

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Informática

Dissertação de Mestrado

Data Warehouse de Trajetórias: um Modelo
Semântico com Suporte à Agregação por Direção
dos Movimentos

Carlos Augusto de Santana Almeida

Orientador
Prof. Dr. Ulrich Schiel

Campina Grande – PB
Novembro – 2010

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Informática

Data Warehouse de Trajetórias: um Modelo
Semântico com Suporte à Agregação por Direção
dos Movimentos

Carlos Augusto de Santana Almeida
<carlos@dsc.ufcg.edu.br>

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Sistemas de Informação e Banco de Dados

Prof. Dr. Ulrich Schiel
<ulrich@dsc.ufcg.edu.br>
(Orientador)

Campina Grande, Paraíba, Brasil
©Carlos Augusto de Santana Almeida, 29/11/2010

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

A447d

2010 Almeida, Carlos Augusto de Santana

Data Warehouse de Trajetórias: um Modelo Semântico com Suporte à Agregação por Direção dos Movimentos / Carlos Augusto de Santana Almeida - Campina Grande, 2010.

120 f.: il. col.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Campina Grande, Centro de Informática e Engenharia Elétrica.

Orientador: Prof. Dr. Ulrich Schiel

Referências

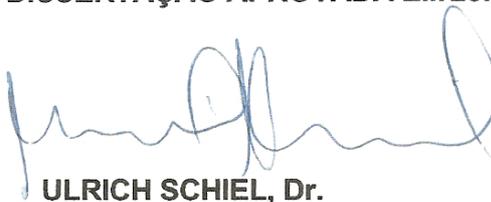
1. Data Warehouse de Trajetórias; 2. OLAP - On-Line Analytical Processing; 3. Objetos Móveis; 4. Banco de Dados. I. Título.

CDU - 004.65 (043)

**"DATA WAREHOUSE DE TRAJETÓRIAS: UM MODELO SEMÂNTICO COM SUPORTE
A AGREGAÇÃO POR DIREÇÃO DOS MOVIMENTOS"**

CARLOS AUGUSTO DE SANTANA ALMEIDA

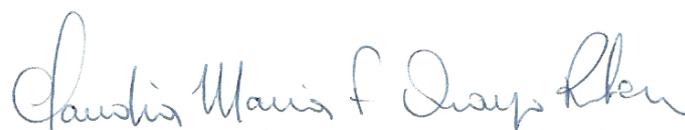
DISSERTAÇÃO APROVADA EM 29.11.2010



**ULRICH SCHIEL, Dr.
Orientador(a)**



**CARLOS EDUARDO SANTOS PIRES, Dr.
Examinador(a)**



**CLAUDIA MARIA FERNANDES ARAUJO RIBEIRO, Dr^a
Examinador(a)**

CAMPINA GRANDE - PB

Resumo

A popularização dos dispositivos móveis cientes de localização, tais como telefones celulares e GPSs, possibilitou o monitoramento em larga escala dos objetos móveis que transportam esses dispositivos, tais como pessoas, carros e aviões. Esse monitoramento tem como resultado a geração de uma grande massa de dados bruta sobre as trajetórias desses objetos. A análise desse tipo de dado permite descobrir padrões de comportamento que podem ser explorados em uma grande número de aplicações, como por exemplo, no gerenciamento de tráfego urbano, no estudo das trajetórias de turistas em uma viagem, para identificar as rotas migratórias de aves, entre outras aplicações.

Para transformar essa massa de dados bruta em informações úteis para tomada de decisão e descoberta do conhecimento, uma forma adequada seria disponibilizá-la em um Data Warehouse (DW), um banco de dados otimizado para lidar com grandes volumes de dados de forma eficiente. Para dados convencionais, as tecnologias de DW têm sido usadas com sucesso durante décadas. Entretanto, a natureza dos dados de trajetórias impõe certos desafios para construção e manutenção do DW, dentre eles: (i) a grande quantidade dos dados de trajetória consome muitos recursos de memória e processamento, tornando o tempo de execução das consultas demasiadamente longo, inviabilizando análise no estilo OLAP; (ii) o suporte oferecido pelas tecnologias de DW para trajetórias ainda está limitada ao armazenamento e recuperação dos dados de trajetórias, não é possível efetuar operações OLAP sobre eles como acontece com os dados espaciais; (iii) escassez de modelos, existem poucos trabalhos relacionados a modelagem multidimensional para trajetórias.

Com o objetivo de resolver parte desses problemas, neste trabalho é proposto: (i) um modelo multidimensional espacial capaz de proporcionar análise OLAP para trajetórias (roll-up para trajetórias), permitindo assim, analisar o comportamento dos objetos móveis *sobre e entre* as regiões no espaço e tempo; (ii) além disso, o modelo permite segmentar trajetórias em diversos componentes semânticos, os quais podem transportar informações que dão significado a trajetória; e (iii) proporcionar compactação de trajetórias, uma forma de reduzir a quantidade de dados armazenada.

Palavras chaves: Data Warehouse de Trajetórias; OLAP - *On-Line Analytical Processing*; Objetos Móveis; Banco de Dados.

Agradecimentos

Este trabalho só foi possível graças ao apoio direto e indireto de outras pessoas e entidades. Por isso, gostaria de agradecer:

A Deus por ter me dado a saúde necessária para vencer os obstáculos da vida.

Aos meus pais, pelo carinho, amor, confiança e pela dedicação de uma vida inteira de trabalho para me dar a educação necessária para que chegasse até aqui. E a minha irmã TaXanata (Marília), por aturar meu estresse e minhas lamúrias nos momentos difíceis.

Ao meu orientador, Prof. Ulrich Schiel, pela paciência e ensinamentos transmitidos ao longo da orientação. Ao Prof. Carlos Eduardo Pires por suas sugestões e ideias.

Aos meus grandes amigos e colegas de mestrado Farinha (Felipe), Preto Boss (Welflen), Negola (Gilson), e Fernando, pela momentos alegres e convivência diária no apartamento.

Aos novos amigos: Carlos Alberto, Danillo Bob (Danillo), Yuri, Halley, Barata (Flávio), Marquinhos (Marcos), Saulo de Tarso, Welmisson, Guiga (Guilherme) pela amizade e momentos de alegria tanto na universidade como fora dela (p.ex. quarta-insana).

Aos demais professores e funcionários da COPIN, em especial, à Aninha, pela gentileza e presteza.

A Capes, pelo apoio financeiro.

E a todos que me ajudaram a seguir o meu caminho e concluir este mestrado.

*“O rio atinge seus objetivos porque
aprendeu a contornar obstáculos.”*

Lao Tsé

Sumário

1	Introdução	1
1.1	Objetivos	4
1.2	Estrutura da Dissertação	5
2	Fundamentação Teórica	7
2.1	Trajетórias	7
2.2	Reconstrução de Trajetórias	12
2.3	Representação do Espaço Geográfico	14
2.3.1	Problema do Relacionamento Parcialmente Contido	15
2.4	Análise de Trajetórias	16
2.4.1	Vantagens oferecidas pelo uso de Trajetórias Semânticas	18
2.5	Formas de Armazenamento para Trajetórias	21
2.5.1	Baseada na sequencia de observações (modelo clássico)	21
2.5.2	Sumarizada em células	22
2.5.3	Na forma vetorial	23
2.6	Tecnologias de Banco de Dados	25
2.6.1	Data Warehouse (DW)	25
2.6.2	Data Warehouse Espacial (DWE)	28
3	Trabalhos Relacionados	30
3.1	Proposta de Braz et al. (2007)	31
3.1.1	Críticas	32

3.2	Proposta de Marketos et al. (2008)	34
3.2.1	Críticas	35
3.3	Proposta de Kuijpers; e Vaisman (2007)	37
3.3.1	Críticas	37
3.4	Proposta de Spaccapietra et al. (2008)	38
3.4.1	Críticas	39
3.5	Proposta de Gomez; Kuijpers; e Vaisman (2007)	40
3.5.1	Críticas	40
3.6	Proposta de Baltzer et al. (2008a)	41
3.6.1	Críticas	43
3.7	Considerações Finais	44
4	Um Modelo Semântico para DWTr com Suporte à Agregação por Direção dos Movimentos	49
4.1	Cenário de Aplicação	50
4.2	Modelo Proposto	53
4.2.1	Fatos	54
4.2.2	Dimensões	55
4.3	Carga de Dados	57
4.4	Roll-up para Trajetórias	59
5	Validação do Modelo	63
5.1	Estudo de Caso	63
5.1.1	Metodologia Utilizada	65
5.1.2	Esquema Conceitual para Gerenciamento de Tráfego Urbano	66
5.2	Base de Trajetórias	70
5.3	Consultas Exemplo	73
5.3.1	Consultas por Região	74
5.3.2	Consultas por Direção do Movimento	79

5.3.3 Consultas por Trajetória	83
5.4 Compactação dos Dados de Trajetória	90
5.5 Tamanho das Dimensões Direção do Movimento	93
6 Conclusão	95
6.1 Trabalhos Futuros	98

Lista de Figuras

2.1	Caminho espaço-temporal de um objeto móvel durante seu ciclo de vida. . .	9
2.2	Diagrama UML para o modelo conceitual de trajetórias proposto por Spaccapietra et al. (2008) (adaptado).	11
2.3	Etapas do processo de Reconstrução de Trajetórias	13
2.4	Trajectoria Interpolada usando a função de interpolação linear.	14
2.5	Espaço geográfico dividido por uma grade regular (Rinzivillo et al., 2008). .	15
2.6	Relacionamento Parcialmente Contido	16
2.7	Análise de trajetórias.	18
2.8	O mesmo conjunto de trajetórias mostrado SEM e COM informações geográficas de fundo, figuras (a) e (b), respectivamente (Palma, 2008, p. 16). . .	19
2.9	Exemplo de uma mesma consulta, resolvida em: (a) SEM o uso de informações semânticas (trajetórias brutas); (b) usando informações semânticas. . .	20
2.10	Objetos móveis se deslocando de um lugar para outro. Considerando em: (a) trajetórias brutas; (b) trajetórias semânticas (Bogorny; Kuijpers; e Alvares, 2009).	21
2.11	Agregação de dados de trajetória, armazenados: (a) de forma NÃO sumari- zada; (b) de forma sumarizada (ocorre o problema da contagem distinta). .	23
2.12	Trajéticas armazenadas na forma de Vetores Movimento	24
3.1	Modelo proposto por Braz et al. (2007). Em: (a) esquema estrela para DWTr. (b) hierarquia baseada em níveis de agregação.	32
3.2	Esquema estrela apresentado por Marketos et al. (2008).	35

3.3	Operador medoid em ação.	42
3.4	Agrupamento em: (a) por Intersecção; e (b) por Sobreposição (Baltzer et al., 2008a) (adaptado).	43
3.5	Tabela de Características dos Trabalhos Relacionados.	48
4.1	Cenário de Aplicação Exemplo.	52
4.2	Sumarização das observações de uma trajetória.	55
4.3	Modelo Proposto ilustrado por um diagrama UML.	56
4.4	Carga de dados para a trajetória H1::C1	60
4.5	Algoritmo da Carga de Dados.	61
4.6	Representação por direção do movimento entre regiões adjacentes, para a trajetória H1::C1, em: (a) por bairro e por RoI; (b) células de H1::C1 e sua ligação com a dimensão DirMovAdjDim.	62
5.1	Esquema Conceitual para Gerenciamento de Tráfego Urbano.	67
5.2	Hierarquias para as dimensões mostradas na Figura 5.1.	68
5.3	Configuração da Base de Trajetórias gerada.	72
5.4	Sintetizador de Trajetórias desenvolvido.	73
5.5	Agregação por região, em termos de: (i) Número de Veículos; (ii) Velocidade Média.	77
5.6	Velocidade Média por região da cidade em detalhes, dentro de uma janela espacial, sobre diferentes formas de visualização.	78
5.7	Deslocamento dos veículos entre <i>bairros adjacentes</i> por intervalo de tempo.	81
5.8	Direção dos movimentos por célula da grade regular.	82
5.9	Deslocamento dos veículos entre bairros por intervalo de tempo.	84
5.10	Diferentes representações para uma mesma trajetória, obtidas através da agregação de seus movimentos entre regiões adjacentes (roll-up para trajetórias).	85

5.11 Diferentes representações para uma mesma trajetória, obtidas através da agregação de seus movimentos entre regiões no tempo (roll-up para trajetórias).	87
5.12 Representação da trajetória por direção dos movimentos entre paradas (roll-up para trajetórias).	89
5.13 Tamanho dos fatos parada e movimento, usando o modelo proposto, <i>variando apenas o tamanho das células da grade</i>	91
5.14 Tamanho dos fatos parada e movimento, usando o modelo proposto, <i>variando apenas o intervalo de captura entre as observações</i>	92
5.15 Mapa da cidade de Aracaju/Sergipe dividido por grades regulares sobre diferentes níveis de granularidade.	92
5.16 Tamanho das dimensões Direção do Movimento entre Regiões Adjacentes (DirMovAdjDim).	93
5.17 Tamanho das dimensões Direção do Movimento entre Regiões (DirMovDim) no tempo.	94

Lista de Símbolos

BD	B anco de D ados
BDET	B anco de D ados E spaço- T emporal
BDOM	B anco de D ados de O bjeto M óveis
DW	D ata W arehouse
DWE	D ata W arehouse E spacial
DWTr	D ata W arehouse de T rajeto r ias
ETL	<i>Extract, Transform, and Load</i> - Extração, Transformação, e Carga
JSON	<i>JavaScript Object Notation</i> - Notação de Objetos JavaScript
KML	<i>Keyhole Markup Language</i>
OGC	<i>Open Geospatial Consortium</i> - Consórcio Geoespacial Livre
OLAP	<i>On-Line Analytical Processing</i> - Processamento Analítico Online
OLTP	<i>On-Line Transaction Processing</i> - Processamento de Transações Online
PGT	P ólos G eradores de T ráfego
RoI	<i>Region of Interest</i> - Região de Interesse
SGBD	Sistema de G erenciamento de B ancos de D ados
SQL	<i>Structured Query Language</i> - Linguagem de Consulta Estruturada
TrOLAP	<i>Trajectory OLAP</i> - OLAP para Trajetórias
XML	<i>eXtensible Markup Language</i> - Linguagem de Marcação Extensível

Capítulo 1

Introdução

A popularização dos dispositivos móveis cientes de localização (*location-aware mobile devices*), tais como telefones celulares e GPS (*Global Positioning System*, Sistema de Posicionamento Global), possibilitou o monitoramento em larga escala dos objetos móveis que transportam esses dispositivos tais como, pessoas, carros e aviões. Esse monitoramento tem como resultado a geração de grandes quantidades de dados sobre as trajetórias desses objetos (Spaccapietra et al., 2008). A análise desse tipo de dado permite descobrir padrões de comportamento que podem ser explorados em uma grande variedade de domínios (Orlando et al., 2007).

Por exemplo, (i) no gerenciamento de tráfego urbano, a medida número de veículos que atravessam um cruzamento é calculada por um sensor que conta os veículos que passam por ele. Essa medida poderia ser substituída por informações detalhadas sobre as trajetórias de cada veículo, incluindo sua origem-destino, rota utilizada, velocidade em cada trecho da trajetória, paradas realizadas, entre outras informações. De forma similar, os dados de trajetórias podem ser usados: (ii) no gerenciamento do transporte público, para melhorar a distribuição das linhas de ônibus; (iii) no estudo das trajetórias de turistas em uma viagem, para descobrir os locais mais visitados por eles; e no estudo da migração de pássaros, para identificar as rotas migratórias, entre outras aplicações.

Para se ter uma noção da importância da análise de trajetórias na economia, considere

as seguintes observações: (i) segundo a Fundação Getúlio Vargas¹ (FGV), o prejuízo gerado pelos congestionamentos na cidade de São Paulo em 2008 foi de aproximadamente R\$ 33 bilhões, o que equivale 20% do que a cidade produz; (ii) de acordo com estimativas da Comissão Europeia de Transportes², o setor de transportes movimenta cerca de 5,4 bilhões de euros por ano, o que representa 13,8% do PIB mundial e entre 10% e 15% do preço final dos produtos.

Como mencionado anteriormente, a análise de trajetórias permite descobrir padrões de comportamento que podem ser usados em uma grande variedade de domínios. Para isso, é necessário analisar uma grande quantidade de dados gerada a partir do monitoramento dos objetos móveis. Para transformar essa massa bruta de dados em informações úteis, uma forma adequada seria disponibilizá-la em um Data Warehouse (DW), um banco de dados otimizado para lidar com grandes volumes de dados de forma eficiente. Para dados convencionais, DWs têm sido usados com sucesso no decorrer das últimas décadas. Entretanto, a natureza dos dados de trajetória e a grande quantidade desses dados impõem certos desafios para construção e manutenção do DW, dentre eles:

1. O monitoramento dos objetos móveis gera apenas dados brutos que para muitas aplicações não são suficientes para extrair informações úteis. Portanto, antes desses dados estarem prontos para uso, eles precisam ser enriquecidos com informações semânticas (Alvares et al., 2007a; Bogorny; Kuijpers; e Alvares, 2009; Palma, 2008).
2. O suporte oferecido pelas tecnologias existentes para dados de trajetória ainda está limitado ao armazenamento e recuperação de observações individuais da trajetória (Spaccapietra et al., 2008; Wolfson et al., 1998). Não existe suporte nativo a trajetórias, como acontece com os dados espaciais em banco de dados espaciais.
3. A grande quantidade dos dados de trajetória consome muitos recursos, tornando o

¹O Estado de São Paulo. **Megacidades Grandes Reportagens**. Disponível em: http://www.estadao.com.br/megacidades/sp_transito.shtm. Acesso em: 20 set 2009.

²Sínteses da Legislação da União Europeia. **A logística do transporte de mercadorias na Europa**. http://europa.eu/legislation_summaries/environment/tackling_climate_change/124456_pt.htm. Acesso em: 20 set 2009.

tempo de processamento das consultas longo, impossibilitando análise no estilo OLAP (Orlando et al., 2007).

4. Escassez de modelos, existem poucos trabalhos relacionados a modelagem de trajetórias, uma questão em aberto, que tem recebido muita atenção nos últimos anos (Macedo et al., 2008, p. 123).

Além do DW Espacial, outras tecnologias poderiam ser utilizadas para analisar trajetórias. Entretanto, elas não são adequadas, principalmente, porque não foram projetadas para analisar grandes quantidades de dados de forma eficiente. Como exemplos dessas tecnologias pode-se mencionar:

Banco de Dados (BD) convencional, projetado para armazenar dados operacionais de forma detalhada e com pouca ou nenhuma redundância, onde as consultas tipicamente se referem a dados correntes, e as operações de inserção, atualização e remoção são muito intensas.

Sistema de Informação Geográfica (SIG), similar a um BD, mas são voltados para análise de dados espaciais.

Banco de Dados de Espaço-Temporais e Banco de Dados de Objetos Móveis (BDET e BDOM): são similares aos banco de dados, mas seu foco é a análise de dados espaço-temporais. A principal diferença entre BDETs e BDOMs, é que apenas os BDOMs foram projetados para lidar com objetos que se movem rapidamente, isto é, alteram sua localização espaço-temporal constantemente. BDETs consideram objetos que não se movem, ou quando se movem é de forma muito lenta (p.ex. um iceberg) (Mokhtar; Su; e Ibarra, 2002). A mesma diferença ocorre entre um DW Espaço-Temporal (DWET) e um DW de Trajetórias (DWTrs).

1.1 Objetivos

O presente trabalho almeja contribuir para o amadurecimento das tecnologias de Data Warehousing de Trajetórias, investigando como as tecnologias existentes podem ser usadas para proporcionar análise de trajetórias no estilo OLAP.

Objetivo Geral

Desenvolver um modelo para Data Warehouse de Trajetórias, (i) que permita analisar o comportamento dos objetos móveis *sobre e entre* as regiões no espaço e tempo, e (ii) que proporcione compactação de trajetórias, para reduzir a quantidade de dados armazenada; (iii) e que mantinha informações semânticas associadas às trajetórias e/ou objetos móveis, o que aumenta o poder de expressividade do modelo, além de simplificar a definição e processamento de algumas consultas. Os trabalhos encontrados na literatura, em geral, (i) não permitem analisar o deslocamento dos objetos *entre* as regiões; (ii) enfrentam problemas de desempenho, devido a grande quantidade dos dados de trajetória; e (iii) não usam informações semânticas associadas as trajetórias e/ou objetos móveis.

Objetivos Específicos

1. Investigar como as tecnologias de DW existentes podem ser usadas para oferecer análise de trajetórias no estilo OLAP.
2. Construir um modelo semântico para DW de Trajetórias, que proporcione: (i) analisar o comportamento dos objetos móveis *sobre e entre* as regiões em diferentes intervalos de tempo, análise orientada a tráfego e análise orientada a trajetórias, respectivamente; (ii) uma redução na quantidade de dados armazenada, compactação de trajetórias, para viabilizar a análise OLAP para trajetórias; (iii) mantér informações semânticas associadas à trajetória.
3. Proporcionar roll-up aplicado à trajetórias (roll-up para trajetórias). A análise orientada a trajetórias – isto é, a análise do comportamento dos objetos móveis entre as

- regiões no espaço e tempo – sobre diferentes perspectivas e níveis de detalhe deverá ser proporcionada através da agregação por direção dos movimentos da trajetória (roll-up para trajetórias).
4. Gerar uma base de dados de trajetórias semânticas, para ser usada durante a validação do modelo. Para isso deverá ser desenvolvido um protótipo, tendo em vista que a literatura conta apenas com geradores para trajetórias brutas, onde não existem dados sobre as paradas, início/fim das trajetórias, entre outros componentes.
 5. Desenvolver um protótipo para povoar o DWTr. Com base em um esquema de dados que segue o modelo proposto e de uma base de dados de trajetórias já enriquecida com informações semânticas, o protótipo deve ser capaz de processar esses dados e carregá-los no DWTr, o que corresponde a fase de Carga de Dados no processo ETL (*Extraction, Transformation, and Load* - Extração, Transformação, e Carga).

1.2 Estrutura da Dissertação

Os demais capítulos desta dissertação estão organizados como segue:

Capítulo 2: é apresentado a fundamentação teórica necessária para uma boa compreensão do trabalho desenvolvido. Seu conteúdo aborda: (i) os conceitos de trajetória, parada, e movimento; (ii) análise de trajetórias; (iii) as formas mais comumente usadas para armazenar trajetórias; (iv) os conceitos e tecnologias de banco de dados, como Data Warehouse (DW), DW Espacial, e DW de Trajetórias.

Capítulo 3: discorre acerca dos principais trabalhos relacionados à modelagem semântica de trajetórias e a análise de objetos móveis envolvendo DWs. Também é apresentada uma discussão sobre os requisitos desejados para um modelo semântico para DW de Trajetórias, adquiridos a partir da análise dos trabalhos relacionados. Ao final do capítulo é apresentada uma tabela de características contendo os requisitos contemplados e almejados pelos trabalhos relacionados e por este trabalho.

Capítulo 4: descreve em detalhes o modelo proposto, o funcionamento dos operadores OLAP aplicados a trajetórias, e como é feita a compactação de trajetórias, para que se consiga um ganho de desempenho a ponto de tornar o DWTr viável.

Capítulo 5: apresenta a metodologia usada para validar o modelo proposto, e os resultados das consultas e testes realizados para avaliar o nível de compactação obtido para dados de trajetória em relação ao modelo clássico. Para isso, o DWTr foi povoado diversas vezes com uma mesma base de dados de trajetórias e um mesmo esquema de dados, mas sobre diferentes configurações para cada teste realizado.

Capítulo 6: conclui o documento com uma avaliação dos resultados obtidos e um conjunto de trabalhos que podem ser desenvolvidos no futuro.

Capítulo 2

Fundamentação Teórica

O presente capítulo apresenta alguns conceitos básicos necessários a uma melhor compreensão desta dissertação. Inicialmente, nas *Seções 2.1, 2.2, e 2.3* são apresentados alguns conceitos importantes usados ao longo do trabalho, como trajetória, parada, e movimento. Na *Seção 2.4* é descrito a análise de trajetória sob dois pontos de vista – análise orientada a tráfego e análise orientada a trajetórias – enquanto na *Seção 2.5* são discutidos as formas mais comumente usadas para armazenar trajetórias. Para encerrar o capítulo, na *Seção 2.6* são apresentados alguns conceitos e tecnologias de banco de dados, como Data Warehouse (DW), DW Espacial (DWE), e DW de Trajetórias (DWTr).

2.1 Trajetórias

Em linhas gerais, o termo trajetória denota a mudança da posição de um objeto móvel no espaço durante um dado intervalo de tempo. Se essa posição é expressada na forma de um conjunto de coordenadas espaço-temporais (isto é, observações na forma (x, y, t)), diz-se que a trajetória é espaço-temporal. Para representar esse tipo de trajetória, buscou-se um modelo capaz de expressar não apenas a sequência de observações da trajetória, mas também informações semânticas que dessem significado a ela. Para isso, foi estendido o modelo de trajetórias semânticas proposto por Spaccapietra et al. (2008), descrito nos parágrafos a seguir.

Definição de Trajetória: uma trajetória é definida pelo usuário, e registra a mudança da posição de um objeto móvel durante um dado intervalo de tempo $[t_{fim_parada_f1}, t_{fim_parada_f2}]$ com um objetivo (p.ex. sair de casa para o trabalho), onde $t_{fim_parada_f1}$ e $t_{fim_parada_f2}$ representam os instantes de tempo do fim da parada $f1$ e $f2$, respectivamente, como ilustrado na Figura 2.1.

Uma trajetória semântica possui duas faces:

Face geométrica: por questões de simplicidade, é representada na forma discreta através de uma sequencia finita de observações na forma $[((x_1, y_1), t_1), ((x_2, y_2), t_2), \dots, ((x_n, y_n), t_n)]$, onde cada **observação** $((x_i, y_i), t_i)$ representa uma localização espaço-temporal capturada da trajetória no instante de tempo t_i , e o par (x_i, y_i) a localização espacial do objeto em um espaço bidimensional, com x_i, y_i e $t_i \in R$, sendo $t_i < t_{i+1}$. Para reconstruir os movimentos da trajetória entre duas observações consecutivas é usada a função de interpolação linear (*linear interpolation*) (Macedo et al., 2008), a qual considera que um objeto móvel se desloca em linha reta a uma velocidade constante entre duas observações.

Face semântica: divide a *face geométrica* em vários componentes, os quais podem transportar informações semânticas definidas pelo usuário, que dão significado à parte da trajetória a qual pertence. Para este trabalho adotou-se os seguintes componentes semânticos: *início da trajetória, fim da trajetória, parada e movimento*.

Considera-se o *início da trajetória* e *fim da trajetória*, as paradas que delimitam o início e fim da trajetória, respectivamente (como o próprio nome diz). Os conceitos de *parada* e *movimento* são descritos logo abaixo.

Definição de Parada: uma parada é a parte de uma trajetória tal que:

- É definida explicitamente pelo usuário no intervalo de tempo $\Delta t = [t_{fim_parada_f1}, t_{fim_parada_f2}]$, com $\Delta t > 0$.
- O objeto móvel não se move, ou quando isso ocorre é de forma muito lenta, permanecendo um longo período de tempo sobre uma mesma faixa espacial, se movimentando

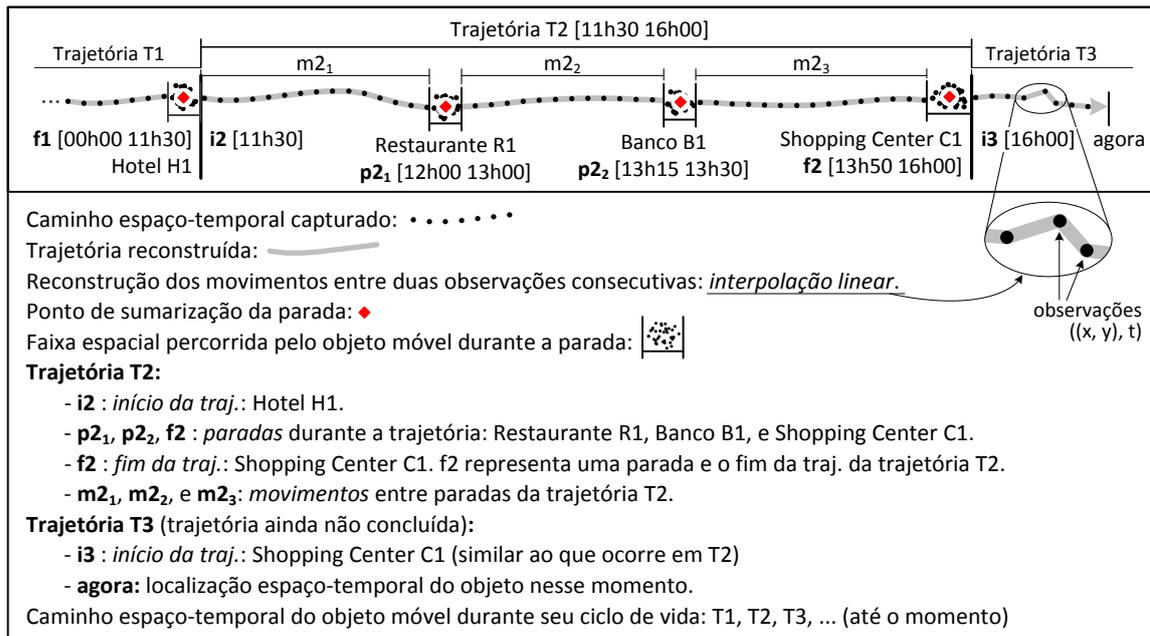


Figura 2.1: Caminho espaço-temporal de um objeto móvel durante seu ciclo de vida.

quase que em círculos. Por exemplo, na parada $p2_1$ da Figura 2.1 o objeto se movimentou dentro de uma pequena faixa espacial delimitada pelo Restaurante R1, durante todo o intervalo de tempo das 12h00 às 13h00.

- Todas as paradas são sempre pontos disjuntos. Isto é, não existem duas paradas consecutivas sem que antes ocorra um movimento entre elas.

Definição de Movimento: um movimento é a parte de uma trajetória tal que:

- É delimitado pelas extremidades de duas paradas consecutivas $[t_{fim_parada_f1}, t_{fim_parada_f2}]$.
- A faixa espacial da trajetória em seu intervalo de tempo é uma linha espaço-temporal, não apenas um simples ponto.

Em aplicações práticas, na maioria das vezes, não é necessário manter todas as observações da trajetória. Nesses casos, as observações podem ser mantidas **de forma sumarizada** (isto é, de forma pré-agregada), onde cada registro representa um resumo de todas as informações contidas em um conjunto de dados. Por exemplo, todas as observações da parada $p2_1$ da Figura 2.1 são expressadas através de um único registro contendo os atributos: local

de parada (Restaurante R1), tempo de parada (das 12h00 às 13h00), e atividades realizadas (parada para almoço). De forma similar acontece com as observações referentes a movimentos, que podem ser expressadas na forma de transição entre paradas, contendo os atributos: origem e destino do movimento (do Restaurante R1 para o Banco B1), velocidade média no percurso, espaço percorrido, tempo gasto, entre outros.

Por simplicidade, na maioria das aplicações, os **objetos móveis** são representados na forma de um simples ponto (x, y) que se desloca no espaço e tempo, desprezando-se suas características espaciais como área e volume.

Não confunda o **conceito de trajetória** com os conceitos de: (i) **caminho espaço-temporal**, um objeto móvel durante todo o seu ciclo de vida possui um único caminho espaço-temporal, o qual é formado por uma sequência de trajetórias. Portanto, cada trajetória representa um segmento (uma parte) de um caminho espaço-temporal, como ilustrado na Figura 2.1; e (ii) **rota**, a qual mantém apenas o percurso realizado pelo objeto móvel para ir de um lugar ao outro, ou seja, é uma sequência de coordenadas espaciais (x, y) da origem ao destino do movimento. Diferente de uma trajetória, uma rota não possui dados temporais.

Baseado nas definições apresentadas nos parágrafos anteriores, Spaccapietra et al. (2008) definiram um modelo conceitual básico para trajetórias semânticas (ver Figura 2.2), descrito nos parágrafos a seguir. Para exemplificar os componentes do modelo, seus atributos são descritos em função de um cenário de aplicação, o gerenciamento de tráfego urbano.

Objeto Móvel (ou *TravelingOT*): mantém as informações sobre os objetos móveis, como por exemplo, nome, sexo, profissão, estado civil, e data de nascimento.

