades (2007), como um instrumento da política de desenvolvimento urbano que aborda a visão democrática de acesso e oportunidades com condições adequadas à mobilidade tanto da população, quanto da distribuição de bens e serviços. E também com a Política Nacional da Mobilidade Urbana (PNMU), que pode ser entendida como:

A reunião das políticas de transporte e de circulação, integrada com a política de desenvolvimento urbano, com a finalidade de proporcionar o acesso amplo e democrático ao espaço urbano, priorizando os modos de transporte coletivo e os não-motorizados, de forma segura, socialmente inclusiva e sustentável (MINISTÉRIO DAS CIDADES, 2007).

A abordagem do Ministério das Cidades compreende a mobilidade urbana como a apropriação do espaço, promovida pelas políticas públicas, de forma articulada e democrática, sem segregação do espaço (BOHUSCH, 2013). Esses princípios são fundamentados na lei nº 12.587, de 03 de janeiro de 2012 (BRASIL, 2012):

Art. 5º A Política Nacional de Mobilidade Urbana está fundamentada nos seguintes princípios:

- I - acessibilidade universal;
- II - desenvolvimento sustentável das cidades, nas dimensões socioeconômicas e ambientais;
- III - equidade no acesso dos cidadãos ao transporte público coletivo;
- IV - eficiência, eficácia e efetividade na prestação dos serviços de transporte urbano;
- V - gestão democrática e controle social do planejamento e avaliação da Política Nacional de Mobilidade Urbana
- VI - segurança nos deslocamentos das pessoas;
- VII - justa distribuição dos benefícios e ônus decorrentes do uso dos diferentes modos e serviços;
- VIII - equidade no uso do espaço público de circulação, vias e logradouros; e
- IX - eficiência, eficácia e efetividade na circulação urbana.

E para que seja possível sua aplicação, é necessário considerar as suas diretrizes:

Art. 6º A Política Nacional de Mobilidade Urbana é orientada pelas seguintes diretrizes:

- I - integração com a política de desenvolvimento urbano e respectivas políticas setoriais de habitação, saneamento básico, planejamento e gestão do uso do solo no âmbito dos entes federativos;
- II - prioridade dos modos de transportes não motorizados sobre os motorizados e dos serviços de transporte público coletivo sobre o transporte individual motorizado;
- III - integração entre os modos e serviços de transporte urbano;
- IV - mitigação dos custos ambientais, sociais e econômicos dos deslocamentos de pessoas e cargas na cidade;
- V - incentivo ao desenvolvimento científico-tecnológico e ao uso de energias renováveis e menos poluentes;
- VI - priorização de projetos de transporte público coletivo estruturadores do território e indutores do desenvolvimento urbano integrado; e
- VII - integração entre as cidades gêmeas localizadas na faixa fronteira com outros países sobre a linha divisória internacional (BRASIL, 2012).

No entanto, diversas questões aplicadas ao conceito de mobilidade urbana adotados pelo governo ainda não estão claras para grande parte da população (MAGAGNIN e DA SILVA, 2008). Apesar do marco regulatório da lei 12.587/12, a Lei Nacional de Mobilidade Urbana (LNMU), estabelecer a obrigatoriedade do Plano de Mobilidade Urbana para municípios com mais de 20 mil habitantes, os planos não são feitos ou não atendem aos requisitos estabelecidos. Apenas 3,8% dos recursos destinados a projetos e obras de transporte nos últimos anos foram utilizados. O planejamento dos transportes, apesar de sustentação na legislação, não pode ser desenvolvido de forma dissociada do planejamento urbano, sob pena de danos às condições de

acessibilidade e padrão de mobilidade contrário à sustentabilidade (MELLO e PORTUGAL, 2017).

Portanto está em curso uma mudança de paradigma para uma mobilidade mais sustentável, considerando seus aspectos qualitativos, com transportes que se preocupam com uma produtividade social. A combinação adequada entre elementos de transporte e o uso do solo permite que se alcance uma mobilidade urbana comprometida com o pensamento sustentável (MELLO e PORTUGAL, 2017).

A qualidade de vida está diretamente ligada com a forma que se faz o planejamento urbano, e, conseqüentemente, associada de forma intrínseca a aspectos de transporte e como estes estão disponíveis para a população. O crescimento urbano desordenado, com aumento das áreas urbanas e uso excessivo de automóveis, falta de infraestrutura urbana, poluição, entre outras questões, impactam fortemente a qualidade de vida. Porém, além dos conceitos de planejamento de transportes associado ao uso do solo, questões sociais e ambientais, o conceito da sustentabilidade também está sendo adotado pelo planejamento de transportes das cidades (MAGAGNIN e DA SILVA, 2008). Ele pode ser entendido como o equilíbrio entre as dimensões ambientais, econômicas e sociais, inclusive nos planejamentos futuros, com eficácia econômica, proteção do meio ambiente e luta contra desigualdades sociais (STEG e GIFFORD, 2005).

Para alcançar o que se espera para mobilidade sustentável, cinco atributos para mobilidade devem ser levados em consideração (MELLO e PORTUGAL, 2017):

- Produtiva: abarcando a eficiência energética e capacidade ótima de infraestrutura de transportes, priorizando modalidades coletivas e diminuam a segregação;
- Inclusiva: pensando nas pessoas que possuem necessidades especiais e restrições físicas, sob uma perspectiva inclusiva;
- Segura: que permita a manutenção da integridade, evitando o risco de acidentes;
- Justa socialmente: com qualidade no serviço, permitindo que outras atividades fundamentais do cotidiano possam ser realizadas, visando o bem-estar social;
- Verde: com transportes amigáveis ambientalmente, focando na redução de emissão de poluentes e utilização de energia limpa.

4 DIAGNÓSTICO DA OPORTUNIDADE IDENTIFICADA

Neste ponto, introduzimos a tendência do conceito de *smart city* e sua relação com a mobilidade urbana, na ideia de mobilidade inteligente. A disseminação das tecnologias da informação e comunicação e desenvolvimentos de tecnologias inovadoras servem como estímulo para a implementação e crescimento das *smart cities*. Portanto, descrevemos algumas características dessas inovações e trabalhos relacionados ao tema.

4.1 SMART CITY E A MOBILIDADE URBANA INTELIGENTE

Mobilidade urbana sustentável também pode ser entendida como “a forma fácil, conveniente, disponível e acessível de viajar para seu destino com mínimo impacto ao meio ambiente e aos outros” (LAM e HEAD, 2012, p. 359). Acessibilidade e conveniência podem ser alcançados “com bom design urbano, mudança de comportamento, tecnologia avançada, políticas de suporte, incentivos econômicos e engajamento e liderança pela cidade” (LAM e HEAD, 2012, p. 359). Portanto, a tecnologia juntamente com o planejamento e engajamento das cidades, unidos de forma inteligente, são de grande importância.

Estamos na era do *smart*, principalmente pela quantidade de dispositivos inteligentes e com tecnologia avançada à disposição. De acordo com o Oxford English Dictionaries, *smart* pode ser entendido como “possuindo ou apresentando uma penetrante inteligência” ou, no caso de um dispositivo, como “assumindo ação independente” (LYONS, 2016, p. 3). Portanto, mobilidade urbana *smart* pode ser definida da seguinte forma (LYONS, 2016, p. 4):

- O uso da tecnologia para gerar e compartilhar dados, informação e conhecimento que influenciam decisões;
- O uso da tecnologia para aprimorar veículos, infraestrutura e serviços; e
- A derivação de aprimoramentos para os operadores de sistemas de transporte e usuários e para acionistas.

Em 2006, o Government’s Foresight Programme examina o futuro a longo prazo da mobilidade urbana *smart*, apesar de ainda não considerar o termo *smart* nas suas ideias, mas

intelligence (LYONS, 2016). É importante destacar a relevância dada à criação de informação que pode suportar a tomada de decisão (OST, 2006, p. 10):

- **Design inteligente**, minimizando a necessidade de se mover através de design urbano, integração eficiente e gestão do transporte público e produção local;
- Um sistema que pode **promover inteligência**, com sensores e mineração de dados, provendo informação para suportar decisões dos indivíduos e provedores de serviço;
- **Infraestrutura inteligente**, que processa a massa de informação coletada e adapta ao mundo real para prover os mais eficientes serviços; e
- **Uso inteligente** do sistema, onde as pessoas modificam seu comportamento para utilizar a infraestrutura de forma sustentável.

Sugerindo uma abordagem para além da ideia de se mover livremente e com facilidade, Lyons (2016) incorpora a capacidade de movimento de materiais e de informação, propondo a mobilidade urbana *smart* como “conectividade nas cidades que é acessível, eficaz, atrativa e sustentável” (LYONS, 2016, p. 1), salientando que o termo deve ser considerado em diferentes níveis e com efeitos combinados entre os diferentes atributos.

O autor explica os termos de sua proposição: Conectividade que considera a distância a ser percorrida, não somente formas de percorrer grandes distâncias com veículo motorizado, onde a conectividade digital se apresenta como uma alternativa; Acessível e Eficaz, respeitando as diferentes necessidades e possibilidades das pessoas, além do que está disponível para os provedores dos serviços de transporte; Atraente para todos os tipos de pessoas em seus diferentes papéis na cidade, nenhum deles pode ser impactado direta nem indiretamente; e finalmente Sustentável, observando que todos os atributos anteriores devem ser realizados considerando a sustentabilidade (LYONS, 2016).

Para se alcançar uma mobilidade urbana *smart* é necessário ponderarmos quanto à forma que nos conectamos, como suportar e manter as necessidades de conexão das pessoas, considerando o design urbano e a maturidade dos sistemas de comunicação. Deve-se minimizar o que é externo à conexão, ou seja, menos veículos e, conseqüentemente, menos espaço utilizado por veículos, e incentivar o suporte à conectividade, numa união entre a tecnologia e equipamentos motorizados, tornando a mobilidade mais atrativa, eficaz e acessível (LYONS, 2016).

Papa et al. (2017) expõem algumas principais definições sobre mobilidade urbana *smart*, como uma mobilidade eficiente e eficaz, independentemente do uso de tecnologias, e que enfatizam o papel fundamental das tecnologias inovadoras, no papel de desenvolvimento e disseminação de meios de transporte sustentável. Apesar de salientarem que as TICs (Tecnologia da Informação e Comunicação) são apenas mais um elemento e que a participação e conscientização dos cidadãos é fundamental na implementação de uma mobilidade urbana sustentável, notamos o grande protagonismo que as tecnologias ganharam para o desenvolvimento da mobilidade urbana *smart*.

Com a difusão das TICs e seu caráter pervasivo, pode-se monitorar tendências nos subsistemas urbanos e como se interconectam. Entendendo a cidade como um sistema complexo, Fistola et al. (2017) discursam sobre três principais subsistemas de uma cidade, dos quais a mobilidade é uma atividade especial: físico, com o componente material; funcional, abrangendo as atividades do contexto urbano; e sócio-antrópico, das relações sociais e com o espaço. Uma cidade *smart* pode fornecer continuamente uma grande quantidade de dados sobre seus subsistemas que precisam ser interpretados corretamente para que esta possa ser gerenciada sob aspectos sustentáveis. Neste contexto, temos grande quantidade de dados (*Big Data*) gerados por uma *smart city*, ou cidade inteligente, sendo analisados e utilizados para auxiliar na tomada de decisão (*Big Data Analytics*).

A concentração nas cidades pode gerar efeitos diversos para a humanidade, tanto positivos quanto negativos. Da mesma forma que vemos o desenvolvimento tecnológico, empregabilidade, aumento do nível cultural e melhoria nas condições econômicas, vemos o desgaste socioambiental nas emissões de gases tóxicos, despejamento de lixo, trânsito, entre outros. O conceito de *smart growth*, ou o crescimento inteligente, começa a aparecer nos anos noventa, pois se vê necessária a ideia de *smart city*, ou cidade inteligente, para ajustar esse crescimento urbano, direcionando-o para uma visão sustentável, preocupada com a vida humana na Terra (COCCHIA, 2014).

Com a elaboração do Protocolo de Kyoto, para atingir seus objetivos, nasce o conceito de *smart city*, com projetos relacionados a sustentabilidade nas cidades. Esse conceito, portanto, está relacionado com iniciativas sustentáveis da cidade, aprimoramento do crescimento verde e qualidade de vida, o uso de infraestruturas de TIC, o envolvimento dos cidadãos na esfera pública, entre outros (COCCHIA, 2014).

O termo *digital city* ainda se confunde com *smart city* – ambos nasceram em meados dos anos 90 – com relação a sua definição e para cada um dos termos, individualmente, ainda se observa pouca conformidade. *Digital city* mantém certo padrão, pois está diretamente relacionado com a presença da TIC na melhoria da cidade. A utilização do termo teve picos nos anos 2000, mas se manteve estável durante a maior parte da sua vida. Já *smart city* possui um viés mais abrangente, abordando questões sociais e ambientais como objetivos da sua implementação. Seu desenvolvimento foi devagar até 2010, quando a União Europeia adota a ideia para seus desenvolvimentos em cidades, fazendo com que o uso desse termo aumentasse muito (COCCHIA, 2014).

As iniciativas tanto de *digital* e *smart city*, originalmente, foram iniciadas de forma *bottom-up*, ou seja, de pequenas ferramentas desenvolvidas pelos cidadãos para solucionar ou estudar problemas urbanos, dando uma característica empírica para ambos. Portanto, como soma de diversas iniciativas individuais, nota-se pouco planejamento estratégico nos desenvolvimentos e implementações, observado apenas nos trabalhos mais recentes. A tecnologia é fator fundamental em ambos, apresentando soluções de compartilhamento de informação, comunicação e envolvimento de cidadãos em projetos, guarda ambiental, redução de poluição e qualidade de infraestrutura (COCCHIA, 2014).

Apesar do significado inovador e transformador do termo *smart*, a questão social é fator central nesse conceito, exigindo um pensamento sociotécnico, com entendimento de toda a complexidade na inter-relação dos ambientes sociais e tecnológicos para se desenvolver uma *smart city* (NAM e PARDO, 2011). Portanto, o objetivo é criar ambientes *smart*, que se adaptem às necessidades dos usuários, com uma arquitetura de camadas, combinando diferentes ambientes numa estrutura complexa (MARSA-MAESTRE, LOPEZ-CARMONA, *et al.*, 2008). Giffinger e Gudrun (2010), em síntese, apresentam seis características relevantes para o conceito *smart city*: economia, pessoas, governança, mobilidade, ambiente e vivência.

Uma *smart city* utiliza tecnologias computacionais inteligentes que tornam a cidade mais interconectada, eficiente (WASHBURN, SINDHU, *et al.*, 2010), instrumentada e inteligente (HARRISON, ECKMAN, *et al.*, 2010), integra as condições de sua infraestrutura, otimiza seus recursos, planeja suas manutenções e maximiza os serviços para os cidadãos (HALL, 2000). É também uma cidade que usa a TIC para fortalecer a liberdade de expressão e acessibilidade ao serviço e informação pública (PARTRIDGE, 2004).

Portanto, Lyons (2016) sintetiza o conceito de *smart city* como:

[...] o reflexo da complexidade dos sistemas da cidade, o conjunto de diferentes atores e comentaristas que buscam articular interpretações contra prioridades e objetivos opostos, e a heterogeneidade pelas cidades em termos de escala, design urbano, população e função (LYONS, 2016, p. 3).

4.2 TECNOLOGIAS INOVADORAS DA *SMART CITY*

Violato et al. (2012) explicam a importância da TIC por constituírem a base de todo desenvolvimento das atividades da sociedade moderna. Atualmente, as TICs são desenvolvidas considerando as necessidades dos usuários, que demandam constantemente soluções por esses meios, seu ciclo está cada vez mais curto, principalmente no que se refere aos aparelhos celulares, e possibilitam a comunicação não apenas com humanos envolvidos, mas entre máquinas. Sob essa perspectiva, os autores vêm três vetores de desenvolvimento: i) mobilidade, um dos fatores mais determinantes no ritmo de demanda de inovação tecnológica; ii) capacidade, pela necessidade cada vez maior de banda larga transparente; e iii) tecnologias centradas nos usuários, orientadas na demanda por segurança e usabilidade por parte dos usuários.

Para suportar as *smart cities* na esfera da TIC, tecnologias inovadoras foram desenvolvidas. Neste cenário, Big Data, Big Data Analytics, Cloud Computing, IoT (Internet of Things) e Crowdsensing são conceitos atuais que auxiliam nos planejamentos e implementações de *smart cities*.

4.2.1 *Big Data*

Com o surgimento de diversas dificuldades geradas pela massificação de dados coletados, pela evolução da TI, pela exigência de escalabilidade, heterogeneidade, tempo real e privacidade de dados, o Big Data emerge como solução. A forma tradicional de gerenciar dados é feita através do banco de dados relacional, que são aplicados em dados estruturados e acabam utilizando cada vez mais espaço de disco rígido. Logo, estes têm muita dificuldade em realizar esse gerenciamento de forma eficiente, com elasticidade, com fácil atualização, armazenamento permanente e com dados desordenados. (CHEN, MAO e LIU, 2014).

Diversas definições de Big Data foram elaboradas na tentativa de diferenciar o conceito da simples armazenagem de grande quantidade de dados. A maioria dos acadêmicos, inicialmente com Doug Laney em 2001, analista da META – Gartner atualmente, enfatizam os três Vs para caracterizar o Big Data: **volume**, referenciando a larga quantidade de dados, **velocidade**, de coleta, processamento e análise, e **variedade**, dados de diferentes tipos e fontes (CHEN, MAO e LIU, 2014). Além disso, o crescimento contínuo de quantidade de dados que são armazenados no auxílio em tomadas de decisão e questões estratégicas também é destacado (MIKALEF, PAPPAS, *et al.*, 2017).

Gantz e Reinsel (2011) incluem a característica de **valor**, com relação aos benefícios que o Big Data oferece quanto a informação e capacidade estratégica. Algumas definições também incluem **veracidade**, representando a confiabilidade dos dados (AKTER, WAMBA, *et al.*, 2016). Mais recentemente, estão sendo consideradas a **variabilidade**, considerando as diferentes interpretações da informação, e a **visualização**, quanto à interpretação dos padrões e tendências das informações (SEDDON e CURRIE, 2017).

Big data envolve armazenamento de dados, gerenciamento, análise e visualização de grande quantidade, variedade e complexidade de dados (RUSSOM, 2011), que vão além da capacidade de processamento de *softwares* e ferramentas comuns (BHARADWAJ, EL SAWY, *et al.*, 2013), em tempo tolerável (CHEN, MAO e LIU, 2014), visando a possibilidade e agilidade de negócios, permitindo ações automáticas em tempo real e tomadas de decisão (WHITE, 2011), e a solução de problemas sociais (KAMIOKA e TAPANAINEN, 2014). Sua cadeia de valores pode ser dividida em quatro fases: geração de dados, aquisição de dados, armazenamento de dados e análise de dados (CHEN, MAO e LIU, 2014).

O crescimento de dados na era do Big Data traz desafios que este deve enfrentar, tais como a representação de dados, redundância e compressão, gerenciamento do tempo de vida, mecanismos analíticos, confidencialidade, gerenciamento de energia, escalabilidade e cooperação. Com programas complexos e de larga escala, o Big Data é compreendido como uma tecnologia analítica avançada, permitindo evoluções de aplicações em diversos campos, como comerciais, de rede e científicas, e análise de dados estruturada, com análise de dados de texto, *web*, multimídia, de rede e de dispositivos móveis (CHEN, MAO e LIU, 2014).

Existem alguns obstáculos que devem ser superados para o desenvolvimento de um Big Data, como a representação dos dados, para que sejam mais significativos para análise e

interpretação; diminuição de redundância e compressão, para reduzir o custo; gerenciamento do ciclo de vida, sobre o que manter ou descartar; mecanismo analítico, análise em curto espaço de tempo; confidencialidade, para assegurar segurança; gerenciamento de energia; escalabilidade, quantidade de dados cada vez maior e mais complexos; e cooperação, sob um viés interdisciplinar (CHEN, MAO e LIU, 2014).

Atualmente, os dados armazenados em Big Data são coletados através de sensores de previsão do tempo, mídias sociais, imagens e vídeos, registros transacionais, sinais de GPS, entre outros. Para conseguir gerenciar e analisar todos esses tipos e quantidades de dados, diversos *frameworks* ferramentais de análise de dados foram propostos. Eles são usados para dimensionar processamentos paralelos que são distribuídos em diferentes máquinas. Esse tipo de arranjo é fundamental para a rápida análise de dados, principalmente face o seu constante crescimento (ELARABI, SHARMA e PAHWA, 2016).

Neste contexto, é importante mencionar a cadeia de valor do Big Data apresentada por Chen et al. (2014), que considera a geração e aquisição de dados como fases de exploração do dado, e o armazenamento e a análise, este último como um processo de criação de valor:

- Geração: dados de diversas fontes, principalmente de operações empresariais, logística e sensoriamento por IoT, interações interpessoais e localização via internet, e via pesquisas científicas;
- Aquisição: contempla a coleta de dados das diversas fontes e meios, o transporte, entre as camadas e infraestruturas computacionais de armazenamento, e o pré-processamento, para que o dado possa ser analisado. O pré-processamento compreende técnicas como a integração de dados, com base no processo de *data warehouse* ETL (Extract, Transform and Load), que conecta as diferentes fontes de dados, padroniza e armazena, além da limpeza de dados desnecessários ou prejudiciais e eliminação de redundâncias.
- Armazenamento: serviço de armazenamento confiável e que possa ser acessado facilmente. Deve ser consistente, ter disponibilidade e permitir particionamento;
- Análise: uso adequando de métodos estatísticos para analisar grande quantidade de dados, extraindo e refinando dados que estão perdidos num conjunto desestruturado, identificando o assunto abordado e maximizando o valor do dado analisado.

Para que se possa processar e analisar com a enorme quantidade de dados de forma independente e simultânea, é necessário que se organize diversos computadores em paralelo. O

Hadoop é um sistema que realiza o processamento de Big Data. É um *software open source* desenvolvido em Java que consiste em duas partes: o Hadoop Distributed File System (HDFS), que armazena os dados, e o MapReduce, que realiza o processamento em paralelo entre as diferentes máquinas organizadas na arquitetura de computadores (RAJARAMAN, 2016).

Por mais pesquisado e estudado que seja o tema Big Data, vários problemas ainda estão em aberto e devem ser discutidos, como a falta de uma definição holística, de um modelo estrutural, de um sistema teórico de *Data Science* mais robusto e de uma padronização de um sistema de avaliação (CHEN, MAO e LIU, 2014). A segurança de dados é outro problema, que concerne à privacidade, qualidade e mecanismos de segurança. Por conta dos avanços tecnológicos e potenciais computacionais, as pessoas estão cada vez mais preocupadas com a maneira com que seus dados são utilizados pelas organizações e seus propósitos comerciais, o que exige maior regulamentação para aumentar a proteção dos indivíduos (WATSON, 2014).

Apesar de estar em um cenário muito inicial e dos diversos problemas técnicos, o Big Data influenciará a vida das pessoas e sua maneira de pensar, abrindo oportunidades para o trabalho com dados em larga escala, performance de recursos de dados, fusão com a ciência, visualização de dados, cenário orientado a dados e uma revolução do pensamento, principalmente na tomada de decisão (CHEN, et al., 2014).

4.2.2 *Big data analytics*

Buscando um foco no caráter analítico do Big Data e sem fazer conexão direta com uma finalidade utilitarista organizacional ou de instituições, os termos *Data Science* e *big data analytics* ganham projeção. Esses termos ressaltam a extensa quantidade e variedade de dados em frequente alteração que precisam ser analisados por novas ferramentas e recursos computacionais poderosos para que se possa extrair informações de valor (RAJARAMAN, 2016).

Apesar de aparecer na literatura de ciências da computação em 1960, o termo *Data Science* começa a emergir pela estatística e mineração de dados apenas no final de 1990 (THE B/A/H REPORT, 2015). Rajaraman explica que *big data analytics* é um componente de *Data Science*, e define o último como:

[...] um campo de estudo preocupado com a coleta, limpeza e anonimização (se necessário) de larga quantidade de dados de variedade diversa relevante para solução de problemas da vida real e sua análise para principiar ações significativas (RAJARAMAN, 2016, p. 698).

No avanço da ciência computacional, *Data Science* emerge como responsável pelo quarto paradigma científico³, que é a descoberta de novas hipóteses, baseando-se em dados, que podem permitir a formulação de novas teorias (RAJARAMAN, 2016).

O termo *analytics* foi inicialmente abordado nos anos 1970 nos conceitos de sistemas de suporte a decisão. Mais tarde, nos anos de 1990, Howard Dresner, um analista do Gartner, popularizou o termo BI (Business Intelligence), uma categoria de aplicações que analisa dados para auxiliar nas tomadas de decisões de usuários de negócio, se tornando referência para todos os tipos de aplicação utilizados para tomadas de decisão. Provavelmente o *big data analytics* evoluiu do BI, e hoje é interpretado como o termo que referencia aplicações de análise de dados (WATSON, 2014).

Big data analytics enfatiza as ferramentas e processos utilizados para se extrair informações de um Big Data e as formas de visualização e apresentação desses dados (MIKALEF, PAPPAS, *et al.*, 2017). Lamba e Dubey o definem como “a aplicação de múltiplos métodos analíticos que endereça a diversidade do Big Data para prover resultados descritivos, preditivos e prescritivos factíveis” (LAMBDA e DUBEY, 2015, p. 5), e Muller et al., como “o modelamento estatístico de conjunto de dados largo, diverso e dinâmico com conteúdo gerado por usuários e rastreo digital” (MULLER, JUNGLAS, *et al.*, 2016, p. 1).

Na literatura, são discutidos quatro tipos de *data analytics*:

- Descritivo: como um dashboard, esse tipo de análise mostra o que ocorreu no passado de forma clara, auxiliando no entendimento do que significam os dados visualizados;
- Preditivo: diz o que é esperado que aconteça num futuro próximo, usando métodos estatísticos, redes neurais e algoritmos de *machine learning*;
- Exploratório: encontra relações inesperadas entre os dados coletados, possibilitando descobertas de novos conhecimentos (WATSON, 2014);

³O quarto paradigma neste contexto é interpretado como um *framework* filosófico/teórico. Este sucedeu os paradigmas da ciência teórica, experimental e de simulação, nesta ordem (RAJARAMAN, 2016).

- Prescritivo: sugere ações a tomar baseando-se nos dados, otimizando a forma de solucionar problemas (RAJARAMAN, 2016).

Alguns autores aprofundam o valor que Big Data e *big data analytics* trazem para as organizações e quais meios devem ser considerados para se alcançar seus benefícios. Para tanto, uma agência governamental deve se fundamentar nas seguintes bases: substituir ou auxiliar a tomada de decisão do ser humano pelas ferramentas; reduzir as ineficiências; instituir transparência; aprimorar a performance; melhorar o retorno sobre o investimento de TI; prover capacidades preditivas para aprimorar resultados; reduzir a criminalidade e ameaças à segurança; eliminar desperdício, fraude e abuso; e inovar nos modelos de negócio e serviços (TECHAMERICA FOUNDATION, 2012).

Uma empresa, por sua vez, deve definir claramente sua necessidade de negócio, com objetivos estreitamente definidos – principalmente orientados ao cliente; ter um patrocinador comprometido; alinhar o negócio com a estratégia da análise; possuir uma cultura de tomada de decisão baseada em fatos, experimentando cenários; ter uma sólida infraestrutura de dados; ferramentas de análise apropriadas; e pessoas capacitadas (WATSON, 2014).

A privacidade é um ponto a ser considerado pelas instituições ao se trabalhar com *big data analytics*, principalmente com relação à publicação de informações. Alguns países possuem leis que obrigam processos de anonimização para os cidadãos, mas nenhuma ferramenta desenvolvida para este fim é infalível. Portanto, os indivíduos devem ficar alerta, pois a negligência quanto à relevância e importância dos dados utilizados pode levar ao uso indevido dos resultados das análises (RAJARAMAN, 2016).

4.2.3 IoT e Cloud Computing

O Cloud Computing é uma tecnologia fortemente relacionada com Big Data, que, a partir de uma operação computacional intensiva, estressa a utilização do sistema Cloud, cujo objetivo é possibilitar o gerenciamento de grande quantidade de processamento e armazenamento de dados e o desenvolvimento de aplicações refinadas de um Big Data, através de tecnologia distribuída de armazenamento e computação paralela. Isso faz com que o desenvolvimento de ambos caminhe conjuntamente. Enquanto o Big Data é usado para fins

estratégicos, auxiliando na tomada de decisão, o Cloud Computing fornece recursos a nível sistêmico e promove uma transformação da arquitetura de TI, assumindo papel fundamental para a boa implementação e utilização de um Big Data (CHEN, MAO e LIU, 2014).

Gubbi et al. (2013) afirmam que, para se desenvolver uma *smart city*, é necessário que se faça a integração de IoT e Cloud Computing, com a combinação de múltiplos serviços, escalabilidade e inteligência de gerenciamento e operação complexa. A IoT foi criada a partir da proliferação de sensores ubíquos das tecnologias WSN (Wireless Sensor Network), com a habilidade de medir, inferir e compreender indicadores do meio ambiente. Impulsionada pelas tecnologias *wireless* de RFID (Radio Frequency IDentification) e sensores embutidos, a IoT está transformando a internet em um ambiente altamente integrado, com mecanismos de busca de dados sofisticados que requerem resposta rápida. Esta tecnologia conta com três componentes: Hardware, composto por sensores embutidos e aparelhos de comunicação; Middleware, com armazenamento sob demanda e ferramentas de análise de dados; e Apresentação, com ferramentas modernas de visualização e interpretação.

De forma geral, os autores concluem que:

A proliferação de dispositivos com capacidades de comunicação-atuação está trazendo para mais perto a visão da IoT, onde as funções de monitoramento e atuação se misturam perfeitamente no pano de fundo e novas capacidades são possibilitadas pelo acesso de novas fontes ricas de informação. A evolução da nova geração do sistema de dispositivo móvel depende da criatividade dos usuários em desenvolverem novas aplicações. IoT é uma tecnologia emergente ideal para influenciar esse ambiente fornecendo novos dados desenvolvidos e recursos computacionais necessários para criar aplicativos revolucionários (GUBBI, BUYYA, *et al.*, 2013, p. 1658).

Dados coletados por IoT diversificam e enriquecem uma plataforma de Big Data, pois costumam possuir perfis bem distintos e heterogêneos, apesar da característica não estruturada e alta redundância. Seu sucesso está diretamente associado à integração bem-sucedida entre Big Data e Cloud Computing (CHEN, MAO e LIU, 2014).

