



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE COMPUTAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**ANTONIO JOSÉ MELO LEITE JÚNIOR**

**UMA ABORDAGEM PARA A INVESTIGAÇÃO DE PADRÕES DE MOVIMENTO E  
DE COMPORTAMENTO DE INDIVÍDUOS EMPREGANDO ANÁLISE VISUAL DE  
PREDICTIVE SUFFIX TREES**

**FORTALEZA**

**2019**

ANTONIO JOSÉ MELO LEITE JÚNIOR

UMA ABORDAGEM PARA A INVESTIGAÇÃO DE PADRÕES DE MOVIMENTO E DE  
COMPORTAMENTO DE INDIVÍDUOS EMPREGANDO ANÁLISE VISUAL DE  
PREDICTIVE SUFFIX TREES

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Computação Gráfica

Orientador: Prof. Creto Augusto Vidal, Ph.D.

Coorientadora: Profa. Emanuele Marques dos Santos, Ph.D.

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

M485a Melo Leite Junior, Antonio José.

Uma Abordagem para a Investigação de Padrões de Movimento e de Comportamento de Indivíduos empregando Análise Visual de Predictive Suffix Trees / Antonio José Melo Leite Junior. – 2019.  
145 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2019.

Orientação: Prof. Dr. Creto Augusto Vidal.

Coorientação: Prof. Dr. Emanuele Marques dos Santos.

1. Predictive Suffix Tree. 2. Análise Visual. 3. Sensemaking. 4. Branching Time. 5. Padrões de Movimento e de Comportamento. I. Título.

CDD 005

---

ANTONIO JOSÉ MELO LEITE JÚNIOR

UMA ABORDAGEM PARA A INVESTIGAÇÃO DE PADRÕES DE MOVIMENTO E DE  
COMPORTAMENTO DE INDIVÍDUOS EMPREGANDO ANÁLISE VISUAL DE  
PREDICTIVE SUFFIX TREES

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Computação Gráfica

Aprovada em:

BANCA EXAMINADORA

---

Prof. Creto Augusto Vidal, Ph.D. (Orientador)  
Universidade Federal do Ceará (UFC)

---

Profa. Emanuele Marques dos Santos,  
Ph.D. (Coorientadora)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. José Antônio F. de Macêdo  
Universidade Federal do Ceará (UFC)

---

Profa. Dra. Aura Conci  
Universidade Federal Fluminense (UFF)

---

Prof. Dr. Anselmo Cardoso de Paiva  
Universidade Federal do Maranhão (UFMA)

À minha hoje grande família.

## AGRADECIMENTOS

Primeiramente, agradeço a Deus e à minha hoje grande família: esposa, Karol, que nunca mediu qualquer esforço para ajudar nesta difícil caminhada; filhas Isabelle e Catarina, que com carinho e atenção retribuíram minhas muitas horas de ausência; pais Melo e Ilsimar, que desde sempre incentivaram meu desenvolvimento, em todos os níveis; irmãs Karine e Caroline, que tanto fizeram para ajudar em tudo o que foi preciso; e sogros Carlos (*in memoriam*) e Eveline, que sempre estiveram presentes.

Agradeço aos meus orientadores Creto, Emanuele, Tonho e Joaquim, que tanto me ensinaram; ao grande amigo George Gomes, que partilhou tanto momentos bons quanto ruins nessa conquista; e a todo o incentivo dos sempre amigos Adriana, Alysson, André, Cadu, Cibelle, Celso, Cleilton, Clemilson, Edgar, Ernesto, Fernando, Gabriel, Geórgia, Henrique Pequeno, Henrique Silva, Inga, Ismael, Lucas, Murakami, Natal, Ney, Rafael e Robson.

Agradeço a todo o apoio recebido do MDCC-Mestrado e Doutorado em Ciência da Computação/Departamento de Computação, do Instituto UFC Virtual e do Curso de Bacharelado em Sistemas e Mídias Digitais da Universidade Federal do Ceará.

E, por fim, agradeço também a Mertens, Einaudi, Glass, Tiersen e tantos outros que me proporcionaram os momentos necessários de calma.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“O homem pensa porque tem mãos.”

(Anaxágoras)



## RESUMO

Predictive Suffix Trees (PSTs) são estruturas de dados capazes de representar simultaneamente espaço, tempo e probabilidade. Elas podem ser utilizadas para prever quando uma pessoa deixará sua posição atual, deslocando-se para um novo provável local. Embora sejam geralmente complexas de ler, PSTs podem ajudar em áreas diversas, tais como investigação criminal; gerenciamento de tráfego rodoviário; ou publicidade baseada em localização, por exemplo. Este trabalho propõe o emprego de análise visual para simplificar a tarefa de encontrar padrões de movimento e de comportamento de indivíduos, com base em dados armazenados em PSTs. Para isso, é apresentada uma abordagem que aplica a metodologia sensemaking e o conceito de branching time para fornecer um caráter menos abstrato às PSTs, permitindo que analistas explorem a dinâmica das combinações espaçotemporais (espaço, tempo e espaço versus tempo), também considerando as probabilidades envolvidas. Para validar a solução proposta, foram desenvolvidas ferramentas de visualização interativa e realizados três testes distintos com usuários, envolvendo um total de 77 participantes e dois datasets específicos. Os resultados obtidos demonstraram a viabilidade da aplicação da solução, permitindo que especialistas e não especialistas resolvesse problemas iniciais, e permitindo que os mesmos também propusessem suas próprias questões a fim de encontrar novas respostas.

**Palavras-chave:** Predictive Suffix Tree. Análise Visual. Sensemaking. Branching Time. Padrões de Movimento. Padrões de Comportamento

## ABSTRACT

Predictive Suffix Trees (PSTs) are data structures capable of simultaneously represent space, time, and probability. They can be used to predict when a person would leave her current position to move to a new probable location. Although they are usually complex to read, PSTs can help crime investigation; management road traffic; or location-based advertising, for example. This work proposes the application of visual analytics to simplify the task of finding movement patterns and possible behaviors of a person, using data stored in PSTs. For that, we introduce an approach that applies sensemaking and branching time to provide a less abstract character to PSTs, allowing analysts to explore the dynamics of space-time combinations (space, time and space versus time relations) considering probabilities. To validate the proposed solution, we developed a visualization tool and performed three distinct user studies, with a total of 77 participants and two different datasets. The obtained results demonstrated the feasibility of applying the solution, allowing expert and non-experts to solve initial problems, but letting they propose their own questions to find new answers.

**Keywords:** Predictive Suffix Tree. Visual Analytics. Sensemaking. Branching Time. Movement Patterns. Behavior Patterns

## LISTA DE FIGURAS

- Figura 1 – Exemplo de uma Probabilistic Suffix Tree (PST), onde cada nó armazena espaço, tempo e probabilidade de uma pessoa estar num local num determinado horário, e cujas arestas representam transições entre as respectivas configurações espaço-tempo-probabilidade. . . . . 22
- Figura 2 – Considerando análise visual, sensemaking e branching time, na abordagem proposta os usuários adotam um workflow específico e operam uma aplicação para inspecionar PSTs, tentando responder perguntas predefinidas e, ao mesmo tempo, estabelecer suas próprias questões para investigação. . . . . 28
- Figura 3 – Visualização de uma Predictive Suffix Tree (PST). O nó selecionado é representado por um marcador branco, enquanto seus nós pais e filhos são representados respectivamente por marcadores em tons de azul e de laranja. A intensidade da cor indica o nível de hereditariedade de cada nó em relação ao nó selecionado. Segmentos de reta coloridos (OD-Links), que seguem a mesma escala cromática anterior, demonstram as relações de hereditariedade entre os nós. Os nós são exibidos em uma grade onde cada célula é colorida em um tom de verde, proporcional ao número de marcadores presentes em sua região. O usuário pode controlar os parâmetros de visualização usando o painel de controle à esquerda. . . . . 29
- Figura 4 – Processo incremental adotado neste trabalho, que considera as contribuições de cada um dos testes procedidos para estabelecer as configurações intermediárias e finais da abordagem e do workflow propostos. . . . . 30
- Figura 5 – Ilustração representando a marcha do exército de Napoleão sobre Moscou, na campanha de 1812-1813, criada por Charles Joseph Minard em 1861. . . . . 35
- Figura 6 – Terminal gráfico IBM 2250 Model 4 com teclado e *light pen*. . . . . 36
- Figura 7 – Exemplo de Lifelines. . . . . 37
- Figura 8 – Exemplo de mural view representando o funcionamento do algoritmo de ordenação *bubble sort*. . . . . 38
- Figura 9 – Dois diferentes arranjos de Dense Pixels Displays, representando o acompanhamento de vinte anos, entre janeiro de 1974 a abril de 1995, de cinquenta empresas no FAZ-Index (índice de ações do mercado alemão): (a) visualização circular de segmentos e (b) tabela de pixels. . . . . 39

Figura 10 – Metáforas visuais baseadas em linhas do tempo coloridas: (a) ThemeRivers e (b) History Flows. . . . .	39
Figura 11 – Interface da aplicação Animal Ecology Explorer. . . . .	40
Figura 12 – Diferentes recursos para a avaliação de ocorrências no tráfego de veículos: (a) observação de velocidades de veículos numa via; (b) seleção de períodos temporais empregando histogramas interativos; e (c) tabela de pixels representando a distribuição de velocidades médias da via no período selecionado. . . . .	42
Figura 13 – Sistema desenvolvido por IVANOV <i>et al.</i> (2007) para vigilância de recintos. Em (a) percebe-se as janelas de visualização sincronizadas, exibindo ocorrências no espaço de acordo com uma linha do tempo. Em (b) é apresentado um dos modos de interação, com base no uso de uma <i>tabletops</i> . . . . .	43
Figura 14 – Sub-trajetórias extraídas de diferentes datasets empregando o framework de LEE <i>et al.</i> (2007). . . . .	43
Figura 15 – Um cluster A sendo dividido em dois outros clusters, B e C, interativamente através da técnica de ANDRIENKO <i>et al.</i> (2009). . . . .	44
Figura 16 – Exemplos de codificações gráficas em mapas: (a) marcadores; (b) segmentos de reta coloridos e (c) tag map associado a tag cloud. . . . .	44
Figura 17 – Fluxo migratório do estado norte-americano do Colorado para outros estados, entre 1995 e 2000. . . . .	45
Figura 18 – OD map demonstrando expectativas migratórias nos Estados Unidos da América, de acordo com o censo norte-americano de 2000. . . . .	46
Figura 19 – Flowstrate representando os fluxos migratórios entre a África Oriental (à esquerda) e a Europa Ocidental (à direita) em 2009. . . . .	47
Figura 20 – Exemplo de uso do sistema proposto por FERREIRA <i>et al.</i> (2013), consultando e comparando viagens iniciadas em aeroportos e acabando em estações de trem na cidade de Nova York. . . . .	48
Figura 21 – Exemplo de integração de dados através de modelos tridimensionais estabelecida em KWAN (2000). . . . .	49
Figura 22 – Apresentação de dados espaçotemporais, de acordo com o modelo de KAPLER; WRIGHT (2005), enfatizando: (a) espaço, representando os principais locais visitados por um indivíduo, e (b) tempo, representando os eventos de uma linha do tempo. . . . .	50

Figura 23 – Exemplo de aplicação, apresentada em EAGLE; PENTLAND (2006), que exibe dados espaçotemporais referentes a ligações entre telefones celulares.	50
Figura 24 – Exemplo de cubo espaçotemporal. . . . .	51
Figura 25 – Exemplo de muro de trajetórias. . . . .	52
Figura 26 – Exemplo de uso da técnica de visualização de CRNOVRSANIN <i>et al.</i> ( 2009): (a) representação gráfica dos deslocamentos de indivíduos numa simulação de evacuação de um prédio devido a uma explosão e (b) gráfico relacionando a distância de cada indivíduo considerado em relação à explosão (eixo horizontal) versus o tempo decorrido (eixo vertical). . . . .	54
Figura 27 – Grafos de visualização de relações de amizade (a) e de proximidade física (b), identificadas entre indivíduos no trabalho realizado por EAGLE; PENTLAND.	55
Figura 28 – Emprego de diferentes técnicas de visualização para avaliação de anomalias em ligações telefônicas por: (a) diagramas de Voronoi associados a gráficos estatísticos e (b) comparação entre tabelas de pixels. . . . .	56
Figura 29 – Dois exemplos de representação tridimensional para identificação de padrões de movimento com base em dados de redes sociais: (a) variações de cubos espaçotemporais e (b) representação de fluxos de comunicação sobre um modelo tridimensional do planeta Terra. . . . .	57
Figura 30 – Dois exemplos de representação tridimensional para identificação de padrões de comportamento: (a) empregando curvas de Bézier num espaço tridimensi- onal, representando ligações entre telefones celulares e (b) através da análise de clusters tridimensionais, representando o fluxo de veículos em uma rodovia.	57
Figura 31 – Exemplos de padrões de movimentação empregados por baleias para alimen- tação, identificados no trabalho de WARE <i>et al.</i> (2006). . . . .	58
Figura 32 – Representação gráfica do conceito de branching time. . . . .	68
Figura 33 – Exemplo de trajetórias oriundas de branching time: (a) conjunto inicial de possibilidades temporais, (b) conjunto derivado de trajetórias e (c) trajetórias individuais. . . . .	68
Figura 34 – A arquitetura da abordagem proposta. . . . .	73

Figura 35 – Elementos referentes à representação visual no mapas virtual: (a) Marcadores em coordenadas geográficas; (b) OD-links que identificam a origem e o destino referentes às arestas de uma PST; (c) Células cujos tons de cor são proporcionais à probabilidade de ocorrência dos respectivos nós da PST representada; e (d) Grupos cujos tons de cor são proporcionais ao número de marcadores presentes em suas respectivas áreas. . . . .	74
Figura 36 – (a) Todos os OD-Links possíveis obtidos de uma determinada PST e (b) os mesmos OD-Links combinados através de edge bundling (ZHOU <i>et al.</i> , 2013).	77
Figura 37 – Exemplo de filtragem empregando f-POIs: (a) acúmulo de pontos correspondentes aos marcadores; (b) POIs primários; (c) POIs primários e secundários e (d) todos os possíveis POIs obtidos. . . . .	78
Figura 38 – Exemplo de filtragem empregando f-Routes: (a) acúmulo de segmentos de reta correspondentes aos OD-Links; (b) configuração inicial de rota; (c) possíveis rotas utilizadas mais frequentemente e (d) todas as rotas obtidas. .	78
Figura 39 – Exemplo de filtragem empregando f-Times: (a) local/probabilidade de um indivíduo estar localizado às 5:00 e (b) às 10:00. . . . .	79
Figura 40 – Exemplo de anotação usando (a) o subsistema de anotação, (b) detalhe de inclusão de uma ilustração esquemática e de rótulo de texto no mapa virtual e (c) fotografia digital oriunda de serviço online. . . . .	80
Figura 41 – Workflow definido para uso da abordagem proposta. . . . .	81

Figura 42 – A interface gráfica da aplicação desenvolvida: A - abrir um arquivo JSON com os dados da PST; B - exibir o overview dos dados da PST; C - selecionar filtros para aplicar; D - exibir o resultado do processo de edge bundling; E - acessar um subsistema de anotação simples; F - selecionar um nó específico da PST; G - exibir a hierarquia dos nós da PST em uma tree list; H - usar uma escala cromática para indicar relações de hereditariedade, referentes a nós pais e filhos, no mapa; I - exibir ou ocultar marcadores, OD-links, células e grupos; J - definir o número de níveis hierárquicos a serem exibidos nas relações pais/filhos; K - selecionar o modo padrão/visão de satélite do mapa virtual; L - visualizar os marcadores correspondentes aos pais do nó selecionado na tree list; M - exibir um marcador correspondente ao item selecionado na tree list; e N - visualizar os marcadores correspondentes aos filhos do nó selecionado na tree list. . . . .	87
Figura 43 – Exemplo de diferenciação de nós pais e filhos, adotando escala cromática para coloração dos respectivos marcadores de acordo com relações de hereditariedade. . . . .	88
Figura 44 – Recursos adicionais para anotação: (a) inclusão de ilustrações esquemáticas, baseada em integração com editor gráfico de terceiros e (b) editor para consulta e alteração de histórias. . . . .	91
Figura 45 – Protótipo inicial da aplicação, desenvolvido especificamente para a realização do primeiro teste com usuários: (a) painel de controle e (b) mapa virtual. . .	94
Figura 46 – Arquitetura do protótipo empregado no primeiro teste com usuários. . . . .	95
Figura 47 – Overview da PST do primeiro usuário avaliado. . . . .	99
Figura 48 – Descoberta de um padrão de movimento com a aplicação desenvolvida: (a) uso de overview para identificar os locais mais frequentes (marcadores brancos e células verdes) e seus respectivos relacionamentos (OD-Links verdes) e (b) ampliação dessa região de interesse para realização de exploração mais detalhada da área tracejada em (a). . . . .	101
Figura 49 – Overview da PST do segundo usuário avaliado. . . . .	102
Figura 50 – Overview parcial da PST do terceiro usuário avaliado. . . . .	103
Figura 51 – Overview total da PST do terceiro usuário avaliado. . . . .	104

Figura 52 – Notas médias (entre 0,0 e 10,0) obtidas nas avaliações realizadas pelos grupos de controle e experimental durante as etapas de pré-teste e de pós-teste. . . . .	110
Figura 53 – Quantidades de respostas válidas nos questionários da etapa de teste. . . . .	111
Figura 54 – Overview da PST do primeiro táxi avaliado. . . . .	117
Figura 55 – Major hubs encontrados na PST do primeiro táxi avaliado. . . . .	117
Figura 56 – Overview da PST do segundo táxi avaliado. . . . .	118
Figura 57 – Agregação de dados, obtidos a partir do uso do filtro f-Times, para definição dos dois conjuntos distintos de rotas principais referentes ao segundo táxi avaliado: (a) percursos diurnos e (b) percursos noturnos. . . . .	119
Figura 58 – Overview da PST do terceiro táxi avaliado. . . . .	120



## **LISTA DE TABELAS**

Tabela 1 – Sumarização de dados dos três testes realizados com usuários . . . . .	121
---	-----

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>21</b>
<b>1.1</b>	<b>Motivação</b>	<b>21</b>
<b>1.2</b>	<b>Caracterização do Problema</b>	<b>23</b>
<i>1.2.1</i>	<i>Por que alguém visita um determinado local?</i>	<i>24</i>
<i>1.2.2</i>	<i>Como alguém investiga uma ocorrência empregando PSTs?</i>	<i>25</i>
<i>1.2.3</i>	<i>Quais são os tipos de padrões esperados nos dados representados por uma PST?</i>	<i>26</i>
<b>1.3</b>	<b>Objetivos</b>	<b>28</b>
<b>1.4</b>	<b>Metodologia</b>	<b>30</b>
<b>1.5</b>	<b>Contribuições</b>	<b>31</b>
<b>1.6</b>	<b>Publicações</b>	<b>31</b>
<b>1.7</b>	<b>Organização</b>	<b>32</b>
<b>1.8</b>	<b>Considerações Finais</b>	<b>32</b>
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>34</b>
<b>2.1</b>	<b>Representação Visual de Dados Espaço-temporais e Atributos</b>	<b>34</b>
<i>2.1.1</i>	<i>Primórdios e Avanços Iniciais</i>	<i>34</i>
<i>2.1.2</i>	<i>Representação Interativa</i>	<i>40</i>
<i>2.1.3</i>	<i>Representação por Clusters</i>	<i>42</i>
<i>2.1.4</i>	<i>Representações de Modelos Origem-Destino</i>	<i>45</i>
<i>2.1.5</i>	<i>Representações Tridimensionais</i>	<i>47</i>
<b>2.2</b>	<b>Identificação de Padrões de Movimento e de Comportamento a partir de Representações Gráficas</b>	<b>52</b>
<i>2.2.1</i>	<i>Representações Bidimensionais</i>	<i>52</i>
<i>2.2.2</i>	<i>Representações Tridimensionais</i>	<i>55</i>
<b>2.3</b>	<b>Considerações Finais</b>	<b>58</b>
<b>3</b>	<b>CONCEITOS PRELIMINARES</b>	<b>60</b>
<b>3.1</b>	<b>Predição de Localização por Predictive Suffix Trees</b>	<b>60</b>
<i>3.1.1</i>	<i>Predição de Localização</i>	<i>60</i>
<i>3.1.2</i>	<i>Predictive Suffix Trees</i>	<i>61</i>
<b>3.2</b>	<b>Análise Visual</b>	<b>62</b>

3.3	Sensemaking . . . . .	63
3.4	Representação de Branching Time . . . . .	66
3.4.1	<i>Modos de organização do tempo</i> . . . . .	67
3.4.2	<i>Branching Time</i> . . . . .	67
3.4.3	<i>Branching Time x Trajetórias</i> . . . . .	68
3.5	Considerações Finais . . . . .	69
4	<b>SOLUÇÃO PROPOSTA</b> . . . . .	70
4.1	Fundamentação . . . . .	70
4.2	Arquitetura . . . . .	72
4.2.1	<i>Extração</i> . . . . .	73
4.2.2	<i>Representação</i> . . . . .	73
4.2.3	<i>Interação</i> . . . . .	75
4.2.4	<i>Filtragem</i> . . . . .	77
4.2.5	<i>Anotação</i> . . . . .	79
4.3	Workflow . . . . .	80
4.3.1	<i>Identificação</i> . . . . .	81
4.3.2	<i>Exploração</i> . . . . .	81
4.3.3	<i>Documentação</i> . . . . .	82
4.4	Processo Cíclico de Avaliação . . . . .	83
4.5	Considerações Finais . . . . .	84
5	<b>APLICAÇÃO DESENVOLVIDA</b> . . . . .	85
5.1	Contextualização . . . . .	85
5.2	Interface Gráfica . . . . .	86
5.3	Funcionalidades . . . . .	86
5.4	Subsistema de Anotação . . . . .	91
5.5	Considerações Finais . . . . .	92
6	<b>VALIDAÇÃO DA SOLUÇÃO</b> . . . . .	93
6.1	Objetivos dos Testes . . . . .	93
6.2	Primeiro Teste com Usuários . . . . .	94
6.2.1	<i>Análise da Primeira PST</i> . . . . .	99
6.2.2	<i>Análise da Segunda PST</i> . . . . .	101
6.2.3	<i>Análise da Terceira PST</i> . . . . .	103

6.2.4	<i>Avaliação de Resultados</i> . . . . .	105
6.2.5	<i>Contribuições à Solução Proposta</i> . . . . .	106
6.3	<b>Segundo Teste com Usuários</b> . . . . .	107
6.3.1	<i>Pré-Teste</i> . . . . .	108
6.3.2	<i>Teste</i> . . . . .	109
6.3.3	<i>Pós-Teste</i> . . . . .	110
6.3.4	<i>Análise de Resultados</i> . . . . .	110
6.3.4.1	<i>Análise Quantitativa</i> . . . . .	110
6.3.4.2	<i>Análise Qualitativa</i> . . . . .	111
6.3.5	<i>Contribuições à Solução Proposta</i> . . . . .	114
6.4	<b>Terceiro Teste com Usuários</b> . . . . .	114
6.4.1	<i>Análise da Primeira PST</i> . . . . .	116
6.4.2	<i>Análise da Segunda PST</i> . . . . .	118
6.4.3	<i>Análise da Terceira PST</i> . . . . .	119
6.5	<b>Discussão Geral</b> . . . . .	121
6.5.1	<i>Validade de Respostas</i> . . . . .	121
6.5.2	<i>Duração das Tarefas</i> . . . . .	123
6.5.3	<i>Quantidade de Histórias Registradas</i> . . . . .	124
6.5.4	<i>Outras Observações</i> . . . . .	124
6.6	<b>Considerações Finais</b> . . . . .	126
7	<b>CONCLUSÕES</b> . . . . .	127
7.1	<b>Resumo</b> . . . . .	127
7.2	<b>Resultados Alcançados</b> . . . . .	128
7.3	<b>Limitações</b> . . . . .	129
7.4	<b>Trabalhos Futuros</b> . . . . .	130
7.5	<b>Considerações Finais</b> . . . . .	131
	<b>REFERÊNCIAS</b> . . . . .	132
	<b>APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLA- RECIDO UTILIZADO NO PRIMEIRO TESTE COM USUÁRIOS</b> . . . . .	144
	<b>APÊNDICE B – QUESTIONÁRIO UTILIZADO NO SEGUNDO TESTE COM USUÁRIOS</b> . . . . .	145

<b>APÊNDICE C – QUESTIONÁRIO UTILIZADO NO TERCEIRO TESTE COM USUÁRIOS . . . . .</b>	<b>146</b>
---	------------

# 1 INTRODUÇÃO

A presente tese propõe uma abordagem para a investigação de padrões de movimento e de comportamento de indivíduos a partir da análise visual dos dados de *predictive suffix trees*, adotando a metodologia *sensemaking* e o conceito de *branching time*.

Neste capítulo, inicialmente a Seção 1.1 contextualiza o trabalho, apresentando sua motivação. Depois a Seção 1.2, com base em *sensemaking*, caracteriza devidamente o problema investigado. A Seção 1.3 expõe os objetivos principal e específicos. A Seção 1.4 discute as etapas metodológicas envolvidas na execução do trabalho. A Seção 1.5 aborda brevemente as contribuições alcançadas. Posteriormente, a Seção 1.6 lista as publicações geradas a partir desse trabalho e por fim, a Seção 1.7 explica como estão organizados os demais capítulos do trabalho. E, por fim, na Seção 1.8 são tecidas considerações finais acerca do Capítulo.

## 1.1 Motivação

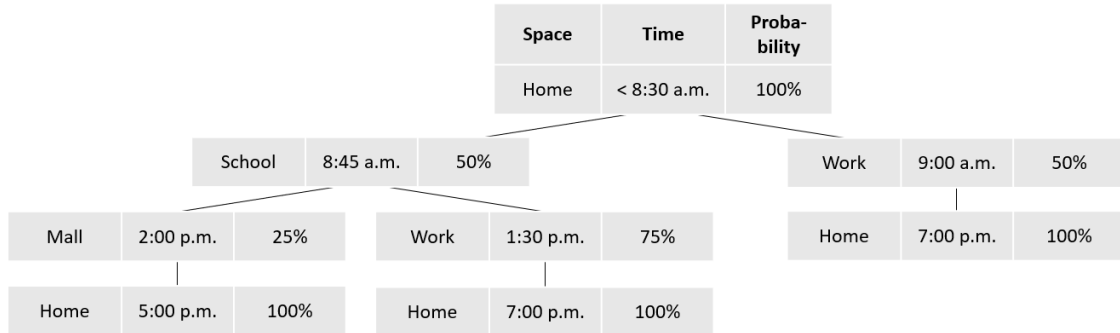
Nos últimos anos, o uso intensivo de *smartphones* tem possibilitado a coleta de dados de pessoas, principalmente suas posições (espaço), as quais podem ser obtidas com auxílio de GPS, e os intervalos temporais em que essas posições foram ocupadas (tempo). O processamento desses dados espaçotemporais, com a finalidade de investigar padrões de movimento e de comportamento de indivíduos, define atualmente toda uma série de aplicações relevantes, que abrangem desde investigação criminal (BROWN *et al.*, 2007) e gestão de vias com acidentes de trânsito (HU; WANG, 2011) até recomendação de dados em mecanismos de busca na web e em redes sociais (BAO *et al.*, 2015) e publicidade baseada em localização (SCCELLATO *et al.*, 2011), por exemplo.

Considerando esses tipos de aplicações, ROCHA *et al.* (2016) propuseram um modelo para prever quando uma pessoa deixaria sua posição atual para se deslocar para um novo local. A principal finalidade desse modelo é aprimorar a confiabilidade das previsões, indicando não apenas *onde* uma pessoa pode ir, algo comum a outros modelos preditivos (MONREALE *et al.*, 2009; XUE *et al.*, 2013; LEI *et al.*, 2013), mas também *quando* ela deve deixar sua localização atual. Dessa forma, pode-se obter ganhos expressivos na qualidade de serviços baseados em localização.

Para tanto, o modelo de Rocha e seus coautores emprega Predictive Suffix Trees, PSTs (Figura 1), estruturas de dados capazes de representar simultaneamente espaço, tempo e

probabilidade (RON *et al.*, 1994).

Figura 1 – Exemplo de uma Probabilistic Suffix Tree (PST), onde cada nó armazena espaço, tempo e probabilidade de uma pessoa estar num local num determinado horário, e cujas arestas representam transições entre as respectivas configurações espaço-tempo-probabilidade.



Fonte: o autor.

No entanto, apesar do grande potencial, geralmente as PSTs são difíceis de avaliar, pois, enquanto se lê e analisa várias linhas de texto preenchidas com dados de latitudes, longitudes, registros de data/hora e porcentagens, dentre outros, é necessário lidar simultaneamente com o espaço, com o tempo e com as probabilidades relacionadas (LEI *et al.*, 2013).

Dessa forma, dado que representações visuais, sejam bidimensionais ou tridimensionais, podem auxiliar a análise de dados espacotemporais (TOMINSKI *et al.*, 2012), uma representação gráfica adequada dos dados armazenados em PSTs pode ser benéfica para tornar sua interpretação mais simples e eficiente. Porém, a maioria das técnicas tradicionais de visualização que combinam dados referentes a espaço, tempo e probabilidades não são muito eficazes, devido a possível poluição visual, onde o excesso de elementos pode acabar atrapalhando o processo de avaliação como um todo (PURCHASE, 1997; AIGNER *et al.*, 2008; SIOW *et al.*, 2018). E, em casos extremos como a exibição simultânea de todos os dados espaciais, temporais e probabilísticos inter-relacionados de uma PST, pode até mesmo ocorrer o fenômeno da oclusão visual, onde dados específicos são sobrepostos por outros, podendo haver confusão na investigação das informações representadas. Assim, com tais restrições, pode ser bastante difícil detectar onde e quando as principais ocorrências estão e quais detalhes (posições, horários, trajetórias etc) devem ser observados mais minuciosamente (ANDRIENKO *et al.*, 2010).

Da mesma forma, soluções interativas podem ser de grande valia para a análise de PSTs, uma vez que o consequente maior controle sobre os dados representados geralmente auxilia na compreensão das respectivas informações armazenadas (SIOW *et al.*, 2018). No

entanto, embora existam várias técnicas para a interação direta com os dados armazenados em árvores de um modo geral (SCHULZ, 2011), as peculiaridades das PSTs, com as inter-relações entre seus nós e arestas, requerem uma solução mais específica.

Assim, com base no modelo preditivo de ROCHA *et al.* (2016), o presente trabalho propõe uma solução de representação visual e interativa específica, explorando os conceitos de análise visual, sensemaking e branching time.

Contextualizando inicialmente, a análise visual objetiva elaborar soluções que explorem e combinem as capacidades de processamento de dados de humanos e de máquinas, permitindo que analistas ultrapassem suas competências de interpretação natural. Já de acordo com o conceito de sensemaking, a análise de dados deve se apoiar na devida contextualização de elementos, para ajudar o analista a compreender ideias abstratas mais facilmente, através de associações com sua própria realidade. E, por sua vez, o branching time estabelece que uma ocorrência probabilística pode ser um ponto a partir do qual dois ou mais eventos possíveis se originam.

Então, combinando análise visual, sensemaking e branching time, busca-se expandir e, ao mesmo tempo, simplificar o escopo de análise de PSTs. Assim, em vez de apenas prever movimentos, permite-se que a avaliação de ocorrências nas PSTs proporcione a analistas, especialistas e não especialistas, a capacidade de identificarem diversos possíveis ou prováveis padrões de movimento e de comportamento de indivíduos.

Dessa forma, dada a devida identificação de tais padrões, além de atenderem às áreas de investigação já citadas, as PSTs podem passar a ser empregadas também para avaliações e análises em diversos outros campos, apresentando possíveis impactos desde na prevenção de epidemias, pela compreensão do espalhamento de agentes biológicos, e na respectiva resposta a emergências, até no planejamento urbano, comercial e financeiro, no fornecimento de conteúdos *online* sob demanda, de acordo com perfis de uso e espaço e tempo, e na modelagem e desenvolvimento de simulações diversas, por exemplo (MARBLE *et al.*, 1997; GONZALEZ *et al.*, 2008; BOYANDIN *et al.*, 2011; SIOW *et al.*, 2018).

## 1.2 Caracterização do Problema

Identificar ocorrências significativas em conjuntos multivariados de trajetórias inter-relacionadas, como as PSTs, geralmente é uma tarefa complicada, porque é complexo agregar espaço, tempo e atributos (as probabilidades, no caso das PSTs) simultaneamente (ANDRIENKO



*et al.*, 2010).

No entanto, apesar da complexidade envolvida, a correta análise dos três componentes associados das PSTs, ou seja, tempo, espaço e probabilidades, pode fornecer diversas informações úteis. Porém, geralmente os métodos tradicionais de visualização interativa não são suficientemente robustos para permitir ao usuário entender os dados representados de forma rápida e efetiva, especialmente no que tange à dinâmica dos atributos inter-relacionados (MARBLE *et al.*, 1997; ANDRIENKO *et al.*, 2010; BOYANDIN *et al.*, 2011). Além disso, geralmente não é simples escolher os devidos pontos representados de origem e de destino a investigar, uma vez que é necessário considerar as múltiplas relações existentes entre os nós e as arestas das PSTs.

No que diz respeito especificamente à aplicação de sensemaking ao problema abordado neste trabalho, isto é, a investigação de padrões de movimento e de comportamentos de indivíduos empregando análise visual de PSTs, entende-se que o modo como os analistas obtêm a informação precisa ser devidamente avaliado antes, sob a ótica do próprio analista, e não simplesmente a partir da perspectiva do observador (PIROLI; CARD, 2005). Portanto, para propor uma solução adequada, deve-se primeiramente conhecer as reais variáveis de interesse para investigação.

Assim, com base na dicotomia estabelecida entre o fazer e o compreender (discutida posteriormente na Seção 3.3 deste trabalho), foram estabelecidas três perguntas iniciais para dar suporte à análise dessas necessidades reais: 1) Por que alguém visita um determinado local? 2) Como alguém investiga uma ocorrência empregando PSTs? e 3) Quais são os tipos de padrões esperados nos dados representados por uma PST?

As respostas a essas questões levantadas são discutidas a seguir, devidamente embasadas na literatura, para definir as variáveis de interesse e suas respectivas importâncias no processo de análise visual pretendido.

### ***1.2.1 Por que alguém visita um determinado local?***

É evidente que pessoas diferentes possuem motivos distintos para ir a um determinado lugar, como, por exemplo, curiosidade, diversão, trabalho, transporte etc (ZHENG *et al.*, 2011). Nesse ínterim, YING *et al.* (2013) categorizam as intenções de um usuário estar em algum lugar em três classes:

- Intenções Acionadas Geograficamente (*Geographically-Triggered Intentions*). Essa classe

se refere a casos em que um indivíduo visita um determinado local devido a uma relação puramente geográfica. Por exemplo, em uma viagem de metrô da estação A para a estação C, uma pessoa pode ter que passar necessariamente pela estação B;

- **Intenções Acionadas Temporalmente (*Temporal-Triggered Intentions*)**. Essa classe se refere à relação direta entre localização e informação temporal, envolvendo as razões pelas quais um indivíduo chega em ou sai de um certo local num determinado momento. Por exemplo, quando uma pessoa vai ao trabalho, ela geralmente sai de casa em um horário específico pela manhã, retornando para casa em um outro horário, normalmente à tarde ou à noite;
- **Intenções Acionadas Semanticamente (*Semantic-Triggered Intentions*)**. Essa classe se refere aos motivos pelos quais pessoas se deslocam de um local a outro. Assim, por exemplo, como uma pessoa geralmente escolhe um local para ir com base em suas necessidades no momento, quando ela vai almoçar, provavelmente irá a um lugar que possua restaurantes ou à sua própria casa.

No geral, porém, é importante observar que o movimento de um indivíduo de um local para outro normalmente é acionado por várias intenções simultaneamente. Assim, frequentemente, os deslocamentos podem ser vistos como um contexto combinado dos comportamentos que são motivados por uma união dessas três classes de intenções (YING *et al.*, 2013). Por exemplo, numa viagem de carro para fins de turismo a uma cidade distante, uma pessoa pode preferir sair pela manhã cedo, para evitar congestionamentos. De todo modo, porém, terá de passar por diversas outras cidades até chegar ao seu destino final. Nesse exemplo, passar por diversas cidades pode ser classificado como uma intenção acionada geograficamente; sair cedo, intenção acionada temporalmente; e viajar a turismo, intenção acionada semanticamente.

### ***1.2.2 Como alguém investiga uma ocorrência empregando PSTs?***

Para responder a essa pergunta, é interessante estender a definição original das três tarefas relacionadas à investigação de ocorrências em modelos espaçotemporais, estabelecidas por ANDRIENKO *et al.* (2010). Assim, incluindo considerações acerca de dados de probabilidade para contemplar PSTs, essas três tarefas seriam:

- **Caracterização de Ocorrências**. Consiste em observar a distribuição de probabilidades ao longo do tempo e do espaço, ou em porções específicas de tempo e de espaço, e caracterizar o comportamento dos atributos considerados. Assim, os comportamentos ob-

servados, em relação às probabilidades envolvidas, podem ser descritos como: constantes ou parcialmente constantes em regiões espaciais específicas ou em intervalos de tempo específicos; graduais ou abruptos, exibindo tendências temporais ou espaciais; periódicos no espaço ou no tempo etc. Um exemplo de uma tarefa de caracterização de ocorrência é o monitoramento de pares { *local visitado* , *tempo de visita* } de uma determinada pessoa durante um certo período, considerando as respectivas probabilidades de ocorrência;

- **Busca de Ocorrências.** Consiste em detectar as ocorrências de um determinado comportamento e determinar como as respectivas probabilidades podem se encontrar em relação ao tempo e ao espaço. Um exemplo de tarefa de busca de ocorrência é a investigação das probabilidades de chegada e de partida de uma determinada pessoa em sua residência em períodos comumente identificáveis, como o início e o final do dia;
- **Comparação de Ocorrências.** Consiste em comparar certas configurações de dados, considerando os valores das probabilidades em diferentes regiões do espaço ou em diferentes intervalos de tempo. Um exemplo de tarefa de comparação de ocorrências é, para uma determinada pessoa, comparar as probabilidades de chegada em e de saída de sua residência, em certos momentos, mantendo registros diferenciados para eventos em dias úteis e em fins de semana.

### ***1.2.3 Quais são os tipos de padrões esperados nos dados representados por uma PST?***

Como não foi encontrado nenhum trabalho que discutisse especificamente a resposta a essa questão, foram aceitos como uma possível solução os três tipos de padrões que os analistas podem revelar e explorar em modelos espaçotemporais (ANDRIENKO *et al.*, 2017):

- **Padrões Temporais.** Consistem em intervalos temporais similares de ocorrências; pequenas e grandes mudanças em intervalos temporais de ocorrências; ruptura de continuidade; repetição periódica em intervalos temporais de ocorrências, referindo-se a ciclos temporais (diário, semanal e sazonal); e valores temporais discrepantes;
- **Padrões espaciais.** Consistem em *major hubs*, locais que apresentam fluxos bem definidos de entrada ou de saída de elementos; tendências de fluxos espaciais, predominando determinados fluxos em certas direções ou a certas distâncias; e regiões de atração ou de repulsão, onde prevalecem significativamente determinados fluxos de entrada ou de saída de elementos;
- **Padrões espaciais de mudanças.** Consistem em locais nos quais os fluxos de saída ou

de entrada aumentam ou diminuem consideravelmente; e regiões em que há mudanças expressivas de direções ou de distanciamentos em fluxos.

Além dos padrões anteriores, extrapolando-se os modelos espaços-temporais, pode-se ainda citar os seguintes padrões identificáveis na investigação de PSTs:

- Padrões de Trajetória. Consistem em um conjunto de trajetórias individuais que compartilham a propriedade de visitar a mesma sequência de locais com tempos de viagem semelhantes (GIANNOTTI *et al.*, 2007);
- Padrões de Movimento. Consistem em qualquer regularidade espacial e temporal reconhecível ou qualquer relacionamento incomum em um conjunto de dados relativos a deslocamentos, considerando algum domínio de aplicação específico, como a interpretação do cotidiano de uma pessoa ou estudos de migrações de animais, por exemplo (DODGE *et al.*, 2008);
- Padrões de Comportamento. Consistem em correlações convergentes entre movimentos, ações ou situações de alguém ou algo que possam ser úteis para prever as respectivas movimentação ou permanência (SONG *et al.*, 2010).

Com base em todas as possíveis respostas à pergunta sobre quais são os tipos de padrões esperados nos dados representados por uma PST, para este trabalho, assume-se que esse tipo particular de árvore armazena grandes padrões de trajetórias inter-relacionadas, e que analistas podem identificar tais padrões de trajetória, podendo estendê-los, a fim de inferir padrões de movimento ou padrões de comportamento de indivíduos.

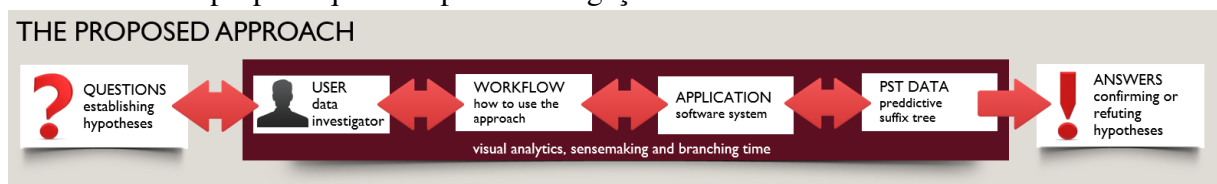
E, considerando as probabilidades envolvidas, bem como outros possíveis indícios observados na investigação, o analista pode classificar tais padrões de comportamento e de movimento como sendo, então, possíveis ou prováveis.

Por fim, dada a associação direta entre homem e máquina preconizada pela análise visual, é importante pontuar que o problema investigado nesta tese, além de abranger as respostas às perguntas anteriormente levantadas, também deve envolver outras duas demandas específicas. A primeira demanda diz respeito à preocupação constante com possíveis questões de desempenho, pois pode-se inviabilizar o uso das plataformas de hardware e de software empregadas caso não se pondere devidamente a quantidade de elementos visuais/interativos utilizados. E a segunda demanda refere-se à simplificação do uso das respectivas aplicações desenvolvidas com base na solução proposta, buscando-se minimizar a complexidade no controle desses mesmos elementos visuais/interativos, a fim de se evitar a poluição e a oclusão visuais.

### 1.3 Objetivos

O objetivo deste trabalho é propor uma solução, incluindo abordagem e respectivo *workflow*, baseada em análise visual, sensemaking e branching time, visando simplificar a avaliação, por usuários especialistas e não especialistas, dos dados representados em PSTs, considerando as inter-relações entre os mesmos (Figura 2).

Figura 2 – Considerando análise visual, sensemaking e branching time, na abordagem proposta os usuários adotam um workflow específico e operam uma aplicação para inspecionar PSTs, tentando responder perguntas predefinidas e, ao mesmo tempo, estabelecer suas próprias questões para investigação.



Fonte: o autor.

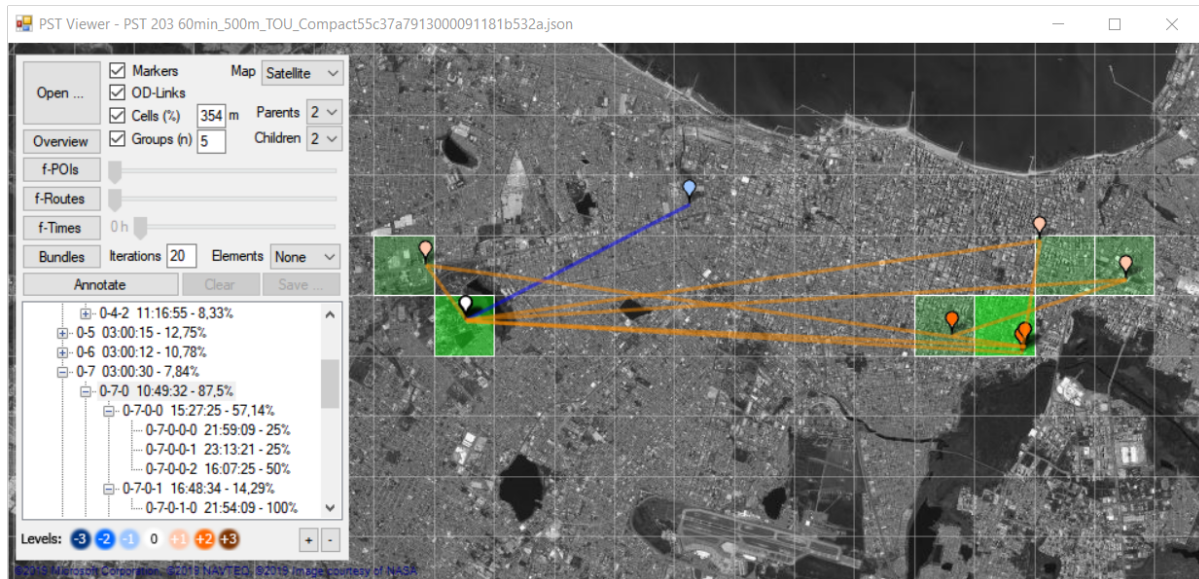
Assim, para o problema específico discutido na seção anterior, ou seja, identificar padrões de movimento e de comportamento de indivíduos com base na análise visual de PSTs, esse tipo particular de árvore passa a definir não mais apenas trajetórias possíveis, mas também ocorrências possíveis.