Trajétória (*Trajectory*): contém informações referentes a trajetória como um todo, tais como origem e destino da trajetória (*localOrigem* e *localDestino*), início e fim da trajetória (*dataInicio* e *dataFim*). Além de informações semânticas como o objetivo da trajetória (*objetivo*).

B.E.S (*Begin, End, e Stop* - ou Início, Fim, e Parada da trajetória): mantém informações sumarizadas sobre as paradas no início, fim, e meio da trajetória – representadas pelos componentes: *início da trajetória*, *fim da trajetória*, e *parada* – respectivamente.

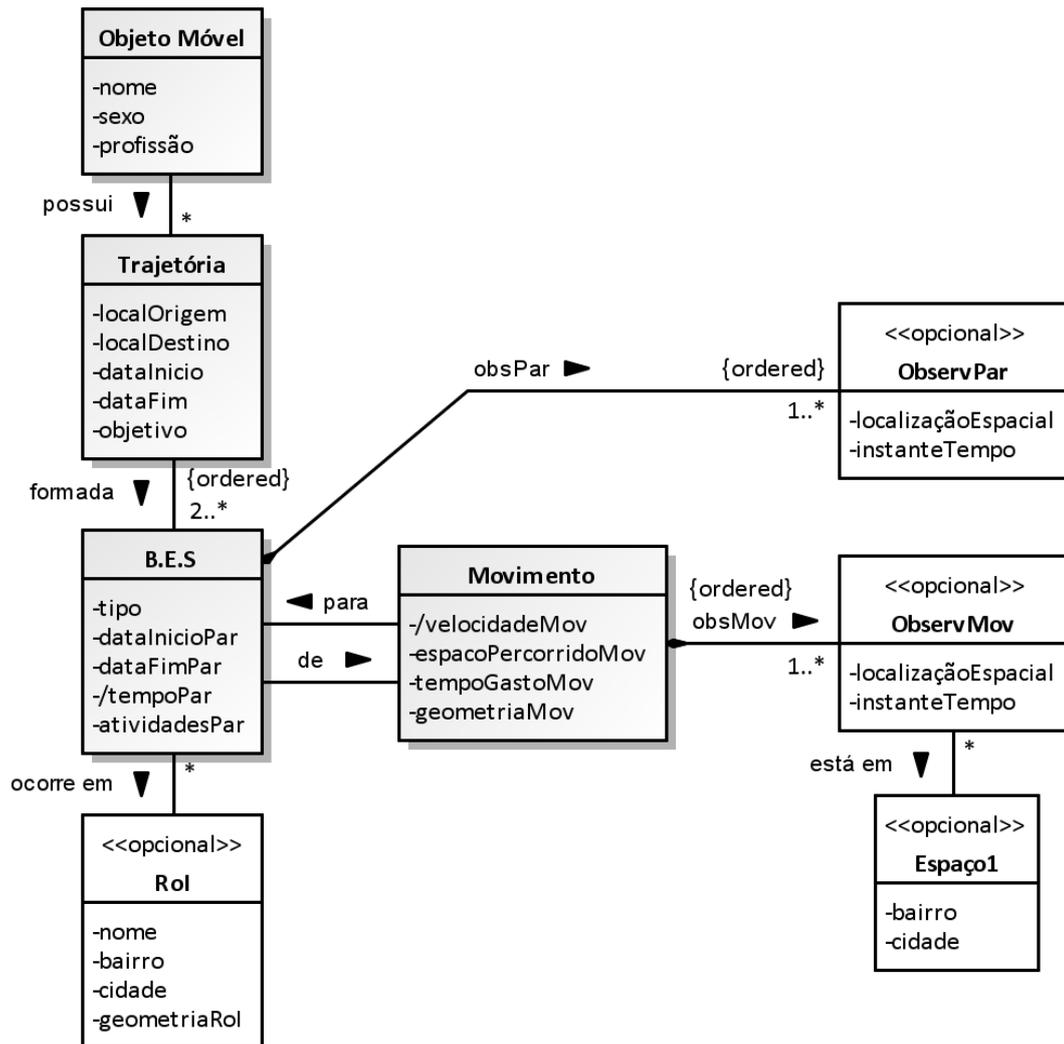


Figura 2.2: Diagrama UML para o modelo conceitual de trajetórias proposto por Spaccapietra et al. (2008) (adaptado).

Exemplos de atributos: início e fim da parada (`dataInicioPar` e `dataFimPar`), tempo de parada (`tempoPar`), e atividades realizadas durante a parada (`atividadesPar`).

Movimento (Move): armazena informações sumarizadas sobre o movimento entre duas paradas consecutivas. Exemplos de atributos: velocidade média do objeto da origem ao destino do movimento (`velocidadeMov`), espaço percorrido (`espacoPercorridoMov`), e tempo gasto (`tempoGastoMov`).

Rol e Espaço1: componentes opcionais, armazenam informações referentes ao espaço geográfico, tais como: nome da região, bairro a qual pertence, cidade, e geometria da região. Um Rol, Região de Interesse, é uma uma região espacial, como por exemplo,

uma casa, ou o campus de uma universidade.

ObservPar e ObservMov: componentes opcionais, armazenam as observações da trajetória referentes as paradas e movimentos, respectivamente. São responsável por manter a face geométrica da trajetória, isto é, a sequencia de observações capturadas. Na maioria das aplicações práticas, não é necessário manter esses dois componentes, as informações sumarizadas mantidas pelos componentes B.E.S e Movimento já são suficientes.

Os componentes – ObservPar, ObservMov, e Espaço1 – não foram definidos em Spaccapietra et al. (2008).

2.2 Reconstrução de Trajetórias

As observações capturadas da trajetória constituem uma sequencia bruta de dados, como ilustrado na Figura 2.3(i). Nessa sequencia, o início e fim de cada trajetória e suas paradas ainda são desconhecidas. Portanto, antes dessa sequencia estar pronta para uso, é necessário identificar quais são as paradas e trajetórias ali contidas, o que é feito através dos métodos para **detecção de paradas**(Bogorny; Kuijpers; e Alvares, 2009; Palma, 2008) e **detecção de trajetórias**, respectivamente.

Para reconstruir os movimentos intermediários da trajetória entre duas observações consecutivas, é necessário usar a **função de interpolação**. Para isso, frequentemente é usado a função de interpolação linear (*linear interpolation*) (Macedo et al., 2008), a qual considera que um objeto móvel se desloca em linha reta a uma velocidade constante entre duas observações, como ilustrado na Figura 2.4. Um método simples que apresenta bons resultados quando o intervalo de captura das observações é pequeno em comparação com a velocidade do objeto móvel, por exemplo, para um veículo circulando em uma cidade, 20 segundos é um bom intervalo de captura.

Todas as etapas necessárias para transformar a sequencia bruta de observações em trajetórias semanticamente enriquecidas, definem o **processo de reconstrução de trajetórias**.

Sendo assim, a detecção de paradas, a detecção de trajetórias, e a interpolação dos movimentos da trajetória são considerados etapas desse processo. Na Figura 2.3 é exemplificado um possível processo para reconstrução de trajetórias, as etapas não precisam ser necessariamente estas, nem precisam estar nessa ordem.

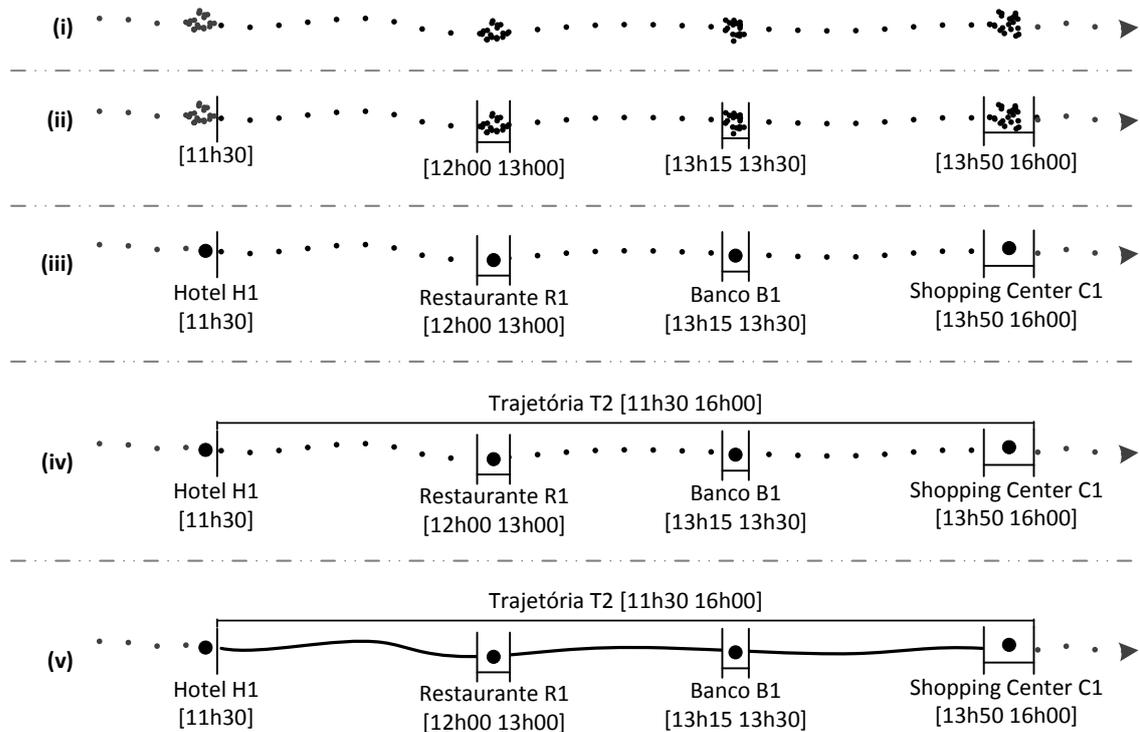


Figura 2.3: Etapas do processo de Reconstrução de Trajetórias. Em: **(i)** sequência bruta de observações gerada a partir do monitoramento do objeto móvel; **(ii)** paradas identificadas e rotuladas; **(iii)** paradas sumarizadas e georeferenciadas; **(iv)** trajetórias identificadas e rotuladas; **(v)** movimentos da trajetória interpolados.

A precisão das trajetórias reconstruídas depende tanto da função de interpolação empregada quanto do intervalo de captura das observações, como exemplificado na Figura 2.4. Em **(a)** a trajetória foi reconstruída com sucesso, pois é similar a trajetória real; mas em **(b)** a reconstrução falhou, a trajetória reconstruída difere muito da trajetória real. Nesse caso, o intervalo de captura das observações usado é muito grande quando comparado com a velocidade do objeto móvel, o que gerou erros grosseiros durante sua reconstrução. Ambas as figuras empregam a função de interpolação linear.

Não confunda os termos “reconstrução de trajetórias” com “detecção de trajetórias”. Na

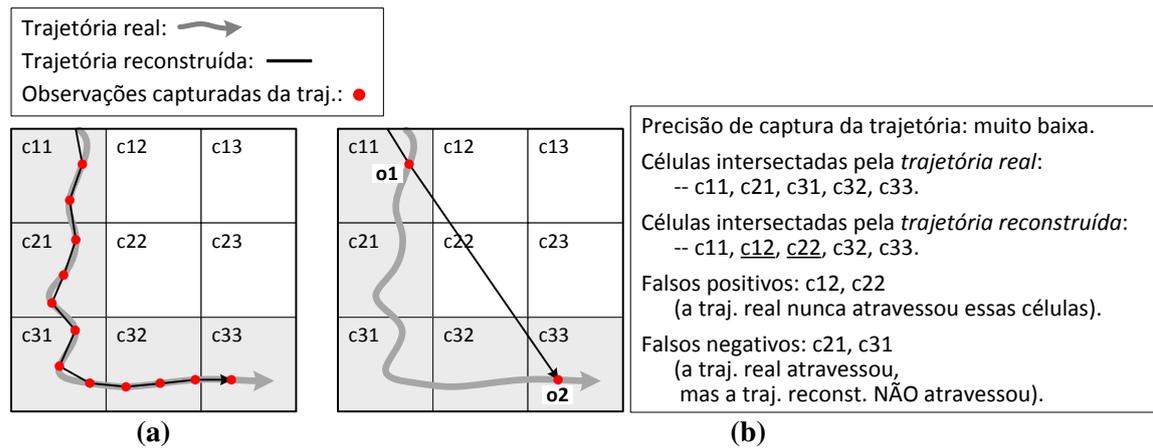


Figura 2.4: Trajetória Interpolada usando a função de interpolação linear.

Em **(a)** Trajetória Interpolada com sucesso, em **(b)** falhou.

literatura, o termo “reconstrução de trajetórias” (*trajectory reconstruction*) frequentemente é usado para referir-se a “detecção de trajetórias”, como em Giannotti; e Pedreschi (2008); Marketos et al. (2008); Marketos (2009), como mostrado no parágrafo anterior, neste trabalho o termo “reconstrução de trajetórias” refere-se a um processo muito mais amplo.

2.3 Representação do Espaço Geográfico

Para representar o espaço geográfico, uma abordagem amplamente utilizada na literatura consiste em dividir o espaço analisado por uma grade regular, formada por um conjunto de células regulares (Tao; e Papadias, 2005). Cada célula, portanto, constitui a menor unidade do espaço geográfico.

Uma das vantagens dessa abordagem é que as células possuem o mesmo tamanho, o que facilita a comparação das características analisadas entre as regiões adjacentes (isto é, outras células) (Rinzivillo et al., 2008), como ilustrado na Figura 2.5. Nessa figura, em cada gráfico de pizza é representado a origem (em amarelo) e o destino (em azul) das trajetórias em um dado intervalo de tempo em termos do número de trajetórias. O tamanho do gráfico é proporcional ao número total de trajetórias na célula (origem + destino).

Entretanto essa forma de representação possui alguns inconvenientes, como descrito na seção abaixo.

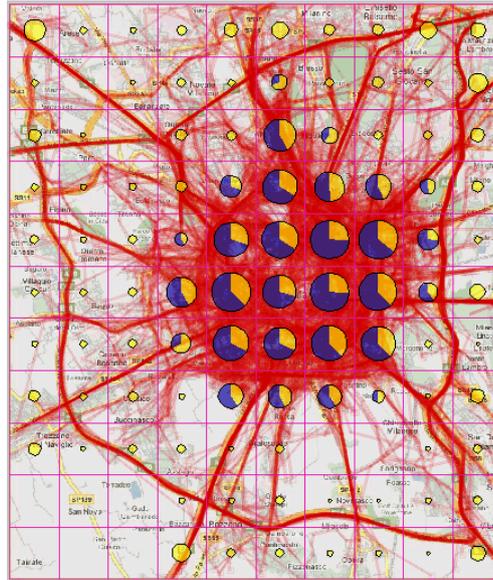


Figura 2.5: Espaço geográfico dividido por uma grade regular (Rinzivillo et al., 2008).

2.3.1 Problema do Relacionamento Parcialmente Contido

Representar o espaço geográfico através de células regulares possui um inconveniente, essa forma de representação adiciona imprecisão as agregações. Isso ocorre porque nem sempre uma célula está contida completamente em uma região, pode acontecer dela pertencer a duas ou mais regiões, como ilustrado na Figura 2.6. Observe na figura que a *célula c23* pertence simultaneamente aos bairros: *bairro1*, *bairro2*, e *bairro3*. Dessa forma, as agregações envolvendo as células do *bairro1*, indiretamente envolvem também dados referentes aos bairros *bairro1* e *bairro2*, o que não deveria acontecer. Isso é conhecido como problema do *relacionamento parcialmente contido* (*partial containment relationships*) (Jensen et al., 2004).

Uma solução para esse problema consiste em dividir as células da grade pelas regiões, de forma a eliminar o relacionamento parcialmente contido existente. Por exemplo, seguindo essa lógica a célula *c23* da Figura 2.6 seria dividida por: *bairro1*, *bairro2*, e *bairro3*. Resultando nas células: *c23-bairro1* (parte da célula *c23* em *bairro1*), *c23-bairro2*, e *c23-bairro3*.

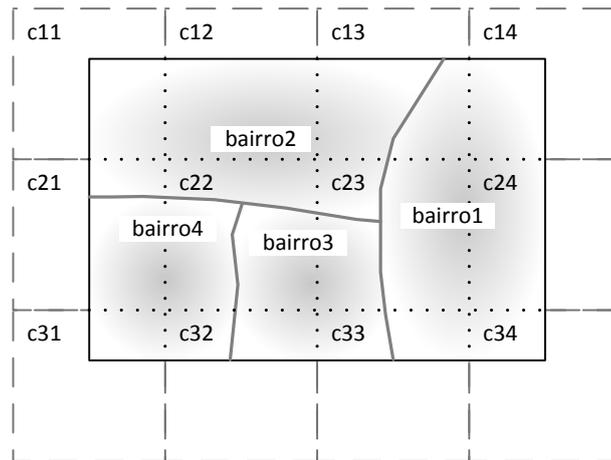


Figura 2.6: Relacionamento Parcialmente Contido, a célula c23 pertence simultaneamente a bairro1, bairro2, e bairro3.

2.4 Análise de Trajetórias

A análise de trajetórias ajuda a compreender melhor o comportamento dos objetos móveis, e suas relações com o ambiente a sua volta, assim como entre eles. Segundo Andrienko; e Andrienko (2008), trajetórias podem ser analisadas sob dois pontos de vista: visão orientada a tráfego (*traffic-oriented view*) e visão orientada a trajetórias (*trajectory-oriented view*), ou como prefere-se chamar neste trabalho, *análise orientada a tráfego* e *análise orientada a trajetórias*, como descritas a seguir:

Análise orientada a Tráfego: tem como objetivo analisar o comportamento dos objetos móveis sobre uma dada região em diferentes intervalos de tempo, em termos de: número de objetos móveis, velocidade média, espaço percorrido, tempo gasto, entre outros atributos. Por exemplo:

- *A velocidade média dos objetos móveis por célula da grade regular?*
- *Os locais mais visitados pelos objetos móveis no centro da cidade por célula da grade regular?*
- *O número de aves circulando na Europa e no norte da África por mês durante o ano de 2000? O resultado dessa consulta é mostrado na Figura 2.7(a).*

Análise orientada a Trajetórias: tem como objetivo analisar a transição dos objetos móveis entre as regiões em termos de origem-destino do movimento, para os atributos: número de objetos móveis, velocidade média, espaço percorrido, entre outros. Os atributos analisados são similares a análise orientada a tráfego. Exemplos de consultas envolvendo esse tipo de análise:

- *O número de objetos móveis que partem para o centro da cidade ao meio-dia?*
- *Qual a parte da população afetada com a interdição de uma determinada ponte?*
- *O número de veículos se deslocaram entre duas paradas consecutivas (origem-destino) a cada 30 minutos?* O resultado dessa consulta é mostrado na Figura 2.7(b).

Dependendo do modelo adotado, os dois tipos de análise mostrados anteriormente estão disponíveis. Por simplificação, este trabalho usa o termo *análise de trajetórias* para expressar ambos os tipos de análise.

Para analisar trajetórias de forma correta, é fundamental oferecer uma distinção clara entre paradas de movimentos em trajetórias. A inclusão de dados sobre paradas na análise de movimentos pode provocar forte discrepância entre os dados analisados e os reais, e vice-versa. Por exemplo, ao se incluir dados sobre paradas no cálculo da velocidade média de uma região, tem-se uma forte impressão de que a velocidade na região analisada está baixa, devido a influência dos dados sobre paradas – cuja velocidade é igual a zero, ou muito próximo a isso – durante as computações.

Em particular, os atributos analisados variam de acordo com o cenário de aplicação, mas analisando os trabalhos de Brakatsoulas; Pfooser; e Tryfona (2004); Braz (2009); Marketos (2009); Spaccapietra et al. (2008), identificou-se que alguns atributos podem ser considerados universais, como segue:

Atributos relacionadas a paradas: número de paradas; tempo de parada; atividades realizadas; e locais de parada.

Atributos relacionadas a movimentos: velocidade; distância percorrida; tempo gasto; e

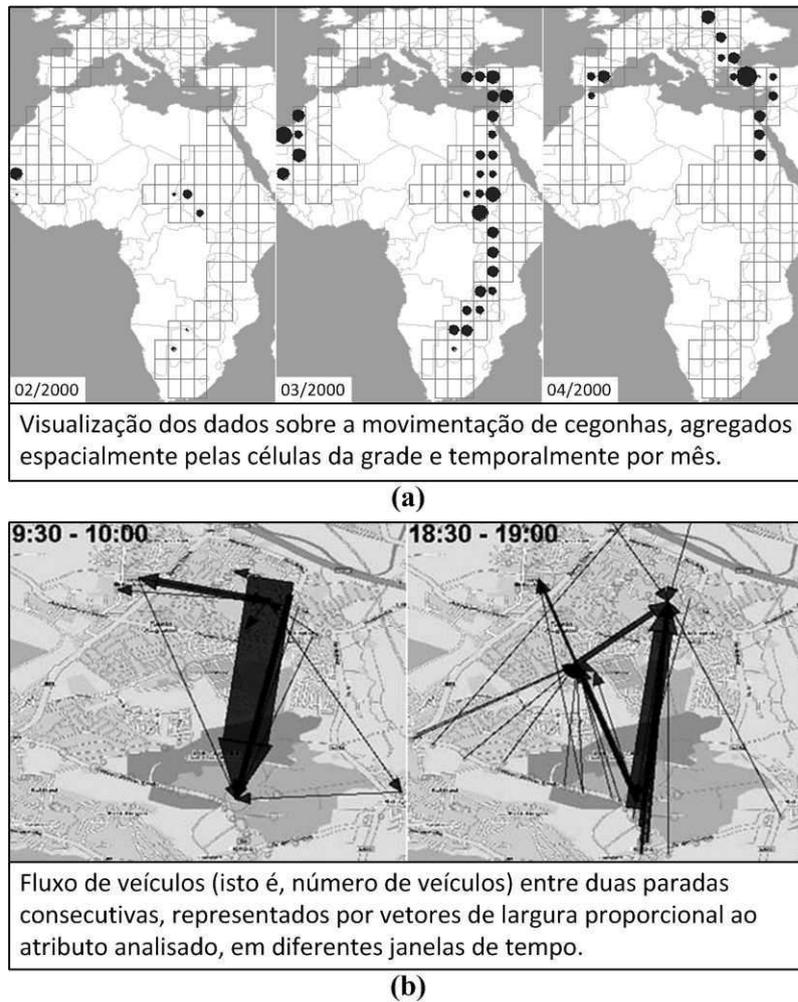


Figura 2.7: Análise de trajetórias. Em: (a) análise orientada a tráfego; (b) análise orientada a trajetórias (Andrienko et al., 2008, p. 396; 397).

direção do movimento entre regiões.

Atributos relacionadas a contagens (quantidade) de objetos móveis e/ou trajetórias:

número de trajetórias que começam ou terminam na região analisada (permite identificar a origem ou destino das trajetórias); e número de objetos móveis que atravessam uma dada região.

2.4.1 Vantagens oferecidas pelo uso de Trajetórias Semânticas

Na maioria das vezes, os objetos móveis se deslocam em um espaço geográfico georeferenciado, e que possui informações semânticas associadas a ele, tais como nome da rua, bairro,

cidade, número de acidentes no trecho, horário de maior congestionamento, entre outras. A esse conjunto de informações dar-se o nome de **informações geográficas de fundo** (*background geography*) (Palma, 2008, p. 15). Essas informações podem ser usadas para ajudar a descobrir informações úteis sobre as trajetórias desses objetos, como por exemplo, os lugares mais visitados por eles e a transição entre esses lugares. Analisando a Figura 2.8, é possível ver que os dados de trajetórias juntamente com as informações geográficas de fundo agregam mais informação útil do que quando as trajetórias são consideradas objetos isolados no espaço (Alvares et al., 2007b; Palma, 2008).

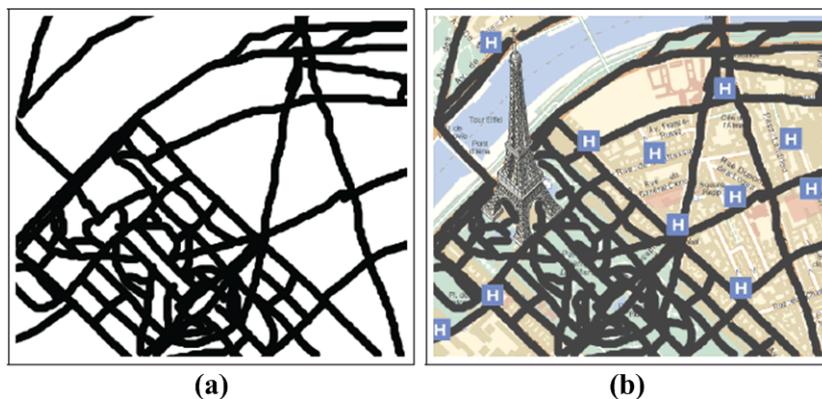


Figura 2.8: O mesmo conjunto de trajetórias mostrado SEM e COM informações geográficas de fundo, figuras (a) e (b), respectivamente (Palma, 2008, p. 16).

Consultas envolvendo trajetórias semânticas podem ser mais simples e eficientes do que consultas usando trajetórias brutas (apenas dados espaço-temporais) (Alvares et al., 2007a; Bogorny; Kuijpers; e Alvares, 2009). Para compreender essa simplificação, considere a consulta exemplo apresentada por Bogorny; Kuijpers; e Alvares (2009): *Quais são os lugares que o objeto móvel “A” passou durante sua trajetória?* Supondo que “A” possui uma única trajetória. Essa consulta é resolvida de duas formas: (i) usando *trajetórias brutas* (informações semânticas não existem) e; (ii) usando *trajetórias semânticas*, como mostrado a seguir:

Trajetoórias Brutas: nesse caso, é preciso testar quais observações da trajetória intersectam a geometria dos lugares conhecidos. Isso é feito através da operação de *interseção* (*intersects*), como ilustrado na consulta SQL da Figura 2.9(a). Durante o processamento

da consulta, esse operador percorre todas as observações da trajetória testando para cada observação se ela intersecta a geometria de algum lugar conhecido.

Trajétórias Semânticas: o processamento da consulta é bem mais simples. Para responder a consulta, apenas é listado os lugares por onde o objeto “A” passou, como ilustrado no SQL da Figura 2.9(b). Nesse caso, antes das trajetória serem carregados no banco de dados, primeiro elas foram enriquecidas com informações semânticas.

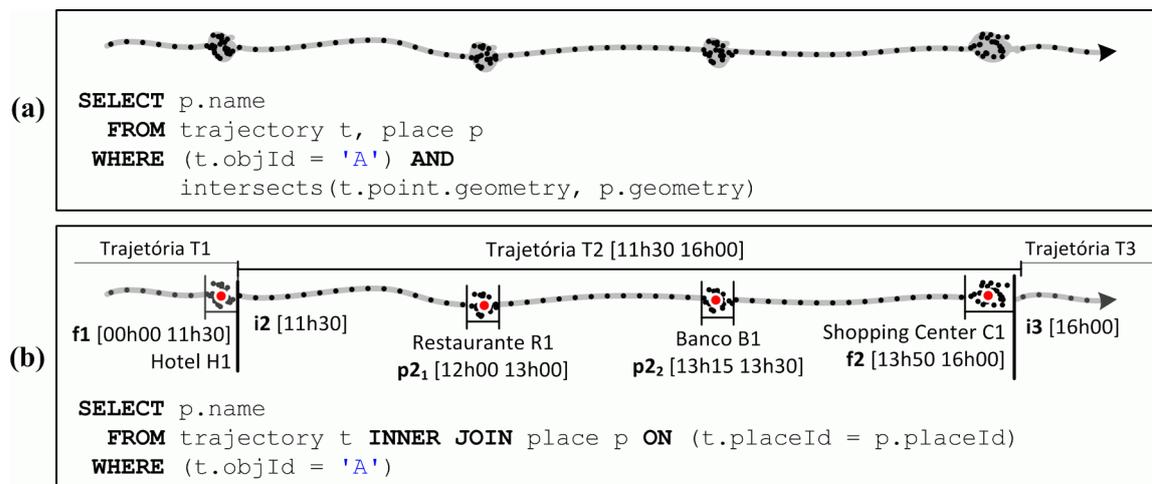


Figura 2.9: Exemplo de uma mesma consulta, resolvida em: (a) SEM o uso de informações semânticas (trajetórias brutas); (b) usando informações semânticas.

Entretanto, essa consulta pode se torna mais complexa quando restrições de tempo são impostas, por exemplo, *Quais são os lugares que o objeto móvel “A” passou durante sua trajetória, e permaneceu por mais de 3 horas?*. Nesse exemplo, tanto a complexidade computacional como a formalização da consulta aumentam muito, pois é preciso verificar para cada trajetória todos os pontos que intersectam as geometrias dos lugares a fim de calcular o tempo de permanência do objeto no local, o que aumenta consideravelmente o espaço de busca quando comparado com a consulta anterior.

Trajétórias semânticas não apenas simplificam o processamento das consultas, mas também **facilitam significativamente a mineração de dados e o processo de descoberta do conhecimento** (Bogorny; Kuijpers; e Alvares, 2009). Por exemplo, na Figura 2.10 é mostrado um conjunto de objetos móveis se deslocando de um lugar para outro. Em (a) as

trajetórias são brutas, nesse caso, não fica claro a origem e o destino das trajetórias. Mas em (b) usando trajetórias semânticas, a origem e o destino das trajetórias ficam evidentes, os objetos móveis partem de escolas (*School*) com destino ao cinema (*Cinema*).

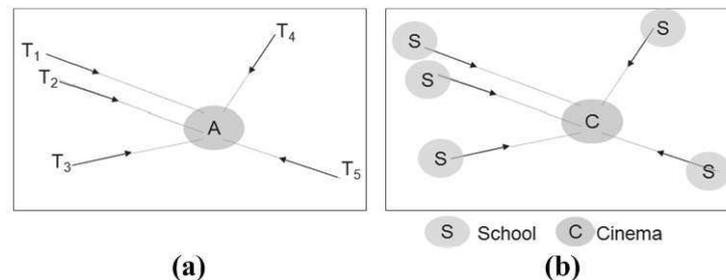


Figura 2.10: Objetos móveis se deslocando de um lugar para outro. Considerando em: (a) trajetórias brutas; (b) trajetórias semânticas (Bogorny; Kuijpers; e Alvares, 2009).

2.5 Formas de Armazenamento para Trajetórias

Um dos desafios enfrentados ao lidar com trajetórias refere-se a forma de armazenamento, este trabalho identificou na literatura, três formas de armazenamento: (i) **baseada na sequencia de observações** (*point-location management*) (modelo clássico) (Macedo et al., 2008); (ii) **sumarizada em células** (Braz et al., 2007; Braz, 2009; Marketos, 2009; Pelekis et al., 2008); e (iii) **na forma vetorial** (E. K. Frenzos, 2008; E. Frenzos, 2003; Guting; Almeida; e Ding, 2006; Macedo et al., 2008), como descrito a seguir.

2.5.1 Baseada na sequencia de observações (modelo clássico)

Uma abordagem simples, amplamente utilizada na indústria, na qual consiste em armazenar toda (ou parte) a sequencia de observações (localizações espaço-temporais (x, y, t)) da trajetória. Essa abordagem possui uma série de inconvenientes, tais como: (i) gera uma enorme quantidade de dados, o que pode gerar sérios problemas de desempenho devido ao grande número de computações necessárias para atender uma simples consulta; (ii) não emprega interpolação, logo, o deslocamento do objeto móvel entre duas observações consecutivas é desconhecido (é como se o objeto teletransportasse, ele deixa um ponto (x_1, y_1) e surge

em (x_2, y_2) e não se sabe como ele fez esse movimento); **(iii)** a ausência de interpolação pode gerar sérios problemas de imprecisão nas consultas (por exemplo, um objeto móvel atravessou uma dada região, mas durante o processamento da consulta essa travessia não é detectada) (Macedo et al., 2008, p. 136).

Para amenizar o problema da grande quantidade de dados, uma solução alternativa consiste em armazenar apenas uma parte das observações capturadas, logo, trata-se de um método de compactação com perdas. Antes da trajetória ser armazenada, um algoritmo varre a trajetória removendo observações não-relevantes. Por exemplo, existem métodos que buscam remover pontos minimizando a deformação espacial da trajetória (Meratnia; e By, 2004), outros mantêm apenas as paradas da trajetória (Gomez; Kuijpers; e Vaisman, 2007).

2.5.2 Sumarizada em células

Em alguns domínios de aplicação, as informações sobre às rotas dos objetos móveis (i) são irrelevantes, (ii) não estão disponíveis, ou (iii) não podem ser mantidas por violar questões de privacidade (Braz et al., 2007; Pelekis et al., 2008; Tao; e Papadias, 2005). Por exemplo, empresas de telefonia móvel não podem manter dados históricos sobre a localização dos telefones móveis de seus clientes por violar questões de privacidade (Tao; e Papadias, 2005).