4.2.4 Crowdsensing

Assumindo papel importante na mobilidade inteligente, um dos componentes críticos da IoT são os telefones móveis, mais propriamente os *smart phones*. Estes dispositivos se

desenvolveram bastante para comunicação e computação, se tornando parte da vida das pessoas na cidade. Seus sensores, como GPS, acelerômetro, conexão de internet e *wifi*, câmera, entre outros, são funcionalidades interessantes para o auxílio na mobilidade (SHENG, TANG, *et al.*, 2013). Nas campanhas de *crowdsensing* eles são fundamentais. Em um cenário de sensoriamento, por exemplo, estão conectados a um sistema *cloud* em uma ponta e a vários sensores na outra (GUBBI, BUYYA, *et al.*, 2013).

O *crowdsensing*, ou o sensoriamento de pessoas em massa, se baseia na coleta de dados de pessoas que estão participando de uma campanha, que estão envolvidas num fenômeno de interesse. O *crowdsensing* pode ser dividido em dois tipos: participativo ou oportunista. Ganti *et al.* (2011) explicam esses conceitos da seguinte forma:

O *sensing* participativo requer o envolvimento ativo do indivíduo para contribuir com o dado sensoriado (ex., tirar uma foto, reportar um fechamento de via) relativo a um fenômeno de larga escala. O *sensing* oportunista é mais autônomo, e o envolvimento do usuário é mínimo (ex., amostra de local contínua sem ação explícita do usuário) (GANTI, *et al.*, p. 32, 2011).

O *sensing* participativo é a utilização cada vez mais intensa, por parte de indivíduos e comunidades, de dispositivos móveis e serviços *cloud*, para coletar e analisar sistematicamente dados. Esse cenário está cada vez mais facilitado pela aproximação da tecnologia e da inovação analítica com o crescente conforto que as pessoas sentem quanto ao uso de dispositivos móveis e mídias sociais (ESTRIN, 2010). O seu custo de implementação é muito baixo, pois faz uso de dispositivos já sendo utilizados, permite uma larga cobertura de espaço/tempo e observação de eventos, conta com o aumento do mercado de desenvolvimento e distribuição de aplicativos para *smart phones*, e pode melhorar a vida das pessoas através da intrínseca inter-relação que essas vivenciam no processo. Por esses motivos, oferece uma série de vantagens sobre os processos de sensoriamento tradicionais (CHRISTIN, REINHARDT, *et al.*, 2011).

Por perseguir um viés paradigmático comunitário, Ganti *et al.* (2011) cunham o termo *mobile crowdsensing* (MCS), afirmando que “o *sensing* comunitário abrange um largo espectro de envolvimento do usuário, com o *sensing* participativo e oportunístico nas duas pontas” (GANTI, YE e LEI, 2011, p. 32). O *mobile crowdsensing* emergiu como um novo paradigma por conta da proliferação dos dispositivos móveis e sensores embutidos (JIANG, GAO, *et al.*, 2017), possibilitando aos usuários coleta, processamento e compartilhamento de dados sensorizados, e suportando aplicações de sensoriamento de larga escala (NI, LIN, *et al.*, 2017).

Há três categorias neste tipo de aplicação: ambiental, onde os fenômenos se limitam ao ambiente natural, tais como medição de poluição de uma cidade, nível de água num afluente, monitoramento de vida selvagem; infraestrutural, com mensuração de fenômenos de larga escala relacionados com infraestrutura pública, tais como congestionamentos, condições de rodovias, disponibilidade de estacionamento; e social, onde os indivíduos compartilham informações sensorizadas entre si, como dados de exercício físico (GANTI, YE e LEI, 2011).

Os sensores e os próprios dispositivos móveis permitem uma consciência contextual muito sofisticada. Através dessas tecnologias, tem se dado muita importância em como as pessoas são capazes de enviar informações do seu dia-a-dia relativas à meio ambiente, comunidade e saúde. Além de dispensar a instalação de sensores de grande custo pela cidade, o motivo de se incluir os cidadãos no processo de coleta de dados por sensoriamento está na precisão e acuracidade desses dados. O que faz com que a participação cidadã seja um requisito fundamental na implantação do conceito de *smart city* (ALKHELAIWI e GRIGORAS, 2016).

O avanço da tecnologia nas cidades busca, entre outras coisas, deixar a vida das pessoas mais conveniente. Porém, desenvolver um conceito de *smart city* é custoso em termos de recurso, investimento e tempo. Mesmo adicionando aspectos simples desse conceito em uma estrutura de transporte público – que é um dos meios de transporte mais utilizados nas cidades modernas – requer uma infraestrutura de acompanhamento de veículo cara. Portanto, a utilização do *crowdsensing* nessas situações se apresenta como uma ótima alternativa.

No entanto, não é simples sensibilizar os cidadãos a utilizarem e colaborarem em uma campanha de *crowdsensing*. Para que ele se disponibilize a baixar um aplicativo no seu celular pessoal, ativar as configurações necessárias para a coleta – como GPS, internet e outras funcionalidades – e participar ativa ou passivamente do processo no seu dia-a-dia, o sistema deve fornecer algum retorno para o usuário. Sua participação é crucial para o sistema de *crowdsensing*, logo, algum mecanismo de incentivo à participação deve ser desenvolvido para que haja o engajamento e que seja valioso para os envolvidos (JAIMES, VERGARA-LAURENS e RAIJ, 2015).

Para o sucesso de um mecanismo de incentivo de *crowdsensing*, algumas condições se fazem necessárias, como viabilidade econômica, fator importante na decisão do modelo de recompensa; qualidade de dados, qual o valor dos dados que um determinado usuário envia; área de cobertura, pela difícil delimitação; equidade, normalmente compreendido como

oportunidade para todos os participantes; número de participantes adequado, mantendo um número mínimo de participantes; adaptação a demandas incrementais, possibilitando escalabilidade; e controle humano independente, com a capacidade de o sistema trabalhar autonomamente (JAIMES, VERGARA-LAURENS e RAIJ, 2015).

Conforme comentado anteriormente, pode-se dividir o *crowdsensing* em participativo e oportunista. Como existem métodos de incentivo que funcionam para ambos os casos, Jaimes et al. (2015) dividiram a classificação desses métodos em monetários, envolvendo recompensa financeira, e não monetários, com outro tipo de recompensa. Os métodos monetários podem ser estáticos, onde o montante do prêmio é definido *a priori*, e dinâmicos, onde se determina um orçamento variável, dependente de condições em tempo real. Os não monetários podem ser coletivos, encorajando o bem comum, sociais, tanto com exposição nas mídias sociais quanto na interação social, e de diversão, como gamificação.

De forma generalizada, *crowdsensing* é um processo simples com dois passos: atribuir tarefas aos usuários e esperar pelos resultados das tarefas atribuídas. Porém, em uma abordagem mais refinada, uma plataforma de *crowdsensing* explora informações sobre potenciais participantes e seu contexto móvel de execução, como a característica do seu deslocamento, para constantemente adaptar o processo de atribuição de tarefas de forma mais efetiva. Esse comportamento permite a divulgação e minimização de carga dos usuários, atribuindo tarefas para pessoas realmente dispostas a executar e compartilhar carga de trabalho de forma justa entre todos os usuários participantes (BELLAVISTA, CARDONE, *et al.*, 2015).

4.3 ESTUDOS CORRELATOS DE SMART CITY/MOBILIDADE URBANA

Cientes desse cenário, muitos estudos foram propostos para melhorar a mobilidade nas cidades. Visando exemplificar o potencial do assunto elaborado, apresentamos alguns projetos realizados que são aderentes ao tema. Na sua maioria, destacamos projetos com foco em transporte público, compartilhado, individuais e de sistemas de trânsito.

Preocupados com o tipo de sistema de trânsito mais adequado, Badia et al. (2016) propõem um estudo que avalia a adequação de um modelo de sistema de transporte público baseado em transferência em uma cidade de característica descentralizada em detrimento da

estrutura de um sistema de viagem em rede, em termos de custo total. Além disso, procuram fornecer orientações em como decidir qual é o melhor sistema para uma determinada idealização de cidade, de acordo com suas características, sistema de trânsito e demanda. Através de uma abordagem analítica, quatro aspectos da estrutura foram analisados, tais como estruturas radiais e baseadas em viagem, e esquemas baseados em transferência em grade e híbrido.

Os autores concluem neste estudo que o melhor modelo de sistema de trânsito depende do padrão de mobilidade da cidade. Identificaram relações entre a demanda e a aplicação dos diferentes sistemas de trânsito: um sistema radial de trânsito é mais adequado para uma cidade com características de alta concentração nas regiões de centro; quando áreas que contornam as regiões centrais começam a ficar mais populosas, deve-se adotar um sistema com novas linhas que conectem diretamente essas áreas; para cidades com padrões de comportamento e concentração dispersos, o mais eficiente seria um sistema híbrido baseado em transferência (BADIA, ESTRADA e ROBUSTÉ, 2016).

Buscando inter-relacionar diferentes políticas de mobilidade, Ambrosino et al. (2016) desenvolveram o FSUM (Flexible and Shared Use Mobility Centre), uma arquitetura conceitual que trabalha com três níveis conjuntamente: transporte coletivo, mobilidade pessoal e sistemas conectados. Essa arquitetura visa integrar serviços de transporte individuais, sob demanda ou compartilhados, juntamente com serviços coletivos, através da coordenação dos diferentes agentes e serviços, estruturados sob uma adhocracia, serviços disponíveis por tecnologia e medidas pouco restritivas.

Essa esquematização possibilita que o compartilhamento de viagens seja definitivamente considerado parte do sistema de transporte flexível e ambos integrados com o sistema de transporte público, aumentando suas possibilidades e aprimorando o serviço, levando este sistema a ser reconhecido como um a mobilidade orientada a serviço, o MaaS (Mobility as a Service). Os autores acreditam que essa é uma forma de se alcançar uma “mobilidade urbana mais sustentável, para pessoas e bens, e benefícios significativos em termos de consumo de energia, impactos ambientais e qualidade do ambiente urbano” (AMBROSINO, NELSON, et al., 2016, p. 179).

Como exemplo de projeto de *crowdsensing*, podemos destacar o Matador, um *framework* que permite especificar em quais situações uma determinada tarefa de

sensoriamento deve ser executada. Isso otimiza a participação dos usuários e minimiza o consumo de recursos do dispositivo móvel, pois conta com uma abordagem de consciência contextual (CARRERAS, MIORANDI, *et al.*, 2013). Na área do transporte público, destacamos o TrackInfo, que permite interações entre três diferentes entidades envolvidas: produtor, consumidor e fornecedor de serviço. Os passageiros, como produtores de dados tanto pelo sensor do dispositivo quanto por sua própria intervenção, reportam incidentes de trânsito, enviando dados da rota para o fornecedor de serviço. O fornecedor, por sua vez, realiza os passos de análise de dados e publica informações em tempo real, que serão utilizadas pelo usuário, no papel de consumidor, através do aplicativo (FARKAS, FEHÉR, *et al.*, 2015).

Com o intuito de se inserir no meio de projetos de *smart cities*, o projeto ParticipACT, idealizado na Universidade de Bolonha, propõe uma nova classe de plataformas sociotécnicas para ajudar no envolvimento das pessoas de forma participativa, onde os usuários são compreendidos como consumidores e produtores de dados e serviços da cidade inteligente. Sob a perspectiva do *mobile crowdsensing*, o projeto permite uma plataforma de sensoriamento onde os dispositivos móveis desempenham papel fundamental na coleta de dados e os usuários atuam ativamente, contribuindo com suas impressões dos acontecimentos, a partir da sua perspectiva individual. Inclusive, cada indivíduo pode interagir com participantes ao seu redor. Em resumo, o ParticipACT é uma solução de *smart city* que propõe diversos elementos, tais como uma plataforma de *mobile crowdsensing* que inclui um aplicativo para dispositivos móveis e uma aplicação *web* que controla esse processo, atuando como *back-end*. (CORRADI, CURATOLA, *et al.*, 2015b).

É pertinente mencionar trabalhos que desenvolveram projetos de ITS (Intelligent Transportation System) sob o pano de fundo *smart city*. ITS são “sistemas que utilizam tecnologias e conceitos de sistemas de engenharia sinergicamente para desenvolver e aprimorar sistemas de transporte de todos os tipos” (WANG, HERGET e ZENG, 2005, p. 261). Toqué et al. (2016) desenvolveram um sistema que realiza a predição do fluxo de passageiros de transporte público, baseando-se na matriz origem/destino das pessoas. A partir da coleta de *timestamp* dos eventos, identificar de veículo público, pontos e parada e padrão de itinerário, Gal et al. (2015) determinam a estimativa de tempo de viagens para soluções de transporte público integradas. A funcionalidade proposta por Dimitrakopoulos et al. (2010) sugere vias e direções para os motoristas e alertas de equipamentos de trânsito baseado no desenvolvimento de um sistema cognitivo. Alguns autores propõem aprimoramentos para o ITS, como o C-ITS

(Cooperative Intelligent Transportation System), que usa padrão de comunicação *vehicle-to-vehicle* (V2V) e *vehicle-to-infrastructure* (V2I) para operar um sistema de transporte (SEREDYNSKI e VITI, 2016), e CTS (Cybernetic Transportantion System), apresentando um módulo que realiza tomadas de decisão para os operadores de ITS, baseando-se em conceitos de automação (VACA, TRAVER, *et al.*, 2016).

5 MODELOS DE ANÁLISE DE DADOS DE MOBILIDADE URBANA

Este capítulo explora pesquisas realizadas no âmbito de análise de dados, sob a perspectiva da *smart city*, que auxiliam no estudo sobre mobilidade urbana. Portanto, primeiramente apresentamos trabalhos que relatam projetos de desenvolvimento de processos que fazem parte de análise de dados como um todo e, posteriormente, trabalhos que propõem arquiteturas completas que sejam pertinentes ao tema da mobilidade urbana. O objetivo é compreender o estado da arte de projetos de análise de dados que podem servir como base para o desenvolvimento do *framework* de análise de dados de mobilidade urbana proposto.

5.1 ETAPAS ENVOLVIDAS NO PROCESSO DE ANÁLISE

Alguns modelos foram propostos com o intuito de realizar análise de dados para estudar a mobilidade urbana. Partindo de tecnologias inovadoras dentro do conceito *smart city* discutido, esses modelos auxiliam na orientação e fornecem substratos para a formulação de um *framework* para análise de dados de mobilidade urbana, como é objetivado neste estudo.

Zheng et al. (2016) entendem que as dinâmicas da organização social e cultural são importantes no estudo da *smart city* e mobilidade urbana, e devem ser utilizadas com tecnologias que permitam a abordagem da diversidade disciplinar. Dados de transporte público podem registrar hora, coordenadas de GPS, velocidade, aceleração, local, textos, vídeos, entre outros. Cada tipo de dado coletado é específico a um aspecto humano, e sua qualidade define as vantagens e limitações que este dado apresenta, fazendo com que cada tipo de dado de transporte social seja utilizado para um fim específico. Por exemplo, sinais de GPS de taxis podem ser usados para estimar a velocidade que se consegue desenvolver nas ruas da cidade, mas não para supor a demanda por transporte, visto que somente considera origem e destino dos próprios taxis.

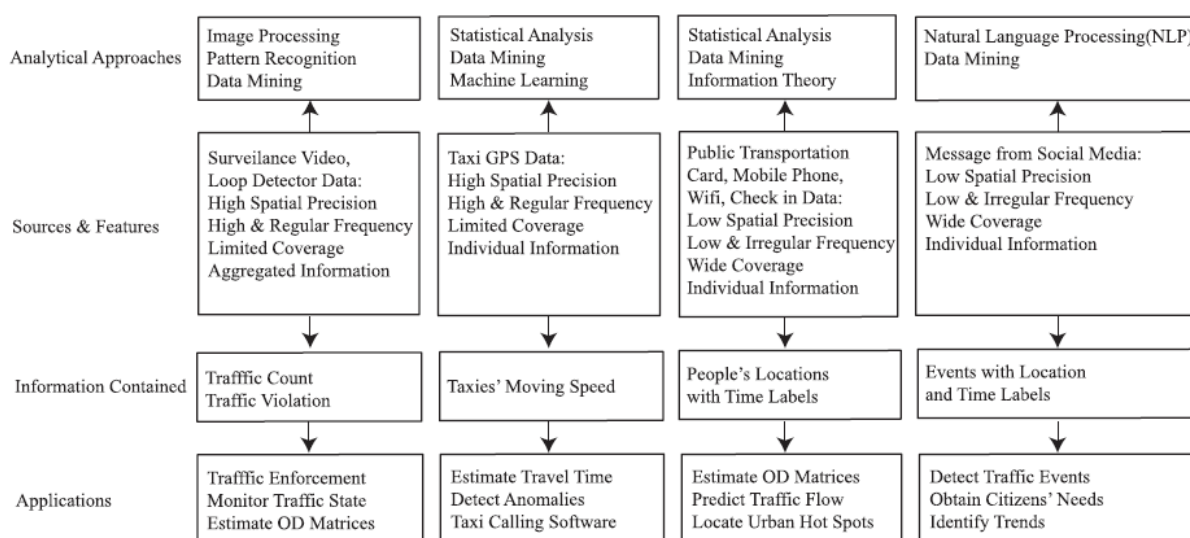
Diferentes tipos de dados de transporte público podem ser reunidos, análises baseadas em um tipo de dado podem ser validadas entre os diferentes tipos, melhorando a assertividade, e novos conhecimentos são gerados a partir dessa organização. O trabalho com variedade de dados ainda está em um estágio inicial, requer comunicação e compartilhamento de diferentes

áreas e disciplinas, e necessita de modelos e metodologias sistemáticas para sua organização. Esses estudos podem auxiliar os sistemas de trânsito na sua performance, descobrindo comportamentos e sentimentos em diferentes escalas, na comunicação com as pessoas de forma mais clara e acessível, e a compreender as necessidades dos cidadãos, tendências e resolver problemas (ZHENG, CHEN, *et al.*, 2016).

Os autores explicam que um *big data analytics* de transporte público pode ser classificado em três categorias: análise estatística, mineração de dados e análise visual. A análise estatística compreende a coleta, análise, interpretação e apresentação dos dados. Além de fazer um estudo descritivo dos dados, também permite realizar inferências sobre a população e identificar padrões e relacionamentos. A mineração de dados possui métodos como agrupamentos, classificação, sumarização, detecção de anormalidade e regressão, buscando identificar padrões desconhecidos e inesperados. Por último, a análise visual tem como cerne a visualização de dados, transformando vários tipos de dados diferentes em uma representação visual harmônica, melhorando o entendimento e permitindo a interação de planejadores para que possam combinar os resultados com seus conhecimentos pessoais nas decisões (ZHENG, CHEN, *et al.*, 2016).

Abaixo podemos verificar diferentes cadeias de utilização de dados de mobilidade urbana, com fontes, análises e aplicações (Figura 1):

Figura 1: Fontes, abordagens analíticas e aplicações



Fonte: (ZHENG, CHEN, *et al.*, 2016).

No entanto, no âmbito do sensoriamento e análise de dados do ambiente urbano, alguns desafios estão em aberto e devem ser discutidos, como o processamento de dados imprecisos e com “sujeira” coletados por sensores, dados faltantes e a construção de uma cadeia que auxilie tanto o fornecedor de serviço quanto os usuários. Todo esse cenário envolve a interação entre dados de fontes heterogêneas, que devem ser integrados face a grande distribuição de sensores, e o pré-processamento desses dados, visto que normalmente apresentam diversas adversidades e podem ser recebidos em várias frações. Desenvolvendo *warehouses* distribuídos, utilizando noSQL (Non-sequel SQL), uma forma não estruturada de organização de banco de dados, e realizando limpeza e sincronização de dados, é possível endereçar esses pontos (SUN, 2017).

A etapa de coleta é crucial para uma arquitetura de análise de dados. A forma que é realizada, suas fontes e tipos de dados coletados são pontos tratados com bastante cuidado nos projetos levantados. A coleta deve ser realizada de forma eficiente e precisa para que as análises sejam confiáveis.

Sun et al. (2016) propõem uma arquitetura sistemática que analisa a mobilidade urbana a partir de dados de 4G dos usuários, que possui grande quantidade de registros HTTP (Hypertext Transfer Protocol), um protocolo de comunicação da internet. Apesar de não ser aplicável para qualquer tipo de dado, os autores afirmam ser adequado para qualquer tipo de dado 4G. Uma das análises é a extração de rotas, onde é realizada a determinação de trajetos brutos a partir dos registros HTTP em sequência, a segmentação desses trajetos brutos através da sua ordenação e avaliação, uma posterior sobreposição e reconstrução dos trajetos. Outra forma de análise é a identificação de locais com maior frequência de usuários, focando nos lugares mais importantes de permanência.

Inspirados em limitações identificadas para os processos de *mobile crowdsensing*, Datta et al. (2016) desenvolveram uma arquitetura baseada em IoT para campanhas de *crowdsensing*. Os autores identificaram a falta de um método uniforme de coleta e compartilhamento de dados de *crowdsensing*, a dependência de constante transmissão de dados dos aparelhos, a falta de padronização nos mecanismos de interpretação dos dados e a dependência de sensores embutidos nos aparelhos.

Com o intuito de minimizar esses problemas, a arquitetura possui uma característica de consciência de energia que se adapta de forma dinâmica de acordo com a bateria dos dispositivos dos usuários. Ela é composta por um mecanismo de monitoramento de contexto e

bateria, um mecanismo de análise e características auto-adaptativas. O mecanismo de monitoramento coleta a bateria restante e o contexto e faz o armazenamento em um módulo estatístico que computa padrões de recarga e localização. A partir de dados de bateria e localização, o sistema pode se adaptar, de acordo com pré-configurações, alterando a forma de solicitação e uso dos recursos do aparelho (DATTA, DA COSTA, *et al.*, 2016).

Vale destacar o subsistema de armazenamento e processamento dos dados. Os metadados dos sensores são tratados via tecnologias de *web semantic*, que visa atribuir significado aos conteúdos analisados, permitindo combinação de dados de fontes e natureza diferentes. De maneira geral, o dado bruto é codificado, submetido a um processo de *web semantic* e aplicações de domínios, e ao final, o conhecimento gerado é combinado para se criar um conhecimento de nível mais complexo. Esse processo resulta uma implementação suficientemente genérica para ser utilizada em diferentes plataformas de IoT (DATTA, DA COSTA, *et al.*, 2016).

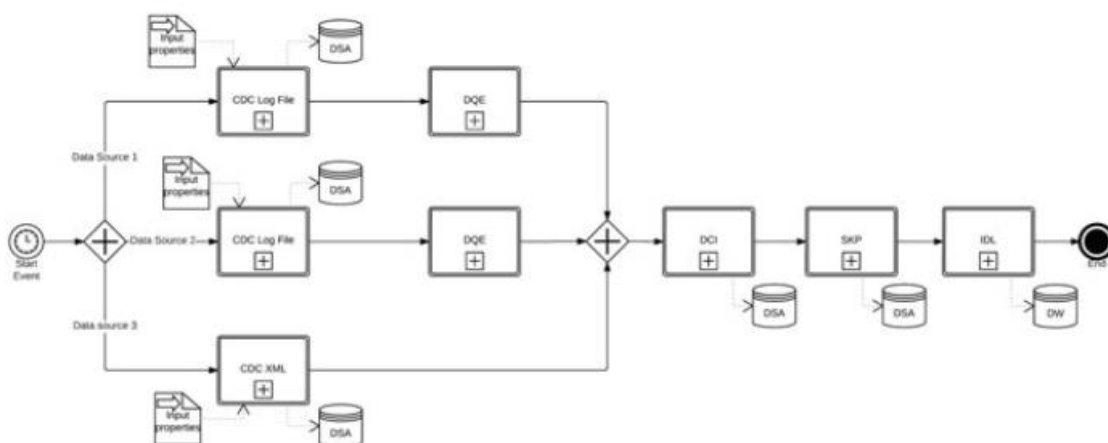
Outra inteligência atenta ao consumo de bateria e identificação das características de atividade dos usuários é o CARDAP, uma plataforma teórica de análise de dados escalável, eficiente e genérica para *mobile crowdsensing*. Cada indivíduo apresenta um comportamento de movimentação característica: correr para uns pode significar andar para outros. Este projeto propõe um sistema adaptativo e evolutivo utilizado para calibração e aprimoramento do reconhecimento de atividades pessoais. Essa técnica é chamada StreamAR, que realiza aprendizagem ativa de reconhecimento de atividade. Como o CARDAP nativamente incorpora essa técnica, é possível utilizá-lo como componente no desenvolvimento de uma aplicação específica de análise de dados (ABDALLAH, GABER, *et al.*, 2012) (JAYARAMAN, GOMES, *et al.*, 2015).

Os dados coletados normalmente são recebidos de forma desorganizada e não uniforme, pois muitas vezes possuem características diferentes, não seguem necessariamente um padrão, o processo de coleta está sujeito a falhas e pode conter dados irrelevantes. Para que se possa submetê-los a uma análise, devem ser organizados de acordo com regras e técnicas definidas, e então processados para servirem de informação para as ferramentas de análise.

Os processos de ETL⁴ são um dos mais importantes componentes do gerenciamento de dados, principalmente na complexidade da *smart city*. Para que seja efetivo no gerenciamento de dados estruturados, atualizados e dentro do escopo definido para tomadas de decisão, um sistema ETL deve contar com alta disponibilidade, baixa latência e escalabilidade. De forma generalizada, o processo envolve o acesso aos dados das fontes, a etapa ETL com limpeza, integração, transporte e canalização que transformam o dado e armazenam no sistema alvo específico (SABTU, AZMI, *et al.*, 2017).

Visando minimizar impactos na estrutura tanto do dado quanto da base de dados, Oliveira e Belo (2017) propõem uma padronização do uso de ETL para construção de pacotes ETL, usando um processo para modelar *workflows* mais conceituais. Foram definidas três categorias de padrões ETL: Extratores, que são usados nas operações de extração, representam processos de extração de dados e aplicações de algoritmos em determinadas estruturas; Transformadores, usados nas transformações de dados como limpeza ou tarefas de adequação; e Carregadores, que carregam dados para o repositório, contando com algoritmos eficientes ou criação de índices e manutenção. Na Figura 2 podemos verificar um arranjo de um *workflow* de ETL com ferramentas representando as características mencionadas:

Figura 2: Padrão de modelo de processos de negócio para ETL



Fonte: (OLIVEIRA e BELO, 2017).

⁴ Extract, Transform and Loading. Processo de extração dos dados de um repositório, transformação via tratamento dos dados e carregamento em posições específicas para posterior processamento (SABTU, AZMI, *et al.*, 2017).

Para extração de dados, por exemplo, vemos alguns padrões CDC⁵ sendo utilizados, cada um atuando em uma específica estrutura de dados, com estratégias para extrair o dado sem prejudicar as rotinas do sistema. Dois padrões CDC atuam em arquivos de *log*, ou seja, que contém informações históricas de alteração, e um em arquivos XML⁶. Para a parte de transformação, padrões DQE processam e resolvem problemas dos processos de extração, como conflito de tipos de dados e redundância, assegurando a qualidade do dado. Padrões DCI mantêm a integridade de dados do mesmo tipo que correspondem a fontes distintas, preservando a relação entre os campos, e SKP se encarregam do processo de alimentação das tabelas, realizando conversões para evitar conflitos. Após essa etapa, padrões de carregamento, como IDL, são executados para inserir os dados nos repositórios corretos. Esse modelo permite a representação de um processo ETL usando níveis de abstração diferentes, produzindo modelos conceituais valiosos (OLIVEIRA e BELO, 2017).

Daniel et al. (2017) desenvolveram um *framework* para análise em tempo real a partir de dados de veículos autônomos, compreendendo mecanismos de armazenamento de dados distribuído para transmissão de processamento de análise em tempo real. No processo de classificação dos dados, o dado bruto não estruturado é coletado dos sensores do veículo. Posteriormente, ocorre um processo de ETL (Extract, Transform and Load) para produzir dados limpos, importantes, semi-estruturados e comprimidos. Por final, o dado pré-processado é classificado para que possa ser analisado.

O *framework* proposto pelos autores mencionados consiste em quatro partes: filtro de dados, análise, processamento e armazenamento distribuído (Figura 3). Na etapa de filtro, é aplicada a técnica Kalman Filter, algoritmo utilizado para remover dados imprecisos contidos nos dados dos sensores, baseando-se na combinação do estado anterior com o estado corrente do dado. Por gerenciar o dado em um ambiente Hadoop, o procedimento Map Reduce⁷ é utilizado para a parte de análise, reduzindo o conjunto de dados anteriormente filtrado. Para processamento em tempo real, o sistema usa Hadoop HBASE, que oferece procedimentos de

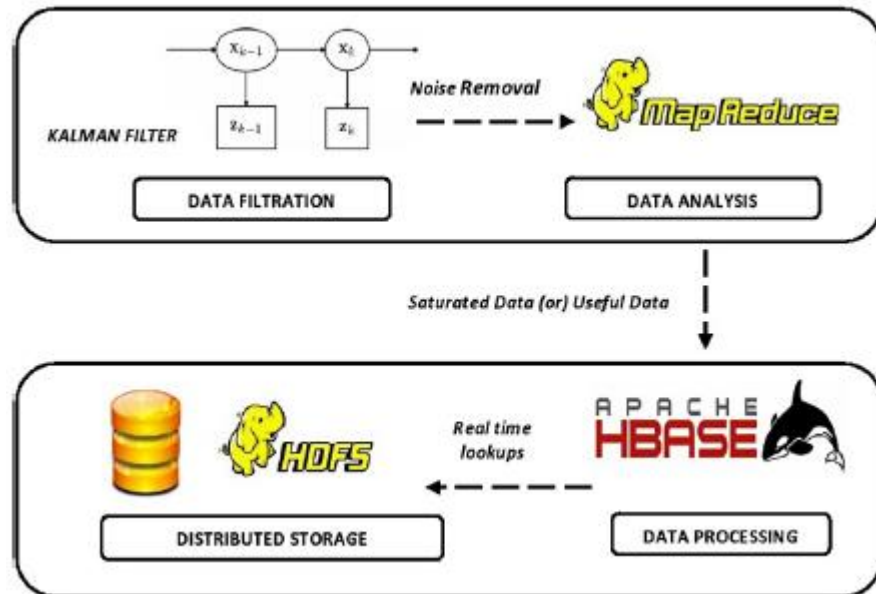
⁵ Sem nomear as siglas, os autores explicam que CDC, DQE, DCI e SKP são padrões ETL de transformação e limpeza de dados para adaptá-los a um determinado objetivo (OLIVEIRA e BELO, 2017).