Dessa forma, a abordagem proposta define componentes específicos para extração, representação, interação, filtragem e anotação das informações obtidas a partir dos dados de uma PST. Além disso, é definido um workflow específico para otimizar o uso dos recursos estabelecidos em cada um desses componentes, considerando operações de identificação, exploração e documentação de ocorrências.

Para avaliar a eficácia da solução proposta, considerando tanto a abordagem quanto o respectivo workflow, foram desenvolvidas aplicações específicas (Figuras 3 e 45), empregando dois *datasets* distintos: o primeiro com dados de estudantes de uma universidade e o segundo referente a viagens de táxi. Posteriormente, para validar o uso dessas aplicações, foram realizados três testes com usuários, envolvendo um total de setenta e sete analistas, sendo quatro especialistas e setenta e três não especialistas. Para tanto, foram adotadas metodologias específicas, baseadas na execução de tarefas predefinidas, e os participantes foram estimulados a propor seus próprios questionamentos acerca dos dados representados, buscando sempre enriquecer o conjunto de descobertas realizadas.

Com a avaliação dos resultados desses testes com usuários, de uma forma geral,

Figura 3 – Visualização de uma Predictive Suffix Tree (PST). O nó selecionado é representado por um marcador branco, enquanto seus nós pais e filhos são representados respectivamente por marcadores em tons de azul e de laranja. A intensidade da cor indica o nível de hereditariedade de cada nó em relação ao nó selecionado. Segmentos de reta coloridos (OD-Links), que seguem a mesma escala cromática anterior, demonstram as relações de hereditariedade entre os nós. Os nós são exibidos em uma grade onde cada célula é colorida em um tom de verde, proporcional ao número de marcadores presentes em sua região. O usuário pode controlar os parâmetros de visualização usando o painel de controle à esquerda.

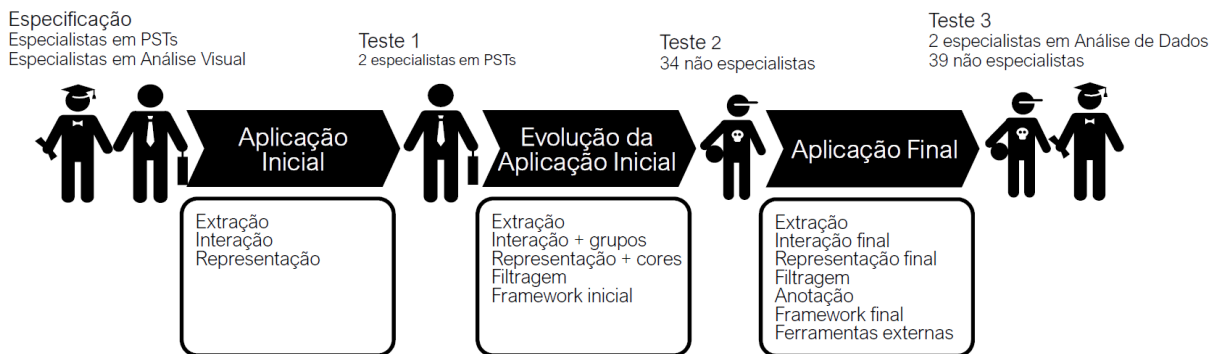


Fonte: o autor.

percebeu-se que os participantes encontraram efetivamente possíveis padrões de movimento e de comportamento dos indivíduos representados nas PSTs utilizadas (estudantes universitários e motoristas de táxi).

Por fim, é importante destacar que, adotando um processo incremental (Figura 4), as contribuições de cada um desses testes procedidos auxiliaram no estabelecimento das configurações intermediárias e finais da abordagem e do workflow aqui propostos. Do mesmo modo, esse processo incremental também incluiu, em paralelo, a realização, pelo próprio autor deste trabalho, de operações de investigação sobre diversas PSTs oriundas dos respectivos datasets utilizados nos testes. Operações essas que foram procedidas tanto para o estabelecimento das tarefas necessárias aos testes quanto para a experimentação de possibilidades e respectiva geração de ideias.

Figura 4 – Processo incremental adotado neste trabalho, que considera as contribuições de cada um dos testes procedidos para estabelecer as configurações intermediárias e finais da abordagem e do workflow propostos.



Fonte: o autor.

## 1.4 Metodologia

O desenvolvimento do presente trabalho foi realizado de acordo com as seguintes etapas metodológicas:

1. Estudo inicial de conteúdos gerais e específicos, referentes às áreas de predição de movimentos, análise visual, sensemaking, branching time e padrões de movimento e de comportamento de indivíduos;
2. Realização de revisão bibliográfica e levantamento de trabalhos relacionados acerca dos tópicos identificados na etapa anterior;
3. Caracterização do problema abordado neste trabalho, com base na metodologia de sensemaking e na discussão de ideias e conceitos com especialistas nas áreas de análise visual e de predição de movimentos empregando PSTs;
4. Especificação de requisitos para definição da solução proposta, baseando-se na evolução de possibilidades de abordagens, arquiteturas e seus componentes, workflows e respectivas aplicações para validação dos resultados intermediários e finais alcançados;
5. Estabelecimento dos modelos de construção necessários à elaboração de protótipos das aplicações;
6. Realização de operações de investigação, pelo próprio autor, como experimentos de investigação e para o estabelecimento das tarefas necessárias à avaliação de uso, sobre PSTs oriundas dos datasets utilizados nos testes com usuários;
7. Avaliação do uso dos protótipos em experimentos baseados em testes com usuários, envolvendo participantes especialistas e não especialistas;
8. Adoção de processo incremental de avaliação das contribuições advindas, paralelamente,

dos testes com usuários realizados e das operações de investigação próprias, para aprimoramento das soluções intermediárias e final propostas;

9. Definição de versão final da solução proposta, devidamente balizada pela análise da evolução do processo incremental definido na etapa anterior; e
10. Escrita final do trabalho.

## 1.5 Contribuições

De forma resumida, as principais contribuições desse trabalho são:

- Propor uma abordagem baseada em análise visual, sensemaking e branching time para auxiliar analistas a investigarem PSTs e a descobrirem padrões de movimento e de comportamento de indivíduos;
- Estabelecer um workflow voltado a otimizar o uso da abordagem proposta;
- Desenvolver diferentes versões de uma aplicação de análise visual, com base na solução proposta (abordagem e workflow), empregando dois datasets distintos;
- Realizar testes com usuários das versões da aplicação desenvolvida, envolvendo analistas especialistas e não especialistas, a fim de avaliar a eficácia das mesmas e, consequentemente, da própria solução proposta; e
- Analisar os resultados dos testes com usuários e discutir como a adoção de sensemaking e de branching time pode ter influenciado o modo como os respectivos participantes realizaram suas descobertas e propuseram seus próprios questionamentos.

Portanto, considera-se que o conteúdo desta tese é relevante porque, dada a complexidade da representação gráfica e interativa das PSTs, é importante estabelecer instrumentos eficientes para a análise visual das ocorrências relacionadas, buscando uma melhor compreensão de todos os dados envolvidos.

Assim, a abordagem aqui proposta permite a criação de instrumentos mais eficientes e diretos para explorar a dinâmica de combinações de dados espaçotemporais (espaço, tempo e espaço versus tempo) considerando simultaneamente as probabilidades associadas, então consequentemente simplificando a análise de PSTs.

## 1.6 Publicações

Ainda como resultado deste trabalho já foram geradas as seguintes publicações:



1. LEITE, A. J. M.; SANTOS, E.; VIDAL, C. A.; MACÊDO, J. A. F. D. Visual analysis of predictive suffix trees for discovering movement patterns and behaviors. In: **IEEE.201730th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S. l.], 2017. p.103–110.
2. LEITE, A. J. M.; SANTOS, E.; VIDAL, C.; ROCHA, C. Empregando análise visual e sensemaking no ensino de predictive suffix trees. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S. l.: SBC, 2018. v. 29, n. 1,p. 1043.

## 1.7 Organização

Este capítulo contextualizou o presente trabalho, incluindo as motivações para o mesmo, a caracterização do problema abordado, os objetivos estabelecidos, as etapas metodológicas procedidas e as contribuições esperadas. O restante desta tese encontra-se organizado da seguinte forma. No Capítulo 2 são apresentados os trabalhos relacionados ao tema abordado. No Capítulo 3 são discutidos conceitos preliminares necessários à devida compreensão do trabalho. No Capítulo 4 é definida a solução proposta, sendo estabelecidos a abordagem e o respectivo workflow. No Capítulo 5 é descrito o conjunto de recursos da configuração final do aplicativo desenvolvido com base na solução proposta. No Capítulo 6 são apresentados os três testes com usuários realizados, envolvendo especialistas e não especialistas, sendo analisados os respectivos resultados e, em particular, as principais respostas fornecidas pelos participantes e o modo como as mesmas foram obtidas. No Capítulo 7 são apresentadas as conclusões sobre este trabalho, bem como limitações observadas, principais resultados alcançados e possíveis continuações, em áreas bem definidas.

## 1.8 Considerações Finais

Este capítulo apresentou a motivação do trabalho, contextualizando-o. Além disso, foi caracterizado o problema investigado, expostos os objetivos principal e específicos, discutidas as etapas metodológicas envolvidas na execução do trabalho e brevemente abordadas as contribuições alcançadas. Foram também listadas as publicações geradas a partir deste trabalho.

Para uma maior compreensão de todos esses tópicos abordados, o próximo capítulo apresenta os principais trabalhos relacionados ao tema desta tese, discutindo a representação

gráfica de dados espaçotemporais e atributos e a identificação de padrões de movimento e de comportamento a partir de representações gráficas.

## 2 TRABALHOS RELACIONADOS

Este capítulo apresenta trabalhos relacionados diretamente ao tema da presente tese. A Seção 2.1 discute a representação visual de dados espaçotemporais e atributos, inicialmente abordando os primórdios e os avanços iniciais na área e, depois, dividindo-se em subseções específicas sobre: representações interativas e gráficas por clusters, de modelos origem-destino e tridimensionais. Depois, a Seção 2.2 aborda trabalhos específicos sobre a identificação de padrões de movimento e de comportamento com base em soluções gráficas, subdividindo-os em representações bidimensionais e tridimensionais. Por fim, a Seção 2.3 apresenta algumas considerações finais acerca deste Capítulo.

### 2.1 Representação Visual de Dados Espaçotemporais e Atributos

Há diversas técnicas para a representação visual de tempo, espaço e atributos (HARROWER, 2004; MACEACHREN, 2004; ANDRIENKO; ANDRIENKO, 2013). No entanto, tais técnicas são geralmente mais adequadas para enfatizar um único tipo de dado por vez ou para combinar os dados diretamente com representações gráficas espaciais estáticas, como mapas ou plantas arquitetônicas, por exemplo (AIGNER *et al.*, 2008; SIOW *et al.*, 2018). Sendo assim, ainda existem poucos trabalhos na literatura propondo soluções de análise visual para explorar especificamente os aspectos probabilísticos inerentes às PSTs, como ocorre particularmente em LEITE *et al.* (2017) e em LEITE *et al.* (2018).

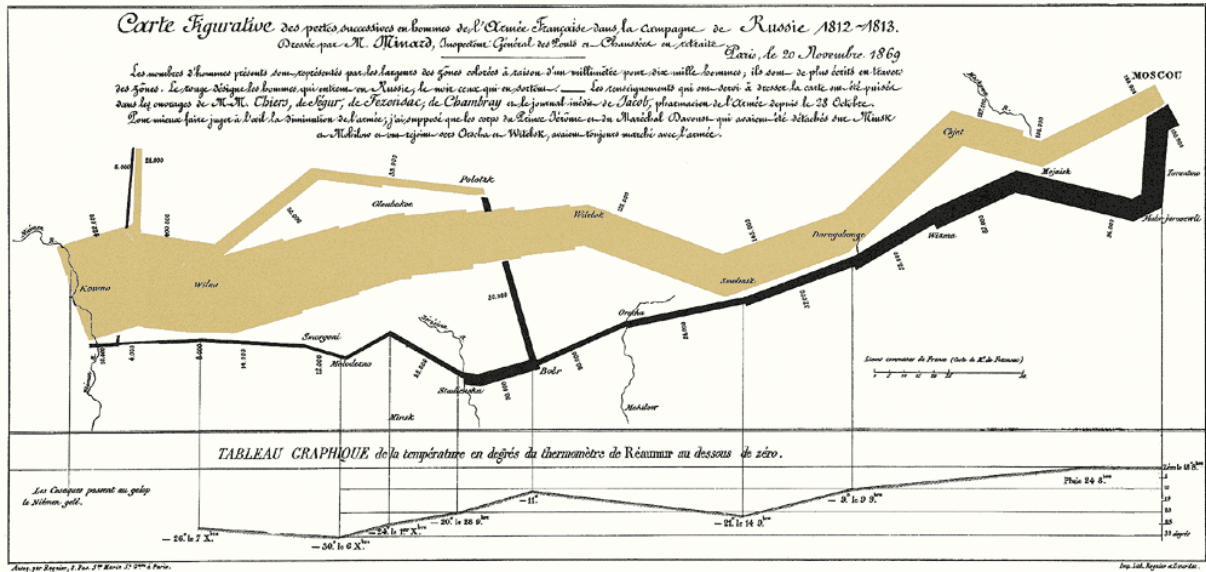
A seguir são apresentadas algumas soluções para representação gráfica de dados espaçotemporais e atributos, abordando os primórdios e avanços iniciais e, depois, dividindo-se em representações interativas e gráficas por clusters, de modelos origem-destino e tridimensionais.

#### 2.1.1 *Primórdios e Avanços Iniciais*

Um dos primeiros e mais famosos trabalhos a considerar especificamente a representação gráfica de relações entre espaço e tempo é a ilustração criada por Charles Joseph Minard em 1861 (Figura 5), representando a marcha do exército de Napoleão sobre Moscou, na campanha de 1812-1813 (GIANNOTTI; PEDRESCHI, 2008).

Em tal ilustração, começando na fronteira polaco-russa, a largura da faixa apresentada se refere ao tamanho do exército em cada posição. A penetração do exército no território inimigo é representada na cor clara, enquanto sua retirada, no frio inverno de Moscou, na cor escura.

Figura 5 – Ilustração representando a marcha do exército de Napoleão sobre Moscou, na campanha de 1812-1813, criada por Charles Joseph Minard em 1861.



Fonte: TUFTE (2001).

Assim, facilmente se percebem as enormes baixas de soldados sofridas.

Apesar de graficamente simples, a imagem consegue representar relações complexas, considerando seis diferentes atributos específicos (TUFTE, 2001):

1. o tamanho do exército, representado pela largura da faixa;
2. a latitude da localização do exército, indicada em pontos da faixa;
3. a longitude da localização do exército, indicada em pontos da faixa;
4. a direção da caminhada do exército, representada pela cor da faixa;
5. a localização do exército de acordo com determinadas datas, indicadas em pontos da faixa;
- e
6. a temperatura ao longo do caminho, indicada em pontos da faixa.

Posteriormente, com a evolução tecnológica, o papel foi sendo substituído por meios eletrônicos. Assim, nos Anos 60 e 70, a NASA utilizou um sistema computacional de visualização para o rastreamento de foguetes e outros objetos enviados ao espaço (SULLIVAN; BURBANK, 1976; LAUNIUS; MCCURDY, 2017). A Missão Apollo XI, por exemplo, empregou cinco mainframes IBM 360-75J com 1 MB de RAM, cada. Quatro desses computadores ficaram responsáveis por realizar todos os cálculos necessários ao controle da missão, ficando o quinto reservado a verificações e simulações.

No entanto, à época, os gastos envolvidos com a representação gráfica digital eram exorbitantes. Por exemplo, um terminal gráfico IBM 2250 Model 4 completo, com teclado e

*light pen* (Figura 6), da época custaria cerca de US\$ 1,8 milhão em 2019 (WIKIPEDIA, 2019).

Figura 6 – Terminal gráfico IBM 2250 Model 4 com teclado e *light pen*.



Fonte: WIKIPEDIA (2019).

Assim, apesar de já conseguir gerar imagens digitalmente, a NASA empregava um sistema de circuito fechado analógico de televisão para replicar dados visuais de telemetria. Utilizando 1.100 *racks* de controle, 140 consoles, 136 câmeras, 384 receptores de TV, 16.000 km de cabos e 20 canais fechados de TV, as imagens eram filmadas nas telas dos mainframe e distribuídas aos operadores (LAUNIUS; MCCURDY, 2017). Para poupar processamento, eram ainda utilizados artifícios simples, como lâminas de vidro com grades e linhas pintadas para definir marcações nos monitores utilizados, por exemplo.

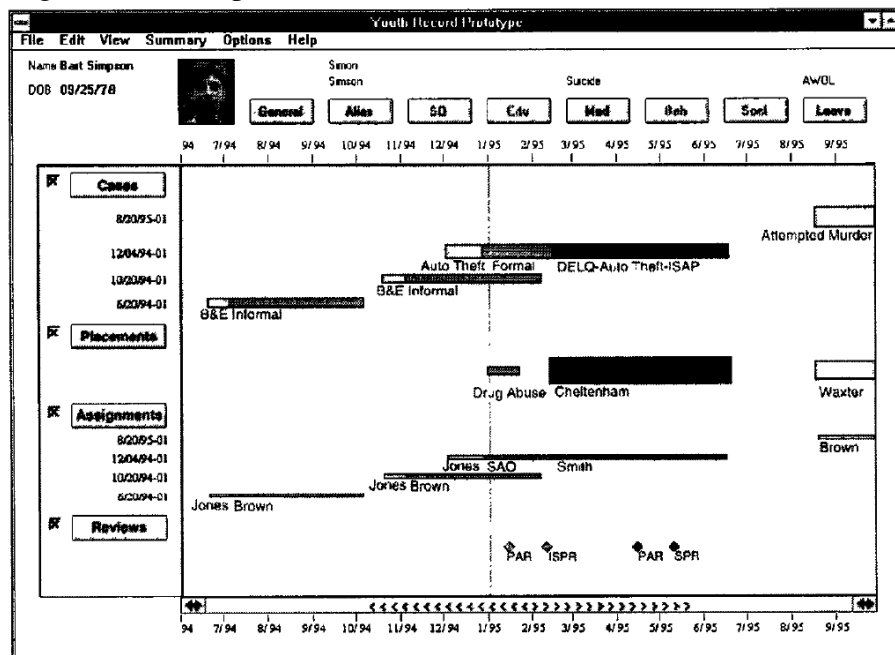
Bem posteriormente, dados os grandes avanços nos campos de hardware e de software, computadores passaram a ser utilizados cada vez mais para representar dados espaçotemporais e atributos, de forma geral. Assim, diversas soluções foram adotadas para representações gráficas, partindo do emprego de setas e linhas em mapas estáticos (VASILIEV, 1997) e de animações de elementos simples (MOELLERING, 1976) até a adoção de metáforas visuais para a representação de linhas do tempo (*timelines*<sup>1</sup>), por exemplo.

Particularmente, dentre essas metáforas, as *Lifelines* (PLAISANT *et al.*, 1995) merecem destaque por buscarem replicar, diretamente na tela do computador, o conceito de linha

<sup>1</sup> linhas temporais que representam a ocorrência de eventos em ordem cronológica (TUFTE *et al.*, 1990).

do tempo (Figura 7).

Figura 7 – Exemplo de Lifelines.



Fonte: PLAISANT *et al.* (1995).

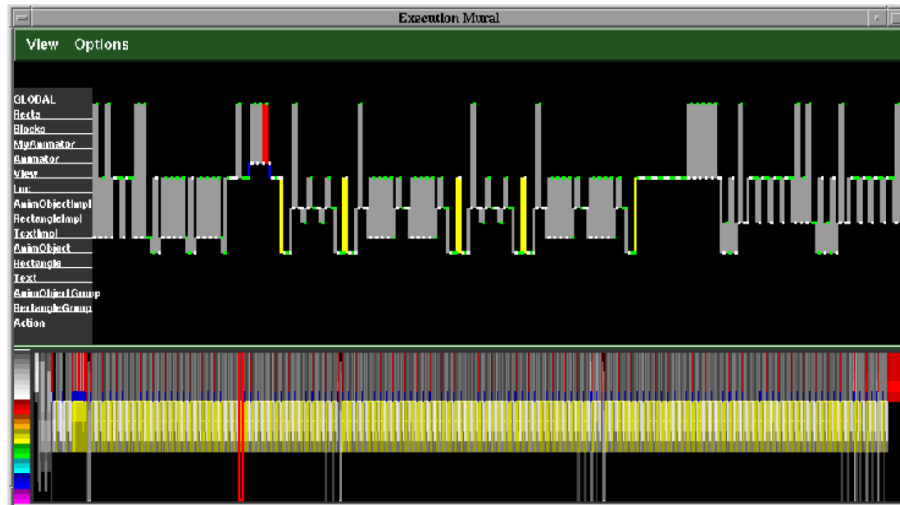
Depois, o emprego de imagens coloridas acabou se tornando um importante elemento na visualização de informações (KEIM *et al.*, 2008; BERTIN, 2011). Assim, em substituição ao preto e branco iniciais e à posterior escala de cinzas, o uso da cor e o maior controle de suas componentes, como matiz, saturação e brilho, puderam fornecer maiores destaques aos dados representados. Isso auxiliou deveras nos processos de percepção, de diferenciação e de compreensão de informações (WURMAN; WURMAN, 1990).

Então, por exemplo, a proposta dos *Mural Views* (JERDING; STASKO, 1998) buscou empregar cores para otimizar o uso de telas para a correta visualização de informações. Para tanto, foram empregados ajustes simultâneos em diversos atributos das imagens utilizadas, como sombreamentos em escala de cinzas, tamanhos de pontos apresentados, alterações de intensidades de cores individuais (ainda em número bastante limitado), além do emprego de *antialiasing*<sup>2</sup> (Figura 8). Tais manipulações de elementos na exibição de dados agilizaram as representações visuais, sobretudo diminuindo a influência do número então bastante limitado de pixels nas telas empregadas (MATKOVIC *et al.*, 2007).

No entanto, dados os resultados de testes utilizando diferentes layouts de mural views, percebeu-se que a simples adoção de uma única cor, trabalhando somente suas intensidades,

<sup>2</sup> técnica de suavização de efeito de serrilhado, em imagens digitais, considerando o ajuste de intensidade de cores (FOLEY *et al.*, 1996).

Figura 8 – Exemplo de mural view representando o funcionamento do algoritmo de ordenação *bubble sort*.



Fonte: JERDING; STASKO (1998).

não necessariamente se mostrava como um instrumento eficiente para a diferenciação dos dados observados (JERDING; STASKO, 1998). Assim, dependendo do padrão de gradação da cor utilizado, os resultados obtidos podiam ser, muitas vezes, inferiores àqueles obtidos com o uso de simples escalas de cinzas (PEDROSA, 1995).

Uma solução para tal problema foi, então, a adoção de escalas cromáticas<sup>3</sup>, como as geralmente adotadas nos mural views mais complexos (Figura 8), para uma melhor percepção da informação visual (LEVKOWITZ; HERMAN, 1992; WURMAN; WURMAN, 1990).

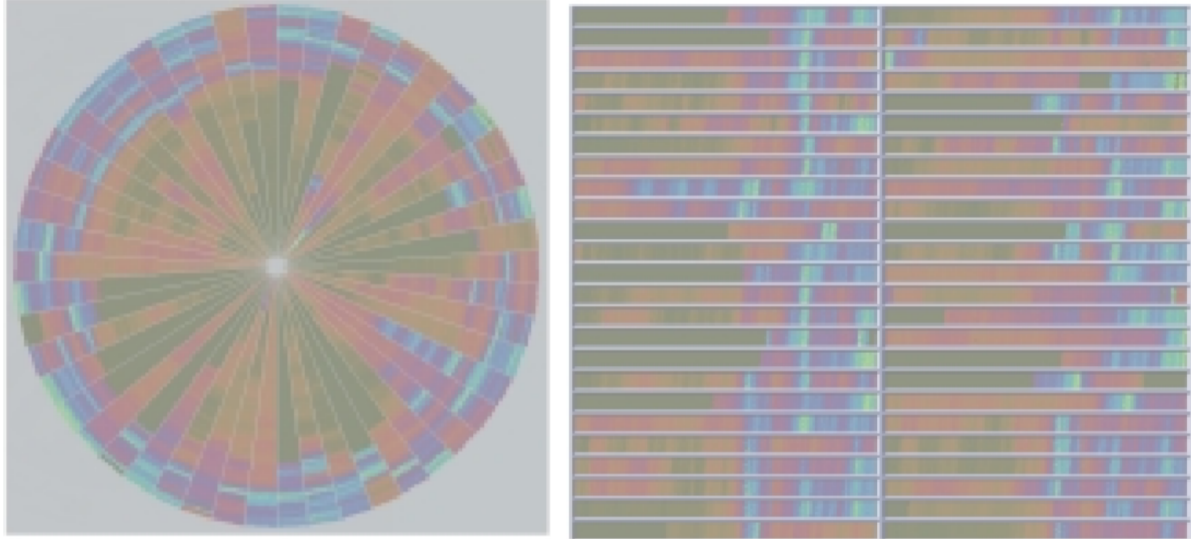
Como complemento, explorando especificamente o agrupamento de pixels coloridos, KEIM (2000) propôs o modelo de *Dense Pixel Displays*, representando vários dados simultaneamente, mapeando cada um deles em sequências de pixels, devidamente organizados em arranjos na tela, como círculos de pixels (Figura 9a) ou tabelas de pixels (Figura 9b), por exemplo.

Nos dense pixel displays, o valor de cada dimensão dos dados era visualizado como um conjunto de pixels coloridos, podendo os mesmos serem ordenados de acordo com as necessidades de análise: em ordem natural de ocorrência, em ordem crescente ou decrescente, por exemplo. Dessa forma, a observação da constância ou mudanças nos padrões de pixels formados podiam, então, fornecer informações importantes sobre o comportamento dos dados observados (Figura 9).

Pouco depois, a ideia anterior de linha do tempo foi ainda estendida, também

<sup>3</sup> sequenciamento de cores, que segue alguma regra preestabelecida, como mudança controlada de matiz ou aumentos/diminuições constantes de saturação ou de brilho, por exemplo (PEDROSA, 1995).

Figura 9 – Dois diferentes arranjos de Dense Pixels Displays, representando o acompanhamento de vinte anos, entre janeiro de 1974 a abril de 1995, de cinquenta empresas no FAZ-Index (índice de ações do mercado alemão): (a) visualização circular de segmentos e (b) tabela de pixels.



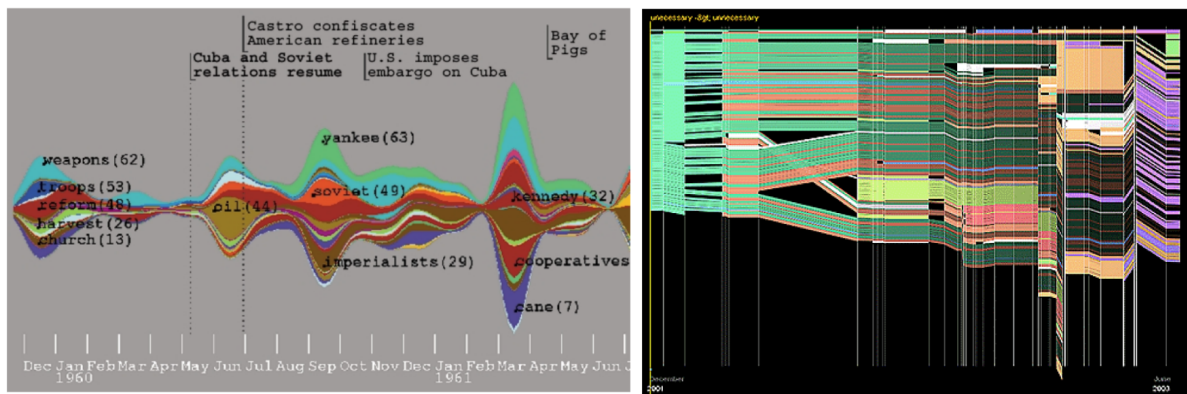
(a)

(b)

Fonte: KEIM (2000).

adotando cores para representar mais ricamente eventos e suas relações. Assim, modelos como os *ThemeRivers* (HAVRE *et al.*, 2002) utilizaram curvas coloridas, em vez de segmentos de reta, para exibir a intensidade e a importância de acontecimentos (Figura 10a); e os *History Flows* (VIÉGAS *et al.*, 2004) empregaram cores para enfatizar relações de causalidade (Figura 10a) (KEIM, 2000).

Figura 10 – Metáforas visuais baseadas em linhas do tempo coloridas: (a) ThemeRivers e (b) History Flows.



(a)

(b)

Fonte: (a) HAVRE *et al.* (2002) e (b) VIÉGAS *et al.* (2004).

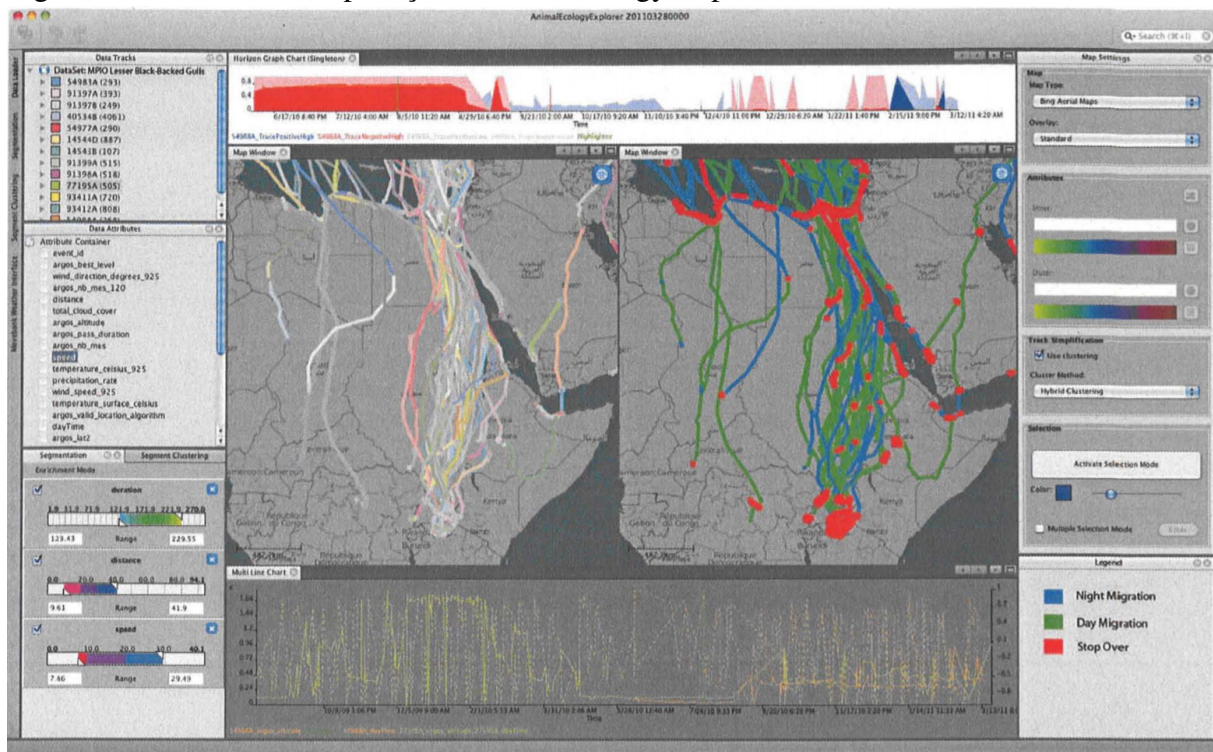


## 2.1.2 Representação Interativa

Já bem mais recentemente, dadas as conquistas técnicas e tecnológicas ocorridas, surgiram também soluções interativas para a representação de dados espaçotemporais e atributos. Tais soluções, além de proporem novas funcionalidades para a visualização de dados, muitas vezes combinam e estendem alguns dos recursos anteriormente apresentados, permitindo uma maior compreensão das informações representadas.

Assim, SPRETKE *et al.* (2011) usam segmentos de reta contíguos e coloridos para mostrar diferentes trajetórias simultaneamente em mapas, possibilitando avaliar individualmente cada caminho percorrido. Para complementar as informações fornecidas, são empregados gráficos para o acompanhamento de atributos específicos, além da integração de informações externas, como dados climáticos ou mapas digitais obtidas diretamente de serviços online.

Figura 11 – Interface da aplicação Animal Ecology Explorer.



Fonte: SPRETKE *et al.* (2011).

A Figura 11 apresenta a aplicação *Animal Ecology Explorer*, que faz uso dessa técnica para avaliar a migração de gaivotas. No lado esquerdo da tela, pode-se selecionar individualmente cada animal rastreado e os respectivos atributos a serem analisados (temperatura ambiente, velocidade da ave, sua altitude etc). Ao centro, duas visualizações paralelas do mesmo

mapa exibem as trajetórias de cada animal selecionado (mapa à esquerda) e a clusterização<sup>4</sup> das mesmas (à direita). Nesse último mapa, é particularmente interessante observar a diferenciação entre voos diurnos e noturnos, marcados respectivamente em azul e em verde; e os pontos de parada, marcados em vermelho, buscando distinguir visualmente os percursos utilizados (TOMINSKI *et al.*, 2012). De forma complementar, gráficos simples auxiliares correlacionam os atributos selecionados, considerando o tempo (gráfico de superfície na porção superior) ou sendo acumulados de acordo com as respectivas distâncias percorridas (gráfico de linha na porção inferior).

No entanto, embora esse sistema consiga mostrar diversas informações de trajetórias, nota-se que a poluição e a oclusões visuais não são inteiramente evitadas, o que pode tornar a análise muitas vezes incômoda (TOMINSKI *et al.*, 2012; CRNOVRSANIN *et al.*, 2009).

Buscando minimizar tal problema, WANG *et al.* (2013) vão além de associarem mapas a gráficos estatísticos. Assim, os autores fazem uso de histogramas interativos, para a filtragem de informações temporais, e tabelas de pixels, para o acompanhamento de ocorrências.

Essa combinação de elementos visuais e interativos é experimentada em estudos envolvendo a avaliação do tráfego de veículos. Assim, por exemplo, a Figura 12 apresenta uma ferramenta para análise de tráfego onde o usuário observa velocidades de veículos numa via representada num mapa (Figura 12a), podendo determinar um período específico para observação, e definir datas e horários para filtragem através da seleção direta nos histogramas interativos (Figura 12b). A distribuição de velocidades médias da via, no período determinado, é representada através de uma tabela de pixels (Figura 12c), seguindo a mesma escala cromática estabelecida no mapa.

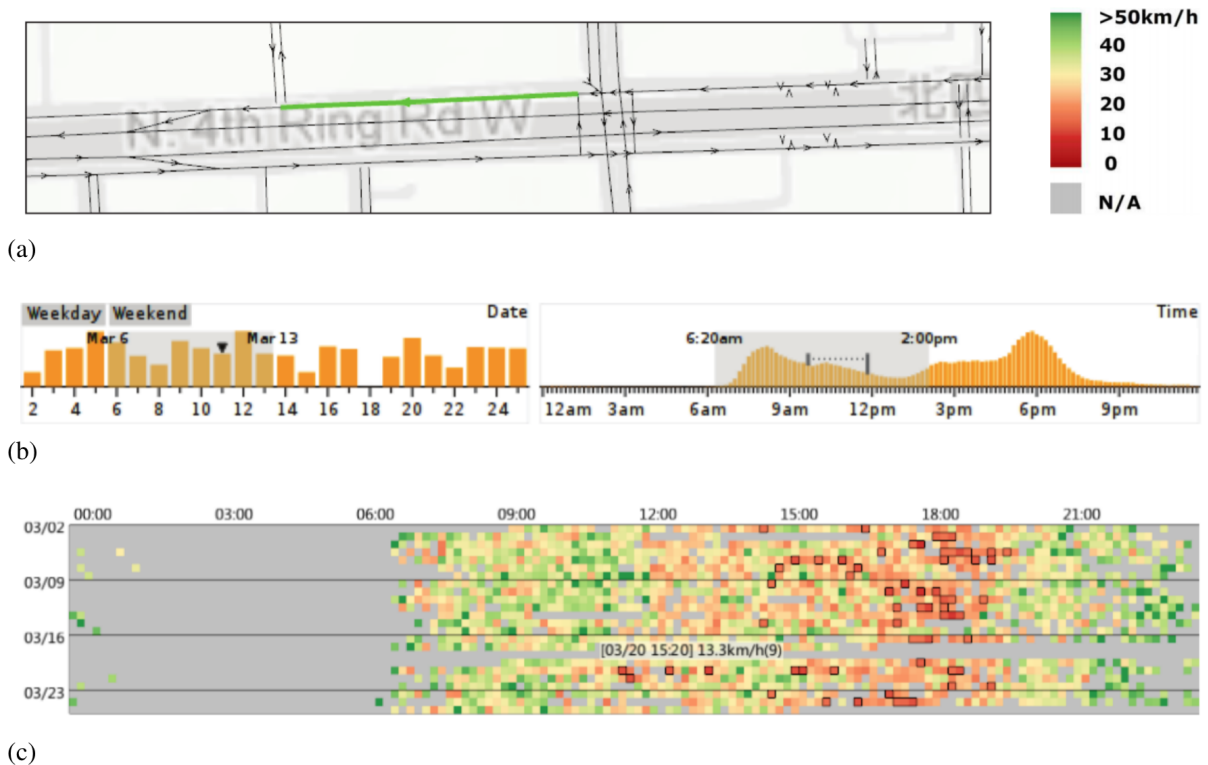
Também buscando minimizar a sobreposição de elementos visuais, IVANOV *et al.* (2007) propõem uma abordagem baseada na divisão dos dados espaçotemporais em diversas janelas de visualização sincronizadas, dentre as quais uma que exhibe especificamente o espaço e outra, o tempo, na forma de uma linha do tempo (Figura 13a).

A Figura 13 apresenta um sistema de vigilância de recintos, desenvolvido com base em IVANOV *et al.* (2007). Dados de câmeras e de sensores de movimento são armazenados e relacionados sincronamente, podendo ser operados diretamente por usuários para o acompanhamento de situações e a verificação de ocorrências passadas (Figura 13b).

---

<sup>4</sup> técnica de agrupamento de instâncias de elementos com determinadas características similares (LEE *et al.*, 2008; LEI *et al.*, 2013).

Figura 12 – Diferentes recursos para a avaliação de ocorrências no tráfego de veículos: (a) observação de velocidades de veículos numa via; (b) seleção de períodos temporais empregando histogramas interativos; e (c) tabela de pixels representando a distribuição de velocidades médias da via no período selecionado.



Fonte: WANG *et al.* (2013).

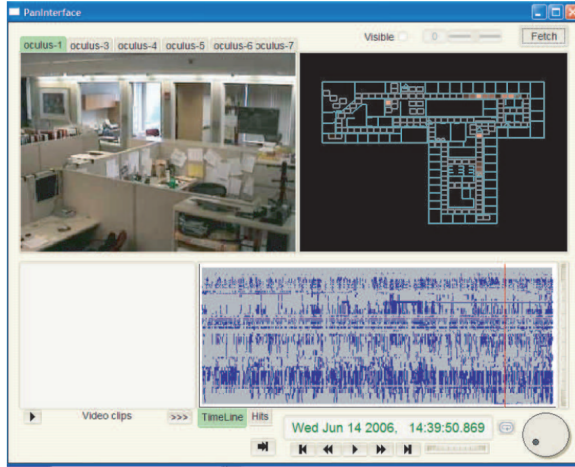
### 2.1.3 Representação por Clusters

Problemas referentes ao excesso de elementos ou à sobreposição dos mesmos, causando respectivamente poluição e oclusão visuais, são notórios em casos onde trajetórias de múltiplas entidades precisam ser exibidas simultaneamente (ZENG *et al.*, 2013). Sendo assim, muitas vezes é comum empregar técnicas de clusterização, substituindo as várias trajetórias originais por *clusters* individuais, que passam a representar as subtrajetórias mais representativas do conjunto inicial. Dessa forma, por exemplo, LEE *et al.* (2007) propõem um *framework* específico, baseado em estratégias de particionamento espacial e no agrupamento visual de elementos, para extrair automaticamente os clusters de diferentes datasets de trajetórias (Figura 14).

Já ANDRIENKO *et al.* (2009) combinam técnicas de clusterização baseada em densidade que permitem refinar o resultado visual, através da intervenção direta do usuário, para tentar estabelecer os respectivos clusters mais representativos (Figura 15).

A título de esclarecimento, há diversos trabalhos que propõem soluções completamente automatizadas, e que podem apresentar resultados bastante semelhantes ao anterior, seja

Figura 13 – Sistema desenvolvido por IVANOV *et al.* (2007) para vigilância de recintos. Em (a) percebe-se as janelas de visualização sincronizadas, exibindo ocorrências no espaço de acordo com uma linha do tempo. Em (b) é apresentado um dos modos de interação, com base no uso de uma *tabletops*.



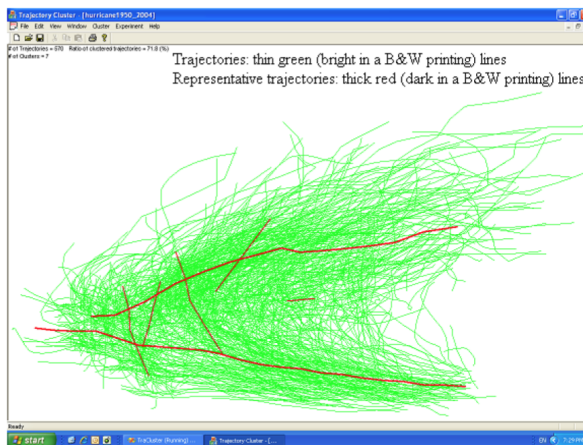
(a)



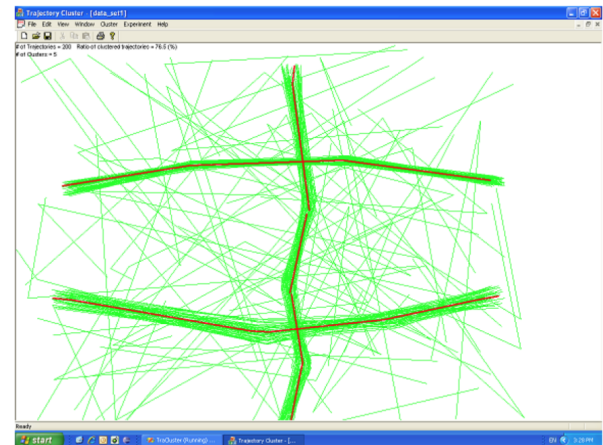
(b)

Fonte: IVANOV *et al.* (2007).

Figura 14 – Sub-trajetórias extraídas de diferentes datasets empregando o framework de LEE *et al.* (2007).



(a)



(b)

Fonte: LEE *et al.* (2007).

empregando ruído Gaussiano (GAFFNEY; SMYTH, 1999; CHUDOVA *et al.*, 2003), cadeias de Markov (ALON *et al.*, 2003), baseados na medida de subsegmentos de trajetórias (HWANG *et al.*, 2005), funções de distância entre elementos (RINZIVILLO *et al.*, 2008) ou mesmo acompanhando a evolução de clusters (SILVA *et al.*, 2016), dentre muitas outras possibilidades. No entanto, pode-se destacar o trabalho de PANAGIOTAKIS *et al.* (2011), que opera exatamente no mesmo campo de atuação da solução de ANDRIENKO *et al.* (2009). De todo modo, porém, dada a possível geração de resultados automáticos incoerentes com as análises realizadas, nota-se que a opção pela intervenção humana ainda é bastante cabível.

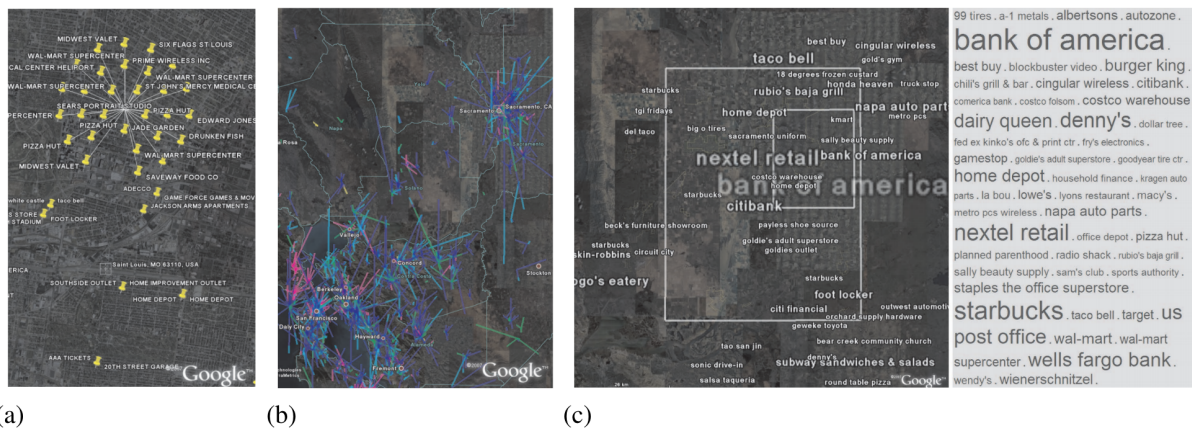
Figura 15 – Um cluster A sendo dividido em dois outros clusters, B e C, interativamente através da técnica de ANDRIENKO *et al.* (2009).