Nesses casos, uma solução comumente usada consiste em manter apenas as informações sumarizadas de todas as trajetórias por célula espaço-temporal. Por exemplo, seguindo essa lógica todas as observações de todas as trajetórias sobre a célula s_1 da Figura 2.11(a) em um dado intervalo de tempo seriam sumarizadas para um único registro, como mostrado em (b). Como nessa sumarização ocorre à perda do identificador do objeto (*obj-id*), logo não é mais possível identificar os objetos móveis, o que resolve o problema da privacidade dos dados de trajetória.

Essa forma de armazenamento também possui a **vantagem** de requerer muito menos espaço de armazenamento quando comparado ao modelo clássico (Braz et al., 2007; Tao; e Papadias, 2005). Entretanto, ela possui uma **desvantagem**, como mostrado a seguir.

Problema da Contagem Distinta

Quando se agrega dados de trajetórias armazenados de forma sumarizada, podem ocorrer erros relacionados com a contagem distinta dos objetos móveis, o que é conhecido como *problema da contagem distinta (distinct count problem)* (Braz et al., 2007; Tao et al., 2004), durante a contagem distinta um mesmo objeto pode sendo contado múltiplas vezes em uma mesma consulta. Isso acontece porque os dados na forma agregada não possuem o identificador do objeto móvel. Logo, não é possível saber durante a agregação se um mesmo objeto se encontra em duas células distintas a serem agregadas, como mostrado na Figura 2.11(b). Em (a), a situação é diferente, os dados não estão agregados. Logo, é possível realizar a contagem distinta a partir do identificador do objeto móvel.

O problema da contagem distinta, também se aplica a outros atributos (p.ex. trajetórias), pelo mesmo motivo, a falta de um identificador. Em seu trabalho, Braz et al. (2007) apresenta uma solução para amenizar esse problema, como discutido na Seção 3.1.

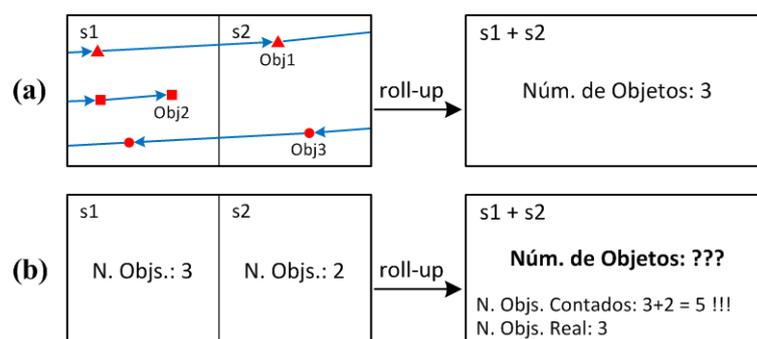


Figura 2.11: Agregação de dados de trajetória, armazenados: **(a)** de forma NÃO sumarizada; **(b)** de forma sumarizada (ocorre o problema da contagem distinta).

2.5.3 Na forma vetorial

Nesta forma de representação, a sequência de observações da trajetória é transformada em uma sequência de vetores movimento na forma $\langle \text{direção do movimento}, \text{velocidade}, \text{distância percorrida} \rangle$, onde *direção do movimento* representa a origem e destino do movimento (Guting; Almeida; e Ding, 2006; Wolfson et al., 1998). Essa forma de representação se divide em duas categorias:

Livre (*unconstrained space* ou *unconstrained movement*): quando a sequência de observações da trajetória é transformada em uma sequência de vetores movimento, sem a imposição de uma restrição para os movimentos do objeto, como ilustrado na Figura 2.12(a). É uma abordagem pouco usada na literatura.

Fixada a uma Rede (*fixed network*): similar a forma livre, mas os vetores movimento são mapeados para uma infra-estrutura de rede (por exemplo, mapa de ruas). Logo, existe uma restrição para o movimento do objeto, como mostrado na Figura 2.12(b). O resultado final dessa transformação passa a ideia de que os objetos móveis estão se movimentando sobre trilhos. Uma das **vantagens** desse método é que há uma redução na quantidade de dados armazenada em comparação com o modelo clássico (Seção 2.5.1), além de facilitar a detecção de relacionamentos entre os objetos móveis (E. K. Frentzos, 2008; Guting; Almeida; e Ding, 2006). Uma das **desvantagens**, é que a maioria das aplicações não conta com uma infra-estrutura de rede completa, por exemplo, não existe uma infra-estrutura de rede associada à rota migratória de aves. É como se os trilhos de uma ferrovia não alcançassem todas as regiões percorridas pelos objetos.

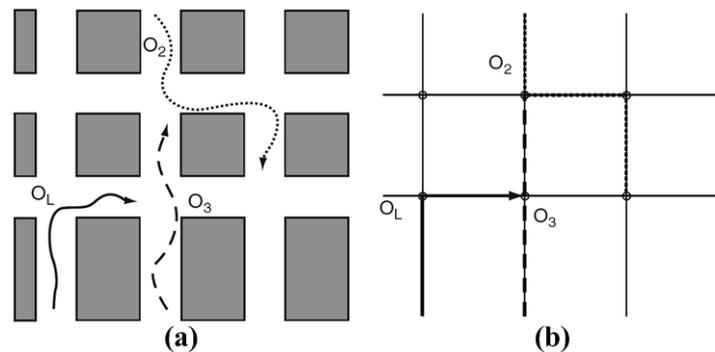


Figura 2.12: Trajetórias armazenadas na forma de Vetores Movimento, em: (a) na forma livre. (b) fixados a uma infra-estrutura de rede. Ambas as figuras correspondem a mesma área urbana (Renso et al., 2008).

2.6 Tecnologias de Banco de Dados

Bancos de Dados (BDs) (Elmasri; e Navathe, 2006) convencionais são sistemas que se encarregam de registrar todas as transações contidas em uma determinada operação organizacional. Tradicionalmente, os bancos de dados transacionais são otimizados para *processamento de transações online* - *On-Line Transaction Processing (OLTP)* - onde consistência e capacidade de recuperação são críticas. Transações são tipicamente pequenas e acessam um pequeno número de registros baseados em chaves primárias. BDs armazenam apenas dados atualizados, não foram projetados com a intenção de manter dados históricos, ou lidar com grandes quantidades de dados, por exemplo, dados na ordem de algumas dezenas de milhões ou bilhões de registros.

2.6.1 Data Warehouse (DW)

Um **Data Warehouse (DW)** (Kimball; Ross; e Merz, 2002) é um banco de dados analítico utilizado para suporte à decisão. Suas principais vantagens são: melhorar a disponibilidade de dados gerenciais, a qualidade das informações obtidas, e integrar dados de múltiplos bancos de dados heterogêneos distribuídos e de múltiplas fontes de informação, oferecendo um enfoque histórico, para permitir um suporte efetivo à decisão. Visto desse modo, um DW é uma visão materializada de dados de diversas fontes (em geral, bancos de dados transacionais), que permite obter informações sumarizadas obedecendo a diferentes níveis de agregação. Para exemplificar, consideremos um DW de vendas. Nas fontes são gravadas as transações de venda feitas durante o dia em uma loja; e no DW, após uma sumarização dessas transações são disponibilizadas as vendas por dia, mês ou ano.

De um modo geral, um DW é modelado em um espaço multidimensional na forma de um *cubo de dados*, com n arestas: cada aresta do cubo representa uma dimensão, e cada ponto interno ao cubo representa uma medida que está associada ao *fato* a ser analisado. Tipicamente, um cubo de dados é implementado através de um *esquema estrela*. De acordo com esse modelo, um DW consiste de uma tabela fato (o centro da estrela), rodeada por

um conjunto de dimensões relacionadas com a tabela fato, a qual contém chaves para as tabelas dimensionais e medidas. As *medidas* são atributos que representam o fato analisado, e as *dimensões* são critérios de agregação para as medidas. Cada dimensão define o nível de granularidade mais refinado os quais se referem aos detalhes dos dados armazenados na tabela fato (Imnon, 2005; Kimball; Ross; e Merz, 2002).

A existência virtual de diversos níveis de agregação leva ao conceito de *hierarquia*. Cada dimensão pode ter uma ou mais hierarquias associadas. Hierarquia pode ser definida como um conjunto de níveis de agregação para uma mesma dimensão nos quais os dados podem ser analisados. Em geral, as medidas são agregadas em níveis mais altos obedecendo algum conceito de hierarquia, para facilitar a compreensão de certos aspectos do negócio. As hierarquias sempre obedecem ao relacionamento um-para-muitos do nível maior (ou mais geral) para o nível menor (ou mais detalhado), como pode ser visto nessa hierarquia - *cidade < estado < país* - na qual um país possui muitos estados, mas um estado pertence a um único país. Entretanto, não é desejável a manutenção física de diversos níveis de agregação devido aos custos inerentes em espaço e tempo. Cada dimensão define o nível de granularidade mais refinado os quais se referem aos detalhes dos dados armazenados na tabela fato. É desejável, entretanto, que as medidas possam ser agregadas em níveis mais altos para facilitar a análise de dados.

Operadores OLAP

As operações OLAP (*On-Line Analytical Processing*, ou processamento analítico online) (Codd; Codd; e Salley, 1993) permitem aos usuários navegarem e explorarem os dados contidos no DW, viabilizando a análise de seus dados sobre várias perspectivas e níveis de granularidade (ou detalhe). Os operadores OLAP mais comuns são:

Roll-up: permitem visualizar os dados sobre diferentes níveis de detalhe. Através da operação roll-up é possível obter dados em um nível menor de detalhe, ou seja, dados mais sumarizados (agregados). Em outras palavras, a partir de dados em um nível maior de detalhe, a operação roll-up através das funções de agregação - tais como sum, count,

avg - agrega esses dados de acordo com um dado critério definido na consulta, tendo como resultado em um nível menor de detalhe. Por exemplo, com base no total de vendas por mês (informações mais detalhadas) calcula-se o total de vendas por ano (informações mais sumarizadas).

Drill-down: a operação *drill-down* tem efeito oposto à operação *roll-up*.

Slice: a operação *slice* permite selecionar quais dimensões farão parte da consulta. Por exemplo, a operação *slice* possibilita um usuário gestor visualizar o total de vendas por nome do produto e por ano (dimensões produto e tempo, respectivamente), omitindo a dimensão filial, a qual não participa da consulta.

Dice: a operação *dice* permite restringir os valores de uma determinada dimensão a partir de um predicado. Por exemplo, considerando a consulta do item anterior, os resultados poderiam ser visualizados apenas para o ano de 2009, incluindo a restrição (ano = 2009) a consulta, as vendas para os demais anos não seriam calculadas.

Não confunda *ferramenta OLAP* com *operação OLAP*. As ferramentas OLAP têm como objetivo permitir a seus usuário consultar e visualizar os dados contidos no DW. Portanto, um usuário explora um DW a partir de uma ferramenta OLAP. É através das operações OLAP que os dados são manipulados para responder a consulta definida pelo usuário. Não confunda também, *operador OLAP* com *operação OLAP*. A operação OLAP é realizada através do operador OLAP.

O Processo ETL

ETL (*Extraction, Transformation, e Load*, ou Extração, Transformação, e Carga) é o processo responsável pela aquisição de dados para o DW, o qual envolve os seguintes passos:

Extração: (i) identificação dos dados relevantes à tomada de decisão; (ii) extração dos dados das múltiplas fontes, heterogêneas (p.ex. banco de dados, arquivos de texto);

Transformação: (i) formatação dos dados para um dado padrão, para manter a consistência

no DW. Por exemplo, em uma empresa cada banco de dados possui uma forma diferente para representar o atributo sexo, alguns bancos usam as literais “masculino” ou “feminino”, outros os caracteres ‘M’ ou ‘F’. Para evitar problemas de consistência, o DW deve possuir apenas uma forma para representação (p.ex. ‘M’ ou ‘F’); (ii) limpeza dos dados, isto é, corrigir dados errôneos ou incompletos. A transformação é uma tarefa muito importante e complexa, de modo geral, consome aproximadamente 80% do tempo na construção de um DW (Inmon, 2002).

Carga: povoamento dos dados no DW. O grande volume de dados torna a carga de dados uma tarefa lenta (horas ou dias). Para evitar que o DW fique ausente nos momentos de maior necessidade, define-se uma política de atualização, a qual determina a forma e a frequência com a qual a carga de dados será executada.

2.6.2 Data Warehouse Espacial (DWE)

Um **Data Warehouse Espacial** (DWE, ou *Spatial Data Warehouse - SDW*) (Bédard; Merrett; e Han, 2001) combina as tecnologias de DW e SIG em um único ambiente. Essa fusão permite explorar o que é de melhor em ambos os sistemas. Enquanto DWs oferecem métodos eficientes para análise e recuperação de grandes quantidades de dados, SIGs oferecem estruturas adequadas para manipular e armazenar **dados espaciais** (Malinowski; e Zimányi, 2004). Esses dados representam as geometrias dos objetos, que possuem uma referência espacial, a qual está associada a algum sistema de coordenadas (p.ex. latitude, longitude). Ponto (simples localização espacial), linha (p.ex. uma rua), e polígono (p.ex. um bairro) são exemplos de tipos de dados espaciais.

O desenvolvimento e barateamento da computação móvel e dos serviços baseados em localização (*location-based services*), tais como satélites e GPS, possibilitam a captura em larga escala de dados espaciais, tendo como resultado a geração de grandes quantidades desses dados. O uso desses dados em um DW pode trazer benefícios ao processo de tomada de decisão, visto que permitem realizar consultas mais abrangentes, através do uso de novos tipos de dados e operações. Por exemplo, questões do tipo: *qual o total de venda de um*

dado produto por bairro da cidade? Considerando apenas os bairros que se encontram a uma distância X da filial. Podem ser respondidas, mostrando aos gestores informações úteis até então desconhecidas. Estima-se que aproximadamente 80% dos dados armazenados em bancos de dados contenham algum tipo de componente com contexto espacial (Franklin, 1992), tais como rua, bairro, cidade, e CEP.

Um DWE pode ser visto como uma extensão natural de um DW convencional que possui dados não-espaciais e espaciais (Stefanovic; Han; e Koperski, 2000). Para um DW ser considerado espacial ele deve possuir ao menos um dado espacial no fato e/ou em uma dimensão. Assim como em DWs, de modo geral, um DWE é modelado em um espaço multidimensional (nesse caso, tem-se um modelo multidimensional espacial). Onde o fato possui medidas: (i) *numéricas* como nos DWs convencionais (p.ex. a quantidade de vendas de um dado produto); e/ou (ii) *espaciais* contém dados espaciais (p.ex. ponto que representa o local de venda). E as dimensões contém critérios: (i) *não-espaciais*: dados convencionais (p.ex. nome do produto, ano de venda); e/ou (ii) *espaciais* (p.ex. polígono que representa o bairro, ou cidade).

Capítulo 3

Trabalhos Relacionados

Data Warehouse de Trajetórias (DWTrs) e os operadores TrOLAP (*Trajectory OLAP*) não têm sido satisfatoriamente abordados na literatura apesar de sua relevância. Porém, algumas propostas apresentadas neste capítulo, podem ser adaptadas para o desenvolvimento de DWTrs e conseqüentemente para análise TrOLAP.

Nos parágrafos abaixo, é descrito de forma sucinta os trabalhos considerados mais relevantes para esta dissertação, os quais serão discutidos em detalhes nas seções a seguir.

Braz et al. (2007) apresentam uma nova forma para agregar dados de trajetória armazenados de forma sumarizada em DWs tradicionais, minimizando o problema da contagem distinta (maior contribuição), e resolvendo o problema da privacidade dos dados de trajetória.

Marketos et al. (2008) descrevem um processo ETL para DWTr baseado no modelo de Braz et al. (2007). A partir de uma base de dados bruta sobre a movimentação dos objetos móveis, os autores investigam como detectar as trajetórias ali contidas, e transformar esses dados de forma eficiente para adequá-los ao modelo de dados.

Kuijpers; e Vaisman (2007) propõem um modelo de dados para trajetórias de objetos móveis usando DW Espacial. Segundo os autores, o uso de medidas e dimensões espaciais aumenta o poder de expressividade do modelo, além de simplificar a construção e processamento de algumas consultas.

Spaccapietra et al. (2008) introduzem o conceito de trajetórias semânticas. De acordo com esse novo conceito, uma trajetória pode ser semanticamente dividida em vários componentes, tais como *parada e movimento*.

Gomez; Kuijpers; e Vaisman (2007) similar ao trabalho de Kuijpers; e Vaisman (2007), mas representam trajetórias através de sua sequência de *paradas e movimentos* ao invés de sua sequência de observações, uma forma de compactar trajetórias.

Baltzer et al. (2008a) propõem um novo operador TrOLAP para agregação de trajetórias similares. Esse operador permite identificar grupos de objetos móveis que se movimentaram em paralelo durante algum tempo, ou que possuem rotas similares.

3.1 Proposta de Braz et al. (2007)

É um dos trabalhos pioneiros em DW de Trajetórias. Neste trabalho, é discutido como as tecnologias de DW existentes podem ser usadas para armazenar e agregar dados de trajetória mantidos de forma sumarizada. Uma preocupação específica dos autores é amenizar o problema da contagem distinta de trajetórias e/ou objetos móveis gerado quando utiliza-se essa forma de armazenamento, como discutido na Seção 2.5.2. Em resumo, o trabalho apresenta como maior contribuição:

1. Uma nova forma de agregar dados de trajetória sumarizados por célula espaço-temporal, apresentando boa aproximação na contagem distinta de trajetórias e/ou objetos móveis.

Neste trabalho propõe-se um modelo para DWTr orientado a tráfego, definido através de um esquema estrela, onde o fato é a célula espaço-temporal de uma grade regular, e as medidas representam as propriedades sumarizadas de todas as trajetórias que por ela passaram. A sumarização de trajetórias por célula espaço-temporal foi discutida na Seção 2.5.2.

O DWTr é formado pelas dimensões: espaciais X e Y; e temporal T. Por simplificação, foi adotado uma grade regular tridimensional obtida através da discretização do espaço e tempo.

Cada dimensão possui uma granularidade mínima de 30 unidades, definida em alguma unidade de medida de acordo com o domínio da aplicação. As dimensões, espaciais e temporal, obedecem a um conceito de hierarquia bem simples, baseada em níveis de agregação, como por exemplo, *nível 0 < nível 1 < nível 2* (ver Figura 3.1(b)).

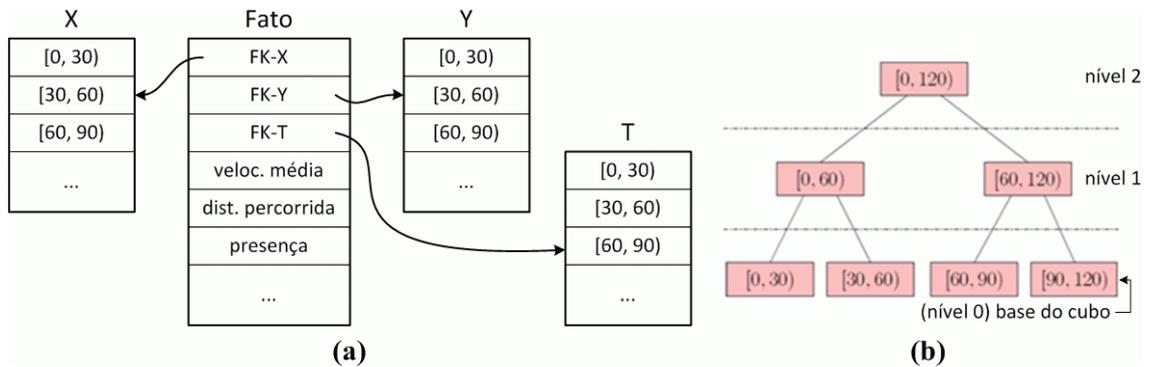


Figura 3.1: Modelo proposto por Braz et al. (2007). Em: (a) esquema estrela para DWTr. (b) hierarquia baseada em níveis de agregação.

O problema da contagem distinta é amenizado através do uso de medidas auxiliares. Essas medidas funcionam como contadores para o número de trajetórias que permaneceram ou deixaram a célula pelas suas diferentes saídas, como segue: (i) *presença* (*presence*): um contador do número de trajetórias permanecem dentro da célula espaço-temporal; (ii) *crossX*: contador do número de trajetórias que deixaram a célula pela borda espacial direita; (iii) *crossY*: contador do número de trajetórias que deixaram a célula pela borda espacial superior; e (iv) *crossT*: contador do número de trajetórias que atravessaram a borda temporal.

3.1.1 Críticas

O modelo de dados definido por Braz et al. (2007) é bem simples, servindo apenas para levantar uma discussão sobre o método proposto para amenizar o problema da contagem distinta. Nos parágrafos a seguir, serão apresentadas os pontos fortes e limitações identificadas neste trabalho.

Pontos fortes

- Apresenta uma solução para amenizar o problema da contagem distinta de trajetórias,

que ocorre quando se usa trajetórias armazenadas de forma sumarizada.

- *Pequena quantidade de dados armazenada.* Trajetórias sumarizadas por célula espaço-temporal, requerem pouquíssimos dados para armazenamento, como discutido na Seção 2.5.2. Essa pequena quantidade proporciona alto desempenho durante a análise TrOLAP.
- *Emprega métodos de interpolação,* utiliza a função de interpolação linear, para reconstruir os movimentos intermediários da trajetória entre duas observações consecutivas, como discutido na Seção 2.2.
- *Privacidade dos dados de trajetória,* trajetórias armazenadas de forma sumarizada sem o identificador do objeto móvel (ver Seção 2.5.2).

Problemas & Limitações

- As hierarquias implementadas, espacial e temporal, são baseadas em níveis de agregação (p.ex. nível 0 < nível 1 < nível 2), o que dificulta a análise dos dados (hierarquias muito abstratas). Os próprios autores assumem a necessidade de se usar hierarquias mais complexas para análise TrOLAP, por exemplo, *célula < bairro < cidade* para a dimensão espacial e *hora < dia < mês < ano* para o tempo.
- Não distingue *paradas de movimentos* em trajetórias.
- Proporciona apenas *análise orientada a tráfego*, já que não mantém as trajetórias dos objetos móveis. Todas as trajetórias de todos os objetos são armazenadas de forma sumarizada por célula espaço-temporal.
- Pode enfrentar o *problema do relacionamento parcialmente contido*, caso novas hierarquias espaciais sejam implementadas. Nada impede que uma mesma célula esteja contida em duas ou mais regiões, o que adiciona imprecisão as agregações, como discutido na Seção 2.3.1.
- Não *armazena informações semânticas*.
- Não *armazena dados espaciais*. Utiliza DW convencional.

Embora os experimentos realizados por Braz et al. (2007), tenham mostrado boa precisão para a contagem distinta, erro inferior a 10% no pior caso. Analisando os resultados de experimentos mais amplos realizados por Marketos et al. (2008) e Orlando et al. (2007) usando outras configurações e bases de dados, conclui-se que o método de Braz et al. (2007) não apresenta boa precisão nas seguintes situações:

- (i) Quando as células são grandes (p.ex. alguns quilômetros ou próximo a isso);
- (ii) Quando uma super-célula (célula em um nível superior) é obtida a partir a agregação de muitas sub-células (células em um nível inferior). Por exemplo, durante o roll-up cada super-célula é gerada a partir da agregação de 128 sub-células.
- (iii) Para cobrir grandes regiões (p.ex. alguns quilômetros), já que para isso é necessário: (a) usar células grandes; ou (b) agregar muitas sub-células pequenas. O que dificulta essa observação, é que Braz et al. (2007) não descrevem o tamanho do espaço territorial que está sendo analisado (p.ex. 100 km^2), apenas Marketos et al. (2008) apresentam esses dados.

Como a literatura entra em contradição, mais experimentos são necessários para verificar até que ponto o modelo de Braz et al. (2007) é eficaz. O modelo proposto por Braz et al. (2007) pode ser encontrado com maiores detalhes em sua tese de doutorado Braz (2009).

3.2 Proposta de Marketos et al. (2008)

Descrevem um processo ETL¹ para DWTr baseado no modelo de Braz et al. (2007). A partir de uma base de dados bruta gerada através do monitoramento dos objetos móveis, os autores investigam como identificar as trajetórias ali contidas (detecção de trajetórias) e transformar esses dados de forma eficiente para adequá-los ao modelo de dados do DW. Em resumo, o trabalho apresenta as seguintes contribuições:

¹ETL (*Extraction, Transformation, and Load* – Extração, Transformação, e Carga), descrito na Seção 2.6.1.

1. Um método para detecção de trajetórias, umas das etapas do processo ETL em DW de Trajetórias.
2. Um método eficiente para sumarização de dados de trajetórias por célula espaço-temporal.

O modelo proposto (ver Figura 3.2) é similar ao de Braz et al. (2007), preserva o mesmo fato (FACT_TBL) e medidas, mas possui dimensões e hierarquias mais adequadas para análise TrOLAP, como descrito no parágrafo a seguir.

De acordo com Marketos et al. (2008) um modelo de dados para DWTr flexível deve incluir pelo menos as dimensões: **Espaço** (SPACE_DIM): dimensão espacial, possui a hierarquia PARTITION_GEOMETRY < DISTRICT < STATE < COUNTRY (*célula da grade* < *bairro* < *estado* < *país*); **Tempo** (TIME_DIM): dimensão temporal, possui hierarquia cronológica (p.ex. hora < dia < mês < ano); e **Perfil do Objeto** (OBJECT_PROFILE_DIM): dimensão demográfica e tecnográfica, contém dados referentes aos objetos móveis (nesse caso, pessoas) e o dispositivo de localização usado (p.ex. GPS) para capturar a localização espaço-temporal do objeto, representado pelo atributo DEVICE_TYPE.

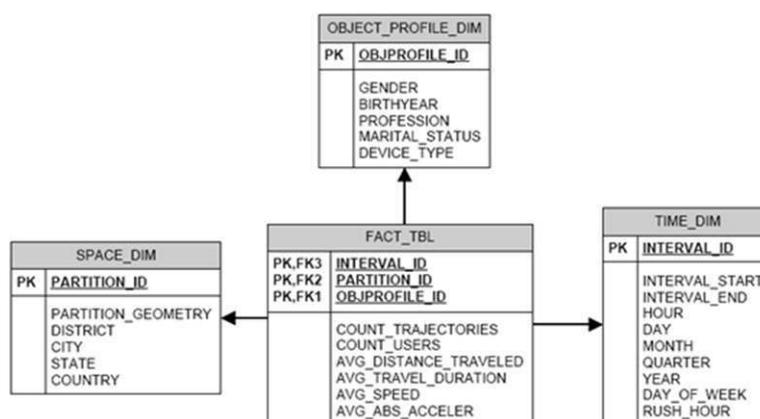


Figura 3.2: Esquema estrela apresentado por Marketos et al. (2008).

3.2.1 Críticas

Foi o único trabalho encontrado à apresentar um método para detecção de trajetórias. Como este trabalho é uma continuação do trabalho de Braz et al. (2007), de modo geral, ele também

herda seus pontos fortes e limitações, como descrito a seguir.

Pontos fortes

- *Pequena quantidade de dados armazenada*, o que proporciona análise TrOLAP eficiente. Assim como em (Braz et al., 2007).
- *Emprega métodos de interpolação*, utilizam interpolação linear.
- *Hierarquias de agregação adequadas* para análise TrOLAP. Por exemplo, hierarquia temporal, *hora < dia < mês < ano*.
- *Privacidade dos dados de trajetória*.
- *Armazena informações semânticas*, os autores defendem a inclusão de vários atributos e dimensões descritivas.
- *Dados de trajetórias georeferenciados*. Embora, não usem dados espaciais, as dimensões mantêm informações que possuem contexto espacial, como por exemplo, o nome da cidade (*city*).

Problemas & Limitações

- Não resolve o *problema do relacionamento parcialmente contido*. Uma mesma célula espaço-temporal pode estar contida em dois ou mais bairro, o que adiciona imprecisão as agregações (ver Seção 2.3.1).
- Não utiliza as informações semânticas durante a detecção de trajetórias, o que poderia aumentar a eficácia do método.
- Não distingue *paradas de movimentos* em trajetórias.
- Proporciona apenas *análise orientada a tráfego*, já que não mantém as trajetórias dos objetos móveis.
- Não *armazena dados espaciais*. Utilizam DW convencional.

Maiores informações sobre este trabalho podem ser encontradas na tese de doutorado de Marketos (2009) e no relatório técnico Theodoridis et al. (2009).

3.3 Proposta de Kuijpers; e Vaisman (2007)

Propõem um modelo de dados para trajetórias de objetos móveis usando DW Espacial, através da ferramenta PIET, que integra de forma federada os dados SIG, DW, e Objetos Móveis em um único ambiente. O modelo de dados proposto é uma extensão natural do modelo proposto em Gómez et al. (2007). Segundo os autores, o uso de medidas e dimensões espaciais aumenta o poder de expressividade do modelo, além de simplificar a construção e processamento de algumas consultas. Neste trabalho, é apresentado como maior contribuição:

1. Um modelo de dados usando DW Espacial para analisar trajetórias.

3.3.1 Críticas

Identificou-se neste trabalho os seguintes pontos fortes e limitações.

Pontos fortes:

- Proporciona *análise orientada a tráfego e trajetórias*. Entretanto, a ausência de métodos de interpolação para as trajetórias prejudica a análise dos dados, como discutido na Seção 2.2 (ver Figura 2.4).
- Permite *roll-up clássico e contagem distinta* para trajetórias e/ou objetos móveis. Entretanto, os dados obtidos são imprecisos, devido a ausência de métodos de interpolação, como no item anterior.
- *Relacionamento totalmente contido*, cada observação da trajetória armazenada está contida em apenas uma única região. Portanto, não enfrentam o problema do relacionamento parcialmente contido, discutido na Seção 2.3.1.
- *Hierarquias de agregação adequadas* para análise TrOLAP.
- *Armazena dados espaciais* nas dimensões. Utiliza DW Espacial.

Problemas & Limitações

- Usa a abordagem de integração federada entre SIG, DW, e Objetos Móveis. Kuijpers; e Vaisman (2007) propõem um modelo que mantém os dados geográficos (SIG), os dados necessários para análise OLAP (DW), e os dados sobre os objetos móveis, em locais distintos (isto é, banco de dados separados), o que gera sérios problemas de desempenho e, principalmente, inconsistência dos dados.
- Não distinguem *paradas de movimentos* em trajetórias.
- Usa o *modelo clássico*, armazenam toda a sequência de observações capturadas da trajetória.
- *Grande quantidade de dados armazenada*, o que gera problemas de desempenho, inviabilizando a análise TrOLAP. Para amenizar esse problema, no cenário de aplicação apresentado, as trajetórias possuem observações capturadas a cada 30 minutos, o que implica em trajetórias muito imprecisas (ver Seção 2.2 Figura 2.4).
- Não *permite roll-up para trajetórias*. Embora o modelo mantenha as trajetórias individuais dos objetos móveis, através de sua sequência de observações, o modelo não proporciona agregação por direção dos movimentos.
- Não *armazena informações semânticas*. Ausência de atributos e/ou dimensões descritivas.
- Não é *independente de SGBD*, requer a ferramenta PIET.

3.4 Proposta de Spaccapietra et al. (2008)

Neste trabalho, buscou-se construir um modelo conceitual para trajetórias que permitisse seu enriquecimento semântico. Para isso, cada trajetória possui duas faces: (i) *face geométrica*: formada pela sequência de observações da trajetória; e (ii) *face semântica*: que divide a face geométrica em diversos componentes, e transporta as informações semânticas, como descrito em detalhes na Seção 2.1. Em resumo, o trabalho apresenta as seguintes contribuições:

1. Os conceitos de *parada e movimento* em trajetórias. Amplamente usados na literatura.

2. Um *modelo conceitual para trajetórias semânticas*, que permite segmentar uma trajetória em diversos componentes, tais como *parada* e *movimento*, os quais podem transportar informações semânticas que dão significado a parte da trajetória a qual pertencem.

3.4.1 Críticas

Ao associar informações semânticas a trajetórias, Spaccapietra et al. (2008) conseguem: (i) aumentar o poder de expressividade do modelo, proporcionando mais informações úteis ao processo de tomada de decisão; e (ii) simplificam consideravelmente a formalização e processamento das consultas, como discutido na Seção 2.4.1. Nos parágrafos a seguir, serão apresentadas os pontos fortes e limitações identificados neste trabalho.