⁶ Arquivos de protocolo para internet.

⁷ Hadoop *software open source* desenvolvido em Java que consiste em duas partes: o Hadoop Distributed File System (HDFS), que armazena os dados, e o MapReduce, que realiza o processamento em paralelo entre as diferentes máquinas organizadas na arquitetura de computadores (RAJARAMAN, 2016).

lookup, *cache* em memória e programação *server-side* – tarefas em paralelo. O armazenamento é realizado pelo HDFS por ser distribuído e suportar classificação de grandes conjuntos de dados e tomadas de decisão em tempo real.

Figura 3: Proposta de *framework* para dados em tempo real



Fonte: (DANIEL, SUBBURATHINAM, *et al.*, 2017).

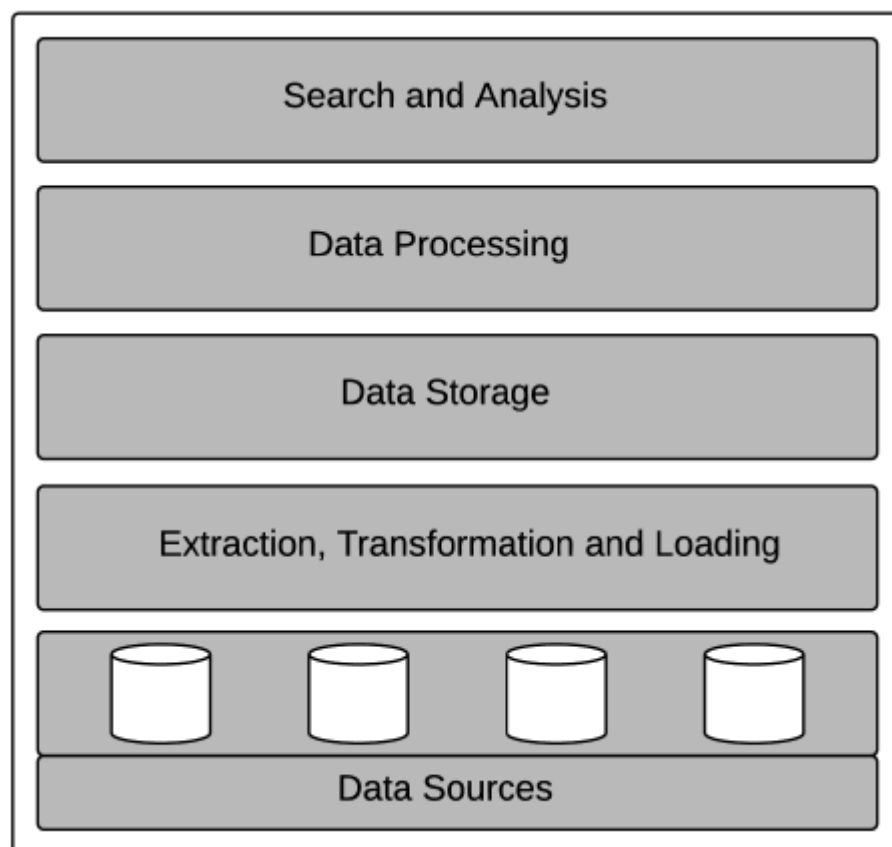
5.2 ARQUITETURAS DE ANÁLISE DE DADOS

Uma arquitetura completa de *big data analytics* deve considerar os aspectos mencionados, tais como coleta, organização e processamento, juntamente com ferramentas de análise e sua disponibilização. Portanto, é necessário gerenciar todo o processo de forma holística, desde a extração de insumos para o estudo até a exposição do conhecimento gerado. Este capítulo tem o intuito de explorar algumas arquiteturas de *big data analytics* propostas por pesquisadores na área de mobilidade urbana, buscando fundamentos teórico/empíricos que auxiliem na elaboração do *framework* pretendido.

No intuito de estruturar um ambiente que represente uma plataforma de larga quantidade e variedade de dados para se realizar um *big data analytics*, com armazenamento eficiente e rápido processamento, Gomes et al. (2016) propuseram um modelo de ambiente de *big data* do projeto ParticipACT Brasil com 5 camadas: a primeira camada possui diferentes dados de

diferentes entidades armazenados em uma base de dados; a segunda camada integra os dados para que sejam armazenados em um mesmo formato; a terceira executa o armazenamento dos dados em um *warehouse*; a quarta camada faz o processamento dos dados; e a última realiza buscas e análises que podem ser visualizadas através de gráficos e interfaces amigáveis, como um portal *web* (Figura 4).

Figura 4: Modelo de ambiente Big Data



Fonte: (GOMES, DANTAS, *et al.*, 2016).

Bellavista et al. (2015) explicam as características de uma campanha de *crowdsensing* – conceito introduzido no subcapítulo 4.2.4 desta dissertação – e as etapas de análises que se pode realizar com esses dados. Adquirir dados de um grupo de pessoas pode trazer complicações sociais, como motivar a participação dos usuários, dar liberdade de escolha na participação e garantir a privacidade dos dados. Portanto, alguns requisitos devem ser atendidos para que a campanha seja exitosa. Deve-se definir as dimensões das tarefas, especificando o quê, quando, onde e de quem o dado será coletado, e seguir os dois passos característicos do *crowdsensing*, a atribuição de tarefas e adquirir os resultados depois da finalização das tarefas.

O sensoriamento de dados inclui todo o gerenciamento e processamento necessários para a coleta de dados. Existem dois tipos de coleta de dados, o passivo, sem a intervenção do usuário, e o ativo, com ação direta do usuário. O usuário deve saber quando uma campanha está disponível, podendo esta ser definida de acordo com o horário ou o local em que ele se encontra. O *crowdsensing* deve ser leve e consciente do contexto e sistema, portanto, deve possibilitar o gerenciamento dos recursos dos aparelhos dos usuários, minimizando o seu consumo, evitando atritos com outras aplicações e atuando de forma não intrusiva (BELLAVISTA, CARDONE, *et al.*, 2015).

Os dados coletados devem ser processados para posterior análise. Num primeiro momento, deve-se realizar a interpolação dos dados, para que alguns dados faltantes possam ser preenchidos por inferência, como dados de localização, por exemplo, e *outliers* possam ser desconsiderados. Outro passo é a integração de dados, a organização de dados complementares, como dados de localização de um mesmo instante. A partir desses dados já tratados, o sistema executa a ferramenta MMA (Mobility Mode Analysis), desenvolvida pelo projeto, que explora a localização e reconhecimento de atividade para inferir rotas preferenciais e como os usuários se deslocam por essas rotas (Figura 5). É possível identificar onde usuários descansam ou param com frequência, conectando esses pontos para identificar como eles se movem e o modal, e utiliza a plataforma Foursquare para descobrir pontos de interesse e estabelecimentos (BELLAVISTA, CARDONE, *et al.*, 2015).

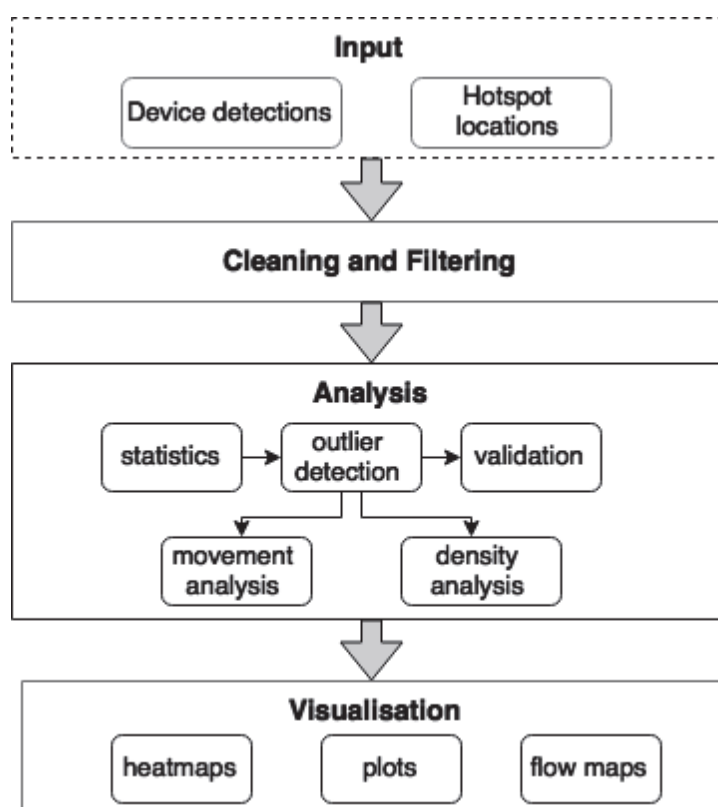
Figura 5: Modo de deslocamento inferido pelo MMA



Fonte: (BELLAVISTA, CARDONE, *et al.*, 2015).

Draghici et al. (2015) apresentam uma ferramenta de visualização de análise da mobilidade urbana. A base para sua coleta é a identificação de *hotspots* de *wi-fi*. Os autores desenvolveram uma arquitetura que aplica métodos analíticos a qualquer estrutura de dados que respeitam uma mesma característica, contando com endereço, *timestamp* e ponto de localização. O sistema de camadas importa dados de vários formatos, armazena e organiza os dados, e conta com vários componentes de análise e métodos de visualização (Figura 6).

Figura 6: Arquitetura de sistema



Fonte: (DRAGHICI, AGIALI e CHILIPREA, 2015).

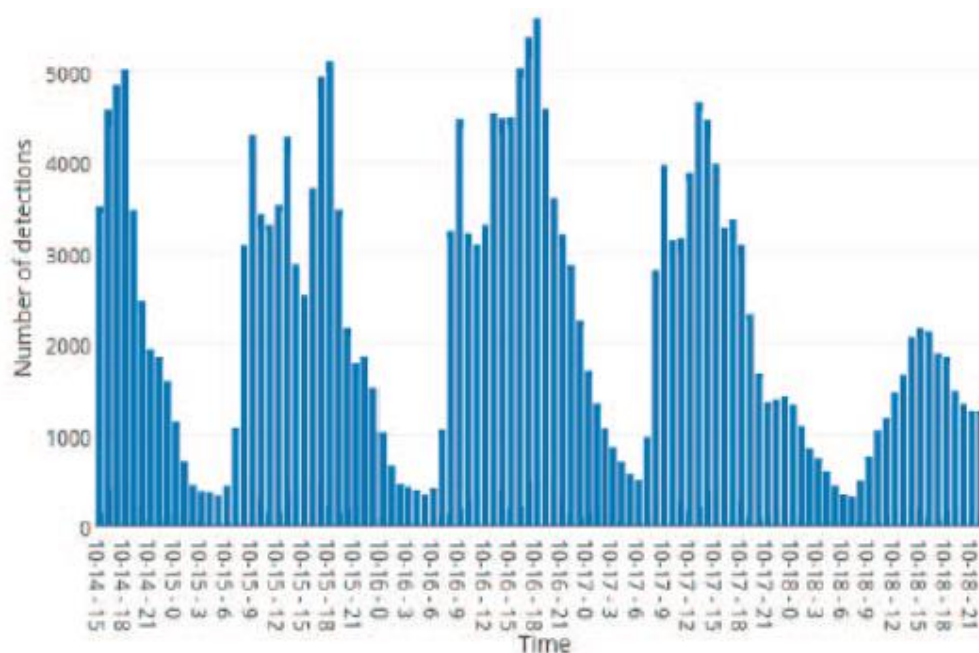
Os dados coletados são compostos por informações de rotas e registros dos dispositivos detectados, respeitando um mínimo de atributos dos registros, como ID do *hotspot*, identificador do dispositivo e *timestamp*. A camada de filtro detecta, corrige ou remove informações corrompidas ou imprecisas, mantendo a consistência dos dados. O sistema executa uma verificação nos dados para evitar informações faltantes ou enganosas. Essa camada também importa e armazena os dados em uma base de dados. Por conta da complexidade e necessidade de desempenho de processamento, o tratamento de *outliers* é realizado na camada

de análise. A eliminação de *outliers*⁸ melhora a precisão dos dados e otimiza o seu processamento, pois diminui a quantidade, removendo dados desnecessários. Uma característica de dados de *wi-fi* de *hotspots* é que estão sujeitos ao efeito *ping-pong*, ou seja, a coleta de um mesmo dispositivo por mais de um ponto. Esse problema pode ser resolvido pela criação de um ponto virtual entre *hotspots* concorrentes (DRAGHICI, AGIALI e CHILIPIREA, 2015).

A camada de visualização tem o propósito de apresentar as informações geradas de forma amigável e intuitiva para gestores e pesquisadores. Ela expõe as características dos deslocamentos e permite a comparação entre momentos diferentes, facilitando o entendimento dos eventos e permitindo a identificação de padrões. Os modelos de visualização podem ser acessados via *web* ou por APIs (Application Programming Interface), que são interfaces que permitem comunicação entre plataformas diferentes, e são apresentados por mapas lineares, de fluxo e de dispersão, mapas de calor dinâmicos, gráficos de barra, pizza e bolha (DRAGHICI, AGIALI e CHILIPIREA, 2015).

Um exemplo é o gráfico de número de detecções a cada três horas. Nele é possível identificar os picos de quantidade de dispositivos identificados desde 14 de outubro, às 15 horas e 18 de outubro, às 21 horas (Gráfico 8).

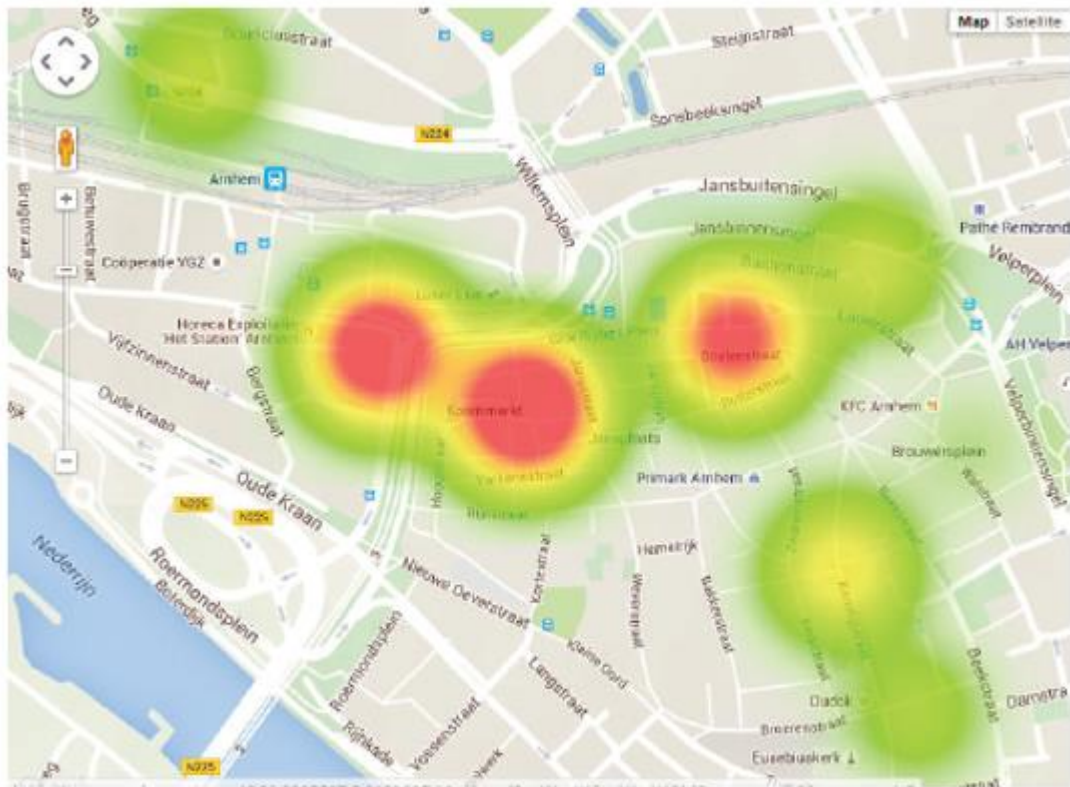
⁸ *Outliers* são valor atípicos, que apresentam grande afastamento dos dados gerais coletados.

Gráfico 8: Quantidade de detecções identificadas a cada três horas

Fonte: (DRAGHICI, AGIALI e CHILIPAREA, 2015).

Outro exemplo de representação dos dados é o mapa de calor, que auxilia na identificação das dinâmicas de movimentação das pessoas e o seu fluxo, pois é apresentado em forma de vídeo (Figura 7). Esse componente faz uma representação tanto de densidade quanto de movimento.

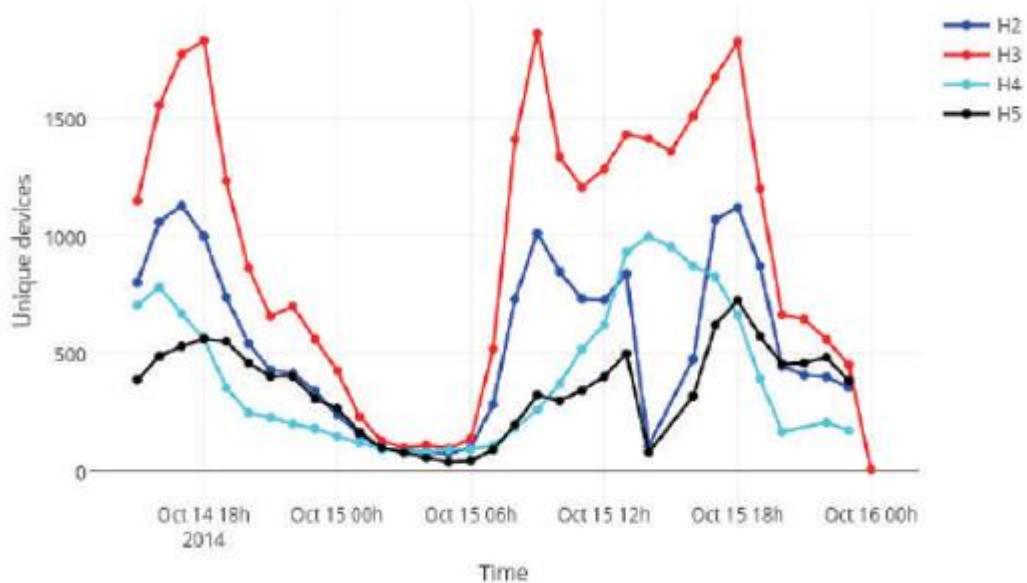
Figura 7: Mapa de calor dinâmico de densidade.



Fonte: (DRAGHICI, AGIALI e CHILIPAREA, 2015).

Também é possível gerar um gráfico comparativo entre os diferentes *hotspots* por intervalo de tempo, para auxiliar nas interpretações da quantidade de dispositivos detectados por localidade (Gráfico 9).

Gráfico 9: Comparativo de densidades de quatro *hotspots* diferentes



Fonte: (DRAGHICI, AGIALI e CHILIPAREA, 2015).

Fazendo uso do Big Data, Wang et al. (2016) desenvolveram um sistema inteligente, que está sendo testado em Fortaleza, para monitorar e avaliar o desempenho do sistema de transporte público, utilizando-se de ferramentas de análise de dados com técnicas de processamento distribuído e que administram grande quantidade de dados. O sistema também é composto por três camadas, onde a primeira camada realiza a coleta de dados de GPS e sinais de registros de coleta de cobrança para saber o tempo de viagem de um ônibus, e localização de embarque de passageiros. A segunda camada identifica conjuntos de pontos de ônibus, de acordo com a localização geográfica e importância, e investiga o que pode estar afetando o fluxo. A terceira camada disponibiliza essas informações em mapa para que planejadores possam monitorar o desempenho e tomar decisões.

Dados de GPS e AFC⁹ são uma forma econômica de se capturar dados de viagens de ônibus e devem ser tratados de forma integrada, pois possuem informações complementares, onde GPS é capaz de informar dados sobre rota e AFC sobre embarque/desembarque de passageiros. Neste estudo, foram utilizados dados de aproximadamente 2000 ônibus, por 7 dias, na cidade de Fortaleza, contendo 300 rotas, todos os veículos equipados com dispositivo de rastreamento por GPS e sistema de AFC (WANG, RAM, *et al.*, 2016).

Através de um algoritmo baseado na lógica MapReduce, o sistema é capaz de associar os dados de GPS dos ônibus, de AFC correspondentes e de pontos de parada. A partir dessas informações de embarque e desembarque de passageiros, o sistema determina, na segunda camada, padrões de deslocamentos compartilhados entre passageiros. Via método *k-means*¹⁰ de agrupamento de conjuntos, as regiões, pontos de parada e deslocamentos compartilhados são determinados. Após isso, um algoritmo modificado de *PageRank*¹¹ classifica as rotas de acordo com a relevância, considerando o peso da região como a quantidade de fronteiras que ela possui. Os resultados são, então, apresentados como interconexão de pontos de parada nas diferentes regiões (Figura 8). Esta análise permite que planejadores de trânsito verifiquem a adequação

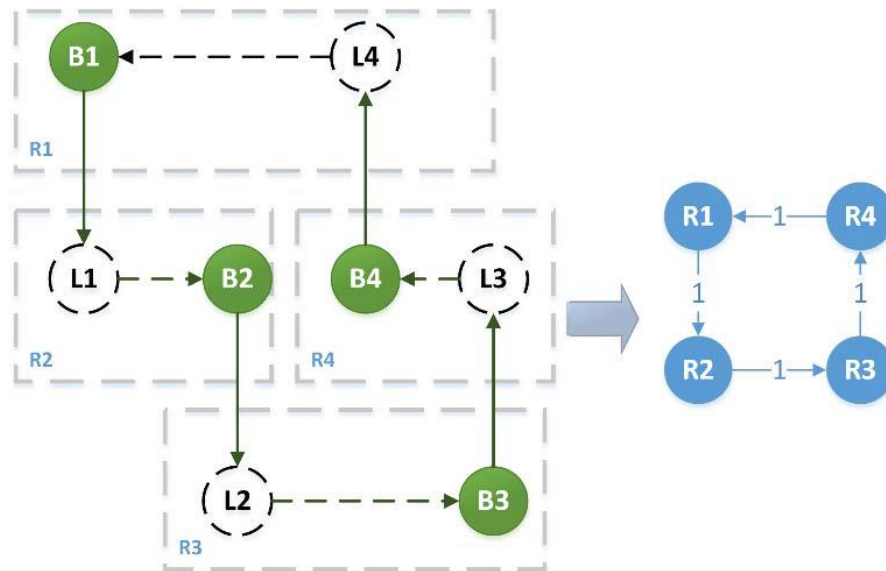
⁹ Automatic Fare Collection, registros de coleta de cobrança dos transportes coletivos (WANG, RAM, *et al.*, 2016).

¹⁰ Método de determinação de conjuntos que particiona os dados observados em grupos de acordo com a média (JAIN, 2010).

¹¹ Algoritmo concebido inicialmente para atribuir uma credibilidade à cada página da internet baseando-se na sua estrutura (BENINCASA, CALDEN, *et al.*, 2006).

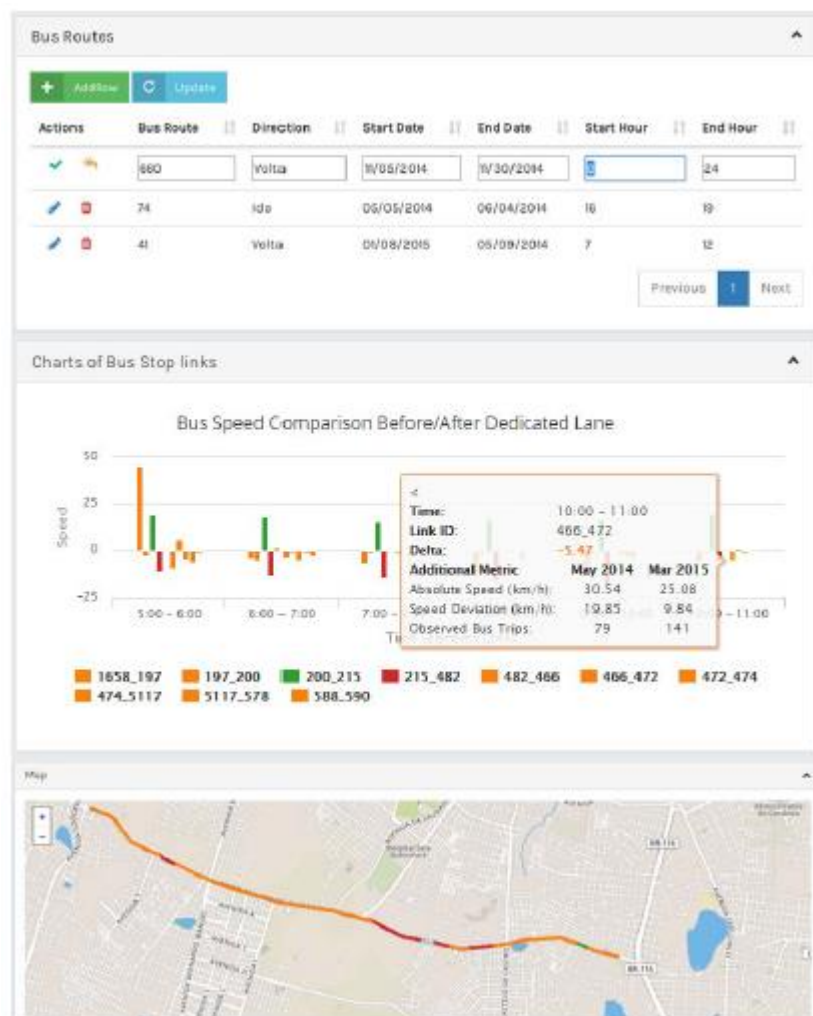
das linhas de ônibus atuais às reais necessidades de deslocamento dos passageiros (WANG, RAM, *et al.*, 2016).

Figura 8: Padrão de mobilidade de passageiros



Fonte: (WANG, RAM, *et al.*, 2016).

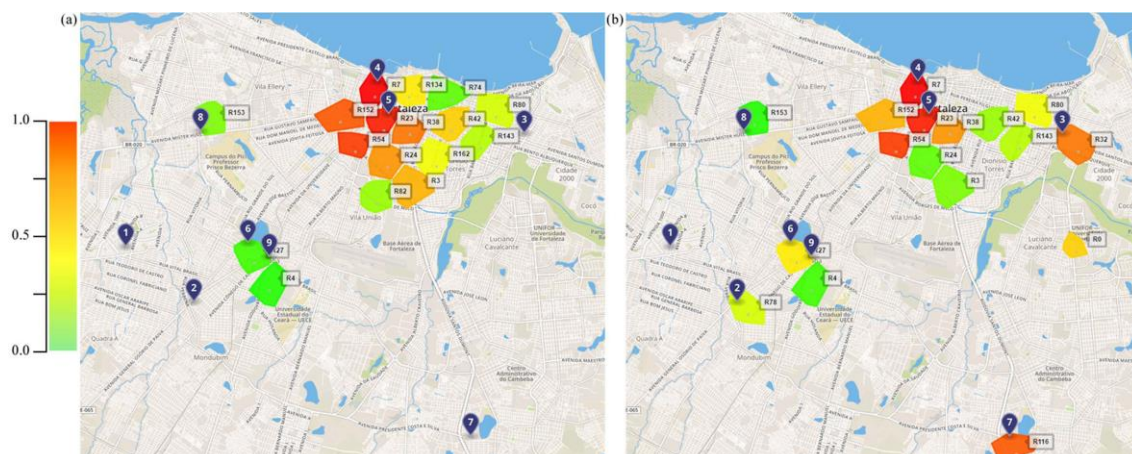
Baseado nos tempos de viagem entre cada parada, é possível realizar análise de atrasos de ônibus. Pode-se identificar atrasos tanto em nível de rota quanto de segmento. Os atrasos de rota são determinados a partir da comparação com os dados de rotas e horários disponibilizados pela prefeitura, e os atrasos de segmento consideram a média de velocidade dos ônibus para inferir o tempo necessário para percorrer um determinado percurso. A análise dos atrasos de segmento é apresentada na terceira camada para utilização dos usuários no mapa (Figura 9), e uma análise de regressão é aplicada à análise de rota para que se entenda os possíveis problemas que podem ter causados tais atrasos. A apresentação dessas análises é importante para que planejadores possam entender melhor os impactos de suas decisões e identificar o que pode estar causando os atrasos (WANG, RAM, *et al.*, 2016).

Figura 9: Interfaces da análise de atrasos

Fonte: (WANG, RAM, *et al.*, 2016).

Realizando uma aplicação em cenário real na cidade de Fortaleza, a camada de visualização permite identificar regiões com maiores demandas de mobilidade a partir de duas métricas: volume médio de embarque de passageiros por parada em cada região (Figura 10-a) e ranqueamento das regiões (Figura 10-b). Essa abordagem permite que planejadores possam reavaliar as regiões que possuem paradas de ônibus de acordo com a demanda.

Figura 10: Identificação de regiões com maior demanda de mobilidade



Fonte: (WANG, RAM, *et al.*, 2016).

Com o intuito de identificar a mobilidade dos turistas de uma cidade, Fistola et al. (2017) desenvolveram um modelo de análise de dados baseados em mídias sociais. Os dados são extraídos do portal TripAdvisor, que possui dicas e publicações de turistas que estão em viagem coletando características de acomodações, restaurantes, museus, entre outros estabelecimentos, e é considerado uma fonte consistente de dados do setor turístico. A base de dados para a análise espacial foi construída com base na técnica Web Scraping aplicada ao portal TripAdvisor. Web Scraping é uma técnica de mineração de dados que permite a extração de dados de páginas HTML. Esses dados, então, são estruturados e sujeitos a processos de análise. O estudo foi realizado na cidade de Nápoles, Itália.