Fonte: ANDRIENKO *et al.* (2009).

Já nos casos onde os dados espacotemporais não constituem especificamente trajetórias, podendo os mesmos serem compreendidos individualmente, WOOD *et al.* (2007) definem um modelo interativo alternativo para codificação gráfica de mapas em resposta a consultas de dados. Para tanto, podem ser empregados desde simples marcadores (Figura 16a) e segmentos de reta coloridos, com cores e espessuras representando atributos dos dados (Figura 16b), até *tag clouds*<sup>5</sup> associadas a *tag maps*<sup>6</sup> (Figura 16c).

Figura 16 – Exemplos de codificações gráficas em mapas: (a) marcadores; (b) segmentos de reta coloridos e (c) tag map associado a tag cloud.



Fonte: WOOD *et al.* (2007).

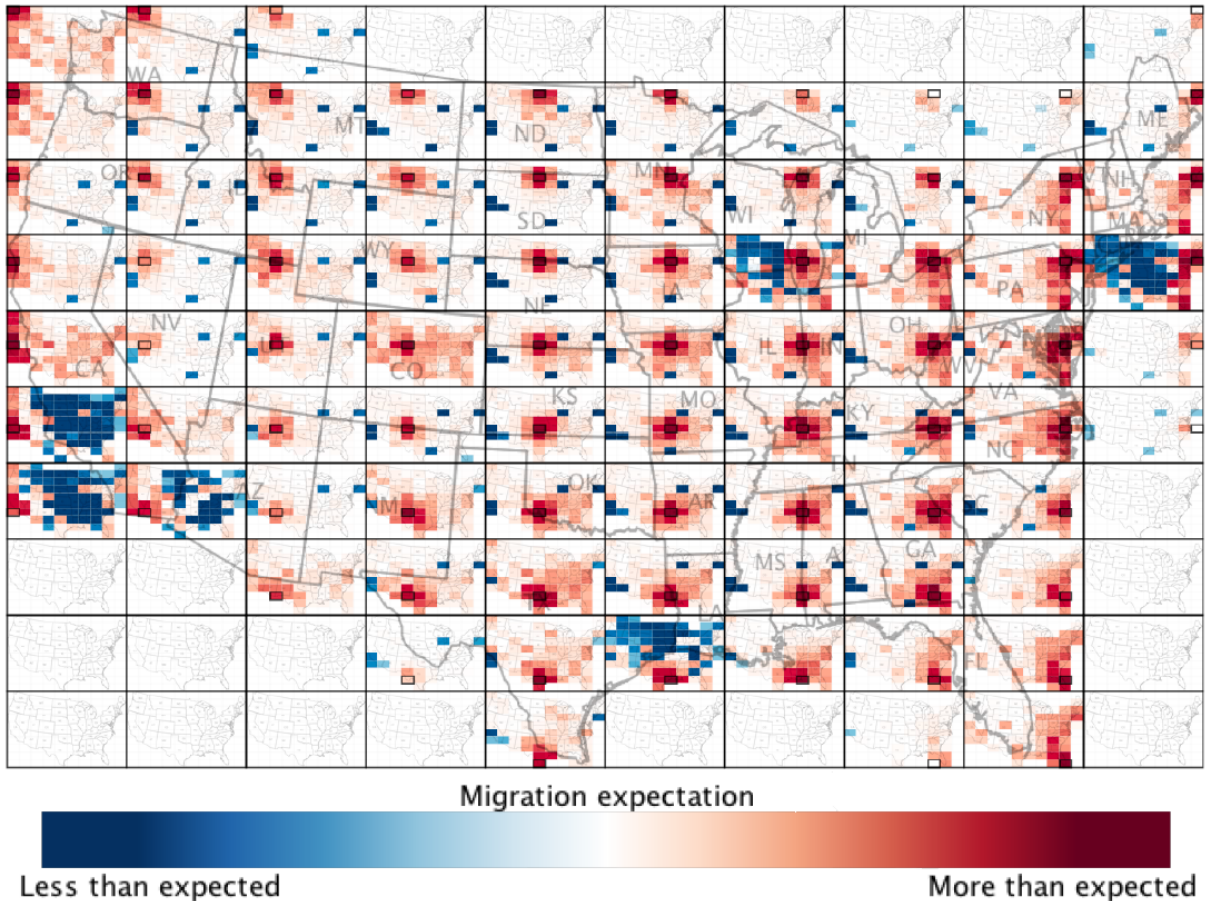
<sup>5</sup> técnica que representa a frequência de uma palavra, em um dado conjunto, alterando o respectivo tamanho de sua representação gráfica (HASSAN-MONTERO; HERRERO-SOLANA, 2006).

<sup>6</sup> versões espaciais de tag clouds em que palavras são dispostas em um mapa, e cujos tamanhos encontram-se relacionados diretamente à frequência na respectiva posição do mapa, de seguindo a escala do mesmo (SLINGSBY *et al.*, 2007).



2000.

Figura 18 – OD map demonstrando expectativas migratórias nos Estados Unidos da América, de acordo com o censo norte-americano de 2000.

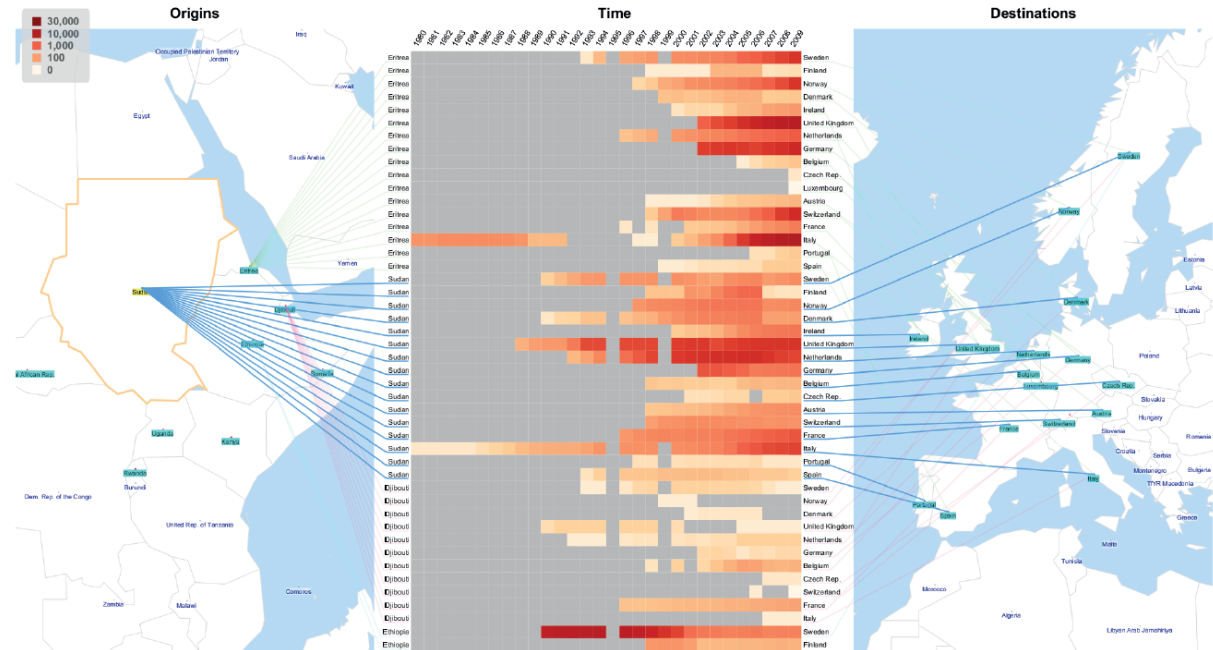


Fonte: WOOD *et al.* (2010).

De forma complementar às duas técnicas anteriores, que somente consideram o espaço, BOYANDIN *et al.* (2011) vão além, estabelecendo os *Flowstrates*, que inter-relacionam locais (espaço) a períodos (tempo) e também empregam tabelas de pixels, para avaliar atributos específicos, como intensidade de ocorrências, velocidades instantâneas etc. Assim, por exemplo, a Figura 19 apresenta fluxos de refugiados entre a África Oriental (à esquerda) e a Europa Ocidental (à direita), com base em dados de TRENDS (2009). Apesar da complexidade inicial, por exemplo, pode-se perceber facilmente que as migrações oriundas do Sudão, país destacado na imagem, são sempre consistentemente altas para o Reino Unido e a Holanda. Isso fica bastante patente ao se consultar a respectiva tabela de pixels, ao centro, relacionando-a à escala cromática no canto esquerdo superior da imagem.

Por último, é importante destacar também o trabalho de FERREIRA *et al.* (2013), que lida com grandes datasets. Ferreira e seus coautores propõem um sistema de consulta

Figura 19 – Flowstrate representando os fluxos migratórios entre a África Oriental (à esquerda) e a Europa Ocidental (à direita) em 2009.



Fonte: BOYANDIN *et al.* (2011).

interativa para dados referentes a origens e destinos de viagens de táxis e limusines na cidade de Nova York, considerando os anos de 2009, 2011 e 2012. A solução proposta oferece suporte à visualização de resultados diretamente em mapas geográficos, podendo utilizar-se ainda de outros recursos visuais, como mapas de calor (CARD, 1999; WILKINSON; FRIENDLY, 2009) ou gráficos estatísticos, por exemplo (Figura 20). Para tanto, os dados são pré-processados, a fim de viabilizar as etapas de análise. Pois, de outro modo, as consultas propostas seriam praticamente impossíveis, já que os dados originais somam cerca de 120 GB, contendo registros de aproximadamente 540 milhões de viagens.

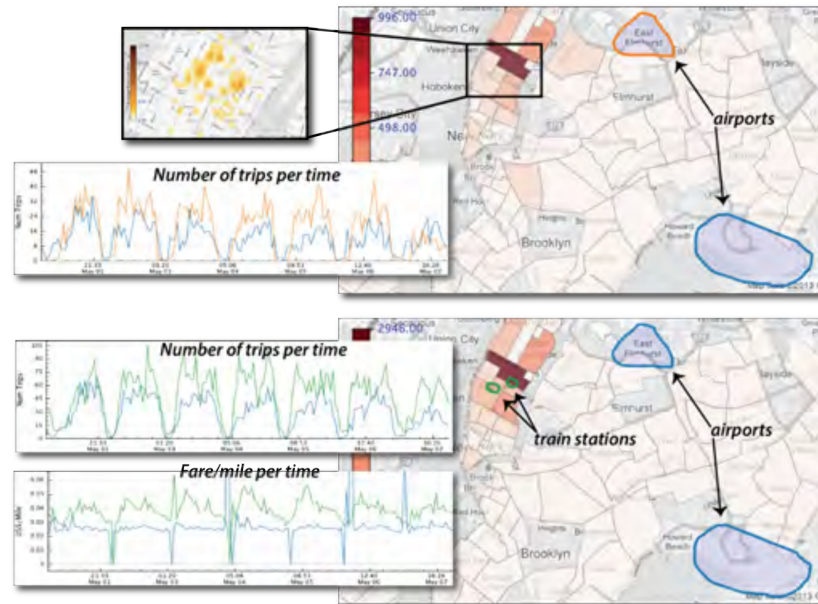
### 2.1.5 Representações Tridimensionais

Com relação especificamente à representação tridimensional de dados espacotemporais e demais atributos, pode-se inicialmente destacar o trabalho de KWAN (2000) na geração de modelos que integram múltiplos dados referentes à realização de atividades, como intensidade, localização, temporização, duração ou sequenciamento, por exemplo.

Assim, por exemplo, a Figura 21 mostra a superfície de um mapa distorcida tridimensionalmente, cujo eixo de alturas tenta refletir o nível de atividade referente aos respectivos locais. É importante salientar que, neste exemplo, o próprio mapa bidimensional, então utilizado como



Figura 20 – Exemplo de uso do sistema proposto por FERREIRA *et al.* (2013), consultando e comparando viagens iniciadas em aeroportos e acabando em estações de trem na cidade de Nova York.



Fonte: FERREIRA *et al.* (2013).

uma textura no modelo tridimensional, foi gerado a partir da datasets individuais, referentes a dados rodoviários, fluviais e urbanos da cidade de Portland, nos Estados Unidos da América.

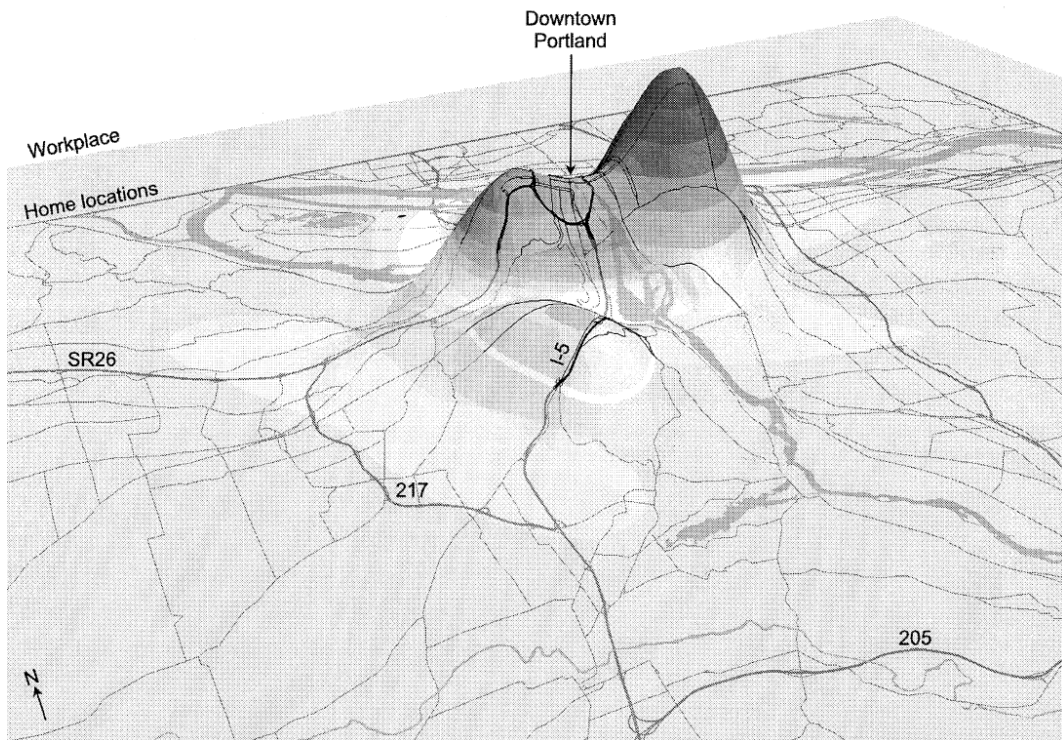
No entanto, apesar do bom resultado alcançado nesse trabalho, nota-se que as representações tridimensionais tem sido adotadas mais como soluções alternativas, principalmente visando minimizar o problema de oclusão visual na exibição de diferentes dados em mapas planos (SIOW *et al.*, 2018).

Assim, os esforços de KAPLER; WRIGHT (2005) focam especificamente na representação de dados espaçotemporais, enfatizando individualmente as informações referentes à dimensão espacial ou temporal (Figura 22).

Dessa forma, por exemplo, a Figura 22a enfatiza o espaço, exibindo os principais locais visitados por um dado indivíduo, sendo possível destacar interativamente os mesmos para visualização no canto direito inferior da tela. Já a Figura 22b enfatiza o tempo, e os mesmos dados espaçotemporais, porém agora considerando os eventos ocorridos, podem ser visualizados com base na linha do tempo disponível na parte superior da tela.

De maneira parecida, EAGLE; PENTLAND (2006) apresentam uma solução, baseada na navegação através do mouse e na interação com *widgets* de uma interface gráfica, que adota o espaço tridimensional como uma simples extensão do bidimensional, tentando

Figura 21 – Exemplo de integração de dados através de modelos tridimensionais estabelecida em KWAN (2000).



Fonte: KWAN (2000).

demonstrar novas possibilidades na visualização de dados espacotemporais, no caso, ligações entre telefones celulares (Figura 23).

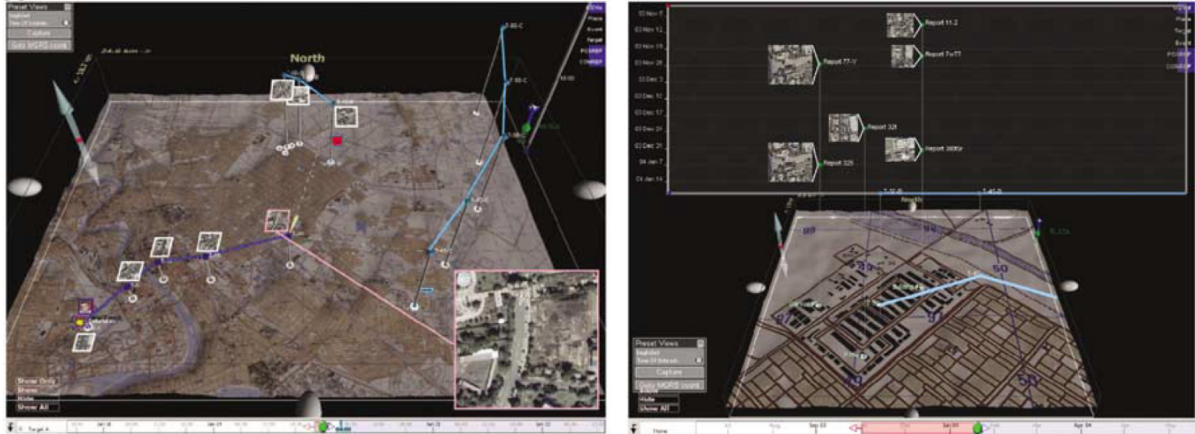
Apesar do grande impacto visual, particularmente à época da publicação do trabalho, nota-se que a simples ideia de estender o plano para o espaço não permite ganhos substanciais na visualização das informações, sendo muitas vezes necessário modificar constantemente o ângulo de visão para compreender corretamente o conjunto dos dados expostos (CRNOVRSANIN *et al.*, 2009; SIOW *et al.*, 2018).

Já fortemente fundamentado na ideia de geografia temporal<sup>8</sup>, KRAAK (2003) apresenta o conceito de cubo espacotemporal. Nesse modelo, o espaço é definido por um mapa geográfico presente na face inferior interna de um cubo imaginário, sendo o tempo representado no sentido ascendente, perpendicular ao plano inferior desse cubo (Figura 24).

Assim, por exemplo, o cubo espacotemporal mostrado na Figura 24 representa os deslocamentos de um determinado indivíduo durante um dia inteiro. Percebe-se que o mesmo permanece em casa (*home*) até aproximadamente 5:10, então indo para uma piscina (*pool*) distante e retornando para casa um pouco antes das 7:55. Depois, ele vai ao trabalho (*work*) por

<sup>8</sup> perspectiva transdisciplinar que considera a relação entre as dimensões espacial e temporal para a análise de processos dinâmicos, como interações entre pessoas e o ambiente em que vivem (HÄGERSTRAND, 1970).

Figura 22 – Apresentação de dados espaçotemporais, de acordo com o modelo de KAPLER; WRIGHT (2005), enfatizando: (a) espaço, representando os principais locais visitados por um indivíduo, e (b) tempo, representando os eventos de uma linha do tempo.



(a)

(b)

Fonte: KAPLER; WRIGHT (2005).

Figura 23 – Exemplo de aplicação, apresentada em EAGLE; PENTLAND (2006), que exibe dados espaçotemporais referentes a ligações entre telefones celulares.

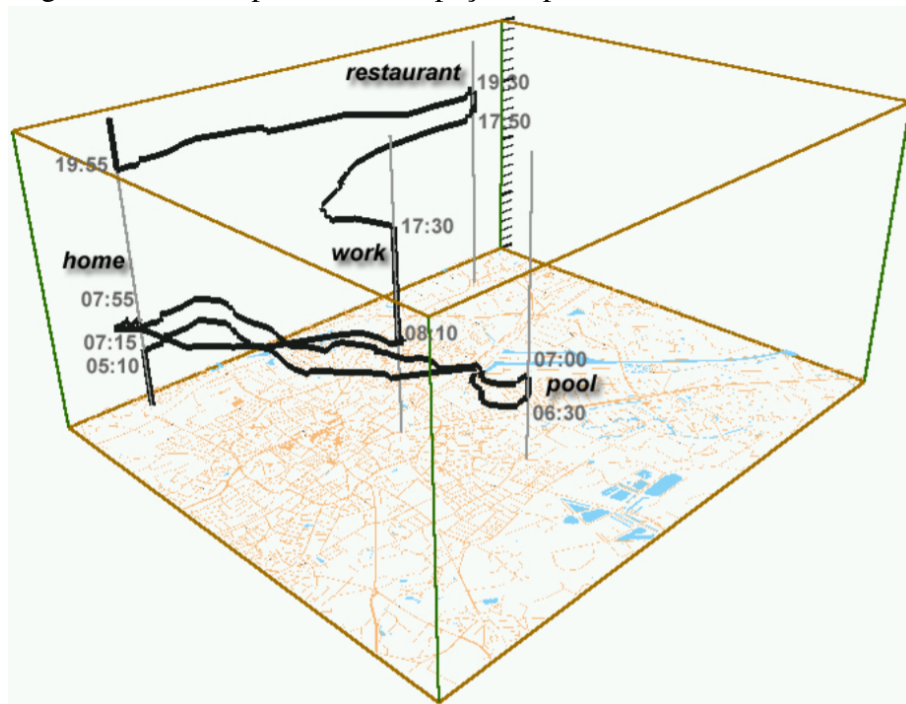


Fonte: EAGLE; PENTLAND (2006).

volta das 8:10, ali ficando até as 17:30. Entre 17:50 e 19:30 ele vai a um restaurante (*restaurant*), então voltando para casa aproximadamente às 19:55.

Considerando essa mesma ideia inicial, os muros de trajetórias (*trajectory walls*) de TOMINSKI *et al.* (2012) expandem o conceito de cubo espaçotemporal, porém apresentando várias trajetórias individuais simultaneamente. Para tanto, essas trajetórias são justapostas

Figura 24 – Exemplo de cubo espaçotemporal.



Fonte: KRAAK (2003).

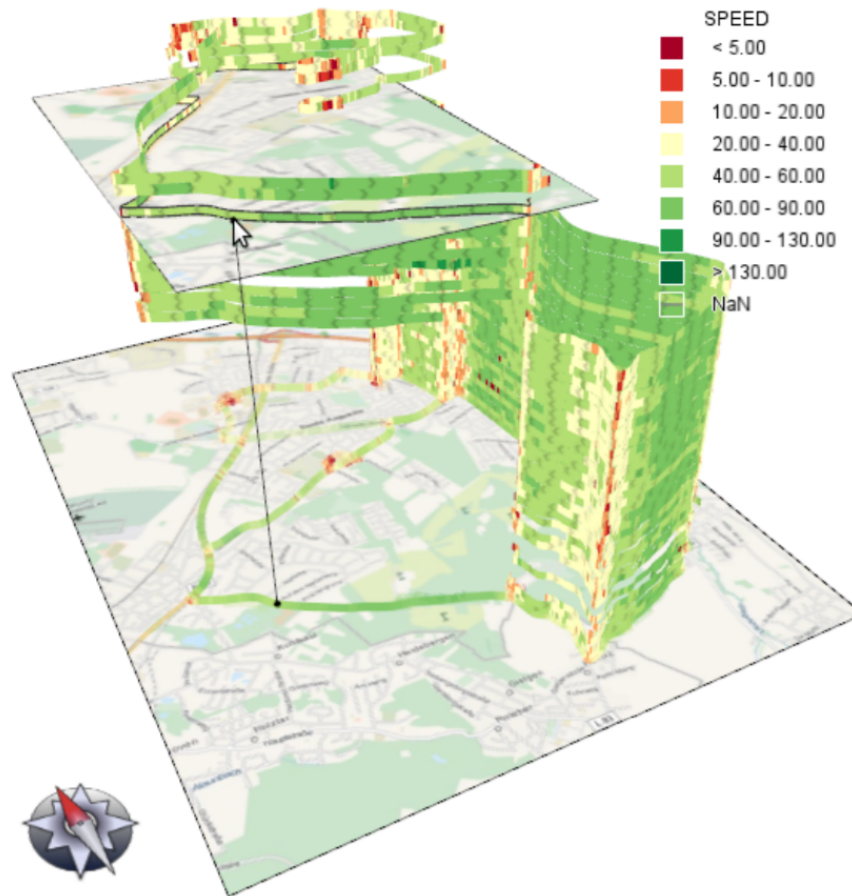
verticalmente sobre um mapa plano presente num espaço tridimensional interativo (Figura 25).

A Figura 25 exhibe um muro de trajetórias que representa os deslocamentos de diversos veículos em um mesmo percurso, destacando suas respectivas velocidades instantâneas através de uma escala cromática. Para avaliar devidamente os dados representados, pode-se selecionar uma trajetória específica, então observando sua projeção direta sobre o mapa. Dada a grande quantidade de elementos gráficos apresentados, também são disponibilizadas ferramentas para reposicionamento, rotação e *zoom* do espaço tridimensional utilizado, permitindo uma melhor observação dos elementos gráficos e dos respectivos dados representados (ANDRIENKO *et al.*, 2014).

Apesar da relativa eficiência das soluções tridimensionais, é bastante comum ainda o emprego paralelo de recursos complementares, como gráficos estatísticos ou tabelas de pixels, por exemplo, para uma melhor avaliação das informações representadas (DYKES; MOUNTAIN, 2003; KRAAK, 2003; PATTERSON *et al.*, 2004; ZENG *et al.*, 2012). Isso decorre principalmente da complexidade de alguns datasets, onde se torna praticamente impossível evitar poluição ou oclusão visuais (ANDRIENKO *et al.*, 2007).

Além disso, por geralmente exigirem maiores esforços computacionais e apresentarem uma maior complexidade no uso, nem sempre essas soluções baseadas em modelos tridimensionais se mostram adequadas à representação de determinados tipos de dados, par-

Figura 25 – Exemplo de muro de trajetórias.



Fonte: TOMINSKI *et al.* (2012).

ticularmente aqueles cuja manipulação exige um tempo de resposta mínimo para a correta avaliação (SIOW *et al.*, 2018).

## 2.2 Identificação de Padrões de Movimento e de Comportamento a partir de Representações Gráficas

A seguir são apresentados trabalhos específicos sobre a identificação de padrões de movimento e de comportamento a partir de representações gráficas, encontrando-se os mesmos subdivididos em representações bidimensionais e tridimensionais.

### 2.2.1 Representações Bidimensionais

Com relação especificamente à análise de padrões de mobilidade, inicialmente pode-se destacar os trabalhos que realizam a análise diretamente sobre os dados disponíveis, sem o emprego de representações gráficas muito elaboradas ou interativas.

Assim, por exemplo, BALACHANDRAN *et al.* (2002) e BALAZINSKA; CASTRO (2003) empregaram modelos lineares e probabilísticos a fim de avaliar os *logs* de redes locais para caracterizar comportamentos de usuários. Os resultados foram interpretados a partir principalmente de gráficos estatísticos simples, de linhas e de colunas, e de algumas tabelas. Apesar disso, porém, foi possível identificar padrões de atividade diárias de diversas pessoas, inclusive inferindo algumas relações de trabalho entre as mesmas.

Já considerando um montante de dados bem maior, GONZALEZ *et al.* (2008) analisaram as trajetórias anonimizadas de usuários de cem mil telefones celulares, rastreados durante um período de seis meses. Esse estudo, em particular, identificou um alto grau de regularidade temporal e espacial para boa parte dos indivíduos representados, indicando grande probabilidade de retorno dos mesmos a alguns locais específicos. Tal resultado contrastou fortemente com os modelos até então esperados, como voos de Levy<sup>9</sup> ou caminhadas randômicas (BROCKMANN *et al.*, 2006). Para demonstrar os padrões encontrados, Gonzalez e seus coautores empregaram gráficos de pontos e de linhas paralelamente ao uso de mapas.

Também avaliando dados oriundos de telefones celulares, porém agora oriundos de vários datasets diferentes, CHAINTREAU *et al.* (2007) e KARAGIANNIS *et al.* (2010) detectaram padrões simples de comportamentos de indivíduos e de grupos, diferenciando, por exemplo, caminhadas randômicas (BROCKMANN *et al.*, 2006) daquelas com intenções bem definidas. Inclusive, aproveitando-se de tal diferenciação, esses autores propuseram, e posteriormente demonstraram, que o tempo médio de contato entre dois indivíduos podia geralmente ser aproximado por uma lei de potência<sup>10</sup>.

SCCELLATO *et al.* (2011) usaram mapas de frequência de dados GPS para identificar pontos de interesse (POIs, *Points of Interest*) em mapas geográficos, focando na previsão de localização com base na análise de séries temporais não-lineares, considerando tanto horários de chegada quanto tempos de permanência de indivíduos. Assim, foi possível identificar uma grande quantidade de lugares possivelmente significativos para esses indivíduos.

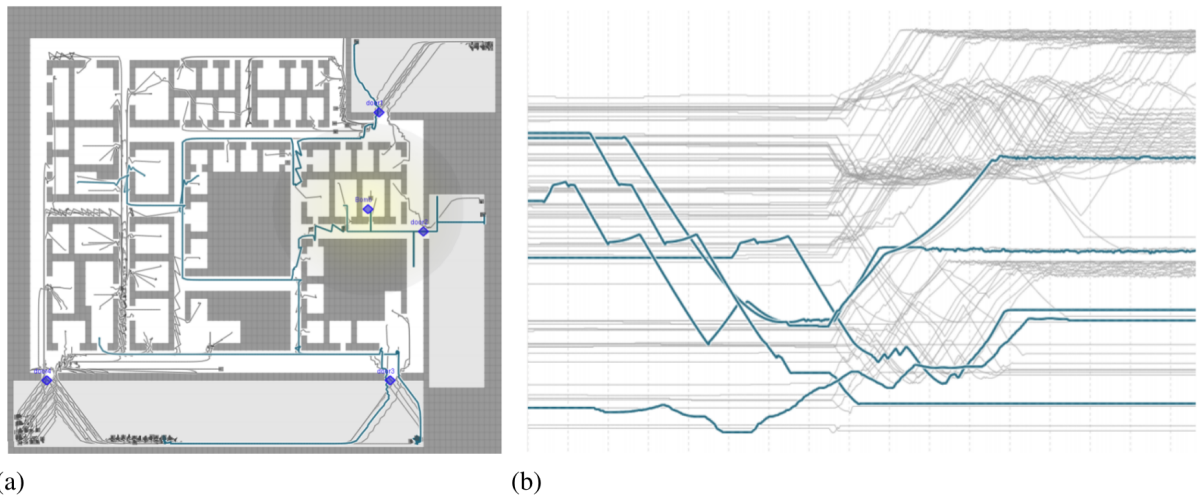
Já no que se refere especificamente à identificação de padrões com base em representações gráficas mais elaboradas, CRNOVRSANIN *et al.* (2009) estabeleceram uma representação bidimensional, que associava uma planta baixa a um gráfico específico, que relacionava a distância a um local pré-selecionado versus o tempo decorrido de um determinado

<sup>9</sup> padrão de movimento, estabelecido pelo matemático Paul Pierre Lévy, que se caracteriza pelo estabelecimento de diferentes trajetórias curtas e longas, mas predominantemente curtas (BROCKMANN *et al.*, 2006).

<sup>10</sup> em Física, uma lei de potência é aquela em que sua relação pode ser descrita matematicamente pela equação  $y = ax^k$ , onde  $a$  (constante de proporcionalidade) e  $k$  (expoente) são constantes (GUERRIERO, 2012).

acontecimento (Figura 26). Para evitar sobreposição visual de informações, os autores empregaram PCA (*Principal Component Analysis*) (PEARSON, 1901) com relação à distância representada, tentando observar diversas trajetória simultaneamente.

Figura 26 – Exemplo de uso da técnica de visualização de CRNOVRSANIN *et al.* (2009): (a) representação gráfica dos deslocamentos de indivíduos numa simulação de evacuação de um prédio devido a uma explosão e (b) gráfico relacionando a distância de cada indivíduo considerado em relação à explosão (eixo horizontal) versus o tempo decorrido (eixo vertical).



Fonte: CRNOVRSANIN *et al.* (2009).

Assim, empregando essa técnica, a Figura 26a apresenta a simulação de uma evacuação em um prédio após uma explosão em uma das salas. Percebe-se que, como o tempo não se encontra devidamente representado na planta baixa empregada, não é possível identificar adequadamente quando cada ocorrência aconteceu (saídas de indivíduos de cada sala, idas aos corredores etc). Já a Figura 26b apresenta o respectivo gráfico que relaciona a distância de cada indivíduo presente na cena em relação à posição onde ocorreu a explosão (eixo vertical) versus o tempo decorrido (eixo horizontal). Dessa forma, pode-se então notar, por exemplo, o movimento de algumas pessoas logo antes da explosão, o que pode sugerir possíveis suspeitos ou testemunhas do ocorrido (CRNOVRSANIN *et al.*, 2009).

Já trabalhando com datasets reais, EAGLE; PENTLAND (2006) avaliaram dados de interações entre identificadores *bluetooth* de cem telefones celulares, coletados durante nove meses, sendo possível modelar ritmos temporais e comportamentos específicos no grupo representado.

Os resultados obtidos, além da geração de um modelo tridimensional para apresentação de dados espacotemporais (Figura 23), já discutido na Seção 2.1.5 deste trabalho, também

Figura 27 – Grafos de visualização de relações de amizade (a) e de proximidade física (b), identificadas entre indivíduos no trabalho realizado por EAGLE; PENTLAND.



(a)

(b)

Fonte: EAGLE; PENTLAND (2006).

incluíram grafos que auxiliaram na identificação de diferentes ligações entre vários indivíduos. Tais grafos identificaram possíveis relações de amizade (Figura 27a) e de proximidade física no espaço considerado (Figura 27b).

E já considerando a investigação de comportamentos coletivos em grande escala, CANDIA *et al.* (2008) analisaram a ocorrência de eventos anormais em ligações entre telefones celulares de cerca de seis milhões de usuários durante um mês. Para tanto, Candia e seus coautores adotaram variadas técnicas de visualização, como diagramas de Voronoi (LEE, 1982) associados a gráficos estatísticos (Figura 28a) e comparações entre tabelas de pixels (Figura 28b).

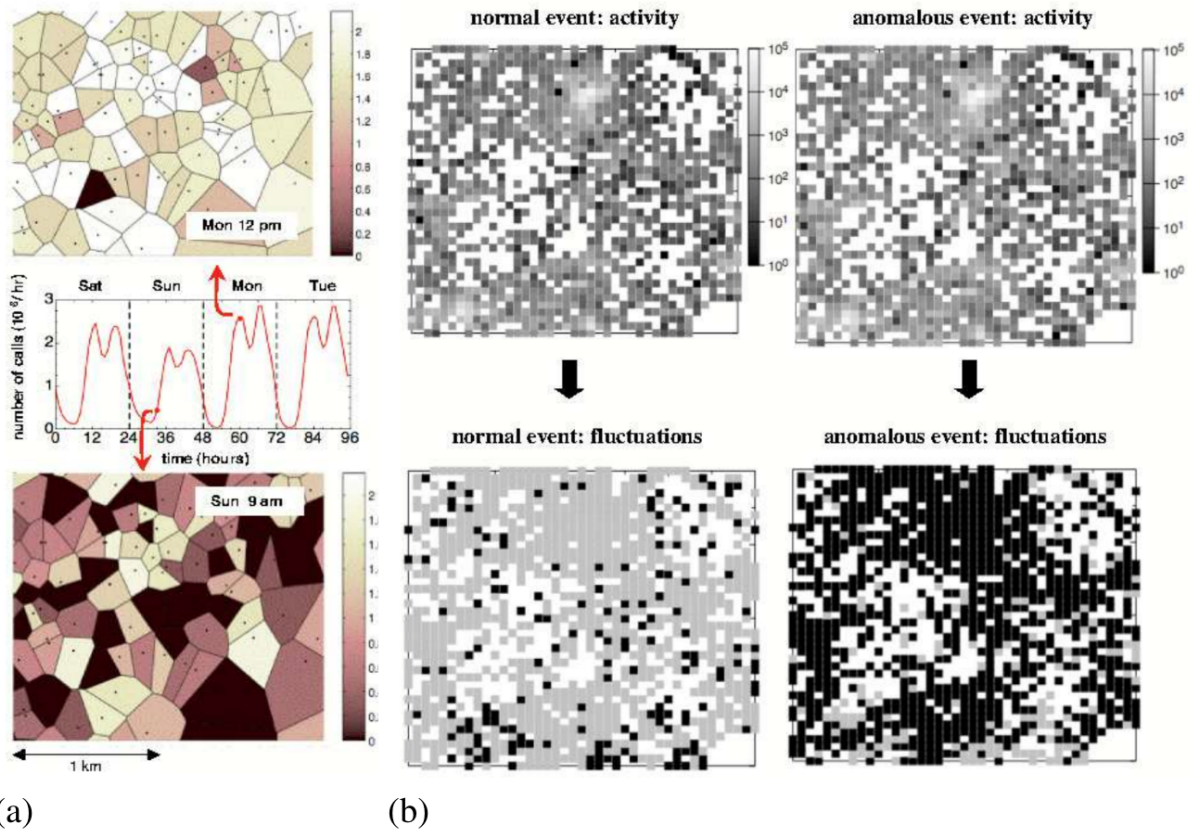
Como resultado, foram detectados alguns mesmos padrões de comportamento presentes em vários grupos de usuários, particularmente demonstrando a ideia de que as chamadas telefônicas realizadas seguiram uma distribuição probabilística de cauda longa (RESNICK, 2007) e não uma de Poisson (CONSUL; JAIN, 1973). Tal descoberta corroborou a ideia de alguns trabalhos teóricos e práticos, em menor escala, anteriores (CANDIA *et al.*, 2008).

## 2.2.2 Representações Tridimensionais

No tocante ao emprego de representações tridimensionais para a análise de padrões de movimento, pode ser inicialmente destacado o trabalho de HUANG; WONG (2015), que identificou padrões de mobilidade humana com base em dados de redes sociais, empregando variações sobre cubos espaçotemporais (Figura 29a). Já YIN *et al.* (2016) realizou o mesmo intento, porém representando os fluxos de comunicação sobre um modelo tridimensional do planeta Terra (Figura 29b).



Figura 28 – Emprego de diferentes técnicas de visualização para avaliação de anomalias em ligações telefônicas por: (a) diagramas de Voronoi associados a gráficos estatísticos e (b) comparação entre tabelas de pixels.



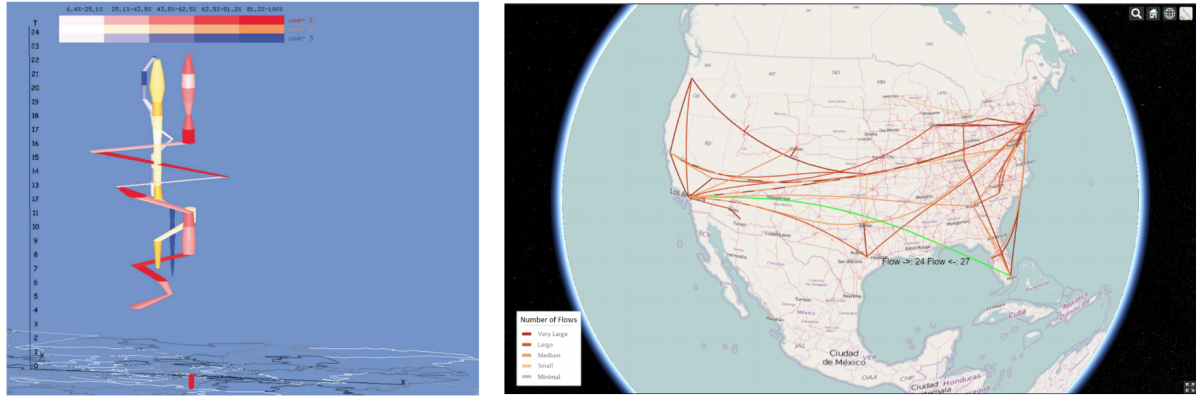
Fonte: CANDIA *et al.* (2008).

Já com relação especificamente à detecção de padrões de comportamento com base na visualização de modelos tridimensionais, GAO (2015) representou conexões entre ligações de telefones celulares, empregando curvas de Bézier num espaço tridimensional, para compreender as dinâmicas de comunicação entre pessoas de acordo com horários específicos. E LOPEZ *et al.* (2017) acompanharam a evolução de clusters tridimensionais, que representavam o fluxo de veículos em rodovias, para detectar pontos de congestionamento, buscando compreender suas respectivas causas.

De um modo geral, apesar das representações tridimensionais muitas vezes não agregarem grandes diferenciais à compreensão das respectivas informações representadas (SIOW *et al.*, 2018), elas podem ser particularmente úteis quando os dados referem-se ao espaço tridimensional real. Assim, por exemplo, pode-se empregar tais modelos para avaliar movimentos no ar ou sob a água, que geralmente são mais difíceis de visualizar que aqueles realizados em superfícies (ANDRIENKO; ANDRIENKO, 2013).

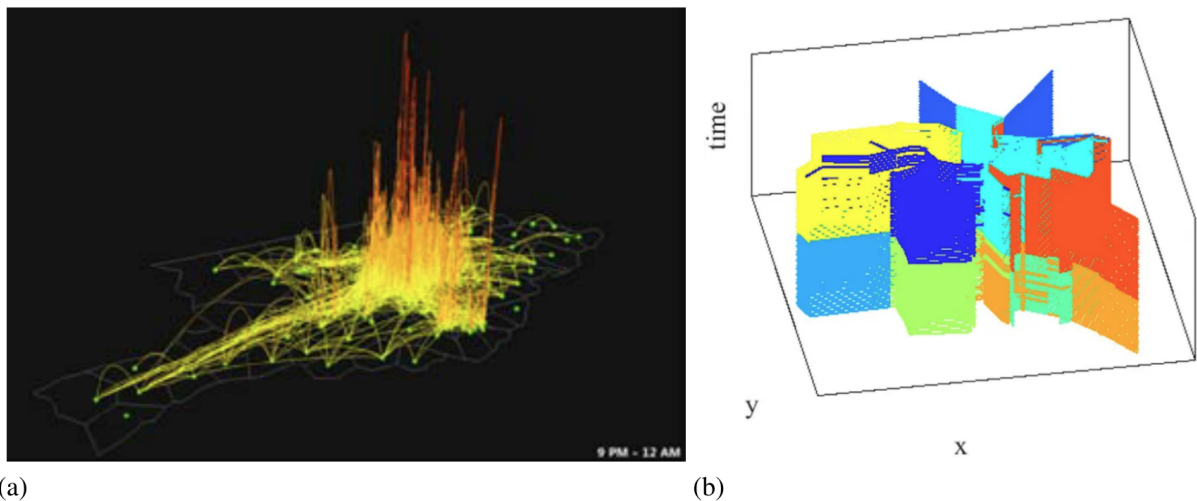
Por exemplo, tal proposta fica bastante evidente no trabalho de WARE *et al.* (2006),

Figura 29 – Dois exemplos de representação tridimensional para identificação de padrões de movimento com base em dados de redes sociais: (a) variações de cubos espaçotemporais e (b) representação de fluxos de comunicação sobre um modelo tridimensional do planeta Terra.



Fonte: (a) HUANG; WONG (2015) e (b) YIN *et al.* (2016).

Figura 30 – Dois exemplos de representação tridimensional para identificação de padrões de comportamento: (a) empregando curvas de Bézier num espaço tridimensional, representando ligações entre telefones celulares e (b) através da análise de clusters tridimensionais, representando o fluxo de veículos em uma rodovia.