Pontos fortes

- Definem os conceitos de: (i) *parada* e *movimento*; e (ii) *trajetórias semânticas*, amplamente usado na literatura. Defendem o uso de informações semânticas para aumentar o poder de expressividade do modelo.
- Proporciona *análise orientada a tráfego e trajetórias*. Entretanto, de forma imprecisa para os movimentos, devido a sua sumarização, que impossibilita obter as medidas de partes dos movimentos.
- Proporciona *contagem distinta* de trajetórias e/ou objetos móveis, já que mantém os identificados da trajetória (traj-id) e objeto móvel (obj-id).
- *Pequena quantidade de dados armazenada*. No modelo proposto as trajetórias são mantidas de forma sumarizada através de suas paradas e movimentos. Os movimentos são representados na forma de transição entre paradas (p.ex. do Hotel H1 para o Banco B1), e os movimentos intermediários (da origem ao destino do movimento) são mantidos através do atributo linha (um dado espacial). Vale ressaltar que o modelo não impede que seja mantido as observações da trajetória.
- *Armazena dados espaciais*.

Problemas & Limitações

- Modelo não adequado para DWTr. Tem-se um bom modelo conceitual para representar trajetórias semânticas, mas não adequado para análise TrOLAP. Os próprios autores, sugerem como trabalhos futuros adaptarem seu modelo para DWTrs.
- Representa os movimentos de forma sumarizada, como transição entre paradas, tem a vantagem de requerer poucos dados, um registro apenas contendo a origem e o destino do movimento. Entretanto, perde-se as propriedades da trajetória ao longo do movimento.
- Enfrenta o *problema do relacionamento parcialmente contido* para os movimentos, devido a sua sumarização, um movimento mesmo pode pertencer a duas ou mais regiões.

3.5 Proposta de Gomez; Kuijpers; e Vaisman (2007)

É uma extensão do modelo proposto em Kuijpers; e Vaisman (2007), mas ao invés de representar trajetórias através de sua sequência de observações (modelo clássico), optaram por representar trajetórias através de sua sequência de *paradas e movimentos*, uma forma de compactar trajetórias. De certo modo, os autores estendem o modelo conceitual de trajetórias idealizado por Spaccapietra et al. (2008) para DW. Em resumo, o trabalho apresenta as seguintes contribuições:

1. Um modelo de dados usando DW Espacial para analisar paradas e movimentos em trajetórias.
2. Uma linguagem de consulta para o modelo de dados proposto.

3.5.1 Críticas

Como este trabalho é uma extensão do trabalho de Kuijpers; e Vaisman (2007), ele também herda alguns de seus pontos fortes e limitações, como descrito a seguir.

Pontos fortes

- Oferece *distinção entre paradas e movimentos*.
- Proporciona *análise orientada a tráfego e trajetórias*, apenas para paradas.
- Permite *roll-up clássico e contagem distinta*, apenas para paradas.
- *Pequena quantidade de dados de trajetória*, difere do trabalho de Kuijpers; e Vaisman (2007). Isso deve-se ao fato de que as paradas e movimentos são mantidos de forma sumarizada.
- *Relacionamento totalmente contido*, cada parada armazenada está contida em apenas uma única região.
- *Armazena dados espaciais*.

Problemas & Limitações

- Não é possível analisar os movimentos da trajetória de forma satisfatória. Porque não é mantido uma geometria representando o deslocamento do objeto móvel entre as paradas, como ocorre em Spaccapietra et al. (2008).
- Abordagem de integração federada entre SIG e DW, como em Kuijpers; e Vaisman (2007).
- Não é *independente de SGBD*, requer a ferramenta PIET.

3.6 Proposta de Baltzer et al. (2008a)

Propõem um novo operador TrOLAP para agregação de trajetórias similares. Esse novo operador permite identificar grupos de objetos móveis que se movimentaram em paralelo durante algum tempo, ou que possuem rotas similares. Através desse operador, por exemplo, pode-se identificar quais são as rotas mais usadas pela população para ir de um lugar ao outro, ou quais são indivíduos da população que possuem trajetórias semelhantes (ver Figura 3.3). Este trabalho tem como maior contribuição:

1. A implementação de um operador *medoid*, para agrupamento de trajetórias similares.

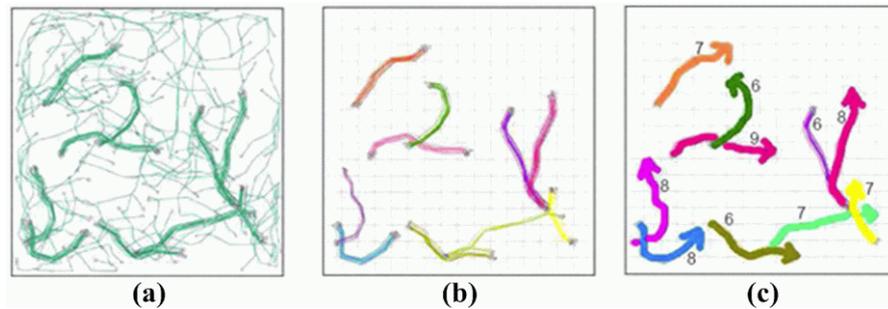


Figura 3.3: Operador medoid em ação. Em: **(a)** todas as trajetórias; **(b)** trajetórias similares identificadas e rotuladas; **(c)** trajetórias similares agregadas e contadas (Baltzer et al., 2008a).

Existem três implementações alternativas para o operador medoid:

Agrupamento por sobreposição (*group by overlap*): busca identificar objetos móveis que se movimentaram em paralelo por algum tempo, como ilustrado na Figura 3.4(b). Para isso, verifica-se o grau de similaridade das sequências de movimentos das trajetórias, logo, para duas trajetórias (t_1 e t_2) serem similares a *sequencia de movimentos de uma parte* da trajetória t_1 deve existir na trajetória t_2 . Por exemplo, considere a trajetória $t_1 = [a \rightarrow b \rightarrow c \rightarrow d \rightarrow e]$ e uma parte dela $[b \rightarrow c \rightarrow d]$, pode-se dizer que t_1 possui movimentos (em sequencia) similares a t_2 , se a sequencia de movimentos $[b \rightarrow c \rightarrow d]$ de t_1 existir em t_2 .

Agrupamento por intersecção (*group by intersection*): busca identificar objetos móveis que possuem trajetórias similares, como ilustrado nas Figura 3.3 e 3.4(a). Para isso, verifica-se para cada par de trajetórias o número de movimentos coincidentes, em outras palavras, o que determina se duas trajetórias t_1 e t_2 são similares é o tamanho do conjunto intersecção de seus movimentos ($t_1 \cap t_2$). Diferente do agrupamento por sobreposição esses movimentos não precisam estar em sequencia.

Agrupamento por sobreposição e intersecção (*group by overlap and intersection*): é uma fusão das duas abordagens anteriores. Para duas trajetórias serem agrupadas: (i) suas trajetórias devem ser similares; e (ii) suas sequencias de movimentos também devem ser similares.

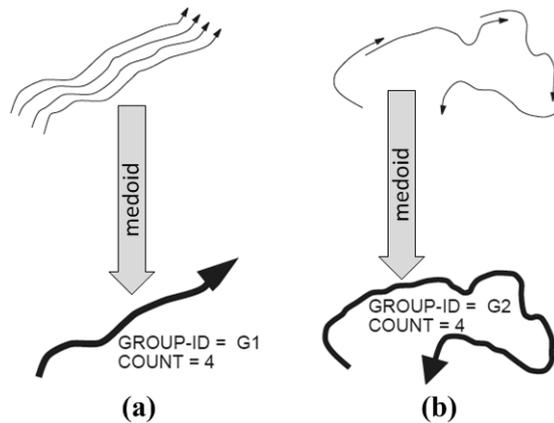


Figura 3.4: Agrupamento em: **(a)** por Intersecção; e **(b)** por Sobreposição (Baltzer et al., 2008a) (adaptado).

3.6.1 Críticas

O modelo de dados definido é bem simples, servindo apenas para validar a solução proposta, o operador *medoid*. Nesta seção serão apresentadas alguns pontos fortes, problemas, e limitações identificados neste trabalho.

Pontos fortes

- Proporciona *análise orientada a tráfego e trajetórias*. Entretanto, a ausência de métodos de interpolação prejudica a análise dos dados.
- Permite *roll-up clássico* e *contagem distinta* para trajetórias e/ou objetos móveis, entretanto, os dados obtidos são imprecisos, devido a ausência de métodos de interpolação.
- Implementa o operador *medoid*, para agrupamento de trajetórias. Mas devido a grande quantidade dos dados de trajetória, pode enfrentar problemas de desempenho.
- *Relacionamento totalmente contido*, cada observação da trajetória está contida em apenas uma única região.

Problemas & Limitações

- Não distingue *paradas* de *movimentos* em trajetórias.
- Os autores não explicam como é feito o *roll-up para trajetórias*, que é fundamental para o sucesso do operador *medoid* proposto.

- Usa o *modelo clássico*, mas não o descrevem em detalhes. Logo, não se pode tirar muitas conclusões a esse respeito.
- *Grande quantidade de dados armazenada*, mantém todas as observações da trajetória.
- Ausência de *informações semânticas*.
- Não *armazena dados espaciais*. Utilizam DW convencional.
- Não é *independente de SGBD*, devido a extensão implementada para o operador *medoid*.

Como ainda não existe um consenso na literatura sobre os nomes dos operadores TrO-LAP, alguns trabalhos utilizam o termo *medoid* para referir-se ao operador que faz agrupamento de trajetórias similares, como em Pelekis et al. (2008). Mas, Baltzer et al. (2008a) preferem usar o termo GROUP_TRAJECTORIES para referir-se a esse operador.

O trabalho de Baltzer et al. (2008a) pode ser encontrado em maiores detalhes em seu relatório técnico Baltzer et al. (2008b).

3.7 Considerações Finais

Analisando os trabalhos relacionados à DW de Trajetórias, identificou-se quais são as características desejáveis para um modelo de DWTr, como descritas a seguir:

1. Distinção entre *paradas* e *movimentos* em trajetórias (abreviado para, **distingue paradas e movimentos**): essa distinção é fundamental para análise correta de trajetórias, pois a inclusão de dados sobre paradas (que possuem velocidade igual a zero, ou próximo a isso) na análise de movimentos pode provocar forte discrepância entre os dados analisados e os reais, e vice-versa. As próprias consultas também deixam bem claro está distinção, horas requerem apenas dados de paradas, horas requerem apenas dados de movimentos.

Dos trabalhos analisados, apenas Spaccapietra et al. (2008) e Gomez; Kuijpers; e Vaisman (2007) atendem ao requisito.

2. Análise orientada a tráfego: o modelo permite analisar o comportamento dos objetos móveis sobre as regiões no espaço e tempo (ver Seção 2.4).

Baltzer et al. (2008a); Kuijpers; e Vaisman (2007) apresentam problemas nesse requisito, por não usarem trajetórias interpoladas, como discutido na Seção 2.2. Spaccapietra et al. (2008) usam trajetórias interpoladas, mas a sumarização por movimento impossibilita obter as medidas de partes de seus movimentos.

3. Análise orientada a trajetórias: o modelo permite analisar o deslocamento dos objetos móveis entre as regiões no espaço e tempo (ver Seção 2.4).

Nenhum dos trabalhos analisados atende ao requisito de forma satisfatória. Em Baltzer et al. (2008a); Spaccapietra et al. (2008) o modelo não permite agregação por direção do movimentos, e em Gomez; Kuijpers; e Vaisman (2007) mantém apenas as paradas, portanto, permitem analisar apenas os movimentos entre paradas.

4. Roll-up/drill-down clássico: permitem visualizar o comportamento dos objetos móveis por região sobre diferentes níveis de detalhe.

Baltzer et al. (2008a); Kuijpers; e Vaisman (2007) não atendem ao requisito de forma satisfatória, porque não usam trajetórias interpoladas.

5. Roll-up/drill-down para trajetórias: através desses operadores TrOLAP, consegue-se visualizar os deslocamentos da trajetória sobre diferentes níveis de detalhe. Por exemplo, visualizar os deslocamentos da trajetória de bairro em bairro, ou de cidade em cidade.

Nenhum dos trabalhos relacionados permitem esse tipo de roll-up.

6. Medoid: um operador TrOLAP para agrupamento de trajetórias similares.

Baltzer et al. (2008a) atendem ao requisito, mas não de forma satisfatória, devido a grande quantidade dos dados de trajetória, que gera problemas de desempenho.

7. Contagem distinta para trajetórias e/ou objetos móveis (abreviado para, **contagem distinta**): devido a natureza dinâmica dos objetos móveis alguns DWTrs enfrentam problemas relacionados com a contagem distinta (ver Seção 2.5.2).

Enfrentam o problema da contagem distinta, os trabalhos de: (i) Baltzer et al. (2008a) e todos que usam o modelo clássico sem usar interpolação; e (ii) todos que usam o modelo orientado a tráfego, mantendo as trajetórias de forma sumarizada por célula espaço-temporal, como em Braz et al. (2007); Marketos et al. (2008).

8. Privacidade dos dados de trajetória (privacidade dos dados): existem aplicações como na telefonia móvel, onde não é permitido manter as trajetória de seus clientes por violarem questões de privacidade (Tao; e Papadias, 2005).

Não apresentam o problema da privacidade dos dados de trajetória, todos os modelos que: (i) descartam o identificador do objeto móvel (obj-id), como em Braz et al. (2007); Marketos et al. (2008), ou (ii) implementam algum artifício que impossibilite identificar o objeto móvel.

9. Pequena quantidade dos dados de trajetória (pequena quantidade de dados): um dos desafios enfrentados quando se lida com trajetórias, é sua grande quantidade de dados, que inviabiliza a maioria das soluções. Portanto, é desejável que um modelo necessite de uma pequena quantidade de dados para representar trajetórias.

Em geral, os modelos orientados a tráfego que mantém trajetórias de forma sumarizada, não apresentam problemas com a quantidade de dados, como em Braz et al. (2007); Marketos et al. (2008); Spaccapietra et al. (2008).

Como será discutido no Capítulo 5, o presente trabalho proporciona uma compactação bastante significativa para paradas, entretanto, o mesmo não pode ser dito para movimentos.

10. Relacionamento totalmente contido (Relac. totalmente contido): para obter precisão nas agregações por região, cada dado deve pertencer exclusivamente a uma única região, e não a duas ou mais, como acontece no relacionamento parcialmente contido, o que gera problemas, como discutido na Seção 2.3.1.

Em geral, os trabalhos que armazenam trajetórias de forma sumarizada enfrentam o problema do relacionamento parcialmente contido, como em Braz et al. (2007);

Marketos et al. (2008); Spaccapietra et al. (2008).

11. Modelo semântico: além dos dados espaço-temporais, para aumentar o poder de expressividade do modelo, é desejável que se mantenha informações semânticas sobre cada uma de suas entidades (p.ex. objetivo da trajetória, as horas de maior congestionamento).

12. Reconstrução de trajetórias: é um processo importante e complexo dentro da fase de extração dos dados no processo ETL em DWTrs. A partir de uma base de dados bruta gerada através do monitoramento dos objetos móveis, por meio desse processo identifica-se quais são as paradas e trajetórias ali contidas (ver Seção 2.2).

Nenhum dos trabalhos relacionados atendem ao requisito.

13. Abordagem integrada entre SIG e DW (SIG + DW (integrado)): existem duas abordagens de integração entre SIG e DW: na (i) *abordagem integrada*, todos os dados são mantidos no DW, tanto os dados convencionais (p.ex. nome do produto, dia da venda) quanto os espaciais (p.ex. geometria do bairro); e na (ii) *abordagem federada*, os dados são mantidos em dois ambientes separados, o SIG mantém os dados espaciais e alguns outros que possuem contexto espacial (p.ex. nome do bairro), e o DW mantém os dados necessários para análise OLAP (p.ex. dia, mês, ano – para proporcionar agregação por tempo). A abordagem federada pode enfrentar problemas de desempenho e, principalmente, inconsistência dos dados (Matousek; Mordacik; e Janku, 2001).

14. Independente de SGBD: o modelo pode ser implementado em qualquer banco de dados espacial, e usa SQL padrão. Ou seja, o modelo não depende de uma tecnologia específica de banco de dados para funcionar.

Dos trabalhos analisados, apenas Gomez; Kuijpers; e Vaisman (2007); Kuijpers; e Vaisman (2007) não atendem ao requisito, pois dependem da ferramenta PIET.

Após conhecer os trabalhos relacionados e suas características mais relevantes, apresentam-se no próximo capítulo as propostas do presente trabalho. Serão discutidos em

	Características		Trabalhos	Braz et al. (2007)	Marketos et al. (2008)	Kuijpers e Vaisman (2007)	Spaccapietra et al. (2008)	Gómez et al. (2007)	Baltzer et al. (2008)	Trabalho Proposto
1	Distinção Paradas e Movimentos						●	●		●
2	Análise orientada a Tráfego			●	●	⊕	⊕	● (p)	⊕	●
3	Análise orientada a Trajetórias					⊕	⊕	● (p)	⊕	●
4	Roll-up/ drill-down	clássico		●	●	⊕		● (p)	⊕	●
5		para trajetórias								●
6	Medoid								⊕	
7	Contagem distinta			⊕	⊕	⊕	●	● (p)	⊕	●
8	Privacidade dos dados			★	★					
9	Pequena quantidade de dados			★	★		●	●		⊕
10	Relac. totalmente contido					●		● (p)	●	●
11	Modelo semântico						★			●
12	Reconstrução de trajetórias									
13	SIG + DW (integrado)									●
14	Independente de SGBD			●	●		●			●
	● atende		★ atende muito bem (solução eficaz)							
	⊕ atende, mas apresenta problemas		(p) apenas para paradas							

Figura 3.5: Tabela de Características dos Trabalhos Relacionados.

detalhes: (i) o modelo proposto para DW de Trajetórias, que distingue paradas de movimentos; é semântico; e proporciona análise orientada a tráfego e trajetórias; (ii) o método para compactação de trajetórias usado, para reduzir a quantidade de dados armazenada; e (iii) como é feito o roll-up para trajetórias.

Capítulo 4

Um Modelo Semântico para DWTr com Suporte à Agregação por Direção dos Movimentos

Existem diferentes formas de se processar dados de trajetórias, em geral, elas se baseiam no armazenamento desses dados em bancos de dados convencionais, que posteriormente podem ser minerados para possibilitar a descoberta do conhecimento. Um dos inconvenientes dessa solução é que esses dados não estão dispostos na forma adequada para sua análise, o que implica em tempos de resposta muito longos (p.ex. horas ou dias), inviabilizando seu uso. Uma forma adequada de processar esses dados seria disponibilizá-los em um Data Warehouse (DW), um banco de dados otimizado para realizar análise sobre grandes volumes de dados.

Entretanto, a modelagem tradicional para dados de trajetória possui alguns inconvenientes: **(i)** a quantidade de dados armazenada é muito grande até mesmo para um DW; e **(ii)** como construir um modelo que possibilite analisar o comportamento dos objetos móveis: *sobre e entre* as regiões no espaço e tempo, **(ii.a)** análise orientada a tráfego e **(ii.b)** análise orientada a trajetórias, respectivamente; e **(iii)** como atender a esses requisitos de forma independente de SGBD. Os trabalhos encontrados na literatura atendem apenas aos requisitos

(i) e (ii.a).

As demais seções deste capítulo estão organizadas como segue. Na *Seção 4.1* é apresentado o cenário de aplicação exemplo usado durante o restante deste trabalho. Na *Seção 4.2* é descrito o modelo proposto para DW de Trajetórias, seus fatos e dimensões. Na *Seção 4.3* são apresentados os procedimentos necessários para preparar os dados antes de serem carregados no DW. Finalmente, na *Seção 4.4* é descrito a forma encontrada para proporcionar agregação por direção dos movimentos da trajetória (roll-up para trajetórias).

4.1 Cenário de Aplicação

Nesta seção é apresentado um cenário de aplicação usando DW de Trajetórias, o **gerenciamento de tráfego urbano**, o qual será utilizado durante o restante deste trabalho. Para esse cenário exemplo, suponha que uma determinada organização governamental esteja disposta a melhorar o tráfego das cidades brasileiras. Para isso, ela está disposta a implantar nas maiores cidades um sistema para monitoramento e análise de tráfego urbano. Esse sistema deverá ajudar as prefeituras municipais na detecção de problemas de tráfego e no processo de tomada de decisões, como por exemplo: quando e onde construir uma nova avenida para ajudar no escoamento do tráfego; na criação de novas linhas de ônibus; para avaliar a necessidade de se construir um metrô na cidade; entre outras questões ligado ao planejamento da cidade. Para viabilizar o projeto, em cada cidade analisada, o governo precisa:

Monitorar os indivíduos de uma parcela representativa da população: para isso, cada indivíduo é monitorado através de seu telefone celular equipado com um GPS, o qual captura sua localização espaço-temporal a cada 20 segundos. Esses dados são armazenados e posteriormente transmitidos a um servidor. Todos os indivíduos monitorados são voluntários, e recebem benefícios do governo por participarem do projeto.

Os voluntários do projeto podem usar diversos meios de transporte para se locomoverem pela cidade, como: ônibus, carro, moto, bicicleta, ou até mesmo a pé. Essa tipo de informação é mantido pela trajetória correspondente ao deslocamento, cada

trajetória se relaciona com apenas um meio de transporte.

Manter informações detalhadas sobre cada voluntário do projeto: para ajudar no enriquecimento semântico das trajetórias, cada voluntário fornece as seguintes informações: (i) *dados pessoais*: sexo, idade, estado civil, profissão, endereço residencial e comercial; (ii) *locais mais frequentados e quando isso ocorre*: local de trabalho, casa, bares, escola das crianças; (iii) *rotas comumente usadas para ir de um lugar ao outro*; (iv) *meio de transporte utilizado*; e qualquer outro tipo de informação que ajude a enriquecer semanticamente a base de dados.

Manter informações sobre cada região analisada: para isso, o governo conta com as informações fornecidas pelo IBGE¹ para cada região, tais como número de habitantes, renda per capita, custo de vida, entre outras informações. Somado a essas informações, o sistema dispõe de *dados espaciais*, como: ruas (representadas através de polilinhas); bairros (polígonos); e regiões de interesse (ROIs) (polígonos), por exemplo, hotéis, restaurantes, escolas, lojas. Associado a essas regiões podem existir eventos que ocorreram no momento em que o objeto atravessava a região, como shows, congestionamentos, alagamentos, acidentes, entre outros.

Na Figura 4.1 é exibido o mapa simplificado do cenário de aplicação exemplo. Neste cenário é ilustrado o comportamento de duas pessoas (objetos móveis Obj1 e Obj2) em seu cotidiano, em termos de suas trajetórias. De acordo com esse cenário, Obj1 e Obj2 possuem o respectivo comportamento:

Obj1: [**Cond1::Rest1**] O objeto móvel Obj1 deixa o condomínio Cond1 com destino ao restaurante Rest1, onde costuma realizar suas refeições, durante sua trajetória (de Cond1 para Rest1, denotado a partir de agora por *Cond1::Rest1*) o objeto para por

¹IBGE (Instituto Brasileiro de Geografia e Estatística): é uma importante empresa pública responsável por coletar e fornecer informações e estudos de natureza econômica, geográfica, cartográfica, demográfica, entre outras informações necessárias ao planejamento econômico e social do país. Disponível em www.ibge.gov.br. Acesso em 01 nov 2010.

cerca de 15 minutos em p1 e p2 devido a congestionamentos. Cerca de 40 minutos depois, [Rest1::Hotel1] Obj1 parte para Hotel1 (seu local de trabalho), trajetória Rest1::Hotel1, durante o trajeto Obj1 atravessa a Ponte1, melhor caminho para alcançar Hotel1. Depois do expediente de trabalho, [Hotel1::Cond1] Obj1 deixa Hotel1 com destino a Cond1 (sua casa), durante o trajeto passa pelo estádio de futebol Estadio1, onde em dias de jogos podem ocorrer fortes congestionamentos.

Obj2: O comportamento do objeto móvel Obj2 é similar a Obj1. [Cond1::Rest1] Obj2 deixa o condomínio Cond1 em direção ao restaurante Rest1 (trajetória Cond1::Rest1), como Obj2 realiza uma rota um pouco diferente de Obj1, ele não enfrenta os congestionamentos enfrentados por Obj1 (paradas p1 e p2). [Rest1::Hotel1] Deixando Rest1, Obj2 vai para Hotel1 (trajetória Rest1::Hotel1), em seguida, [Hotel1::Cond1] em vez de voltar direto para casa Cond1 como sempre faz, Obj2 foge a rotina e realiza uma parada de 15 minutos em uma região desconhecida R1. Para os dois últimos deslocamentos, poderia-se pensar em representa-los através das trajetórias Hotel1::R1 e R1::Cond1, respectivamente, mas como a parada em R1 é uma exceção à rotina, assume-se que a finalidade de Obj2 é atingir Cond1, portanto, o que ocorre em R1 é considerado uma parada na trajetória Hotel1::Cond1.

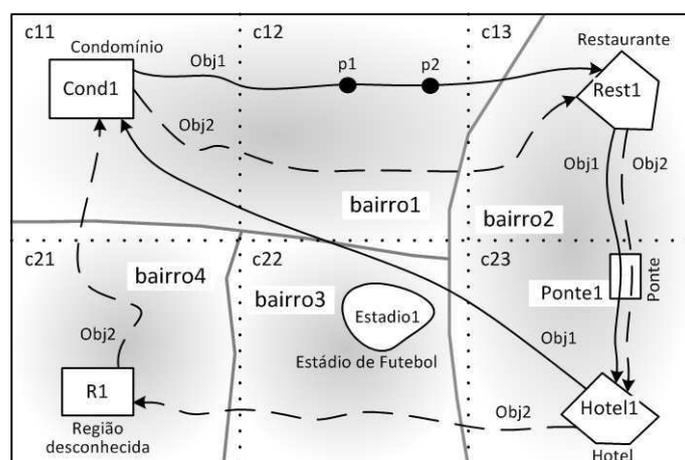


Figura 4.1: Cenário de Aplicação Exemplo.

Considerando cenários como esse, gestores do transporte estão interessados em consultas do tipo:

- A velocidade média dos veículos circulando na cidade por região a cada 20 minutos?
- O número de veículos que deixam o Bairro1 em direção ao Bairro2 por hora?
- Qual é a parte da população afetada com a interdição da Ponte1?
- A velocidade média dos veículos que circulam próximos ao estádio de futebol Estadio1 em dias de jogos a cada 20 minutos?
- Os locais mais visitados e o tempo de parada dos indivíduos que circulam no Bairro1 por dia da semana?

4.2 Modelo Proposto

O modelo proposto é uma extensão de um DW Espacial (Bédard; Merrett; e Han, 2001), o qual adota o conceito de trajetórias semânticas, idealizado por Spaccapietra et al. (2008) descrito na Seção 2.1. De acordo com esse conceito, uma trajetória possui duas faces: (i) *face geométrica*: que corresponde a sequência de observações (x, y, t) capturadas da trajetória real; e a (ii) *face semântica*: que divide a face geométrica em vários componentes (p.ex. parada e movimento), os quais podem transportar informações semânticas, que dão significado à parte da trajetória a qual pertence. Sendo assim, na perspectiva de banco de dados, pode-se representar uma trajetória através de sua sequência de *paradas* e *movimentos*, que caracterizam dois fatos e um conjunto de dimensões, onde os fatos mantem a face geométrica e as dimensões a face semântica, como descrito nas seções a seguir.

Para representação de trajetórias, o *espaço geográfico* é discretizado por uma grade regular formada por um conjunto de células espaciais (ver Seção 2.3), e o *tempo* é discretizado em intervalos de tempo regulares. Em geral, intervalos com duração de algumas dezenas de minutos (p.ex. 20 minutos).

4.2.1 Fatos

Para analisar trajetórias de forma correta, é fundamental oferecer uma distinção clara entre paradas de movimentos em trajetórias. A inclusão de dados sobre paradas na análise de movimentos pode provocar forte discrepância entre os dados analisados e os reais, e vice-versa. Por exemplo, ao se incluir dados sobre paradas no cálculo da velocidade média de uma região, tem-se uma forte impressão de que a velocidade na região analisada está baixa, devido a influência dos dados sobre paradas (cuja velocidade é igual a zero, ou muito próximo a isso) durante as computações. Dos trabalhos analisados (ver Capítulo 3), apenas Gomez; Kuijpers; e Vaisman (2007) e Spaccapietra et al. (2008) fazem essa distinção. Dessa forma, no modelo proposto são adotados dois fatos: *fato movimento* e *fato parada*, descritos a seguir:

Fato Movimento: armazena os dados referentes aos movimentos na forma sumarizada por célula espaço-temporal. É representado pelas medidas velocidade média, espaço percorrido, tempo decorrido, entre outras.

Como a análise OLAP para Trajetórias está interessada em analisar o comportamento dos objetos móveis sobre regiões espaço-temporais, que em seu maior nível de detalhe corresponde a algumas centenas ou milhares de metros. Logo, armazenar todas as observações capturadas da trajetória além de desnecessário, gera problemas de desempenho relacionados a grande quantidade de dados armazenada (Braz et al., 2007; Jensen, 2002; Orlando et al., 2007). Portanto, a solução proposta consiste em armazenar dados de trajetória na forma sumarizada por célula espaço-temporal, isto é, a sequencia de observações que atravessam uma dada célula espaço-temporal são agregadas e armazenadas na forma de um único registro.

Fato Parada: armazena os dados referentes às paradas na forma sumarizada por parada. Em geral, é representado pelas medidas *tempo de parada* e *local de parada* (um dado espacial, representado por um simples ponto).

De forma análoga ao que acontece no fato movimento, as observações referentes a paradas são armazenadas na forma sumarizada por Região de Interesse (RoI), com a

diferença de que para paradas outras características são analisadas, como por exemplo, número de paradas, tempo de parada, atividades realizadas, entre outras.

Para se ter uma noção de como funciona a sumarização das observações, considere a sumarização dos atributos *velocidade* e *tempo de parada* para a trajetória ilustrada na Figura 4.2. As observações dessa trajetória sofrem as seguintes sumarizações: (i) as observações de M_1 referentes a movimentos são agregadas e armazenadas na forma de um único registro no fato movimento, com referência a célula $c31$. Dessa forma, a medida velocidade em $c31$ representa a velocidade média calculada a partir de todas as observações de M_1 ; (ii) a sequencia de observações P_1 , referentes a paradas sobre o RoI $R1$ são agregadas e armazenadas no fato parada na forma de um único registro, com referência a $R1$; esses mesmos passos são repetidos para as sequencias M_2 , M_3 , e P_2 , (iii) M_2 e M_3 são agregados por $c22$ e $c12$, respectivamente, e armazenados no fato movimento, as observações P_2 são agregadas por $R2$ e armazenada no fato parada, como mostrado na figura. A sumarização das observações de uma trajetória é descrito em detalhes na Seção 4.3.

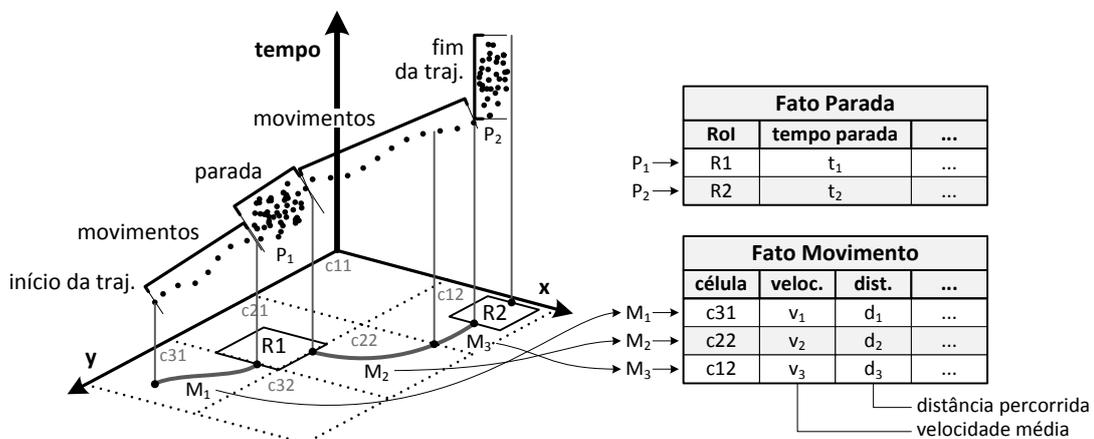


Figura 4.2: Sumarização das observações de uma trajetória.