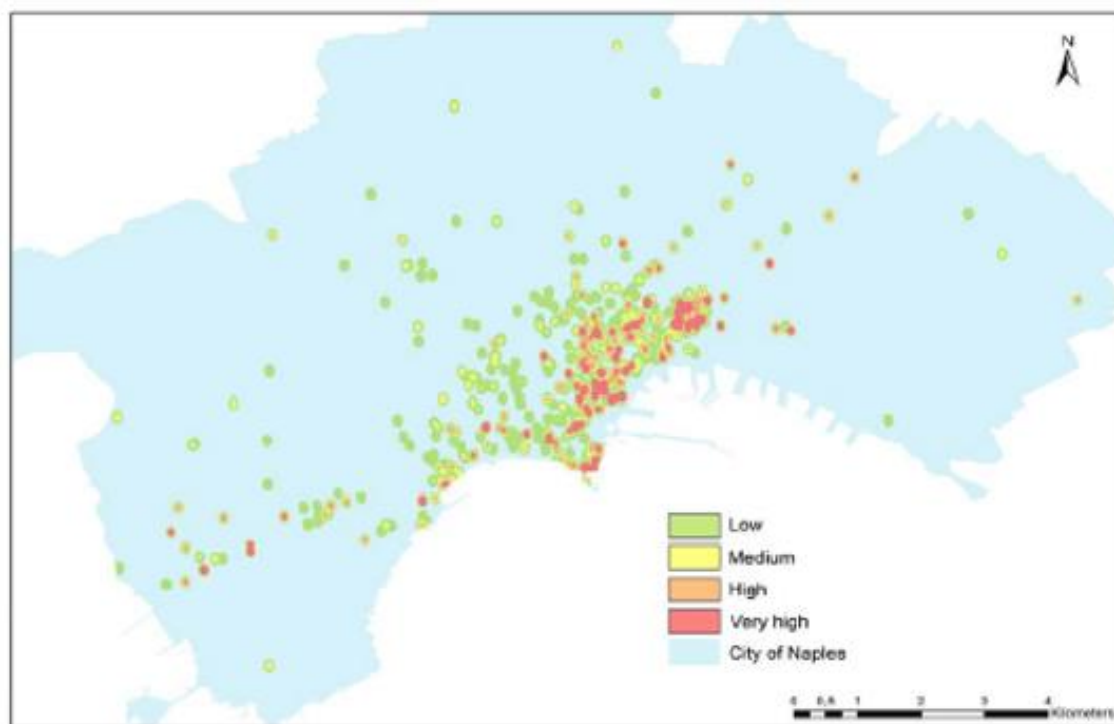
Como o TripAdvisor não fornece dados de geolocalização, foi aplicado o método KDE (Kernel Density Estimation), que permite estimar a concentração de um fenômeno em um determinado território. O seu resultado é uma tabela utilizada para se calcular a densidade individualmente por célula, de acordo com o tamanho estipulado pela pessoa que está realizando a análise. Neste estudo, a densidade foi determinada considerando o número de visitantes aos pontos de interesse (Points Of Interest, POI) e hóspedes nos estabelecimentos de acomodação. O peso é atribuído de acordo com o número de avaliações dos usuários. No Quadro 1, pode-se verificar um exemplo para o museu Capela San Severo:

Quadro 1: Exemplo de identificação de peso por número de avaliações

Type of traveler	Number of reviews	Coefficient	Product	Tot. n. of reviews for Kernel
Family	1868	4	7472	15254
Twosome	3563	2	7126	
Singular	656	1	656	

Fonte: (FISTOLA, RAIMONDO e LA ROCCA, 2017).

A densidade por região é então apresentada no mapa e classificada como muito alta, alta, média e baixa. Isso permite identificar áreas urbanas sensíveis que podem ser consideradas no planejamento estratégico para mobilidade dos turistas, integrando-se com objetivos para a mobilidade urbana e aprimoramentos da mobilidade sustentável, juntamente com a avaliação de eficiência das políticas de mobilidade já implementadas. Vemos na Figura 11 um exemplo de apresentação das regiões com maior densidade a partir dos dados de estabelecimentos de acomodações:

Figura 11: Regiões com maior densidade

Fonte: (FISTOLA, RAIMONDO e LA ROCCA, 2017).

Relativo à identificação de regiões mais frequentes, Corradi et al. (2015a) apresentaram uma integração do já discutido ParticipACT que utiliza dados de geolocalização coletados via

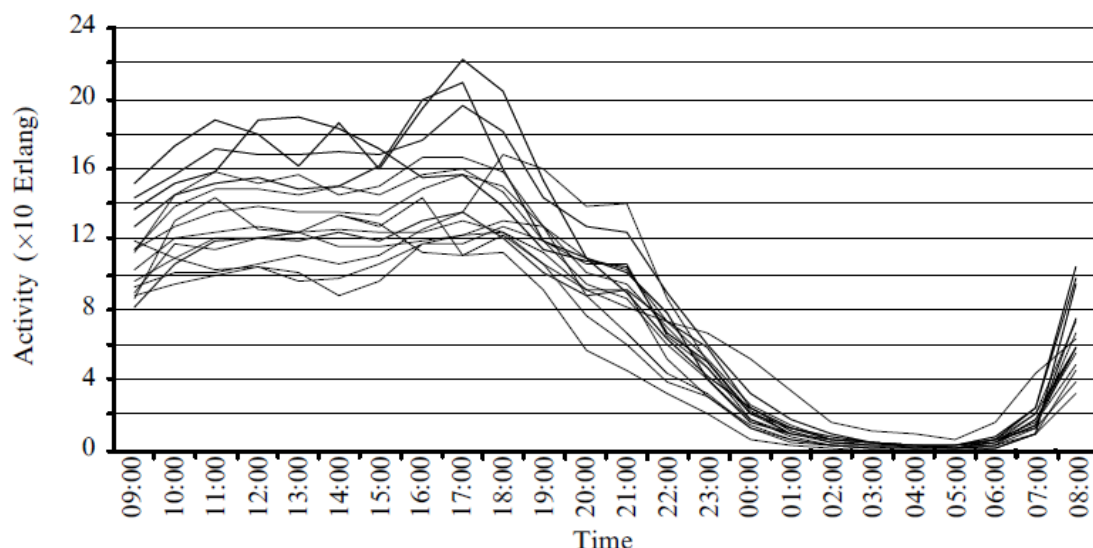
crowdsensing para promover a identificação de POIs. Inicialmente, o sistema coleta todos os dados de GPS dos usuários tanto parados quanto se locomovendo. Para aprimorar a precisão, esses dados são processados juntamente com dados de atividade coletados dos sensores de acelerômetro dos celulares, que permitem uma inferência das características de movimentação do usuário. Após essa coleta, é necessário fazer uma integração com a localização do usuário correspondente, uma tarefa complicada pois depende da combinação do usuário específico e do *timestamp*. Esse procedimento é custoso, assim como a própria coleta de GPS, e um procedimento de coleta com uma frequência espaçada, algo como a cada 5 minutos, é razoável para precisão e não consome tanta bateria do dispositivo. A segunda etapa aproxima todos os dados de localização para determinar uma vizinhança onde os usuários estão mais presentes e permanecem por mais tempo. Por fim, a terceira etapa conta o número de visitas num mesmo período e elege pontos considerados como prováveis POIs.

No intuito de caracterizar o fluxo do tráfego urbano de forma apropriada, Lana et al. (2016) desenvolveram um sistema que utiliza um processo de *clustering* para determinar padrões no tráfego a partir de dados de sensores de rua. Esses sensores são os ATRs (Automatic Traffic Recorder), que contam os veículos que passam sobre eles. Os dados passam por uma etapa de pré-processamento: com o agrupamento do fluxo do tráfego por dia, tem-se um conjunto de dados por iteração, após processo de limpeza. Para cada iteração há uma distribuição por dia do fluxo em diferentes partes do dia. Essa distribuição muda em dias especiais, como fins de semana e feriados, e são resultados do processo de *clustering*. Portanto, todos os dias no mesmo *cluster* irão caracterizar a mesma distribuição de fluxo de veículos. Dessa forma é possível diferenciar os dias de acordo com sua característica de tráfego. Na etapa de *clustering*, o algoritmo DBSCAN é utilizado para descobrir a relevância dos *clusters* identificados, classificando cada dia em grupos baseados em métricas de similaridade, com modelo que define as distâncias entre regiões para considerar quais dias se referem a quais *clusters*. Esse modelo e parâmetros são refinados a cara execução do processo, para melhor interpretação do analista. Posteriormente, o resultado desses processamentos é utilizado para execução de rotinas de previsão do tráfego.

Baseando-se em dados do uso de telefones celulares, Ratti et al. (2006) desenvolveram o projeto Mobile Landscape, que possibilita uma representação gráfica da intensidade das atividades urbanas e sua evolução pelo tempo e espaço, que, neste estudo, foi testado na cidade de Milão, Itália. A partir da quantidade de atividades de uso dos celulares, neste caso por

chamadas telefônicas medidas em Erlang¹², pode-se verificar a intensidade do tráfego em determinada posição por dia, ou qualquer outra medida (Gráfico 10).

Gráfico 10: Atividade de telefone celular em Erlang

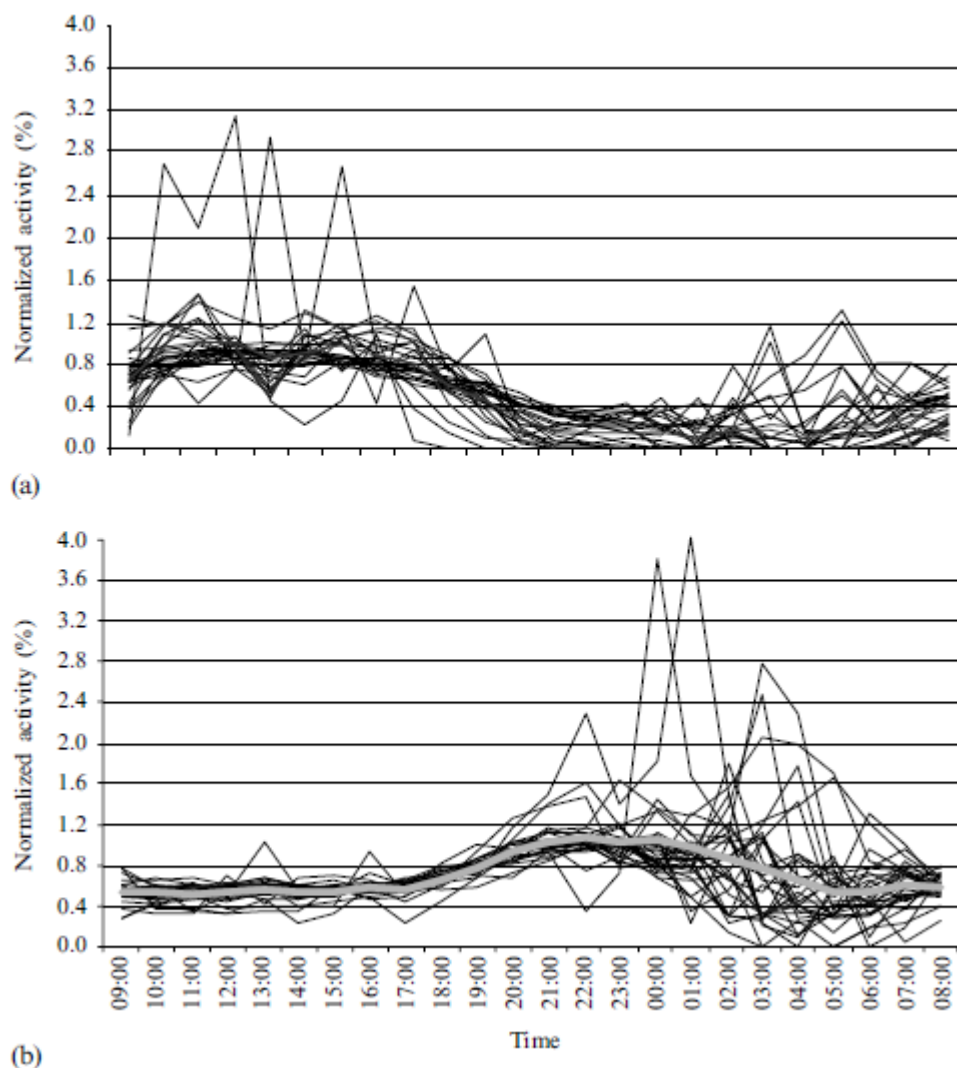


Fonte: (RATTI, FRENCHMAN, *et al.*, 2006).

O que se pode inferir dessa informação é a característica de uma determinada região. Uma região que apresenta grande quantidade de atividade durante o horário de expediente provavelmente é uma região com muitos escritórios, comércio, ou seja, uma área de negócios. Já uma região com grande concentração de atividade a noite e nas manhãs podem ser identificadas como vizinhanças residenciais. Porém, uma região residencial com considerável atividade durante o horário de expediente pode significar o aparecimento de situações de *home office*. Com a normalização do gráfico 10, dividindo-se pela quantidade total das chamadas em determinado horário, é possível verificar essa suposição. Para padronizar as assinaturas e permitir comparação, pode-se normalizar dividindo os resultados pela intensidade média de atividades. O Gráfico 11 mostra (a) mais intensidade de atividade durante o dia e (b) mais intensidade à noite. Essa análise pode auxiliar no descobrimento de tendências de atividades na cidade, possibilitando ações regulatórias (RATTI, FRENCHMAN, *et al.*, 2006).

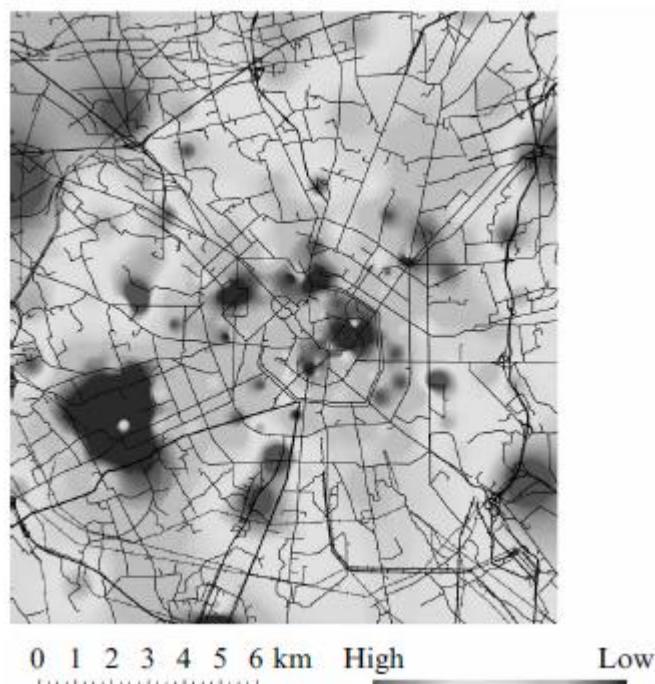
¹² Erlang é uma unidade de medida de intensidade de tráfego que não possui dimensão, utilizada em sistemas de telecomunicação. Um Erlang equivale à uma chamada telefônica de uma hora em um aparelho (RATTI, FRENCHMAN, *et al.*, 2006).

Gráfico 11: Atividades de celulares normalizados



Fonte: (RATTI, FRENCHMAN, *et al.*, 2006).

Com os dados das atividades normalizados pela média e a obtenção da intensidade relativa, é possível gerar uma apresentação da intensidade em mapa georreferenciado por representação de mapa de calor. Com a informação de localização das antenas de comunicação, determinam-se setores com a cobertura de sinal de cada antena em volta de um único centro, que seria a base da estação, divididos em regiões – normalmente uma região com três antenas – e apresentados no mapa (Figura 12).

Figura 12: Mapa de áreas com diferentes densidades

Fonte: (RATTI, FRENCHMAN, *et al.*, 2006).

Com essas informações, é possível criar representações com variações no tempo, verificando os locais onde a intensidade das chamadas é maior e em quais períodos. Também pode-se verificar essa atividade em locais estratégicos específicos, com grande atividade de deslocamentos de pessoas e baldeação de transportes públicos. Apesar de não poder afirmar a quantidade de pessoas no espaço e tempo baseando-se estritamente na quantidade de chamadas telefônicas, os padrões dessas chamadas podem ser correlacionados com a dinâmica geral, permitindo um panorama do comportamento da mobilidade urbana (RATTI, FRENCHMAN, *et al.*, 2006).

Angelakis et al. (2013) desenvolveram uma arquitetura para extrair conhecimento sobre a mobilidade urbana através dos dados da rede de telefonia celular. Esta arquitetura é flexível, genérica, configurável, mantém a integridade e privacidade dos usuários envolvidos, e possui uma interface para manter a interação com diferentes aplicativos e serviços. Conta com três camadas: Coleta, Análise e Exposição do Conhecimento, que são isoladas umas das outras, o que assegura a modularidade do processo, permitindo customizações de acordo com a necessidade (Figura 13).

Figura 13: Arquitetura de três camadas isoladas



Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

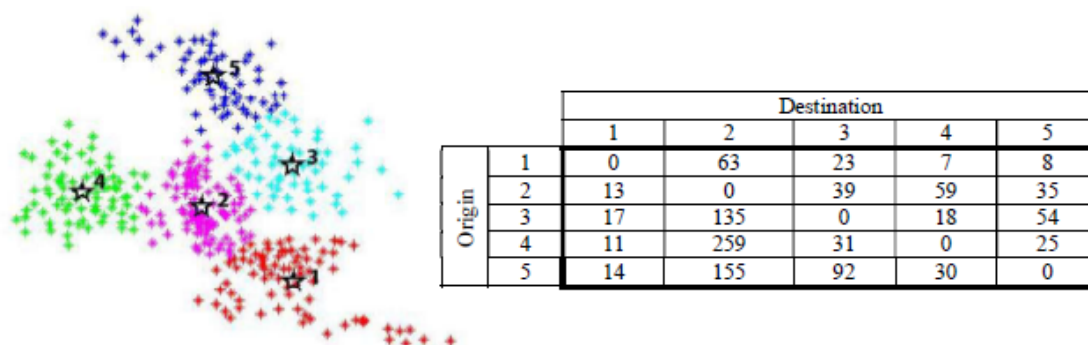
A camada de coleta tem como objetivo principal coletar os dados brutos das redes de telefonia móvel e transformá-los em dados uniformes para serem analisados. Portanto buscam os dados das diferentes fontes e converter os tipos diferentes de dados em formatos uniformes, resultando um conjunto de dados com a informação de localização dos usuários. Mais de um método de análise pode ser implementado na camada de análise, dependendo do tipo de estudo que se deseja. A camada de análise possui três principais funções: precisão analítica, processamento eficiente e preservação da privacidade. Após esse processo, a camada de apresentação disponibiliza os resultados para consulta e tomada de decisão (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Os autores destacam, neste estudo, as diferentes maneiras de se apresentar as análises efetuadas, que podem ser aplicadas às diversas necessidades no contexto de tomada de decisão quanto à mobilidade urbana. Uma das análises é a demanda de viagem, baseada na matriz de origem/destino. Dada a divisão de uma área geográfica em zonas, a matriz descreve as viagens ocorridas entre as diferentes zonas determinadas com a demanda de viagens em um determinado espaço de tempo (Figura 14).

Figura 14: Exemplo de uma matriz O/D

Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Um dos tipos de métricas de apresentação da demanda de viagens mencionada é o tipo estático, que fixa o tempo e as zonas analisadas e define origem e destino como os lugares mais comuns dos usuários, tais como casa e trabalho. Usando método *k-means* de formação de conjuntos é possível determinar as origens e destinos no mapa baseado nas antenas envolvidas, calculando deslocamentos entre diferentes regiões. A Figura 15 apresenta as regiões no mapa e a quantidade de deslocamentos entre elas nos períodos “casa” – das 22:00 até 07:00 – e “trabalho” – das 09:00 até as 16:00. Essa forma estática de análise permite a escolha de cenário pela pessoa que está utilizando o sistema, juntamente com a combinação com resultados de outros estudos. Todas as aplicações foram realizadas na cidade de Abidjan, Costa do Marfim.

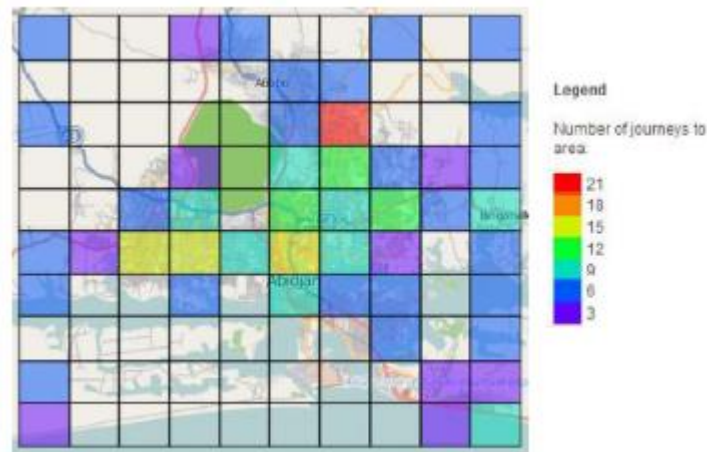
Figura 15: Clusters de regiões e sua quantidade de deslocamentos

Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Visando capturar quantas viagens forem possíveis pelos usuários, os autores apresentam o tipo de métrica dinâmico. Essa forma requer uma técnica que determina locais onde o usuário acaba se fixando, sem se locomover, contando posteriormente as viagens realizadas entre esses locais, podendo gerar a matriz O/D, no mesmo molde já mencionado, com deslocamentos em um período de, por exemplo, uma hora. O resultado contém as viagens realizadas para, a partir de, ou do fluxo interno de cada zona (Figura 16). Essa abordagem é importante em análises que

requerem uma resolução de espaço e tempo maiores, tais como diferenciação de tipos de deslocamento.

Figura 16: Exemplo de movimentações entre regiões



Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Retirando os locais fixos dos usuários, tratando cada par de locais acessados como uma viagem e reduzindo o tamanho das zonas, tem-se uma métrica que informa quantos usuários passam pelas regiões em um espaço de tempo específico, definida como métrica de fluxo de viagem. Os dados são agregados em dados de fluxo entre os locais que foram definidos de forma generalizada nas redondezas das antenas de comunicação. Os dados das chamadas telefônicas são ordenados cronologicamente para formar uma trajetória do usuário no espaço. Locais das redondezas das antenas são generalizados, formando polígonos que dividem a área explorada em um conjunto de regiões. As rotas, então, são formadas por deslocamentos entre cada par de regiões determinadas, contando o número de transições entre elas (Figura 17).

Figura 17: Fluxo de viagens entre regiões agregadas.

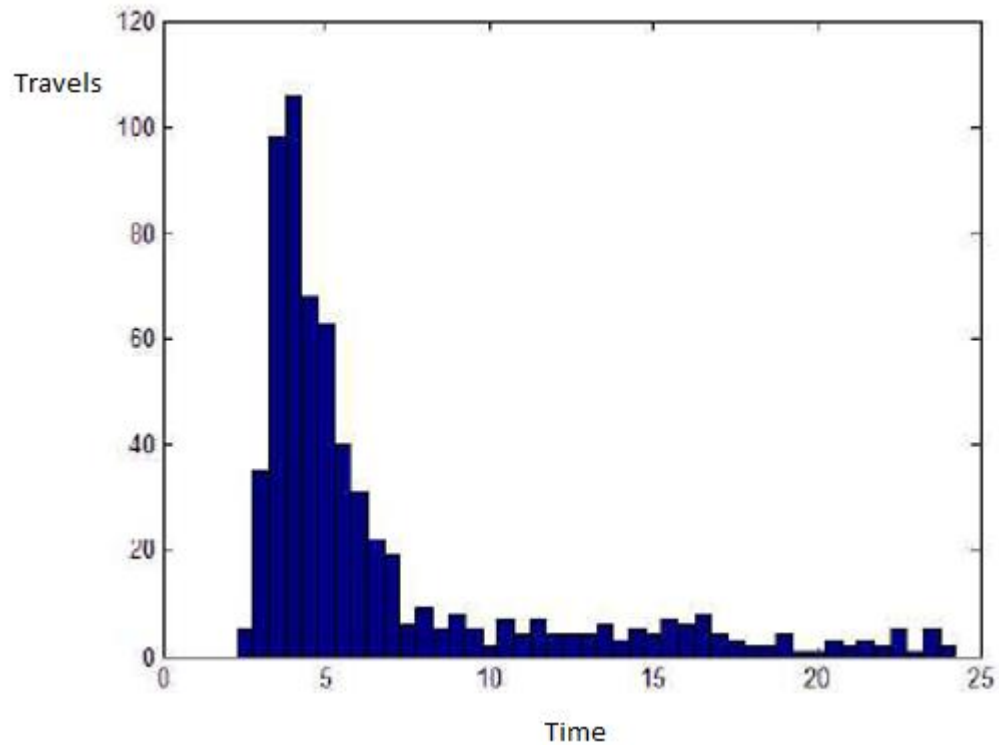


Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Obter locais através dos dados de antenas limita a análise, principalmente por requerer uma atividade de uso dos aparelhos. No entanto, é possível inferir tempos de viagens entre os pontos de forma satisfatória, sendo um bom indicador da qualidade da viagem. O Gráfico 12 apresenta um histograma com a quantidade de viagens entre duas regiões específicas por faixa de tempo de duração (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

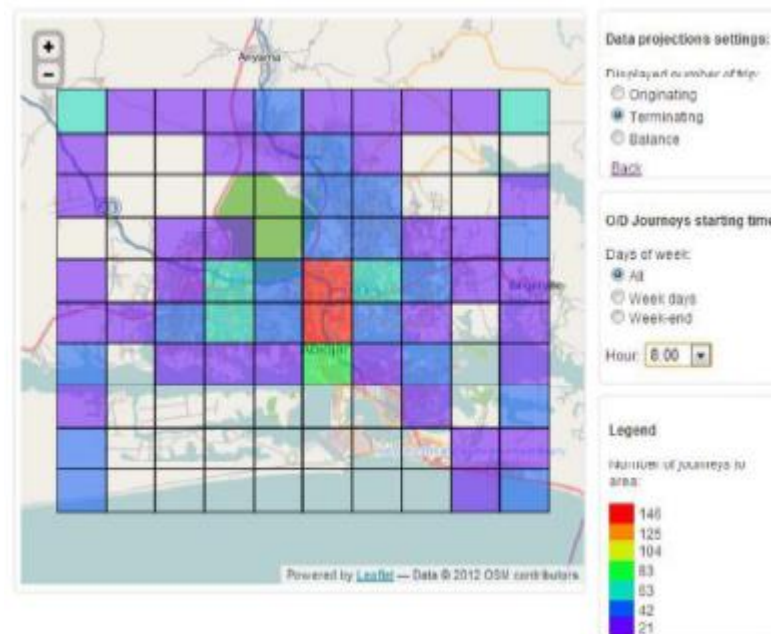
Para apresentar os resultados de forma acessível e inteligente, os autores desenvolveram uma ferramenta interativa, o Output Data Explorer, também disponível por comunicação via API, que permite configuração das fontes de dados, criação de novas rotinas analíticas, configuração de parâmetros dos algoritmos e a visualização gráfica dos resultados nas formas discutidas (Figura 18).

Gráfico 12: Quantidade de viagens entre duas regiões



Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Figura 18: Visão do Output Data Explorer



Fonte: (ANGELAKIS, GUNDLEGÅRD, *et al.*, 2013).

Foi possível verificar diversos trabalhos que endereçam o assunto análise de dados para mobilidade urbana. Notamos que isso é feito a partir de diversas perspectivas, ou seja, tratando

a mobilidade por vieses diferentes. Alguns trabalhos enfocam partes específicas do processo, o que traz um aprofundamento importante, possibilitando a evolução e aprimoramento das análises. Outros propõem arquiteturas completas, sugerindo um padrão de desenvolvimento a ser seguido.

Diferentes aspectos da mobilidade são tratados nas pesquisas mencionadas. Há variedade de fontes de dados e formas de processamento que possibilitam verificar dimensões distintas, ou, até mesmo, complementares da mobilidade, tais como a concentração de pessoas residentes na cidade e turistas, igualmente como o trânsito na figura do veículo particular e coletivo. O arcabouço teórico/empírico apresentado nos municia com informações valiosas para a proposta do *framework* pretendido.

6 PROCEDIMENTOS METODOLÓGICOS

Conforme discutido anteriormente, o trabalho pressupõe a possibilidade de aprofundamento do estudo da mobilidade urbana a partir da análise de dados das pessoas que se locomovem na cidade. Portanto, sob o pano de fundo da *smart city*, vislumbramos aprofundar os conceitos de mobilidade urbana e *data analytics*, trabalhos relacionados ao tema, realizar um experimento nos padrões da literatura estudada, e propor um *framework* de análise de dados dentro da abordagem seguida.

Para esse objetivo, buscamos uma metodologia que melhor auxiliasse na condução do estudo. Essa metodologia deve orientar a análise complementar e comparativa entre a literatura levantada e o pré-experimento realizado, visando produzir insumos para a formulação do *framework* objetivado.

Primeiramente, explicamos a caracterização da pesquisa como um todo, explorando suas bases ontológicas, epistemológicas e paradigmáticas. Posteriormente, apresentamos as etapas e métodos utilizados durante o trabalho, descrevendo a forma de condução dos desenvolvimentos e sua natureza.

Na terceira parte do capítulo, explicamos as categorias de análise consideradas no estudo, tais como a coleta de dados, organização e armazenamento, processamento, análise e publicação. Por fim, projetamos as possíveis limitações que o presente estudo pode apresentar, considerando a abordagem de cunho científico do trabalho.

6.1 CARACTERIZAÇÃO DA PESQUISA

O presente trabalho pretende formular um *framework* de análise de dados de mobilidade urbana que possibilita o desenvolvimento de uma arquitetura de análise de dados consistente e abrangente, permitindo o estudo dos hábitos de mobilidade dos indivíduos de uma cidade. Essa proposição se baseou nos trabalhos realizados por pesquisadores do tema, buscando-se técnicas de tratamento de dados, juntamente com arquiteturas e *frameworks* desenvolvidos dentro do cenário discutido. Além do levantamento bibliográfico, também foi realizado um pré-experimento desenvolvido a partir dos fundamentos da literatura levantada.

Estudando a literatura levantada, os pesquisadores compreenderam como o tema é tratado e abordado no seu estado-da-arte, como de fato são desenvolvidos e implementados os projetos de análise de dados de mobilidade urbana. Essa perspectiva permite o conhecimento de como o tema é conduzido na realidade, portanto possui um viés **realista**.

No pré-experimento realizado, desenvolvemos e implementamos um processo próprio de análise de dados, seguindo alguns aspectos da literatura. Apesar de seguir modelos propostos, houve deliberação em algumas decisões específicas, como parâmetros dos programas de processamento de dados desenvolvidos, que foram definidos com o intuito de melhor traduzir o comportamento identificado, aprimorando a visualização do resultado. Essas informações no sistema são fruto da opinião dos pesquisadores quanto ao melhor arranjo na visualização, visando tornar a interface mais amigável, melhorando a experiência. Essa abordagem possui um caráter de **interação sujeito-objeto**.