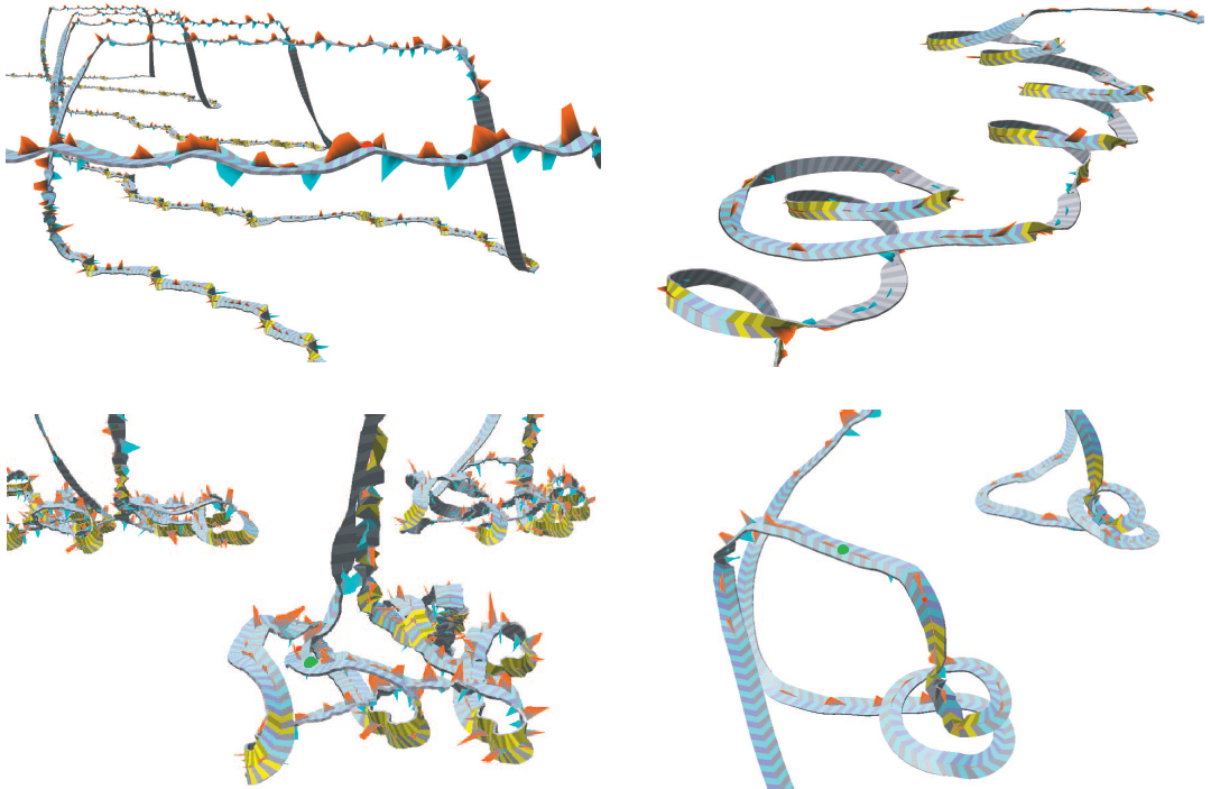


Fonte: (a) GAO (2015) e (b) LOPEZ *et al.* (2017).

que acompanharam por diversas horas a movimentação de baleias no mar, rastreadas através de sensores específicos. Analisando o conjunto de registros coletados, com base num sistema de visualização tridimensional, foi possível identificar diversos padrões de movimentação para alimentação adotados pelos animais (Figura 31).

No mesmo trabalho, apesar de inicialmente voltado para a análise de padrões de movimento, Ware e seus coautores ainda conseguiram diferenciar padrões de comportamento também referentes à alimentação dos cetáceos estudados. Assim, com as mudanças climáticas ocorridas na região estudada em dois anos consecutivos, pôde ser notado que, em 2004, as baleias

Figura 31 – Exemplos de padrões de movimentação empregados por baleias para alimentação, identificados no trabalho de WARE *et al.* (2006).



Fonte: WARE *et al.* (2006).

concentraram-se mais em explorar o fundo do mar; e, em 2005, elas passaram a se alimentar mais de krill<sup>11</sup>, encontrado próximo à superfície. Assim, possivelmente as alterações no ambiente acabaram repercutindo em alterações no próprio comportamento dos animais (WARE *et al.*, 2006).

### 2.3 Considerações Finais

Neste capítulo foram apresentados diversos trabalhos diretamente relacionados ao tema desta tese, sendo os mesmos divididos em duas áreas principais: a representação gráfica de dados espaçotemporais e atributos; e a identificação de padrões de movimento e de comportamento a partir de representações gráficas.

Como resultado deste capítulo, nota-se que a representação de dados espaçotemporais pode ser uma tarefa bastante complexa. Percebe-se ainda que, embora algumas das técnicas citadas possam ser adaptadas para a análise visual de PSTs, elas possivelmente ainda apresentariam os seguintes problemas: incapacidade de lidar adequadamente com a visualização

<sup>11</sup> animais invertebrados, semelhantes a camarões, que fazem parte do zooplâncton.

de dados específicos e seus relacionamentos, particularmente a inter-relação entre nós e arestas; baixa interatividade e baixo desempenho, devido à complexidade de armazenar e processar todos os dados inter-relacionados das PSTs. Além disso, como tais técnicas originalmente não foram estabelecidas para esse fim específico, provavelmente haveria problemas de oclusão e poluição visuais quando fosse necessário apresentar todos os dados das PSTs simultaneamente.

Além disso, deve-se ainda levar em consideração a avaliação das probabilidades armazenadas nas PSTs em relação ao espaço, ao tempo, e a conjuntos que relacionam simultaneamente espaço e tempo.

Considerando todas as dificuldades citadas, bem como as questões acerca da avaliação das probabilidades das PSTs, o próximo capítulo discute alguns conceitos específicos, necessários ao devido embasamento da abordagem proposta nesta tese.

### 3 CONCEITOS PRELIMINARES

Neste capítulo são apresentados conceitos específicos necessários ao devido embasamento do presente trabalho. Inicialmente a Seção 3.1 expõe o conceito de predição de localização por PSTs, discutindo inicialmente a predição de localização como um todo e, depois, a estrutura das PSTs e como as mesmas são geralmente utilizadas. Depois, as Seções 3.2 e 3.3 abordam respectivamente os conceitos de análise visual e de sensemaking. Posteriormente, a Seção 3.4 aborda a representação de branching time, discutindo os modos de organização do tempo em geral, o branching time propriamente dito e, por fim, sua relação com trajetórias. Ao final são tecidas considerações gerais acerca do Capítulo, especificamente na Seção 3.5.

#### 3.1 Predição de Localização por Predictive Suffix Trees

A seguir é apresentado o conceito geral de predição de localização por PSTs. Para uma melhor compreensão, no entanto, o mesmo encontra-se dividido em tópicos mais específicos, quais sejam: predição de localização e PSTs propriamente ditas.

##### 3.1.1 Predição de Localização

A trajetória de um corpo pode ser descrita como sendo uma sequência de dados espaçotemporais, ou seja, posições específicas no espaço estabelecidas em determinados pontos temporais (GIANNOTTI *et al.*, 2007). Assim, quando se analisa uma certa quantidade de trajetórias de um ou mais corpos, muitas vezes é possível estabelecer relações espaçotemporais, compostas por dados espaciais, geralmente consistindo nas latitudes e longitudes de pontos geográficos, e por dados temporais, os *timestamps*<sup>1</sup>, que indicam quando a coleta de cada ponto geográfico foi realizada (LIU *et al.*, 2006).

Muitas vezes, tais relações espaçotemporais podem apresentar algum tipo de comportamento repetitivo, que é conhecido como padrão de trajetória (GIANNOTTI *et al.*, 2007). Considerando-se esses padrões de trajetórias, pode-se estabelecer o conceito de predição de localização de um ou mais corpos, que em termos gerais pode ser descrito como o desafio de prever a próxima localização de um indivíduo ou objeto com base em seus dados históricos espaçotemporais (ROCHA *et al.*, 2016).

Normalmente, as soluções para predição de localização desempenham papéis im-

<sup>1</sup> cadeias de caracteres que geralmente denotam a hora e a data em que certo evento ocorreu.

portantes em áreas diversas, por exemplo, abrangendo desde a análise de popularidade de locais específicos (MONTOLIU; GATICA-PEREZ, 2010) e estabelecimento de rotas de viagens (ZHENG *et al.*, 2011) até a avaliação de vias para a otimização do fluxo de veículos (LU *et al.*, 2011).

O cálculo de predição de localização baseada em padrões de movimento é geralmente feito a partir de uma função de movimentação, com base na velocidade e na direção dos movimentos recentes realizados pelo corpo considerado. No entanto, esse tipo de função normalmente só se mostra útil para intervalos de tempo futuro bastante curtos e próximos ao tempo atual considerado, sendo então geralmente ineficazes para grandes intervalos temporais futuros (JEUNG *et al.*, 2008).

Assim, para evitar tal problema, pode-se tentar prever a localização de um determinado objeto com base na análise de seus locais mais frequentados. A partir dessa análise, pode-se então criar regras de inferência (MORZY, 2007) ou sequências bem definidas de padrões de trajetórias (JEUNG *et al.*, 2007; GUYET; QUINIOU, 2008) que associem tais locais entre si. Essa abordagem particular, inclusive, tende a gerar maior precisão para intervalos maiores de tempo futuro (MONREALE *et al.*, 2009). No entanto, tanto as regras de inferência quanto as sequências de padrões podem se mostrar tendenciosas caso também não considerem os dados temporais associados (GONZALEZ *et al.*, 2008).

### 3.1.2 *Predictive Suffix Trees*

Uma alternativa mais confiável para uma projeção de maior prazo da localização de um corpo é basear-se na análise de seus lugares mais visitados, considerando os respectivos períodos envolvidos (tempos de chegada, permanência e saída). Sendo assim, particularmente, para prever localizações com maior precisão, ROCHA *et al.* (2016) propuseram o uso de PSTs<sup>2</sup>, armazenando dados espaciais, temporais e de probabilidade simultaneamente. Tal solução estabelece relações diretas de inferência entre os dados espaçotemporais registrados, simultaneamente também armazenando e representando um número significativo de dados do tipo origem-destino (WOOD *et al.*, 2010), obtidos com base nas trajetórias prováveis, associando diretamente deslocamentos, timestamps e dados de probabilidade (LEI *et al.*, 2013).

Uma PST é um caso particular de árvore de sufixos<sup>3</sup>, cujos nós representam os locais

<sup>2</sup> detalhes acerca da formalização matemática e do processo de construção das PSTs empregadas neste trabalho podem ser obtidos em ROCHA (2016).

<sup>3</sup> estruturas de dados inicialmente concebidas para a busca rápida de substrings (MCCREIGHT, 1976; UKKONEN,

visitados e as arestas, as transições entre esses locais visitados (Figura 1). Assim, por meio de uma rápida inspeção nos diferentes níveis de uma PST, nota-se que diversas informações podem ser diretamente inferidas. Dados espaciais, temporais e suas relações espaçotemporais encontram-se associados de forma bastante compacta e intuitiva, podendo-se perceber facilmente informações sobre os movimentos prováveis do corpo representado, permitindo seu uso de forma geralmente rápida e simples.

Por exemplo, a Figura 1 ilustra uma PST que representa os possíveis movimentos diários de uma determinada pessoa. Nesse caso, torna-se fácil notar que o indivíduo geralmente permanece em casa até as 8:30. Mais tarde, há uma probabilidade de 50% de que ele estará na escola às 8:45 e também uma probabilidade de 50% dele estar no trabalho às 9:00. Além disso, pode-se inferir diretamente que, se essa pessoa vai trabalhar, ela volta para casa às 19:00. A menos que ela vá ao shopping, caso em que ela então chegará em casa mais cedo, às 17:00.

### 3.2 Análise Visual

A Análise Visual tem como objetivo desenvolver métodos, tecnologias e práticas que explorem e combinem as capacidades de processamento de dados de humanos e de máquinas (COOK; THOMAS, 2005; KEIM *et al.*, 2008). Dessa forma, a visualização interativa se torna o meio pelo qual homens e computadores cooperam, empregando suas capacidades distintas, para obter resultados efetivos na geração de imagens e extração de conhecimento a partir das mesmas (ANDRIENKO *et al.*, 2007).

Assim, pode-se perceber que a análise visual é diferente de outras abordagens padrão de análise, uma vez que se baseia na premissa de que representações visuais interativas podem amplificar a capacidade humana de detectar padrões, estabelecer relações e criar inferências (FABRIKANT; LOBBEN, 2009).

Em um contexto mais específico de GIS <sup>4</sup>, intimamente relacionado ao presente trabalho, ANDRIENKO *et al.* (2007) consideram a *análise geovisual* como sendo a atividade humana assistida por computador com base no processamento de dados espaçotemporais. Como resultado, os analistas obtêm informações úteis extraíndo conhecimento a partir da interação com imagens digitais. Dessa forma, o uso de gráficos interativos pode acelerar processos de

---

1995), que acabaram sendo adaptadas para uso em diversas outras áreas, tais como Genética (BRĀZMA *et al.*, 1998), Música (CHEN *et al.*, 2000) e Comunicação de Dados (LI *et al.*, 2010).

<sup>4</sup> *Geographic Information Systems*, sistemas voltados à representação, análise e gestão de dados geográficos, incluindo aqueles de caráter espaçotemporal.

investigação, focando no projeto e no uso de diferentes tipos de soluções gráficas e respectivas funcionalidades interativas para estimular a geração de ideias (ANDRIENKO *et al.*, 2010).

Deve-se ressaltar que a análise visual possui sutilezas que a diferenciam particularmente de outras abordagens investigativas, como a mineração de dados ou a visualização científica, por se basear especificamente em representações visuais interativas, buscando que analistas ultrapassem suas capacidades de interpretação natural (FABRIKANT; LOBBEN, 2009).

A análise visual pode apresentar vantagens sobre procedimentos automáticos, pois permite que analistas utilizem seus próprios conhecimentos prévios para elaborar questões (estabelecer hipóteses) e buscar respostas (confirmar ou refutar as hipóteses estabelecidas), à medida que observam mudanças em eventos nos contextos avaliados (FABRIKANT; LOBBEN, 2009). De todo modo, a análise visual pode - e deve - empregar ferramentas complementares automáticas ou semiautomáticas para ajudar usuários a realizarem investigações mais rápidas e eficientes (ANDRIENKO *et al.*, 2010; SIOW *et al.*, 2018), devendo tais ferramentas serem configuráveis para se adaptarem mais facilmente ao modo como os analistas trabalham.

Nesse ínterim, o desenvolvimento de ferramentas gráficas interativas cada vez mais poderosas tem estimulado o surgimento de ideias, estabelecendo os seguintes tópicos como os eixos gerais da pesquisa em análise visual (ANDRIENKO *et al.*, 2010):

- Ênfase em análise de dados, resolução de problemas e tomada de decisões;
- Emprego de técnicas automatizadas para processamento de dados e algoritmos de descoberta de conhecimento;
- Envolvimento do ser humano como entidade ativa do processamento, particularmente através do estabelecimento de interfaces visuais interativas;
- Apoio à obtenção, tratamento e correto fornecimento de dados iniciais; e
- Suporte à comunicação dos resultados obtidos aos destinatários mais adequados.

Em particular, deve-se destacar que esses tópicos balizaram a execução da presente tese.

### **3.3 Sensemaking**

A metodologia sensemaking (*dar sentido*, em tradução livre para o Português) foi formalmente estabelecido na Ciência da Computação nos anos 80 (DERVIN, 1983), embora muitas de suas ideias já tivessem sido empregadas desde os primeiros experimentos que buscaram a simplificação da interação humano-computador. De forma resumida e direta, DERVIN (1983)



define que sensemaking é o modo como as pessoas dão sentido às coisas a partir de suas próprias experiências. Assim, inicialmente tal metodologia procura tornar claras e simples as representações de elementos para facilitar a execução de tarefas com base nos mesmos. Sensemaking mostra-se essencialmente interdisciplinar, no sentido de que se apropria e estende conceitos de muitos campos diferentes do conhecimento, como Educação, Psicologia, Filosofia e Sociologia, por exemplo (RUSSELL *et al.*, 1993).

Dessa forma, sensemaking tem sido aplicada para simplificar a compreensão de ideias em muitas áreas diferentes, abrangendo desde a área de Negócios (AGUINIS; GLAVAS, 2019) e Educação (DUFFY, 1995) até Saúde (RHODES *et al.*, 2016), por exemplo.

Mas, apesar de seu uso geral, sensemaking vem sendo particularmente empregada na criação de interfaces gráficas, principalmente na adaptação de contextos de uso, normalmente empregando conhecimentos prévios do usuário. Tal utilização visa principalmente um maior envolvimento do humano, buscando que o mesmo apresente um maior engajamento para com o software utilizado.

Nesse caso, a ideia por trás da sensemaking torna-se basicamente agregar sentido à experiência do usuário, permitindo que o mesmo possa acessar, manipular e explorar dados de diversas maneiras, porém do modo mais intuitivo possível e sempre de forma ativa (PIROLI; CARD, 2005; KLEIN *et al.*, 2006).

Como exemplo prático, pode-se citar, na *metáfora de desktop*<sup>5</sup>, a ação de levar um documento até a lixeira virtual, na área de trabalho, para excluir arquivos no computador (BLACKWELL, 2006). Nesse caso, o usuário adota conceitos e conhecimentos do mundo físico (documento e lixeira) para realizar operações virtuais, substituindo a aprendizagem de ideias abstratas (arquivo de computador) e comandos textuais para lidar com o sistema operacional (OSWALD, 2013).

Em um contexto semelhante, merece ser citado ainda o conceito de *esqueuomorfismo*<sup>6</sup>, que na área da Computação concebe interfaces gráficas com base na mimetização de objetos físicos. Assim, por exemplo, o aspecto visual de aplicações como *media players* tem sido muitas vezes baseado em antigos gravadores de fitas cassete, mantendo inclusive os respectivos símbolos dos botões para as operações de pausa, reprodução etc. Da mesma forma, softwares de agenda,

<sup>5</sup> conceito de interface gráfica, definida por Alan Kay no Xerox Palo Alto Research Center nos Anos 70 e hoje bastante difundida em diversas plataformas de software, que se baseia na utilização de objetos do cotidiano para a manipulação de elementos virtuais (KOVED; SELKER, 1999).

<sup>6</sup> quando um objeto derivado mantém elementos que eram realmente necessários somente no original, o objeto simulado: por exemplo, recipientes de plástico adornados com rebites, presentes originalmente em similares de metal.

livros, calculadoras e muitos outros acabaram herdando o *layout* dos respectivos produtos físicos originais.

Portanto, pode-se notar que essa opção de *design* aproveitou o conhecimento prévio de toda uma geração inicial de usuários para propor a simplificação do uso de novas soluções digitais (CHO *et al.*, 2015). Tanto é assim que hoje o emprego do esquemorfismo tem diminuído substancialmente nos últimos tempos, cedendo seu espaço ao *flat design*<sup>7</sup>, *material design*<sup>8</sup> e outras linguagens visuais e interativas. Isso decorre principalmente da busca por inovações, mas também pelas novas gerações de usuários muitas vezes sequer terem tido qualquer tipo de contato com os produtos físicos originais simulados.

No entanto, sensemaking não se deve restringir ao *compreender para fazer*. Assim, particularmente, MOORE (2017) discute que sensemaking deve ir além do simples conhecimento de conteúdos individuais e específicos:

"por exemplo, ao estudar história, saber quando e onde algo aconteceu é necessário, mas insuficiente para a compreensão de por que e de como. Conhecer fatos científicos, como a forma e a localização das mitocôndrias, é necessário, mas insuficiente para a compreensão de seu papel na célula. Conhecer e identificar estruturas narrativas é necessário, mas insuficiente para a compreensão da mensagem do escritor de um livro."

Assim, saber em geral é necessário, mas insuficiente para um real entendimento (GRENO, 1994). Então, sensemaking, deve ser também *compreender para compreender mais e melhor*.

Particularmente importante nesta tese, para casos específicos de análise visual, sensemaking necessariamente deve envolver o processo de *autoavaliação*, que consiste em alguém fazer perguntas válidas sobre os dados investigados a fim de, então, buscar suas próprias respostas (SHUM; CRICK, 2012; VERBERT *et al.*, 2013).

Assim, DERVIN (1983), em seu trabalho seminal sobre sensemaking, afirma que a busca e uso da informação devem ser atividades "construtoras", referindo-se especificamente à necessidade constante de o analista estabelecer suas próprias percepções pessoais. Ele, em hipóteses alguma, pode se tornar um mero observador, mas sim um ente sempre crítico e atuante e, portanto, necessariamente pensante. Logo, o usuário não deve só se aproveitar de seus

<sup>7</sup> <https://developer.apple.com/design/human-interface-guidelines/>

<sup>8</sup> <https://material.io/>

conhecimentos prévios. Ele tem também de ser capaz de apropriar-se dos novos conhecimentos apresentados.

Então, dada essa dicotomia estabelecida entre *o fazer e o compreender*, o grande desafio do uso de sensemaking em análise visual acaba sendo buscar um equilíbrio. Esse equilíbrio deve, inicialmente, permitir o estabelecimento e a compreensão de conceitos para que uma investigação seja realizada corretamente; e, posteriormente, conferir suficiente liberdade para se ir além, fornecendo ferramentas para que o analista possa criar e explorar seus próprios conceitos durante a investigação (WISE *et al.*, 2016).

Por último, é importante salientar que tem se tornado cada vez mais comum a adoção de sensemaking como base para a apresentação de conceitos abstratos (SIEMENS; LONG, 2011), como é o caso da presente tese. Isso fica patente numa série recente de trabalhos, que envolvem diferentes públicos e áreas do conhecimento. Por exemplo, CAGLAYAN (2018) avalia como o uso do software MATLAB<sup>9</sup>, incluindo seu uso na geração de imagens, pode ser adotado por alunos universitários no estudo de conceitos de álgebra linear. Já BOBIS; WAY (2018) discutem como diferentes meios de representação empregados por crianças, destacando suas produções gráficas digitais, podem ser explorados na compreensão de conceitos matemáticos. Na mesma área, porém adotando tecnologias mais recentes, NG *et al.* (2018) descrevem como interações, mediadas por software, entre modelos tridimensionais virtuais e físicos podem ser uma alternativa para simplificar a interpretação de conceitos da Geometria e da Trigonometria. Já no campo da Física, BEHESHTI *et al.* (2018) utilizam mesas interativas num museu para apresentar a ideia de circuitos eletrônicos para pais e filhos visitantes do local. E WARBURTON; LAUGHLIN (2018) vão além, e exploram a fusão entre dança e imagens digitais projetadas num palco, buscando facilitar a interpretação de problemas relacionados à Lei da Gravitação Universal de Newton.

### 3.4 Representação de Branching Time

A seguir é discutido o conceito de branching time e apresentadas considerações acerca de sua representação. Para simplificar o entendimento, optou-se por dividir as discussões em torno de três tópicos principais: modos de organização do tempo, o conceito de branching time propriamente dito e a relação entre branching time e trajetórias.

<sup>9</sup> <http://www.mathworks.com/products/matlab.html>

### 3.4.1 *Modos de organização do tempo*

Compilando os esforços de FRANK (1998) e de AIGNER *et al.* (2008) para categorizar os modos de organização do tempo, os seguintes modelos gerais podem ser definidos:

- Tempo Linear x Tempo Cíclico. O tempo linear corresponde à percepção natural do tempo, representando uma sequência de primitivas temporais que são estabelecidas necessariamente do passado para o futuro. Já o tempo cíclico é composto por um conjunto de primitivas temporais recorrentes (dias da semana, meses do ano etc.), onde dados eventos podem acabar ocorrendo num padrão repetitivo específico;
- Ponto Temporal x Intervalo Temporal. O ponto temporal representa um evento cuja duração não é considerada, sendo importante somente para a ordenação de ocorrências, por exemplo. Já o intervalo temporal é aquele onde a duração do tempo assume um papel relevante, sendo necessária a indicação de um início e de um final bem definidos para o estabelecimento de um determinado evento;
- Tempo Ordenado x Branching Time. O tempo ordenado corresponde a eventos únicos que ocorrem necessariamente um após o outro. Já o modelo de branching time estabelece que um determinado evento pode dar origem a outros dois ou mais eventos distintos (THOMASON, 1970).

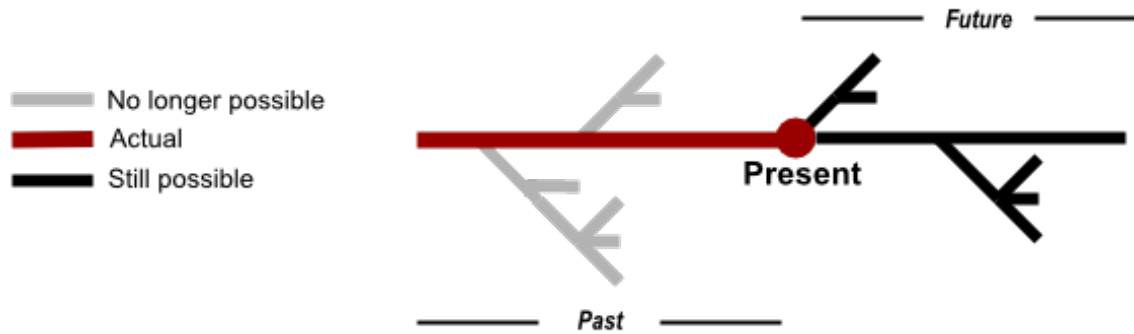
Dadas as intenções específicas deste trabalho, faz-se necessário um maior aprofundamento especificamente no conceito de branching time, discutido mais adiante.

### 3.4.2 *Branching Time*

Inicialmente, o termo branching time relacionava-se à classificação de estruturas narrativas presentes em algumas obras literárias, nas quais um dado evento poderia ser a origem de dois ou mais outros eventos distintos (MARTÍNEZ, 2011). Mais tarde, considerando a evolução da comunicação e das mídias como um todo, o conceito de branching time, mesmo adotando outros nomes, passou a ser aplicado também a filmes, peças teatrais, videogames e mesmo revistas e livros (WULLSCHLAGER, 2002; GOSCIOLA, 2003; SCHUYTEMA, 2008).

Em particular, o branching time tem se tornado importante também como ferramenta para avaliar cenários em atividades de planejamento, onde é possível analisar situações nas quais sequências de ações podem ser previstas, muitas vezes com múltiplas alternativas, para então embasar devidamente decisões futuras (Figura 32) (AIGNER *et al.*, 2008).

Figura 32 – Representação gráfica do conceito de branching time.

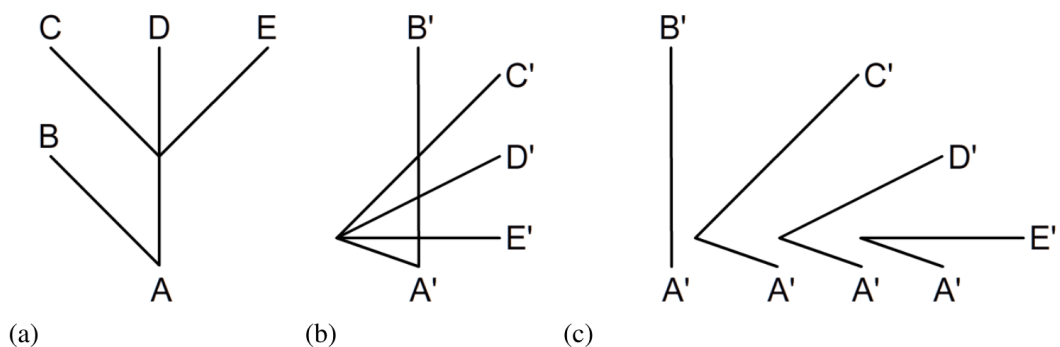


Fonte: o autor.

Aproveitando-se dessa multiplicidade de aplicações, o branching time também pode ser empregado em modelos espaçotemporais, a fim de representar problemas complexos. Tal observação é particularmente pertinente a questões presentes nesta tese, em que se busca determinar a probabilidade de um determinado indivíduo estar presente numa determinada localidade num determinado horário, por exemplo.

### 3.4.3 Branching Time x Trajetórias

Figura 33 – Exemplo de trajetórias oriundas de branching time: (a) conjunto inicial de possibilidades temporais, (b) conjunto derivado de trajetórias e (c) trajetórias individuais.



Fonte: o autor.

Considerando a definição de trajetória da Seção 3.1, pode-se aplicar o conceito de branching time em modelos espaçotemporais, assumindo-se a alocação de pontos espaciais respectivamente a cada ponto temporal considerado. Assim, percebe-se que o conjunto de possibilidades temporais representadas (Figura 33a) pode, então, ser diretamente mapeado em um conjunto de trajetórias (Figura 33b); que, por sua vez, pode definir uma ou mais trajetórias individuais (Figura 33c).

No entanto, na visualização de cada uma dessas trajetórias individuais, a observação

de todo o conjunto de possibilidades (relações de tempo, de espaço e de tempo versus espaço) pode se tornar muito complexa. Isso ocorre, pois, a representação simultânea de todas as trajetórias individuais definidas pode muitas vezes levar à oclusão visual dos dados espaçotemporais, inviabilizando processos de visualização, análise e compreensão dos dados representados (SIOW *et al.*, 2018).

De todo modo, é relativamente simples estabelecer uma solução inicial e direta para a solução de tal problema, de processar e visualizar simultaneamente os dados temporais e espaciais do conjunto de trajetórias individuais possíveis oriundas de branching time: representar os dados espaçotemporais através de nós em um grafo diretamente sobre um mapa virtual (dados espaciais), com suas arestas significando as possíveis relações estabelecidas entre os mesmos (dados temporais, então mantendo-se as mesmas relações de passado e de futuro).

Então, com a projeção dos dados de posição, armazenados no grafo, diretamente sobre o mapa virtual, obtém-se facilmente todo o conjunto de trajetórias possíveis representadas.

Esta definição inicial de solução, baseada em grafo, será estendida posteriormente, no Capítulo 4 deste trabalho, e servirá de base à proposta de abordagem proposta nesta tese.

### **3.5 Considerações Finais**

O presente capítulo apresentou os principais conceitos necessários ao devido embasamento para a proposição da abordagem apresentada nesta tese: predição de localização por PSTs, análise visual, sensemaking e branching time.

A partir desses conceitos, e particularmente adaptando a ideia de representação de branching time a partir de grafos, inicialmente apresentada na Seção 3.4.3, o próximo capítulo estabelece devidamente a proposta de análise visual de PSTs para a investigação de padrões de movimento e de comportamento de indivíduos considerando sensemaking e branching time.

## 4 SOLUÇÃO PROPOSTA

O presente capítulo apresenta a abordagem proposta nesta tese. Inicialmente, a Seção 4.1 discute a fundamentação desta solução, a princípio, com base na ideia de que o branching time pode ser representado através de um grafo e, posteriormente, envolvendo os conceitos de análise visual e de sensemaking, e então adotando o emprego das PSTs. Depois, a Seção 4.2 apresenta a arquitetura da solução, destacando seus componentes específicos definidos: extração, representação, interação, filtragem e anotação. A Seção 4.3 expõe o respectivo workflow definido para a otimização do uso da abordagem proposta, destacando suas etapas de identificação, exploração e documentação, bem como as possíveis interações entre as mesmas. Já a Seção 4.4 discute o processo cíclico de avaliação que se espera ser sempre adotado, buscando que o analista seja necessariamente ativo em relação ao processo de investigação, propondo seus próprios questionamentos à medida que avalia os eventos descobertos. Ao final, na Seção 4.5 constam algumas considerações finais sobre o Capítulo.

### 4.1 Fundamentação

Conforme discutido na Seção 3.4.3 deste trabalho, uma opção inicial para a representação de branching time é a adoção de um grafo associado diretamente a um mapa virtual.

Porém, é importante salientar que os tamanhos do grafo gerado e do respectivo mapa virtual utilizado podem se tornar críticos num possível processo de análise visual. Se o número de nós no grafo ou de pixels ou vetores no mapa virtual for muito grande, o desempenho das plataformas de hardware e de software utilizadas pode ser comprometido. E mesmo que seja possível exibir a representação gráfica total do modelo, problemas de visibilidade e de uso podem ocorrer, como a poluição e a oclusão visuais na exibição dos dados ou o aumento da complexidade na navegação e no controle de elementos interativos (ANDRIENKO *et al.*, 2007; SIOW *et al.*, 2018).

Para evitar tais problemas, buscando garantir uma análise visual viável, faz-se necessário definir uma solução mais específica, que seja capaz de representar corretamente todas as possíveis trajetórias individuais geradas, permitindo a exploração de seus dados relacionadas. Assim, já que não há ciclos no modelo de branching time, o grafo proposto originalmente pode ser substituído por uma árvore, uma vez que, em sua essência, árvores são grafos acíclicos conectados (BOOTH; LUEKER, 1976; ROBERTSON; SEYMOUR, 1986). E, considerando

a hierarquia dos nós de uma árvore, pode-se ainda trabalhar a ideia de nós pais e filhos, estabelecendo relações de hereditariedade. Assim, dado um determinado nó é possível identificar nós que têm algum grau de parentesco anterior (ascendência: pai, pai do pai etc) ou posterior (descendência: filho, filho do filho etc).

Dessa forma, operando a substituição do grafo inicial pela árvore, a representação visual do conceito de branching time pode tornar-se mais clara e agradável (SHIN *et al.*, 2011; SCHULZ, 2011). Além disso, uma vez que o princípio de construção de uma árvore baseia-se essencialmente no processo de hierarquização de seus nós, permite-se ainda a possível simplificação do armazenamento e do processamento dos referidos dados (HERMAN *et al.*, 2000).

Sob esse aspecto, cada caminho mínimo entre a raiz e uma determinada folha da árvore representa, então, uma trajetória individual possível.

Pode-se ainda acrescentar dados de probabilidade a essa árvore, transformando-a numa PST, conforme já discutido na Seção 3.1. Assim, considerando as ideias de ROCHA *et al.* (2016), ao se avaliar o conjunto de nós internos inter-relacionados (nós com seus ascendentes e descendentes), torna-se então possível prever onde o corpo, que descreve as trajetórias, estaria num determinado momento.

No entanto, para o devido uso da PST, é essencial que se possa compreender todos os seus dados da forma mais natural e eficiente possível. Assim, a correta visualização de seus componentes associados (tempo, espaço e probabilidades) pode fornecer uma grande quantidade de informações.

Porém, conforme já discutido no Capítulo 2, utilizando os métodos tradicionais de visualização, é difícil compreender devidamente os dados representados, principalmente a dinâmica dos atributos armazenados na PST em relação ao espaço, ao tempo e ao conjunto espaço-tempo como um todo (ANDRIENKO *et al.*, 2010) e ainda as probabilidades associadas (ROCHA *et al.*, 2016).

Assim, propõe-se, então, uma abordagem que aplica sensemaking e branching time para dar suporte ao processo de análise visual de PSTs, relacionando-o às três questões anteriormente estabelecidas na Seção 1.2 deste trabalho: 1) por que alguém visita um determinado local?; 2) como alguém investiga uma ocorrência usando PSTs?; e 3) quais são os tipos de padrões esperados nos dados de uma PST?.

Dessa forma, emprega-se o sensemaking a fim de fornecer um caráter menos abstrato



a uma PST, denotando que seus nós inter-relacionam três informações associadas a um dado evento: o local de ocorrência do evento, o tempo de ocorrência do evento e a probabilidade de ocorrência do evento; e suas arestas representam possíveis mudanças de configurações de estados, ou seja, transições entre eventos, estabelecidas como relações entre espaço, tempo e probabilidade.

Já o conceito de branching time, por sua vez, é utilizado para estabelecer que cada nó de uma PST corresponda a um evento individual, que pode ser a origem de um ou mais outros eventos distintos (nós filhos).

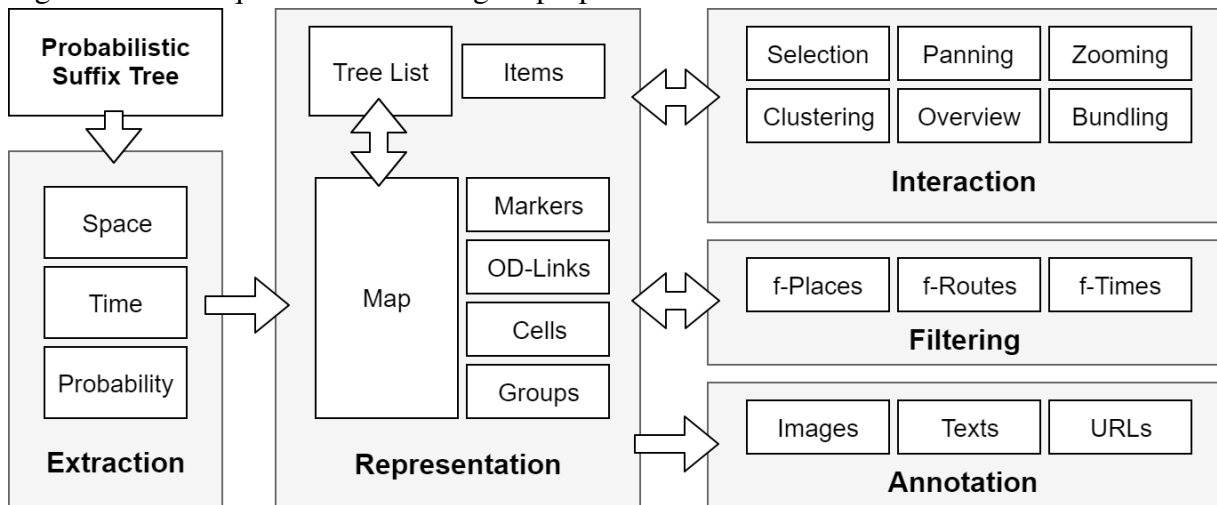
Então, considerando tais conceitos (análise visual, sensemaking e branching time) para avaliar as ocorrências de eventos e as transições entre os mesmos, permite-se investigar possíveis padrões de movimento e de comportamento de indivíduos, especialmente quando as PSTs são examinadas em combinação com um mapa geográfico. Além disso, analisando ocorrências passadas e futuras, e considerando suas respectivas probabilidades e outros indícios levantados na análise, cada um desses padrões de movimento e de comportamento podem, então, ser ou não classificados como padrões prováveis, em vez de simplesmente possíveis.

O estabelecimento dessa solução, incluindo a abordagem, a arquitetura e o workflow, baseou-se no desenvolvimento de um processo incremental (Figura 4), que considerou as contribuições de testes com usuários, realizados por analistas especialistas e não especialistas, exposto mais adiante, no Capítulo 6 deste trabalho. Esse processo incremental também considerou a realização paralela de operações de investigação, pelo próprio autor, sobre as PSTs oriundas dos respectivos datasets utilizados nos testes, para o estabelecimento das tarefas necessárias à avaliação de uso ou como experimento para a obtenção de ideias.

## **4.2 Arquitetura**

A arquitetura da abordagem proposta é mostrada na Figura 34, definindo interações entre cinco componentes específicos: Extração, Representação, Interação, Filtragem e Anotação. Esses componentes fornecem as funcionalidades de investigação necessárias à identificação, à exploração e à documentação de eventos em PSTs e são devidamente apresentados a seguir.

Figura 34 – A arquitetura da abordagem proposta.



Fonte: o autor.

#### 4.2.1 Extração

Dado que as PSTs armazenam tempo (timestamps), espaço (coordenadas geográficas) e dados de probabilidade simultaneamente de forma inter-relacionada, é necessário extrair esses três tipos específicos de dados independentemente para realizar o posterior processamento dos mesmos. No entanto, é necessário preservar todas as relações pré-existentes, pois o resultado desse processo de extração deve dar suporte à execução das três tarefas necessárias relacionadas à investigação de ocorrências utilizando PSTs: caracterização, busca e comparação, as quais já foram estabelecidas na Seção 1.2 deste trabalho.

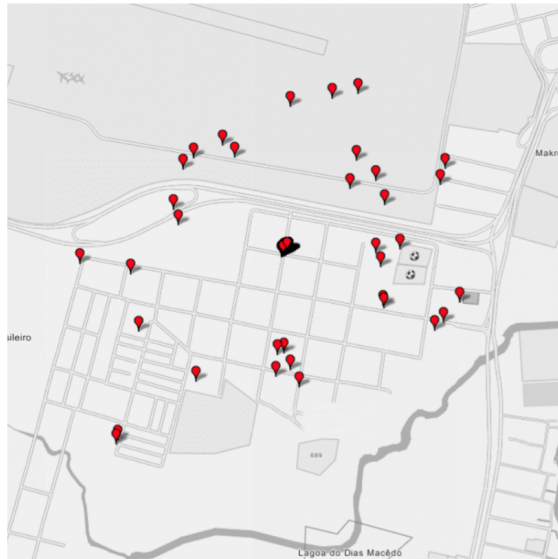
#### 4.2.2 Representação

A representação dos dados extraídos da PST, obtidos na etapa de Extração, encontra-se dividida em dois modelos paralelos e integrados: uma *tree list* e um mapa virtual. O uso simultâneo desses dois modelos visuais e interativos visa auxiliar o analista a compreender melhor as relações estabelecidas entre os eventos armazenados nas PSTs, permitindo que seus respectivos dados sejam tratados de forma mais adequada e efetiva.

A tree list agrega itens que representam os elementos da PST: nós, que armazenam dados específicos sobre os eventos armazenados; e arestas, que representam as relações de hereditariedade estabelecidas entre nós pais e filhos, respectivamente eventos passados e futuros.

O mapa virtual, por sua vez, permite a identificação geográfica dos eventos armazenados na PST. Para tanto, definem-se os seguintes elementos:

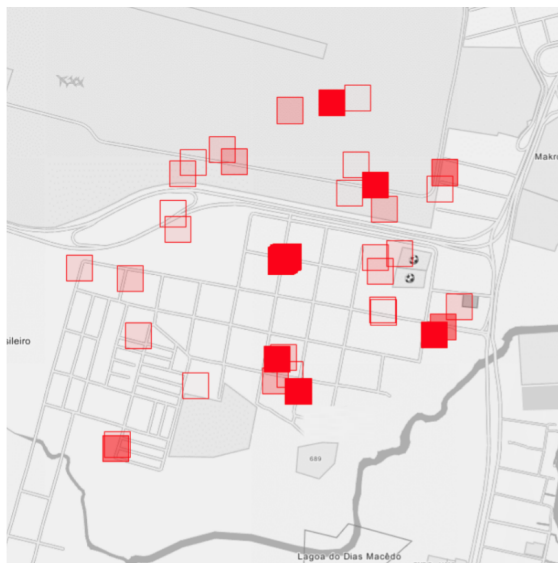
Figura 35 – Elementos referentes à representação visual no mapas virtual: (a) Marcadores em coordenadas geográficas; (b) OD-links que identificam a origem e o destino referentes às arestas de uma PST; (c) Células cujos tons de cor são proporcionais à probabilidade de ocorrência dos respectivos nós da PST representada; e (d) Grupos cujos tons de cor são proporcionais ao número de marcadores presentes em suas respectivas áreas.



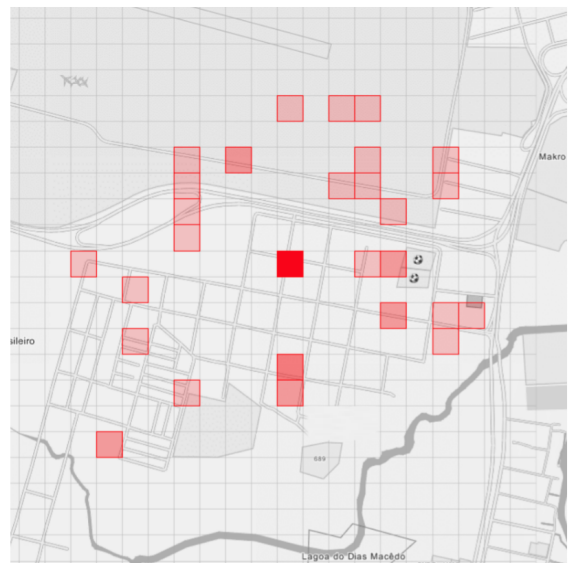
(a)



(b)



(c)



(d)

Fonte: o autor.

- Marcador. Elemento gráfico que é colocado em uma coordenada geográfica específica do mapa virtual, referindo-se a um determinado nó da PST (Figura 35a);
- OD-Link (*Origin-Destination Link*). Segmento de reta que representa uma ligação direta entre origem e destino no mapa virtual, associada a uma determinada aresta da PST (Figura 35b);
- Célula. Quadrado colorido, no mapa virtual, que representa informações de probabilidade

de um determinado marcador (Figura 35c). O tom da cor da célula é proporcional à probabilidade de ocorrência do respectivo evento armazenado no nó da PST e representado pelo marcador;

- Grupo. Quadrado colorido, no mapa virtual, que envolve um ou mais marcadores, e cujo tom da cor é proporcional à quantidade de marcadores envolvidos em sua respectiva área (Figura 35d).

Considerando os elementos definidos anteriormente e adotando os conceitos de sensemaking e de branching time, pode-se, então, estabelecer as seguintes premissas para a representação visual dos dados de PSTs no mapa virtual:

1. Cada nó da PST representa um evento temporal individual, cujos possíveis nós pais e filhos representam respectivamente eventos anteriores e posteriores;
2. Cada nó da PST deve ser apropriadamente representado por um marcador nas respectivas coordenadas geográficas de um mapa virtual;
3. Cada aresta de uma PST deve ser apropriadamente representada por um OD-Link nas respectivas coordenadas geográficas do mapa virtual, exibindo claramente suas relações de hereditariedade;
4. Todo o modelo espaçotemporal, ou parte do mesmo, deve ser visualizado sempre que necessário, podendo apresentar individualmente suas células (relacionadas aos valores de probabilidade de nós) e seus grupos (relacionados ao número de ocorrências de nós em regiões específicas);
5. Porções limitadas do modelo espaçotemporal podem ser examinadas individualmente para estabelecer regiões de interesse (ROIs, *Regions of Interest*) cujos dados inter-relacionados (intervalos de tempo, localizações geográficas e valores de probabilidade) podem ser explorados individualmente ou em grupo.

A partir desses pressupostos, entende-se que seja possível perceber ocorrências considerando individualmente espaço, tempo e probabilidades e, mais importante, sendo ainda permitido relacionar direta e simultaneamente essas três componentes para análise.