4.2.2 Dimensões

Para atingir os objetivos propostos, o modelo adotado incorpora as seguintes dimensões, como ilustrado na Figura 4.3:

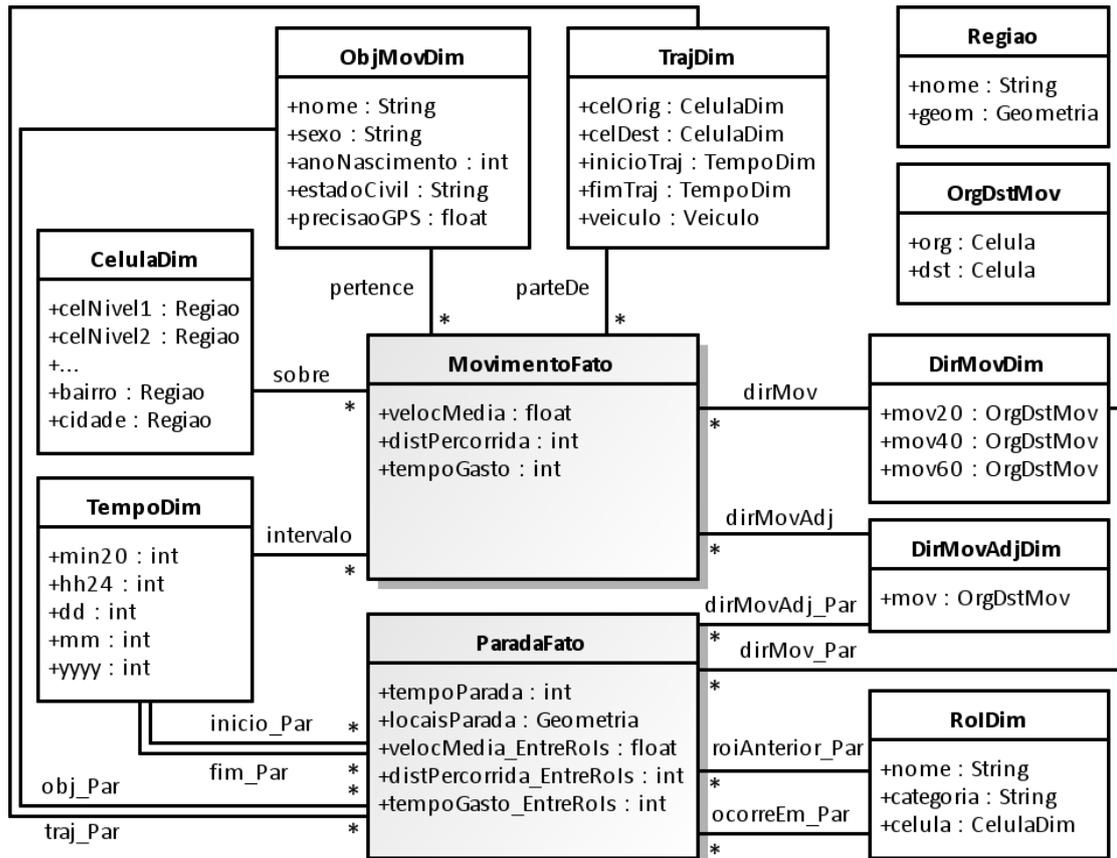


Figura 4.3: Modelo Proposto ilustrado por um diagrama UML.

Objeto Móvel: (ObjMovDim) dimensões demográfica e tecnográfica. A *face demográfica* mantém dados sobre os objetos móveis. Por exemplo, no caso de indivíduos, nome, sexo, idade, profissão e estado civil (estadoCivil). A *face tecnográfica* mantém dados sobre o dispositivo de localização usado, tais como a precisão do GPS usado (precisaoGPS).

Trajectoria: (TrajDim) dimensão descrita, contém as informações sobre a trajetória como um todo. Basicamente possui informações: (i) *espaciais*: origem e destino da trajetória; (ii) *temporais*: início e fim da trajetória; e (iii) *descrita*: objetivo da trajetória (p.ex. indo do trabalho para casa), meio de transporte usado (carro, moto, ônibus).

Célula: (CelulaDim) dimensão espacial, armazena as células espaciais da grade regular. Em geral, possui a hierarquia *célula* < *bairro* < *cidade*.

Tempo: (TempoDim) dimensão temporal, definida em intervalos de tempo. Mantém os

eventos que ocorreram para cada intervalo de tempo, como por exemplo, shows, partidas de futebol, acidentes de trânsito, entre outros.

Região de Interesse: (RoIDim) dimensão espacial, armazena os dados sobre as regiões de interesse (RoIs) tais como nome, categoria (hotel, shopping, universidade, entre outros) e dados espaciais (polígono que representa o RoI). Em geral, possui a hierarquia *roi < célula < bairro < cidade*.

Direção do Movimento: (DirMovDim) e (DirAdjMovDim), representada pelas dimensões: (i) DirMovDim: (direção do movimento entre regiões), mantém a direção do movimento entre as regiões no tempo, através do par origem-destino do movimento (por exemplo, do bairro1 para o bairro3); e (ii) DirMovAdjDim: (direção do movimento entre regiões adjacentes) similar a DirMovDim.

A *análise orientada a tráfego* é proporcionada pelas dimensões CelulaDim e TempoDim, e a *análise orientada a trajetórias* é pelo conjunto de dimensões direção do movimento: DirMovDim, DirMovAdjDim e RoIDim. São essas dimensões que tornam possível o roll-up para trajetórias, como descrito em detalhes na Seção 4.4.

4.3 Carga de Dados

Nesta seção são descritos os passos necessários para transformar a sequência de observações capturadas da trajetória de forma a se adequarem ao modelo proposto. Este trabalho pressupõe que os componentes da trajetória, tais como paradas e movimentos, já foram previamente identificados e as anotações semânticas incluídas. Os passos para realizar a carga de dados são os seguintes:

Passo 1 – Sumarização das observações referentes às paradas: as observações referentes a cada parada da trajetória são sumarizadas e armazenadas na forma de um único registro no fato *parada*.

Exemplo: na Figura 4.4(a) as observações no intervalo (00h53, 01h33) referentes

à parada p1 são sumarizadas e armazenadas como um único registro, como é mostrado Figura 4.4(b). Para facilitar a compreensão dos exemplos, o identificador das observações da trajetória coincide com o momento de captura da observação.

Passo 2 – Identificação e exclusão dos movimentos dentro de RoIs: a análise dos movimentos da trajetória está interessada nos dados sobre a movimentação dos objetos na célula (ou seja, nas ruas), mas não dentro de regiões de interesse (RoIs). Para distinguir os movimentos que ocorreram *dentro* e *fora* dessas regiões, novas observações são acrescentadas à trajetória, nos pontos de intersecção dela com as bordas espaciais dos RoIs.

Exemplo: na 4.4(a) é acrescentada a observação 00h14 ao intervalo (00h13, 00h33) para dividi-lo em (00h13, 00h14) e (00h14, 00h33), movimentos *dentro* e *fora* do RoI H1, respectivamente.

Passo 3 – Divisão dos movimentos *por intervalo de tempo e por célula espacial*: para que os movimentos em um dado intervalo se encaixem perfeitamente dentro dos limites de cada célula, novas observações são acrescentadas à trajetória nesse intervalo, nos pontos que intersectam as bordas espaciais e temporais das células. Isso é necessário para uma correta análise dos dados de trajetória. De forma análoga é feito a *divisão por intervalo de tempo*.

Exemplo (divisão por célula espacial): na Figura 4.4(a), os movimentos no intervalo (00h13, 00h33) ultrapassam os limites espaciais das células c23 e c22. Para que os movimentos se encaixem dentro dessas células, acrescenta-se a observação 00h18 ao intervalo (00h13, 00h33) para distribuir seus movimentos entre (00h13, 00h18) e (00h18, 00h33), os quais respeitam os limites de c23 e c22, respectivamente.

Exemplo (divisão por intervalo de tempo): na Figura 4.4(a), os movimentos no intervalo (01h53, 02h13) ultrapassam a barreira temporal, sendo assim, é acrescentada a observação 02h00 a esse intervalo. Considerando células espaço-temporais com duração de uma hora. É mostrada na Figura 4.4(d) a divisão dos movimentos da trajetória H1::C1.

Passo 4 – Sumarização dos movimentos por célula espaço-temporal: após a divisão dos movimentos por célula espaço-temporal (passo 3), os movimentos dentro de cada célula são sumarizados e armazenados na forma de um único registro no fato movimento.

Exemplo: na Figura 4.4(d) são ilustrados os movimentos da trajetória H1::C1, divididos e sumarizados por célula espaço-temporal.

Para possibilitar agregação por direção dos movimentos (roll-up para trajetórias), após a execução dos passos descritos acima, é necessário extrair e armazenar as direções do movimento, como discutido na seção a seguir.

Por questões de simplificação, para reconstruir os movimentos intermediários da trajetória entre duas observações consecutivas, foi usado a função de interpolação linear (ver Seção 2.2). Isso possibilita a detecção de movimentos sobre as células que seriam impossíveis de serem detectados sem a interpolação da trajetória. Por exemplo, suponha que a observação 33 na Figura 4.4(a) não existisse. Sem essa observação, e sem a interpolação da trajetória, seria impossível detectar o movimento do objeto móvel sobre a célula c22.

4.4 Roll-up para Trajetórias

Um dos desafios dos DWTrs é proporcionar *múltiplas representações para trajetórias* (*multiple representation of trajectories*) (Pelekis et al., 2008, p. 207), ou seja, proporcionar a representação de trajetórias e movimentos sobre diferentes perspectivas e níveis de granularidade. Por exemplo, para uma mesma trajetória, pode-se desejar visualizar o deslocamento do objeto móvel entre bairros, ou entre RoIs, ou de hora em hora, entre outras representações. Para resolver esse problema, a solução proposta consiste em obter múltiplas representações através de agregações das células espaço-temporais da trajetória (seu elemento mais básico), proporcionada pelo conjunto de dimensões direção do movimento.

Por exemplo, na Figura 4.4(a), é exibida a trajetória H1::C1, armazenada no fato movimento através da sequencia de células espaço-temporais [c23:00h, c22:00h, c21:00h, c21:01h, c11:01h, c11:02h]. Os *movimentos entre regiões adjacentes* de H1::C1 podem

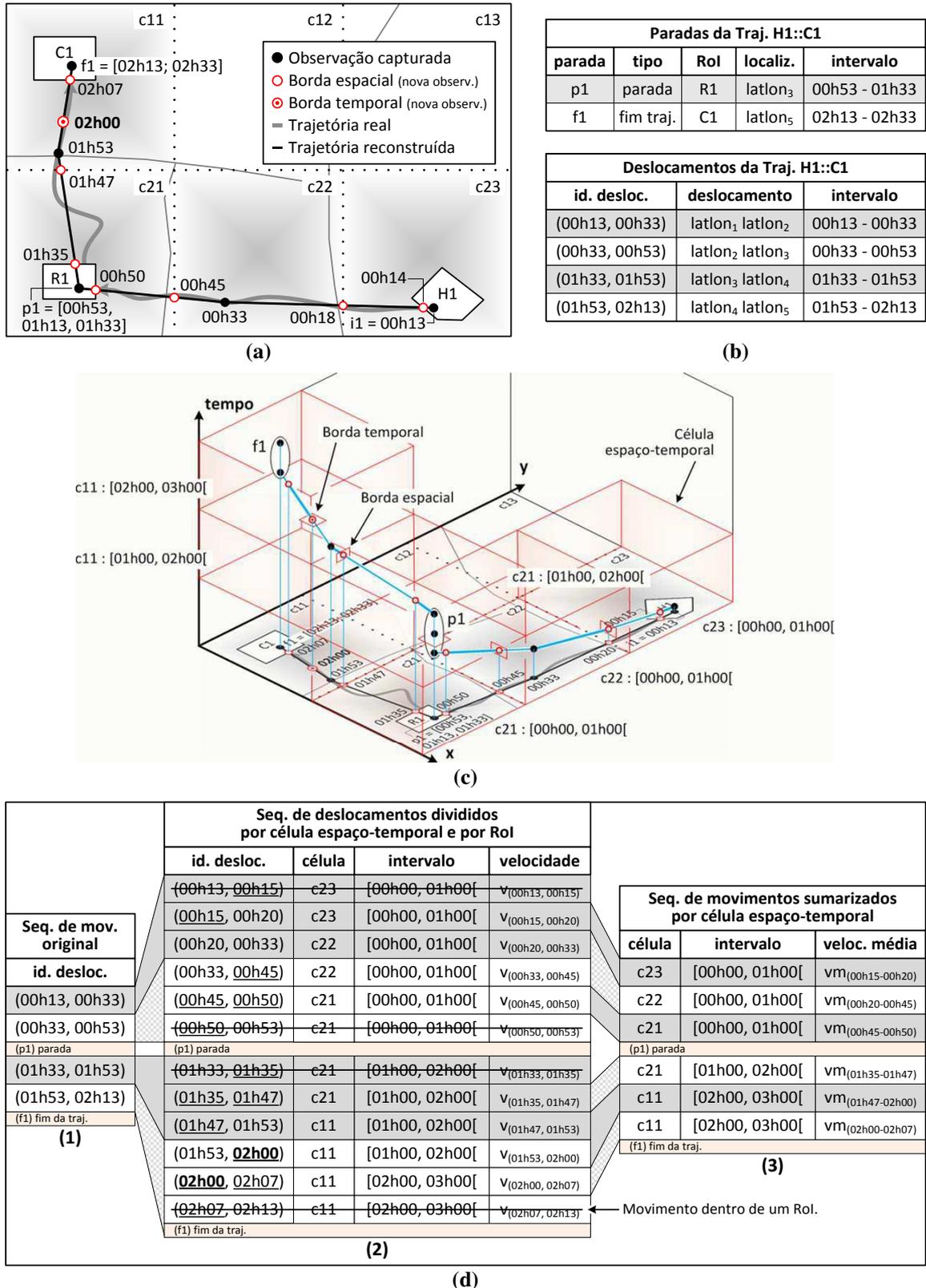


Figura 4.4: Carga de dados para a trajetória H1::C1: **(a)** representação de H1::C1 sobre o mapa; **(b)** observações capturadas de H1::C1 divididas entre paradas e movimentos; **(c)** células espaço-temporais de H1::C1 em um espaço tridimensional; **(d)** divisão completa dos movimentos de H1::C1 por célula espaço-temporal em (2), e sua respectiva sumarização em (3).

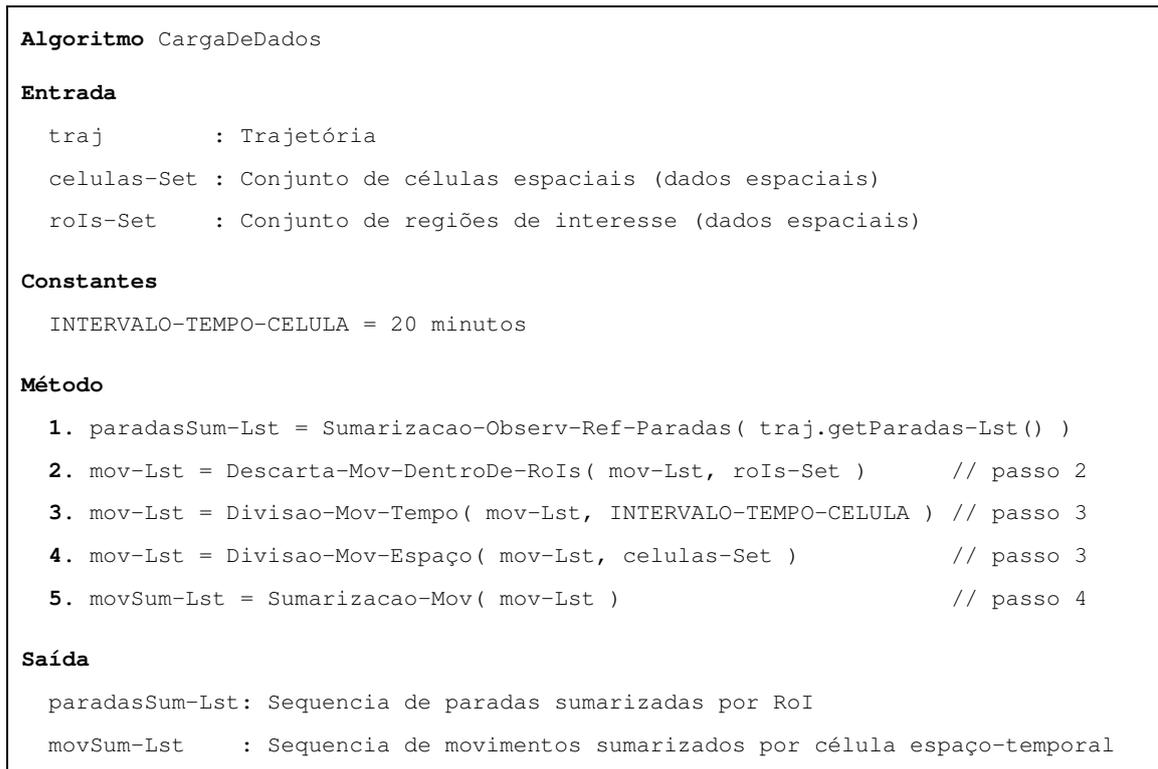


Figura 4.5: Algoritmo da Carga de Dados.

ser representados *por bairro e por RoI*, através das sequencias de movimentos [(bairro1, bairro3), (bairro3, bairro4), (bairro4, bairro2)] e [(H1, R1), (R1, C1)], respectivamente, como mostrado na Figura 4.6(a). Para obter essas representações através do uso de agregações, cada célula de H1::C1 deve estar ligada (indicado por \rightarrow) a um registro direção do movimento, como segue: (ligação por bairro) [c23:00h \rightarrow (bairro1, bairro3)], [c22:00h, e c21:00h \rightarrow (bairro3, bairro4)], [c21:01h, c11:01h, e c11:02h \rightarrow (bairro3, bairro4)]; e (ligação por RoI) [c23:00h, c22:00h, e c21:00h \rightarrow (H1, R1)], [c21:00h, c11:01h, c11:02h \rightarrow (R1, C1)]. Sendo assim, para representar os movimentos de H1::C1 por bairro, basta selecionar a trajetória e colocar como critérios de agregação: o bairro-origem e o bairro-destino.

Da forma apresentada, para cada representação desejada, é necessário manter uma dimensão direção do movimento, e uma chave estrangeira para relacionar o fato as dimensões, o que pode aumentar significativamente o volume do fato. Para reduzir o número de dimensões necessárias, uma forma adequada é unir essas dimensões (Kimball; Ross; e Merz, 2002). Dessa forma, as células de H1::C1 passam a ser ligadas aos registros: [c23:00h \rightarrow [(bairro1, bairro3), (H1, R1)]], [c22:00h, e c21:00h \rightarrow [(bairro3, bairro4), (H1, R1)]], e

[c21:01h, c11:01h, e c11:02h → [(bairro4, bairro2), (R1, C1)]], como mostrado na Figura 4.6(b). O inconveniente dessa solução, é que a união de dimensões pode gerar uma dimensão muito grande, como será discutido na Seção 5.5.

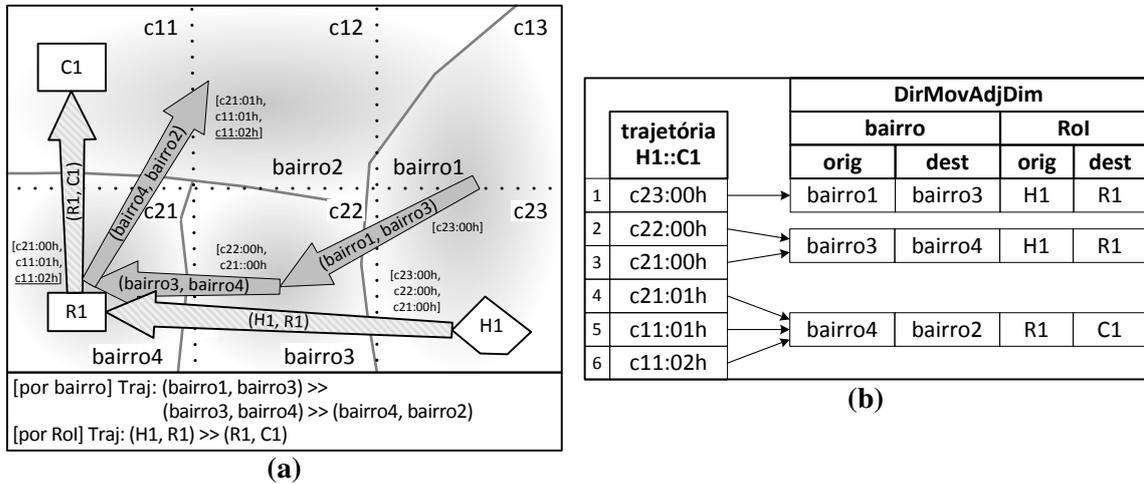


Figura 4.6: Representação por direção do movimento entre regiões adjacentes, para a trajetória H1::C1, em: **(a)** por bairro e por Rol; **(b)** células de H1::C1 e sua ligação com a dimensão DirMovAdjDim.

Capítulo 5

Validação do Modelo

Este capítulo tem como finalidade descrever a validação do modelo proposto para DW de Trajetórias (DWTr), do operador roll-up para trajetórias, e avaliar o nível de compactação para dados de trajetória proporcionado por esse modelo em relação ao clássico.

O *modelo proposto* e o *operador roll-up para trajetórias* são validados através de um conjunto de consultas típicas para DWTrs baseadas em um estudo de caso, o gerenciamento de tráfego urbano. O estudo de caso e as consultas exemplo são apresentados nas Seções 5.1 e 5.3, respectivamente. A *compactação dos dados de trajetória* é avaliada por meio de diversos testes de carga realizados sobre uma mesma base de trajetórias, mas usando diferentes configurações para cada carga de dados executada, como discutido na Seção 5.4. A *base de dados de trajetórias usada* nos experimentos e consultas foi gerada a partir de um protótipo desenvolvido, o sintetizador de trajetórias semânticas, descrito na Seção 5.2.

5.1 Estudo de Caso

Nesta seção é apresentada uma descrição sucinta de um cenário de aplicação, o *gerenciamento de tráfego urbano*, uma extensão do cenário discutido na Seção 4.1. Neste cenário exemplo, informações sobre os indivíduos (objetos móveis), meios de transporte usados (p.ex. carro, ônibus, metrô), dados de trajetória, paradas, rotas mais usadas pela população

para ir de um lugar a outro, entre outras, são consideradas no processo de tomada de decisão e descoberta do conhecimento relacionados à logística de transporte, para facilitar a circulação de pessoas na cidade, como descrito em detalhes na Seção 4.1.

Como parte dos requisitos necessários para planejamento e gerenciamento de tráfego urbano, considera-se necessário recuperar as seguintes informações:

- r1:** O comportamento dos indivíduos sobre uma dada região em diferentes intervalos de tempo, em termos do: número de indivíduos, velocidade, locais de parada, entre outros.
- r2:** A impedância de uma região, isto é, obstrução do movimento.
- r3:** Acompanhar o número de acidentes e atropelamentos por região.
- r4:** Identificar quais são os Pólos Geradores de Tráfego (PGT), e quais são as origens e destinos dos indivíduos que por ele atravessam.
- r5:** O fluxo médio de pessoas (número de pessoas) entre as regiões e suas variações em diferentes intervalos de tempo e na presença de eventos extraordinários como jogos de futebol, shows, entre outros acontecimentos que fogem a rotina.
- r6:** A proporção de veículos que deixam uma avenida em suas diferentes saídas.
- r7:** As rotas mais usadas pela população para ir de um lugar ao outro.

Todos esses requisitos foram levados em conta para elaboração das consultas. Os requisitos (r3 e r4) foram definidos pelo DENATRAN¹, e os demais pelo projeto GeoPKDD (disponível em Andrienko; Andrienko; e Wrobel (2007)).

¹Departamento Nacional de Trânsito (DENATRAN) / Fundação Getúlio Vargas (FGV) (2001). *Manual de Procedimentos para o Tratamento de Pólos Geradores de Tráfego*. Brasília/DF. Disponível em www.denatran.gov.br/publicacoes/download/PolosGeradores.pdf. Acesso em 01 nov 2010.

5.1.1 Metodologia Utilizada

Assumindo que os requisitos apresentados na seção anterior para planejamento e gerenciamento de tráfego urbano podem ser atendidos com base na análise de trajetórias. Os requisitos de (r1 à r3) podem ser atendidos através de consultas envolvendo *agregações por região* (análise orientada a tráfego), e os requisitos de (r4 à r7) por consultas envolvendo *agregações por direção do movimento*, e *por trajetória* (análise orientada a trajetória). Essas consultas serão descritas na Seção 5.3.

Para o estudo de caso, o tempo foi dividido em intervalos de 20 minutos e o espaço geográfico por uma grade regular com células de 40.000 m² (200 × 200 m) cada. Supõe-se que essa configuração é suficiente para o gerenciamento de tráfego urbano. Portanto, outro nível de granularidade espaço-temporal poderia ter sido adotado. Para evitar o problema do relacionamento parcialmente contido (Jensen et al., 2004), descrito na Seção 2.3.1, cada célula pode estar subdividida por bairro.

Para realizar as consultas e experimentos almejados, foi escolhido o SGBD PostgreSQL². A sua escolha deu-se em virtude do elevado grau de maturidade do projeto e por possuir uma extensão espacial bastante robusta, o PostGIS³, que contém todos os recursos almejados para esse trabalho. Com relação à visualização dos resultados, as agregações são realizadas no banco de dados e os resultados são carregados e transformados por um protótipo em imagens ou arquivos KML⁴, para que possam ser visualizados de várias formas, incluindo mapas estáticos ou animados e também através de representações não-cartográficas (p.ex. gráficos). Todas as tecnologias usadas neste trabalho possuem distribuição livre e código aberto.

²PostgreSQL. Disponível em: www.postgresql.org. Acesso em: 01 nov 2010.

³PostGIS. Disponível em: www.postgis.org. Acesso em: 01 nov 2010.

⁴KML (*Keyhole Markup Language*) é uma linguagem definida pela Open Geospatial Consortium (OGC) (*KML*), baseada no XML e foi projetada para expressar anotações geográficas e conteúdos que podem ser visualizados através mapas. É uma linguagem amplamente difundida e suportada pela maioria dos SIGs, como o Google Maps (<http://maps.google.com.br/>. Acesso em: 20 set 2010) e Google Earth (<http://earth.google.com/intl/pt/>. Acesso em: 20 set 2010.).

Nas seções seguintes discute-se como foi modelado o DW de Trajetórias para suportar as agregações e visões propostas no parágrafo anterior, e quais técnicas de visualização são mais adequadas para exibir e explorar os resultados das agregações.

5.1.2 Esquema Conceitual para Gerenciamento de Tráfego Urbano

O cenário de aplicação discutido na seção anterior é definido pelo esquema de dados ilustrado na Figura 5.1, o qual foi projetado para atender aos requisitos apresentados para planejamento e gerenciamento de tráfego urbano.

Esse cenário impõe dois fatos a serem analisados: o *fato movimento* (MovimentoFato) e o *fato parada* (ParadaFato), descritos abaixo:

Fato Movimento (MovimentoFato): descreve a movimentação dos objetos móveis (pessoas) na forma sumarizada sobre as células espaço-temporais, em termos de: *velocidade média* (velocMedia), *tempo gasto* (tempoGasto), *distância percorrida* (distPercorrida), e pela medida espacial *rota do movimento* (rotaMov).

Fato Parada (ParadaFato): representa na forma sumarizada os atributos referentes a uma parada em uma região de interesse (RoI) em uma janela de tempo (data inicial e final da parada), em termos de: *locais de parada* (locaisDeParada), e *tempo total de parada* (sumTempoParada). Informações detalhadas sobre o que ocorre em cada parada podem ser encontradas na medida complexa (VisitaParada) correspondente. A qual pode armazenar as paradas (ou *visitas* como este trabalho prefere chamar) realizadas dentro de um RoI, como por exemplo, os departamentos visitados por um aluno em sua parada no campus da universidade, como será visto na Seção 5.3.1 (Consulta 1.3).

Os fatos são caracterizados pelas dimensões:

Objeto Móvel (ObjMovDim): representa as dimensões *demográfica* e *tecnográfica*, armazena dados referentes aos indivíduos [atributos: nome, sexo, anoNascimento, estadoCivil]; os meios de transporte utilizados por estes [atributo veículo]; e sobre o dispositivo de localização usado, nesse caso, precisão do GPS usado [precisaoGPS].

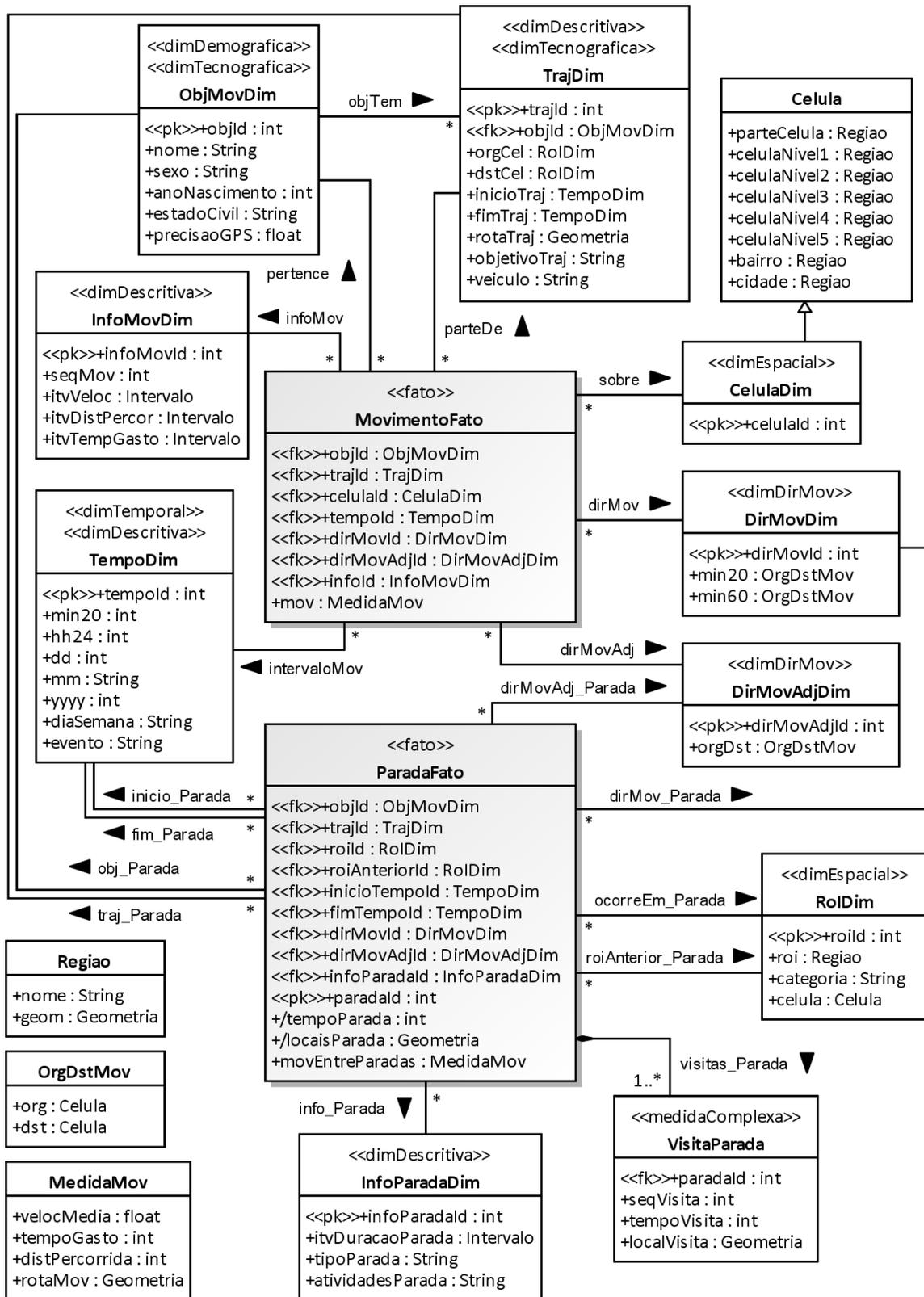


Figura 5.1: Esquema Conceitual para Gerenciamento de Tráfego Urbano.