Conforme mencionado, o trabalho tem o intuito de formular uma proposta de *framework* com base nos requisitos levantados a partir da observação dos pesquisadores da literatura estudada e do pré-experimento realizado. Essa abordagem de pesquisa se orienta pelo que se desenvolve na realidade sobre o tema proposto e o que os pesquisadores perceberam e interpretaram como necessário e valioso. Logo, a pesquisa possui uma ontologia de **interação sujeito-objeto** na sua plenitude, pois entendemos que:

a realidade não é considerada nem como algo totalmente externo e independente da mente humana (objetiva), nem como fruto somente da percepção individual de cada indivíduo isoladamente (idealista, subjetiva), porém, a realidade é percebida e “criada” numa instância coletiva – as percepções do mundo que compartilhamos em sociedade (SACCOL, 2009, p. 252).

O conhecimento, portanto, é construído pelos pesquisadores, não apenas extraído de trabalhos anteriores sem questionar sua real finalidade ou de resultados obtidos através do pré-experimento que indiquem se determinada técnica é ou não satisfatória através de números. Buscamos entender o contexto em que estamos inseridos, pela observação da realidade analisada, e então realizar a proposta de acordo com a nossa experiência, acreditando que “não existe uma realidade objetiva esperando por ser descoberta. Verdades e significados só passam a existir a partir do nosso engajamento com o mundo” (SACCOL, 2009, p. 253).

Destacamos a orientação do trabalho pelo paradigma **interpretativista** de pesquisa, que em suas etapas propositivas ressalta métodos de natureza qualitativa ao analisar os

conhecimentos estudados e experiências vivenciadas. O trabalho busca a combinação de aspectos **objetivos e subjetivos**. As técnicas de coleta e levantamentos são caracterizadas pelo aspecto **fenomenológico**, pois não se fixam em questões a serem descobertas ou que possam emergir da análise, mas no dado em si. Mas ao se observar o conteúdo estudado, resultados de ações experimentais realizadas e fazer proposições, a visão **interpretativista** é predominante (GIL, 2008) (SACCOL, 2009).

6.2 DELINEAMENTO E MÉTODOS DE PESQUISA

Em consonância com os objetivos apresentados na introdução, a dissertação consiste em quatro etapas de pesquisa: Revisão bibliográfica, campanha de *crowdsensing*, levantamento de requisitos e proposta do *framework*.

Na primeira etapa, a revisão bibliográfica descrita no item 5.2 do documento, foram pesquisadas teorias e propostas de *framework* na literatura que pudessem auxiliar para a maior compreensão do assunto discutido, principalmente com informações de arquiteturas e *frameworks* já desenvolvidos por outros pesquisadores pertinentes à proposta pretendida. Esses estudos levantados deveriam fornecer insumos para o desenvolvimento de um processo e implementação de programas computacionais para o estudo. Neste ponto, buscamos entender quais técnicas são utilizadas e como implementá-las para viabilizarmos o pré-experimento subsequente.

Esta etapa foi conduzida através de **revisão narrativa** dos estudos científicos publicados. Com base na literatura estudada que abrange a temática da mobilidade urbana sob o pano de fundo da *smart city* como um todo, buscamos artigos científicos que discutissem os termos “*smart city*” e “*urban mobility*”, “*big data platform*”, “*data analytics process*”, “*dbscan*”, “*extract, transform and load*”, “*crowdsensing*”, “*Intelligent Transportation Systems*”, artigos identificados através de referenciais teóricos, e considerados artigos que discutissem esses temas com a mobilidade urbana. As bases de dados científicos que indexaram mais artigos relevantes foram Scopus e EBSCO. Devido à grande quantidade de literatura internacional sobre o tema, a pesquisa foi realizada com termos em inglês para melhor assertividade de resultados. Escolhemos esse método devido à abrangência do tema, tanto no que se refere a arquiteturas de análise de dados já desenvolvidas, quanto a estudos de análise

da mobilidade urbana. Esse aspecto limitou uma verificação mais completa e minuciosa dos trabalhos realizados, mas ainda se apresenta como uma técnica de revisão bibliográfica valiosa.

Portanto, a etapa de levantamento bibliográfico possui características de uma pesquisa **exploratória**, visando proporcionar uma visão geral do assunto, e foi realizada com uma **leitura interpretativa** dos textos, avaliando a qualidade, aderência e pertinência à presente dissertação, de acordo com os fundamentos que oferecem (AMBONI, 1997) (GIL, 2008).

A segunda etapa se refere ao pré-experimento da aplicação de uma campanha de *crowdsensing* com pessoas envolvidas na pós-graduação da ESAG/UDESC. O objetivo foi realizar um *crowdsensing* para coletar dados de localização dos usuários envolvidos, submetê-los a rotinas de organização e processamento de dados, e por fim analisar o conhecimento gerado através de representações em mapas. Todos os processos e técnicas desenvolvidas para tanto foram baseadas nas literaturas estudadas, principalmente em técnicas de *crowdsensing* e determinação de conjuntos de dados em forma de *clusters*.

Esta etapa contou com a aplicação de um questionário, através da ferramenta de formulários do Google, para interpretar como o usuário se envolveu com a campanha, seu entendimento e expectativas. Com esse pré-experimento, foi possível realizar testes funcionais e técnicos das ferramentas envolvidas, avaliando os métodos e processos utilizados, refinando parâmetros e interpretando resultados apresentados.

A campanha foi realizada somente uma vez, com um número limitado de participantes por problemas técnicos e de engajamento, e não contou com uma equipe que executasse o teste em ambiente controlado. Portanto, interpretamos esta etapa como um **pré-experimento**, cujos resultados e observações dos pesquisadores são utilizados complementar e comparativamente com outras fontes, neste caso, a revisão bibliográfica. É uma pesquisa **descritiva**, pois utilizou técnicas padronizadas de coleta e análise de dados, buscando descrever as características dos elementos envolvidos no processo, sem interferir na realidade observada a fim de modificá-la. Além do *crowdsensing*, outros métodos de pesquisa utilizados foram o experimento, **análise estatística** e **questionário** (RUDIO, 2007) (GIL, 2008).

Algumas necessidades técnicas e de processo, e novas funcionalidades foram identificadas na observação da literatura estudada e do pré-experimento realizado. A partir disso, na terceira etapa, foi possível realizar um **levantamento de requisitos** que servirá como diretriz para a formulação do *framework* desejado. O levantamento de requisitos é uma

metodologia da Engenharia de Software aderente à característica do atual trabalho pois é utilizada em desenvolvimentos de *softwares* e refere-se às necessidades e funcionalidades que um sistema deve, ou não, apresentar. Os requisitos funcionais descrevem o comportamento do sistema, o que eles devem fazer, como reagir em determinadas situações e até mesmo o que não fazer. Os requisitos não funcionais não estão diretamente relacionados com os serviços oferecidos pelo sistema, ou seja, com as suas funcionalidades. Dizem respeito a propriedades do sistema, como restrições, capacidades, desempenho, entre outros aspectos. A utilização desta metodologia depende da característica do desenvolvimento, portanto, seguimos uma postura de elaboração de requisitos de usuário, descrito de forma abstrata, considerando uma linguagem natural e comunicativa com o público não especialista na área da computação (SOMMERVILLE, 2011).

Não adotamos uma conduta de obediência absoluta às regras e práticas do levantamento de requisitos técnicos, pois não buscamos desenvolver e implementar uma arquitetura ou sistema computacional, o que exigiria maior critério nesta etapa. Apenas nos inspiramos no método pela sua aderência ao tema. Logo, esta etapa buscou fornecer substratos que fundamentem a proposta pretendida nas dimensões quali e quantitativa. Além do embasamento teórico extraído da realidade estudada e os resultados obtidos das análises da pré-experimentação, os requisitos surgiram da observação e interpretação dos pesquisadores das etapas anteriores. Se caracteriza como uma pesquisa **interpretativa**, com métodos de **comparação e observação** (SACCOL, 2009).

Por fim, a etapa de proposta do *framework* possui um caráter avaliativo de pesquisa. Ela consiste na apresentação formal de um *framework* que realiza estudos de mobilidade urbana idealizado pelos pesquisadores, seguindo o perfil identificado nos estudos que seguem o tema e como este é representado. Sendo uma fase propositiva, onde os pesquisadores qualificam as suas experiências e observações nos levantamentos realizados nas etapas anteriores, é pertinente seguir a perspectiva do tipo de pesquisa **avaliativa**, que se apoia na análise que as mudanças sugeridas pelo estudo podem causar na realidade observada (CARO, 1982).

A fim de proporcionar uma visão geral da metodologia seguida, o Quadro 2 apresenta o quadro resumo das características metodológicas relativas a cada etapa da pesquisa:

Quadro 2: Resumo da característica metodológica das etapas

Etapa	Tipo de pesquisa	Método
Pesquisa de teorias e propostas de <i>framework</i> na literatura	1. Exploratória	1. Pesquisa bibliográfica 2. Leitura interpretativa
Pré-experimento de campanha de <i>crowdsensing</i>	1. Descritiva	1. <i>Crowdsensing</i> 2. Experimento 3. Análise estatística 4. Questionário
Levantamento de requisitos	1. Interpretativa	1. Comparativo 2. Observacional
Elaboração de <i>framework</i> para análise de dados	1. Avaliativa	1. Comparativo 2. Observacional

Fonte: Quadro elaborado com base em Amboni (1997) , Saccol (2009), Gil (2008) e Caro (1982).

6.3 CATEGORIAS DE ANÁLISE E DIMENSÕES

A pesquisa considerou analisar, nas etapas mencionadas anteriormente, as seguintes dimensões que competem ao tema em pauta, principalmente pela pertinência com estudos de mobilidade urbana e com o processo de *data analytics* na abordagem de proposta de um *framework*:

Coleta de dados: são as técnicas para coletar dados para a análise, como obtê-los, de quais fontes, de quais tipos, com que frequência, entre outras características. Esta categoria fornece os insumos brutos recolhidos do mundo real para que o mecanismo de análise possa gerar conhecimento.

Organização e armazenamento: Diversos tipos de dados podem ser coletados e devem ser armazenados de forma padronizada para viabilizar um processo de análise sistemático. Essa dimensão compreende processos de organização de dados, como limpeza e aglutinação, e métodos de armazenamento.

Processamento: Apesar de organizados, os dados ainda não oferecem clareza para serem interpretados. O processamento se encarrega de executar rotinas que transformam os dados em informações passíveis de análise, abrangendo diversos métodos e técnicas de processamento de grande quantidade de dados.

Análise: As informações devem ser apresentadas de forma amigável para a interpretação de um analista. Esta dimensão considera, portanto, as várias formas de apresentação dos dados processados e como elas podem auxiliar na tomada de decisão, gerando conhecimento muitas vezes desconhecido.

Publicação: O conhecimento produzido pela análise deve ser disseminado para agentes interessados e para a sociedade. Métodos e veículos de comunicação e divulgação do estudo realizado são importantes para que isso ocorra.

6.4 POSSÍVEIS LIMITAÇÕES DA PESQUISA

Como o levantamento bibliográfico foi realizado de forma narrativa, ou seja, na busca não sistematizada de literatura nas bases de dados, alguns estudos certamente foram ignorados. Portanto, algum conhecimento oportuno pode ter sido desconsiderado nas avaliações. A pouca literatura nacional também limita o conhecimento sobre a condução do assunto no Brasil, principalmente na relação com entidades fontes de dados e cidadãos.

No pré-experimento foi possível realizar testes funcionais do processo de análise, conforme mencionado, além de testes técnicos, importantes para o aprimoramento da ferramenta de *crowdsensing* desenvolvida. Apesar de fornecer material importante para observação e avaliação dos pesquisadores, experimentos controlados e robustos, que analisem a qualidade dos resultados acerca da mobilidade urbana, certamente permitiriam maior capacidade de interpretação e proposta.

Esse cenário foi prejudicado por limitações de engajamento na coleta de dados e pela falta de tempo dos pesquisadores, certamente pretendemos uma maior quantidade de dados passíveis de análise. Por conta de limitações técnicas, o pré-experimento não contou com a automatização desejada de todos os processos, portanto, algumas etapas foram executadas manualmente.

Apesar de utilizar a técnica de levantamento de requisitos como orientação metodológica, o trabalho não adota total concordância ao método e não segue um padrão de planejamento de projeto de desenvolvimento de sistemas computacionais. Esse caminho despende grande quantidade de tempo e exige uma equipe engajada. O trabalho não propõe uma implementação de arquitetura de sistemas completa, que demandaria um levantamento de requisitos funcionais e não funcionais consistente, inclusive contando com informações de cunho técnico. Ainda assim, serve como um panorama em uma possível implementação.

O estudo considera a mobilidade urbana de forma geral e propõe métodos e processos que se enquadrariam em cenários diferentes de aplicação. No entanto, a diversidade de dados e contextos podem não ser comportados pelo *framework* vislumbrado, necessitando adequações.

7 **FRAMEWORK: UMA PROPOSTA**

Apresentamos neste capítulo o pré-experimento de campanha de *crowdsensing* realizada, a característica do seu desenvolvimento, seus resultados de análise e do questionário aplicado aos usuários participantes. Após isso, listamos os requisitos levantados a partir do levantamento bibliográfico das pesquisas de modelos de análise de dados de mobilidade urbana e da experiência vivenciada na campanha de *crowdsensing*, tanto pela perspectiva dos pesquisadores quanto dos participantes. Com base no levantamento de requisitos, elaboramos o *framework* de análise de dados para mobilidade urbana, objetivo geral deste estudo.

7.1 DESCRIÇÃO PRÉ-EXPERIMENTAL DA CAMPANHA DE *CROWDSENSING*

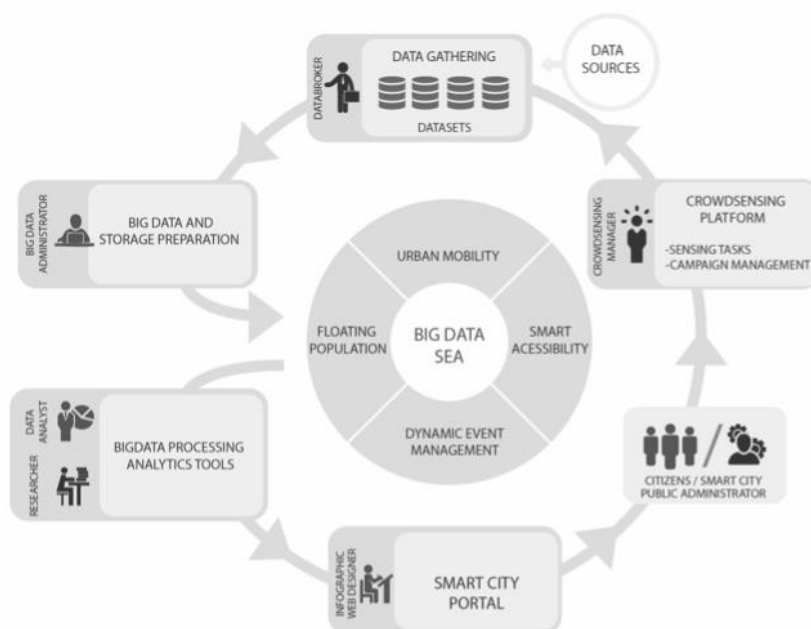
Para contribuir com estudos na área de *smart city* e vislumbrar substratos para uma solução para problemáticas de mobilidade urbana, realizamos uma campanha como teste de uma aplicação de *crowdsensing* com alunos e professores da pós-graduação da ESAG/UDESC de Florianópolis, no período de 15/05 até 14/06 de 2017. Os participantes foram cadastrados no sistema ParticipACT Brasil e convidados, via e-mail, apresentação nas salas de aula, redes sociais, entre outros meios de comunicação, para participar da campanha, para o qual deveriam instalar o aplicativo no dispositivo móvel e aceitar a campanha para terem seus dados de localização monitorados, juntamente com outros requisitos técnicos, tal como manter o sensor de GPS do celular ativo. Dos 112 potenciais participantes, 50 usuários aderiram de fato à campanha. Destes, 27 enviaram dados de localização corretamente durante boa parte da campanha, número que foi reduzido posteriormente para 15 usuários por conta de deficiências técnicas da aplicação ou do próprio dispositivo móvel do participante. Os dados coletados são armazenados no servidor do ParticipACT Brasil para posterior análise.

7.1.1 ParticipACT Brasil, a plataforma de *smart city* utilizada

Para apoiar e suportar a campanha de *crowdsensing* realizada, utilizamos a plataforma já mencionada ParticipACT Brasil. Esta plataforma é uma iniciativa do Laboratório de

Tecnologias de Gestão, LabGES, da ESAG/UDESC, que busca contribuir com a expansão de conhecimento dos problemas urbanos e encontrar soluções para se construir uma *smart city*, usando de tecnologias como Big Data. Ele integra dados de diferentes bases, tais como agências de governo, instituições de ensino e coleta por *crowdsensing*, com ferramentas de análise de dados ao final da coleta. Também permite pesquisas acadêmicas em três frentes: acesso a uma grande quantidade de dados; envolvendo o cidadão na colaboração participativa; legitimando e justificando decisões da esfera pública e privada na solução de problemas urbanos. A visão sistêmica do funcionamento pode ser verificada na Figura 19:

Figura 19: Representação do fluxo dos dados

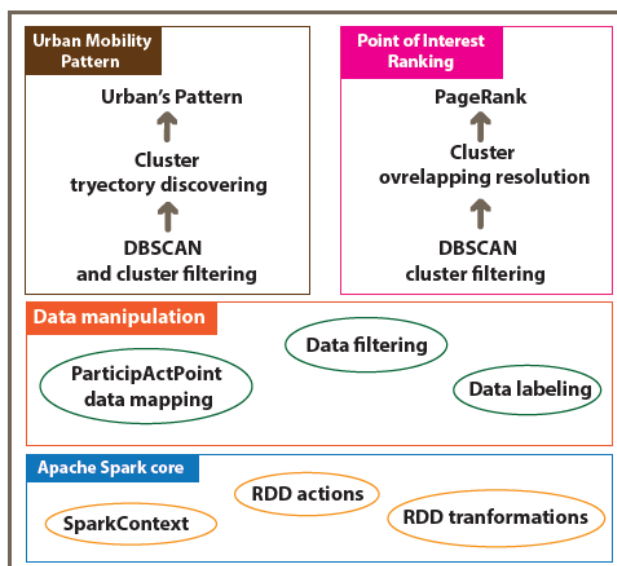


Fonte: PARTICIPACT BRASIL. **Tecnologias inovadoras na gestão da cidade inteligente.** Disponível em: <<http://labges.esag.udesc.br/participact/>>. Acesso em: mar. 2018.

Para este estudo específico, estamos utilizando como coleta de dados somente o *crowdsensing* para a área da Mobilidade Urbana. No entanto, a plataforma vislumbra o estudo de outras problemáticas urbanas, tais como a População Flutuante, a Acessibilidade Inteligente e o Gerenciamento de Eventos Dinâmico, com dados de diferentes fontes, conforme explicado anteriormente nos trabalhos de Corradi et al. (2015a). Para analisar os dados coletados,

desenvolvemos uma ferramenta de análise de dados, baseada na plataforma de desenvolvimento Apache Spark¹³ (Figura 20).

Figura 20: Arquitetura para desenvolvimento de análise de dados.



Fonte: (ZITO e CORRADI, 2017).

7.1.2 Métodos de processamento de dados e análise

Dois tipos diferentes de análise são realizados para os dados de GPS coletados dos participantes da campanha de *crowdsensing*. Na primeira análise, chamada de **UMP (Urban mobility Pattern)**, nos interessa encontrar as regiões da cidade onde os usuários permanecem por maior tempo e construir um mapa que represente o padrão de mobilidade desses usuários. Na segunda análise, chamada de **PIR (Point of Interest Ranking)**, nós propusemos refinar a detecção de *clusters* através de uma dupla execução da rotina DBSCAN¹⁴ com diferentes parâmetros, atribuindo uma pontuação para cada *cluster*, baseada no número de conexões com as áreas externas a ele e ordenando-os por importância, considerando as trajetórias dos usuários.

¹³ Plataforma de desenvolvimento que atua como um mecanismo de processamento de dados em larga escala (ZITO e CORRADI, 2017).

¹⁴ Density-Based Spatial Clustering of Applications with Noise. Algoritmo que compara densidades de pontos de localização dentro e fora dos *clusters* determinados (ESTER, KRIEGL, *et al.*, 1996).

Um algoritmo de *PageRank* foi ajustado para o trabalho com *clusters*, contabilizando a quantidade e qualidade de conexões apontando para cada um deles.

Clustering é o processo de agrupamento de um conjunto de objetos físicos ou abstratos em classes de objetos similares. Um *cluster* é caracterizado pelos objetos similares entre si e não similares do restante (HAN, PEI e KAMBER, 2011). DBSCAN (Density-Based Spatial Clustering of Applications with Noise) é um algoritmo que analisa se a densidade de pontos internos de um *cluster* é maior que de um existente do lado de fora. O conjunto de dados que não faz parte de nenhum *cluster* é chamado de barulho (*noise*). Para se determinar um *cluster*, a ideia é que cada um deve ter sua própria vizinhança dentro de determinado raio, com um mínimo número de pontos (ESTER, KRIEGEL, *et al.*, 1996).

A ferramenta faz a leitura do dado bruto, no formato .csv, baixado do servidor do ParticipACT Brasil. Cada registro do arquivo representa um ponto de GPS coletado. Esses pontos coletados são filtrados de acordo com a acurácia do dado, ou seja, de acordo com a fidelidade do seu posicionamento identificado. Os registros são, então, mapeados e transformados em um novo objeto Java, num novo formato, para poderem ser utilizados pelo sistema desenvolvido em Java.

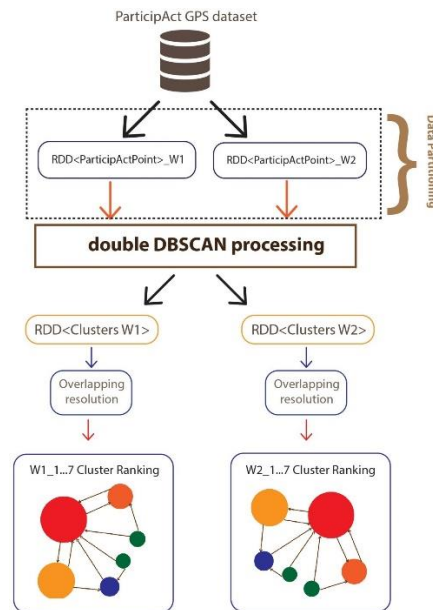
Na análise **UMP**, o programa separa os dados por períodos, definidos em manhã (7:01 às 12:00), representado pela cor vermelha no mapa, tarde (12:01 às 18:00), representado pela cor laranja, noite (18:01 às 23:00), representado pela cor azul, e madrugada (23:01 às 7:00), representado pela cor cinza. São definidos grupos de índice a partir do agrupamento de dias, uma parte contando como as três primeiras semanas de campanha (semanas 20, 21 e 22), chamada aqui de recorte W1, e outra com as demais semanas (23 a 25), recorte W2. Cada pedaço de dados, ou seja, período e grupo, - podendo ser, por exemplo, segunda-feira de manhã – conceituamos como uma partição de dados. Para cada partição de dados, é executada a rotina DBSCAN para identificar os conjuntos de posição, chamados aqui de *clusters*. Cada *cluster* é uma área onde há bastante concentração de pontos, onde as pessoas, representadas por posições, estão mais frequentemente concentradas. Na rotina DBSCAN, são informados como parâmetros o *eps*, que representa a distância entre os pontos dentro de um *cluster*, e o *min number*, que define a quantidade mínima de pontos em cada *cluster*.

Com essa rotina, sabemos os pontos que estão aproximados e quais são os *clusters* encontrados. Após isso, o programa determinou o ponto central de cada *cluster*, calculando o

ponto médio entre latitude e longitude dos pontos extremos de cada *cluster*, para que se possa utilizar um ponto de conexão entre eles. Comparamos os diferentes pontos dos usuários para determinar interconexões: se dois *clusters* compartilharem usuários em comum, significa que há conexão entre eles. O programa realiza essa rotina com *clusters* da mesma partição ou de partições com períodos seguidos – como, por exemplo, manhã com tarde. Finalmente, foi desenvolvida uma ferramenta *web-based* que apresenta os dados determinados em um mapa do Google Maps, através de uma página *web*.

Diferentemente da análise UMP, a análise **PIR** não faz a separação por período, apenas divide os dados coletados por grupo de índice. Neste caso, dividimos cada partição de dados como cada dia útil da semana das primeiras três semanas (20 a 22), recorte W1, e cada dia da semana das três últimas semanas (23 a 25), recorte W2. Nesta análise propusemos uma abordagem baseada na dupla execução da rotina DBSCAN, realizada em cada partição de dados com dois pares diferentes de parâmetros para identificar *clusters* com grande densidade e grande variância de usuários. Esta também é uma solução viável para corresponder qualquer ponto de interesse possível com uma densidade pequena e grande variância de usuários.

Essa abordagem pode gerar resultados com *clusters* duplicados ou sobrepostos, exigindo um processo posterior de eliminação; para cada *cluster* determinado, uma pontuação é calculada, definida pelo número de usuários que o cruzaram. Esse cálculo é então usado para computar uma métrica, chamada *ratio*, que é uma relação entre a pontuação e a extensão do *cluster*, ajudando na estimativa do quanto ele representa. Para cada par de *cluster* cruzado, o programa compara se o número de pontos de localização na intersecção foi maior que 20% dos pontos de localização do *cluster* maior. Se for o caso, o *cluster* menor pode ser considerado para representar a região se apresentou um *ratio* maior. Isso ocorre pois um *cluster* de tamanho menor que tenha um maior *ratio* contribui mais para região. Após esse passo, a rotina *PageRank* foi utilizada para ranquear e organizar os *clusters*. Por fim, a informação é apresentada no Google Maps através da ferramenta mencionada anteriormente. Na Figura 21 podemos visualizar o fluxo de informações da análise PIR.

Figura 21: Fluxo de informações do módulo de análise de dados PIR

Fonte: (ZITO e CORRADI, 2017).

O método descrito neste tópico e os conceitos que o fundamentaram no que diz respeito às plataformas e ferramentas adotadas e à escolha pela utilização de *clusters* para inferir a concentração de pessoas são explicados com mais detalhes no trabalho de Zito et al. (2017).

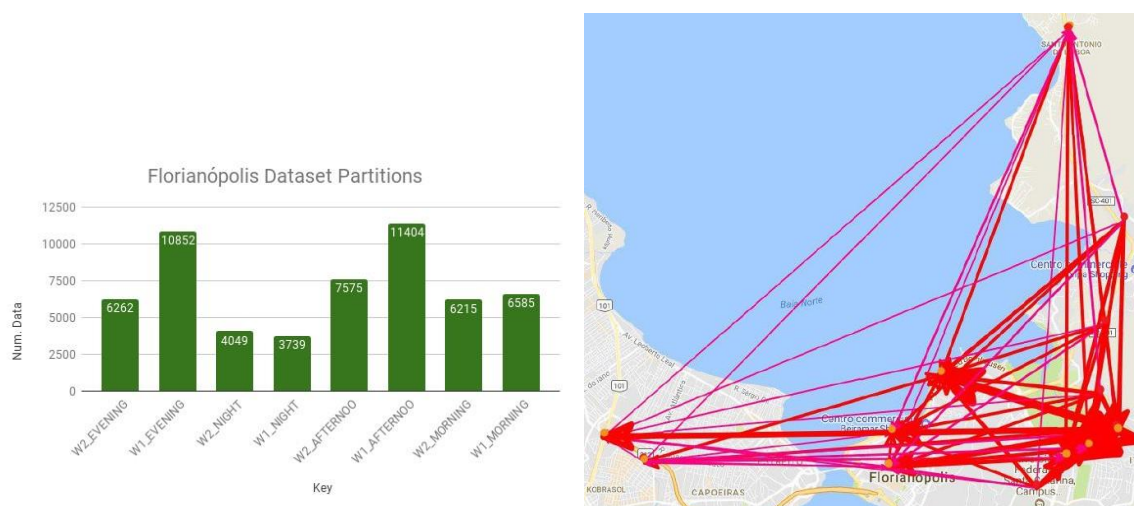
7.1.3 Resultados experimentais

Os procedimentos de organização de dados, tais como limpeza e alguns ajustes necessários, assim como o armazenamento posterior ao recebimento do dado bruto no servidor do ParticipACT Brasil foram realizados manualmente para cada situação. Apesar de não contarmos com ferramentas para esse fim que estejam devidamente configuradas e implementadas, esses processos não foram automatizados, mas são fundamentais para a operação.

O sistema foi utilizado para executar a análise **UMP** nos dados de localização coletados a partir dos voluntários da campanha em Florianópolis, conforme explicado. Os dados são divididos por recortes (W1 e W2) e por período do dia. O histograma faz a relação desses períodos com sua respectiva quantidade de dados, permitindo a identificação de 35 pontos de interesse espalhados pela cidade, conectados por 291 trajetos, onde 10 padrões de movimentação podem ser extraídos dos resultados.

O resultado é apresentado em uma interface *web-based* intuitiva, baseada no Google Maps, mostrando os padrões detectados. Esse mapa permite que administradores da cidade observem o deslocamento das pessoas no seu dia-a-dia, usando cores para representar os trajetos identificados de acordo com o período do dia correspondente e se são individuais ou compartilhados por vários usuários (Figura 22).

Figura 22: Distribuição da quantidade de dados e trajetos W1



Fonte: Elaborado pelos pesquisadores do laboratório LabGES, 2017.

Para se ter uma perspectiva de concentração das pessoas, podemos visualizar a distribuição dos pontos de localização na forma de *heatmap*. Através dele, podemos verificar a variação da intensidade das cores entre os diferentes lugares onde as pessoas transitaram com base na determinação dos *clusters* realizada, inclusive permitindo inferências sobre vias mais utilizadas (Figura 23). Mesmo não detalhando as trajetórias que os usuários adotaram é uma forma interessante de representar a concentração de pessoas de maneira geral.

Figura 23: Heatmap de concentração de pessoas



Fonte: Elaborado pelos pesquisadores do laboratório LabGES, 2017.

Para detectar as áreas mais relevantes de visitação dos usuários envolvidos, a análise **PIR** foi executada para os dados da campanha em questão. O resultado identificou e ranqueou 33 pontos de interesse, que foram representados no Google Maps. Cores diferentes representam regiões diferentes geradas pelo processo de *clustering* das partições de dados processadas. Essa representação permite uma clara e rápida visualização de onde as pessoas costumam se concentrar durante a semana, com ao UDESC e o centro da cidade. Esse resultado se confirma, principalmente pelo perfil dos voluntários envolvidos na campanha (Figura 24).