### 4.2.3 Interação

Para evitar problemas de processamento (limitações na capacidade do hardware empregado, por exemplo), minimizar problemas de visibilidade (poluição ou oclusão visuais, por exemplo) e simplificar o uso (diminuir as complexidades decorrentes de navegação e de

controle, por exemplo) na exibição e na exploração de todas as informações armazenadas em uma PST (HERMAN *et al.*, 2000; ARTERO *et al.*, 2004; SHIN *et al.*, 2011), oferece-se suporte aos seguintes recursos interativos:

- **Seleção.** Para que o analista escolha um item da tree list ou um marcador no mapa virtual, relacionando-o diretamente aos seus respectivos dados na PST;
- **Panning.** Para que o analista altere a posição da janela de visualização do mapa virtual sem perda da respectiva correlação de espaço;
- **Zooming.** Para que o analista amplie ou reduza o zoom no mapa virtual, observando instantaneamente o resultado;

Os recursos descritos acima, além de permitirem uma visualização mais adequada dos dados representados graficamente, ainda podem possibilitar um melhor aproveitamento das capacidades de processamento e de armazenamento em memória disponíveis. Isso ocorre pois o consequente particionamentos espacial do mapa virtual limita o acesso somente aos dados realmente necessários, desconsiderando o conjunto total dos mesmos (YOUNG, 2001). Além disso, permite-se que o usuário possa navegar mais livremente pelo mapa virtual, podendo realizar uma exploração mais incremental, selecionando gradativamente espaços que despertem algum tipo de interesse, estabelecendo POIs e ROIs à medida que seja julgado necessário.

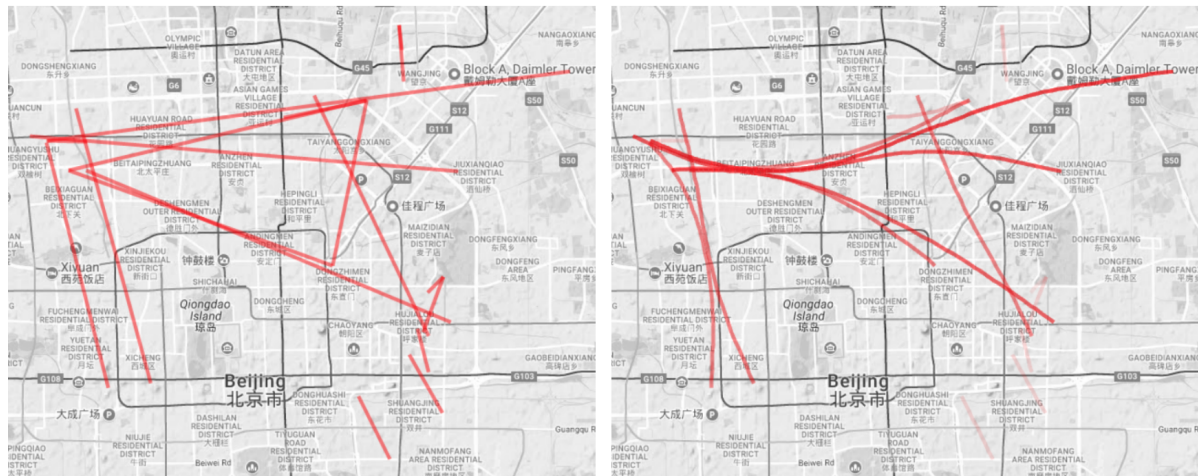
E voltados especificamente à interação com a representação total e simultânea dos dados de uma PST, são propostos ainda os três recursos seguintes:

- **Agregação.** Para que o analista defina atributos de agregação no mapa (tamanho/área e número máximo de eventos a serem considerados, por exemplo) para consolidar dados em grupos ou células (LEE *et al.*, 2008; LEI *et al.*, 2013).
- **Overview.** Para que o analista controle a apresentação simultânea, no mapa virtual, de todos os dados armazenados na PST, isto é, exiba ou oculte, individualmente ou coletivamente, marcadores, OD-Links, células e grupos (Figura 35);
- **Bundling.** Para que o analista, empregando técnicas de *edge bundling display* (ZHOU *et al.*, 2013), controle a agregação de todos os OD-Links em *bundles*<sup>1</sup>, visando fornecer uma visão dos possíveis fluxos de deslocamento presentes numa PST (Figura 36).

De uma forma geral, todos esses recursos interativos foram estabelecidos com a finalidade de fornecer suporte à realização das três tarefas relacionadas à investigação de PSTs, definidas anteriormente na Seção 1.2: caracterização, busca e comparação de ocorrências.

<sup>1</sup> linhas que combinam fluxos visuais dispersos, buscando definir os fluxos principais (ZHOU *et al.*, 2013)

Figura 36 – (a) Todos os OD-Links possíveis obtidos de uma determinada PST e (b) os mesmos OD-Links combinados através de edge bundling (ZHOU *et al.*, 2013).



(a)

(b)

Fonte: o autor.

#### 4.2.4 Filtragem

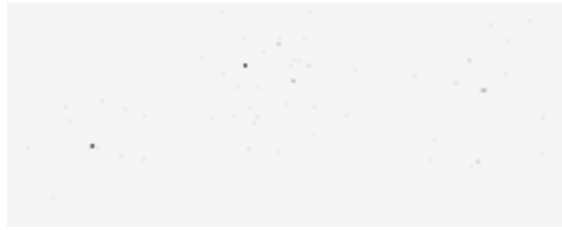
Adotando estratégias para o acúmulo de pixels referentes a pontos ou segmentos de retas no mapa virtual (LEI *et al.*, 2013; SCELLATO *et al.*, 2011), foram criados filtros visuais específicos, estabelecendo três ferramentas semi-automáticas para auxiliar o analista a identificar mais facilmente informações relevantes nos dados representados das PSTs:

- f-POIs. Acumula todos os pixels referentes aos pontos relacionados às posições dos marcadores no mapa virtual para indicar POIs (Figura 37);
- f-Routes. Acumula todos os pixels referentes aos segmentos de reta relacionados aos OD-Links no mapa virtual para indicar as rotas utilizadas mais frequentemente (Figura 38);
- f-Times. Mostra quadrados coloridos no mapa virtual que destacam apenas as localizações dos eventos que ocorrem em um dado intervalo de tempo, especificado pelo analista. O tom da cor de cada quadrado é proporcional à probabilidade de cada nó considerado (Figura 39).

Então, por exemplo, na sequência apresentada na Figura 37 pode-se observar POIs primários (Figura 37b), secundários (Figura 37c) e gerais (Figura 37d), obtidos a partir dos pixels presentes na Figura 37a empregando f-POIs. A Figura 37a foi gerada a partir do acúmulo dos pixels referentes a todos os pontos ocupados pelos marcadores do mapa virtual, relacionados aos nós da PST considerada.

Do mesmo modo, na Figura 38 observa-se a exibição gradual do conjunto de OD-Links correspondentes às rotas utilizadas por um indivíduo, empregando a ferramenta f-Routes.

Figura 37 – Exemplo de filtragem empregando f-POIs: (a) acúmulo de pontos correspondentes aos marcadores; (b) POIs primários; (c) POIs primários e secundários e (d) todos os possíveis POIs obtidos.



(a)



(b)



(c)



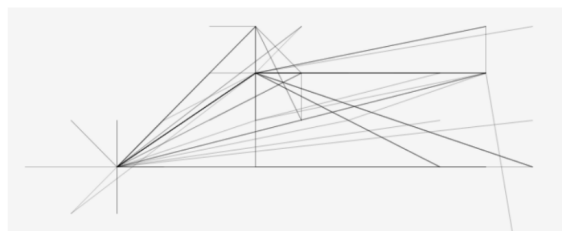
(d)

Fonte: o autor.

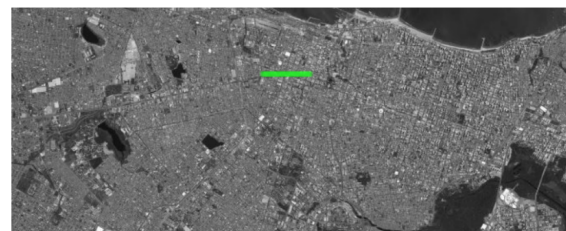
Nesse caso, o analista estabeleceu um limiar de filtragem<sup>2</sup> específico para acumular os pixels referentes aos segmentos de reta relacionados aos OD-Links presentes na respectiva PST. Assim, pode-se perceber desde uma configuração bem inicial de rota (Figura 38b), até um modelo bastante complexo de ligações (Figura 38d), passando por um estágio intermediário que expõe

<sup>2</sup> intensidade mínima ou máxima de pixels que deve ser considerada para exibição no mapa (GONZALEZ; WOODS, 2002)

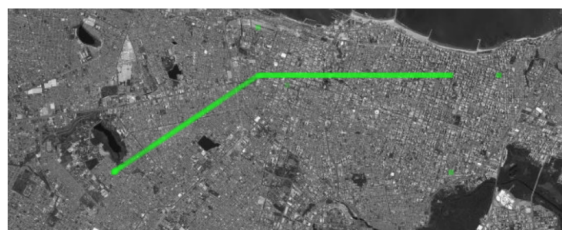
Figura 38 – Exemplo de filtragem empregando f-Routes: (a) acúmulo de segmentos de reta correspondentes aos OD-Links; (b) configuração inicial de rota; (c) possíveis rotas utilizadas mais frequentemente e (d) todas as rotas obtidas.



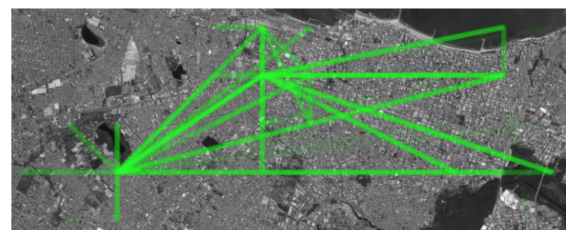
(a)



(b)



(c)



(d)

Fonte: o autor.

as prováveis rotas utilizadas mais frequentemente (Figura 38c), que merecem ser devidamente exploradas no processo de investigação.

Figura 39 – Exemplo de filtragem empregando f-Times: (a) local/probabilidade de um indivíduo estar localizado às 5:00 e (b) às 10:00.



Fonte: o autor.

Já na Figura 39, onde um analista estabeleceu diferentes horários para filtragem empregando a ferramenta f-Times, nota-se facilmente que os dados da PST avaliada indicam que uma pessoa provavelmente reside na região apresentada na Figura 39a, referente ao horário de 5:00. Tal constatação se deve às poucas ocorrências no período e à alta probabilidade de presença, indicada pelo retângulo de cor vermelha intensa. Da mesma forma, os dados da mesma pessoa representam que, às 10:00, ela frequenta as regiões apresentadas na Figura 39b, mais separadas e com probabilidades mais distribuídas, podendo essas serem seus locais de trabalho ou de estudo, por exemplo.

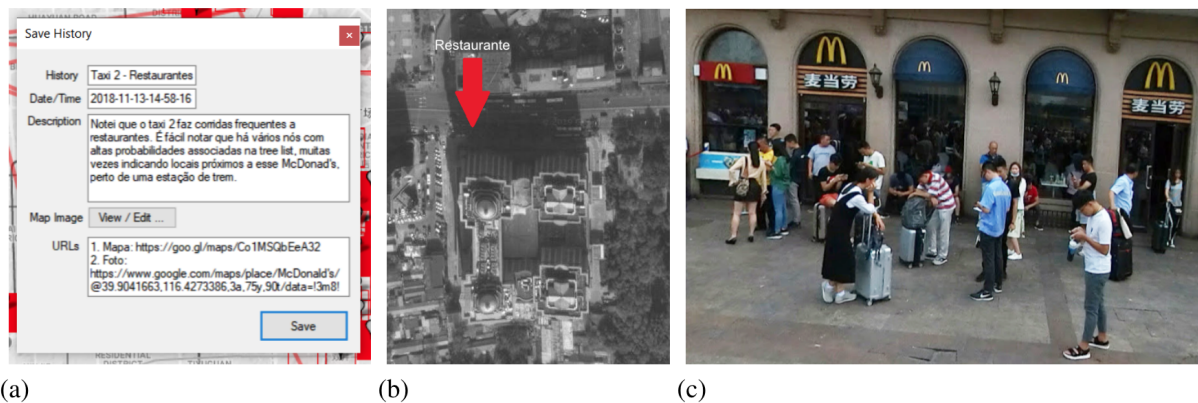
Por último, é importante destacar que tais ferramentas são consideradas semi-automáticas, e não automáticas, pois, conforme os exemplos apresentados, elas precisam ser devidamente configuradas pelo analista, com base no ajuste de valores, para então apresentarem os resultados adequados.

#### 4.2.5 Anotação

Os resultados do processo de análise (mapa virtual com marcadores, OD-links, células, grupos, overview, bundling e filtragens) podem ser editados para inclusão de ilustrações esquemáticas e de rótulos textuais, por exemplo (Figura 40b).



Figura 40 – Exemplo de anotação usando (a) o subsistema de anotação, (b) detalhe de inclusão de uma ilustração esquemática e de rótulo de texto no mapa virtual e (c) fotografia digital oriunda de serviço online.



Fonte: (a) o autor e (b) e (c) <https://www.google.com/maps>.

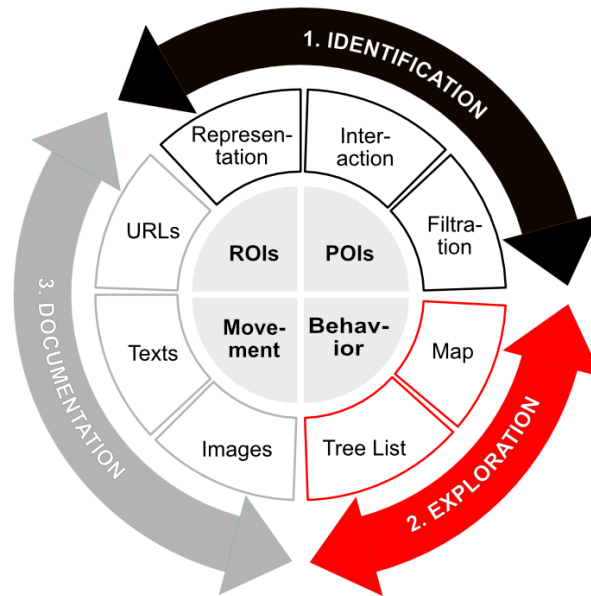
Além de todos esses elementos, os analistas podem ainda usar serviços online para adicionar informações complementares, como mapas externos, fotografias digitais, visitas virtuais e URLs com informações relacionados às ocorrências documentadas (Figura 40c). Os resultados anotados, incluindo possíveis textos para a descrição dos padrões de movimento e de comportamento identificados (Figura 40a), podem ser usados como registros de ocorrências para consultas ou comparações posteriores. Posteriormente, tais registros podem também servir de base para a elaboração de outras questões, próprias, a serem respondidas pelos analistas.

### 4.3 Workflow

O workflow sumarizado na Figura 41 foi projetado para otimizar o processo de investigação, levando em consideração o uso mais apropriado da abordagem proposta. Para elaborar esse workflow, inicialmente, foi observado como especialistas em análise de PST realizavam seu trabalho pelo método tradicional, inspecionando cada nível da PST para isolar dados (coordenadas geográficas, timestamps e probabilidades) e identificar ROIs e POIs, tentando relacionar os mesmos posteriormente (LEITE *et al.*, 2017). Posteriormente, após a análise dos resultados dos testes com usuários descritos no Capítulo 6 deste trabalho, foi definida a configuração final do workflow, que se baseia na execução de tarefas divididas em três etapas específicas: Identificação, Exploração e Documentação.

Essas etapas do workflow, que estabelecem os procedimentos necessários à análise visual de PSTs com base na abordagem proposta, encontram-se devidamente descritas a seguir.

Figura 41 – Workflow definido para uso da abordagem proposta.



Fonte: o autor.

#### 4.3.1 Identificação

Inicialmente, considerando-se a representação geral da PST, deve-se empregar os recursos Overview ou Bundling para exibir todos os OD-Links no mapa simultaneamente. Em seguida, observam-se as quantidades mais significativas de sobreposições ou junções nos respectivos segmentos de reta oriundos da representação dos OD-Links no mapa. Depois, identificam-se os movimentos mais frequentes, buscando-se principalmente os padrões espaciais e os padrões espaciais de mudanças, já discutidos na Seção 1.2 deste trabalho. Com base nessa identificação, podem ser estabelecidos possíveis ROIs e os primeiros POIs a serem inspecionados mais minuciosamente. Nesse momento específico, podem ser empregadas as ferramentas semi-automáticas de filtragem (f-POIs, f-Routes e f-Times) para validar as observações realizadas e encontrar outros POIs em potencial. Isso facilita a decisão sobre quais regiões do mapa virtual merecem ser observadas em mais detalhes na etapa seguinte.

#### 4.3.2 Exploração

Depois da identificação das regiões do mapa virtual a serem investigadas mais minuciosamente, grupos e células podem ser usados para detectar ROIs ainda mais específicos (ARTERO *et al.*, 2004). Para tanto, deve-se proceder uma macro-avaliação, com grupos e células grandes, que podem ser diminuídos à medida que marcadores e OD-Links vão sendo

contextualizados. Assim, para começar a pesquisar as ocorrências identificadas, examinando POIs e ROIs com mais precisão, as relações diretas estabelecidas entre a tree list e o mapa virtual devem ser consideradas. Então, inicialmente pode-se interagir com os itens da tree list, mostrando ou ocultando os respectivos marcadores e OD-Links no mapa virtual. A ação inversa também é possível e, então, pode-se clicar em um marcador do mapa virtual para selecionar automaticamente o respectivo item da tree list. Assim, podem ser identificados os nós da PST, e suas localizações geográficas, ou vice-versa.

Para uma análise ainda mais aprofundada dos POIs identificados, pode-se explorar a hierarquia da PST, investigando nós pais e filhos em vários níveis, procurando relacionar os itens da tree list (tempo e probabilidades associadas) aos marcadores correspondentes no mapa virtual (posições geográficas), conforme pode ser observado na Figura 3. Nesse momento, considerando em particular o tempo, pode-se verificar quando os eventos avaliados parecem ocorrer, buscando identificar padrões temporais, discutidos na Seção 1.2 deste trabalho.

Depois, após identificar os eventos que devem ser realmente considerados, o analista pode finalmente caracterizá-los devidamente. Para tanto, devem-se empregar células e grupos no mapa virtual e observar marcadores, bem como seus respectivos itens da tree list. Assim, considerando as probabilidades envolvidas, avalia-se cada evento específico (nó da PST) em relação a eventos passados e futuros (respectivos nós pais e filhos na PST), de forma a poder classificar as ocorrências analisadas como possíveis ou prováveis.

Então, com base nos resultados obtidos, pode-se tentar determinar as respectivas intenções das ocorrências identificadas (acionadas geograficamente, temporalmente ou semanticamente, conforme apresentado na Seção 1.2 desta tese, a fim de inferir os padrões de movimento e de comportamento associados à PST avaliada.

Para reforçar as descobertas realizadas, opcionalmente, podem ser empregadas ferramentas externas, visitando websites ou usando mapas online, passeios virtuais ou fotografias digitais de locais, por exemplo, para medir distâncias, identificar lugares e, então, compreender os eventos encontrados ainda mais profundamente.

### **4.3.3 Documentação**

Paralelamente às etapas de identificação e de exploração, pode-se proceder à anotação do conjunto de POIs, ROIs, padrões de movimento e de comportamento identificados, incluindo as informações opcionais, coletadas por meio das ferramentas externas utilizadas.

Essas informações opcionais podem abranger desde imagens (mapas e fotografias) e textos (descrição das descobertas e observações sobre o processo de investigação) até URLs associadas a observações (websites específicos e links para mapas online, por exemplo).

Posteriormente, as anotações realizadas podem ainda ser editadas para, por exemplo, incluir rótulos de textos e ilustrações esquemáticas (Figura 40).

Além disso, dado que todo o processo de investigação pode ser registrado através dos recursos de anotação, depois pode-se comparar ocorrências, buscando identificar possíveis relações entre as mesmas.

Por fim, é essencial destacar que a adoção deste workflow não é, de forma alguma, obrigatória. Conforme dito anteriormente, a construção do mesmo baseou-se na observação inicial do trabalho de especialistas e no uso prático, pelo autor deste trabalho e por participantes dos testes com usuários realizados, de aplicações elaboradas com base na abordagem proposta. Sendo assim, usar esse workflow é opcional, uma vez que o mesmo é somente uma proposta de trabalho otimizada, com o intuito de obter resultados de forma eficiente. Sendo assim, analistas podem – e devem – usar livremente todos os recursos da abordagem proposta para explorar a combinação espaço-tempo-probabilidade conforme desejarem e precisarem.

#### **4.4 Processo Cíclico de Avaliação**

É importante salientar que, considerando a abordagem proposta, e adotando ou não o respectivo workflow, à medida que os procedimentos de investigação são realizados, e o analista tenta resolver problemas, incentiva-se que o mesmo, a qualquer momento, estabeleça seus próprios questionamentos a fim de buscar obter novas respostas. Espera-se que isso ocorra independentemente do analista ser um especialista na área de investigação ou não.

Tal prática, incentivada tanto pela análise visual (FABRIKANT; LOBBEN, 2009; ANDRIENKO *et al.*, 2010) quanto pela metodologia sensemaking (DERVIN, 1983; MOORE, 2017), busca a manutenção de um processo de avaliação cíclico e mais rico em resultados, sejam intermediários ou finais. Assim, deve-se almejar sempre estabelecer novas hipóteses, a partir dos questionamentos adicionais, e posteriormente comprovar ou refutar tais hipóteses, de forma a determinar sempre novas possibilidades de respostas (KLEIN *et al.*, 2006).

#### **4.5 Considerações Finais**

Este capítulo estabeleceu a solução de avaliação de PSTs para a investigação de padrões de movimento e de comportamento de indivíduos com base nos conceitos de sensemaking e branching time. Foram apresentados a abordagem proposta, destacando a interação entre os elementos de sua arquitetura, e seu respectivo workflow, que pode ser utilizado opcionalmente, a fim de otimizar as investigações pretendidas. Além disso foi também discutido o processo cíclico de avaliação, com base na ideia de que o analista deve assumir um papel sempre ativo na investigação dos padrões de movimento e de comportamento. O próximo capítulo descreve uma aplicação específica, desenvolvida com base na solução aqui apresentada, que servirá de base para a validação da solução proposta.

## 5 APLICAÇÃO DESENVOLVIDA

Este capítulo apresenta a versão final da aplicação desenvolvida para validação da abordagem proposta nesta tese. A Seção 5.1 inicia contextualizando o desenvolvimento da aplicação e a Seção 5.2 apresenta os vários widget presentes em sua interface gráfica. Depois, a Seção 5.3 expõe as funcionalidades da aplicação, apresentando alguns detalhes de sua implementação. Já a Seção 5.4 foca especificamente no subsistema de anotação e, por fim, são abordadas algumas considerações finais sobre o Capítulo na Seção 5.5.

### 5.1 Contextualização

Após estabelecida a solução proposta, devidamente descrita no Capítulo 4, foi criada uma aplicação de visualização interativa (Figura 3) para avaliar a combinação pretendida de análise visual, sensemaking e branching time. Essa aplicação foi desenvolvida em C#/Windows Forms <sup>1</sup>, utilizando o ambiente de desenvolvimento Visual Studio Community <sup>2</sup> e o componente de visualização de mapas online GMap.NET <sup>3</sup>.

A aplicação desenvolvida relaciona os três componentes de investigação (espaço, tempo e probabilidades) contidos em uma PST armazenada em um arquivo JSON. Ela permite a um usuário explorar dados, criando conexões entre os mesmos, para identificar ocorrências e realizar descobertas. Com base na abordagem proposta, e opcionalmente no respectivo workflow, pode-se extrair e manipular dados (coordenadas geográficas, timestamps e probabilidades associadas) oriundos de uma PST para encontrar padrões de movimento e de comportamento de indivíduos.

Os dados necessários são organizados, de forma sincronizada, em uma tree list (Figura 42F), que preserva todos os níveis da PST original, e em um mapa virtual, que pode ser exibido tanto num modo padrão (Figura 42) quanto numa visão de satélite (Figura 3). Os elementos exibidos no mapa virtual podem empregar diferentes cores para melhorar o contraste visual: uma paleta de vermelhos para o modo padrão ou uma paleta de verdes para a visão de satélite.

<sup>1</sup> <https://docs.microsoft.com/pt-br/dotnet/csharp>

<sup>2</sup> <https://visualstudio.microsoft.com/pt-br/vs/community>

<sup>3</sup> <https://archive.codeplex.com/?p=greatmaps>

## 5.2 Interface Gráfica

Considerando todos os elementos da abordagem proposta apresentados no Capítulo 4 deste trabalho, a interface gráfica da aplicação desenvolvida (Figura 42) é dividida em três visões: configuração e filtragem no canto superior esquerdo; tree list no canto inferior esquerdo; e mapa virtual à direita.

Empregando os widgets na visão de configuração e filtragem, o usuário pode executar um conjunto de ações específicas, tais como: abrir um arquivo JSON com os dados da PST (Figura 42A); apresentar um overview dos dados da PST (Figura 42B), mostrando ou ocultando marcadores, OD-Links, células ou grupos (Figura 42I); definir o número de níveis hierárquicos a exibir nas relações pai/filho (Figura 42J); aplicar filtros visuais aos dados da PST (Figura 42C); utilizar o recurso de edge bundling (Figura 42D); e anotar as observações realizadas (Figura 42E), usando o modo padrão ou a visão de satélite do mapa virtual (Figura 42K).

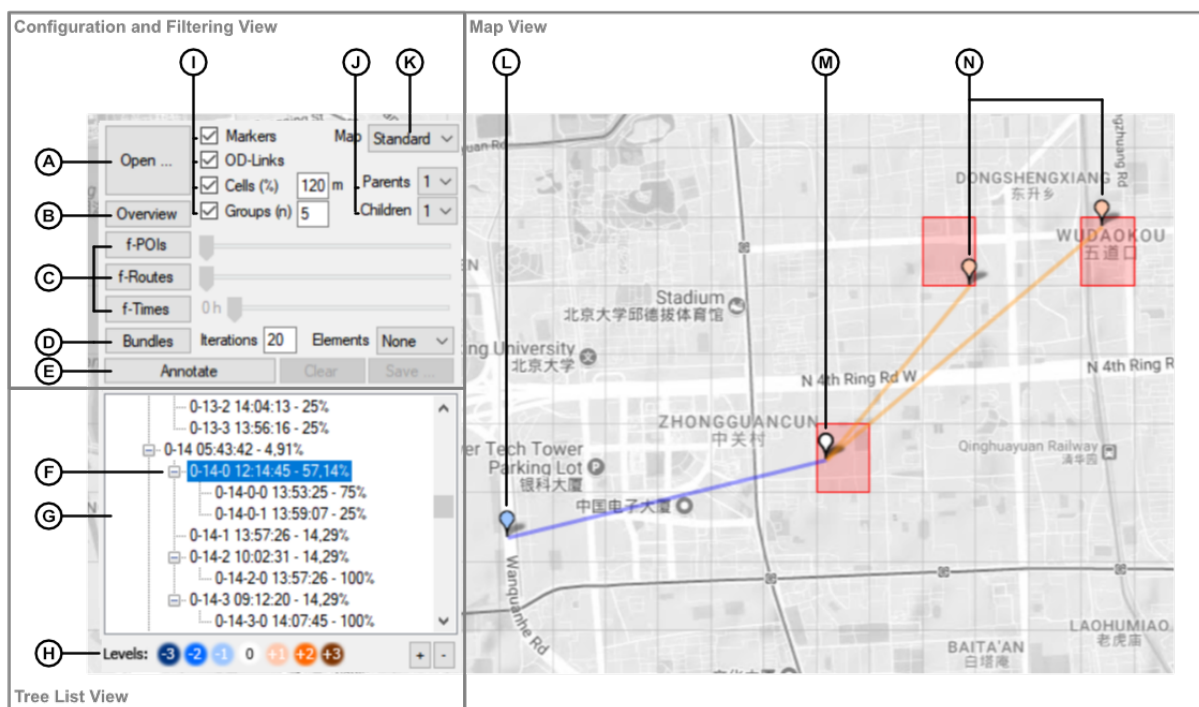
Já na visão de tree list, a hierarquia dos nós da PST é exibida (Figura 42G). Seguindo os pressupostos estabelecidos na abordagem proposta, discutidos na Seção 4.2.2 desta tese, a informação representada em cada item da tree list corresponde a um determinado evento, relacionado a um dado nó da PST. Cada um desses itens encontra-se dividido em três porções: a porção à esquerda apresenta um identificador do evento, indicando sua relação de hereditariedade; a porção central apresenta a hora em que ocorreu o evento, no formato hh:mm:ss; e a porção à direita apresenta a probabilidade de ocorrência do evento. Quando se seleciona um dado evento da PST (Figura 42F), clicando sobre o item correspondente na tree list, sua respectiva representação gráfica é mostrada no mapa virtual, através de um marcador (Figura 42M).

A visão de mapa virtual acessa mapas obtidos diretamente de serviços online que, conforme já dito, podem ser exibidos em modo padrão ou visão de satélite. Nessa visão ocorrem os processos de caracterização, busca e comparação de ocorrências, discutidos na Seção 1.2 deste trabalho, sempre em estreita combinação com a visão de tree list e com as ações executadas a partir da visão de configuração e filtragem.

## 5.3 Funcionalidades

Considerando a abordagem e o workflow definidos na Seção 4.3, a aplicação desenvolvida oferece ao usuário diversas funcionalidades visuais interativas referentes aos processos de identificação, exploração e documentação de ocorrências. Coletivamente, todas essas funcio-

Figura 42 – A interface gráfica da aplicação desenvolvida: A - abrir um arquivo JSON com os dados da PST; B - exibir o overview dos dados da PST; C - selecionar filtros para aplicar; D - exibir o resultado do processo de edge bundling; E - acessar um subsistema de anotação simples; F - selecionar um nó específico da PST; G - exibir a hierarquia dos nós da PST em uma tree list; H - usar uma escala cromática para indicar relações de hereditariedade, referentes a nós pais e filhos, no mapa; I - exibir ou ocultar marcadores, OD-links, células e grupos; J - definir o número de níveis hierárquicos a serem exibidos nas relações pais/filhos; K - selecionar o modo padrão/visão de satélite do mapa virtual; L - visualizar os marcadores correspondentes aos pais do nó selecionado na tree list; M - exibir um marcador correspondente ao item selecionado na tree list; e N - visualizar os marcadores correspondentes aos filhos do nó selecionado na tree list.



Fonte: o autor.

nalidades suportam a investigação dos diferentes padrões de movimento e de comportamento de indivíduos representados na PST.

Assim, por exemplo, um usuário pode avaliar os nós de uma PST que indicam a casa de uma determinada pessoa e os respectivos marcadores correspondentes no mapa, a fim de verificar seus horários habituais de saída para o trabalho. Então, já identificando os nós referentes ao local de trabalho dessa pessoa, o usuário pode analisar se ela geralmente volta para casa ao meio-dia ou almoça em um restaurante. Além disso, considerando as probabilidades armazenadas na PST, é possível também aprofundar a avaliação, procurando padrões de movimento e, posteriormente, investigando o comportamento geral dessa pessoa. Assim, através da análise de sua rotina diária, pode-se então diferenciar as atividades realizadas em dias úteis, geralmente

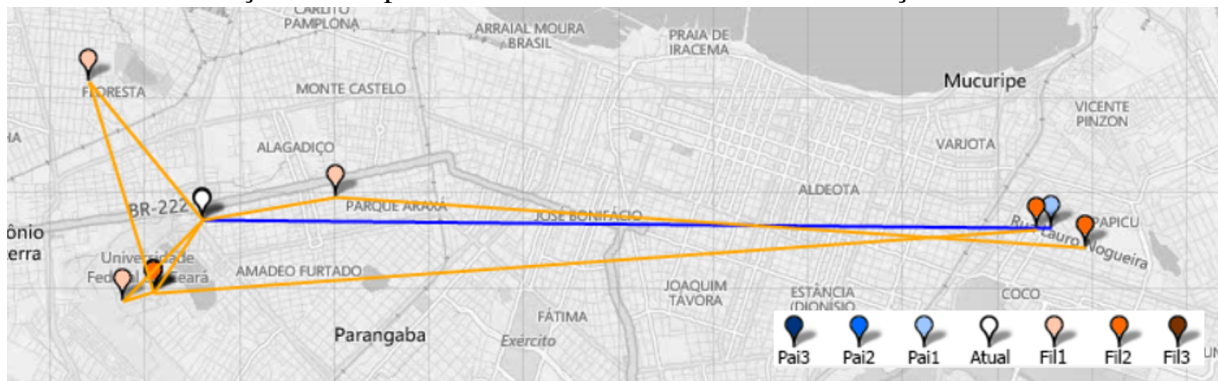


relacionadas aos nós da PST correspondentes ao local de trabalho, daquelas que ocorrem em finais de semana.

As funcionalidades estabelecidas para a aplicação desenvolvida são:

- Pan e zoom. Permitem ao usuário deslocar, aumentar ou diminuir o nível de aproximação do mapa virtual, reduzindo o número de elementos exibidos, buscando minimizar a poluição visual, e concentrando o foco de investigação em áreas geográficas específicas. Uma visualização clara permite realizar uma avaliação mais precisa, com observação mais detalhada dos eventos (ANDRIENKO *et al.*, 2010; SIOW *et al.*, 2018);
- Sincronização de seleções entre tree list e mapa. Dada a organização sincronizada dos dados extraídos da PST na tree list e no mapa virtual, essa funcionalidade permite ao usuário clicar diretamente em um item da tree list (Figura 42F) para selecionar uma ocorrência da PST e exibir automaticamente um marcador (Figura 42M) no mapa virtual, nas coordenadas geográficas correspondentes. Da mesma forma, o processo de seleção inversa também pode ser realizado, clicando-se em um marcador no mapa e automaticamente selecionando o respectivo item na tree list;

Figura 43 – Exemplo de diferenciação de nós pais e filhos, adotando escala cromática para coloração dos respectivos marcadores de acordo com relações de hereditariedade.



Fonte: o autor.

- Visualização de nós pais e filhos. Permite que, quando o usuário selecionar um item na tree list que possua pais (Figura 42L) ou filhos (Figura 42N), seus respectivos marcadores também sejam mostrados no mapa, considerando seus respectivos níveis na PST. Usando *combo boxes* (Figura 42J), o usuário pode definir quantos níveis de ancestralidade ou descendência a aplicação deve exibir. Nesse caso, os marcadores do mapa indicando nós pais e filhos são coloridos usando uma escala cromática complementar, variando de azul escuro para o nó pai de terceiro nível até laranja escuro para o nó filho de terceiro nível

(Figuras 42H e 43);

- Representação de branching-time. Essa funcionalidade considera o conceito de branching time, em conjunto com as relações de hereditariedade, de forma que nós pais e filhos representam, respectivamente, eventos que ocorreram antes e depois do evento associado a um dado nó selecionado na tree list. Assim, os OD-Links representam graficamente as arestas da PST, denotando relações temporais entre ocorrências (Figura 43). Para diferenciar ocorrências anteriores ou posteriores, a visualização de OD-Links segue a mesma escala cromática apresentada anteriormente (Figura 42H);
- Uso de células. Permite que a aplicação desenvolvida exiba, ao redor de um marcador, uma célula (Figura 35c), cujo tom de cor é proporcional à probabilidade de acontecimento presente em seu nó correspondente na PST. Dessa forma, torna-se fácil descobrir a probabilidade de uma pessoa se encontrar em uma determinada região, por exemplo;
- Uso de grupos. Permite que dados possam ser agregados em grupos, alinhados a uma grade visual cujo tamanho pode ser definido pelo usuário. Nesse caso, o mapa mostra quadrados (Figura 35d), os grupos, cujas tonalidades de cor são proporcionais ao número de marcadores presentes em sua área. É possível também estabelecer um limite máximo para a contagem das ocorrências agregadas, controlando o cálculo das tonalidades de cores. Esse recurso, de limitação de contagem, é particularmente útil para enfatizar visualmente presenças mínimas ou mudanças abruptas de quantidades nos grupos (JOHANSSON *et al.*, 2005);
- Exibição de overview. Permite a exibição de overviews no mapa (Figura 42B), que consideram simultaneamente todos os dados de uma PST, podendo-se exibir ou ocultar, coletivamente ou individualmente, marcadores, OD-links, grupos e células (Figura 35);
- Emprego de edge bundling display. Adotando o algoritmo de *Kernel Density Estimation-based Edge Bundling* (HURTER *et al.*, 2014), permite que o usuário combine todos os OD-Links do mapa, a fim de exibir possíveis fluxos principais, de acordo com diferentes níveis de ajuste para a aproximação visual dos OD-Links (Figura 36);
- Identificação semiautomática de POIs. Permite ao usuário empregar o filtro visual f-POIs (Figura 42C) para identificar POIs, acumulando marcadores (Figura 37). Para isso, a aplicação desenvolvida inicialmente cria e armazena automaticamente uma imagem que acumula todos os pixels referentes aos pontos associados às posições de todos os marcadores do mapa virtual; e posteriormente avalia cada pixel dessa imagem, enfatizando

os dados visuais acumulados de acordo com o valor de um widget de controle deslizante (Figura 42C), manipulado pelo usuário;

- Identificação semiautomática de rotas mais frequentes. De forma semelhante à anterior, permite ao usuário aplicar o filtro visual f-Routes (Figura 42C), auxiliando-o na identificação das possíveis rotas mais frequentadas representadas na PST (Figura 38). Para tanto a aplicação gera automaticamente uma imagem que acumula os pixels referentes aos segmentos de reta correspondentes às posições dos OD-Links; e posteriormente analisa cada pixel dessa imagem, enfatizando os dados visuais acumulados de acordo com um limite estabelecido através de um widget de controle deslizante (Figura 42C), manipulado pelo usuário;
- Exibição de eventos em intervalos de tempo. Essa funcionalidade permite ao usuário utilizar o filtro visual f-Times (Figura 39), que apresenta quadrados coloridos referentes à presença de eventos, de acordo com um intervalo de tempo especificado, utilizando-se um widget de controle deslizante (Figura 42C). Os tons das cores dos quadrados obtidos são proporcionais às probabilidades dos nós da PST correspondentes aos eventos considerados no período de tempo determinado. Assim, em última instância, empregando f-Times, o conjunto de quadrados visíveis no mapa virtual sinaliza, então, os locais com as respectivas probabilidades dos eventos que ocorreram no intervalo de tempo especificado.

Para a implementação dessas funcionalidades, sempre que possível, foram adotados procedimentos para a otimização do processamento e do armazenamento em memória necessários. Assim, por exemplo, ao desenvolver as ferramentas referentes às operações de pan e de zoom sobre o mapa virtual, aplicou-se a técnica de particionamento espacial (YOUNG, 2001) para operar somente sobre o respectivo subconjunto de elementos visíveis, minimizando o custo computacional envolvido.

Da mesma forma, para o emprego de edge bundling display foi adotada uma versão do algoritmo de Kernel Density Estimation-based Edge Bundling <sup>4</sup> que podia fazer uso de processamento por GPU <sup>5</sup>, caso o microcomputador que executasse a aplicação dispusesse de tal recurso.

E, no caso dos filtros f-POIs e f-Routes, foram criadas rotinas otimizadas para as operações de geração de imagens e de acesso a pixels das mesmas, baseadas na técnica de

<sup>4</sup> <http://recherche.enac.fr/~hurter/KDEEB.html>

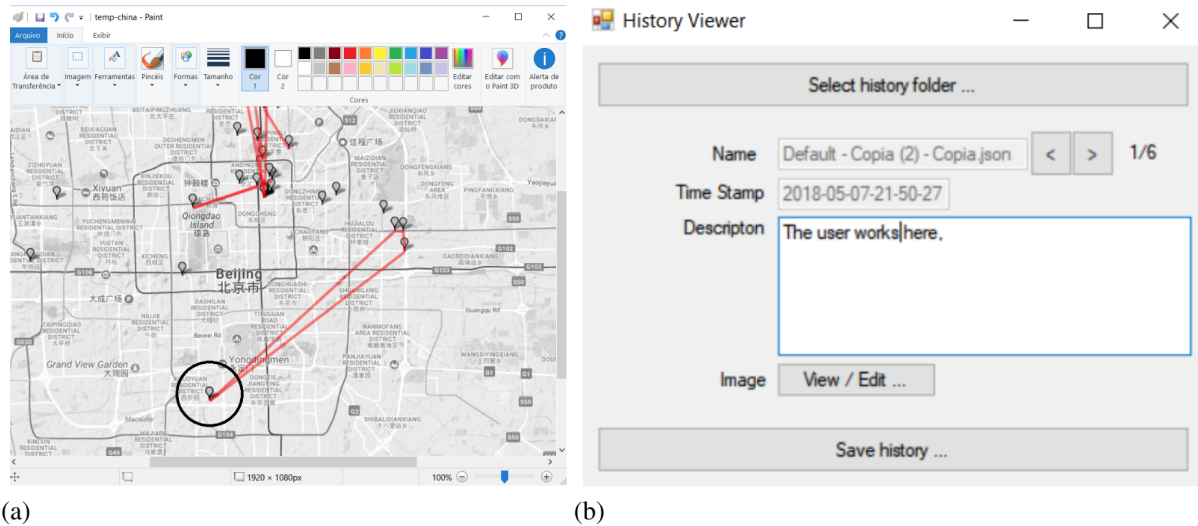
<sup>5</sup> *Graphics Processing Unit*, unidade especializada no processamento de dados gráficos.

LockBitmap<sup>6</sup> (HUANG; DU, 2014).

#### 5.4 Subsistema de Anotação

Além das as funcionalidades descritas anteriormente, a aplicação desenvolvida inclui ainda um subsistema simples para realizar anotações do processo de análise visual (Figura 40), adotando o conceito de *história*. Nesse caso específico, história consiste num modelo padronizado, baseado num *template* específico (Figura 40a), para armazenamento e indexação das anotações realizadas.

Figura 44 – Recursos adicionais para anotação: (a) inclusão de ilustrações esquemáticas, baseada em integração com editor gráfico de terceiros e (b) editor para consulta e alteração de histórias.



(a) Fonte: o autor.

Essas histórias ajudam a registrar as observações realizadas, podendo guardar diretamente textos e URLs com informação relacionadas às ocorrências documentadas ou incluir dados visuais, a partir dos mapas virtuais utilizados. Nesse último caso, para simplificar a edição de ilustrações esquemáticas pelo usuário, a aplicação desenvolvida integra a ferramenta Microsoft Paint<sup>7</sup>, através de uma chamada direta à mesma (Figura 44a).

Adicionalmente, o subsistema de anotação apresenta também recursos para edição das histórias (Figura 44b), permitindo realizar consultas e alterações nos textos e imagens utilizados, mais uma vez amparando os almejados processos de caracterização, busca e comparação de ocorrências.

<sup>6</sup> <https://www.codeproject.com/Tips/240428/Work-with-Bitmaps-Faster-in-Csharp-3>

<sup>7</sup> [https://en.wikipedia.org/wiki/Microsoft\\_Paint](https://en.wikipedia.org/wiki/Microsoft_Paint)

## **5.5 Considerações Finais**

Este capítulo apresentou a versão final da aplicação desenvolvida com base na solução proposta nesta tese. Foi descrita a interface gráfica de tal aplicação, bem como suas principais funcionalidades. Foi destacado ainda o subsistema de anotação, essencial para a devida documentação de ocorrências com base no conceito proposto de histórias.

O próximo capítulo apresenta o processo de validação da solução proposta. Tal solução baseou-se na realização de testes com usuários que, além de fornecerem subsídios para a validação propriamente dita, ainda serviram de base ao processo incremental que estabeleceu a abordagem, a arquitetura e o workflow propostos neste trabalho.

## 6 VALIDAÇÃO DA SOLUÇÃO

O presente capítulo apresenta o processo de validação da abordagem proposta neste trabalho, com base na realização de três testes com usuários. A Seção 6.1 discute os objetivos específicos desses testes e as Seções 6.2, 6.3 e 6.4 apresentam individualmente cada um deles, expondo os perfis dos participantes, as características de cada PST empregada, os resultados obtidos e as respectivas contribuições à solução proposta. A Seção 6.5 realiza uma discussão geral sobre as principais observações realizadas, incluindo validade de respostas obtidas, duração das tarefas, quantidade de histórias registradas e outras constatações julgadas pertinentes. Ao final, são apresentadas algumas considerações finais sobre o Capítulo, na Seção 6.6.

### 6.1 Objetivos dos Testes

Adotando um processo incremental (Figura 4) para aprimorar versões intermediárias da abordagem proposta e do respectivo workflow, bem como validar as configurações finais dos mesmos, foram realizados três testes distintos com usuários para avaliar a eficiência da combinação de análise visual, branching time e sensemaking. Esses testes empregaram diferentes metodologias e modelos de análise (quantitativa ou qualitativa), de acordo com as necessidades específicas dos mesmos.

Deve-se destacar que, embora o emprego de PSTs seja bastante específico, sendo mais voltado à investigação profissional, optou-se por conduzir testes com dois datasets diferentes, explorados por um total de setenta e sete analistas, sendo quatro usuários especialistas na área de análise de dados e setenta e três não especialistas.