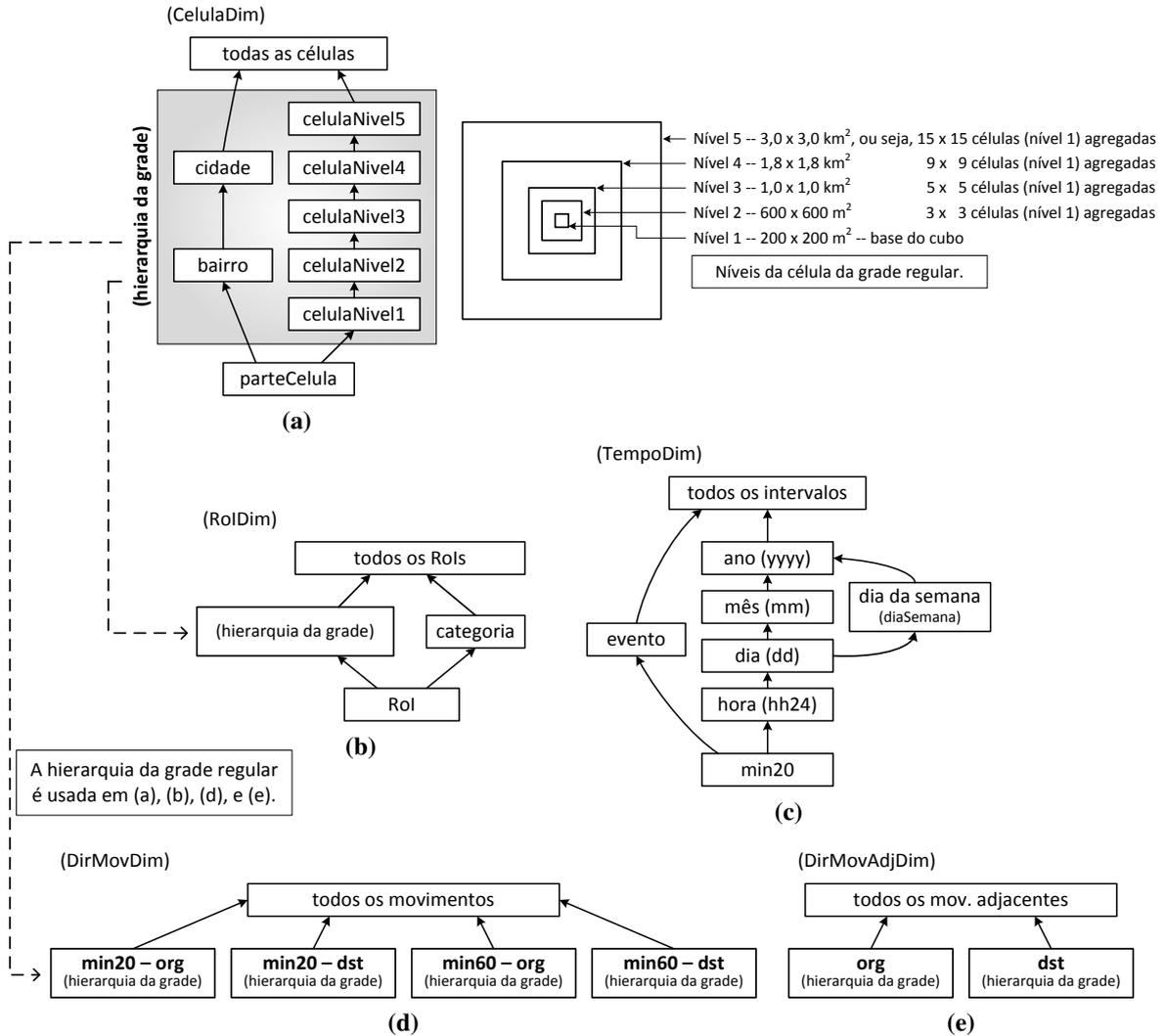


Figura 5.2: Hierarquias para as dimensões mostradas na Figura 5.1.

Trajétoria (TrajDim): armazena informações semânticas sobre as trajetórias. Atributos: RoI de origem e destino da trajetória [orgRoI e dstRoI, respectivamente]; data de início e fim da trajetória [inicioTraj e fimTraj]; representação espacial da rota da trajetória [rotaTraj, dado espacial do tipo linha]; e objetivo da trajetória [objetivoTraj] (p.ex. sair de casa para o trabalho).

Célula (CelulaDim): dimensão espacial. É responsável por guardar os dados referentes às células da grade regular. Sua hierarquia é ilustrada pela Figura 5.2(a), onde cada nível representa a agregação de um dado conjunto de células vizinhas. Por exemplo, cada agregação no nível 2 (3×3) agrega 9 células vizinhas da base do cubo (nível 1),

gerando uma grade regular cujas células possuem $600 \times 600 \text{ m}^2$ (isto é, $3 * (200 \times 200 \text{ m}^2)$).

Tempo (TempoDim): dimensão temporal e descritiva. Mantém informações referentes aos intervalos de tempo, e os eventos que ocorreram nessa janela de tempo. Seu maior nível de detalhe corresponde ao intervalo de 20 minutos. Exemplo de eventos: um jogo de futebol, um show, um engarrafamento, momento de tráfego intenso, entre outros. Hierarquia apresentada na Figura 5.2(c).

Região de Interesse (RoIDim): dimensão espacial. Armazena informações referentes às regiões de interesse (RoIs). Para esse estudo de caso, todas os RoIs foram pré-estabelecidos, mas nada impede que em aplicações práticas novas regiões sejam definidas durante o processo de carga dos dados. Exemplo de RoIs: o campus de uma universidade, um condomínio, o prédio de uma empresa, entre outros. Hierarquia exibida na Figura 5.2(b).

Para o cenário de aplicação, adotou-se RoIs que representam as entidades em sua totalidade, não partes dela. Exemplificando, as paradas realizadas em um departamento da universidade são interpretadas como paradas que ocorreram no campus da universidade, e nessa parada no campus ocorreu uma visita ao departamento em questão, como discutido no início desta seção.

Através dos relacionamentos `roiAnterior_Parada` e `ocorreEm_Parada`, é mantido a transição entre paradas, isto é, a origem e o destino do movimento que originou a parada corrente. Com isso, pode-se visualizar o deslocamento dos objetos móveis entre paradas, como será visto na Seção 5.3.3 (Consulta 3.3).

Informações sobre movimentos e paradas (InfoMovDim e InfoParadaDim): mantém informações semânticas referentes ao fato associado. São expressadas pelos atributos:

Informações sobre o Movimento (InfoMovDim): [`seqMov`] número de sequencia do movimento na trajetória (similar ao índice em um vetor); [`faixaVeloc`] a faixa de velocidade que o movimento pertence (por exemplo, 0..30, 30..60, 60..90, entre outros).

Informações sobre as Paradas (InfoParadaDim): [seqParada] número de sequência da parada na trajetória (útil para regras de associação); [faixaDuracaoParada] a faixa de tempo de duração da parada (por exemplo, 15 minutos, 30 minutos, 1 hora); [atividades] atividades realizadas durante a parada (por exemplo, pausa para refeição, trabalho, descanso); [tipoParada] se a parada ocorreu durante ou no fim da trajetória.

Direção do Movimento: Existem dois tipos: *Direção do Movimento Espaço-Temporal* (DirMovDim) e *Direção do Movimento entre Regiões Adjacentes* (DirAdjMovDim). Hierarquias apresentadas nas Figuras 5.2 (d) e (e).

DirMovDim: permite realizar roll-up por direção do movimento entre duas células em um dado intervalo de tempo pré-definido. O esquema definido permite obter a direção que os objetos móveis tomam a cada 20 minutos ou 1 hora através dos atributos min20 e min50, respectivamente. Em termos de: número de veículos, velocidade, tempo gasto para realizar o percurso, número de paradas realizadas no percurso, entre outros atributos.

DirMovAdjDim: similar à direção do movimento espaço-temporal (DirMovDim), mas se aplica a movimentação entre regiões adjacentes. Maiores informações estão disponíveis na Seção 5.3.2.

5.2 Base de Trajetórias

Para realizar as consultas e os experimentos almejados, o DW de Trajetórias foi povoado com uma base de dados de trajetórias sintéticas gerada a partir de um *sintetizador de trajetórias semânticas* desenvolvido. Para dar um caráter mais realístico à base de dados, o protótipo desenvolvido permite criar problemas de tráfego a partir de um conjunto de configurações pré-estabelecidas definidas pelo usuário. Através dessas configurações é possível: o número de objetos móveis; os locais de parada; a velocidade máxima que os objetos podem atingir; o tempo de parada; a origem das trajetórias; além de configurações específicas para cada região, válidas para um dado intervalo de tempo, por exemplo: as horas com tráfego mais

intenso; quando e onde devem ocorrer congestionamentos; as regiões com maior ou menor velocidade; e mais alguns parâmetros que ajudam a diferenciar o comportamento dos objetos móveis entre as regiões e no tempo. Além disso, cada objeto móvel está associado a um perfil, onde é possível determinar seus locais de paradas; os locais mais visitados; o tempo de visita; a precisão da trajetória capturada; entre outros atributos. Para dar um caráter ainda mais realista a base gerada, a rota de cada trajetória - isto é, a sequência de coordenadas espaciais que vão da origem ao destino da trajetória - é obtida do Google Maps. De posse desses dados, o protótipo simula o deslocamento do objeto móvel.

Em geral, todas as configurações estão baseadas nos atributos velocidade e número de objetos móveis, por exemplo, para criar regiões com maior tráfego, o gerador desenvolvido altera a rota padrão de algumas trajetórias para que elas passem obrigatoriamente por um dado trecho, forçando assim o aparecimento de um tráfego mais intenso na região; de forma similar são criados os congestionamentos, para isso, é definida uma região cuja velocidade é muito baixa, mas de tráfego muito intenso. Por questões de simplificação, no estado atual de desenvolvimento, o protótipo não é capaz de detectar colisão de objetos móveis.

Para gerar a base de dados sintética, simulou-se o comportamento de um conjunto de 2.000 objetos móveis, que se movimentaram na cidade de Aracaju/Sergipe, durante os meses de janeiro a junho de 2009. São aproximadamente 6.400 RoIs, e 1,8 milhão de trajetórias, uma média de 1.000 trajetórias por objeto. No mundo real, isso equivale a cerca de 1 bilhão de observações, sendo 82 milhões referentes a *movimentos* e 931 milhões a *paradas*. Considerando observações capturadas a cada 20 segundos, e objetos móveis realizando em média 4 paradas de 2 horas por dia. Para tornar a leitura dos dados mais rápida e reduzir o espaço ocupado em disco, as observações referentes a paradas já são armazenadas na forma sumariada. Dessa forma, consegue-se armazenar todas as *paradas* em 4 milhões de registros. A base de dados possui aproximadamente 15 GB de arquivos de texto no formato JSON⁵,

⁵JSON (*JavaScript Object Notation* – Notação de Objetos JavaScript): é uma linguagem de marcação similar ao XML, só que mais leve. Ela é descrita pela RFC 4627, disponível em: <http://tools.ietf.org/html/rfc4627>, link permanente. O site oficial está disponível em: <http://www.json.org/>. Acesso em: 01 nov 2010.

como ilustrado na Figura 5.3.

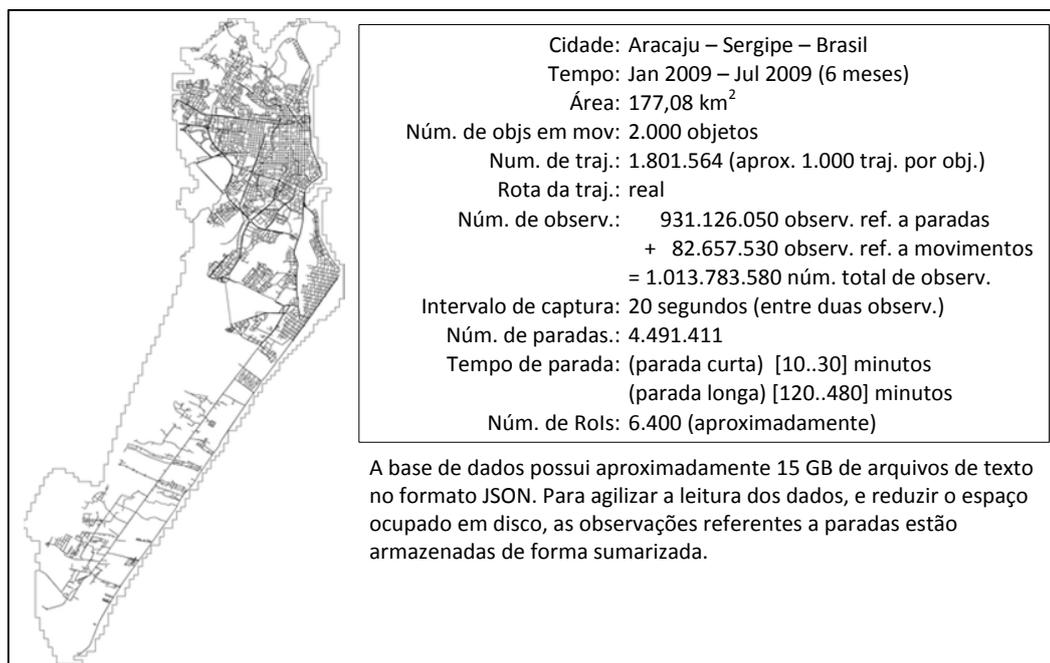


Figura 5.3: Configuração da Base de Trajetórias gerada.

Com relação ao comportamento dos objetos, ao longo de seu ciclo de vida um objeto móvel sai de sua origem (seu ponto de apoio), realiza algumas trajetórias em sequencia, e volta para sua origem. Ao longo de cada trajetória ele realiza uma série de paradas curtas em Regiões de Interesse (RoIs), e uma parada longa no final da trajetória, no RoI destino. Durante seu deslocamento o objeto varia sua velocidade de forma aleatória. Para evitar brechas no mapa – isto é, regiões sem a presença de trajetórias. Durante a geração da base buscou-se percorrer todo o território da cidade.

O protótipo desenvolvido (ver Figura 5.4) possui interface Web, e foi escrito em *Java Applet*⁶ e *JSP (JavaServer Pages)* rodando sobre o servidor Web *Apache Tomcat*⁷.

As rotas obtidas do Google Maps têm a característica de sempre se encaixarem perfeitamente com as ruas da cidade, é como se os objetos móveis andassem sobre trilhos, o que não

⁶Java Applets. Disponível em: <http://java.sun.com/applets/>. Acesso em: 01 nov 2010.

⁷Apache Tomcat. Disponível em: <http://tomcat.apache.org/>. Acesso em: 01 nov 2010.

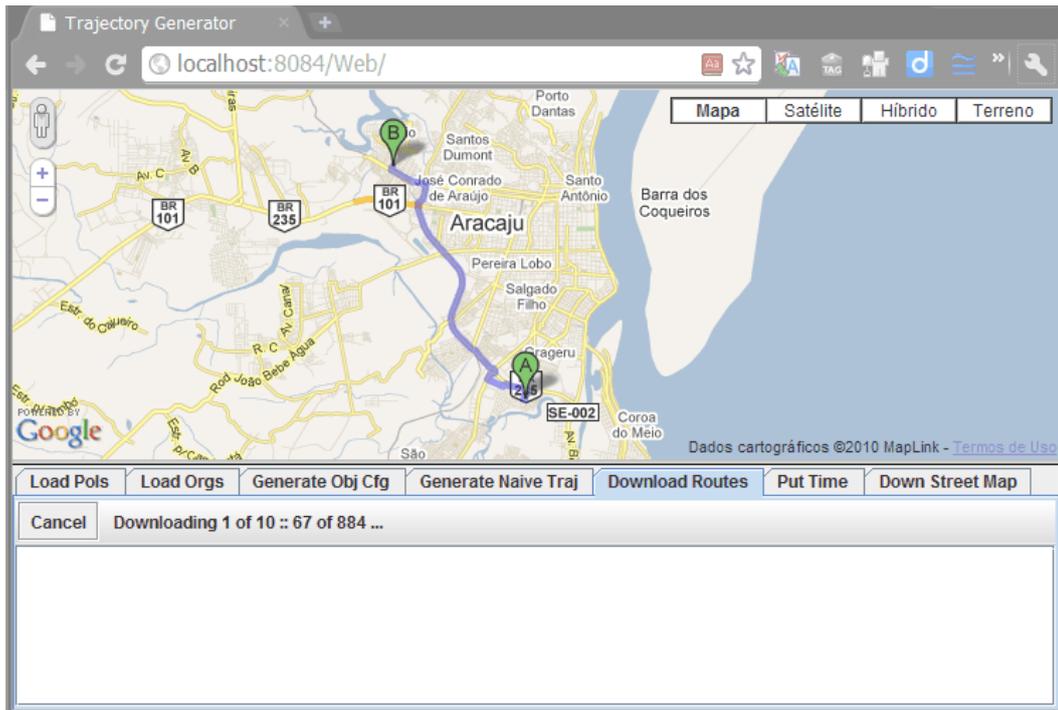


Figura 5.4: Sintetizador de Trajetórias desenvolvido.

acontece com uma trajetória real capturada por um GPS, devido aos problemas de imprecisão gerados por este. Para se aproximar ainda mais de um cenário real, o protótipo poderia injetar imprecisão na rota obtida, imprecisão essa que durante o processo ETL buscava-se reduzir. Logo, optou-se por não alterar a rota baixada.

Antes de construir o sintetizador de trajetórias buscou-se usar uma base de dados de trajetórias reais existente, mas alguns problemas inviabilizaram seu uso para esse trabalho: (i) devido a questões de privacidade poucas dessas bases estão disponíveis para livre uso; (ii) geralmente são pequenas, se resumem a apenas alguns milhares de registros, o que é muito pouco para avaliar a compactação dos dados de trajetória; e (iii) todas as bases encontradas são brutas, não possuem informações semânticas (p.ex. início/fim da trajetória, paradas).

5.3 Consultas Exemplo

O poder de expressividade do modelo proposto é mostrado através das consultas definidas a seguir, e buscam atender aos requisitos apresentados na Seção 5.1. Para facilitar sua compreensão as consultas estão divididas em três categorias:

Consultas por Região: (Seção 5.3.1) estão relacionadas a análise orientada a tráfego. Logo, têm como objetivo analisar o comportamento dos objetos móveis sobre uma dada região espacial em diferentes intervalos de tempo. Por exemplo: *a velocidade média dos veículos por célula da grade regular?*. Atendem aos requisitos (r1 à r3).

Consultas por Direção do Movimento: (Seção 5.3.2) envolvem agregação por direção dos movimentos. Permitem visualizar a direção que os objetos móveis tomam em diferentes intervalos de tempo. Por exemplo: *o fluxo de veículos (número de veículos) entre os bairros da cidade a cada 20 minutos?*. Atendem aos requisitos (r4 à r7).

Consultas por Trajetória: (Seção 5.3.3) envolvem roll-up para trajetórias (em outras palavras, envolvem múltiplas representações para uma mesma trajetória). Com base nessas consultar é possível identificar quais são as rotas mais usadas pela população para ir de um lugar a outro (requisito r7); os trechos de menor/maior velocidade da trajetória, devido à obstrução de movimento nas vias (p.ex. pista esburacada) (requisito r2). Exemplo de consulta: *a velocidade média por bairro das trajetórias com origem no restaurante R1 e banco B1?* (mostra a velocidade média de cada movimento da trajetória entre bairros). Atendem aos requisitos (r4 à r7).

5.3.1 Consultas por Região (requisitos r1 à r4)

A agregação por região é adequada para analisar o comportamento dos objetos móveis no espaço em diferentes intervalos de tempo. Considerando o cenário de aplicação, pode-se analisar o comportamento dos objetos em termos de: (movimento) velocidade, número de objetos móveis, distância percorrida, e tempo gasto; (parada) tempo de parada, número de objetos parados, e atributos derivados a partir do uso de funções de agregação (p.ex. mínimo (min), máximo (max), e média (avg)).

Consulta 1.1: Agregação por Região

Para exemplificar a agregação das medidas por região, considere a consulta: (i) *O número*

de veículos e a velocidade média dos carros circulando na cidade de Aracaju no dia 10-jan-2009, por hora e por região da cidade? Considere uma região da cidade como uma célula da grade regular no nível 2. Se levarmos em conta apenas os veículos em movimento, a consulta pode ser respondida apenas com o *fato movimento* (MovimentoFato) de acordo com os passos:

(Sobre o fato Movimento): restrição da data para “10-jan-2009”; **agregação** dos atributos *identificador do objeto* (objId) e *velocidade* (velocMedia) através das funções *contagem distinta* e *média*, respectivamente, tendo como **critério de agregação** a *célula da grade regular* no nível 2 (celulaNivel2) e a hora (hh24). Na Figura 5.5(a) é mostrada essa consulta em SQL.

Levando em consideração veículos parados e em movimento, é necessário realizar uma consulta sobre os fatos (parada e movimento) e unir seus resultados. O número de veículos parados em uma região pode ser computado da seguinte forma:

(Sobre o fato Parada): restrição temporal: para a data “10-jan-2009”; **agregação** do atributo *identificador do objeto* (objId) através da função *contagem distinta*, usando como **critério de agregação** a *célula da grade regular* no nível 2 (celulaNivel2).

De posse desses resultados, é feita sua união: **agregação** através das funções de agregação *contagem distinta* e *média* para os atributos *identificador do objeto* (objId) e *velocidade* (velocMedia), usando como **critério de agregação** a célula da grade regular. Durante a união dos resultados, considera-se nulo a velocidade para objetos parados, o que excluir esse resultado do cálculo da velocidade média.

Para facilitar a visualização dos resultados da consulta, seus dados foram divididos em duas imagens, na Figura 5.5(b)(i) é mostrado o número de objetos móveis por célula da grade regular no nível 3 (1×1 km), os quadrados vermelhos indicam as regiões com maior concentração do tráfego (isto é, número de veículos), onde o número de veículos é superior à média por célula da cidade, e os quadrados azuis indicam que a concentração do tráfego é inferior a média. O tamanho dos quadrados é proporcional ao número de veículos, observe que em algumas regiões a concentração de veículos é muito baixa. Em alguns casos, isso

se deve ao fato de serem regiões pouco povoadas, principalmente na região sul. Na Figura 5.5(b)(ii) é mostrada a velocidade média por célula da grade no nível 2 (600×600 m), as regiões em vermelho indicam os trechos problemáticos onde o tráfego tem velocidade inferior a 30 km/h, observe que alguns desses trechos apresentam a maior concentração do tráfego de veículos.

Consulta 1.2: Restrição Espacial & Medida Espacial

Existem ocasiões onde é necessário realizar consultas baseadas em uma janela espacial, e em outras onde é preciso analisar o comportamento detalhado dos objetos móveis por região. Para exemplificar uma consulta que envolve esses requisitos, considere: *A velocidade média por região da cidade em detalhes, dentro de uma janela espacial, no dia 20-jan-2009 das 12h00 às 13h00?* Considere uma região como uma célula da grade regular no seu nível mais refinado (nível 1). Para responder essa consulta os seguintes passos devem ser seguidos:

Restrição espacial: sobre a dimensão célula para obter as medidas do movimento que pertencem à janela espacial definida, e **temporal:** para a data 12-jan-2009 entre 12h00 e 13h00; **agregação** da medida espacial *rota do movimento* (rotaMov) através do operador para *união de geometrias*; adotando como **critério de agregação** por célula da grade regular no nível 1, e por faixa de velocidade, cada faixa de comprimento de 15 km/h cada (por exemplo: [0,15[, [15,30[, [30,45[, ...).

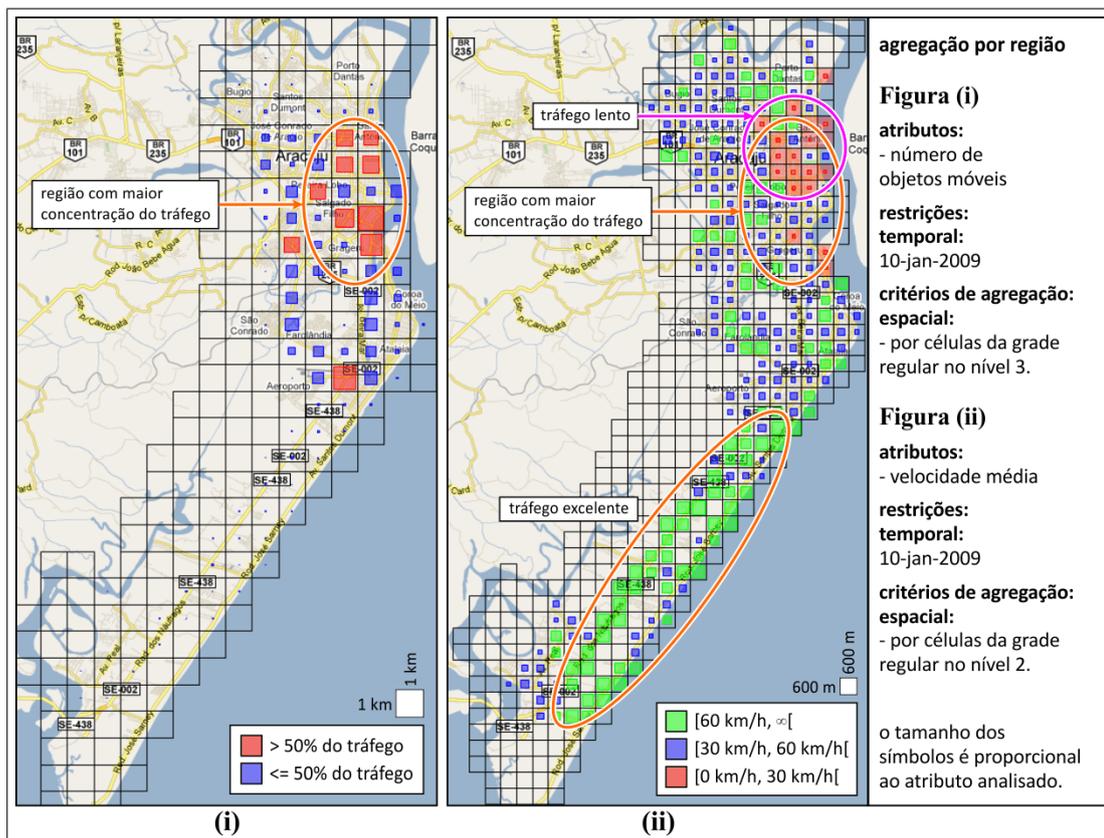
O resultado dessa consulta pode ser visualizado na Figura 5.6(b)(iii), as figuras (b)(i) e (b)(ii) representam uma pequena variação dessa consulta com o objetivo de mostrar algumas das diferentes formas de visualização para o resultado das consultas. Cada forma de visualização possui suas peculiaridades e ajudam de diferentes formas no processo de descoberta do conhecimento. Na Figura 5.6(b)(i) temos a consulta em sua forma convencional (não detalhada), como na Consulta 1.1, a medida usada durante as agregações é a velocidade média; em (b)(i) temos a mesma consulta, mas a restrição temporal foi definida para uma semana (de 18-jan-2009 até 24-jan-2009), portanto, é mostrado em (b)(i) a variação da velocidade por região ao longo da semana e das horas do dia; e em (b)(iii) temos a consulta em sua forma detalhada, como explicitada no parágrafo anterior.

```

1 SELECT
2   cd.celulaNivel3_latlon AS latlon,
3   td.hh24 AS intervaloTempo, -- 1 hora
4   COUNT(DISTINCT mf.objId) AS numObjetosMoveis,
5   AVG(mf.velocMedia) AS velocidadeMedia
6 FROM
7   MovimentoFato AS mf INNER JOIN
8   CelulaDim AS cd ON (mf.celulaId = cd.celulaId) INNER JOIN
9   TempoDim AS td ON (mf.tempoId = td.tempoId)
10 WHERE
11   td.tempo BETWEEN
12     to_timestamp('10-jan-2009 00:00', 'DD-mon-YYYY HH24:MI') AND
13     to_timestamp('10-jan-2009 23:40', 'DD-mon-YYYY HH24:MI')
14 GROUP BY
15   latlon, intervalotempo

```

(a)



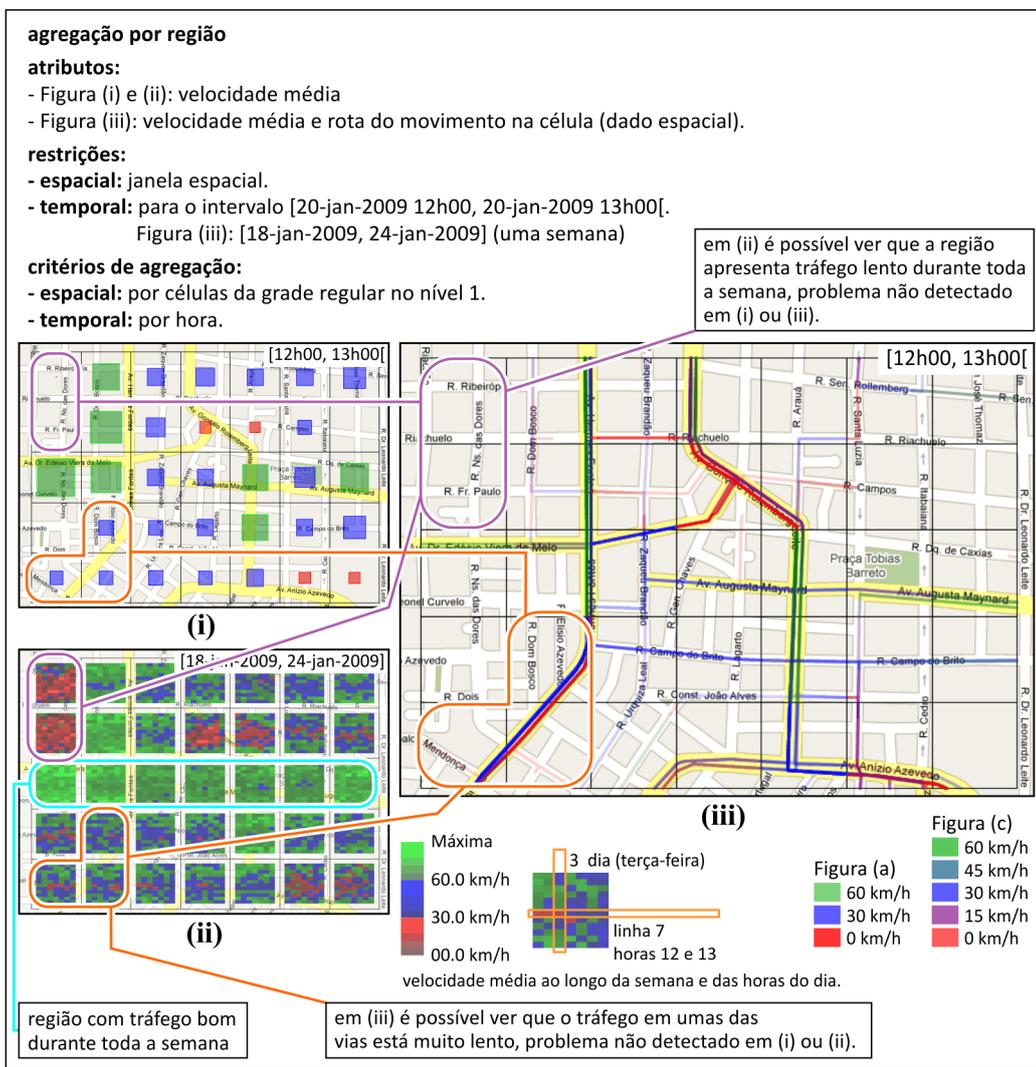
(b)

Figura 5.5: (Consulta 1.1) Agregação por região, em termos de:
(i) Número de Veículos; (ii) Velocidade Média.

```

1 SELECT td.hh24 AS hora24, ifd.faixaVeloc AS faixaVeloc,
2     ST_Multi(ST_Union(mf.rotaMov)) AS rotaMov
3 FROM MovimentoFato AS mf INNER JOIN
4     CelulaDim AS cd ON (mf.celulaId = cd.celulaId) INNER JOIN
5     TempoDim AS td ON (mf.tempoId = td.tempoId) INNER JOIN
6     InfoMovDim AS ifd ON (mf.infoMovId = ifd.infoMovId)
7 WHERE ( td.tempo BETWEEN
8     to_timestamp('20-jan-2009 00:00', 'DD-mon-YYYY HH24:MI') AND
9     to_timestamp('20-jan-2009 23:40', 'DD-mon-YYYY HH24:MI') )
10 AND ( ST_Within(cd.celulaNivel1_poly, ST_GeomFromText(
11     'SRID=4326;POLYGON((-37.061743 -10.91917, -37.048289 -10.91917,
12     -37.048289 -10.92871, -37.061743 -10.92871,
13     -37.061743 -10.91917))', 4326)) )
14 GROUP BY hora24, faixaVeloc
    
```

(a)



(b)

Figura 5.6: (Consulta 1.2) Velocidade Média por região da cidade em detalhes, dentro de uma janela espacial, sobre diferentes formas de visualização: (i) convencional; (ii) mosaico; (iii) detalhado.