Figura 24: Distribuição dos dados e pontos de interesse (continua)

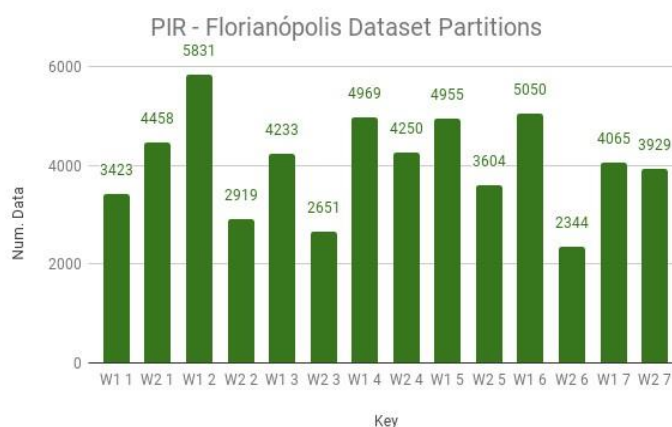


Figura 25: Distribuição dos dados e pontos de interesse (conclusão)



Fonte: Elaborado pelos pesquisadores do laboratório LabGES, 2017.

Com o intuito de divulgar o estudo realizado e seus resultados, utilizamos alguns meios de comunicação disponíveis que apresentavam um poder de disseminação aceitável para o público envolvido. Desenvolvemos um *banner* com informações sobre os objetivos do projeto ParticipACT Brasil, o resultado do estudo realizado e os apoiadores (Anexo 1), que foi exposto na sala do laboratório na ESAG/UDESC. Um e-mail de agradecimento e breve explicação foi enviado para os envolvidos, juntamente com o *banner* mencionado.

7.1.4 Questionário aplicado

Ao final da coleta dos dados na campanha de *crowdsensing*, foi enviado, via e-mail, um questionário criado através da ferramenta Google Forms para os participantes envolvidos. Dentre os 50 que participaram da campanha, 24 responderam ao questionário, que contou com questões de múltipla escolha e abertas. O resultado do questionário não irá pautar as propostas do trabalho, nem é objetivo utilizá-lo como etapa de coleta de dados que irá fundamentar o estudo, será apenas mais um aspecto considerado na experiência da etapa de campanha de *crowdsensing*. Por conta disso, iremos apenas discorrer, neste capítulo, sobre nossas interpretações das respostas coletadas, permitindo sua verificação em detalhes no anexo 2.

O objetivo da aplicação do questionário era entender a característica do envolvimento dos participantes, abordando a sua compreensão com relação ao propósito da campanha, quais eram os reais motivadores da sua participação, que expectativas possuíam antes de se engajarem

e alguns aspectos da sua experiência na prática da campanha. Essas informações são úteis para se ter uma visão da dinâmica da campanha de *crowdsensing* sob a perspectiva dos próprios participantes. Essa visão auxilia na melhoria do processo com relação à divulgação, formas de incentivo, à interface e experiência do usuário, vislumbrando posteriores execuções com um público mais abrangente.

As apresentações em salas de aula e propaganda boca-a-boca foram os meios de divulgação e comunicação mais efetivos da campanha conforme as respostas, seguidos pelo e-mail, sem tanta representatividade. Por se tratar de um público reduzido para os padrões de uma campanha de *crowdsensing*, esse tipo de abordagem foi possível ser feita.

Mais da metade das pessoas (62,5%) entenderam se tratar de um teste de um aplicativo de *crowdsensing* – real objetivo do estudo – mas por volta de 30% pensaram estar participando de um projeto de monitoramento dos padrões de mobilidade. Esse último dado nos atenta para a falha na comunicação da campanha, pois não desejamos relacionar nenhum tipo de monitoramento com a proposta do estudo. O projeto ParticipACT conta, inclusive, com uma política de privacidade dos dados, atestando que “tanto as bases de dados que irão compor o Big Data quanto os participantes da campanha de *crowdsensing* terão sua segurança e privacidade garantidos, desde a coleta e armazenamento até a divulgação da informação, respeitando-se o anonimato dos dados” (PARTICIPACT BRASIL). No entanto, 12,5% das pessoas afirmam não se sentir confortável com a disponibilização dos seus dados de localização, mesmo que para fins acadêmicos. O restante não vê nenhuma restrição neste sentido.

Como expectativa de *feedback*, os participantes gostariam de receber os resultados do estudo, com o conhecimento gerado sobre a mobilidade urbana, com as regiões mais problemáticas, características da demanda do transporte público, propostas de melhorias, entre outros. Esperavam receber um serviço por parte da plataforma, como informações sobre os meios de transporte, otimização de rotas, problemas e eventos urbanos nas regiões. Algumas questões técnicas e de design de interface com o usuário, como dificuldade no login, notificações, apresentação de resultados via aplicativo e diminuição do consumo de bateria, também foram sugeridos pelos participantes.

A maioria das pessoas (86,4%) informou que se engajou na pesquisa para ajudar no atual estudo do projeto de dissertação de mestrado e por volta de metade das pessoas também admite ter sido motivada pela oportunidade de auxiliar no teste do aplicativo de *crowdsensing*

do ParticipACT e em estudos de problemáticas de mobilidade urbana. Esse dado mostra a importância de uma campanha de incentivo para sensibilizar as pessoas quanto ao projeto. Por fim, aproximadamente 30% das pessoas perceberam um aumento no consumo da bateria durante o período em que a campanha esteve ativa.

7.2 LEVANTAMENTO DE REQUISITOS

Com base nas literaturas estudadas sobre plataformas ou arquiteturas de análise de dados de mobilidade urbana e nas observações dos pesquisadores durante a experiência de campanha de *crowdsensing* realizada, com posterior processamento e análise de dados, podemos formular requisitos que municiam a elaboração do *framework* de análise de dados de mobilidade urbana. O levantamento de requisitos é uma metodologia para desenvolvimento de sistemas que especifica o comportamento externo do sistema, sem detalhamentos técnicos. Nos inspiramos nesse método pela sua aderência com o assunto, mas como não é intenção deste estudo desenvolver uma arquitetura ou sistema computacional completo, apesar de o *framework* objetivado permitir vislumbrar essa hipótese, adotamos uma postura de elaboração de requisitos de usuário. Dessa forma, foi possível produzir diretrizes que orientam e embasam a proposta desejada, conservando uma comunicação clara e abrangente.

De acordo com a norma IEEE (ISO, IEC, 2011), o levantamento de requisitos compreende funções intersetoriais que realizam a mediação entre quem está adquirindo e quem está fornecendo um sistema, software ou serviço, considerando a “descoberta, dedução, desenvolvimento, análise, determinação de métodos verificadores, validação, comunicação, documentação e gerenciamento de requisitos”. Segue a seguinte hierarquia de requisitos: “Permite um entendimento acordado entre *stakeholders* [...]; é validado contra necessidades do mundo real, podendo ser implementado; fornece designs de base de verificação e aceitação de soluções” (ISO, IEC, 2011).

Os **requisitos não funcionais** possuem relação com propriedades do sistema, restrições de implementação, características gerais do sistema, necessidades do produto e da organização, entre outros aspectos (SOMMERVILLE, 2011). Identificamos os seguintes requisitos não funcionais para a elaboração do *framework*:

- A organização que irá se valer do *framework* deve desenvolver parcerias com entidades de dados abertos, com interfaces tecnológicas de comunicação de dados, para que haja uma extração sistemática de dados a serem analisados;
- A organização deve formular mecanismos de disseminação e incentivo ao uso do sistema por parte das pessoas, principalmente do *front-end*¹⁵ do aplicativo de celular, para que possam se engajar na iniciativa tanto como fornecedores de dados quanto consumidores do conhecimento gerado;
- Devem ser desenvolvidas políticas de parceria com agências públicas responsáveis ou interessadas pelo ambiente urbano, para que o conhecimento gerado pelas análises e os apontamentos através do sistema pelos usuários participantes enquanto habitantes do meio urbano possam chegar a esse público por um canal confiável. No sentido contrário, para que quaisquer projetos de melhorias e retornos dessas agências possam ser consultados através do sistema pelos cidadãos;
- As campanhas de *crowdsensing* a serem disparadas devem ser bem planejadas, com propósitos claros e bem definidos, evitando o desperdício de coleta de dados e eventual desinteresse;
- As interfaces com os usuários devem ser claras, intuitivas, amigáveis e apresentar características gráficas atraentes;
- O acesso aos dispositivos móveis dos participantes deve manter uma latência ótima, a fim de conseguir coletar os dados da melhor forma e não exigir que os dispositivos despendam de muitos recursos;
- O *framework* deve manter uma dinâmica de funcionamento sistemática, permitindo recorrência de análises para que os cidadãos envolvidos possam contar com conhecimento atualizado constantemente;
- O *framework* deve apresentar uma característica de sistema escalável, seguro, rápido e confiável;
- Deve garantir a privacidade dos dados coletados das bases de dados de entidades e de pessoas que contribuírem com dados pessoais, apenas utilizando-os para fins de estudo;

¹⁵ *Front-end* é um tipo de aplicação por onde o usuário interage diretamente com o sistema (TECHTARGET NETWORK).

- Dados que exprimem opiniões de pessoas, tais como texto, foto ou vídeo, devem ser submetidos a um processo de moderação por parte de operadores.

Os **requisitos funcionais** descrevem o que o sistema deve fazer e devem ser elaborados de acordo com o tipo do desenvolvimento que se pretende realizar. Podem possuir uma característica mais abstrata quando expresso como requisitos de usuário – forma adotada no presente estudo – ou informações mais detalhadas, buscando uma especificação mais técnica, variando entre requisitos gerais, que abrangem todo o sistema, e específicos, com detalhes de funcionamento (SOMMERVILLE, 2011). Os seguintes requisitos funcionais foram levantados:

- O *framework* deve coletar dados de mobilidade urbana de fontes diversas, como dados de sensores de dispositivos móveis – como localização via GPS, *wifi*, acelerômetro, foto, vídeo –, dados de acesso à internet via 4G, POIs fornecidos por plataformas *web*, *hotspots* de *wifi*, *roaming* e registros de chamadas via celular nas antenas de comunicação;
- Os aplicativos de dispositivos móveis devem possuir mecanismos de economia de energia com consciência de contexto e de sistema, evitando possíveis conflitos com outros aplicativos;
- Reconhecer atividades das pessoas via sensores de celular, com inteligência adaptativa e evolutiva para calibração e aprimoramento;
- A plataforma de crowdsensing deve permitir campanhas participativas, com ações diretas dos usuários envolvidos, e oportunistas, coletando dados de acordo com o contexto em que os usuários se encontram;
- Coletar dados de transporte público, como localização dos ônibus, registros de cobrança de passageiros, pontos de parada de ônibus, rotas por linha e sensores de tráfego urbano;
- Coletar os dados das agências fontes de dados via APIs ou outras interfaces de comunicação e armazená-los em catálogos de dados próprios;
- Filtrar os dados imprecisos, *outliers*, faltantes e redundantes;
- Integrar os diferentes tipos de dados, com rotinas que interpretam as características dos dados, realizando as correspondências e estruturação necessárias para que possam ser utilizados em análises comparativas ou complementares;
- Validar e processar dados que sejam conflitantes após a integração, evitando barulhos nas etapas de processamento e análise;

- Transportar e armazenar os dados pré-processados em locais específicos para posterior processamento e análise, mantendo um fluxo automático de utilização dos dados;
- Identificar os modais utilizados pelos usuários através de sensores dos celulares, considerando a velocidade de deslocamento;
- Processamento inteligente de identificação de regiões com maior concentração de pessoas;
- A partir de campanhas específicas de crowdsensing, com público direcionado e janela de tempo definida, o framework deve identificar as rotas principais, via agrupamentos de rotas, e modais preferenciais, via identificação de POIs e interpretação de sensores dos dispositivos;
- Analisar o desempenho do transporte público, considerando o tempo despendido nos percursos e o que foi planejado pelo departamento de trânsito;
- Visualização da quantidade de pessoas concentradas por região e período;
- Permitir a comparação entre a concentração de pessoas em determinadas regiões com a quantidade/localização de pontos de ônibus disponíveis na cidade;
- Visualização da quantidade de viagens que ocorrem entre as diferentes regiões da cidade e o tempo que se despende nos deslocamentos;
- Analisar a concentração de turistas na cidade em comparação com atrativos e infraestrutura turística;
- Disponibilizar a consulta dos resultados por parte das pessoas via aplicativo de celular, portal *web* e mídias sociais, com possibilidade de escolha entre os diferentes tipos de análise e filtros de seleção;
- Possuir uma funcionalidade de interação do usuário com o projeto via aplicativo celular de reclamação de problemas urbanos identificados, como um canal de exercício da cidadania.

7.3 APRESENTAÇÃO DO *FRAMEWORK*:

Neste ponto, apresentamos a idealização do trabalho, um *framework* para análise de dados de mobilidade urbana e classificação de pessoas por hábito de mobilidade. Após muita reflexão no estudo dos trabalhos relacionados e no pré-experimento da campanha de

crowdsensing, e posterior orientação dos requisitos de *software* levantados, é possível vislumbrar um *framework* para tal fim.

Não é objetivo da proposta fornecer uma arquitetura completa para pronta implementação de um *big data analytics*, mas sim substratos que orientem tal projeto. Chamamos o produto de *framework* pela aderência do seu significado à proposta intencionada: “uma estrutura de suporte pela qual algo pode ser construído; um sistema de regras, ideias ou crenças que é usado para planejar ou decidir alguma coisa” (CAMBRIDGE DICTIONARY).

Sua estrutura respeita as categorias de análise mencionadas no item 6.3 deste documento. São elas a **coleta, organização e armazenamento, processamento, análise e publicação**. Na etapa de coleta, os dados são extraídos das fontes desenvolvidas, como as pessoas envolvidas e entidades parceiras. Na organização e armazenamento, o dado é submetido a rotinas de limpeza e organização, e são armazenados em locais alvo de acordo com sua posterior utilização. O processamento realiza rotinas de determinação de conjuntos e correlações entre variáveis complementares. Na análise, essa informação é organizada e apresentada de forma clara, via ferramentas de visualização, para que possam ser interpretadas, gerando conhecimento e auxiliando em tomadas de decisão. Visando disseminar esse conhecimento, o *framework* faz uso de vias de comunicação via internet facilitando o acesso ao público interessado.

O fluxo de funcionamento é sistemático, ou seja, frequentemente, de acordo com a periodicidade de cada análise, os dados são coletados e utilizados no processo. Esse modelo de trabalho é importante para esse tipo de estudo, pois a mobilidade urbana é sazonal e sofre alterações com o passar do tempo por conta da característica dinâmica das cidades. Também é cíclico para que as próprias pessoas envolvidas no processo de coleta de dados, como cidadãos e entidades interessadas, possam receber o conhecimento que foi gerado ao final das análises. Portanto, o conhecimento sobre a mobilidade urbana estará sempre acessível para as pessoas e atualizado (Figura 25).

Figura 26: Fluxo macro do processo do *framework*



Fonte: Elaboração própria, 2018.

Existem quatro tipos de análise que irão nortear o fluxo de dados e informações no *framework*. São elas:

- **Análise MCS:** este tipo de análise se baseia em dados gerados via campanhas de **Mobile Crowdsensing**. Nele, o *framework* apresenta uma análise com conexões entre regiões mais frequentadas, com pontos de interesse, que expressam os locais com maior concentração de pessoas, e de padrões de mobilidade, com as rotas mais utilizadas e meio de transporte;
- **Análise CPT:** baseada em dados de companhias telefônicas, como chamadas telefônicas, e de plataformas de turismo, a análise CPT (**Concentração de População e Turistas**) apresenta a concentração de pessoas, a quantidade de viagens entre regiões determinadas e a comparação da concentração de turistas com locais turísticos da cidade;

- **Análise DTP:** visando avaliar o **Desempenho do Transporte Público**, apresenta uma análise da pontualidade dos ônibus com base no planejamento e uma comparação da demanda de passageiros de ônibus com terminais de ônibus. Se baseia em dados da companhia de transporte coletivo da cidade;
- **Análise PU:** este portal interativo permite o reporte de **Problemas Urbanos** por parte dos usuários e os apresenta de forma amigável no mapa da cidade, juntamente com informações adicionais explicativas.

A partir dessa orientação, podemos descrever com detalhes o que cada categoria do *framework* irá desempenhar.

7.3.1 Coleta

De acordo com os tipos de análise mencionados, os dados coletados serão organizados em quatro grupos, cada um direcionado a cada tipo de análise. Os dados destinados à análise MCS são coletados dos sensores dos aparelhos celulares dos usuários participantes de campanhas de crowdsensing gerenciadas pela plataforma ParticipACT Brasil. Durante as campanhas, os dados de GPS, localização de *wi-fi* e acelerômetro são armazenados no servidor para serem utilizados e sua frequência de coleta deve ser a cada 5 minutos, visando minimizar o consumo de energia do aparelho.

Para o tipo de análise CPT, os dados são coletados de companhias de telefone e de plataformas de turismo, como o TripAdvisor. Via desenvolvimento de APIs de comunicação com os sistemas das agências mencionadas, os dados são armazenados em um catálogo no servidor próprio do *framework* para posterior extração e utilização. Serão coletados dados de chamadas telefônicas – medidas em Erlang –, *hotspots* de *wi-fi* e acesso via internet 4G pelas companhias telefônicas, e POIs da plataforma de turismo, com uma periodicidade mensal.

Também coletados através de APIs e armazenados em catálogos mensalmente, os dados utilizados para análise DTP têm origem nas companhias de transporte coletivo. São eles os dados de planejamento de transporte, como rotas, linhas e pontos de parada de ônibus e terminais, e GPS e AFC – dados de pagamento de passageiros – de ônibus.

Os dados da análise PU são coletados via aplicativo de *crowdsensing*, pela mesma plataforma ParticipACT Brasil já mencionada. Ao informar um problema urbano identificado, o sistema coleta o tipo do problema reportado no aplicativo, além da localização via GPS e possíveis informações de texto, foto ou vídeos enviados pelo usuário. As campanhas que gerenciam essa coleta são abertas, de tempo indeterminado, ou seja, estão sempre em funcionamento, e são participativas, pois necessitam uma intervenção direta dos usuários.

Para simplificar a visualização da etapa de coleta, elaboramos um quadro resumo com as características de cada bloco de dados coletados (Quadro 3):

Quadro 3: Características da etapa de coleta

Dados	Fontes	Finalidade	Frequência	Via de coleta
GPS, <i>wi-fi</i> , acelerômetro	Dispositivos móveis	Análise Mobile Crowdsensing	A cada 5 minutos durante a campanha	Aplicativo ParticipACT Brasil
Chamadas telefônicas, <i>hotspots</i> , 4G e POIs	Companhias telefônicas e plataformas de turismo	Análise Concentração de População e Turistas	Mensal	API
Rotas, linhas, pontos de parada, GPS e AFC de ônibus	Companhias de transporte público	Análise Desempenho de Transporte Público	Mensal	API
GPS, texto, foto, vídeo	Dispositivos móveis	Análise Problemas Urbanos	A cada 5 minutos	Aplicativo ParticipACT Brasil

Fonte: Elaboração própria, 2018.

7.3.2 Organização e Armazenamento

Os dados coletados devem ser organizados de certa forma para posterior utilização nas rotinas de processamento. Conforme mencionado na etapa de coleta, estes são armazenados em locais que permitem acesso por ferramentas internas do sistema (Figura 26). O *framework* possui um passo de **extração** que recolhe os dados desses locais de acordo com a sua periodicidade de análise e disponibiliza-os para uso. Portanto, os dados coletados via dispositivos celulares são recolhidos das bases de dados do servidor onde foram armazenados e outros dados recebidos através de APIs são extraídos dos respectivos catálogos, respeitando a frequência de uso estipulada.

Por se tratar de dados brutos, ou seja, que foram coletados direto da fonte com o mínimo de tratamento possível, os dados são submetidos a um processo de **filtro**. São desconsiderados os seguintes dados: incorretos, como dados não solicitados, preenchidos em desacordo com a característica do campo ou corrompidos; *outliers*, que apresentam realidades muito distantes da normalidade, principalmente para dados de localização; e redundantes, como dados duplicados ou que representam a mesma coisa. Este passo também filtra dados que possam inferir a identificação de pessoas envolvidas, como dados pessoais ou de identificação de aparelhos, residência, veículo, excluindo-os da base. Apenas dados de identificação que mantêm uma unicidade para o tratamento e não inferem identificação pessoal são conservados, assegurando a anonimização dos dados.

Esta manutenção da unicidade do dado é importante para o passo de **integração**. Nele, dados de tipos diferentes e complementares são combinados para possível correlação nas etapas posteriores. Em outras palavras, dados utilizados para um mesmo fim, na mesma linha de análise, são integrados considerando-se o contexto em que se encontram. Dados utilizados para as análises MCS e PU são integrados por *timestamp*, ID – identificação do usuário mantendo a privacidade – e localização. Dados para a análise CPT, além de considerarem os mesmos conteúdos na integração, também sofrem o tratamento contra o efeito *ping-pong*. Neste momento, dois registros que contém informação para o mesmo ID de mesmo local e *timestamp* são interpretados como um só dado. Os dados para a análise DTP sofrem integração respeitando a correlação entre os elementos estáticos do planejamento de transporte público e dos dados de sensores dos ônibus.

Por fim, após organização dos dados, estes são armazenados em bases ou locais específicos para posterior utilização de acordo com o tipo de análise aos quais estão destinados. Esta é a etapa de **armazenamento**, que assegura o fluxo sistemático do funcionamento do

framework, pois os dados estão posicionados em locais pré-determinados e são direcionados para o uso por parte das rotinas de processamento, mantendo os dados acessíveis para os elementos e com a frequência necessária. Portanto, quatro locais são mantidos para a finalidade de armazenamento após a organização dos dados.

Figura 27: Fluxo da etapa de organização e armazenamento



Fonte: Elaboração própria, 2018.

7.3.3 Processamento

Nesta etapa, a enorme quantidade de dados é processada por algoritmos que os interpretam e extraem informações a partir deles. Através de computação distribuída e processamento paralelo, métodos estatísticos são aplicados para extrair, classificar e refinar dados que estão aparentemente perdidos e desestruturados, afim de identificar o conteúdo que realmente expressam.

Inicialmente, um processo de identificação e formação de conjuntos, aqui chamado de **CLUSTERING**, é executado para todos os dados, exceto para dados utilizados na **análise PU**, pois estes não requerem nenhum tipo de processamento. As rotinas de *clustering* são baseadas no algoritmo **DBSCAN** já mencionado, considerando um conjunto com base em sua densidade

em detrimento de demais pontos circunvizinhos pouco relevantes. Cada *cluster* representa locais com maior concentração de pessoas, onde se fixaram ou percorreram com mais intensidade, e são definidos com a orientação dos parâmetros de distância entre pontos de um *cluster* e a quantidade mínima de pontos em cada um. No *framework* sugerido, este processamento é realizado para concentração de pessoas em locais específicos e para rotas mais utilizadas.

Após a identificação dos *clusters*, alguns processamentos são realizados de acordo com o tipo de análise interessado. Para a **análise MCS**, que visa analisar a mobilidade urbana a partir de dados coletados em campanhas pontuais de *crowdsensing*, considerando rotas e concentração de pessoas, o *framework* contempla:

- **Conexões entre pontos:** Após a determinação de pontos que estão aproximados e os respectivos *clusters*, calcula-se o ponto central de cada um pelo ponto médio entre latitude e longitude dos extremos, indicando de onde parte a conexão. Para todos os *clusters* que possuem usuários em comum em um mesmo período, ou em período subsequentes, tais como manhã e tarde, define-se a interconexão;
- **Rotas preferenciais:** Através de campanhas de *crowdsensing* fechadas, com público determinado, tempo limitado e oportunistas de acordo com a localização, são determinadas as rotas de cada usuário por período do dia, considerando as conexões entre regiões determinadas no processamento anterior e os pontos de deslocamento entre essas regiões. *Clusters* de rotas são, então, identificados, correspondendo às rotas mais densas e preferidas pelos participantes;
- **Reconhecimento de atividade:** Com os dados de acelerômetro coletados, é possível identificar de que forma as pessoas estão se deslocando a partir da velocidade de deslocamento. Logo, para cada rota preferencial determinada, o tipo de modal preferido é correlacionado, podendo ser a pé, bicicleta ou veículo motorizado;
- **Pontos de interesse:** São as regiões que apresentaram maior concentração de pessoas. Para este processamento, a rotina DBSCAN de determinação de *clusters* é executada duas vezes, uma para identificação de clusters com pequena densidade e outra para grande densidade. Como certamente gerará *clusters* sobrepostos, uma relação de pontuação é definida de acordo com o número de usuários que cruzaram por eles e sua extensão. Se a quantidade de pontos de *cluster* menor for maior que 20% dos pontos do *cluster* maior, então ele é o escolhido para representar a região, pois tem mais

contribuição na concentração de pessoas. Posteriormente, uma rotina de PageRank é executada para ranquear e organizar os *clusters*.

Entre as rotinas de processamento para o tipo de **análise CTP**, buscando interpretar a concentração de pessoas via utilização do aparelho celular e de plataformas de turismo, estão as seguintes:

- **Concentração de população:** São as regiões com maior concentração de pessoas. Como os dados coletados para este processamento não possuem coordenadas geográficas, determina-se as regiões de acordo com a posição da antena utilizada pelo aparelho. Se houver mais de uma antena, o espaço intermediário é considerado. A rotina de *clustering* é então executada para definir quais regiões são mais relevantes para a concentração;
- **Quantidade de viagens:** Aqui são calculadas a quantidade de viagens entre regiões. Para calcular a quantidade de viagens que ocorreram no espaço urbano analisado, regiões são definidas através de representação de grade no mapa, de acordo com o parâmetro de distância definido pelo operador. Cada usuário que cruza regiões diferentes em um mesmo período ou em períodos subsequentes contabiliza uma viagem entre as regiões em questão;
- **Concentração de turistas:** O mesmo processamento realizado para a concentração da população é executado neste ponto, com a diferença do dado utilizado, que aqui são considerados dados de chamadas em *roaming*, ou seja, de dispositivos de cidades diferentes da cidade em questão.

Para a **análise DTP**, através de dados de planejamento e realização dos veículos de transporte públicos, o *framework* apresenta os seguintes processamentos:

- **Planejado/Realizado:** Esse processamento realiza a determinação do desempenho de cada linha de ônibus por rota a ser analisado posteriormente. Os dados de percurso dos ônibus são organizados de acordo com as rotas que realizaram nos períodos, o tempo decorrido e a velocidade. O mesmo é executado para os dados que representam o que foi planejado para cada linha;
- **Demanda de transporte:** Identifica as regiões com maior demanda de passageiros. Com base nos dados de AFC e GPS dos ônibus, determinam-se os pontos onde houve registro de embarque de passageiros. Após isso, esses pontos são submetidos ao

processo de *clustering* para determinar as regiões mais relevantes para demanda de passageiros.

A figura 27 apresenta uma disposição ilustrativa das rotinas executadas na etapa de processamento:

Figura 28: Representação da etapa de processamento



Fonte: Elaboração própria, 2018.

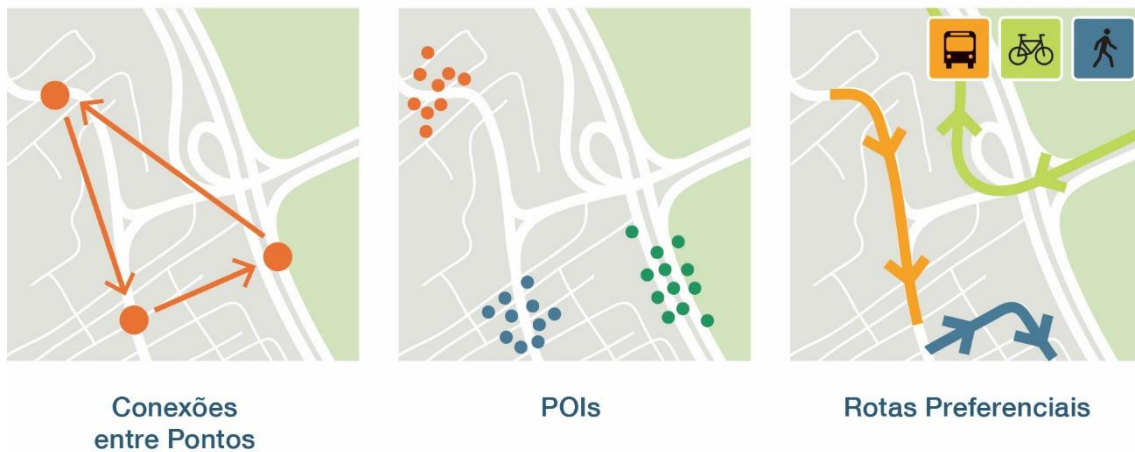
7.3.4 Análise

Os dados processados devem ser organizados e apresentados de tal forma que possam ser analisados, interpretados e que ajudem em tomadas de decisão. Ou seja, conjuntos de *clusters* que foram determinados devem ser apresentados nos seus locais específicos, suas devidas conexões, rotas, correlação com outros dados, tudo considerando o mapa do local correspondente. Conforme comentado desde o início da proposta do *framework*, os tipos de

análise orientaram a estruturação das outras etapas. Portanto, agora descrevemos como serão apresentadas as análises de cada tipo, começando com as **análises MCS** (Figura 28):

- **Conexões entre pontos:** Cada interconexão identificada para os *clusters* determinados, a partir das coletas de *crowdsensing*, são apresentadas no mapa. Logo, essa análise mostra locais de concentração, identificados por apenas uma execução da rotina DBSCAN de determinação de *cluster*, com suas interconexões correspondentes indicadas por uma seta;
- **POIs:** Nesta análise, os pontos de interesse dos participantes, determinados via dupla execução do algoritmo DBSCAN com parâmetros diferentes, são apresentados no mapa. Esses locais expressam, portanto, os lugares onde mais pessoas permaneceram por maior quantidade de tempo durante a campanha de *crowdsensing* correspondente;
- **Rotas preferenciais:** Após a identificação das rotas mais utilizadas no período de sensoriamento, os respectivos modais utilizados podem ser correlacionados. Portanto, para cada rota preferencial, o modal mais utilizado – determinado através do processamento de reconhecimento de atividade – é referenciado. As rotas são apresentadas no mapa e cada uma conterà a informação do modal correspondente.