Cada um desses três testes apresentou finalidades específicas:

- Teste 1. Identificação do comportamento de dois especialistas em análise de PSTs para teste do protótipo inicial da aplicação; definição de possíveis ferramentas auxiliares de investigação e estabelecimento de uma proposta inicial de workflow, que fosse adequada ao processo de investigação;
- Teste 2. Identificação do comportamento de trinta e quatro não especialistas analisando PSTs, em situações relativamente familiares, para avaliação de uma versão evoluída do protótipo testado anteriormente. Tal protótipo já contava com as respectivas ferramentas auxiliares de investigação definidas nos Capítulos 4 e 5. Esse teste também buscou validar uma proposta inicial de workflow, definida ao fim do primeiro teste, a fim de aprimorá-lo;

- Teste 3. Identificação do comportamento, em situações não familiares, de dois especialistas em análise de dados, não necessariamente em PSTs, e mais trinta e nove não especialistas para a avaliação das versões finais da aplicação desenvolvida e do respectivo workflow final estabelecido.

Nas três próximas seções, cada um dos testes é apresentado individualmente, incluindo a descrição do perfil dos usuários participantes, datasets empregados, respectivos procedimentos executados e principais descobertas realizadas. Deve-se ainda informar que, no texto a seguir, os termos "possível" e "provável" encontram-se empregados de acordo com o grau de certeza das respostas fornecidas pelos participantes dos testes realizados.

## 6.2 Primeiro Teste com Usuários

O primeiro teste com usuários foi realizado junto a especialistas em análise de PSTs, no intuito de compreender melhor como os mesmos trabalhavam. Para tanto, foi criado um protótipo inicial (Figura 45), desenvolvido com base no modo usualmente empregado para identificar ocorrências em PSTs. Assim, considerando as recomendações apresentadas em LEI *et al.* (2013) e em ROCHA *et al.* (2016), e também a partir de conversas iniciais com especialistas na área de análise visual e de predição de movimentos, percebeu-se que, de uma forma geral, o método tradicional para investigação de PSTs consistia em abrir o arquivo que continha a PST e inspecionar cada nível da árvore, tentando isolar dados (coordenadas geográficas, timestamps e probabilidades). Com base na repetição desse processo, eram estabelecidos ROIs e POIs, bem como possíveis relações entre os mesmos.

Figura 45 – Protótipo inicial da aplicação, desenvolvido especificamente para a realização do primeiro teste com usuários: (a) painel de controle e (b) mapa virtual.



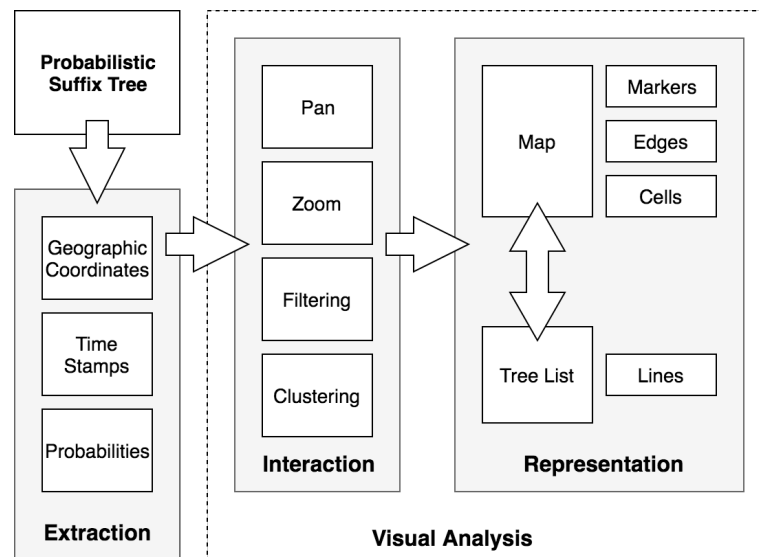
Fonte: o autor.

Esse primeiro protótipo foi desenvolvido com base nas citadas conversas com espe-

cialistas e nas especificidades desse método tradicional, mas já incluindo a ideia de identificação de localizações geográfica num mapa virtual.

A Figura 46 apresenta a arquitetura estabelecida para este primeiro protótipo, e cujos componentes serviram de base inicial à abordagem proposta, apresentada no Capítulo 4 deste trabalho.

Figura 46 – Arquitetura do protótipo empregado no primeiro teste com usuários.



Fonte: o autor.

De uma forma geral, esse protótipo apresentava somente os recursos mais básicos que posteriormente viriam a compor a aplicação final desenvolvida (Figura 45a):

- Abertura de arquivos contendo PSTS;
- Exibição ou ocultação de overviews, marcadores, OD-Links e células em mapas;
- Apresentação de marcadores correspondentes a pais e filhos em diferentes níveis, porém ainda sem qualquer diferenciação cromática de suas relações de hereditariedade (Figura 45b).

Como ainda não eram utilizadas cores para distinguir visualmente pais e filhos, foi adotado um sistema simples de diferenciação diretamente nos respectivos OD-Links: linhas tracejadas para pais, e linhas contínuas para filhos (Figura 45b).

Para a realização do teste, foi utilizado um dataset gerado pelo Projeto Eai?<sup>1</sup>, que fornecia o cardápio diário do Restaurante Universitário da Universidade Federal do Ceará (UFC), através de um aplicativo para dispositivos móveis. Mantido pelos grupos de pesquisa ARiDa<sup>2</sup> e

<sup>1</sup> <https://apkpure.com/br/eai/br.ufc.appei>

<sup>2</sup> <http://www.arida.ufc.br/site>



GREat<sup>3</sup>, do Departamento de Computação da UFC, o Projeto Eaí? recebia dados de posição e tempo oriundos do sistema de GPS dos dispositivos móveis utilizados pelos usuários do aplicativo disponibilizado (ROCHA *et al.*, 2016). Esses dados eram enviados ao servidor do Projeto, que os processava para gerar as respectivas PSTs dos usuários. Por questões de privacidade, as PSTs geradas eram armazenadas em arquivos JSON sem qualquer identificação direta dos usuários originais.

Dado que o método tradicional de avaliação de PSTs exige uma boa experiência no uso desse tipo de árvore, foram selecionados dois especialistas em predição de movimento baseada em PSTs para a realização desse primeiro teste com usuários:

- X1. Especialista do sexo masculino que é estudante de graduação em Ciência da Computação na UFC. Ele é um dos responsáveis pelo gerenciamento de dados oriundos da aplicação móvel utilizada pelo Projeto Eaí?, e já costumava empregar mapas virtuais a fim de localizar ocorrências nas PSTs dos usuários;
- X2. Especialista do sexo masculino que é mestre em Ciência da Computação pela UFC e trabalhou com PSTs. Ele é desenvolvedor sênior numa empresa particular, e usualmente lê arquivos JSON a fim de identificar ocorrências. Porém, durante esse processo, raramente verifica a localização dos eventos detectados em mapas.

Novamente, dada a natureza bastante particular do modo tradicional de investigar PSTs, nesse primeiro teste, resolveu-se preterir métodos quantitativos na avaliação do protótipo a ser testado, uma vez que os mesmos não apresentariam resultados adequados, principalmente devido à pequena quantidade de participantes envolvidos inicialmente (NIELSEN, 1994).

Sendo assim, resolveu-se adotar um método específico de avaliação qualitativa (POLKINGHORNE, 2005), o Thinking Aloud Protocol (TAP)<sup>4</sup>. Tal abordagem avaliativa, além de permitir a observação do uso natural de aplicações, é relativamente fácil de ser implementada (KRUG, 2000), e permite documentar o desempenho dos participantes com base em gravações audiovisuais e nas anotações de um supervisor do teste.

De acordo com o TAP, os participantes do teste deveriam executar determinadas tarefas predefinidas. Para estabelecer tais tarefas, foram então avaliados diversos arquivos JSON fornecidos pelo Projeto Eaí?, dos quais foram selecionadas três PSTs que apresentaram diferentes

<sup>3</sup> <https://www.great.ufc.br>

<sup>4</sup> teste com usuários no qual cada participante deve verbalizar seus pensamentos à medida que realiza operações sobre um software, visando executar tarefas pré-estabelecidas. Um ou mais supervisores podem acompanhar diretamente o participante e seu diálogo, que é geralmente gravado, encorajando-o a discutir particularidades das operações realizadas (KRUG, 2000; JASPERS *et al.*, 2004).

características, discutidas mais adiante, consideradas relevantes para a realização do teste formal proposto.

As PSTs selecionadas referiam-se especificamente a dados coletados nos dias úteis do período de 27/08/2015 a 01/10/2015, durante o qual, por conta de uma greve na Universidade, várias aulas foram suspensas. Isso foi particularmente interessante para o teste por modificar substancialmente a rotina do local, a qual os especialistas X1 e X2 já deviam estar bastante familiarizados. Isso permitiria, então, a identificação de possíveis novos padrões de movimento nas três PSTs selecionadas, tornando as tarefas estabelecidas para o teste mais estimulantes e desafiadoras.

Para a devida realização do teste, os especialistas foram contatados separadamente para duas etapas específicas de uso da aplicação. A primeira, de sensibilização, se deu através da disponibilização do protótipo da aplicação a partir de um link online para seu respectivo código-fonte. Optou-se por fornecer o código-fonte para que os próprios especialistas pudessem avaliar mais concretamente as soluções ali propostas.

Os especialistas realizaram o *download* do código fornecido, compilaram-no, e geraram uma cópia da aplicação que foi utilizada pelo prazo de uma semana. Ao final desse período, cada especialista foi solicitado por *e-mail* a descrever as principais facilidades e dificuldades no uso da aplicação, bem como a apresentar seus prós e contras quando comparada ao método tradicional de análise de PSTs.

As respostas referente a essa primeira fase de uso da aplicação, recebidas por e-mail, demonstraram que os participantes realmente utilizaram a aplicação para proceder a análise visual dos arquivos JSON de que dispunham, inclusive comprovando ou refutando algumas das hipóteses que eles próprios haviam formulado a partir de avaliações anteriores das respectivas PSTs. Durante essa etapa do teste, os especialistas consideraram a aplicação como sendo uma ferramenta de uso simples, com uma interface visual agradável e que fornecia um meio adequado à análise pretendida. Tais afirmações confirmaram o engajamento e o interesse dos especialistas em participar do teste formal, seguindo o TAP.

Para a segunda fase de uso da aplicação, o teste propriamente dito, os especialistas foram novamente contatados após a avaliação de suas resposta por e-mail. Duas semanas depois, eles foram visitados separadamente em seus respectivos locais de trabalho. Na ocasião, receberam para uso um notebook (Apple MacBook Pro 17" Early 2011 com processador Intel i7, 8 GB de RAM, 240 GB de SSD, executando Microsoft Windows 10 e equipado com mouse

externo) contendo o protótipo a ser utilizado e as três PSTs selecionadas para análise.

Para evitar problemas na avaliação, como conversas casuais durante o teste, o supervisor, que é o autor da presente tese, não conhecia pessoalmente nenhum dos dois especialistas participantes (HOLZINGER, 2005).

Seguindo as recomendações do TAP, durante o teste os participantes foram instruídos a seguir um roteiro de tarefas a executar, as quais continham perguntas a serem respondidas enquanto usavam o protótipo. Tal roteiro foi definido com base nas observações realizadas inicialmente sobre as três PSTs selecionadas. Assim, as quatro perguntas seguintes foram estabelecidas para serem respondidas pelos especialistas:

1. Onde o usuário reside?
2. Qual curso o usuário frequenta?
3. Onde o usuário trabalha?
4. Quando o usuário vai ao Restaurante Universitário?

Todas as perguntas foram elaboradas seguindo as orientações de ANDRIENKO *et al.* (2008) sobre como formular questões para a avaliação de dados espaçotemporais, considerando particularmente os conceitos estabelecidos de restrição (*constraint*, a informação já conhecida) e de alvo (*target*, a informação aguardada como resposta).

Nenhuma observação prévia sobre os três arquivos selecionados, incluindo o período representado, foi fornecida aos participantes do teste. E, devido à privacidade de dados e às características probabilísticas das PSTs, o teste não pôde se concentrar especificamente na resolução formal de problemas. Assim, o principal objetivo das quatro perguntas estabelecidas foi o de ajudar os participantes a elaborar novas questões (hipóteses) e a propor soluções (provar ou refutar hipóteses), de acordo com os princípios de análise visual e de sensemaking, discutidos no Capítulo 3 deste trabalho. Dessa forma, os especialistas puderam operar os dados disponíveis para sugerir novas possibilidades de interpretação da informação representada, então extrapolando os contextos originais predefinidos (MOORE, 2017; WISE *et al.*, 2016).

O teste propriamente dito iniciou com as assinaturas dos termos de consentimento livre e esclarecido (Apêndice A) e com explicações gerais sobre os procedimentos do TAP. A partir desse momento, o supervisor passou a gravar o som da sessão, utilizando um smartphone, e a tomar notas. Os participantes usaram o protótipo livremente por cinco minutos, para que pudessem recordar as principais funções do mesmo. Em seguida, após a apresentação das quatro perguntas a serem respondidas, os especialistas passaram a investigar cada uma das três PSTs

em sequência, podendo responder cada pergunta na ordem e do modo que desejassem.

Após a finalização do processo de resposta às perguntas, o supervisor discutiu algumas observações específicas realizadas pelos especialistas, a fim de compreender melhor o modo como as mesmas se deram. Além disso, esse momento também foi aproveitado para debater ideias e conceitos pertinentes ao entendimento geral do uso de PSTs e possíveis aprimoramentos para o protótipo então utilizado.

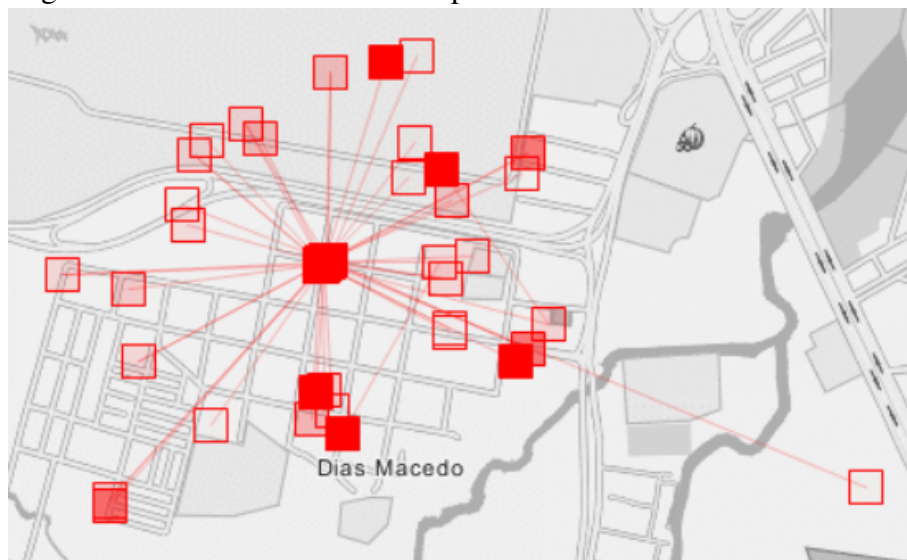
Posteriormente, o áudio das sessões foi transcrito, sendo acrescentadas as notas tomadas pelo supervisor, buscando enfatizar e embasar, quando necessário, as observações realizadas pelos especialistas.

A seguir são apresentadas as três PSTs selecionadas para o teste, suas principais características e as observações mais relevantes realizadas pelos especialistas, com base na análise da transcrição dos diálogos e nas referidas notas tomadas.

### 6.2.1 *Análise da Primeira PST*

A primeira PST refere-se a um usuário que não foi à Universidade durante o período considerado (Figura 47). Foi observado que os especialistas, no início, tentaram adotar a metodologia tradicional, já discutida, para responder às perguntas, concentrando-se principalmente em utilizar a tree list da aplicação. Posteriormente, no entanto, ao realmente compreenderem os recursos disponíveis no protótipo, eles passaram a trabalhar de maneira mais eficiente.

Figura 47 – Overview da PST do primeiro usuário avaliado.



Fonte: o autor.

Assim, considerando o mapa virtual, com base no overview dos dados da PST, particularmente a exibição de todos os seus OD-Links simultaneamente, os dois especialistas identificaram o provável local de residência da pessoa. Suas análises consistiram principalmente em observar os pontos de origem mais importantes e as frequências de movimento mais constantes, procurando as maiores quantidades de segmentos de reta sobrepostos. Essas interseções definiram POIs e ROIs que, posteriormente, foram confirmados empregando células para considerar as probabilidades envolvidas. Sempre que necessário, o tamanho das células foi alterado para dar um destaque visual às regiões consideradas, comparando-as aos POIs e às ROIs em avaliação no momento (Figura 48).

Após confirmarem POIs e ROIs no mapa virtual, os especialistas iniciaram uma exploração mais detalhada, então voltando a interagir com a tree list, observando o tempo e as probabilidades de presença e visualizando marcadores referentes a itens, relacionando-os aos respectivos locais no mapa virtual. O processo inverso também se tornou comum, isto é, aquele em que os especialistas clicam sobre marcadores no mapa, principalmente ao observarem overviews, para terem acesso aos respectivos itens da tree list.

Em seguida, a investigação controlada dos nós pais e filhos, em vários níveis na tree list, ajudou os participantes do teste a detectarem padrões de movimento, acompanhando os respectivos marcadores e OD-Links no mapa virtual.

Dessa forma, por exemplo, X1 identificou que o usuário permanecia em um determinado lugar desde tarde da noite até o amanhecer, e concluiu que esse local tinha uma alta probabilidade de ser sua casa. X2 também usou os mesmos recursos para reconhecer o provável local de residência do usuário. No caso específico desse primeiro usuário, dada a distribuição radial de sua PST (SCHULZ, 2011), pôde-se perceber facilmente o local provável de sua residência (Figura 47a). E, consultando os tempos representados na tree list, ambos os especialistas também identificaram que, de uma forma geral, o usuário passava muito tempo em casa, porém deslocando-se relativamente muito a noite. Como o respectivo overview não apresentava marcadores na região da Universidade, tanto X1 quanto X2 concluíram rapidamente que o usuário não frequentou nenhum curso ou o Restaurante Universitário.

Além disso, X1 notou que o usuário frequentemente ia a uma igreja e a alguns outros lugares em sua vizinhança, possivelmente residências de amigos e alguns locais de compras (Figura 48), muitas vezes a noite ou mesmo de madrugada. Vale ressaltar que, como já estava acostumado, X1 foi capaz de detectar a igreja e os demais locais utilizando serviços online. Em

Figura 48 – Descoberta de um padrão de movimento com a aplicação desenvolvida: (a) uso de overview para identificar os locais mais frequentes (marcadores brancos e células verdes) e seus respectivos relacionamentos (OD-Links verdes) e (b) ampliação dessa região de interesse para realização de exploração mais detalhada da área tracejada em (a).



Fonte: o autor.

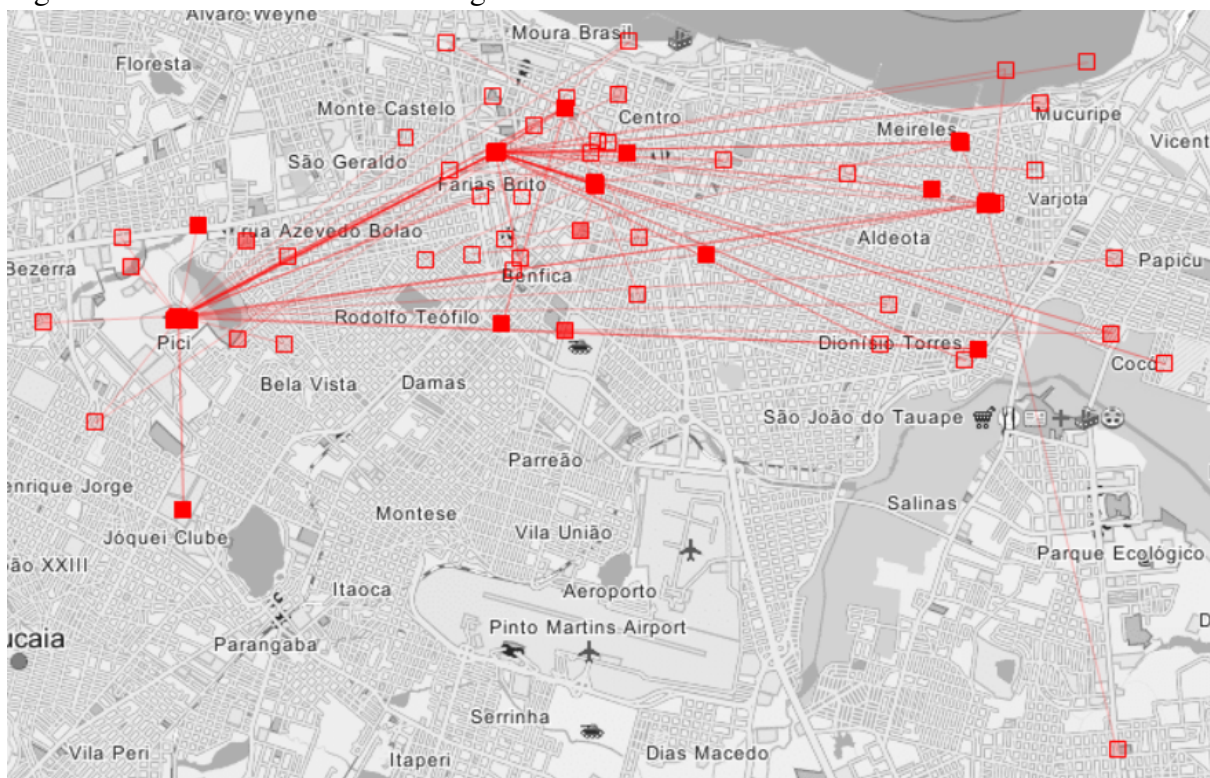
particular, tal observação criou o precedente para incluir o acesso à internet como um possível recurso, em paralelo, para o estabelecimento da abordagem proposta nesta tese.

X1 presumiu que a pessoa não trabalhava, pois não encontrou um padrão de movimento significativo no mapa virtual. Já X2, por outro lado, analisou a tree list mais profundamente e detectou que a pessoa estava constantemente indo a um local específico. Ele, então, supôs que aquele seria o possível local de trabalho do usuário. No entanto, ao não concentrar sua atenção no mapa virtual, X2 acabou não percebendo que a referida área era a igreja detectada por X1. Devido à privacidade dos dados, não é possível saber se a pessoa trabalhou realmente na igreja ou apenas a frequentou durante o período. De todo modo, nota-se que os especialistas distinguiram com precisão possíveis padrões de movimento, conseguindo fazer inferências válidas sobre os prováveis comportamentos da pessoa representada.

### 6.2.2 *Análise da Segunda PST*

Na análise da segunda PST (Figura 49), os especialistas preferiram já iniciar com a avaliação do overview. Depois, interagindo com a tree list e com o mapa virtual simultaneamente, eles rapidamente identificaram o provável local da residência do usuário representado, basicamente repetindo o processo estabelecido na análise da primeira PST.

Figura 49 – Overview da PST do segundo usuário avaliado.



Fonte: o autor.

X1 afirmou que o segundo usuário frequentou a Universidade durante o período considerado, provavelmente estudando Biologia à noite. Tal informação, no entanto, inicialmente parecia não ser provável, pois o curso indicado só é ofertado no período diurno. No entanto, devido à greve do corpo docente, é realmente possível que a pessoa tenha ido à Universidade à noite. X2, por outro lado, disse que o indivíduo provavelmente estudou Ciência da Computação ou outro curso oferecido no Centro de Ciências, que, no caso da UFC, inclui o curso de Biologia, cujas instalações ficam próximas às do Curso de Ciência da Computação. No entanto, X2 não indicou a que horas a pessoa geralmente parecia ir à Universidade.

X1 notou um padrão de movimento repetitivo entre 14:00 e 17:00, conectando uma região específica num bairro nobre de Fortaleza com a Universidade. Então, ele informou que nessa região poderia estar um provável local de trabalho do indivíduo representado. X2 também encontrou o mesmo padrão, mas construiu uma narrativa mais completa, afirmando que o usuário ia de casa ou da Universidade para um possível local de trabalho, o mesmo indicado por X1, chegando lá ao final da manhã, possivelmente para o almoço, continuando até o final da tarde, e depois geralmente retornando direto para casa.

De acordo tanto com X1 quanto com X2, a pessoa não frequentou o Restaurante Universitário durante esse período, uma vez que eles não identificaram nenhum padrão de

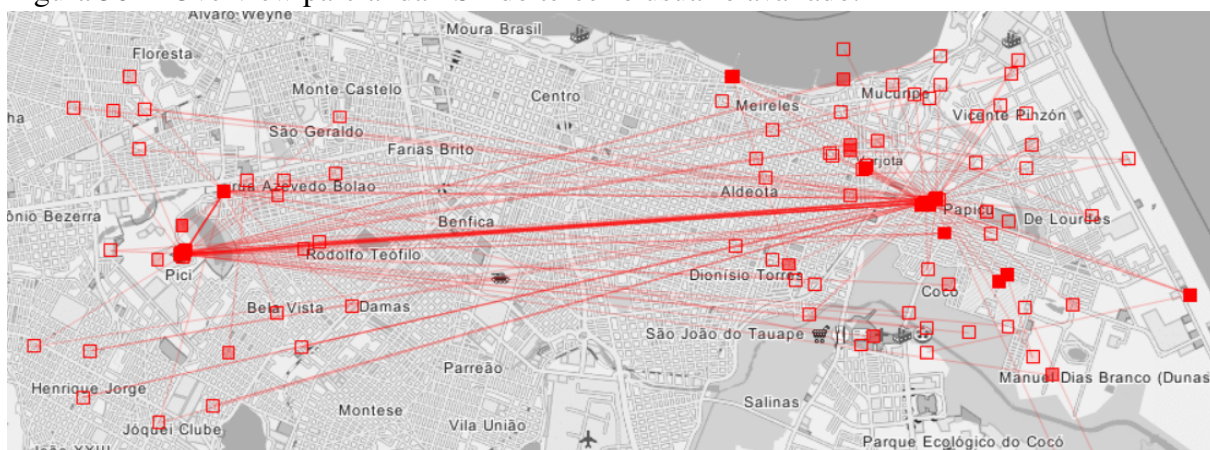
movimento no mapa que levasse a tal conclusão.

Ambos os especialistas ainda detectaram alguns prováveis erros de GPS, dada a presença de marcadores no mar (Figura 49). Isso provavelmente ocorreu porque optou-se por não realizar qualquer tratamento nos dados a serem visualizados, a fim de evitar intervenções nas possíveis informações originais.

### 6.2.3 *Análise da Terceira PST*

O terceiro usuário apresentou uma PST (Figura 50) maior e mais complexa do que a dos outros dois anteriores. Na análise dessa árvore, mais uma vez, os especialistas partiram do overview e, posteriormente, utilizaram a relação entre os itens da tree list e o mapa virtual para distinguir POIs e ROIs, geralmente confirmados utilizando células.

Figura 50 – Overview parcial da PST do terceiro usuário avaliado.



Fonte: o autor.

X1 indicou a provável residência do usuário, chegando inclusive a apontar paradas de ônibus frequentemente usadas para viagens de ida e de volta para a Universidade, devidamente identificadas mais uma vez empregando os serviços online. X2 identificou a mesma provável residência e também notou o movimento recorrente do usuário para os mesmos dois pontos apontados por X1, um perto da provável residência e outro próximo à Universidade. No entanto, como X2 não se concentrou muito em procurar tais pontos no mapa virtual, nem paralelamente acessou serviços online, não percebeu que eram paradas de ônibus, e acabou informando que poderiam ser casas de familiares ou de amigos do usuário representado na PST.

Com relação especificamente ao possível curso o qual o usuário estudava, X1 inferiu que a pessoa provavelmente estudava Ciência da Computação no período da tarde. X2 chegou à



mesma conclusão e identificou ainda um padrão de movimento muito bem definido, em que a pessoa representada se deslocava, no início da noite, para um prédio próximo à Universidade duas vezes por semana, passando cerca de duas horas no local.

X1 não conseguiu detectar um local de trabalho para o usuário, mas X2 indicou que ele provavelmente trabalhava no prédio próximo à Universidade, detectado anteriormente.

Como X1 não observou um padrão de movimento distinto, o mesmo se restringiu a informar que o usuário provavelmente não frequentou o Restaurante Universitário. Já X2, por outro lado, foi além e intensificou as buscas na tree list, concluindo que a pessoa realmente não havia frequentado o Restaurante Universitário. No entanto, X2 indicou que o usuário pode ter almoçado alguns dias numa determinada lanchonete no campus da Universidade, geralmente chegando à mesma entre 11:45 e 12:00.

Deve-se destacar ainda que, algumas vezes, os especialistas chegaram a algumas conclusões inusitadas como, nas palavras de X2, "o usuário parece ser uma pessoa bastante sociável, devido à presença de muitos nós com pouca convergência".

Figura 51 – Overview total da PST do terceiro usuário avaliado.



Fonte: o autor.

E curiosamente X1 identificou que provavelmente já havia lidado com esse mesmo arquivo JSON, devido à observação de uma viagem realizada pelo usuário a Maceió, Estado de Alagoas, no início do período considerado (Figura 51). X2 não conseguiu identificar tal

viagem, provavelmente pela mesma só ser perceptível quando o zoom do mapa virtual encontra-se ajustado para a visualização de uma grande área, ou talvez pelo fato de que a tree list, elemento no qual esse especialista geralmente se concentrava, não identificar diretamente dados geográficos.

Novamente os dois especialistas encontraram alguns prováveis erros de captura de GPS, mais uma vez dada a presença de marcadores no mar (Figura 50).

#### **6.2.4 Avaliação de Resultados**

Os resultados desse primeiro teste foram avaliados seguindo a metodologia de análise de conteúdo<sup>5</sup>, organizando sistematicamente os dados em um formato estruturado: tempo necessário para executar cada tarefa, identificação de eventos, comportamentos significativos dos especialistas, afirmações realizadas e assim por diante.

Assim, qualitativamente, os especialistas consideraram o protótipo desenvolvido satisfatório para análise visual; pois, em geral, eles puderam encontrar possíveis padrões de movimento e de comportamento para os três usuários representados. Isso se deu, principalmente, empregando o recurso overview e usando as associações entre a tree list e o mapa virtual.

Nota-se que os dois especialistas começaram a analisar as PSTs da mesma forma, adotando a metodologia tradicional. Porém, isso mudou ao passarem a considerar o mapa virtual. Ao perceberem os recursos disponíveis, e resolverem os primeiros problemas, assumiram uma nova metodologia de trabalho. Dessa forma, primeiro, para iniciar a investigação, eles usaram overviews para identificar POIs e ROIs. Em seguida, eles interagiram com os itens da tree list para explorar os dados em detalhes, ocultando ou mostrando simultaneamente marcadores e OD-Links no mapa virtual, facilitando a decisão sobre quais regiões deveriam ser observadas mais minuciosamente.

Sempre que os especialistas precisavam considerar as probabilidades de presença, eles usaram células, alterando seus tamanhos quando necessário, para tentar confirmar regiões mais específicas. Posteriormente, os especialistas passaram a analisar relações entre os itens da tree list e o mapa virtual. E para uma análise mais profunda dos dados, eles se concentraram na avaliação das relações de hereditariedade entre os nós da PST avaliada.

Assim, analisando as observações desse primeiro teste, percebe-se que os especialistas conseguiram investigar as PSTs, encontrando respostas para as perguntas formuladas e indo

---

<sup>5</sup> oriunda das ciências sociais e humanas, para estudo de aspectos comunicacionais de conteúdos específicos, considerando a frequência de ocorrência de termos, incluindo seus possíveis equivalentes, em construções textuais (LIAMPUTTONG; EZZY, 2005; TAYLOR *et al.*, 2015).

além, buscando soluções para seus próprios questionamentos.

Pode-se notar, no entanto, que ambos os especialistas assumiram perfis distintos de análise: X1 concentrou-se mais no mapa virtual da aplicação e no acesso a serviços online, em paralelo, enquanto X2 focou mais na interação direta com a tree list. Assim, por exemplo, X1 identificou as visitas da primeira pessoa a uma igreja e detectou as paradas de ônibus que o terceiro indivíduo provavelmente usou. Da mesma forma, X2 sugeriu almoços esporádicos numa lanchonete para o terceiro usuário e ainda indicou padrões de movimento válidos, alguns bastante complexos, para discutir os possíveis locais de trabalho para todos os usuários, apesar da questão aqui levantada com relação à igreja para o primeiro usuário.

Dessa forma, pode-se concluir que esse perfil distinto dos especialistas acabou influenciando como as propostas e descobertas foram realizadas, respectivamente privilegiando o uso do mapa virtual ou da tree list.

Maiores detalhes sobre esse primeiro teste com usuários podem ser encontrados ainda em LEITE *et al.* (2017).

### **6.2.5 Contribuições à Solução Proposta**

Deve-se destacar que os resultados desse primeiro teste definiram a base para o estabelecimento de todos os recursos da solução proposta neste trabalho, particularmente a posterior criação das ferramentas semiautomáticas, os filtros f-POIs, f-Routes e f-Times. Note-se que, em última instância, tais ferramentas tentam replicar parte do comportamento de análise dos especialistas desse primeiro teste, buscando localizar POIs e ROIs com base na aglomeração gradativa respectivamente de marcadores e de OD-Links no mapa virtual, ou seja, considerando o acúmulo de seus respectivos pixels.

Além disso, a observação desse primeiro teste acabou ajudando a definir também uma primeira versão do workflow proposto. Para tanto, foi inicialmente avaliado todo o processo aqui descrito, observando o comportamento e o modo como os especialistas lidaram com os recursos presentes nesse primeiro protótipo testado. Posteriormente, o próprio autor deste trabalho realizou ainda diversos experimentos, analisando várias outras PSTs, adotando esse mesmo método observado. Assim, em seguida, a arquitetura do primeiro protótipo foi estendida para contemplar um novo conjunto de possíveis recursos. Tais novos recursos foram devidamente implementados, gerando uma nova aplicação, que foi empregada no segundo teste com usuários, apresentado na seção a seguir.

### 6.3 Segundo Teste com Usuários

Uma vez estabelecidas as versões iniciais da abordagem proposta para análise visual de PSTs e de seu respectivo workflow, elaborados com base nos resultados do primeiro teste, uma versão evoluída do protótipo foi desenvolvida para a validação dos mesmos. Em comparação ao protótipo empregado no primeiro teste, foram incluídas as seguintes funcionalidades nessa nova aplicação:

- Adoção de escala cromática para diferenciação de pais e filhos em diversos níveis;
- Visualização de grupos para contagem de ocorrências em regiões específicas, já com a possibilidade de estabelecimento de limite para tal contagem;
- Seleção de visualização de mapa no modo padrão ou em visão de satélite;
- Acesso opcional às ferramentas semiautomáticas de análise (f-POIs, f-Routes, f-Times e Bundles).

Para avaliar a eficiência dessa proposta intermediária, realizou-se um novo teste com usuários, agora voltado especificamente a analistas não especialistas. Optou-se, assim, por comparar as respostas obtidas nesse segundo teste com aquelas obtidas no primeiro, realizado com especialistas, buscando analisar possíveis similaridades.

Dada essa necessidade de comparação entre os testes, dessa vez preferiu-se trabalhar simultaneamente com avaliações quantitativa e qualitativa dos resultados. Sendo assim, a metodologia adotada consistiu num experimento de uso prático da aplicação por dois grupos distintos em laboratório de informática (*lab test*) (DAVIS, 1985; BERNSEN *et al.*, 2012), com posterior resposta a um questionário em papel (ABRAS *et al.*, 2004; NIELSEN, 1994; KRUG, 2000) (Apêndice B). Preferiu-se o uso de questionário eletrônico devido à necessidade de assinatura do termo de consentimento livre e esclarecido e à garantia de entrega das respostas.

Participaram do experimento um total de 34 alunos do Curso de Bacharelado em Sistemas e Mídias Digitais da UFC, 26 dos quais eram do sexo masculino e 8, do sexo feminino, e que tinham idades entre 17 e 22 anos. Esses alunos foram divididos em dois grupos: um grupo de controle, com 17 participantes, que utilizou a nova aplicação desenvolvida, porém sem acesso às suas funções semiautomáticas para identificação de POIs e ROIs; e um grupo experimental, também com 17 participantes, que teve acesso irrestrito a todos os recursos da aplicação, incluindo as funções semiautomáticas. Note-se que, dado o caráter aplicado do trabalho, optou-se por empregar o grupo de controle especificamente para verificar se o uso dos semiautomatismos poderia influenciar o modo como o processo de análise visual seria

desenvolvido.

Para cada grupo, o teste foi realizado durante uma aula, com duração total de 1 hora e 40 minutos, sobre análise visual da disciplina de Matemática Aplicada a Multimídia I, do Curso de Bacharelado em Sistemas e Mídias Digitais da UFC. Devido ao seu caráter multidisciplinar, essa disciplina é geralmente ministrada por dois professores com perfis diferentes (programação e design) simultaneamente, dos quais um é o autor da presente tese.

Sempre que possível, essa disciplina emprega sensemaking no intuito de simplificar a compreensão de conceitos abstratos, particularmente algoritmos, estruturas de dados e outros assuntos relacionados à geração de interfaces gráficas, jogos eletrônicos, sistemas de visualização de dados e afins.

O teste foi realizado em laboratório de informática com acesso à internet e contando com microcomputadores padrão IBM PC, executando Microsoft Windows 10, com processadores Intel i5 3470 de 3.2 GHz, 8 GB de memória RAM e monitores de 23". Dada a necessidade de comparação de resultados, os mesmos três arquivos JSON empregados no primeiro teste foram utilizados nesse experimento, o qual foi dividido em três etapas sequenciais: pré-teste, teste e pós-teste. A execução dessas três etapas encontra-se descrita a seguir.

### **6.3.1 Pré-Teste**

Antes da realização do teste propriamente dito, durante uma apresentação de quinze minutos, os professores da disciplina discutiram o conceito de PST, expondo os dados que as mesmas armazenam, os significados de seus nós e arestas e seus principais usos. Após essa apresentação, os alunos responderam a uma avaliação, com notas de 0,0 a 10,0, com as seguintes perguntas:

1. O que são PSTs?
2. Quais dados uma PST armazena?
3. O que os nós e as arestas de uma PST representam?
4. E para que as PSTs são utilizadas?

O objetivo dessa etapa foi contextualizar o conceito de PSTs para os participantes não especialistas, avaliando seu conhecimento inicial do assunto com base nas quatro perguntas anteriores.

### 6.3.2 *Teste*

Antes de iniciar o teste propriamente dito, por cerca de cinco minutos foi demonstrado o uso do segundo protótipo desenvolvido, empregando uma PST genérica e seguindo o workflow então estabelecido com base na observação do primeiro teste. Também foi brevemente apresentada a origem das três PSTs que seriam analisadas, porém novamente sem a indicação do período de greve. Em seguida, os participantes assinaram os termos de consentimento livre e esclarecido, presentes num questionário em papel (Apêndice B), que também foi utilizado para registrar as respostas. Feito isso, os alunos passaram a analisar as PSTs fornecidas, na ordem que desejassem, a fim de responder às mesmas perguntas do teste anterior: 1. Onde o usuário reside? 2. Qual curso o usuário frequenta? 3. Onde o usuário trabalha? e 4. Quando o usuário vai ao Restaurante Universitário?

Apesar da breve apresentação do workflow, durante a demonstração de uso do segundo protótipo, os participantes foram instruídos a utilizarem a aplicação como quisessem. Isso se mostrou particularmente importante, pois assim poderia ser avaliado posteriormente quão natural e intuitivo o workflow se mostrava. Dado que não houve qualquer formalismo em sua apresentação, caso os participantes do teste replicassem a sequência das etapas de uso da demonstração, provavelmente eles o teriam compreendido devidamente.

Novamente, devido à política de privacidade do Projeto Eaí? e às características probabilísticas das PSTs, o teste não se preocupou com a resolução formal das questões elencadas. O intuito principal foi agora avaliar se os participantes seriam capazes de apresentar respostas válidas<sup>6</sup> para a posterior comparação com os resultados dos especialistas do primeiro teste.

Um importante diferencial em relação ao primeiro teste é que, dado o acesso à internet do laboratório utilizado, os participantes puderam utilizar serviços online, como visualização de mapas ou consultas a endereços, dentre outros, paralelamente à aplicação. Os alunos responderam às quatro perguntas no questionário em papel, sempre explicitando e justificando suas observações, incluindo hipóteses estabelecidas e como as mesmas foram avaliadas. Além disso, os participantes anotaram, nos próprios questionários, suas observações gerais sobre o uso da aplicação e os possíveis serviços online utilizados.

---

<sup>6</sup> no caso, respostas consistentes, adequadas aos dados das PSTs analisadas e devidamente justificadas.

### 6.3.3 Pós-Teste

Após a realização do teste, os participantes responderam novamente à mesma avaliação apresentada na etapa de pré-teste, para fins de análise da eficiência do uso do protótipo evoluído. Tal etapa foi particularmente importante para detectar possíveis diferenças na recepção dos conceitos pelos grupos experimental e de controle.

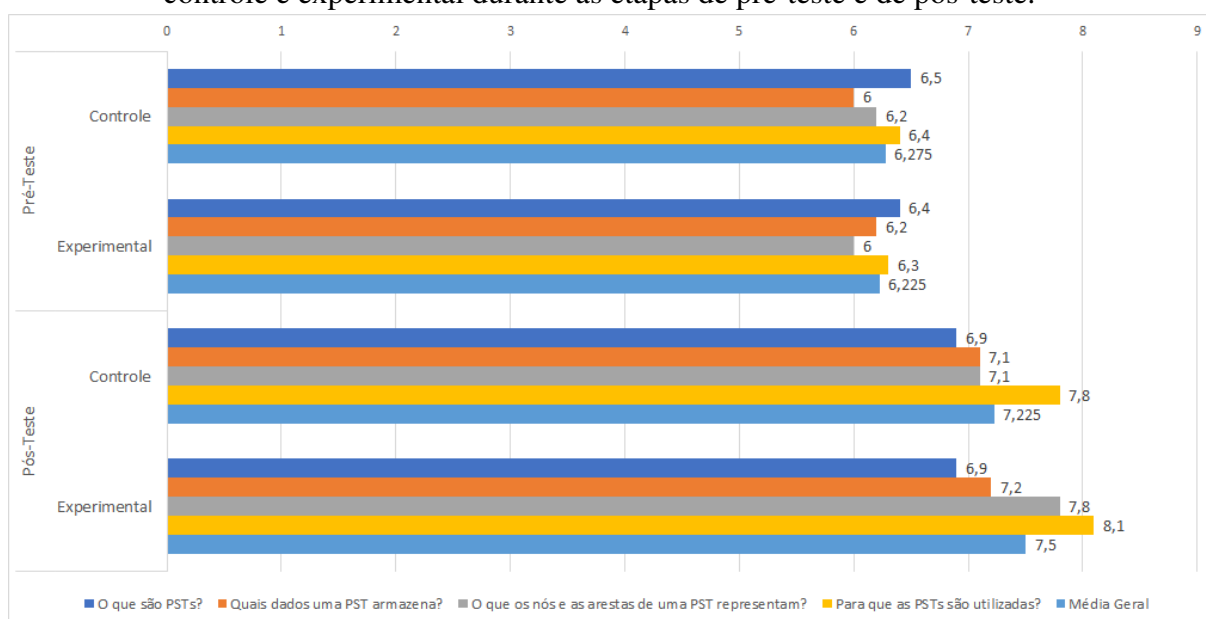
### 6.3.4 Análise de Resultados

A seguir são apresentadas análises quantitativa e qualitativa referentes aos resultados do segundo teste com usuários realizado.

#### 6.3.4.1 Análise Quantitativa

Do ponto de vista quantitativo, observa-se que as notas das avaliações na etapa de pós-teste apresentaram incremento em relação às de pré-teste (Figura 52), com aumento de 0,95 na média geral para o grupo de controle e 1,275 para o experimental, com um leve ganho para o último. Tais acréscimos já eram esperados, uma vez que todos os participantes praticaram o uso das PST através do protótipo utilizado, o que provavelmente influenciou a resposta da segunda aplicação da avaliação, na fase de pós-teste (BROWN *et al.*, 1989). Isso pode ser particularmente

Figura 52 – Notas médias (entre 0,0 e 10,0) obtidas nas avaliações realizadas pelos grupos de controle e experimental durante as etapas de pré-teste e de pós-teste.