Observe na Figura 5.6 que apenas algumas formas de visualização conseguem destacar algumas características que passam despercebidas em outras formas. Por exemplo, em **(b)(i)** é mostrada uma visão geral do tráfego em uma dada região em um intervalo de tempo, destaca de forma clara regiões com problemas, mas não apresenta os detalhes do comportamento dos objetos móveis dentro das regiões, o que fica mais claro em **(b)(iii)**, a forma de visualização em mosaico facilita a detecção das regiões mais problemáticas, cujo tráfego se encontra ruim durante toda a semana.

Consulta 1.3: Medida Complexa

Existem situações onde é necessário conhecer os detalhes sobre o que ocorreu em uma parada (ou em um conjunto de paradas). Por exemplo, considere a visita de um indivíduo ao campus de uma universidade, esse trabalho considera o campus como sendo uma região de interesse (RoI), portanto, todos os dados de movimentação sobre essa região são considerados como paradas, logo, são sumarizados e armazenados no fato parada (ParadaFato) através das medidas locaisDeParada e tempoParada, mas as visitas feitas aos departamentos da universidade em cada parada são mantidas na medida complexa correspondente. Essa foi uma forma encontrada de manter o detalhamento do que ocorre nas paradas, sem sobrecarregar o fato parada com informações inúteis para a maioria das consultas, já que essas consultas não estão interessadas nos detalhes do que ocorreu em cada parada, mas no contexto geral. Para exemplificar o uso da medida complexa, considere a consulta: *Quais são os locais mais visitados e o tempo de visita das paradas dentro do campus da universidade?*

5.3.2 Consultas por Trajetória (requisitos r1 à r4)

Uma consulta típica no contexto da análise de objetos móveis é visualizar seu deslocamento entre as regiões em diferentes intervalos de tempo.

Consulta 2.1: Por Direção do Movimento entre Regiões Adjacentes

Para exemplificar consultas que envolvem agregação *por direção do movimento entre regiões*

adjacentes, considere a consulta: *Mostre o deslocamento dos veículos na cidade entre os bairros adjacentes*. Para simplificar a visualização dos resultados, considere apenas o bairro Getúlio Vargas e seus adjacentes. Consultas como essa, envolvem basicamente o uso do fato movimento juntamente com a dimensão direção do movimento, nesse caso, movimento entre regiões adjacentes, e são muito úteis para analisar o deslocamento dos objetos móveis entre as regiões. Essa consulta pode ser respondida com os seguintes passos:

Restrição: movimentos que pertencem às células do bairro “Getúlio Vargas” e bairros adjacentes, e sobre a dimensão tempo (TempoDim) para o intervalo “10-jan-2009 10h00-10h20”; **agregação:** usando contagem distinta para o identificador do objeto móvel (objId); **critérios de agregação** *por origem e destino do movimento entre bairros adjacentes* (atributos *org.bairro.geom* e *dst.bairro.geom* da dimensão *DirMovAdjDim*).

Na Figura 5.7 é exibido o deslocamento dos veículos entre bairros adjacentes, o tamanho das setas é proporcional ao número de veículos, e a seta em vermelho indica o deslocamento com maior número de veículos. A tabela à direita representa uma forma de visualização não-cartográfica para o resultado, o tamanho dos quadrados é proporcional ao número de veículos se deslocando na respectiva direção. Na coluna origem, é exibido nas barras horizontais o número de veículos que *deixam* a respectiva região e, no destino, as barras verticais representam o número de veículos que *entram* em um determinado bairro.

Para exemplificar consultas similares a anterior, mas em um nível de granularidade mais refinado, considere a consulta: *Exiba o número de veículos por direção do movimento e por hora ao longo dia 10-jan-2009, para as células dentro de uma janela espacial que detenham a maior concentração do tráfego na região no intervalo de tempo analisado*. Considere a *direção do movimento* entre células adjacentes em seu grau mais refinado (nível 1), e as células que detenham a *maior concentração do tráfego na região* para cada intervalo de tempo analisado, isto é, células onde o número de veículos está acima da média para a janela espacial analisada no intervalo de tempo de 1 hora. O resultado dessa consulta é ilustrado na Figura 5.8, em **(a)** é mostrado apenas a direção principal do movimento na célula, e em **(b)** é exibido todas as direções do movimento dentro da célula, o tamanho das barras é

proporcional ao número de veículos por direção. Observe nessa figura que em seu intervalo de tempo a maioria dos veículos se desloca da região norte para o sul da cidade, e o tráfego se concentra nas avenidas, destacadas em amarelo.

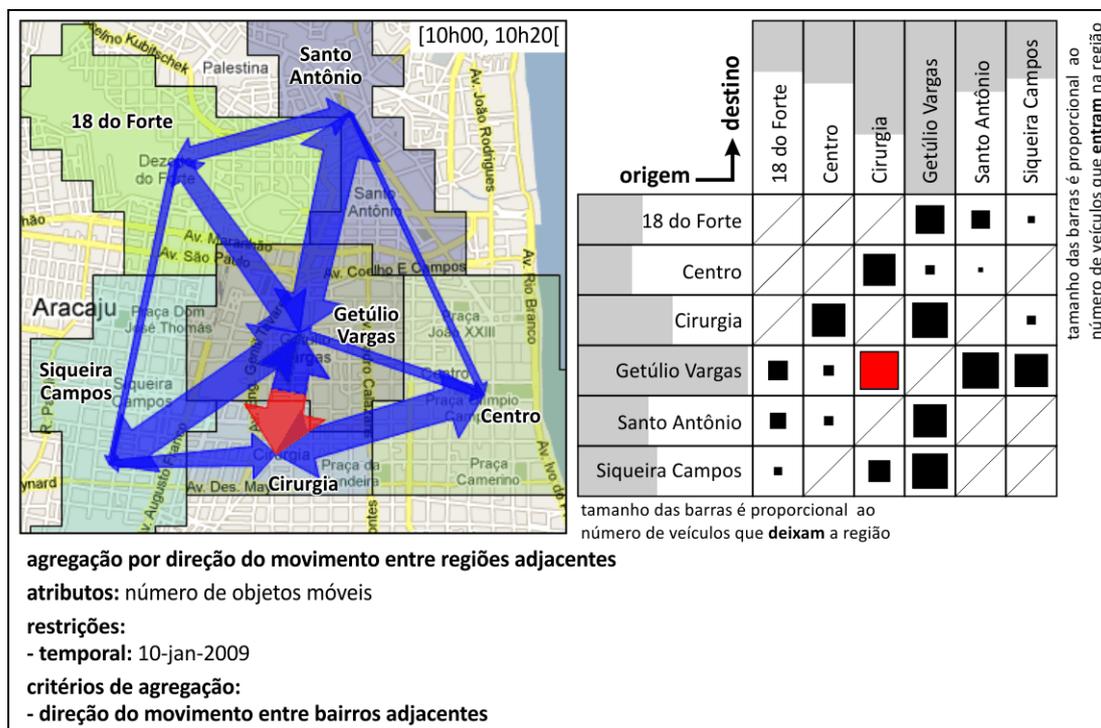
Consulta 2.2: Por Direção do Movimento Espaço-Temporal

```

1  SELECT
2      dmad.org_bairro_nome      AS org,
3      dmad.dst_bairro_nome     AS dst,
4      td.tempo                 AS intervaloTempo, -- 20 minutos
5      COUNT(DISTINCT mf.objId) AS numObjetosMoveis
6  FROM
7      MovimentoFato AS mf  INNER JOIN
8      DirMovAdjDim  AS dmad ON (mf.dirMovAdjId = dmad.dirMovAdjId) INNER JOIN
9      TempoDim      AS td  ON (mf.tempoId      = td.tempoId)
10 WHERE
11     td.tempo BETWEEN
12         to_timestamp('10-jan-2009 00:00', 'DD-mon-YYYY HH24:MI') AND
13         to_timestamp('10-jan-2009 23:40', 'DD-mon-YYYY HH24:MI')
14 GROUP BY
15     org, dst, intervaloTempo

```

(a)



(b)

Figura 5.7: (Consulta 2.1) Deslocamento dos veículos entre *bairros adjacentes* por intervalo de tempo. (a) Consulta em SQL; (b) Número de veículos se deslocando entre *bairro adjacentes* no intervalo de tempo das 10h00..10h20.

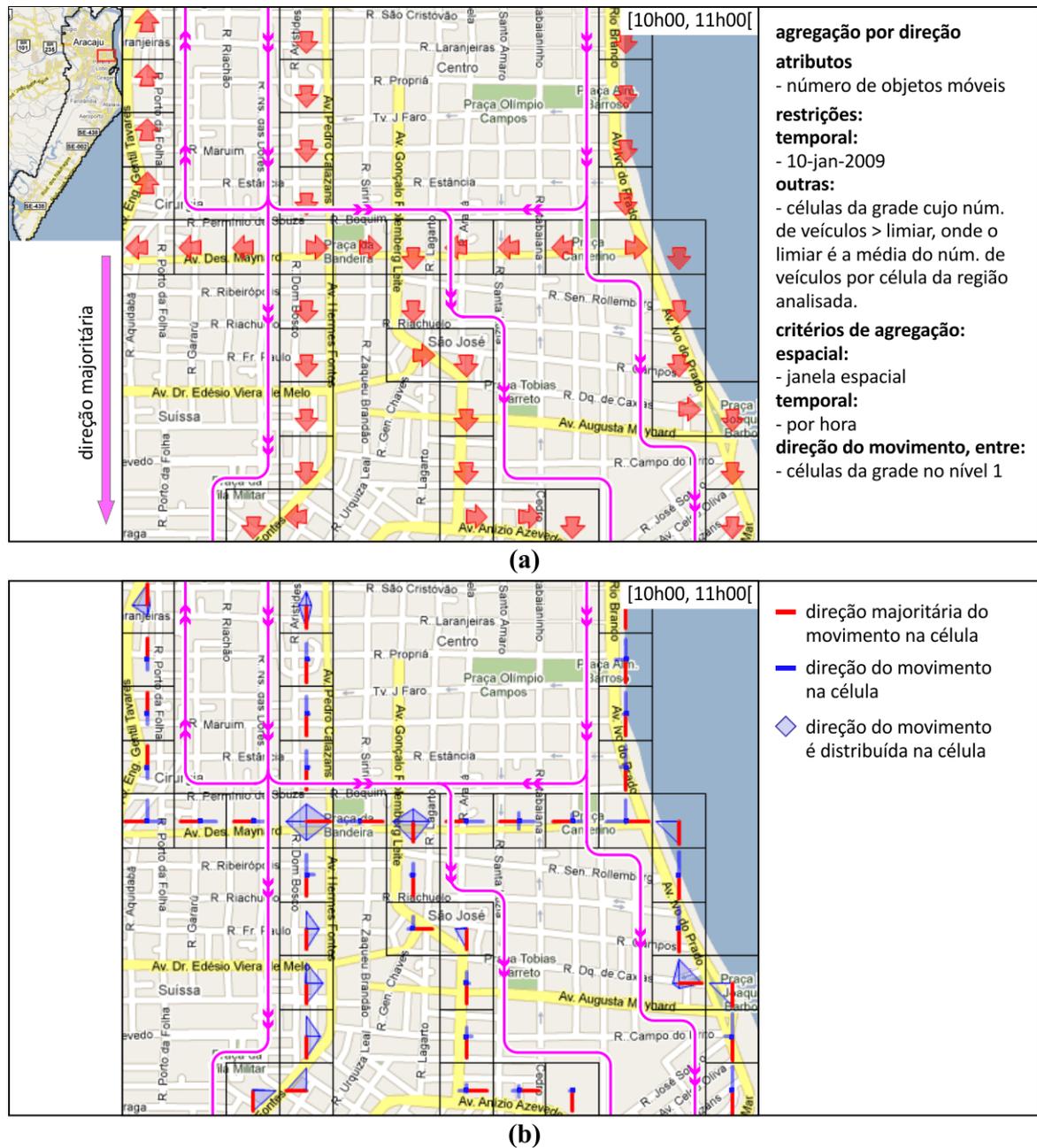


Figura 5.8: (Consulta 2.1) Direção dos movimentos por célula da grade regular. Em (a) é mostrado apenas a direção do movimento majoritária na célula; e em (b) a direção do movimento em todas as direções.

Além do deslocamento entre regiões adjacentes, o esquema definido permite visualizar o deslocamento dos veículos entre regiões a cada 20 ou 60 minutos, uma forma de análise muito similar a anterior, mas tem como objetivo visualizar a direção em que os veículos tomam em cada intervalo de tempo. Para exemplificar esse tipo de consulta, considere: *Mostre o deslocamento dos veículos na cidade a cada 20 minutos entre bairros*. Como essa consulta é bastante similar a Consulta 2.1, seus passos para resolução foram suprimidos. A única diferença entre as duas consultas está no critério de agregação. Consultas envolvendo deslocamento espaço-temporal dos objetos móveis são respondidas através da *dimensão direção do movimento espaço-temporal* (DirMovEspTempoDim). É exibido na Figura 5.9 o resultado dessa consulta.

5.3.3 Consultas por Trajetória (requisitos r4 à r7)

Muito similar a *consultas por direção do movimento* (Seção 5.3.2), mas sua finalidade é analisar o deslocamento completo do objeto móvel, isto é, analisar a trajetória adotada pelo objeto desde de sua origem até seu destino. Nos parágrafos a seguir serão apresentados algumas consultas nesse sentido.

Consulta 3.1: Por Direção dos Movimentos Adjacentes da Trajetória

Para exemplificar múltiplas representações para uma mesma trajetória, envolvendo seus movimentos entre regiões adjacentes, considere a consulta: *Exiba a trajetória #100 sobre diferentes níveis de granularidade por direção dos movimentos adjacentes*.

Sobre o fato *movimento*, e a dimensão *direção dos movimentos adjacentes*: **restrição:** para a trajetória #100; **agregação:** para a medida velocidade (velocMedia), usando a função de agregação média; **critérios de agregação:** *por origem e destino do movimento entre células adjacentes* no nível 2 (atributos: org_celulaNivel2_latlon e dst_celulaNivel2_latlon). Na Figura 5.10(a) é mostrada essa consulta escrita em SQL.

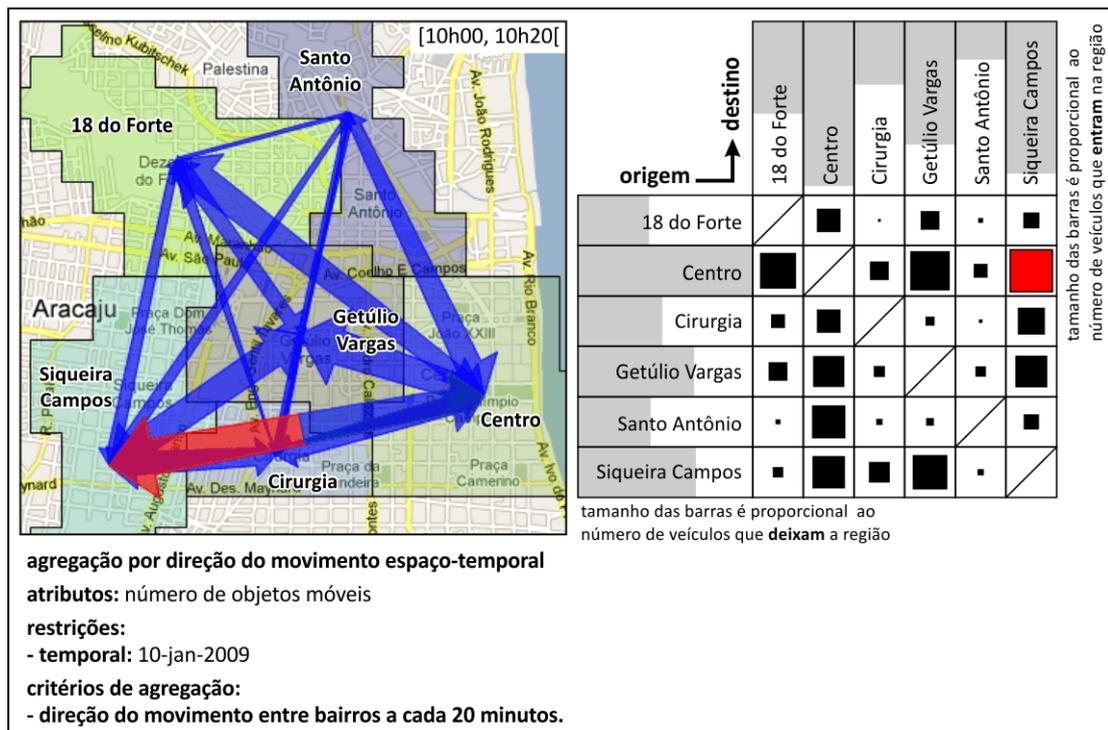
Na Figura 5.10(b) é exibido algumas das múltiplas representações para a trajetória #100

```

1 SELECT
2     dmd.min20_org_bairro_nome AS org,
3     dmd.min20_dst_bairro_nome AS dst,
4     td.tempo                 AS intervaloTempo, -- 20 minutos
5     COUNT(DISTINCT mf.objId) AS numObjetosMoveis
6 FROM
7     MovimentoFato AS mf INNER JOIN
8     DirMovDim     AS dmd ON (mf.dirMovId = dmd.dirMovId) INNER JOIN
9     TempoDim      AS td  ON (mf.tempoId = td.tempoId)
10 WHERE
11     td.tempo BETWEEN
12         to_timestamp('10-jan-2009 00:00', 'DD-mon-YYYY HH24:MI') AND
13         to_timestamp('10-jan-2009 23:40', 'DD-mon-YYYY HH24:MI')
14 GROUP BY
15     org, dst, intervaloTempo

```

(a)



(b)

Figura 5.9: (Consulta 2.3) Deslocamento dos veículos entre bairros adjacentes por intervalo de tempo. (a) Consulta em SQL; (b) Número de veículos se deslocando entre regiões no intervalo de tempo das 10h00..10h20.

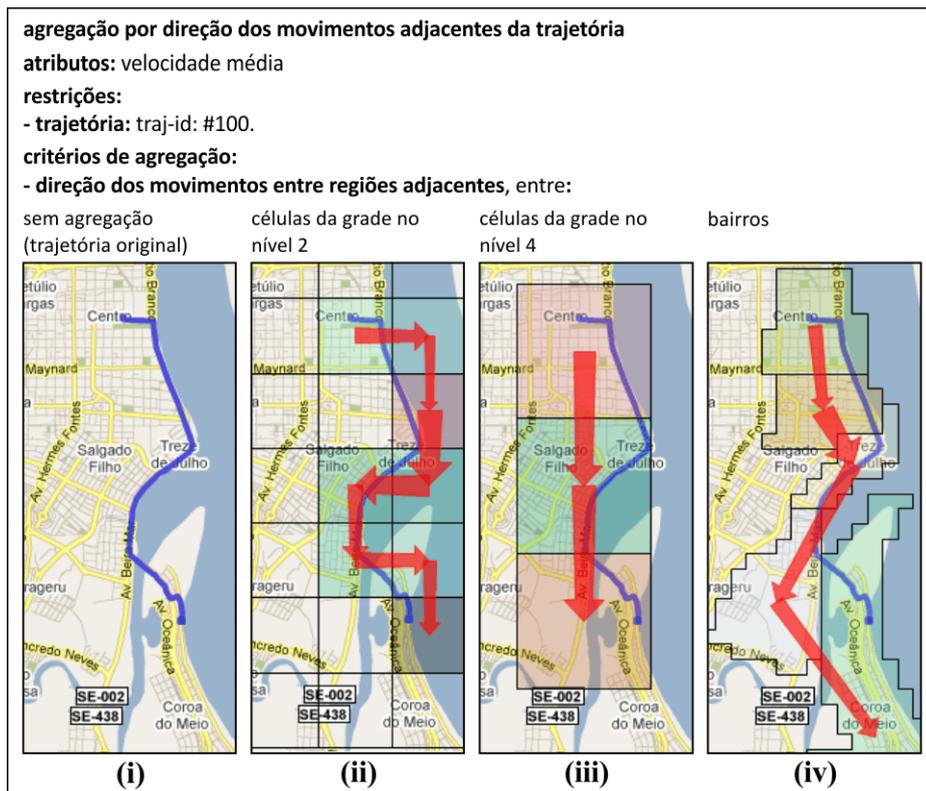
proporcionadas pelo esquema definido (Figura 5.1). Em **(b)(i)** é mostrada a trajetória original, e em **(b)(ii)** pode-se ver a trajetória representada em outro nível de granularidade, nesse caso, por células da grade regular no nível 2 (celulasNivel2). Nas figuras **(b)(iii)** e **(b)(iv)**, é mostrada a mesma trajetória por células da grade regular no nível 4 (celulasNivel4) e por bairro, respectivamente. A largura das setas é proporcional a velocidade média na região.

```

1  SELECT
2      dmad.org_celulaNivel2_latlon AS org,
3      dmad.dst_celulaNivel2_latlon AS dst,
4      AVG(mf.velocMedia)          AS velocidadeMedia
5  FROM
6      MovimentoFato AS mf      INNER JOIN
7      DirMovAdjDim AS dmad ON (mf.dirMovAdjId = dmad.dirMovAdjId)
8  WHERE
9      mf.trajId = 100
10 GROUP BY
11     org, dst

```

(a)



(b)

Figura 5.10: (Consulta 3.1) Diferentes representações para uma mesma trajetória, obtidas através da agregação de seus movimentos entre regiões adjacentes (roll-up para trajetórias). **(a)** SQL correspondente a figura **(b)(ii)**. **(b)** Trajetória **(b)(i)** exibida sobre diferentes níveis de agregação.

Consulta 3.2: Por Direção dos Movimentos da Trajetória

Consultas envolvendo agregação *por direção dos movimentos entre regiões no tempo* seguem a mesma lógica empregada nas agregações envolvendo a *direção dos movimentos adjacentes*, a única diferença entre as representações está na perspectiva. O primeiro, permite visualizar o deslocamento (origem-destino) sofrido pelos objetos móveis dentro de um intervalo de tempo, e o segundo, vê para esse mesmo intervalo de tempo o deslocamento sofrido pelos objetos entre as regiões adjacentes. São muito similares, mas têm propósitos diferentes.

Para exemplificar consultas envolvendo agregações por direção do movimento entre regiões no tempo, considere a consulta: *Exiba a trajetória #100 sobre diferentes níveis de granularidade por direção dos movimentos entre regiões no tempo.*

Sobre o fato *movimento*, e a dimensão *direção dos movimentos adjacentes*: **restrição:** para a trajetória #100; **agregação:** para a medida velocidade (*velocMedia*), usando a função de agregação média; **critérios de agregação:** *por origem e destino do movimento entre células adjacentes* no nível 2 (atributos: *org_celulaNivel2_latlon* e *dst_celulaNivel2_latlon*). Na Figura 5.10(a) é mostrada essa consulta escrita em SQL.

Na Figura 5.10(b) é exibido algumas das múltiplas representações para a trajetória #100 proporcionadas pelo esquema definido (Figura 5.1). Em **(b)(i)** é mostrado a trajetória original, e em **(b)(ii)** pode-se ver a trajetória representada em outro nível de granularidade, nesse caso, *por células da grade regular* no nível 2 (*celulasNivel2*). Nas figuras **(b)(ii)** e **(b)(iii)**, é mostrada a mesma trajetória por células da grade regular no nível 4 (*celulasNivel4*) e *por bairro*, respectivamente. A largura das setas é proporcional à velocidade média na região.

Consulta 3.3: Por Direção dos Movimentos entre Paradas da Trajetória

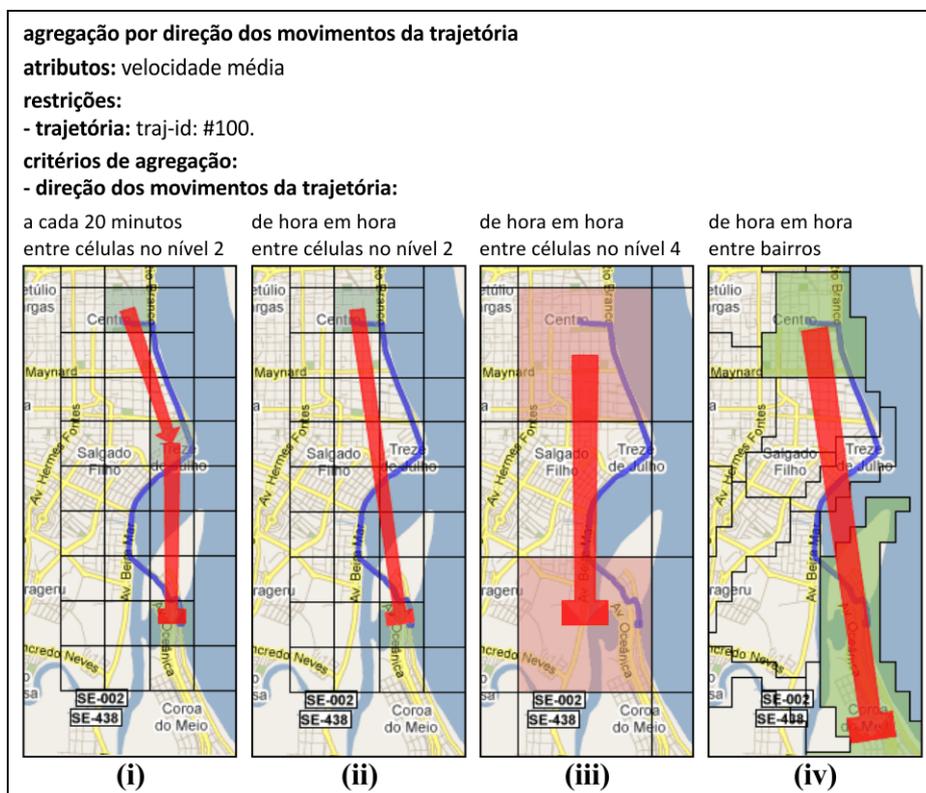
Para exemplificar consultas envolvendo agregações por direção do movimento entre duas paradas consecutivas, considere a consulta: *Mostre a velocidade média e o tempo de parada da trajetória #100 por direção do movimento entre paradas consecutivas.* Essa consulta é respondida através dos seguintes passos:

```

1  SELECT
2      dmd.celulaNivel2_latlon AS org,
3      dmd.celulaNivel2_latlon AS dst,
4      AVG(mf.velocMedia)      AS velocidadeMedia
5  FROM
6      MovimentoFato AS mf INNER JOIN
7      DirMovDim      AS dmd ON (mf.dirMovId = dmd.dirMovId)
8  WHERE
9      mf.trajId = 100
10 GROUP BY
11     org, dst

```

(a)



(b)

Figura 5.11: (Consulta 3.2) Diferentes representações para uma mesma trajetória, obtidas através da agregação de seus movimentos entre regiões no tempo (roll-up para trajetórias). (a) SQL correspondente a figura (b)(i). (b) Trajetória representada sobre diferentes níveis de agregação.

Sobre o fato *parada* e a dimensão *região de interesse* (RoIDim): **restrição:** para a trajetória #100; **agregação:** para a medida *velocidade* (*velocMedia*) através da função de agregação *média*; **critérios de agregação:** *por direção do movimento entre duas paradas consecutivas*, usando os dois relacionamentos com a dimensão *região de interesse* (RoIDim), *roi anterior* e *roi corrente*, usando as chaves estrangeiras *roiAnteriorId* e *roiId*, respectivamente. Na Figura 5.12(a) é apresentada essa consulta em SQL.

Na Figura 5.12(b) é mostrado o resultado dessa consulta, as setas representam o deslocamento do objeto móvel entre duas paradas consecutivas, e sua largura é proporcional à velocidade média entre as paradas. O tamanho dos quadrados nas paradas é proporcional ao tempo de parada.

```

1  SELECT
2      rd1.roi_latlon           AS org,
3      rd2.roi_latlon           AS dst,
4      pf.tempoParada           AS tempoParada,
5      AVG(pf.movEntreRoIs_velocMedia) AS velocidadeMedia
6  FROM
7      ParadaFato AS pf INNER JOIN
8      RoIDim     AS rd1 ON (pf.roiAnteriorId = rd1.roiId) INNER JOIN
9      RoIDim     AS rd2 ON (pf.roiId        = rd2.roiId)
10 WHERE
11     mf.trajId = 100
12 GROUP BY
13     org, dst, tempoParada

```

(a)



(b)

Figura 5.12: (Consulta 3.3) Representação da trajetória por direção dos movimentos entre paradas (roll-up para trajetórias). Em (a) SQL correspondente; (b) Representação para a trajetória.

5.4 Compactação dos Dados de Trajetória

Um dos maiores desafios enfrentados para a construção de um data warehouse de trajetórias, é a grande quantidade de dados de trajetória gerada pelo monitoramento dos objetos móveis. Para se ter uma noção do tamanho do problema em números, considere a base de dados gerada, o monitoramento dos 2.000 objetos móveis usados, durante 6 meses, gerou o equivalente a mais de *1 bilhão* de observações. A solução proposta para resolver esse problema envolve a sumarização dos dados de trajetória por parada e por célula espaço-temporal, como discutido na Seção 4.3.

Para avaliar o nível de compactação proporcionado pelo modelo proposto para dados de trajetória em relação ao modelo clássico, duas baterias de testes foram realizadas: (i) *variando apenas o tamanho das células*, o intervalo de captura é sempre o mesmo, 20 segundos; e (ii) *variando apenas o intervalo de captura entre as observações*, o tamanho de cada célula foi fixado em $200 \times 200 \text{ m}^2$. Os resultados dos experimentos (i) e (ii) são exibidos nas Figuras 5.13 e 5.14, respectivamente.

Do **experimento (i)** é possível concluir que: (a) a sumarização das paradas é a maior responsável pela compactação das trajetórias, sua proporção de compactação (quantidade de dados compactada dividido pela quantidade de dados original) foi de 0,5%; e (b) para movimentos, como esperado, quanto maior o tamanho das células, maior a taxa de compactação, pois mais movimentos são sumarizados em uma mesma tupla.

Analisando os resultados do **experimento (ii)**, é possível perceber a grande quantidade de dados armazenada usando o modelo clássico, até mesmo para intervalos de captura longos (acima de 1 minuto), onde as trajetórias capturadas são imprecisas. Para facilitar a comparação dos modelos clássico e proposto, nesse experimento, a solução proposta não usa a interpolação linear local para reconstruir os movimentos das trajetórias. Ao invés disso, supõe-se um método capaz de recuperar sempre os mesmos movimentos da trajetória, independente do intervalo de captura usado. Dessa maneira, a forma espacial da trajetória reconstruída é sempre a mesma.

Embora, nos experimentos (i) para células menores que $200 \times 200 \text{ m}^2$, e (ii) para intervalos de captura maiores que 20 segundos, a compactação dos movimentos gera mais dados que o número de observações. Isso ocorre porque o intervalo de captura usado é grande demais para o tamanho da célula, conseqüentemente, durante a carga de dados, novas observações são acrescentadas à trajetória para que seus movimentos se encaixem dentro dos limites das células, como discutido na Seção 4.3 (passo 3).