Figura 29: Representações das análises MCS



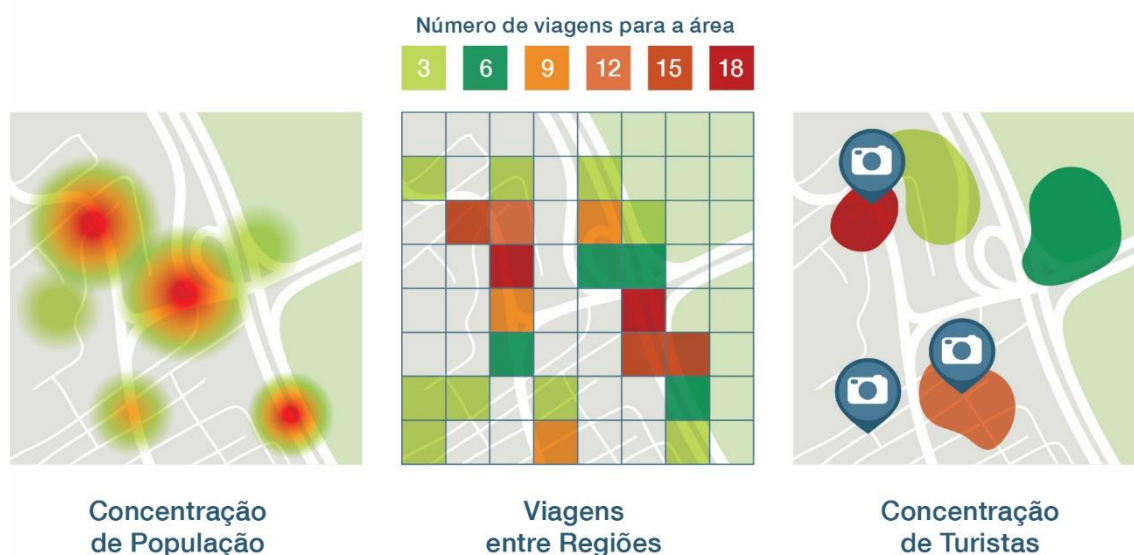
Fonte: Elaboração própria, 2018.

As análises do tipo **CPT** (Figura 29) serão apresentadas da seguinte forma:

- **Concentração de população:** Os dados de concentração são apresentados no formato de mapa de calor. Após a determinação dos clusters, as regiões mais populosas são informadas no mapa considerando a densidade como diferenciação de coloração;

- **Viagens entre regiões:** Nesta análise, o mapa é representado por uma grade que separa o espaço em regiões, identificadas anteriormente na etapa de processamento. Os quadrados representantes de cada região são apresentados com colorações diferentes correspondendo a quantidade de viagens até aquela região, ou seja, contando a região como o destino, conforme legenda explicativa;
- **Concentração de turistas:** Os locais turísticos coletados das plataformas de turismo e as regiões de concentração de pessoas de fora da cidade são apresentados no mapa. Essa análise permite a visualização comparativa da concentração de turistas com os locais turísticos cadastrados.

Figura 30: Representações das análises CPT



Fonte: Elaboração própria, 2018.

Para as **análises DTP** (Figura 30) temos:

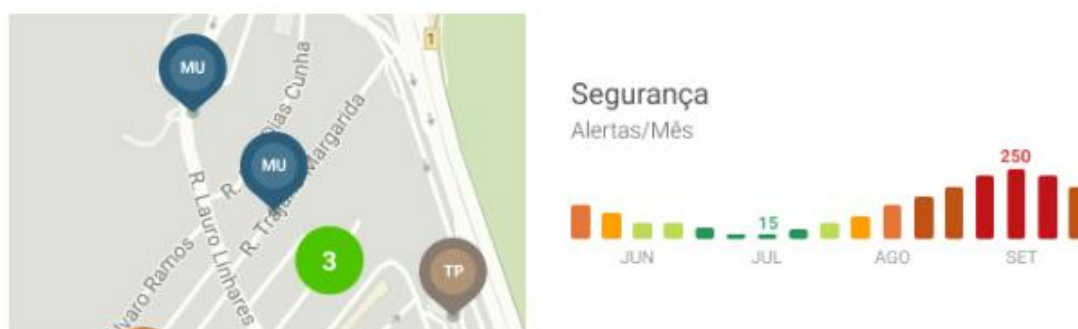
- **Desempenho:** O que foi realizado em cada rota de ônibus, de acordo com velocidade e tempo, é apresentado no mapa, respeitando um filtro de seleção do cenário que se quer visualizar. Os tempos planejados entre cada parada de ônibus são correlacionados com o que foi realizado. A coloração da rota analisada é apresentada de acordo com uma coloração, respeitando o desempenho de acordo com atrasos, adiantamentos e desempenho normal por cada trecho;
- **Demanda de passageiros:** As regiões identificadas como maior concentração de passageiros são apresentadas no mapa juntamente com a localização dos terminais de ônibus da cidade, permitindo uma análise comparativa entre esses elementos.

Figura 31: Representações das análises DTP

Fonte: Elaboração própria, 2018.

O tipo de **análise PU** (Figura 31) contém somente uma análise:

- **Problemas urbanos:** Essa análise não necessita de rotina de processamento, pois recebe diretamente os dados de reporte de problemas urbanos dos participantes e usuários do aplicativo e os apresenta no mapa, juntamente com outros dados, tais como fotos, vídeos e textos correspondentes, também informados pelas pessoas. É possível gerar gráficos com os dados de acordo com a quantidade das ocorrências e o tipo do problema.

Figura 32: Representações da análise PU

Fonte: Elaborado pelos pesquisadores do laboratório LabGES, 2018.

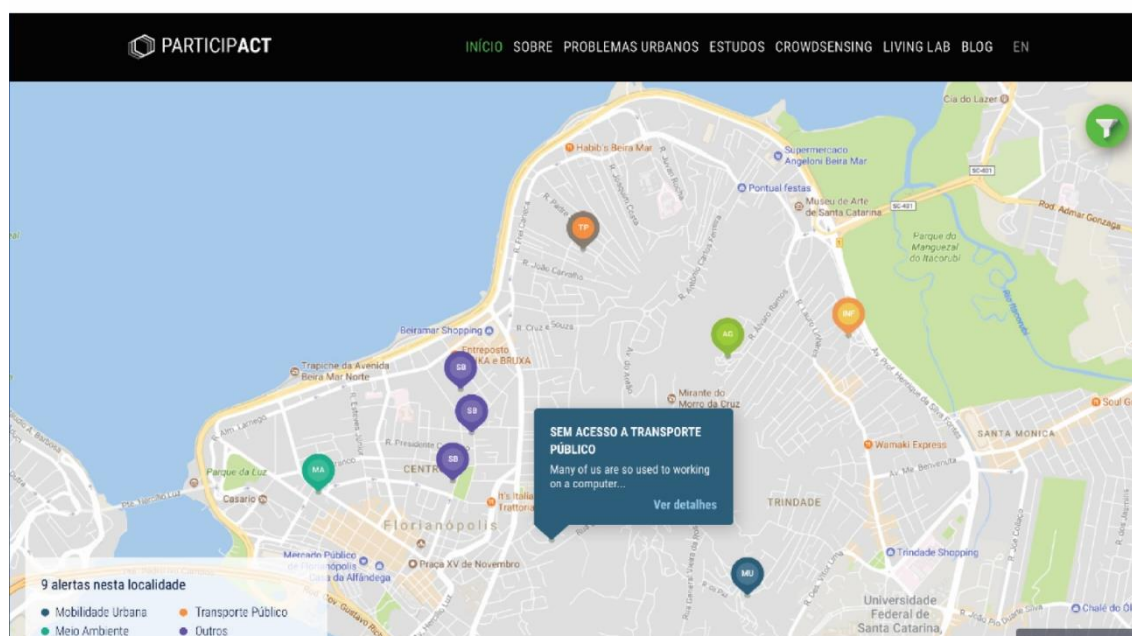
7.3.5 Publicação

Para que possam ser acessados e utilizados pela comunidade, respeitando a característica cíclica do *framework*, os resultados das análises são disponibilizados em plataformas amigáveis e atraentes (Figura 32). Nessas plataformas, os usuários são capazes de

visualizar as análises que desejarem, quais resultados querem verificar e escolher o filtro de seleção do cenário, possibilitando o acesso a todos os tipos de análises efetuadas. Isso pode ser realizado através de um portal *web*, por onde os usuários interagem via *browser* de internet, e pelo mesmo aplicativo por onde participam das campanhas de *crowdsensing* e reportam os problemas urbanos, através de aparelhos celulares. Neste trabalho, contamos com a plataforma ParticipACT Brasil, já utilizada no pré-experimento no item 7.1 deste documento.

Outra forma de disponibilizar os resultados é através de interfaces de comunicação computacional. Essa comunicação é realizada através de APIs desenvolvidas com agências interessadas em adquirir as informações trabalhadas para que possam realizar suas próprias verificações, se assim desejarem. Portanto, as mesmas informações que são utilizadas no momento da apresentação do mapa são disponibilizadas para coleta via API pelos canais autorizados. Esta é uma forma de manter o canal direto com entidades que são fornecedoras e consumidoras de dados através da parceria desenvolvida.

Figura 33: Plataformas para disponibilização de resultados e interação

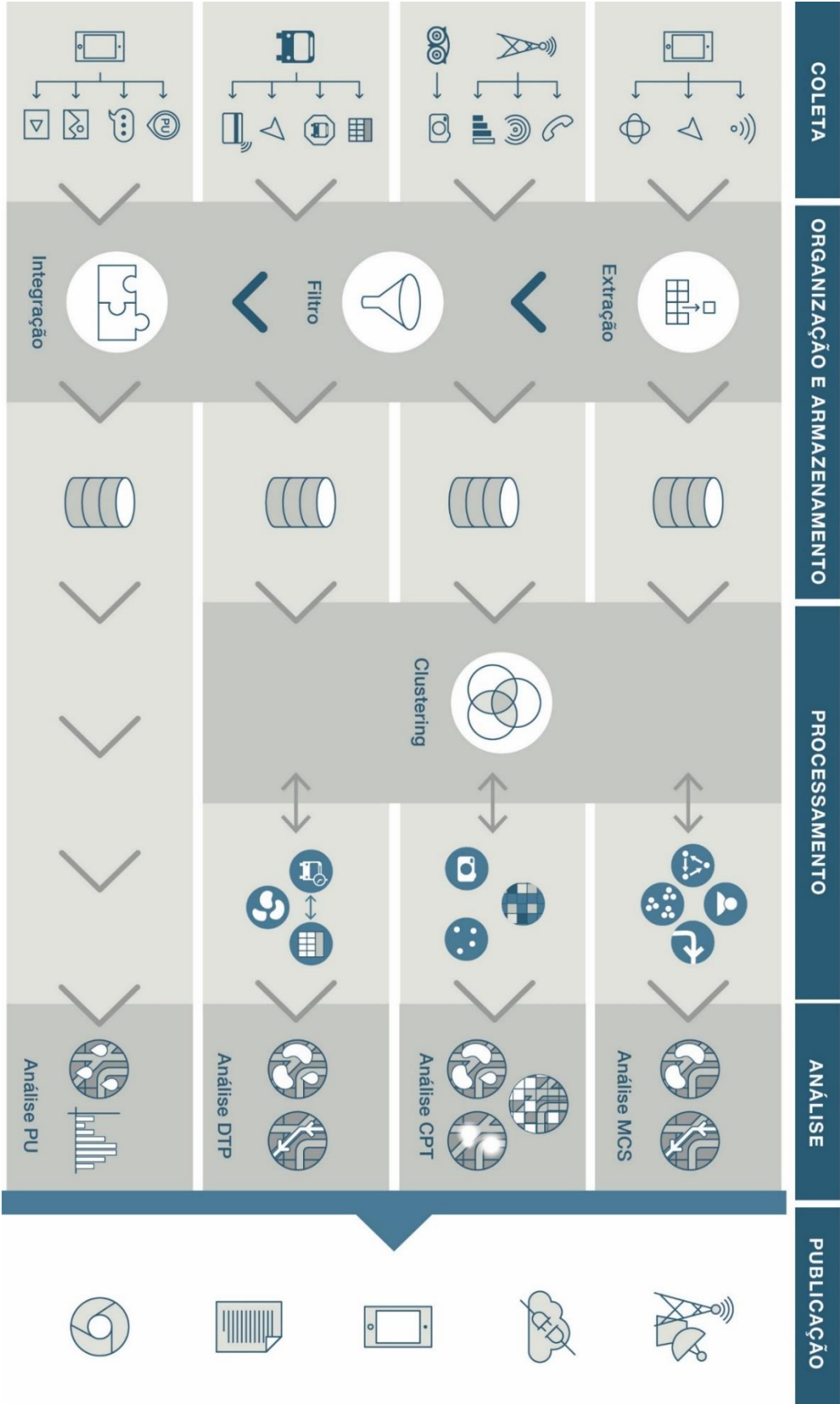


Fonte: Elaboração própria, 2018.

7.3.6 *Framework*

Portanto, o *framework* compreende o caminho dos dados e sua transformação em conhecimento desde o momento da coleta até a publicação dos resultados das análises, com característica cíclica e sistemática. Para que se possa visualizar o fluxo das suas informações como um todo, elaboramos uma figura ilustrativa, apresentando todas as etapas descritas anteriormente e seus subprocessos, de forma resumida e exemplificada (Figura 33).

Figura 34: Fluxo geral do framework



Fonte: Elaboração própria, 2018

7.4 DISCUSSÃO DOS RESULTADOS

A elaboração do *framework* foi fruto de uma análise comparativa entre a literatura estudada sobre o tema e a experiência de uma campanha de *crowdsensing*, que contou com processos de análise de dados de mobilidade urbana. A partir da nossa interpretação desses panoramas distintos, foi possível propor uma ideia de *framework* que contemplasse o que identificamos como pertinente e o que sentimos falta num processo de análise de dados de mobilidade urbana.

Na revisão da literatura, buscamos pesquisas que apresentaram propostas de *framework* ou arquiteturas de análise de dados para mobilidade urbana com o intuito de estudar fundamentos para o desenvolvimento do pré-experimento de análise e coletar insumos para comparação com a análise realizada no pré-experimento. Adotamos uma leitura interpretativa dos textos, avaliando a qualidade do estudo e a pertinência ao tema. Os estudos buscados se encontram nas áreas de arquitetura de análise de dados e análise de mobilidade urbana sob o pano de fundo da *smart city*.

Notamos a necessidade de se analisar o comportamento e eficiência do transporte público com a ajuda do *data analytics*. Diversas pesquisas propõem a utilização do ITS – conforme descrito no subcapítulo 4.3 dessa dissertação – como controle, uma ferramenta muito valiosa, no entanto, uma análise *a priori* de como o transporte público se desenvolve é fundamental para se ter base analítica e para a tomada de decisão. As necessidades técnicas que este tipo de análise requer não exigem muito desembolso e já estão sendo implementadas em algumas cidades do Brasil. Notamos a importância da utilização de dados de chamadas via dispositivos móveis. Essa abordagem de análise permite um bom panorama da concentração de pessoas, principalmente quando utilizada em conjunto com dados de *hotspots*, uma tendência crescente de utilização nas cidades. Ela permite o conhecimento, mesmo que aproximado, da concentração das pessoas sem necessitar da interação direta dos participantes. A análise para pessoas que não são locais também é de grande valor, principalmente para cidades com características de concentração sazonal de pessoas, devido à atratividade turística.

Os trabalhos analisados nos permitiram também identificar a importância da utilização de métodos de organização dos dados. Focamos muito em etapas chave do processo, como coleta e análise, mas muitos estudos realçam as etapas de organização e armazenamento dos

dados de forma estruturada e coordenada. Essa orientação ajuda no desenvolvimento de outra característica necessária para a análise de dados de mobilidade urbana, como o funcionamento sistemático. Como é um fenômeno dinâmico, a mobilidade urbana deve ser analisada frequentemente. Por conta disso, um *framework* dessa natureza deve ser sistemático.

Na etapa metodológica de pré-experimento, enfrentamos dificuldades no engajamento e comunicação clara com os usuários participantes. É essencial que o projeto seja atrativo e desperte a consciência de cidadania nas pessoas. Portanto, deve focar no design das interfaces de *front-end*, visando grande aderência ao uso, e na publicação, buscando a disseminação do projeto de forma a sensibilizar os cidadãos e informando concisa e claramente sobre os resultados alcançados. Uma funcionalidade de prestação de serviço é igualmente fundamental na busca de maior engajamento, atuando como um mecanismo de incentivo pela participação. Essa abordagem compreende a ideia do usuário no papel de produtor/consumidor, mantendo a relação de parceria com a comunidade.

Quanto a questões técnicas relativas ao pré-experimento, notamos a falta de organização sistêmica do fluxo dos dados para análise. Para o desenvolvimento de uma arquitetura para análise do tipo de dados em questão, o processo deve ser bem definido e estruturado para que seja executado de forma suave e sistemática, contando com organização de armazenamento que permita o funcionamento automático. A padronização do processo de análise é um artifício que auxilia neste ponto. Já contar com ferramentas desenvolvidas para realizar cada etapa do processo e que mantenham comunicação genuína entre si permite uma fácil implementação e utilização da inteligência.

Conforme mencionado, o ParticipACT Brasil tem a pretensão de se tornar uma plataforma de *big data analytics* para *smart city* mais completa, contemplando análises de problemáticas urbanas diversas. Seguindo essa diretriz, o *framework* propõe complementos à plataforma na área de mobilidade urbana ao endereçar os pontos discutidos, vislumbrando uma arquitetura robusta que comporta um processo cíclico e sistêmico de análise de dados.

8 CONCLUSÃO/RECOMENDAÇÕES

Dentro do cenário atual de crescimento urbano e seus impactos no bem-estar social e sustentabilidade do meio ambiente, identificamos a grande parcela de contribuição a essa problemática por parte da mobilidade urbana deficiente. O uso da tecnologia é um dos caminhos a serem adotado na busca por uma solução, tanto na abordagem do planejamento urbano quanto do tráfego inteligente. Inspirados neste cenário, buscamos nesse trabalho propor um *framework* para realização de análise de dados de mobilidade urbana em uma *smart city*, visando auxiliar na avaliação do comportamento dos deslocamentos urbanos e possibilitando insumos para tomadas de decisão.

Realizamos uma revisão da literatura acerca do tema, buscando estudos científicos realizados dentro a temática *data analytic* para mobilidade urbana sob o pano de fundo da *smart city*. Encontramos pesquisas de desenvolvimento de processos e arquiteturas que construíram não somente algumas etapas de análise de dados para o tema em pauta de forma isolada, tais como plataformas que somente realizam coletas de dados ou que apresentam modelos de análise, mas também arquiteturas completas, contemplando diversos passos do início ao fim do processo. Identificamos projetos que enfocam a análise da mobilidade urbana sob diferentes aspectos e abordagens. De maneira geral, os estudos contemplam o transporte público, o deslocamento independente do modal, a concentração de pessoas, fontes de dados, tipos de processamento e análise diversas, e formatos de fluxos de informação particulares. Esse contexto diversificado possibilitou o conhecimento abrangente do tema e como tratá-lo tanto na amplitude quanto no detalhamento.

O conhecimento proporcionado pela revisão da literatura fundamentou a realização do pré-experimento executado. O pré-experimento de mobilidade urbana consistiu numa campanha de *crowdsensing* com pessoas envolvidas na pós-graduação da ESAG/UDESC com posterior análise de dados e publicação. Nele, pudemos imprimir as opiniões, visões e interpretações dos pesquisadores, e desenvolver o projeto de acordo com a abordagem proposta, sem se limitar a uma padronização pré-determinada. Necessidades de ordem técnica e conceitual, tais como organização dos dados e definição de processo, e na esfera de relacionamento, principalmente com cidadãos, foram identificadas no decorrer desta etapa.

As pessoas que participaram como usuários na campanha, ou seja, que tiveram seus dados de localização e deslocamento coletados via dispositivo celular próprio e analisados, possuem informações importantes com relação a essa experiência, pois interagiram com o projeto tanto ativa quanto passivamente. Através de questionários aplicados, pudemos coletar e interpretar suas impressões e perspectivas sobre o mecanismo de comunicação e entendimento da campanha, suas expectativas com relação a resultados e prestação de serviço, suas opiniões e sugestões. Esse método revela aspectos observados por agentes externos aos pesquisadores que estão envolvidos na idealização do projeto, fundamental para se ter uma visão distinta e complementar.

Com base na interpretação e observação dos pesquisadores da literatura estudada e do que foi vivenciado no pré-experimento, nos municiamos de informação para elaborar um levantamento de requisitos para o *framework*. Os requisitos não funcionais delimitaram características e restrições do sistema, além de necessidades do ambiente no qual ele está envolvido. Os funcionais retrataram o que o sistema deve fazer, definindo o funcionamento de maneira geral. Foram estabelecidos para todas as etapas do processo, de modo a apresentar diretrizes que nortearam a formulação de um fluxo de dados/informações, estabelecendo suas funcionalidades necessárias. Por conta do escopo do trabalho, detalhes técnicos não foram aprofundados, apenas atentando-se para o funcionamento do *framework*.

O *framework* proposto foi, portanto, estruturado em cinco categorias: coleta, organização e armazenamento, processamento, análise e publicação. Essas categorias são responsáveis por tratar os dados, de acordo com sua aplicabilidade, considerando quatro tipos de abordagem de análise: análise MCS, voltada para *mobile crowdsensing*; CPT, para analisar a concentração de população e turistas; DTP, para transporte público; e PU, para reporte de problemas urbanos.

Na categoria de coleta, definimos os tipos de dados a serem coletados, as suas fontes, qual é a finalidade de cada conjunto de dados, com qual frequência devem ser coletados e por qual via de coleta. Dados de sensores de celulares serão utilizados para as análises de campanha de *crowdsensing* e para registrar os problemas urbanos. Dados de chamadas telefônicas e informações turísticas para análise de concentração de pessoas e turistas, e dados de localização e planejamento de transporte público para as análises de transporte público. Logo, em cidades que possibilitam coletar esse tipo de dados poderão se beneficiar desse arranjo.

Foram definidos quatro passos de organização e armazenamento de dados: extração, filtro, integração e armazenamento. Os métodos de extração servem como um mecanismo que ajuda no fluxo contínuo e natural do processo, recolhendo os dados coletados dos catálogos de armazenamento e disponibilizando para os processamentos de filtro correspondentes. Este, por sua vez, descarta dados incorretos, *outliers*, redundantes e que possam inferir na identificação pessoal dos participantes, mantendo a privacidade. A parte de integração de dados tem o intuito de combinar dados de tipos diferentes que são complementares para determinada análise. Por fim, o armazenamento posiciona os dados em locais específicos que direcionam sua exploração por parte das rotinas de processamento, assegurando o fluxo sistemático do *framework*.

Para interpretar e extrair informações a partir dos dados coletados e organizados, a etapa de processamento compreende um bloco de rotinas de determinação de *clusters* e rotinas de processamento específicas. Processamentos baseados no algoritmo DBSCAN são executados para a determinação de conjuntos com maior concentração de pontos de localização. O resultado desse processamento é utilizado pelas rotinas posteriores onde, de acordo com a abordagem de análise alvo, conexões, rotas, locais de concentração entre outras características são determinadas.

Em alguns estudos, a etapa de processamento é realizada juntamente com a análise. No entanto, pela variedade de abordagens e diferença de análises, consideramos a importância do tratamento em separado dessas etapas. Na análise, vislumbramos uma preocupação mais visual, que atentasse ao cuidado de oferecer uma demonstração dos resultados de forma intuitiva e amigável, auxiliando da melhor forma na tomada de decisão por agentes responsáveis. Por se tratar de mobilidade urbana, os resultados são apresentados em mapas com a representação do comportamento da mobilidade para cada tipo de análise.

Para atender à característica cíclica do *framework*, à ideia de prestação de serviço e disseminação do projeto, a etapa de publicação foi idealizada com a disponibilização dos resultados através do aplicativo móvel, portal *web* e APIs de comunicação. Dessa forma, mantém-se uma parceria com a comunidade que fornece dados e consome o conhecimento gerado. Portanto, de acordo com as características apresentadas, a proposição elaborada atende ao objetivo desse estudo de propor um *framework* para processos de análise de dados de mobilidade urbana em uma *smart city*.

Como trabalhos futuros e recomendações, sugerimos que se desenvolvam mecanismos de incentivo para o uso dos aplicativos de *crowdsensing*. Notamos uma dificuldade em sensibilizar as pessoas a se envolverem na campanha, e, quanto maior o número de pessoas engajadas, mais fiéis serão os dados coletados. Com o intuito de amenizar essa questão, sugerimos no *framework* a funcionalidade de análise de problemas urbanos, servindo como uma conexão direta e constante entre o projeto e os participantes. No entanto, outros modelos de incentivo devem ser implementados, tanto na dimensão técnica quanto comunicativa. Sensibilizar entidades fontes de dados que formam uma rede de apoio ao projeto, principalmente no âmbito acadêmico, é igualmente fundamental. Além do apoio ao desenvolvimento, a rede serve para reforçar a importância e legitimar o trabalho junto a sociedade para maior engajamento, principalmente nos momentos de coleta de dados e publicação de resultados.

Naturalmente, recomendamos que seja continuado o processo de desenvolvimento do *framework* com o intuito de concretizar uma plataforma robusta de análise de dados de mobilidade urbana. Para isso, o levantamento de requisitos deve ser refinado e aprofundado, buscando endereçar questões técnicas e gerenciais da implementação: de infraestrutura, desenvolvimento, dinâmica de trabalho e relação com agentes externos. Uma parceria interdisciplinar com o curso de Ciências e Engenharia da Computação é um caminho interessante para que isso possa ser materializado. Além de realizar o trabalho de desenvolvimento técnico, essas áreas podem auxiliar na padronização do processo e da arquitetura de análise de dados, como definição de parâmetros de execução das rotinas de identificação de *clusters*, quais ferramentas/plataformas de desenvolvimentos devem ser utilizadas em cada etapa, entre outras grandes dificuldades identificadas nesse estudo. A visão disciplinar fragmentada do assunto, tanto no isolamento da área administrativa quanto tecnológica, cria barreiras que dificultam a implementação de um projeto desse porte.

Com a implementação do sistema vislumbrado, é importante que se realizem testes mais abrangentes na cidade. Com uma coleta de dados robusta, de diversas fontes e tipos, com uma estrutura de análise bem determinada, resultados de grande qualidade serão alcançados. Assim, será possível avaliar e interpretar o comportamento da mobilidade urbana e sugerir aprimoramentos mais assertivos do processo. De antemão, entendemos que a apresentação do histórico do comportamento já é um bom indicativo para se fazer inferências, mas um experimento mais abrangente pode fornecer insumos para concepção de modelos preditivos

e prescritivos da mobilidade urbana, com base em informações econômico/político/sociais. O conceito de *big data analytics* compreende esses modelos, pois tem a pretensão de antecipar o que irá acontecer e sugerir decisões.

Conforme mencionado no início do capítulo, esse trabalho foi inspirado nos impactos do crescimento urbano no bem-estar social, na mobilidade urbana e na sustentabilidade do meio ambiente. Portanto, foi concebido para ser utilizado com um apelo socioambiental, visando melhorar a vida das pessoas no ambiente urbano, sob uma política de conservação ecológica e pensamento sustentável.

9 REFERÊNCIAS

ABDALLAH, Z. et al. **StreamAR**: incremental and active learning with evolving sensory data for activity recognition. In: Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference. [S.l.]: IEEE. 2012. p. 1163-1170.

AKTER, S. et al. How to improve firm performance using big data analytics capability and business strategy alignment? **International Journal of Production Economics**, v. 182, p. 113-131, 2016.

ALKHELAIWI, A.; GRIGORAS, D. **Scheduling crowdsensing data to smart city applications in the cloud**. Intelligent Computer Communication and Processing (ICCP), 2016 IEEE 12th International Conference. [S.l.]: IEEE. 2016. p. 395-401.

ALVES, P.; RAIA JUNIOR, A. **Mobilidade e Acessibilidade Urbanas sustentáveis: A gestão da mobilidade no Brasil**. VI Congresso de Meio Ambiente da Associação de Universidades Grupo de Montevidéu - AUGM ambiente. São Carlos: Anais de Eventos da UFSCar. 2009.

AMBONI, N. **O Caso da cecrisa s.a.:** uma aprendizagem que deu certo. Florianópolis: Tese - Engenharia de Produção - UFSC, 1997.

AMBROSINO, G. et al. Enabling intermodal urban transport through complementary services: From Flexible Mobility Services to the Shared Use Mobility Agency: Workshop 4. Developing inter-modal transport systems. **Research in Transportation Economics**, v. 59, p. 179-184, 2016.

ANGELAKIS, V. et al. **Mobility modeling for transport efficiency**: analysis of travel characteristics based on mobile phone data. In: Netmob 2013-Third International Conference on the Analysis of Mobile Phone Datasets. Cambridge: MIT. 2013. p. 1-11.

ANTP. **Sistema de Informação da Mobilidade Urbana: Relatório Comparativo 2003-2014**. São Paulo. 2016.

ASCHER, F. **Metapolis**: Acerca do futuro da cidade. Oeiras: Celta Editora, 1998.

BADIA, H.; ESTRADA, M.; ROBUSTÉ, F. Bus network structure and mobility pattern: A monocentric analytical approach on a grid street layout. **Transportation Research Part B: Methodological**, v. 93, p. 37-56, 2016.

BELLAVISTA, P. et al. Crowdsensing in Smart Cities: Technical Challenges, Open Issues and Emerging Solution Guidelines. **Handbook of Research on Social, Economic, and Environmental Sustainability in the Development of Smart Cities**, p. 316, 2015.

BENINCASA, C. et al. Page Rank Algorithm. **Department of Mathematics and Statics, University of Massachusetts, Amherst, Research**, 2006.

BHARADWAJ, A. et al. Digital business strategy: toward a next generation of insights. **MISQ**, 2013. 471–482.

BOHUSCH, G. **Mobilidade urbana sustentável: uma proposta de visão ampliada do conceito**. Florianópolis: Dissertação – Programa de pós-graduação em Geografia, Universidade Federal de Santa Catarina, 2013.

BRASIL. lei 9053 de 23 de setembro de 1997. Institui o Código de Trânsito Brasileiro. **Diário Oficial [da] República Federativa do Brasil**, Brasília, 1997. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9503.htm>.

BRASIL. Lei nº 12.587, de 03 de janeiro de 2012. Institui as diretrizes da Política Nacional de Mobilidade Urbana e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, jan 2012.

BUENO, A. P. **PATRIMÔNIO PAISAGÍSTICO E TURISMO NA ILHA DE SANTA CATARINA: A PREMÊNCIA DA PAISAGEM NO DESENVOLVIMENTO SUSTENTÁVEL DA ATIVIDADE TURÍSTICA**. São Paulo: Tese - Pós-graduação em Arquitetura e Urbanismo, Universidade de São Paulo, 2006.