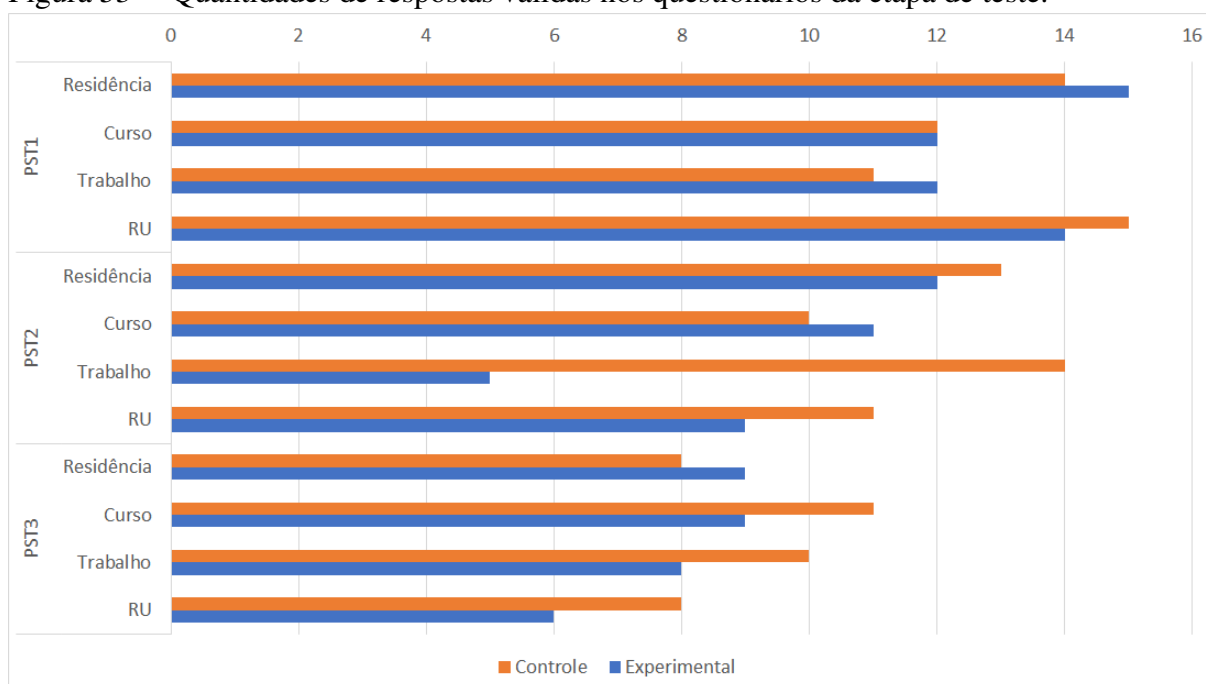


Fonte: o autor.

notado na questão sobre para que as PSTs são utilizadas, onde houve um ganho de 1,4 pontos para o grupo de controle e 1,8 para o experimental.

Percebe-se também que o grupo experimental obteve um aumento substancial (1,8 pontos) em comparação com o grupo de controle (0,9 pontos) na questão sobre o que representam nós e arestas de PSTs. Tal diferença pode advir do acesso diferenciado dos grupos às funções semiautomáticas do protótipo.

Figura 53 – Quantidades de respostas válidas nos questionários da etapa de teste.



Fonte: o autor.

Com relação à quantidade total de respostas válidas (Figura 53), nota-se que quase não houve diferenças significativas entre os grupos nos questionários da etapa de teste. A única exceção, porém, refere-se ao local de trabalho do usuário da segunda PST avaliada: das 17 respostas válidas possíveis em cada grupo (quantidade total de participantes em cada grupo), o grupo de controle forneceu somente 5 enquanto o grupo experimental, 14. Mais uma vez as ferramentas semiautomáticas podem ter influenciado no desempenho dos grupos.

#### 6.3.4.2 Análise Qualitativa

Avaliando as respostas e anotações dos questionários da etapa de teste, pode-se destacar que, independente de pertencerem ao grupo de controle ou ao grupo experimental, os participantes estabeleceram hipóteses bem fundamentadas a fim de confirmá-las ou refutá-las



posteriormente. Por exemplo, apesar de não terem sido perguntados, diversos participantes indicaram que o primeiro e o segundo usuários das PSTs analisadas frequentaram locais de culto religioso, dos quais um é a igreja identificada no primeiro teste (Figura 48). Utilizando os serviços online, foram inclusive apontados alguns dos prováveis locais mais frequentados: uma padaria para o primeiro usuário, um shopping para o segundo e uma academia de ginástica para o terceiro.

Os participantes desse segundo teste também conseguiram indicar as possíveis paradas de ônibus geralmente utilizadas pelo terceiro usuário; e o fato de que o primeiro usuário não frequentou a Universidade durante o período avaliado, passando bastante tempo em casa e deslocando-se várias vezes à noite ou de madrugada, fatos esses também observados pelos especialistas no primeiro teste.

Houve observações ainda mais específicas, como a de que o terceiro usuário poderia ser um aluno novato na cidade. Tal indicação foi baseada na identificação da viagem ao Estado de Alagoas logo no início do período considerado, também feita por um dos especialistas do primeiro teste, bem como na constatação de que esse usuário frequentou diversos pontos turísticos próximos à orla marítima, inclusive à noite. Esse último detalhe não foi comentado pelos especialistas no primeiro teste.

Alguns participantes observaram ainda a existência de possíveis erros de GPS nas PSTs do segundo e do terceiro usuário (Figuras 47 e 49), corroborando a ideia dos especialistas do primeiro teste, pois indicaram que havia pontos de presença em pleno mar, a cerca de cem metros da costa. No entanto, um participante foi além da simples observação; pois, devido à sua experiência própria com aquele local específico, indicou que tal área não era adequada à prática de surfe ou de outros esportes aquáticos. Assim, de fato, deve ter havido algum tipo de inconsistência nos dados representados. Vale destacar que tal tipo de avaliação particular, que considera as experiências prévias dos analistas, é um diferencial bastante importante da análise visual em comparação a outras técnicas de avaliação, como visualização científica ou aquelas baseadas em *machine learning*, por exemplo.

No geral, os participantes do experimento indicaram corretamente possíveis locais de residência para os três usuários, todos compatíveis com as observações dos especialistas do primeiro teste. E, muitas vezes, os não especialistas conseguiram, inclusive, informar os respectivos logradouros e números das residências, devido ao uso dos serviços online disponíveis no laboratório utilizado no teste.

No entanto, percebeu-se que a não informação de que os dados se referiam a uma época de greve gerou muitas dúvidas sobre o curso frequentado pelo primeiro usuário, que, conforme dito anteriormente, realmente não compareceu à Universidade no período avaliado. De todo modo, diversos participantes do segundo teste apontaram que ou o usuário não frequentou um curso ou que o mesmo podia ser um aluno semipresencial, sendo ambas as respostas então consideradas válidas.

Com relação aos prováveis locais de trabalho, muitos participantes enfatizaram que o primeiro usuário fazia visitas esporádicas a determinados locais no período noturno, nos arredores de sua residência, como os especialistas também haviam indicado no primeiro teste. Essa constatação ficou ainda mais patente no grupo experimental, cujos participantes tiveram acesso à ferramenta semiautomática f-Times.

Isso levou a possibilidades de profissões bastante diversas para esse usuário, todas devidamente justificadas, abrangendo desde celebrante de cerimônias religiosas, na igreja já citada, e vigilante noturno a traficante de drogas. Essa última conclusão pode advir particularmente do fato de que o bairro em que se localiza a provável residência apresenta grandes índices de criminalidade, fato esse conhecido por boa parte dos alunos que participaram do teste.

E ainda no que se refere ao local de trabalho, vários participantes do grupo experimental, que tiveram acesso às ferramentas semiautomáticas, apontaram que o segundo usuário provavelmente trabalhava numa universidade particular, que pôde ser devidamente identificada com base no acesso aos serviços online. Vale salientar que tal universidade fica localizada no bairro nobre que os especialistas do primeiro teste haviam identificado. No entanto, o mesmo não aconteceu com os participantes do grupo de controle, o que mais uma vez pode justificar as diferenças de ganhos discutidas anteriormente na avaliação quantitativa do experimento, na seção anterior.

No que se refere às idas ao Restaurante Universitário, é importante lembrar que as PSTs analisadas apontavam que nenhum dos três usuários frequentou o referido local. Porém, alguns participantes do grupo experimental encontraram um padrão que indicava que o segundo usuário almoçou algumas vezes numa lanchonete na Universidade, fato esse que só um dos especialistas conseguiu também identificar no primeiro teste. Tal descoberta, no entanto, provavelmente só foi possível mais uma vez graças ao uso das ferramentas semiautomáticas disponibilizadas.

Maiores detalhes sobre esse segundo teste com usuários podem ser encontrados

ainda em LEITE *et al.* (2018).

### **6.3.5 Contribuições à Solução Proposta**

Uma vez que os resultados desse segundo teste justificaram o emprego das ferramentas semiautomáticas como instrumentos para agilizar a obtenção de respostas, as mesmas foram adotadas como parte da solução final proposta. Do mesmo modo, esse teste demonstrou que o acesso paralelo a serviços online também foi bastante útil, permitindo aos usuários não especialistas superarem, em certos pontos, as observações dos especialistas do primeiro teste.

Isso ocorreu, por exemplo, quando os não especialistas conseguiram indicar locais geralmente frequentados pelos usuários representados nas PSTs ou informaram os endereços completos das residências dos mesmos. Assim, resolveu-se incluir a opção do acesso a tais serviços também no próprio workflow estabelecido.

Por último, com a finalidade de simplificar a tomada de notas sobre ocorrências pelos participantes, propôs-se adicionar componentes de anotação na abordagem proposta. Assim, foi estabelecida a etapa específica de documentação no workflow, bem como implementado o subsistema de anotação para a versão final da aplicação. Desse modo, abordagem, workflow e aplicação desenvolvida puderam receber suas configurações definitivas, já devidamente apresentados nos Capítulos 4 e 5 deste trabalho. A eficiência dessas configurações finais foi avaliada no terceiro teste com usuários, descrito a seguir.

## **6.4 Terceiro Teste com Usuários**

O terceiro teste com usuários foi realizado com a configuração definitiva da aplicação desenvolvida, ou seja, a versão empregada no segundo teste, porém acrescida das ferramentas semiautomáticas, agora sempre disponíveis, e também já contando com o subsistema de anotação (Figura 42) para a geração de histórias. Dessa forma, dado que a abordagem, o workflow e a aplicação desenvolvida receberam suas versões definitivas, esse terceiro teste apresentou um objetivo bastante específico: avaliar a eficiência da solução proposta em contextos menos conhecidos tanto por analistas especialistas quanto por não especialistas.

Assim, optou-se por adotar um novo dataset, o T-Drive<sup>7</sup>, que representa trajetórias de 10.357 táxis da cidade de Beijing no período entre 02/02/2008 e 08/08/2008 (YUAN *et al.*, 2010; YUAN *et al.*, 2011). Essa opção por utilizar dados da China permitiu, então, observar o

<sup>7</sup> <https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/>

comportamento dos participantes do teste investigando dinâmicas geralmente bem diferentes daquelas presentes em seu cotidiano.

Para tanto, inicialmente foram selecionados 100 táxis aleatórios do T-Drive, sendo geradas suas respectivas PSTs. Dado que as viagens de táxis são geralmente mais dispersas do que os movimentos de indivíduos, geralmente apresentando poucos padrões detectáveis, utilizou-se a versão final da própria aplicação desenvolvida para selecionar cuidadosamente três dessas 100 PSTs geradas. As PSTs selecionadas apresentavam graus crescentes de difusão de rotas (Figuras 54, 56 e 58) e alguns possíveis padrões de movimento e de comportamento.

Como o dataset empregado era diferente do utilizado no primeiro teste, foram estabelecidas novas perguntas, mais uma vez elaboradas com base nas orientações de ANDRIENKO *et al.* (2008), para a execução de tarefas sobre cada uma das três PSTs selecionadas:

1. Quais são as principais rotas utilizadas pelo táxi?
2. Quando as viagens de táxi costumam ocorrer?
3. Quais são as principais paradas de táxi representadas?

Como esse terceiro teste com usuários envolveu a participação de especialistas e de não especialistas simultaneamente, fez-se necessário proceder alguns ajustes e readequações em relação aos dois testes anteriormente realizados.

Assim, para os especialistas, foram seguidos os mesmos passos do primeiro teste, adotando TAP com dois participantes especialistas em análise de dados, porém agora sem qualquer experiência no uso de PSTs:

- Y1. Professor do Curso de Sistemas e Mídias Digitais da UFC, com doutorado em análise de padrões de trajetória e perito no desenvolvimento de aplicações interativas para visualização de dados;
- Y2. Professor de uma universidade particular de Fortaleza, mestre em Arquitetura com especialização em comunicação em novas mídias, ex-sócio de uma empresa na área de criação de serviços online e com grande experiência em design de informação.

Já para os participantes não especialistas, foi adotada uma versão simplificada do segundo teste, considerando-se somente sua etapa de teste propriamente dita. Assim, foi formado um único grupo de não especialistas, composto por 39 estudantes (34 do sexo masculino e 5 do sexo feminino), com idades variando entre 17 e 28 anos, novamente do Curso de Sistemas e Mídias Digitais da UFC.

Dado o prévio conhecimento dos comportamentos dos analistas investigando PSTs

com base nos testes anteriores, optou-se por realizar uma avaliação mais objetiva, já buscando comparar diretamente as respostas de especialistas com as de não especialistas. Assim, apesar dos especialistas terem trabalhado empregando o TAP, tanto eles quanto os não especialistas responderam a um mesmo questionário em papel (Apêndice C), voltado tanto a receber as respostas quanto a descrever a maneira como as mesmas foram obtidas.

Como complemento, todos os participantes foram instruídos a fazer uso do sistema de anotação, documentando suas respostas através da criação de histórias. Dessa forma, tornou-se mais fácil avaliar os resultados obtidos, em especial de forma numérica, pois boa parte das informações necessárias para a avaliação desse terceiro teste já pôde ser, então, armazenada em meio digital. Apesar da necessidade de conferência entre questionários e histórias, de um modo geral, isso acabou facilitando sobremaneira o processo de compilação das respostas e as consequentes análises necessárias.

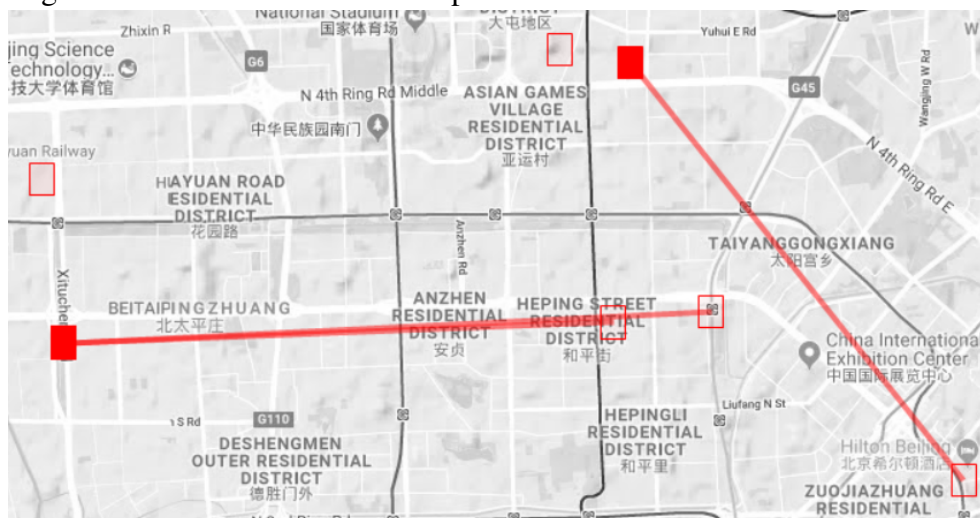
Antes de realizarem o teste, especialistas e não especialistas assistiram a apresentações de vinte minutos que: expuseram o conceito de PSTs; discutiram brevemente os dois testes realizados anteriormente; e mostraram o modo como a abordagem, o workflow e a aplicação final desenvolvida foram estabelecidos e empregados. Essa apresentação foi realizada individualmente para os especialistas, novamente em seus respectivos locais de trabalho e agora dotando o notebook utilizado de acesso a internet através de um smartphone; e em grupo para os não especialistas, mais uma vez durante uma aula da disciplina de Matemática Aplicada a Multimídia I. Novamente os participantes foram instruídos a utilizarem a aplicação como quisessem, não necessariamente seguindo o workflow apresentado.

Em geral, pôde-se perceber que, considerando o retorno dos questionários e das respectivas histórias geradas, especialistas e não especialistas chegaram a diversas conclusões similares. A seguir são apresentados os resultados da avaliação desse terceiro teste com usuários, de acordo com cada uma das três PST selecionadas.

#### **6.4.1 *Análise da Primeira PST***

Com relação à pergunta sobre quais foram as principais rotas utilizadas pelo táxi, ambos os especialistas e 15 não especialistas identificaram, principalmente usando os filtros f-POIs e f-Routes, e verificando informações em serviços online, que esse táxi parecia ser usado principalmente pelos hóspedes de um hotel, fazendo viagens regulares a uma estação de metrô e a um shopping center a partir desse hotel (Figura 54).

Figura 54 – Overview da PST do primeiro táxi avaliado.



Fonte: o autor.

Já no que se refere à pergunta sobre quando as viagens do táxi costumavam ocorrer, principalmente utilizando o filtro f-Times e posteriormente inspecionando os itens da tree list, os dois especialistas e 14 não especialistas comentaram que as viagens desse primeiro táxi aconteciam principalmente das 7:00 às 15:00.

Figura 55 – Major hubs encontrados na PST do primeiro táxi avaliado.



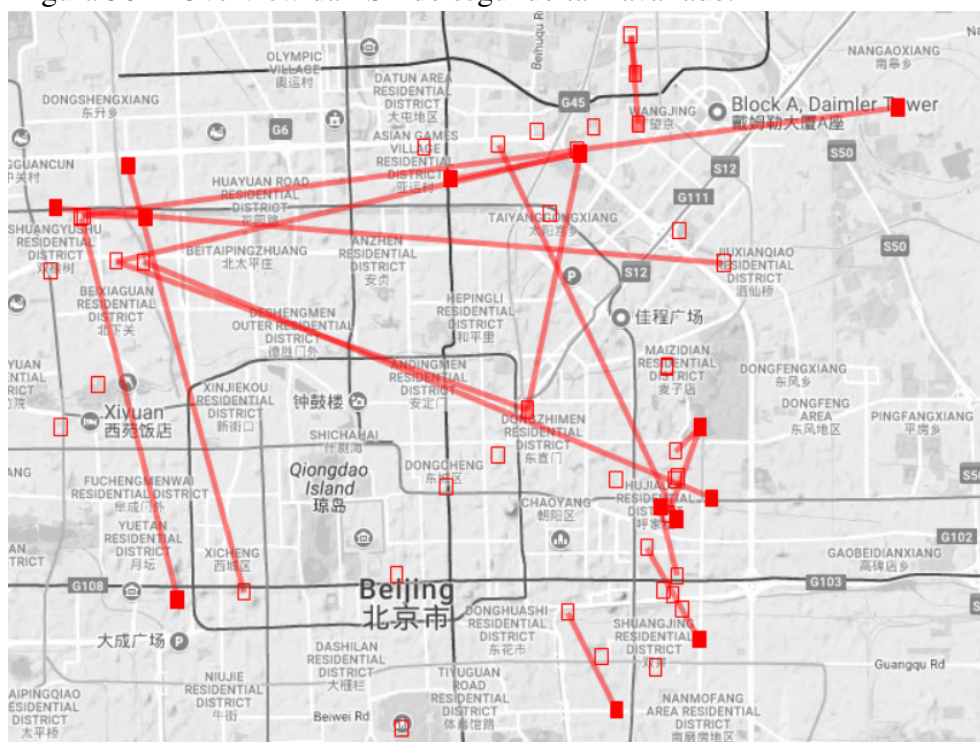
Fonte: o autor.

E dadas as pequenas profundidade e extensão da PST referente a esse táxi, foi considerado bastante simples identificar suas principais paradas utilizadas, já que a representação no mapa apresentava major hubs (conceito apresentado na Seção 1.2 deste trabalho) facilmente identificáveis no mapa virtual. Assim, tanto os especialistas quanto 17 não especialistas puderam indicar as paradas com precisão, já que os referidos major hubs puderam ser enfatizados com o emprego do filtro f-POIs e seus respectivos itens da tree list apresentavam altas probabilidades de ocorrência (Figura 55).

#### 6.4.2 Análise da Segunda PST

De um modo geral, pôde-se perceber que os participantes do teste tiveram um pouco mais de dificuldade para identificar os padrões de movimento e de comportamento do Táxi 2, uma vez que sua PST era mais ampla e apresentava comportamentos bem mais variados do que os do primeiro táxi avaliado, devido principalmente às suas altas frequências de deslocamento (Figura 56).

Figura 56 – Overview da PST do segundo táxi avaliado.

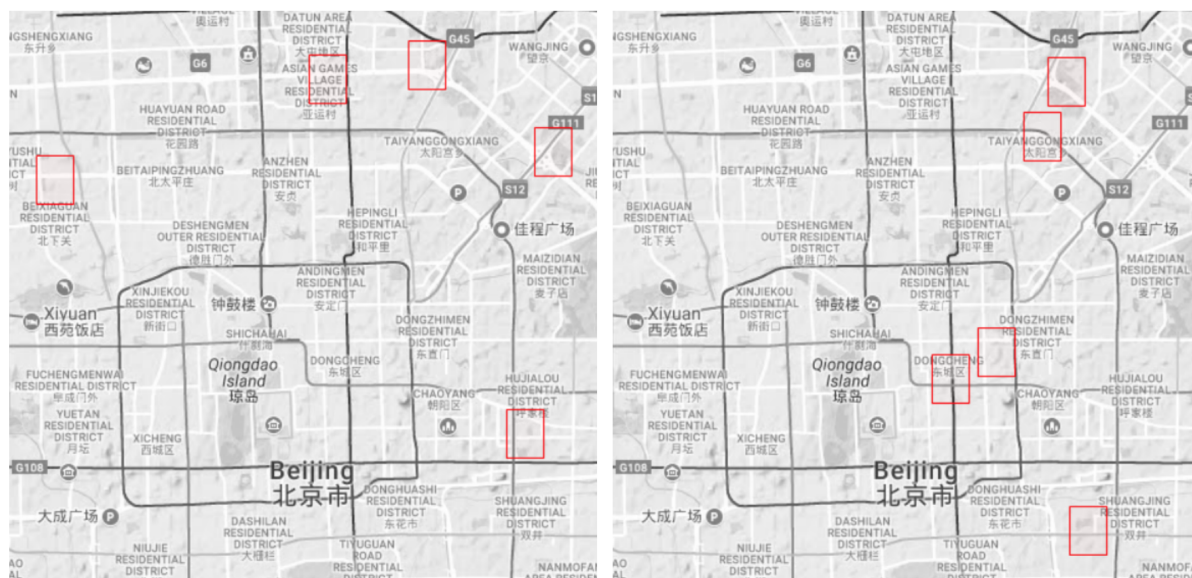


Fonte: o autor.

No entanto, principalmente observando o overview de OD-links, tanto os dois especialistas quanto 12 não especialistas conseguiram identificar dois conjuntos de percursos bem definidos para responder à questão sobre quais eram as principais rotas utilizadas pelo táxi. Curiosamente, em complemento ao uso de f-POIs e f-Routes, foi o filtro f-Times que acabou ajudando a definir esses dois conjuntos distintos, destacando um no período diurno e outro no noturno (Figura 57).

Ainda com base diretamente nos resultados do emprego de f-Times, para responder à pergunta sobre quando as viagens do táxi costumavam ocorrer, os dois especialistas e 12 não especialistas detectaram que as viagens do primeiro conjunto de rotas iniciavam basicamente pela manhã, estendendo-se até o meio da tarde; e que as viagens do segundo conjunto aconteciam

Figura 57 – Agregação de dados, obtidos a partir do uso do filtro f-Times, para definição dos dois conjuntos distintos de rotas principais referentes ao segundo táxi avaliado: (a) percursos diurnos e (b) percursos noturnos.



(a)

(b)

Fonte: o autor.

praticamente somente à noite. Nesse caso, devido à existência desses dois conjuntos bem definidos de rotas, Y2 e 8 não especialistas especularam que possivelmente esse táxi fosse compartilhado por dois ou mais motoristas, que então trabalhariam em períodos distintos.

Já com relação à pergunta sobre quais seriam as principais paradas de táxi representadas, os participantes identificaram várias possibilidades, mas ambos os especialistas e 18 não especialistas evidenciaram um mesmo provável ponto de parada principal, nas proximidades de uma lanchonete McDonald's (Figura 40), devidamente identificada com base no acesso aos serviços online utilizados.

### 6.4.3 Análise da Terceira PST

A terceira PST (Figura 58) apresentava uma árvore consideravelmente mais complexa para ser analisada, sendo bem mais profunda e larga que as outras duas avaliadas anteriormente. Os filtros f-POIs e f-Routes mais uma vez contribuíram para as descobertas iniciais. Porém, foi principalmente trabalhando com a relação entre a tree list e o mapa virtual que ambos os especialistas e 19 não especialistas conseguiram indicar viagens frequentes desse táxi a uma universidade e a um instituto de ciências, devidamente identificados através do uso de serviços online.

No que se refere a quais seriam as principais rotas utilizadas, os dois especialistas e



Figura 58 – Overview da PST do terceiro táxi avaliado.



Fonte: o autor.

20 não especialistas concluíram que as viagens desse terceiro táxi ocorriam principalmente de manhã e à tarde, variando bastante em seus horários específicos. Devido à grande complexidade dessa PST, os participantes do teste buscaram estabelecer padrões principalmente avaliando as maiores probabilidades observadas, usando células e grupos e, então, inspecionando os respectivos itens da tree list.

Já com relação à avaliação das principais paradas de táxi utilizadas, os participantes forneceram diversas respostas diferentes, dada a grande difusão dessa PST. A única resposta similar, fornecida pelos dois especialistas e por 18 não especialistas, indicou uma parada em particular, próxima a um restaurante numa grande estação de trens. É importante destacar que, ao se analisar as trajetórias representadas no mapa virtual, tal estação acabava realmente se sobressaindo devido ao seu formato e à sua grande área.

## 6.5 Discussão Geral

De uma maneira geral, os três testes com usuários visaram compreender como os diferentes perfis de analistas, aplicando a abordagem proposta e o respectivo workflow, poderiam fazer inferências válidas sobre padrões de movimento e de comportamento de indivíduos. Dados os resultados dos testes, nota-se que as questões relacionadas à avaliação de PSTs, levantadas ainda na Seção 1.2 desta tese, parecem ter orientado devidamente o desenvolvimento da solução aqui proposta. Percebe-se isso com base na análise das observações dos participantes dos três testes realizados, principalmente verificando o êxito obtido pelos analistas ao mapearem ideias abstratas (dados das PSTs) em conceitos familiares (possíveis ou prováveis ocorrências em locais específicos), mapeamento esse tão almejado pelo sensemaking.

A seguir são apresentadas algumas considerações mais específicas acerca da validade das respostas, da duração das tarefas realizadas e das histórias geradas empregando o subsistema de anotação. Esses dados encontram-se resumidos na Tabela 1. Além disso, são feitas ainda algumas observações julgadas pertinentes sobre o comportamento dos participantes dos três testes como um todo.

Tabela 1 – Sumarização de dados dos três testes realizados com usuários

<b>Teste</b>	<b>Participantes</b>	<b>Duração das Tarefas (minutos)</b>	<b>Respostas Válidas</b>	<b>Total de Respostas Possíveis</b>	<b>Taxa de Respostas Válidas (%)</b>	<b>Quantidade de Histórias</b>
<b>1</b>	2 especialistas	15	24	24	100	**
<b>2</b>	34 não especialistas	20*	259	408	63	**
<b>3</b>	2 especialistas	24	18	18	100	19
	39 não especialistas	20*	201	351	57	204

Fonte: o autor.

Nota: \*Limitado ao tempo da aula; \*\*Não se aplica devido ao subsistema de anotação ainda não disponível.

### 6.5.1 Validade de Respostas

Segundo a Tabela 1, considerando as respostas válidas obtidas no primeiro teste, ambos os especialistas responderam consistentemente a todas elas. No segundo, o grupo de não especialistas alcançou uma taxa de respostas válidas<sup>8</sup> de 63%, ou seja, 259 respostas válidas para 408 respostas possíveis no total (respostas totais = 34 participantes x 4 perguntas x 3 PSTs). E no terceiro teste, os especialistas apresentaram respostas válidas a todas as questões, e os não

<sup>8</sup> proporção entre as respostas obtidas consideradas válidas e o número total de respostas esperadas.

especialistas atingiram uma taxa de respostas válidas de 57%, ou seja, 201 respostas válidas para um total de 351 respostas (total de respostas = 39 participantes x 3 questões x 3 PSTs).

Conforme já citado anteriormente, o interesse dos testes realizados não foi especificamente obter as respostas em si, mas principalmente registrar como os participantes trabalharam para obtê-las. De todo modo, é importante destacar que eles conseguiram responder a boa parte das perguntas formuladas. Tal fato é particularmente importante no que diz respeito à participação dos analistas não especialistas, que nunca haviam avaliado PSTs ou mesmo outros tipos abstratos de dados espaçotemporais.

Em geral, as respostas válidas representaram interpretações individuais dos diversos participantes. No entanto, nota-se que em casos específicos, os analistas, especialistas ou não especialistas, chegaram a conclusões bastante interessantes e similares. Fora os diversos exemplos já citados, considerando o primeiro e o segundo teste, um especialista e vários não especialistas inferiram que o terceiro usuário provavelmente usava principalmente ônibus para se locomover, sendo inclusive indicadas as respectivas paradas geralmente utilizadas.

E no terceiro estudo, merece destaque a menção de diversos participantes à cadeia de lanchonetes McDonald's, ao tentarem estabelecer pontos de parada para os táxis representados. Tais observações são particularmente importantes, pois, segundo os questionários e as respectivas histórias registradas, muitos participantes especialistas e não especialistas informaram ter reconhecido o respectivo logotipo ao usar serviços online para visitas virtuais às ruas relacionadas (Figura 40). A identificação dessa imagem, em particular, demonstra como o conhecimento prévio dos participantes do terceiro teste pode ter influenciado a investigação numa realidade tão distante daquela dos analistas. Assim, essa imagem deve ter se destacado sobremaneira nas ruas de Beijing, onde não é tão comum avistar-se caracteres ocidentais.

Do mesmo modo, conforme já mencionado, a identificação de um ponto de parada do terceiro táxi próximo a um restaurante numa estação de trens pode ter sido induzida pelo destaque visual dessa grande construção no mapa virtual. Além disso, o mesmo tipo de reconhecimento, provavelmente baseado na experiência prévia dos participantes com o ambiente universitário, pode ter ajudado também a reforçar menções a uma universidade e a um instituto científico quando os analistas buscavam as principais rotas desse mesmo táxi.

### 6.5.2 *Duração das Tarefas*

De acordo com a Tabela 1, com relação ao tempo gasto para explorar os dados fornecidos, no primeiro teste, os dois especialistas levaram cerca de 15 minutos para cada uma das PSTs. No terceiro teste, por outro lado, cada especialista levou cerca de 24 minutos para explorar cada PST. O tempo extra necessário no terceiro teste deve-se possivelmente à maior complexidade na análise, já que os deslocamentos de um táxi são geralmente bem mais dispersos que os do cotidiano de uma única pessoa. Assim, encontrar padrões de movimento e de comportamento para um táxi torna-se bem mais complexo. Além disso, é importante lembrar que os especialistas do primeiro teste já possuíam bastante experiência na avaliação de PSTs, ao contrário dos especialistas do terceiro teste, que até então só haviam lidado com outros tipos de estruturas para a análise de dados.

Particularmente, ainda com relação ao terceiro teste, foi notado que um dos especialistas, Y1, perito em análise de trajetória, acabou levando um pouco mais de tempo para se familiarizar com o aplicativo fornecido, em comparação ao especialista Y2. Isso possivelmente aconteceu porque, a princípio, Y1 não percebeu que a abordagem proposta se referia especificamente à predição de movimentos e de comportamentos, sendo diferente de sua área de trabalho original, na qual as trajetórias correspondentes são observadas somente com relação ao tempo presente. Mas, notou-se que, após compreender o real propósito da aplicação desenvolvida, considerando as probabilidades de ocorrência dos eventos, esse especialista passou a investigar as PSTs de forma adequada e bem mais rápida, mostrando-se inclusive muito entusiasmado com os resultados obtidos. Assim, em relação ao sensemaking, a experiência prévia de Y1 parece ter prejudicado seu desempenho no início. Porém, essa mesma experiência veio, possivelmente, auxiliá-lo posteriormente na realização mais rápida das tarefas predefinidas.

Com relação aos não especialistas, como os respectivos testes foram realizados durante aulas, houve grande restrição no tempo disponível para operação dos protótipos fornecidos – apenas cerca de uma hora, no total. Assim, no caso dos não especialistas do segundo teste, observou-se que o tempo médio gasto pelos membros do grupo experimental foi de aproximadamente 19 minutos para analisar cada PST, enquanto o grupo de controle precisou ter seu tempo limitado aos 20 minutos disponíveis para a investigação de cada árvore, devido à restrição de horário da aula. De todo modo, mais uma vez, o uso das ferramentas semiautomáticas pode ajudar na explicação de tal diferença no tempo necessário para as análises entre os dois grupos.

Já no terceiro teste, devido à complexidade das PSTs fornecidas, à não familiaridade

das situações representadas e também às limitações de tempo durante a aula, o grupo de não especialistas atingiu o limite total de tempo disponível, também de vinte minutos por árvore. Isso acabou influenciando na quantidade de respostas fornecidas e, conseqüentemente, na taxa de respostas válidas, já discutida na Seção 6.5.1. Provavelmente, no geral, os não especialistas teriam alcançado maiores taxas caso dispusessem de mais tempo para trabalhar.

### **6.5.3 *Quantidade de Histórias Registradas***

Conforme a Tabela 1, no terceiro teste, o único onde o uso do subsistema de anotação esteve disponível, foram registradas um total de 223 histórias, sendo 19 feitas pelos dois especialistas e 204 pelos não especialistas. Todas as anotações realizadas apresentaram links para serviços online. Apesar da restrição do uso somente no terceiro teste, esses números são particularmente significativos, pois demonstram o engajamento dos participantes em tentar estabelecer hipóteses e validá-las ou refutá-las. Isso fica demonstrado principalmente quando se considera a limitação de tempo imposta aos participantes não especialistas, ou se observa que o número de anotações realizadas pelos analistas especialistas, 19, foi até superior ao número total de 18 respostas possíveis no teste.

### **6.5.4 *Outras Observações***

De uma forma geral, há ainda algumas observações sobre os testes realizados e sobre a análise de suas respectivas respostas que merecem ser apresentadas.

Curiosamente, notou-se que avaliar uma PST mais profunda, às vezes, é mais fácil que avaliar uma mais rasa. Por exemplo, no terceiro teste, para encontrar as paradas do segundo táxi a partir de sua PST, mais rasa, os analistas tiveram que se concentrar em muitos nós diferentes ao mesmo tempo, procurando pelas probabilidades mais altas de presença. Por outro lado, para encontrar as paradas do terceiro táxi com base em sua PST, mais profunda, os analistas apenas avaliaram, diretamente na tree list, alguns nós referentes a POIs. Então, eles verificaram os respectivos descendentes, a fim de identificar se os movimentos retornavam ao mesmo local inicial, ou a suas proximidades.

Também no terceiro teste, os participantes comentaram que observar o mapa virtual em visão de satélite foi muito importante para a localização correta de lugares nos serviços online, particularmente devido à dificuldade de entender os ideogramas chineses observados, geralmente presentes na visão em modo padrão (Figuras 54, 56 e 58). O mesmo não aconteceu

no primeiro e no segundo testes, possivelmente porque os participantes já conheciam os locais que estavam investigando, em Fortaleza, a própria cidade em que a grande maioria deles residia.

Nos três testes, muitos participantes, tanto especialistas quanto não especialistas, relataram que o recurso de edge bundling acabou introduzindo uma certa poluição visual quando as PSTs avaliadas apresentavam uma distribuição geograficamente mais dispersa. Isso provavelmente aconteceu porque a técnica de edge bundling é geralmente mais adequada para mostrar fluxos oriundo de apenas alguns locais (ERSOY *et al.*, 2011), o que não foi o caso da maioria das PSTs fornecidas. De todo modo, optou-se por manter o edge bundling na abordagem proposta, pois analistas podem se deparar com casos nos quais tal recurso ainda possa ser útil.

Vale ressaltar também que, no segundo e no terceiro testes, a maioria dos participantes acabou geralmente seguindo os passos dos workflows apresentados, com pequenas variações, para operar os aplicativos. Sendo assim, então os mesmos parecem ser intuitivos e ter sido bem compreendidos, apesar de suas breves apresentações somente durante as demonstrações de uso dos protótipos no início dos testes.

De uma forma geral, as análises dos questionários indicaram que as aplicações desenvolvidas foram consideradas úteis e relativamente fáceis de operar. Assim, com relação à classificação da experiência de uso, os especialistas do primeiro teste consideraram a operação do protótipo inicial "particularmente agradável", "fácil de usar" e "que forneceu uma maneira bastante adequada de analisar as PSTs". Os não especialistas do segundo teste também gostaram da experiência de usar o protótipo evoluído, empregando muitos adjetivos distintos para classificá-lo, incluindo: "interessante", "cativante", "intrigante" e "divertido". Já no terceiro teste, dada a maior complexidade no contexto de análise das PSTs fornecidas e a pouca familiaridade com a região explorada, os especialistas classificaram o experimento como "desafiador, porém intuitivo" e "intenso", enquanto os não especialistas acharam a experiência "confusa para usuários não treinados<sup>9</sup>, podendo causar frustração", "extremamente interessante" e "difícil, mas divertida".

Por último, houve também reações gerais inusitadas, porém ainda dignas de nota, tais como "acho que eu até já consigo andar na China", "Deus, eu sei até qual é a religião do sujeito!", "imagine usar isso pra ganhar dinheiro" e "nunca mais vou usar um celular".

<sup>9</sup> no caso, referindo-se especificamente à utilidade da demonstração de uso realizada durante o início do teste.

## 6.6 Considerações Finais

O presente capítulo apresentou o processo de validação da solução proposta, envolvendo abordagem, workflow e aplicações desenvolvidas. Para tanto, foram descritos os procedimentos para a realização de três testes com usuários, envolvendo dois datasets específicos e a participação de um total de 77 analistas, sendo 4 especialistas e 73 não especialistas.

Os resultados de tais testes, também devidamente discutidos, além de servirem como base para a validação, também foram essenciais ao processo incremental empregado para o estabelecimento da solução como um todo, que envolveu a evolução da abordagem, do workflow e das diferentes versões da aplicação de análise de PSTs empregadas.

O próximo capítulo conclui esta tese, apresentando os resultados alcançados, as principais limitações enfrentadas e ainda possíveis trabalhos futuros.

## 7 CONCLUSÕES

Este capítulo finaliza a tese aqui apresentada. A Seção 7.1 resume os esforços realizados como um todo. Na Seção 7.2 são listados os principais resultados alcançados pelo trabalho e, na Seção 7.3, expostas as limitações enfrentadas em sua execução. A Seção 7.4 discute possíveis trabalhos futuros e, por fim, a Seção 7.5 apresenta as últimas considerações.

### 7.1 Resumo

Esta tese propôs uma abordagem, baseada em análise visual, na metodologia sensemaking e em branching time, para simplificar a tarefa de investigar possíveis padrões de movimento e de comportamentos de indivíduos, empregando dados armazenados em PSTs, estruturas capazes de representar simultaneamente espaço, tempo e probabilidade.

Para tanto, foram inicialmente apresentados trabalhos relacionados ao tema e discutidos conceitos considerados importantes para a devida compreensão desta tese, abordando predição de localização por PSTs, análise visual, sensemaking e representação de branching time.

Posteriormente, foi devidamente apresentada e discutida a abordagem proposta, incluindo sua arquitetura e o respectivo workflow, voltado à otimização de seu uso.

Para validar a abordagem proposta e o respectivo workflow definido, foram desenvolvidas versões de uma aplicação específica de visualização interativa, sendo realizados testes com usuários, envolvendo a participação de analistas especialistas e não especialistas e dois datasets distintos. Apesar da natureza bastante particular do uso das PSTs e da relativa complexidade em analisá-las, os resultados obtidos na avaliação dos testes realizados demonstraram a viabilidade de se aplicar a solução para identificar padrões de movimento e de comportamento de indivíduos.

Assim, demonstrou-se que a solução proposta permitiu que os analistas respondessem a questões inicialmente estabelecidas, sendo também capazes de propor seus próprios questionamentos, buscando novas respostas.

É importante salientar que este trabalho se baseou na adoção de um processo incremental, que considerou as contribuições de cada um dos testes procedidos e a realização paralela de operações de investigação, pelo próprio autor da tese, sobre diversas PSTs. Tal processo visou uma maior vivência, ativa e prática, no que tange ao problema enfrentado e auxiliou sobremaneira no estabelecimento das configurações intermediárias e finais da abordagem e do



workflow ora propostos.

## 7.2 Resultados Alcançados

Os principais resultados alcançados por este trabalho foram:

- Proposição de uma abordagem específica, baseada em análise visual, sensemaking e branching time, que auxilie analistas na investigação de PSTs, permitindo a descoberta de padrões de movimento e de comportamento de indivíduos;
- Estabelecimento de um workflow específico, voltado a otimizar o uso da abordagem proposta, dividido em etapas bem definidas e de possível execução em paralelo;
- Desenvolvimento de diferentes versões de uma aplicação para análise visual de PSTs, com base na solução proposta (abordagem e workflow), empregando algumas técnicas de otimização e operações em dois datasets distintos. O primeiro dataset foi fornecido pelo Projeto Eaí?, que gerava PSTs a partir de dados espacotemporais capturados de dispositivos móveis de usuários que desejavam informações acerca do cardápio do Restaurante Universitário da Universidade Federal do Ceará. Já o segundo dataset utilizado foi o T-Drive, que representava trajetórias de 10.357 táxis da cidade de Beijing, no período entre 02/02/2008 e 08/02/2008;
- Realização de três testes com usuários das diferentes versões da aplicação desenvolvida para este trabalho, com a participação total de 77 analistas, sendo 4 especialistas e 73 não especialistas. Tais testes, realizados de acordo com diferentes metodologias, serviram de base para a avaliação da eficiência da solução proposta;
- Análise dos resultados dos três testes realizados com usuários, envolvendo abordagens quantitativa e qualitativa, para uma melhor compreensão sobre como a adoção de sensemaking e de branching time pode ter influenciado o modo como os participantes realizaram suas descobertas e propuseram seus próprios questionamentos; e
- Identificação de alguns procedimentos específicos, observados nos testes, que podem servir de base à criação de novas ferramentas, automáticas ou semiautomáticas, para a análise visual de PSTs.

### 7.3 Limitações

De um modo geral, percebeu-se que os resultados obtidos demonstraram a viabilidade do uso dessa solução nos processos de caracterização, busca e comparação de comportamentos em PSTs individuais.

No entanto, com relação especificamente aos testes com usuários não especialistas, as taxas de respostas válidas obtidas ainda podem ser consideradas relativamente baixas, particularmente em comparação àquelas atingidas pelos participantes especialistas, que beiraram os 100%: 63% para os não especialistas do segundo teste e 57%, para os do terceiro.

Assim, embora se reconheça que avaliar PSTs possa ser uma tarefa difícil e que tiveram de ser impostas grandes restrições de tempo aos não especialistas, devido aos testes serem realizados durante aulas de graduação, acredita-se que aprimoramentos na abordagem proposta podem levar a resultados melhores.

Também se notou que, apesar dos esforços voltados à otimização do processamento das PSTs utilizadas e da respectiva representação gráfica das mesmas, ainda há várias possibilidades para a implementação de métodos mais rápidos e eficientes para a realização de tais tarefas, particularmente considerando o emprego de GPU.

A observação de alguns comportamentos particulares dos analistas durante os testes merecem, futuramente, servir de base a estudos mais aprofundados, no intuito de estabelecer possíveis técnicas de investigação mais específicas, que possam simplificar e agilizar o processo de detecção de ocorrências. Dentre esses comportamentos, pode-se destacar aqueles que envolveram a identificação de POIs e de ROIs levando em consideração a diferenciação entre PSTs rasas e profundas, como ocorreu particularmente com a busca de paradas de táxis no terceiro teste.

Dadas das limitações impostas com relação ao recrutamento de participantes, principalmente de analistas especialistas, e apesar do número relativamente alto de 77 usuários presentes nos testes realizados, julga-se necessária ainda a execução de outros experimentos, principalmente adotando outras metodologias específicas para realização e para avaliação.

E, face a importância demonstrada pelo uso das PSTs, porém dadas as políticas de privacidade normalmente adotadas na geração das mesmas, entende-se que uma avaliação mais ampla e de cunho estatístico ainda seja importante para atestar a robustez dos resultados ora obtidos.

## 7.4 Trabalhos Futuros

Como trabalho futuro, propõe-se inicialmente incluir novos recursos interativos, como, por exemplo, algum tipo de widget na própria interface gráfica, que facilite a navegação cruzada entre a tree list e o mapa virtual, buscando agilizar o processo de investigação como um todo.

Espera-se também adotar técnicas de machine learning, ou semelhantes, a fim de automatizar alguns padrões de uso observados nos testes com usuários, como investigar diferenciadamente PSTs rasas ou profundas, ajudando os analistas a identificar POIs e ROIs mais rapidamente e tentando tornar tal processo mais autônomo e rápido.

Também como trabalho futuro, pretende-se estender a abordagem para lidar com a avaliação simultânea de diversas PSTs, buscando identificar padrões gerais de movimento e de comportamento, visando estabelecer perfis comuns a grupos de usuários.

Um outro trabalho futuro seria ainda aprimorar o subsistema de anotação proposto, trabalhando em três frentes distintas: adotar recursos semânticos, baseados no uso de etiquetas; integrar o acesso a serviços online diretamente à própria aplicação de análise visual; e armazenar na nuvem os resultados obtidos, viabilizando consultas posteriores por um ou mais analistas em paralelo.