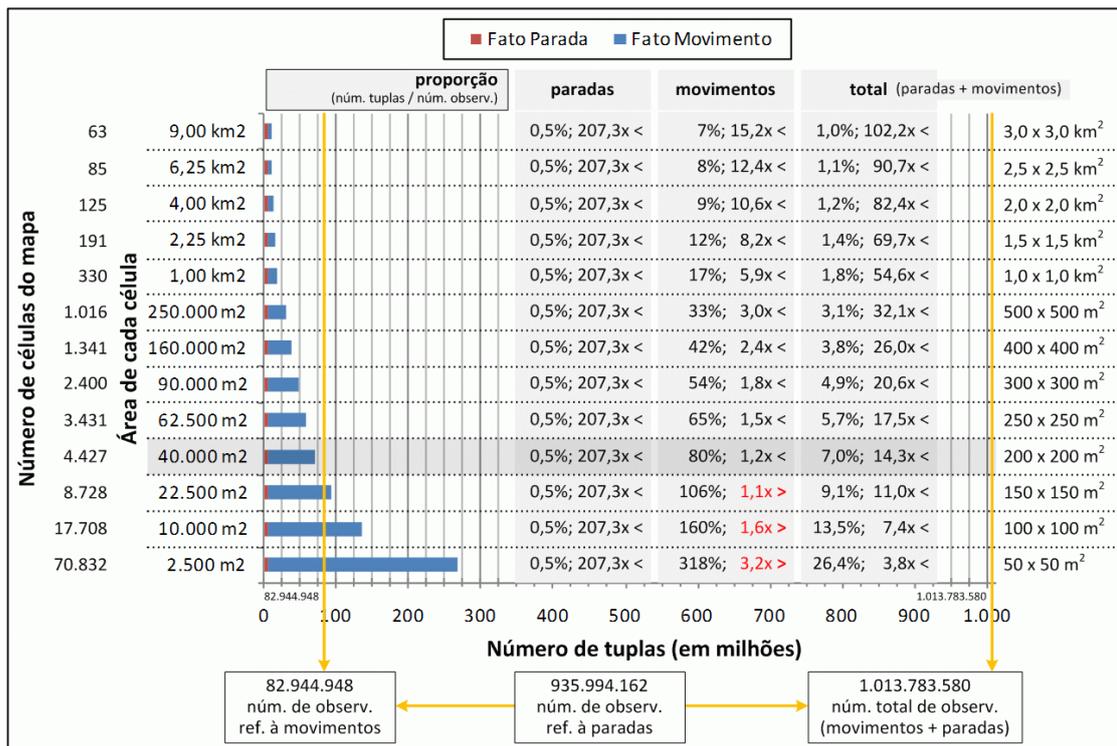


Figura 5.13: Tamanho dos fatos parada e movimento, usando o modelo proposto, variando apenas o tamanho das células da grade, o intervalo de captura é sempre o mesmo, 20 segundos.

Para se ter uma noção de como fica o mapa de Aracaju dividido por uma grade regular em diferentes níveis de granularidade, observe a Figura 5.15. O nível de granularidade ideal para a grade depende da aplicação, para o gerenciamento de tráfego urbano, este trabalho supõe que células de $200 \times 200 \text{ m}^2$ são suficientes. Quanto mais refinado for a grade regular, maior será a precisão das trajetórias armazenadas, e conseqüentemente maior a quantidade de dados necessária.

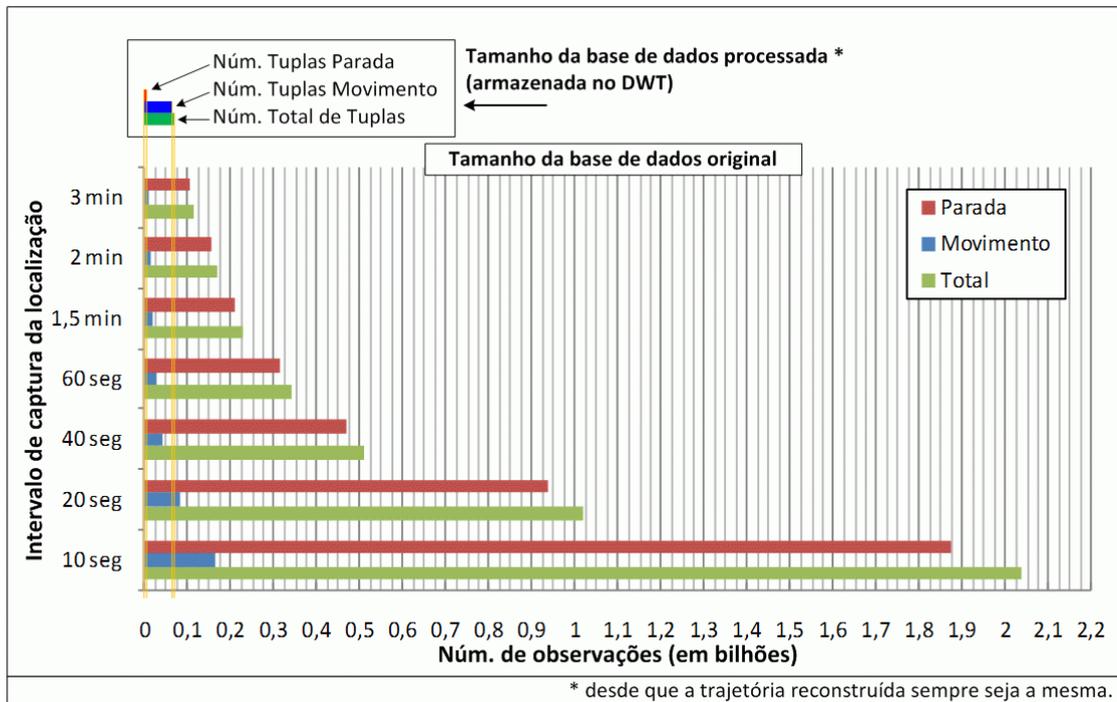


Figura 5.14: Tamanho dos fatos parada e movimento, usando o modelo proposto, *variando apenas o intervalo de captura entre as observações*, o tamanho de cada célula foi fixado em $200 \times 200 \text{ m}^2$.

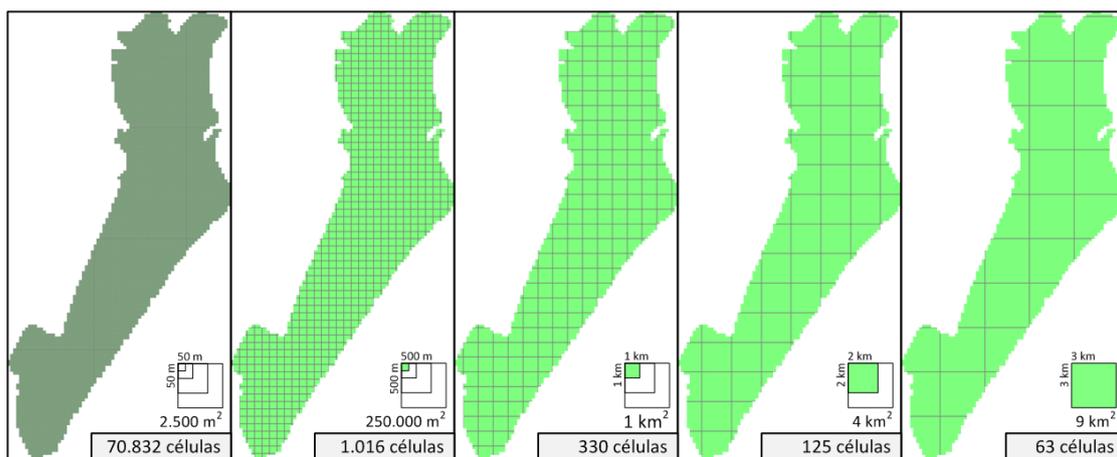


Figura 5.15: Mapa da cidade de Aracaju/Sergipe dividido por grades regulares sobre diferentes níveis de granularidade.

5.5 Tamanho das Dimensões Direção do Movimento

Para reduzir o número de dimensões *direção do movimento* e, conseqüentemente, o número de chaves estrangeiras no fato, a solução encontrada consiste em unir algumas dessas dimensões em uma só. Entretanto, essa solução possui um inconveniente: essas uniões podem gerar uma dimensão grande, o que pode levar a perda de desempenho. Para avaliar o tamanho dessas dimensões, vários experimentos foram realizados, envolvendo: (i) a *união das dimensões direção do movimento entre regiões adjacentes*; e (ii) a *união das dimensões direção do movimento no tempo*. Os resultados dos experimentos (i) e (ii) são mostrados nas Figuras 5.16 e 5.17, respectivamente.

Não foram encontrados problemas relacionados à união das dimensões no experimento (i), o número de registros gerados ficou abaixo de 1 milhão de tuplas, o que é recomendado para uma dimensão (Kimball; Ross; e Merz, 2002), mesmo quando se deseja analisar a movimentação entre pequenas regiões (por exemplo, regiões com $200 \times 200 \text{ m}^2$). Entretanto, no experimento (ii), a união de duas ou mais dimensões ultrapassou o número de tuplas recomendado. Portanto, para analisar a direção do movimento entre regiões menores que $1 \times 1 \text{ km}^2$, não é recomendada a união de dimensões.

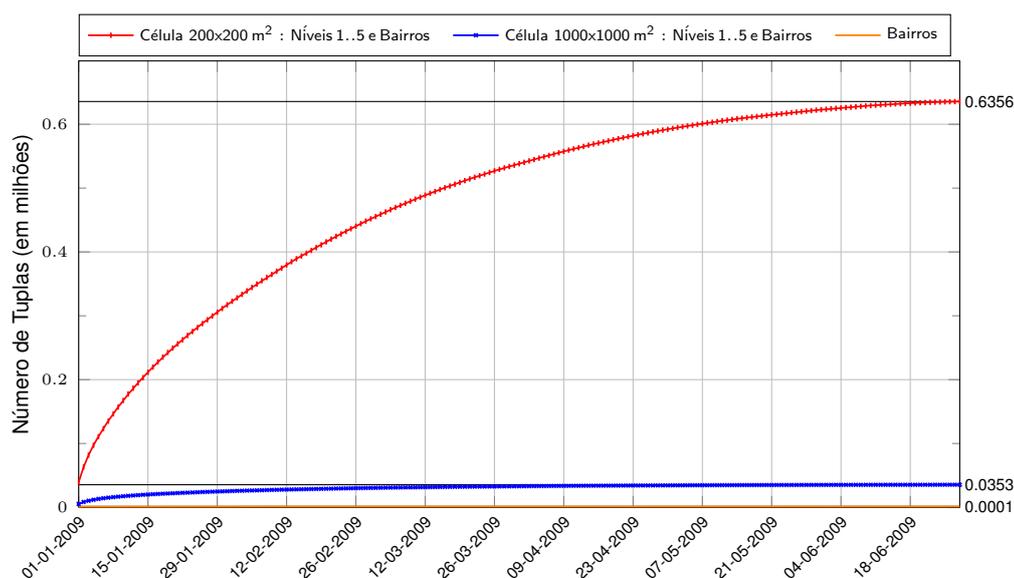


Figura 5.16: Tamanho das dimensões Direção do Movimento entre Regiões Adjacentes (DirMovAdjDim).

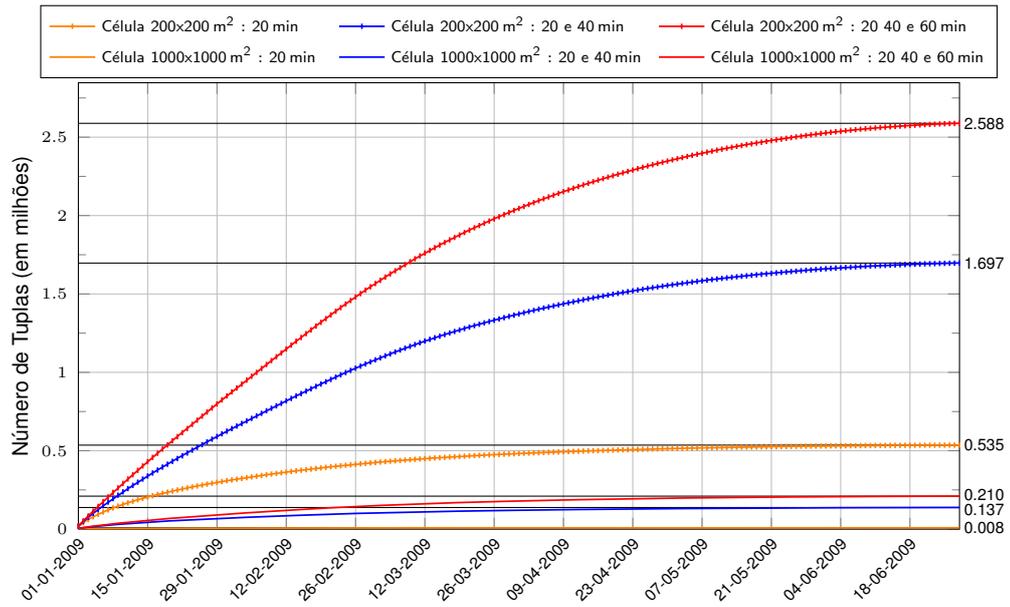


Figura 5.17: Tamanho das dimensões Direção do Movimento entre Regiões (DirMovDim) no tempo.

Capítulo 6

Conclusão

Nos parágrafos a seguir, são apresentadas as principais contribuições deste trabalho:

Modelo Proposto

Uma preocupação específica neste trabalho foi a de construir um modelo que permitisse analisar a movimentação dos objetos móveis *sobre* e *entre* as regiões no espaço e no tempo, análise orientada a tráfego e análise orientada a trajetórias, respectivamente. A primeira é obtida agregando-se as medidas das trajetórias por espaço e tempo. A análise orientada a trajetórias é proporcionada por um *conjunto de dimensões direção do movimento*, que possibilitam efetuar roll-up de trajetórias, permitindo assim analisar os movimentos da trajetória sobre diferentes níveis de granularidades.

Outra preocupação, foi a modelagem de trajetórias semânticas em DWTrs. Tal modelagem permite segmentar trajetórias em diversos componentes, tais como *paradas* e *movimentos*, que podem transportar informações que dão significado ao componente que pertencem.

Distinguir paradas de movimentos em trajetórias é fundamental para a análise correta de trajetórias. A inclusão de dados sobre paradas na análise de movimentos pode provocar forte discrepância entre os dados analisados e os reais, e vice-versa. Por exemplo, ao se incluir dados sobre paradas no cálculo da velocidade média de uma região, tem-se uma forte impressão de que a velocidade naquela região está baixa, devido a influência dos dados de paradas, cuja velocidade é igual a zero (ou muito próxima a isso), durante as computações.

Compactação de Trajetórias

Para amenizar o problema da grande quantidade dos dados de trajetória, propõe-se compactar trajetórias mediante a sumarização de suas paradas e movimentos. Dessa forma, consegue-se reduzir drasticamente o tamanho do *fato parada*, e de forma significativa o tamanho do *fato movimento*. Para avaliar o nível de compactação obtido, diversos experimentos foram realizados sobre uma mesma base de trajetórias, mas usando diferentes configurações para cada carga de dados executada (ver Seção 5.4). Com base nos resultados obtidos, conclui-se:

Paradas: A sumarização das paradas mostrou-se a maior responsável pela redução na quantidade de dados armazenada. Isso deve-se ao fato de que objetos parados permanecem muito tempo sobre uma mesma faixa espacial, isto é, uma região de interesse (RoI). Como a sumarização é por RoI, um grande número de observações são sumarizadas para um mesmo registro.

Movimentos: A sumarização dos movimentos mostrou-se eficaz apenas para células grandes (algumas centenas ou milhares de metros), e fica melhor a medida que o tamanho das células aumenta. Por exemplo, para células acima de: (i) $200 \times 200 \text{ m}^2$ a proporção *número de tuplas / número de observações* é menor que 70%; (ii) $300 \times 300 \text{ m}^2$ é menor que 50%; e (iii) $1.5 \times 1.5 \text{ km}^2$ é menor que 10%.

Tamanho das dimensões Direção do Movimento

Para possibilitar roll-up para trajetórias, a solução proposta consiste em armazenar todos os movimentos de todas as trajetórias em um conjunto de dimensões direção do movimento. Para minimizar o número de dimensões necessárias, os movimentos de um mesmo tipo são unidas em uma única dimensão (ver Seção 4.4). O inconveniente dessa solução é que o tamanho da dimensão resultante dessa união pode aumentar muito (acima de 1 milhão de tuplas), o que gera problemas de desempenho (Kimball; Ross; e Merz, 2002).

Para eficácia da solução proposta, diversos experimentos foram realizados, para calcular o tamanho das dimensões direção do movimento. Com base nos resultados obtidos dos experimentos, conclui-se que a solução adotada é eficaz para a *dimensão direção do movimento*

entre regiões adjacentes (DirAdjMovDim), cujo tamanho permaneceu abaixo do valor máximo recomendado (1 milhão de tuplas) (Kimball; Ross; e Merz, 2002), mesmo quando se analisa o deslocamento entre células pequenas (algumas dezenas de metros). Entretanto, o mesmo não acontece para a *dimensão direção do movimento entre regiões no tempo* (DirMovDim), a solução foi eficaz apenas para células grandes (acima de $1 \times 1 \text{ km}^2$), quando se analisa o deslocamento entre células pequenas, o tamanho da dimensão ultrapassou o valor máximo recomendado. Portanto, para esse tipo de movimento é recomendado cautela na união das dimensões, nesse caso, não é recomendável unir várias dimensões em uma só.

Base de Trajetórias

Inicialmente, buscou-se usar uma base de dados de trajetórias reais existente. Entretanto, existem alguns problemas que impossibilitaram seu uso: (i) por questões de privacidade poucas dessas bases estão disponíveis para livre uso; (ii) as bases encontradas são muito pequenas para avaliar a eficácia do modelo proposto e da compactação de trajetória; e (iii) essas bases são brutas, o que demanda um esforço muito grande para tratamento de dados, antes de estarem prontos para uso.

Devido a esses problemas, optou-se por desenvolver um protótipo para gerar a própria base de trajetórias usada neste trabalho, o *sintetizador de trajetórias semânticas*. Para dar um caráter mais realístico à base de dados, o protótipo desenvolvido permite criar alguns problemas de tráfego a partir de um conjunto de configurações pré-estabelecidas. Através dessas configurações é possível definir: o número de objetos móveis; os locais de parada; a velocidade máxima que os objetos podem atingir; o tempo de parada; entre outras. Para dar um caráter ainda mais realista a base gerada, a rota de cada trajetória - isto é, a sequência de coordenadas espaciais que vão da origem ao destino da trajetória - é obtida do Google Maps. De posse desses dados, o protótipo simula o deslocamento do objeto móvel.

6.1 Trabalhos Futuros

Durante o desenvolvimento deste trabalho, surgiram várias questões e ideias interessantes a serem incorporadas ao *modelo proposto* e ao *sintetizador de trajetórias desenvolvido*. A seguir, é apresentada uma lista de melhorias que foram deixadas para trabalhos futuros.

Referentes à *Carga de Dados*

Implementar métodos mais robustos para *reconstrução dos movimentos da trajetória*:

por questões de simplicidade, para reconstruir os movimentos da trajetória entre duas observações consecutivas, foi usada a *interpolação linear*, a qual considera que um objeto se desloca em linha reta a uma velocidade constante entre duas observações. Entretanto, existem alguns cenários onde é possível contar com informações adicionais para reconstruir esses movimentos, como a infra-estrutura de rede sobre a qual os objetos móveis se movem (p.ex. mapa de ruas).

Estudar métodos para *deteção de trajetórias e paradas*: as observações capturadas da movimentação dos objetos móveis formam uma massa bruta de dados, onde as trajetórias e suas paradas ainda não desconhecidas. Não se sabe o início e fim de cada trajetória, nem ao menos quantas trajetória existem ali. Portanto, antes dessa massa de dados ser usada, as trajetórias e suas paradas precisam ser previamente identificadas e rotuladas, processo conhecido como *deteção de trajetórias e paradas* (ver Seção 2.2). Por questões de simplificação, este trabalho adota que as trajetórias e suas paradas já são conhecidas.

Referentes aos *Operadores TrOLAP*

Implementar operadores para agrupamento de trajetórias similares: existem aplicações, onde é necessário identificar grupos de objetos móveis que se movimentaram em paralelo durante algum tempo, ou que adotaram rotas muito similares em suas viagens. Através do agrupamento de trajetórias similares, por exemplo, pode-se identificar quais são as rotas mais usadas pela população para ir de um lugar ao outro.

Referentes ao Sintetizador de Trajetórias

Gerar a rota da trajetória: em seu estado atual de desenvolvimento, a rota da trajetória – isto é, a sequência de coordenadas espaciais da origem ao destino da trajetória – é obtida do Google Maps. De posse desses dados, o protótipo simula o deslocamento do objeto móvel sobre essa rota. A vantagem dessa solução é que ela não necessita do mapa de ruas da cidade para gerar a base de trajetórias. A desvantagem é manter uma relação de dependência com serviços prestados por terceiros (Google Maps).

Como trabalho futuro, o protótipo deverá além de permitir baixar a rota do Google Maps, permitir gerar a rota da trajetória com base em uma mapa de ruas.

Melhorar o módulo para simulação de tráfego: no momento, quase todas as configurações estão baseadas nos atributos velocidade e número de objetos móveis por região no espaço e tempo. Isso já permite gerar alguns problemas de tráfego (p.ex. quando e onde devem ocorrer congestionamentos). Entretanto, com apenas esses atributos não é possível simular um tráfego real de forma convincente.

Referências

- Alvares, L. O.; V. Bogorny; B. Kuijpers; J. A. F. de Macedo; B. Moelans; e A. Vaisman (2007a). “A Model for Enriching Trajectories with Semantic Geographical Information”. Em: *GIS '07: Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. Seattle, Washington: ACM, páginas 1–8. ISBN: 978-1-59593-914-2. DOI: <http://doi.acm.org/10.1145/1341012.1341041>.
- Alvares, L. O.; V. Bogorny; B. Kuijpers; J. A. F. de Macelo; B. Moelans; e A. T. Palma (2007b). *Towards Semantic Trajectory Knowledge Discovery*. Relatório técnico. Hasselt University, página 12.
- Spatio-Temporal Aggregation for Visual Analysis of Movements* (2008), páginas 51–58. DOI: 10.1109/VAST.2008.4677356. URL: <http://dx.doi.org/10.1109/VAST.2008.4677356>.
- Andrienko, G.; N. Andrienko; e S. Wrobel (2007). “Visual Analytics Tools for Analysis of Movement Data”. Em: *SIGKDD Explor. Newsl.* 9.2, páginas 38–46. ISSN: 1931-0145. DOI: <http://doi.acm.org/10.1145/1345448.1345455>.
- Andrienko, G.; N. Andrienko; I. Kopanakis; A. Ligtenberg; e S. Wrobel (2008). “Visual Analytics Methods for Movement Data”. Em: *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Editado por F. Giannotti; e D. Pedreschi. Springer Publishing Company, Incorporated. Capítulo 13, páginas 375–410. ISBN: 9783540751762.
- Baltzer, O.; F. Dehne; S. Hambrusch; e A. Rau-Chaplin (2008b). *OLAP for Trajectories, Technical Report TR-08-11*. Relatório técnico. School of Computer Science, Carleton University, <http://www.scs.carleton.ca>.

- Baltzer, O.; F. Dehne; S. Hambrusch; e A. Rau-Chaplin (2008a). “OLAP for Trajectories”. Em: *Database and Expert Systems Applications*. Editado por S. S. Bhowmick; J. Küng; e R. Wagner. Volume 5181. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, páginas 340–347.
- Bédard, Y.; T. Merrett; e J. Han (2001). “Geographic Data Mining and Knowledge Discovery”. Em: editado por H. J. Miller; e J. Han. CRC Press. Capítulo Fundamentals of spatial data warehousing for geographic knowledge discovery, páginas 53–73. URL: <http://www.informaworld.com/smpp/content~content=a728339576>.
- Bogorny, V.; B. Kuijpers; e L. O. Alvares (2009). “ST-DMQL: A Semantic Trajectory Data Mining Query Language”. Em: *International Journal of Geographical Information Science* 23.10, páginas 1245–1276. DOI: 10.1080/13658810802231449. URL: <http://dx.doi.org/10.1080/13658810802231449>.
- Brakatsoulas, S.; D. Pfoser; e N. Tryfona (2004). “Modeling, Storing and Mining Moving Object Databases”. Em: páginas 68–77. DOI: 10.1109/IDEAS.2004.1319779.
- Braz, F.; S. Orlando; R. Orsini; A. Raffaetà; A. Roncato; e C. Silvestri (2007). “Approximate Aggregations in Trajectory Data Warehouses”. Em: *ICDE Workshops*, páginas 536–545.
- Braz, F. J. (2009). “Warehousing and Mining Aggregate Measures Concerning Trajectories of Moving Objects”. Tese de doutorado. Via Torino, 155, 30172 Venezia Mestre – Italia: Università Ca Foscari di Venezia.
- Codd, E. F.; S. B. Codd; e C. T. Salley (1993). *Providing OLAP (On-Line Analytical Processing) to User-Analysis: An IT Mandate*.
- Elmasri, R.; e S. B. Navathe (2006). *Fundamentals of Database Systems (5th Edition)*. Addison Wesley. URL: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike04-20{\&}path=ASIN/0321369572>.
- Franklin, C. (1992). “An introduction to geographic information systems: linking maps to databases”. Em: *Database* 15 (2), páginas 12–21. ISSN: 0162-4105. URL: <http://portal.acm.org/citation.cfm?id=135779.135781>.

- Frentzos, E. K. (2008). “Trajectory Data Management in Moving Object Databases”. Tese de doutorado. University of Piraeus.
- Frentzos, E. (2003). “Indexing Objects Moving on Fixed Networks”. Em: *SSTD*. Editado por T. Hadzilacos; Y. Manolopoulos; J. F. Roddick; e Y. Theodoridis. Volume 2750. Lecture Notes in Computer Science. Springer, páginas 289–305.
- Giannotti, F.; e D. Pedreschi (2008). *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Editado por F. Giannotti; e D. Pedreschi. Springer Publishing Company, Incorporated. ISBN: 9783540751762. URL: <http://portal.acm.org/citation.cfm?id=1355342>.
- Gómez, L. I.; S. Haesevoets; B. Kuijpers; e A. A. Vaisman (2007). “Spatial Aggregation: Data Model and Implementation”. Em: *CoRR* abs/0707.4304.
- Gomez, L.; B. Kuijpers; e A. A. Vaisman (2007). “Aggregation Languages for Moving Object and Places of Interest Data”. Em: *CoRR* abs/0708.2717. Tema: Turismo; Journal.
- Guting, H.; T. de Almeida; e Z. Ding (2006). “Modeling and Querying Moving Objects in Networks”. Em: *The VLDB Journal* 15.2, páginas 165–190. ISSN: 1066-8888. DOI: <http://dx.doi.org/10.1007/s00778-005-0152-x>.
- Inmon, W. H. (2005). *Building the Data Warehouse, Fourth Edition*. Wiley Publishing, Inc.
- Inmon, W. H. (2002). *Building the Data Warehouse*. 3ª edição. New York, NY, USA: John Wiley e Sons, Inc. ISBN: 0471081302.
- Jensen, C. S.; A. Kligys; T. B. Pedersen; e I. Timko (2004). “Multidimensional Data Modeling for Location-Based Services”. Em: *The VLDB Journal The International Journal on Very Large Data Bases* 13, páginas 1–21.
- Jensen, C. S. (2002). “Research Challenges in Location-Enabled M-Services”. Em: *Mobile Data Management*, páginas 3–7.
- Kimball, R.; M. Ross; e R. Merz (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. 2ª edição. Wiley Computer Publishing. ISBN: 0-471-20024-7.
- Kuijpers, B.; e A. A. Vaisman (2007). “A Data Model for Moving Objects Supporting Aggregation”. Em: *ICDE Workshops*, páginas 546–554.

- Macedo, J.; C. Vangenot; W. Othman; N. Pelekis; E. Frenzos; B. Kuijpers; I. Ntoutsis; S. Spaccapietra; e Y. Theodoridis (2008). “Trajectory Data Models”. Em: *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Editado por F. Giannotti; e D. Pedreschi. Springer Publishing Company, Incorporated. Capítulo 5, páginas 123–150. ISBN: 9783540751762.
- Malinowski, E.; e E. Zimányi (2004). “Representing spatiality in a conceptual multidimensional model”. Em: *GIS*, páginas 12–22.
- Marketos, G.; N. Pelekis; E. Frenzos; A. Raffaetà; I. Ntoutsis; e Y. Theodoridis (2008). “Building Real-World Trajectory Warehouses”. Em: *Proc. 7th International ACM SIGMOD Workshop on Data Engineering for Wireless and Mobile Access (MobiDE’08)*.
- Marketos, G. D. (2009). “Data Warehousing & Mining Techniques for Moving Object Databases”. Tese de doutorado. University of Piraeus.
- Matousek, K.; J. Mordacik; e L. Janku (2001). “On Implementing the Data Warehouse - GIS Integration”. Em: *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics: Information Systems Development-Volume I - Volume I*. ISAS-SCI ’01. IIS, páginas 206–210. ISBN: 980-07-7541-2. URL: <http://portal.acm.org/citation.cfm?id=646789.704053>.
- Meratnia, N.; e R. A. de By (2004). “Spatiotemporal Compression Techniques for Moving Point Objects”. Em: *EDBT*, páginas 765–782.
- Mokhtar, H.; J. Su; e O. Ibarra (2002). “On Moving Object Queries: (Extended Abstract)”. Em: *PODS ’02: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. Madison, Wisconsin: ACM, páginas 188–198. ISBN: 1-58113-507-6. DOI: <http://doi.acm.org/10.1145/543613.543638>.
- Open Geospatial Consortium (OGC). *KML*. URL: <http://www.opengeospatial.org/standards/kml/>.
- Orlando, S.; R. Orsini; A. Raffaetà; e A. R. C. Silvestri (2007). “Trajectory Data Warehouses: Design and Implementation Issues”. Em: *Journal of Computing Science and*

- Engineering (JCSE)* 1.2, páginas 211–232. URL: http://jcse.kiise.org/PublishedPaper/topic_abstract.asp?idx=11.
- Palma, A. L. T. (2008). “A Clustering-Based Approach for Discovering Interesting Places in Trajectories”. Dissertação de mestrado. Universidade Federal do Rio Grande do Sul.
- Pelekis, N.; A. Raffaeta; M. L. Damiani; C. Vangenot; G. Marketos; E. Frenzos; I. Ntoutsis; e Y. Theodoridis (2008). “Towards Trajectory Data Warehouses”. Em: *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Editado por F. Giannotti; e D. Pedreschi. Springer Publishing Company, Incorporated. Capítulo 7, páginas 189–211. ISBN: 9783540751762.
- Renso, C.; S. Puntoni; E. Frenzos; A. Mazzoni; B. Moelans; N. Pelekis; e F. Pini (2008). “Wireless Network Data Sources: Tracking and Synthesizing Trajectories”. Em: *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. Editado por F. Giannotti; e D. Pedreschi. Springer Publishing Company, Incorporated. Capítulo 3, páginas 73–100. ISBN: 9783540751762.
- Rinzivillo, S.; D. Pedreschi; M. Nanni; F. Giannotti; N. Andrienko; e G. Andrienko (2008). “Visually-Driven Analysis of Movement Data by Progressive Clustering”. Em: *Information Visualization 7.3-4*, páginas 225–239. ISSN: 1473-8716. DOI: 10.1057/palgrave.ivs.9500183. URL: <http://dx.doi.org/10.1057/palgrave.ivs.9500183>.
- Spaccapietra, S.; C. Parent; M. L. Damiani; J. A. de Macedo; F. Porto; e C. Vangenot (2008). “A Conceptual View on Trajectories”. Em: *Data & Knowledge Engineering* 65.1. Including Special Section: Privacy Aspects of Data Mining Workshop (2006) - Five invited and extended papers, páginas 126–146. ISSN: 0169-023X. DOI: DOI:10.1016/j.datak.2007.10.008. URL: <http://www.sciencedirect.com/science/article/B6TYX-4R68N7K-1/2/c1610033369ca327c7c20711a5c4014a>.
- Stefanovic, N.; J. Han; e K. Koperski (2000). “Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes”. Em: *IEEE Trans. on Knowl. and Data*

- Eng.* 12 (6), páginas 938–958. ISSN: 1041-4347. DOI: <http://dx.doi.org/10.1109/69.895803>. URL: <http://dx.doi.org/10.1109/69.895803>.
- Tao, Y.; e D. Papadias (2005). “Historical Spatio-Temporal Aggregation”. Em: *ACM Trans. Inf. Syst.* 23.1, páginas 61–102. ISSN: 1046-8188. DOI: <http://doi.acm.org/10.1145/1055709.1055713>.
- Tao, Y.; G. Kollios; J. Considine; F. Li; e D. Papadias (2004). “Spatio-Temporal Aggregation Using Sketches”. Em: *Data Engineering, International Conference on* 0, página 214. ISSN: 1063-6382. DOI: <http://doi.ieeecomputersociety.org/10.1109/ICDE.2004.1319998>.
- Theodoridis, Y.; M. L. Damiani; A. Gkoulalas-Divanis; B. Kuijpers; J. A. F. de Macêdo; G. Marketos; N. Pelekis; A. Raffaetà; Y. Saygin; e V. S. Verykios (2009). *Technical Report Deliverable 1.6 – Final report WPI*. Relatório técnico 014915. GeoPKDD consortium. URL: http://www.geopkdd.eu/files/dev_public/D1.6.pdf.
- Wolfson, O.; B. Xu; S. Chamberlain; e L. Jiang (1998). “Moving Objects Databases: Issues and Solutions”. Em: *Statistical and Scientific Database Management*, páginas 111–122. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.8292>.