CAMBRIDGE DICTIONARY. **Cambridge Dictionary**. [S.l.]: [s.n.]. Disponível em: <<https://dictionary.cambridge.org/dictionary/english/framework>>. Acesso em: mar. 2018.

CARO, F. Pesquisa avaliativa: uma visão panorâmica. In: GOLDBERG, M.; SOUZA, C. **Avaliações de programas educacionais: vicissitudes, controvérsias, desafios**. São Paulo: EPU, 1982. p. 10-14.

CARRERAS, I. et al. **Matador**: Mobile task detector for context-aware crowd-sensing campaigns. Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference. [S.l.]: IEEE. 2013. p. 212-217.

CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile Networks and Applications**, v. 19, n. 2, p. 171-209, 2014.

CHRISTIN, D. et al. A survey on privacy in mobile participatory sensing applications. **Journal of systems and software**, v. 84, n. 11, p. 1928-1946, 2011.

COCCHIA, A. Smart and digital city: A systematic literature review. **Springer International Publishing**, 2014. 13-43.

CONSELHO DAS CIDADES. Resolução nº 34, de 01 de julho de 2005. **Diário Oficial [da] República Federativa do Brasil**, Brasília, 2005.

CORRADI, A. et al. **Automatic extraction of pois in smart cities**: Big data processing in participact. In: Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on. IEEE. [S.l.]: IEEE. 2015a. p. 1059-1064.

CORRADI, A. et al. **Smartphones as smart cities sensors**: MCS scheduling in the ParticipAct project. Computers and Communication (ISCC), 2015 IEEE Symposium. [S.l.]: IEEE. 2015b. p. 222-228.

DANIEL, A. et al. Big autonomous vehicular data classifications: Towards procuring intelligence in ITS. **Vehicular Communications**, v. 9, p. 306-312, 2017.

DATTA, S. et al. **oneM2M architecture based IoT framework for mobile crowd sensing in smart cities**. In: Networks and Communications (EuCNC), 2016 European Conference. [S.l.]: IEEE. 2016. p. 168-173.

DE ÁVILA GOMIDE, A.; GALINDO, E. A mobilidade urbana: uma agenda inconclusa ou o retorno daquilo que não foi. **Estudos avançados**, v. 27, n. 79, p. 27-39, 2013.

DIMITRAKOPOULOS, G.; DEMESTICHAS, P. Intelligent transportation systems. **IEEE Vehicular Technology Magazine**, v. 5, n. 1, p. 77-84, 2010.

DRAGHICI, A.; AGIALI, T.; CHILIPAREA, C. **Visualization system for human mobility analysis**. In: RoEduNet International Conference-Networking in Education and Research (RoEduNet NER). [S.l.]: IEEE. 2015. p. 152-157.

ELARABI, T.; SHARMA, B.; PAHWA, K. **Big data analytics concepts and management techniques**. Inventive Computation Technologies (ICICT), International Conference on. IEEE. [S.l.]: [s.n.]. 2016. p. 1-6.

ESTER, M. et al. **A density-based algorithm for discovering clusters in large spatial databases with noise**. In: KDD. [S.l.]: [s.n.]. 1996. p. 226-231.

ESTRIN, D. Participatory sensing: applications and architecture. **IEEE Internet Computing**, v. 14, n. 1, p. 12-14, 2010.

FARKAS, K. et al. Crowdsending based public transport information service in smart cities. **IEEE Communications Magazine**, v. 53, n. 8, p. 158-165, 2015.

FISTOLA, R.; RAIMONDO, M.; LA ROCCA, L. **The smart city and mobility: The functional polarization of urban flow**. Models and Technologies for Intelligent Transportation Systems (MT-ITS). [S.l.]: 5th IEEE International Conference. 2017. p. 532-537.

GAL, A. et al. Traveling time prediction in scheduled transportation with journey segments. **Information Systems**, v. 64, p. 266-280, 2015.

GANTI, R.; YE, F.; LEI, H. Mobile crowdsensing: current state and future challenges. **IEEE Communications Magazine**, v. 49, n. 11, 2011.

GANTZ, J.; REINSEL, D. Extracting value from chaos. **IDC iView**, p. 1-12, 2011.

GIFFINGER, R.; GUDRUN, H. Smart cities ranking: An effective instrument for the positioning of cities? **ACE: Architecture, City and Environment**, 2010. 7-25.

GIL, A. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GOMES, E. et al. **Towards an Infrastructure to Support Big Data for a Smart City Project**. In: Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016 IEEE 25th International Conference. [S.l.]: IEEE. 2016. p. 107-112.

GUBBI, J. et al. Internet of Things (IoT): A vision, architectural elements, and future directions. **Future generation computer systems**, v. 29, n. 7, p. 1645-1660, 2013.

HALL, R. **The vision of a smart city**. Proceedings of the 2nd International Life Extension Technology Workshop. Paris: [s.n.]. 2000.

HAN, J.; PEI, J.; KAMBER, M. Data mining: concepts and techniques. **Elsevier**, 2011.

HARRISON, C. et al. Foundations for Smarter Cities. **IBM Journal of Research and Development**, 2010.

HEISKALA, M.; JOKINEN, J.; TINNILÄ, M. Crowdsensing-based transportation services—An analysis from business model and sustainability viewpoints. **Research in Transportation Business & Management**, v. 18, p. 38-48, 2016.

IBGE. **Censo demográfico 2010**. IBGE: Instituto Brasileiro de Geografia e Estatística. [S.l.]. 2010.

ISO, IEC. **IEEE 29148: Systems and software engineering-Requirements engineering**. Technical report. 2011.

JAIMES, L.; VERGARA-LAURENS, I.; RAIJ, A. A survey of incentive techniques for mobile crowd sensing. **IEEE Internet of Things Journal**, v. 2, n. 5, p. 370-380, 2015.

JAIN, A. Data clustering: 50 years beyond K-means. **Pattern recognition letters**, v. 31, n. 8, p. 651-666, 2010.

JAYARAMAN, P. et al. Scalable energy-efficient distributed data analytics for crowdsensing applications in mobile environments. **IEEE Transactions on Computational Social Systems**, v. 2, n. 3, p. 109-123, 2015.

JIANG, C. et al. Scalable Mobile Crowdsensing via Peer-to-Peer Data Sharing. **IEEE Transactions on Mobile Computing**, 2017.

KAMIOKA, T.; TAPANAINEN, T. **Organizational use of big data and competitive advantage - exploration of antecedents**. Pacific Asia conference on information systems (PACIS). [S.l.]: [s.n.]. 2014. p. 372.

LAM, D.; HEAD, P. Sustainable Urban Mobility. **Energy, Transport, & the Environment**, Londres, 2012.

LAMBA, H.; DUBEY, S. **Analysis of requirements for big data adoption to maximize IT business value in reliability**. 4th International conference on infocom technologies and optimization. [S.l.]: IEEE. 2015. p. 1-6.

LANA, I.; DEL SER, J.; OLABARRIETA, I. **Understanding daily mobility patterns in urban road networks using traffic flow analytics**. In: Network Operations and Management Symposium (NOMS), 2016 IEEE/IFIP. [S.l.]: IEEE. 2016. p. 1157-1162.

LODER, C. O processo do crescimento Urbano no Brasil. **Pesquisa e Planejamento Econômico**, 1977. 459-484.

LORA, R. **Por uma Construção da Mobilidade Urbana: Metodologia e Indicadores na Cidade de Vitória-ES**. Vitória: Dissertação de Mestrado, Universidade Federal do Espírito Santo, 2012.

LYONS, G. Getting smart about urban mobility—aligning the paradigms of smart and sustainable. **Transportation Research Part A: Policy and Practice**, 2016.

MAGAGNIN, R.; DA SILVA, A. A percepção do especialista sobre o tema mobilidade urbana. **Transportes**, v. 16, n. 1, 2008.

MANIAK, T.; IQBAL, R.; DOCTOR, F. **Traffic Modelling, Visualisation and Prediction for Urban Mobility Management**. In: Advances in Hybridization of Intelligent Methods. Cham: Springer. 2018. p. 57-70.

MARSA-MAESTRE, I. et al. Mobile agents for service personalization in smart environments. **Journal of Networks**, 2008. 30-41.

MARTINE, G.; OJIMA, R.; FIORAVANTE, E. Transporte individual, dinâmica demográfica e meio ambiente. In: MARTINE, G.; CARMO, R. **População e sustentabilidade na era das mudanças ambientais globais**. Campinas: ABEP, 2012. p. 175-185.

MAY, A. et al. Appropriate national policy frameworks for sustainable urban mobility plans. **European Transport Research Review**, v. 9, n. 1, p. 7, 2017.

MELLO, A.; PORTUGAL, L. Um procedimento baseado na acessibilidade para a concepção de Planos Estratégicos de Mobilidade Urbana: o caso do Brasil. **EURE**, Santiago, v. 43, n. 128, p. 99-125, 2017.

MIKALEF, P. et al. Big data analytics capabilities: a systematic literature review and research agenda. **Information Systems and e-Business Management**, 2017. 1-32.

MINISTÉRIO DAS CIDADES. **Cadernos MCidades Desenvolvimento Urbano: Política Nacional de Desenvolvimento Urbano**. Brasília: [s.n.], 2004a.

MINISTÉRIO DAS CIDADES. **Política Nacional de Mobilidade Urbana Sustentável**. Brasília: SeMob, 2004b.

MINISTÉRIO DAS CIDADES. **PlanMob: Caderno de referência para elaboração do Plano de Mobilidade Urbana**. Brasília: SeMob, 2007.

MITCHELL, W.; BORRONI-BIRD, C.; BURNS, L. **Reinventing the automobile. Personal Urban Mobility for the 21st Century**. Cambridge: The MIT Press, 2010.

MULLER, O. et al. Utilizing big data analytics for information systems research: challenges, promises and guidelines. **European Journal of Information Systems**, v. 25, n. 4, p. 289-302, 2016.

NAM, T.; PARDO, T. **Conceptualizing smart city with dimensions of technology, people, and institutions**. 12th Annual International Conference. College Park: Digital Government Research. 2011.

NI, J. et al. **Dual-anonymous reward distribution for mobile crowdsensing**. Communications (ICC), IEEE International Conference. [S.l.]: IEEE. 2017. p. 1-6.

NOBRE, E. Ampliação da Marginal do Tietê: demanda real ou rodoviarismo requeentado? **AU. Arquitetura e Urbanismo**, v. 191, p. 58-63, 2010.

OLIVEIRA, B.; BELO, O. On the specification of extract, transform, and load patterns behavior: A domain-specific language approach. **Expert Systems**, v. 34, n. 1, p. 1-9, 2017.

OST. **Intelligent Infrastructure Futures: Project Overview**. Foresight Programme, Office of Science and Technology, Department of Trade and Industry. London. 2006.

PAPA, R.; CARAGIULO, C.; RUSSO, L. **The evolution of smart mobility strategies and behaviors to build the smart city**. Models and Technologies for Intelligent Transportation Systems (MT-ITS). [S.l.]: 5th IEEE International Conference. 2017. p. 409-414.

PARTICIPACT BRASIL. **Tecnologias inovadoras na gestão da cidade inteligente.** [S.l.]: [s.n.]. Disponível em: <<http://labges.esag.udesc.br/participact/>>. Acesso em: mar. 2018.

PARTRIDGE, H. **Developing a human perspective to the digital divide in the smart city.** Proceedings of the Biennial Conference of Australian Library and information Association. Queensland: [s.n.]. 2004.

PAVONE, M. et al. Robotic load balancing for mobility-on-demand systems. **The International Journal of Robotics Research**, v. 31, n. 7, p. 839-854, 2012.

PEREIRA, M. Redes, sistemas de transportes e as novas dinâmicas do território no período atual: notas sobre o caso brasileiro. **Soc. nat. (Online)**, Uberlândia, v. 21, n. 1, abril 2009.

PLAMUS. **Plano de Mobilidade Urbana Sustentável da Grande Florianópolis.** Florianópolis. 2014.

QUADROS JUNIOR, H. **Entre o ônibus e o carro: a questão da prioridade do transporte público na mobilidade urbana brasileira.** Florianópolis: Dissertação – Programa de pós-graduação em Engenharia Civil, Universidade Federal de Santa Catarina, 2011.

RAJARAMAN, V. Big data analytics. **Resonance**, v. 21, p. 695-716, 2016.

RATTI, C. et al. Mobile landscapes: using location data from cell phones for urban analysis. **Environment and Planning B: Planning and Design**, v. 33, n. 5, p. 727-748, 2006.

RUDIO, F. **Introdução ao projeto de pesquisa científica.** 34. ed. Petrópolis: Vozes, 2007.

RUSSO, F.; COMI, A. From the analysis of European accident data to safety assessment for planning: the role of good vehicles in urban area. **European Transport Research Review**, v. 9, n. 1, p. 9, 2017.

RUSSOM, P. Big data analytics. **TDWI Best Practices Report, Fourth Quarter**, 2011. 1-35.

SABTU, A. et al. **The challenges of Extract, Transform and Loading (ETL) system implementation for near real-time environment.** In: Research and Innovation in Information Systems (ICRIIS), 2017 International Conference on. IEEE. [S.l.]: IEEE. 2017. p. 1-5.

SACCOL, A. Um retorno ao básico: compreendendo os paradigmas de pesquisa e sua aplicação na pesquisa em administração. **Revista de Administração da Universidade Federal de Santa Maria**, Santa Maria, v. 2, n. 2, p. 250-269, mai/ago 2009.

SCHATZINGER, S.; LIM, C. **Taxi of the future**: Big data analysis as a framework for future urban fleets in smart cities. International conference on Smart and Sustainable Planning for Cities and Regions. Cham: Springer. 2015. p. 83-98.

SEDDON, J.; CURRIE, W. A model for unpacking big data analytics in high-frequency trading. **Journal of Business Research**, v. 70, p. 300-307, 2017.

SEREDYNSKI, M.; VITI, F. **A survey of cooperative ITS for next generation public transport systems**. Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference. [S.l.]: IEEE. 2016. p. 1229-1234.

SHENG, X. et al. Sensing as a service: Challenges, solutions and future directions. **IEEE Sensors journal**, v. 13, n. 10, p. 3733-3741, 2013.

SILVA, P. **Engenharia de Tráfego**: Elementos dos Sistemas de Tráfego. Brasília: Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, 2001. Disponível em:

<ftp://ftp.cefetes.br/cursos/transportes/Michellygf/SegurancaeEngenhariadeTrafego/apostilas/APOSTILA1_elementos%20do%20sistema%20de%20trafego.pdf>.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson Prentice Hall, 2011.

STEFANELLI, V.; PERO, V. A questão da mobilidade urbana nas metrópoles brasileiras. **Revista de Economia Contemporânea**, v. 19, n. 3, p. 366-402, 2015.

STEG, L.; GIFFORD, R. Sustainable transportation and quality of life. **Journal of Transport Geography**, v. 13, p. 59-69, 2005.

SUN, F. **PhD Forum**: Robust Sensing and Analytics in Urban Environment. In: Smart Computing (SMARTCOMP), 2017 IEEE International Conference. [S.l.]: IEEE. 2017. p. 1-2.

SUN, W. et al. **Characterizing User Mobility from the View of 4G Cellular Network**. In: Mobile Data Management (MDM), 2016 17th IEEE International Conference. [S.l.]: IEEE. 2016. p. 34-39.

TECHAMERICA FOUNDATION. **Demystifying Big Data: A Practical Guide to Transforming the Business of Government**. Federal Big Data Commission. Washington: TechAmerica Foundation. 2012.

TECHTARGET NETWORK. **WhatIs.com**. [S.l.]: [s.n.]. Disponível em: <<http://whatis.techtarget.com/definition/front-end>>. Acesso em: mar. 2018.

TELES, P. **Os territórios (sociais) da Mobilidade: um desafio para a área metropolitana do Porto**. Porto: Lugar do Plano, 2005.

THE B/A/H REPORT. **The Field Guide to DATA SCIENCE**. The Booz/Allen/Hamilton. [S.l.]. 2015.

TOQUÉ, F. et al. **Forecasting dynamic public transport Origin-Destination matrices with long-Short term Memory recurrent neural networks**. Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference. [S.l.]: IEEE. 2016. p. 1071-1076.

VACA, M. et al. **Automated global planner for cybernetic transportation systems**. Control, Automation, Robotics and Vision (ICARCV), 2016 14th International Conference. [S.l.]: IEEE. 2016. p. 1-6.

VIOLATO, C.; LOURAL, C. Desafios para o desenvolvimento das TICs no Brasil. **Parcerias Estratégicas**, v. 15, n. 31, p. 283-288, 2012.

WANG, F.; HERGET, C.; ZENG, D. Developing and improving transportation systems: the structure and operation of IEEE intelligent transportation systems society. **IEEE Transactions on Intelligent Transportation Systems**, v. 6, n. 3, p. 261-264, 2005.

WANG, Y. et al. **A Big Data Approach for Smart Transportation Management on Bus Network**. Smart Cities Conference (ISC2), 2016 IEEE International. [S.l.]: IEEE. 2016.

WASHBURN, D. et al. Helping CIOs Understand “Smart City” Initiatives: Defining the Smart City, Its Drivers, and the Role of the CIO. **Inc. Retrieved April**, Cambridge, v. 12, p. 2014, 2010.

WATSON, H. Tutorial: Big data analytics: Concepts, technologies, and applications. **Communications of the Association for Information Systems**, v. 34, p. 65, 2014.

WEISS, R. **A dinâmica urbana:** da compreensão à construção de um modelo de identificação dos padrões de crescimento urbano por meio das métricas espaciais da paisagem: o caso do distrito da Lagoa da Conceição. Florianópolis: Tese – Programa de pós-graduação em arquitetura e urbanismo, Universidade Federal de Santa Catarina, 2016.

WHITE, C. Using big data for smarter decision making. **BI research**, New York, 2011.

ZHENG, X. et al. Big data for social transportation. **IEEE Transactions on Intelligent Transportation Systems**, v. 17, n. 3, p. 620-630, 2016.

ZITO, P.; CORRADI, A. **Recognition of urban mobility patterns by using Spark on crowdsensing data.** [S.l.]: Tesi di Laurea - Corso di Ingegneria Informatica - Università di Bologna, 2017.

10 ANEXOS

10.1 ANEXO 1: BANNER DE FEEDBACK DA CAMPANHA DE CROWDSENSING

ParticipACT

TECNOLOGIAS INOVADORAS NA GESTÃO DA CIDADE INTELIGENTE

O **Projeto ParticipACT Brasil** objetiva projetar, promover e desenvolver tecnologias inovadoras para a gestão participativa da cidade inteligente.

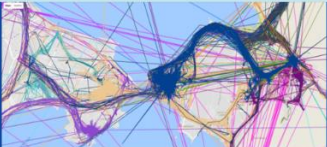
Baseado nos conceitos de *Smart City*, o **ParticipACT Brasil** irá captar informações por meio do uso inovador de Tecnologias da Informação e Comunicação (TIC) para facilitar a tomada de decisão de gestores públicos e privados.

Coleta de dados por meio de campanhas de **crowdsensing**, através da participação cooperada e voluntária da população, visando integrar um **Big Data** referente a características urbanas.





Estudo da mobilidade urbana: Envolve a coleta de dados de GPS via smartphone de pessoas engajadas de forma voluntária em uma campanha de crowdsensing para a identificação de padrões de hábitos de mobilidade.



Visualização da **rota consolidada** de todos os participantes durante a campanha.



Retrato das **regiões com maior concentração** dos participantes (clusters) e suas respectivas conexões, de acordo com o período do dia (**verde** = manhã, **laranja** = tarde, **azul** = noite).



Heat map com os pontos de maior concentração de acordo com os deslocamentos dos participantes.

Alunos da ESAG: estudo dos hábitos de mobilidade

Dados da Campanha
Período: 15/05 a 15/07
Número de participantes: 67
Sensores utilizados: GPS

PRÓXIMOS DESAFIOS:

- **Expandir** quantidade de participantes;
- **Refinar** a ferramenta de análise, visando estabelecer padrões de mobilidade.

REALIZAÇÃO E APOIO:

LabGES

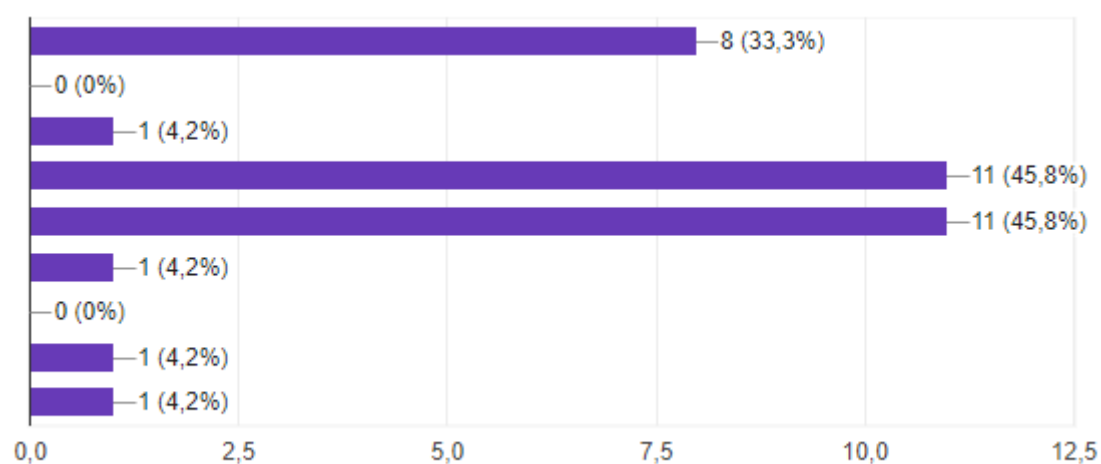


Fonte: Elaborado pelos pesquisadores do laboratório LabGES, 2017.

10.2 ANEXO 2: QUESTIONÁRIO APLICADO AOS PARTICIPANTES DA CAMPANHA

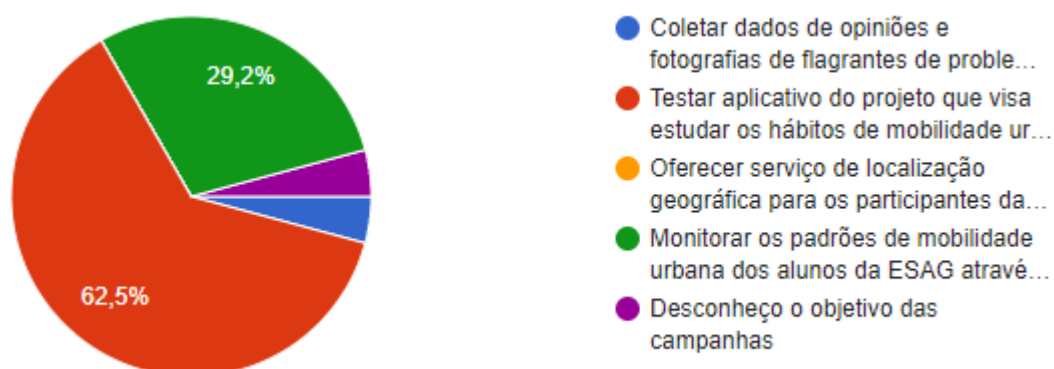
Por quais meios de comunicação ficou sabendo da campanha de crowdsensing "Mobilidade na ESAG" do projeto ParticipACT Brasil?

- ☐ Email
- ☐ Panfletos nos murais
- ☐ TV no saguão da ESAG
- ☐ Apresentação nas salas de aula
- ☐ Boca-a-boca
- ☐ Facebook da ESAG
- ☐ Site da ESAG



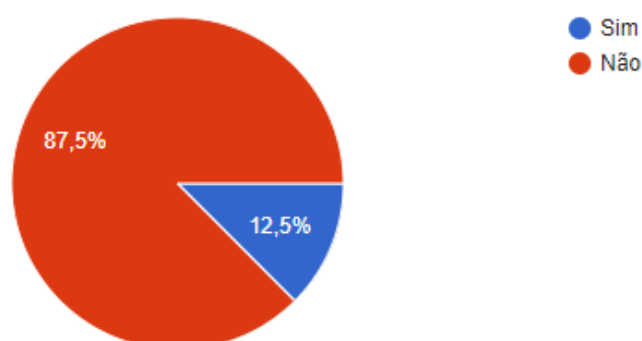
O objetivo da campanha "Mobilidade na ESAG" é: *

- ☐ Coletar dados de opiniões e fotografias de flagrantes de problemas de mobilidade urbana dos participantes
- ☐ Testar aplicativo do projeto que visa estudar os hábitos de mobilidade urbana através da coleta de dados de localização
- ☐ Oferecer serviço de localização geográfica para os participantes da campanha de crowdsensing
- ☐ Monitorar os padrões de mobilidade urbana dos alunos da ESAG através de coleta de dados via sensoramento por IoT
- ☐ Desconheço o objetivo das campanhas



Considerando o fato de coletarmos dados de posição geográfica através do seu dispositivo móvel para estudo de mobilidade urbana, você vê alguma restrição em disponibilizar os seus dados de localização para fins de estudo acadêmico?

24 respostas



Que tipo de feedback você gostaria de receber após a campanha de crowdsensing e análise dos dados do nosso estudo?

15 respostas

- Gostaria de receber informações das regiões mais problemáticas relacionadas a mobilidade em Florianópolis.
- o resultado do teste ou do estudo
- Resultados e próximos passos do estudo
- Um relatório com os padrões de mobilidade urbana dos alunos da ESAG.
- qual o fluxo e a porcentagem de pessoas que utilizam ônibus
- Gostaria de ter acesso ao banco de dados criado através da campanha.
- Análise macro
- Vizualização do relatório final.
- Tudo que foi descoberto
- Como os dados serão utilizados para se propor melhorias
- Mostrar quais as contribuições que foram alcançadas com esta campanha
- Receber relatório de mobilidade
- Um relatório objetivo sobre os resultados alcançados
- Apresentação dos resultados da campanha
- Trajetos mais tumultuados e que requerem investimento de infra estrutura do governo, horários de pico a serem evitados

Que tipo de serviço você gostaria que o aplicativo oferecesse para o usuário que estiver participando das campanhas de crowdsensing?

10 respostas

- Informações ao vivo de situações específicas próximas a ESAG
- na minha visao aplicativo nao precisa ter um serviço pois os resultados serão os estudos realizados
- Mapa sobre mobilidade
- Opção de indicar qual meio de transporte está utilizando.
- disponibilização de linhas e horários de ônibus próximo a localização
- Algum tipo de brinde ou ainda pode ser melhorado através da gamificação.
- Melhores rotas.
- Mostar padrões por região, seja de coisas boas ou ruins
- Se tem outros usuários nas imediações
- Serviço de cidadania, problemas urbanos

Com quais sugestões de melhorias e/ou críticas sobre o aplicativo você pode contribuir?

9 respostas

facilitar o login,

Havia durante o período da campanha uma notificação constante no celular sem possibilidade de ser retirada informando que os dados estavam sendo coletados pelo app, o que incomodava bastante. Seria bom se possível retirar essa notificação.

não achei fácil saber se deveria ou não postar fotos. e se sim, quais fotos seriam de interesse para a campanha.

Notei que o app utilizou bastante a bateria, sugiro verificar este aspecto.

O fato de ter que ficar com o GPS do celular ligado faz com que a bateria do celular acabe muito rápido.

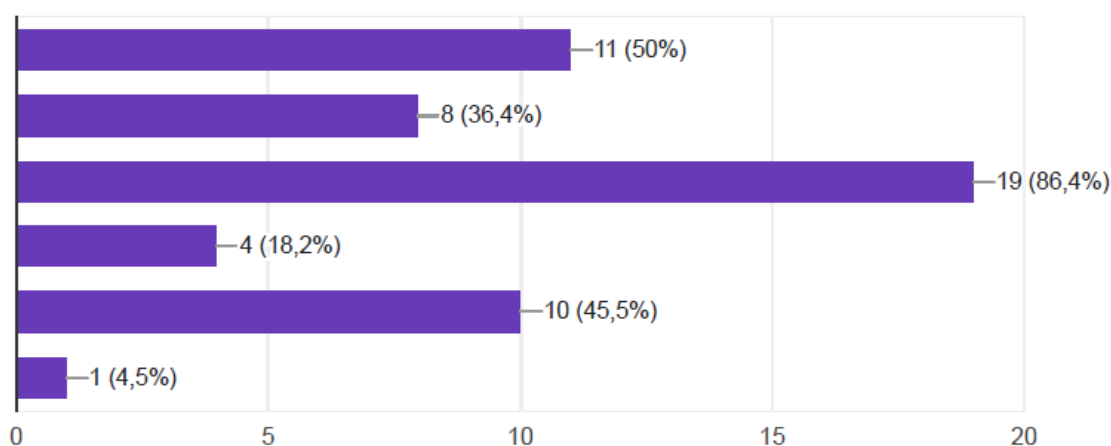
Melhorar o design do aplicativo, algumas mensagens apareciam cortadas. Possibilitar enquanto a campanha esta em andamento que se possa editar os questionários respondidos.

Não utilizei o aplicativo. Foi instalado no meu celular, mas acho que não enviei informações para a pesquisa. Não tenho o que sugerir.

Apresentar melhor os resultados, interface mais intuitiva, melhor descrição da campanha e atividades

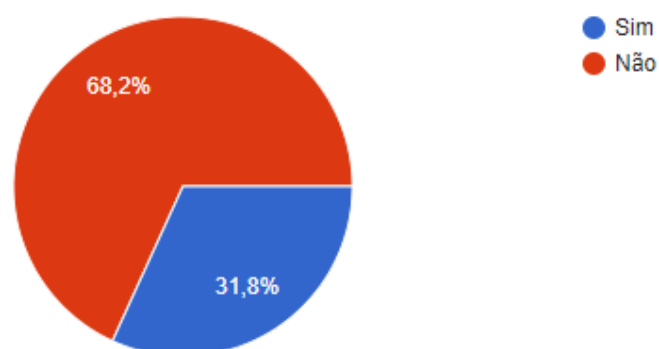
Você se engajou na campanha pois quer:

- ☐ Ajudar no estudo de problemáticas de mobilidade urbana
- ☐ Ajudar no estudo de tecnologias de crowdsensing
- ☐ Ajudar na pesquisa de projeto do Mestrado em Administração
- ☐ Receber informações sobre mobilidade urbana
- ☐ Ajudar nos testes do aplicativo ParticipACT Brasil
- ☐ Outros...



Identificou alguma alteração no consumo de bateria do seu celular durante a utilização do aplicativo?

22 respostas



Fonte: Elaboração própria, 2018.