Uma outra opção seria estender a ideia inicial de PSTs, buscando especializá-las para considerar questões mais específicas sobre padrões de movimento e de comportamento de indivíduos. Assim, as PSTs poderiam incluir, além da probabilidade de ocorrência de eventos, também propriedades quantitativas ou qualitativas associadas, tais como velocidade, direção, nível de estresse de indivíduos etc. Apesar da complexidade envolvida, principalmente na captura e no processamento de tais dados, essa inclusão poderia repercutir muito positivamente, servindo de base a futuros modelos mais avançados e precisos de análise visual para as ocorrências representadas.

Além disso, planeja-se realizar novos estudos, trabalhando com outros datasets, para avaliar o emprego da abordagem proposta, contemplando diferentes circunstâncias, como, por exemplo: *time and location-based advertisement*<sup>1</sup>, combinando consultas ao log de uso da internet de um usuário aos dados de GPS de seu smartphone; investigação policial, utilizando dados de GPS de tornozeleiras eletrônicas; ou sistemas de *emergency healthcare delivery*<sup>2</sup>,

<sup>1</sup> publicidade fornecida com base no horário e na localização de usuários.

<sup>2</sup> prestação de cuidados de emergência em saúde.

referenciando dados de atendimento de pacientes, de médicos especializados e de instituições de consulta e tratamento. Essas novas possibilidades de dados poderiam, então, ser avaliadas em novos testes com usuários, a fim de se avaliar a eficácia da solução proposta nesses outros contextos.

Por fim, o uso desses novos datasets, bem como a avaliação dos resultados dos respectivos experimentos realizados, poderia dar origem a novos recursos para a abordagem proposta, servindo de base ao estabelecimento de um *framework* específico. Tal *framework*, por sua vez, poderia definir uma solução, baseada na customização de componentes, para simplificar o desenvolvimento de novas aplicações para análise visual em outros áreas.

## 7.5 Considerações Finais

Em resumo, esta tese propôs a adoção de sensemaking para fornecer um caráter menos abstrato às PSTs. Assim, cada nó desse tipo de árvore passa a representar a probabilidade de um dado evento ocorrer em um local específico em um determinado momento; e cada aresta, então, configura possíveis mudanças de estado entre os eventos representados, estabelecidas como relações entre espaço, tempo e probabilidade. E, também considerando o conceito de branching time, cada nó da PST passa a corresponder a um evento individual, que pode ser a origem de um ou mais outros eventos, então representados por seus respectivos possíveis nós filhos.

Dessa forma, uma PST define não mais apenas trajetórias presumíveis, mas também possíveis eventos.

Então, investigando ocorrências passadas e futuras e suas probabilidades associadas, bem como outros indícios levantados, cada padrão de movimento ou de comportamento detectado pode ser devidamente avaliado, podendo ser classificado como possível ou provável.

Com isso, acredita-se que a união proposta de análise visual, sensemaking e branching time possa fornecer alternativas reais à simplificação da avaliação de dados que relacionem espaço, tempo e probabilidade de uma maneira mais natural e eficiente.

Assim, espera-se estar contribuindo para a difusão do uso de PSTs e para a aceleração e o enriquecimento do processo de descoberta de informações como um todo.

## REFERÊNCIAS

- ABRAS, C.; MALONEY-KRICHMAR, D.; PREECE, J. User-centered design. **Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications, Citeseer**, v. 37, n. 4, p. 445–456, 2004.
- AGUINIS, H.; GLAVAS, A. On corporate social responsibility, sensemaking, and the search for meaningfulness through work. **Journal of Management**, SAGE Publications Sage CA: Los Angeles, CA, v. 45, n. 3, p. 1057–1086, 2019.
- AIGNER, W.; MIKSCH, S.; MÜLLER, W.; SCHUMANN, H.; TOMINSKI, C. Visual methods for analyzing time-oriented data. **IEEE transactions on visualization and computer graphics**, IEEE, v. 14, n. 1, p. 47–60, 2008.
- ALON, J.; SCLAROFF, S.; KOLLIOS, G.; PAVLOVIC, V. Discovering clusters in motion time-series data. In: IEEE. **2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.** [S. l.], 2003. v. 1, p. I–I.
- ANDRIENKO, G.; ANDRIENKO, N.; DEMSAR, U.; DRANSCH, D.; DYKES, J.; FABRIKANT, S. I.; JERN, M.; KRAAK, M.-J.; SCHUMANN, H.; TOMINSKI, C. Space, time and visual analytics. **International Journal of Geographical Information Science**, Taylor & Francis, v. 24, n. 10, p. 1577–1600, 2010.
- ANDRIENKO, G.; ANDRIENKO, N.; FUCHS, G.; WOOD, J. Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. **IEEE Transactions on Visualization & Computer Graphics**, IEEE, v. 1, n. 1, p. 1–1, 2017.
- ANDRIENKO, G.; ANDRIENKO, N.; JANKOWSKI, P.; KEIM, D.; KRAAK, M.-J.; MACEACHREN, A.; WROBEL, S. Geovisual analytics for spatial decision support: Setting the research agenda. **International Journal of Geographical Information Science**, Taylor & Francis, v. 21, n. 8, p. 839–857, 2007.
- ANDRIENKO, G.; ANDRIENKO, N.; RINZIVILLO, S.; NANNI, M.; PEDRESCHI, D.; GIANNOTTI, F. Interactive visual clustering of large collections of trajectories. In: IEEE. **2009 IEEE Symposium on visual analytics science and technology.** [S. l.], 2009. p. 3–10.
- ANDRIENKO, G.; ANDRIENKO, N.; SCHUMANN, H.; TOMINSKI, C. Visualization of trajectory attributes in space–time cube and trajectory wall. In: **Cartography from Pole to Pole.** [S. l.]: Springer, 2014. p. 157–163.
- ANDRIENKO, G.; ANDRIENKO, N.; WROBEL, S. Visual analytics tools for analysis of movement data. **ACM SIGKDD Explorations Newsletter**, ACM, v. 9, n. 2, p. 38–46, 2007.
- ANDRIENKO, N.; ANDRIENKO, G. Visual analytics of movement: An overview of methods, tools and procedures. **Information Visualization**, Sage Publications Sage UK: London, England, v. 12, n. 1, p. 3–24, 2013.
- ANDRIENKO, N.; ANDRIENKO, G.; PELEKIS, N.; SPACCAPIETRA, S. Basic concepts of movement data. In: **Mobility, data mining and privacy.** [S. l.]: Springer, 2008. p. 15–38.
- ARTERO, A. O.; OLIVEIRA, M. C. F. de; LEVKOWITZ, H. Uncovering clusters in crowded parallel coordinates visualizations. In: IEEE. **Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On.** [S. l.], 2004. p. 81–88.

BALACHANDRAN, A.; VOELKER, G. M.; BAHL, P.; RANGAN, P. V. Characterizing user behavior and network performance in a public wireless lan. In: ACM. **ACM SIGMETRICS Performance Evaluation Review**. [S. l.], 2002. v. 30, p. 195–205.

BALAZINSKA, M.; CASTRO, P. Characterizing mobility and network usage in a corporate wireless local-area network. In: ACM. **Proceedings of the 1st international conference on Mobile systems, applications and services**. [S. l.], 2003. p. 303–316.

BAO, J.; ZHENG, Y.; WILKIE, D.; MOKBEL, M. Recommendations in location-based social networks: a survey. **GeoInformatica**, Springer, v. 19, n. 3, p. 525–565, 2015.

BEHESHTI, E.; VILLANOSA, K.; HORN, M. Understanding parent-child sensemaking around interactive museum exhibits. **Annual Meeting of the American Educational Research Association (AERA 2018)**, 2018.

BERNSEN, N. O.; DYBKJÆR, H.; DYBKJÆR, L. **Designing interactive speech systems: From first ideas to user testing**. [S. l.]: Springer Science & Business Media, 2012.

BERTIN, J. **Graphics and graphic information processing**. [S. l.]: Walter de Gruyter, 2011.

BLACKWELL, A. F. The reification of metaphor as a design tool. **ACM Transactions on Computer-Human Interaction (TOCHI)**, ACM, v. 13, n. 4, p. 490–530, 2006.

BOBIS, J.; WAY, J. Building connections between children's representations and their conceptual development in mathematics. In: **Forging Connections in Early Mathematics Teaching and Learning**. [S. l.]: Springer, 2018. p. 55–72.

BOOTH, K. S.; LUEKER, G. S. Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms. **Journal of Computer and System Sciences**, Elsevier, v. 13, n. 3, p. 335–379, 1976.

BOYANDIN, I.; BERTINI, E.; BAK, P.; LALANNE, D. Flowstrates: An approach for visual exploration of temporal origin-destination data. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S. l.], 2011. v. 30, n. 3, p. 971–980.

BRĀZMA, A.; JONASSEN, I.; VILO, J.; UKKONEN, E. Predicting gene regulatory elements in silico on a genomic scale. **Genome research**, Cold Spring Harbor Lab, v. 8, n. 11, p. 1202–1215, 1998.

BROCKMANN, D.; HUFNAGEL, L.; GEISEL, T. The scaling laws of human travel. **Nature**, Nature Publishing Group, v. 439, n. 7075, p. 462, 2006.

BROWN, J. S.; COLLINS, A.; DUGUID, P. Situated cognition and the culture of learning. **Educational researcher**, Sage Publications Sage CA: Thousand Oaks, CA, v. 18, n. 1, p. 32–42, 1989.

BROWN, T. M.; MCCABE, S. A.; WELLFORD, C. F. **Global positioning system (GPS) technology for community supervision: Lessons learned**. [S. l.]: Noblis, 2007. (1, 219376).

CAGLAYAN, G. Coordinating analytic and visual approaches: Math majors' understanding of orthogonal hermite polynomials in the inner product space pnp in a technology-assisted learning environment. **The Journal of Mathematical Behavior**, Elsevier, 2018.

CANDIA, J.; GONZÁLEZ, M. C.; WANG, P.; SCHOENHARL, T.; MADEY, G.; BARABÁSI, A.-L. Uncovering individual and collective human dynamics from mobile phone records. **Journal of physics A: mathematical and theoretical**, IOP Publishing, v. 41, n. 22, p. 224015, 2008.

CARD, M. **Readings in information visualization: using vision to think**. [S. l.]: Morgan Kaufmann, 1999.

CHARENTREAU, A.; HUI, P.; CROWCROFT, J.; DIOT, C.; GASS, R.; SCOTT, J. Impact of human mobility on opportunistic forwarding algorithms. **IEEE Transactions on Mobile Computing**, IEEE, n. 6, p. 606–620, 2007.

CHEN, A. L.; CHANG, M.; CHEN, J.; HSU, J.-L.; HSU, C.-H.; HUA, S. Y. Query by music segments: An efficient approach for song retrieval. In: IEEE. **2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)**. [S. l.], 2000. v. 2, p. 873–876.

CHO, M.; KWON, S.; NA, N.; SUK, H.-J.; LEE, K. The elders preference for skeuomorphism as app icon style. In: ACM. **Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems**. [S. l.], 2015. p. 899–904.

CHUDOVA, D.; GAFFNEY, S.; MJOLSNESS, E.; SMYTH, P. Translation-invariant mixture models for curve clustering. In: ACM. **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**. [S. l.], 2003. p. 79–88.

CONSUL, P. C.; JAIN, G. C. A generalization of the poisson distribution. **Technometrics**, Taylor & Francis, v. 15, n. 4, p. 791–799, 1973.

COOK, K. A.; THOMAS, J. J. **Illuminating the path: The research and development agenda for visual analytics**. [S. l.], 2005.

CRNOVRSANIN, T.; MUELDER, C.; CORREA, C.; MA, K.-L. Proximity-based visualization of movement trace data. In: IEEE. **Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on**. [S. l.], 2009. p. 11–18.

DAVIS, F. D. **A technology acceptance model for empirically testing new end-user information systems: Theory and results**. Tese (Doutorado) — Massachusetts Institute of Technology, 1985.

DERVIN, B. **An overview of sense-making research: Concepts, methods, and results to date**. [S. l.]: A autora, 1983.

DODGE, S.; WEIBEL, R.; LAUTENSCHÜTZ, A.-K. Towards a taxonomy of movement patterns. **Information visualization**, SAGE Publications Sage UK: London, England, v. 7, n. 3-4, p. 240–252, 2008.

DUFFY, M. Sensemaking in classroom conversations. **Openness in research: The tension between self and other**, Van Gorcum, p. 119–132, 1995.

DYKES, J. A.; MOUNTAIN, D. M. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. **Computational Statistics & Data Analysis**, Elsevier, v. 43, n. 4, p. 581–603, 2003.

EAGLE, N.; PENTLAND, A. Reality mining: sensing complex social systems. **Personal and ubiquitous computing**, Springer-Verlag, v. 10, n. 4, p. 255–268, 2006.

ERSOY, O.; HURTER, C.; PAULOVICH, F.; CANTAREIRO, G.; TELEA, A. Skeleton-based edge bundling for graph visualization. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 17, n. 12, p. 2364–2373, 2011.

FABRIKANT, S. I.; LOBBEN, A. Introduction: Cognitive issues in geographic information visualization. **Cartographica: The International Journal for Geographic Information and Geovisualization**, International Cartographic Association/Association Cartographique internationale, v. 44, n. 3, p. 139–143, 2009.

FERREIRA, N.; POCO, J.; VO, H. T.; FREIRE, J.; SILVA, C. T. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 19, n. 12, p. 2149–2158, 2013.

FOLEY, J. D.; VAN, F. D.; DAM, A. V.; FEINER, S. K.; HUGHES, J. F.; HUGHES, J.; ANGEL, E. **Computer graphics: principles and practice**. [S. l.]: Addison-Wesley Professional, 1996. v. 12110.

FRANK, A. U. Different types of “times” in gis. **Spatial and temporal reasoning in geographic information systems**, Oxford University Press, New York, p. 40–62, 1998.

GAFFNEY, S.; SMYTH, P. Trajectory clustering with mixtures of regression models. In: **KDD**. [S. l.]: University of California, Irvine, 1999. v. 99, p. 63–72.

GAO, S. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. **Spatial Cognition & Computation**, Taylor & Francis, v. 15, n. 2, p. 86–114, 2015.

GIANNOTTI, F.; NANNI, M.; PINELLI, F.; PEDRESCHI, D. Trajectory pattern mining. In: **ACM. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S. l.], 2007. p. 330–339.

GIANNOTTI, F.; PEDRESCHI, D. Mobility, data mining and privacy: A vision of convergence. In: **Mobility, data mining and privacy**. [S. l.]: Springer, 2008. p. 1–11.

GONZALEZ, M. C.; HIDALGO, C. A.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, Nature Publishing Group, v. 453, n. 7196, p. 779–782, 2008.

GONZALEZ, R. C.; WOODS, R. E. Digital image processing [m]. **Publishing house of electronics industry**, v. 141, n. 7, 2002.

GOSCIOLA, V. **Roteiro para as novas mídias**. São Paulo: Senac, 2003.

GREENO, J. G. Gibson’s affordances. **Psychological Review**, v. 101, n. 2, p. 336–342, 1994.

GUERRIERO, V. Power law distribution: Method of multi-scale inferential statistics. **Journal of Modern Mathematics Frontier**, Science and Engineering Publishing Company, v. 1, n. 1, p. 21–28, 2012.

GUYET, T.; QUINIOU, R. Mining temporal patterns with quantitative intervals. In: IEEE. **2008 IEEE International Conference on Data Mining Workshops**. [S. l.], 2008. p. 218–227.



- HÄGERSTRAND, T. What about people in regional science? **Papers in regional science**, Springer, v. 24, n. 1, p. 6–21, 1970.
- HARROWER, M. A look at the history and future of animated maps. **Cartographica: The International Journal for Geographic Information and Geovisualization**, University of Toronto Press, v. 39, n. 3, p. 33–42, 2004.
- HASSAN-MONTERO, Y.; HERRERO-SOLANA, V. Improving tag-clouds as visual information retrieval interfaces. In: CITESEER. **International conference on multidisciplinary information sciences and technologies**. [S. l.], 2006. p. 25–28.
- HAVRE, S.; HETZLER, E.; WHITNEY, P.; NOWELL, L. Themeriver: Visualizing thematic changes in large document collections. **IEEE transactions on visualization and computer graphics**, IEEE, v. 8, n. 1, p. 9–20, 2002.
- HERMAN, I.; MELANÇON, G.; MARSHALL, M. S. Graph visualization and navigation in information visualization: A survey. **IEEE Transactions on visualization and computer graphics**, IEEE, v. 6, n. 1, p. 24–43, 2000.
- HOLZINGER, A. Usability engineering methods for software developers. **Communications of the ACM**, ACM, v. 48, n. 1, p. 71–74, 2005.
- HU, Q.; WANG, H. A framework for traffic accident scene investigation with gps vrs, road database and stereo vision integration. In: IEEE. **Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011 International Workshop on**. [S. l.], 2011. p. 1–5.
- HUANG, Q.; WONG, D. W. Modeling and visualizing regular human mobility patterns with uncertainty: An example using twitter data. **Annals of the Association of American Geographers**, Taylor & Francis, v. 105, n. 6, p. 1179–1197, 2015.
- HUANG, X.; DU, X. Achieving big data privacy via hybrid cloud. In: IEEE. **2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)**. [S. l.], 2014. p. 512–517.
- HURTER, C.; ERSOY, O.; FABRIKANT, S. I.; KLEIN, T. R.; TELEA, A. C. Bundled visualization of DynamicGraph and trail data. **IEEE TVCG**, v. 20, n. 8, p. 1141–1157, 2014.
- HWANG, S.-Y.; LIU, Y.-H.; CHIU, J.-K.; LIM, E.-P. Mining mobile group patterns: A trajectory-based approach. In: SPRINGER. **Pacific-Asia Conference on Knowledge Discovery and Data Mining**. [S. l.], 2005. p. 713–718.
- IVANOV, Y.; WREN, C.; SOROKIN, A.; KAUR, I. Visualizing the history of living spaces. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 13, n. 6, p. 1153–1160, 2007.
- JASPERS, M. W.; STEEN, T.; BOS, C. V. D.; GEENEN, M. The think aloud method: a guide to user interface design. **International journal of medical informatics**, Elsevier, v. 73, n. 11–12, p. 781–795, 2004.
- JERDING, D. F.; STASKO, J. T. The information mural: A technique for displaying and navigating large information spaces. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 4, n. 3, p. 257–271, 1998.

JEUNG, H.; LIU, Q.; SHEN, H. T.; ZHOU, X. A hybrid prediction model for moving objects. In: IEEE. **Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on**. [S. l.], 2008. p. 70–79.

JEUNG, H.; SHEN, H. T.; ZHOU, X. Mining trajectory patterns using hidden markov models. In: SPRINGER. **International Conference on Data Warehousing and Knowledge Discovery**. [S. l.], 2007. p. 470–480.

JOHANSSON, J.; LJUNG, P.; JERN, M.; COOPER, M. Revealing structure within clustered parallel coordinates displays. In: IEEE. **IEEE Symposium on Information Visualization, 2005. INFOVIS 2005**. [S. l.], 2005. p. 125–132.

KAPLER, T.; WRIGHT, W. Geotime information visualization. **Information visualization**, SAGE Publications Sage UK: London, England, v. 4, n. 2, p. 136–146, 2005.

KARAGIANNIS, T.; BOUDEEC, J.-Y. L.; VOJNOVIC, M. Power law and exponential decay of intercontact times between mobile devices. **IEEE Transactions on Mobile Computing**, IEEE, v. 9, n. 10, p. 1377–1390, 2010.

KEIM, D.; ANDRIENKO, G.; FEKETE, J.-D.; GÖRG, C.; KOHLHAMMER, J.; MELANÇON, G. Visual analytics: Definition, process, and challenges. In: **Information visualization**. [S. l.]: Springer, 2008. p. 154–175.

KEIM, D. A. Designing pixel-oriented visualization techniques: Theory and applications. **IEEE Transactions on visualization and computer graphics**, IEEE, v. 6, n. 1, p. 59–78, 2000.

KLEIN, G.; MOON, B.; HOFFMAN, R. R. Making sense of sensemaking 2: A macrocognitive model. **IEEE Intelligent systems**, IEEE, v. 21, n. 5, p. 88–92, 2006.

KOVED, L.; SELKER, T. **Room with a view (RWAV): A metaphor for interactive computing**. [S. l.]: Citeseer, 1999.

KRAAK, M.-J. The space-time cube revisited from a geovisualization perspective. In: **Proc. 21st International Cartographic Conference**. [S. l.]: ICACI, 2003. p. 1988–1996.

KRUG, S. **Don't make me think!: a common sense approach to Web usability**. [S. l.]: Pearson Education India, 2000.

KWAN, M.-P. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 8, n. 1-6, p. 185–203, 2000.

LAUNIUS, R. D.; MCCURDY, H. E. **NASA Spaceflight: A History of Innovation**. [S. l.]: Springer, 2017.

LEE, D.-T. On k-nearest neighbor voronoi diagrams in the plane. **IEEE transactions on computers**, IEEE, v. 100, n. 6, p. 478–487, 1982.

LEE, J.-G.; HAN, J.; LI, X.; GONZALEZ, H. Traiclass: trajectory classification using hierarchical region-based and trajectory-based clustering. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 1, n. 1, p. 1081–1094, 2008.

- LEE, J.-G.; HAN, J.; WHANG, K.-Y. Trajectory clustering: a partition-and-group framework. In: ACM. **Proceedings of the 2007 ACM SIGMOD international conference on Management of data**. [S. l.], 2007. p. 593–604.
- LEI, P.-R.; LI, S.-C.; PENG, W.-C. Qs-stt: Quadsection clustering and spatial-temporal trajectory model for location prediction. **Distributed and Parallel Databases**, Springer, v. 31, n. 2, p. 231–258, 2013.
- LEITE, A. J. M.; SANTOS, E.; VIDAL, C. A.; MACÊDO, J. A. F. D. Visual analysis of predictive suffix trees for discovering movement patterns and behaviors. In: IEEE. **2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. [S. l.], 2017. p. 103–110.
- LEITE, A. M.; SANTOS, E.; VIDAL, C.; ROCHA, C. Empregando análise visual e sensemaking no ensino de predictive suffix trees. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S. l.]: SBC, 2018. p. 1043.
- LEVKOWITZ, H.; HERMAN, G. T. Color scales for image data. **IEEE Computer Graphics and Applications**, IEEE, n. 1, p. 72–80, 1992.
- LI, Y.; THOMASON, M.; PARKER, L. E. Detecting time-related changes in wireless sensor networks using symbol compression and probabilistic suffix trees. In: IEEE. **2010 IEEE/RSJ International Conference on Intelligent Robots and Systems**. [S. l.], 2010. p. 2946–2951.
- LIAMPUTTONG, P.; EZZY, D. **Qualitative Research Methods**. [S. l.]: Oxford University Press, 2005. (Qualitative Research Methods). ISBN 9780195517446.
- LIU, J.; WOLFSON, O.; YIN, H. Extracting semantic location from outdoor positioning systems. In: IEEE. **7th International Conference on Mobile Data Management (MDM'06)**. [S. l.], 2006. p. 73–73.
- LOPEZ, C.; LECLERCQ, L.; KRISHNAKUMARI, P.; CHIABAUT, N.; LINT, H. V. Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. **Scientific Reports**, Nature Publishing Group, v. 7, n. 1, p. 14029, 2017.
- LU, C.-T.; LEI, P.-R.; PENG, W.-C.; SU, J. A framework of mining semantic regions from trajectories. In: SPRINGER. **International Conference on Database Systems for Advanced Applications**. [S. l.], 2011. p. 193–207.
- MACEACHREN, A. M. **How maps work: representation, visualization, and design**. [S. l.]: Guilford Press, 2004.
- MARBLE, D.; GOU, Z.; LIU, L.; SAUNDERS, J. Recent advances in the exploratory analysis of interregional flows in space and time. In: **Innovations in GIS**. [S. l.]: CRC Press, 1997. p. 93–106.
- MARTÍNEZ, M. Travelling in branching time. **Disputatio**, Centro de Filosofia da Universidade de Lisboa, v. 4, n. 31, p. 271–287, 2011.
- MATKOVIC, K.; GRACANIN, D.; KONYHA, Z.; HAUSER, H. Color linesview: An approach to visualization of families of function graphs. In: IEEE. **Information Visualization, 2007. IV'07. 11th International Conference**. [S. l.], 2007. p. 59–64.

MCCREIGHT, E. M. A space-economical suffix tree construction algorithm. **Journal of the ACM (JACM)**, Citeseer, v. 23, n. 2, p. 262–272, 1976.

MOELLERING, H. The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. **Accident Analysis & Prevention**, Elsevier, v. 8, n. 4, p. 215–227, 1976.

MONREALE, A.; PINELLI, F.; TRASARTI, R.; GIANNOTTI, F. Wherenext: a location predictor on trajectory pattern mining. In: ACM. **Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S. l.], 2009. p. 637–646.

MONTOLIU, R.; GATICA-PEREZ, D. Discovering human places of interest from multimodal mobile phone data. In: ACM. **Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia**. [S. l.], 2010. p. 12.

MOORE, C. **Creating scientists: Teaching and assessing science practice for the NGSS**. [S. l.]: Routledge, 2017.

MORZY, M. Mining frequent trajectories of moving objects for location prediction. In: SPRINGER. **International Workshop on Machine Learning and Data Mining in Pattern Recognition**. [S. l.], 2007. p. 667–680.

NG, O.-L.; SINCLAIR, N.; DAVIS, B. Drawing off the page: How new 3d technologies provide insight into cognitive and pedagogical assumptions about mathematics. **The Mathematics Enthusiast**, v. 15, n. 3, p. 563–578, 2018.

NIELSEN, J. **Usability engineering**. [S. l.]: Elsevier, 1994.

OSWALD, D. Dynamic sense-making in use processes of digital products—a semiotic approach to user interface design. **Proceedings of AISDR 2013**, p. 1652–1663, 2013.

PANAGIOTAKIS, C.; PELEKIS, N.; KOPANAKIS, I.; RAMASSO, E.; THEODORIDIS, Y. Segmentation and sampling of moving object trajectories based on representativeness. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 24, n. 7, p. 1328–1343, 2011.

PATTERSON, D. J.; LIAO, L.; GAJOS, K.; COLLIER, M.; LIVIC, N.; OLSON, K.; WANG, S.; FOX, D.; KAUTZ, H. Opportunity knocks: A system to provide cognitive assistance with transportation services. In: SPRINGER. **International Conference on Ubiquitous Computing**. [S. l.], 2004. p. 433–450.

PEARSON, K. Liii. on lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901.

PEDROSA, I. Da cor à cor inexistente. In: **Da cor à cor inexistente**. São Paulo: SENAC, 1995.

PHAN, D.; XIAO, L.; YEH, R.; HANRAHAN, P. Flow map layout. In: IEEE. **IEEE Symposium on Information Visualization, 2005. INFOVIS 2005**. [S. l.], 2005. p. 219–224.

PIROLI, P.; CARD, S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: MCLEAN, VA, USA. **Proceedings of international conference on intelligence analysis**. [S. l.], 2005. v. 5, p. 2–4.

PLAISANT, C.; MILASH, B.; ROSE, A.; WIDOFF, S.; SHNEIDERMAN, B. **LifeLines: visualizing personal histories**. [S. l.], 1995.

POLKINGHORNE, D. E. Language and meaning: Data collection in qualitative research. **Journal of counseling psychology**, American Psychological Association, v. 52, n. 2, p. 137, 2005.

PURCHASE, H. Which aesthetic has the greatest effect on human understanding? In: SPRINGER. **International Symposium on Graph Drawing**. [S. l.], 1997. p. 248–261.

RESNICK, S. I. **Heavy-tail phenomena: probabilistic and statistical modeling**. [S. l.]: Springer Science & Business Media, 2007.

RHODES, P.; MCDONALD, R.; CAMPBELL, S.; DAKER-WHITE, G.; SANDERS, C. Sensemaking and the co-production of safety: a qualitative study of primary medical care patients. **Sociology of health & illness**, Wiley Online Library, v. 38, n. 2, p. 270–285, 2016.

RINZIVILLO, S.; PEDRESCHI, D.; NANNI, M.; GIANNOTTI, F.; ANDRIENKO, N.; ANDRIENKO, G. Visually driven analysis of movement data by progressive clustering. **Information Visualization**, SAGE Publications Sage UK: London, England, [S. l.], v. 7, n. 3-4, p. 225–239, 2008.

ROBERTSON, N.; SEYMOUR, P. D. Graph minors. ii. algorithmic aspects of tree-width. **Journal of algorithms**, Elsevier, v. 7, n. 3, p. 309–322, 1986.

ROCHA, C. L. Tpred: um framework espaço-temporal de predição de localização. **Dissertação, Universidade Federal do Ceará, Departamento de Computação, Brasil**, 2016.

ROCHA, C. L.; BRILHANTE, I. R.; LETTICH, F.; MACEDO, J. A. F. D.; RAFFAETÀ, A.; ANDRADE, R.; ORLANDO, S. Tpred: a spatio-temporal location predictor framework. In: ACM. **Proceedings of the 20th International Database Engineering & Applications Symposium**. [S. l.], 2016. p. 34–42.

RON, D.; SINGER, Y.; TISHBY, N. Learning probabilistic automata with variable memory length. In: ACM. **Proceedings of the seventh annual conference on Computational learning theory**. [S. l.], 1994. p. 35–46.

RUSSELL, D. M.; STEFIK, M. J.; PIROLI, P.; CARD, S. K. The cost structure of sensemaking. In: ACM. **Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems**. [S. l.], 1993. p. 269–276.

SCELLATO, S.; MUSOLESI, M.; MASCOLO, C.; LATORA, V.; CAMPBELL, A. T. Nextplace: a spatio-temporal prediction framework for pervasive systems. In: SPRINGER. **International Conference on Pervasive Computing**. [S. l.], 2011. p. 152–169.

SCHULZ, H.-J. Treevis. net: A tree visualization reference. **IEEE Computer Graphics and Applications**, IEEE, v. 31, n. 6, p. 11–15, 2011.

SCHUYTEMA, P. **Design de games: uma abordagem prática**. [S. l.]: Cengage Learning, 2008.

SHIN, H.; PARK, G.; HAN, J. Tablorer—an interactive tree visualization system for tablet pcs. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S. l.], 2011. v. 30, n. 3, p. 1131–1140.

- SHUM, S. B.; CRICK, R. D. Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In: ACM. **Proceedings of the 2nd international conference on learning analytics and knowledge**. [S. l.], 2012. p. 92–101.
- SIEMENS, G.; LONG, P. Penetrating the fog: Analytics in learning and education. **EDUCAUSE review**, ERIC, v. 46, n. 5, p. 30, 2011.
- SILVA, T. L. C. da; ZEITOUNI, K.; MACÊDO, J. A. de. Online clustering of trajectory data stream. In: IEEE. **2016 17th IEEE International Conference on Mobile Data Management (MDM)**. [S. l.], 2016. v. 1, p. 112–121.
- SIOW, E.; TIROPANIS, T.; HALL, W. Analytics for the internet of things: A survey. **ACM Computing Surveys (CSUR)**, ACM, v. 51, n. 4, p. 74, 2018.
- SLINGSBY, A.; DYKES, J.; WOOD, J.; CLARKE, K. Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets. In: IEEE. **2007 11th International Conference Information Visualization (IV'07)**. [S. l.], 2007. p. 497–504.
- SONG, J.; TANG, E. Y.; LIU, L. User behavior pattern analysis and prediction based on mobile phone sensors. In: SPRINGER. **IFIP International Conference on Network and Parallel Computing**. [S. l.], 2010. p. 177–189.
- SPRETKE, D.; BAK, P.; JANETZKO, H.; KRANSTAUBER, B.; MANSMANN, F.; DAVIDSON, S. Exploration through enrichment: a visual analytics approach for animal movement. In: ACM. **Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems**. [S. l.], 2011. p. 421–424.
- SULLIVAN, C. J.; BURBANK, L. Apollo experience report: Real-time display system. In: **Apollo experience report: Real-time display system**. [S. l.]: NASA, 1976.
- TAYLOR, S. J.; BOGDAN, R.; DEVAULT, M. **Introduction to qualitative research methods: A guidebook and resource**. [S. l.]: John Wiley & Sons, 2015.
- THOMASON, R. H. Indeterminist time and truth-value gaps<sup>1</sup>. **Theoria**, Wiley Online Library, v. 36, n. 3, p. 264–281, 1970.
- TOMINSKI, C.; SCHUMANN, H.; ANDRIENKO, G.; ANDRIENKO, N. Stacking-based visualization of trajectory attribute data. **IEEE TVCG**, IEEE, v. 18, n. 12, p. 2565–2574, 2012.
- TRENDS, G. **Refugees, Asylum-seekers, Returnees, Internally Displaced and Stateless Persons.(2010)**. United Nations High Commissioner for Refugees. 2009.
- TUFTE, E. R. **The visual display of quantitative information**. [S. l.]: Graphics press Cheshire, CT, 2001. v. 2.
- TUFTE, E. R.; GOELER, N. H.; BENSON, R. **Envisioning information**. [S. l.]: Graphics press Cheshire, CT, 1990. v. 126.
- UKKONEN, E. On-line construction of suffix trees. **Algorithmica**, Springer, v. 14, n. 3, p. 249–260, 1995.
- VASILIEV, I. Mapping time. **Cartographica: The international journal for geographic information and geovisualization**, University of Toronto Press, v. 34, n. 2, p. 1–51, 1997.

VERBERT, K.; DUVAL, E.; KLERKX, J.; GOVAERTS, S.; SANTOS, J. L. Learning analytics dashboard applications. **American Behavioral Scientist**, Sage Publications Sage CA: Los Angeles, CA, v. 57, n. 10, p. 1500–1509, 2013.

VIÉGAS, F. B.; WATTENBERG, M.; DAVE, K. Studying cooperation and conflict between authors with history flow visualizations. In: ACM. **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S. l.], 2004. p. 575–582.

WANG, Z.; LU, M.; YUAN, X.; ZHANG, J.; WETERING, H. V. D. Visual traffic jam analysis based on trajectory data. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 19, n. 12, p. 2159–2168, 2013.

WARBURTON, E. C.; LAUGHLIN, G. A performed solution to the pythagorean problem: The three bodies project. **Leonardo**, MIT Press, p. 1–12, 2018.

WARE, C.; ARSENAULT, R.; PLUMLEE, M.; WILEY, D. Visualizing the underwater behavior of humpback whales. **IEEE Computer Graphics and Applications**, IEEE, v. 26, n. 4, p. 14–18, 2006.

WIKIPEDIA, t. f. e. **IBM 2250**. 2019. Disponível em: <[https://en.wikipedia.org/wiki/IBM\\_2250](https://en.wikipedia.org/wiki/IBM_2250)>.

WILKINSON, L.; FRIENDLY, M. The history of the cluster heat map. **The American Statistician**, Taylor & Francis, v. 63, n. 2, p. 179–184, 2009.

WISE, A. F.; VYTASEK, J. M.; HAUSKNECHT, S.; ZHAO, Y. Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. **Online Learning**, ERIC, v. 20, n. 2, p. 155–182, 2016.

WOOD, J.; DYKES, J.; SLINGSBY, A. Visualisation of origins, destinations and flows with od maps. **The Cartographic Journal**, Taylor & Francis, [S. l.], v. 47, n. 2, p. 117–129, 2010.

WOOD, J.; DYKES, J.; SLINGSBY, A.; CLARKE, K. Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. **IEEE transactions on visualization and computer graphics**, IEEE, v. 13, n. 6, p. 1176–1183, 2007.

WULLSCHLAGER, J. **Hans Christian Andersen: the life of a storyteller**. [S. l.]: University of Chicago Press, 2002.

WURMAN, R. S.; WURMAN, R. S. **Information anxiety: What to do when information doesn't tell you what you need to know**. [S. l.]: Bantam New York, 1990.

XUE, A. Y.; ZHANG, R.; ZHENG, Y.; XIE, X.; HUANG, J.; XU, Z. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In: IEEE. **2013 IEEE 29th international conference on data engineering (ICDE)**. [S. l.], 2013. p. 254–265.

YIN, J.; GAO, Y.; DU, Z.; WANG, S. Exploring multi-scale spatiotemporal twitter user mobility patterns with a visual-analytics approach. **ISPRS International Journal of Geo-Information**, Multidisciplinary Digital Publishing Institute, v. 5, n. 10, p. 187, 2016.

YING, J. J.-C.; LEE, W.-C.; TSENG, V. S. Mining geographic-temporal-semantic patterns in trajectories for location prediction. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM, v. 5, n. 1, p. 2, 2013.

YOUNG, T. Optimizing points-of-visibility pathfinding. **Game Programming Gems**, v. 2, p. 324–329, 2001.

YUAN, J.; ZHENG, Y.; XIE, X.; SUN, G. Driving with knowledge from the physical world. In: ACM. **Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S. l.], 2011. p. 316–324.

YUAN, J.; ZHENG, Y.; ZHANG, C.; XIE, W.; XIE, X.; SUN, G.; HUANG, Y. T-drive: driving directions based on taxi trajectories. In: ACM. **Proceedings of the 18th SIGSPATIAL International conference on advances in geographic information systems**. [S. l.], 2010. p. 99–108.

ZENG, W.; FU, C.-W.; ARISONA, S. M.; QU, H. Visualizing interchange patterns in massive movement data. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S. l.], 2013. v. 32, p. 271–280.

ZENG, W.; ZHONG, C.; ANWAR, A.; ARISONA, S. M.; MCLOUGHLIN, I. V. Metrobuzz: Interactive 3d visualization of spatiotemporal data. In: IEEE. **2012 International Conference on Computer & Information Science (ICCIS)**. [S. l.], 2012. v. 1, p. 143–147.

ZHENG, K.; TRAJCEVSKI, G.; ZHOU, X.; SCHEUERMANN, P. Probabilistic range queries for uncertain trajectories on road networks. In: ACM. **Proceedings of the 14th International Conference on Extending Database Technology**. [S. l.], 2011. p. 283–294.

ZHOU, H.; XU, P.; YUAN, X.; QU, H. Edge bundling in information visualization. **Tsinghua Science and Technology**, Citeseer, v. 18, n. 2, p. 145–156, 2013.



## APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO UTILIZADO NO PRIMEIRO TESTE COM USUÁRIOS



UNIVERSIDADE FEDERAL DO CEARÁ  
INSTITUTO UFC VIRTUAL

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Convidamos o(a) Sr(a) para participar da Pesquisa **Uma Proposta de Solução para Análise Visual de Árvores Probabilísticas de Predição de Trajetórias**, sob a responsabilidade do pesquisador **Antonio José Melo Leite Júnior**, a qual pretende avaliar o uso de uma aplicação específica para análise visual de árvores probabilísticas de predição de trajetórias.

Sua participação é voluntária e se dará por meio da realização de tarefas empregando a aplicação supracitada. Se você aceitar participar, estará contribuindo para a geração de importantes dados de uso para análise.

Se depois de consentir em sua participação o Sr (a) desistir de continuar participando, tem o direito e a liberdade de retirar seu consentimento em qualquer fase da pesquisa, seja antes ou depois da coleta dos dados, independente do motivo e sem nenhum prejuízo a sua pessoa. O(a) Sr(a) não terá nenhuma despesa e também não receberá nenhuma remuneração. Os resultados da pesquisa serão analisados e publicados, mas sua identidade não será divulgada, sendo guardada em sigilo. Para qualquer outra informação, o(a) Sr(a) poderá entrar em contato com o pesquisador através do e-mail [melojr@virtual.ufc.br](mailto:melojr@virtual.ufc.br), pelo telefone (85) 98805.7400, ou contatar diretamente o **Instituto Universidade Virtual/Universidade Federal do Ceará, Av. Humberto Monte, s/n, bloco 901, 1º andar, CEP: 60.440-554, Fone (85)3366-9457.**

#### Consentimento Pós-Infomação

Eu, \_\_\_\_\_, fui informado sobre o que o pesquisador pretende fazer e o motivo de necessitar minha colaboração, e entendi a explicação. Por isso, eu concordo em participar da Pesquisa, sabendo que não vou receber qualquer tipo de pagamento e que posso sair quando quiser. Este documento é emitido em duas vias que serão ambas assinadas por mim e pelo pesquisador, ficando uma via com cada um de nós.

\_\_\_\_\_  
Assinatura do participante

Data: \_\_\_/\_\_\_/\_\_\_\_

\_\_\_\_\_  
Assinatura do Pesquisador Responsável

## APÊNDICE B – QUESTIONÁRIO UTILIZADO NO SEGUNDO TESTE COM USUÁRIOS

**1. Caso concorde em participar do teste, favor preencher os dados abaixo:**

*Termo de Consentimento Livre e Esclarecido*

Sua participação é voluntária e se dará por meio da realização de tarefas de **análise visual de árvores probabilísticas para avaliação do comportamento de pessoas**. Caso você concorde em participar, estará contribuindo para a geração de importantes dados de uso para avaliação. Os resultados da pesquisa serão analisados e publicados, mas sua identidade não será divulgada, sendo guardada em sigilo.

*Consentimento Pós-Informação*

Eu, \_\_\_\_\_ (nome completo e legível), fui informado(a) sobre o que a pesquisa pretende e o motivo de necessitar minha colaboração. Por isso, concordo em participar do teste, sabendo que não vou receber qualquer tipo de pagamento e que posso solicitar o não uso de meus dados a qualquer momento, caso julgue necessário.

\_\_\_\_\_  
Assinatura do participante

\_\_\_\_\_  
Local

\_\_\_\_\_  
Data

**2. Responder e justificar brevemente as questões a seguir:**

*Usuário 1*

Onde a pessoa mora? \_\_\_\_\_

Qual a pessoa frequenta? \_\_\_\_\_

Onde a pessoa trabalha? \_\_\_\_\_

A que horas a pessoa vai ao RU? \_\_\_\_\_

*Usuário 2*

Onde a pessoa mora? \_\_\_\_\_

Qual a pessoa frequenta? \_\_\_\_\_

Onde a pessoa trabalha? \_\_\_\_\_

A que horas a pessoa vai ao RU? \_\_\_\_\_

*Usuário 3*

Onde a pessoa mora? \_\_\_\_\_

Qual a pessoa frequenta? \_\_\_\_\_

Onde a pessoa trabalha? \_\_\_\_\_

A que horas a pessoa vai ao RU? \_\_\_\_\_

*Metodologia*

Como você obteve as respostas acima? \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**3. Comentar abaixo sobre a realização do experimento e sua participação no mesmo:**

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**4. Opinar brevemente sobre a utilidade das seguintes ferramentas:**

POI \_\_\_\_\_

ROI \_\_\_\_\_

Bundles \_\_\_\_\_

**5. Comentar abaixo sobre a realização do experimento e sua participação no mesmo:**

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

## APÊNDICE C – QUESTIONÁRIO UTILIZADO NO TERCEIRO TESTE COM USUÁRIOS

### FORMULÁRIO DE PARTICIPAÇÃO EM PESQUISA EXPERIMENTAL

#### 1. Caso concorde em participar do teste, favor preencher os dados abaixo:

##### *Termo de Consentimento Livre e Esclarecido*

Sua participação é voluntária e se dará por meio da realização de tarefas de **análise visual de árvores probabilísticas para avaliação do comportamento de pessoas**. Caso você concorde em participar, estará contribuindo para a geração de importantes dados de uso para avaliação. Os resultados da pesquisa serão analisados e publicados, mas sua identidade não será divulgada, sendo guardada em sigilo.

##### *Consentimento Pós-Informação*

Eu, \_\_\_\_\_ (nome completo e legível), fui informado(a) sobre o que a pesquisa pretende e o motivo de necessitar minha colaboração. Por isso, concordo em participar do teste, sabendo que não vou receber qualquer tipo de pagamento e que posso solicitar o não uso de meus dados a qualquer momento, caso julgue necessário.

\_\_\_\_\_  
Assinatura do participante

\_\_\_\_\_  
Local

\_\_\_\_\_  
Data

#### 2. Responder e justificar brevemente as questões a seguir:

##### *Taxi 1*

Quais as principais rotas (localizações e extensões) utilizadas pelo taxi? \_\_\_\_\_

As viagens ocorrem em quais períodos do dia? \_\_\_\_\_

É possível identificar algum ponto de parada do taxi? \_\_\_\_\_

Quais eventos/fatos adicionais podem ser identificados? \_\_\_\_\_

##### *Taxi 2*

Quais as principais rotas (localizações e extensões) utilizadas pelo taxi? \_\_\_\_\_

As viagens ocorrem em quais períodos do dia? \_\_\_\_\_

É possível identificar algum ponto de parada do taxi? \_\_\_\_\_

Quais eventos/fatos adicionais podem ser identificados? \_\_\_\_\_

##### *Taxi 3*

Quais as principais rotas (localizações e extensões) utilizadas pelo taxi? \_\_\_\_\_

As viagens ocorrem em quais períodos do dia? \_\_\_\_\_

É possível identificar algum ponto de parada do taxi? \_\_\_\_\_

Quais eventos/fatos adicionais podem ser identificados? \_\_\_\_\_

#### 3. Comentar sobre o método utilizado para a obtenção das respostas, incluindo o acesso a serviços online:

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

#### 4. Comentar sobre a realização do experimento e sua participação no mesmo:

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_