UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Mineração de trajetórias em redes sociais geolocalizadas

Ricardo Miguel Puma Alvarez

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Ricardo Miguel Puma Alvarez

Mineração de trajetórias em redes sociais geolocalizadas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Alneu de Andrade Lopes

USP – São Carlos
Julho de 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

A473m Alvarez, Ricardo Miguel Puma
Mineração de trajetórias em redes sociais
geolocalizadas / Ricardo Miguel Puma Alvarez;
orientador Alneu de Andrade Lopes. -- São Carlos,
2017.
142 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Ciências de Computação e Matemática
Computacional) -- Instituto de Ciências Matemáticas
e de Computação, Universidade de São Paulo, 2017.

1. Mineração de trajetórias. 2. Reconstrução de
trajetórias. 3. Redes sociais baseadas em
localização. I. de Andrade Lopes, Alneu , orient.
II. Título.

Ricardo Miguel Puma Alvarez

Trajectory mining in location-based social networks

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the degree of the Master Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Alneu de Andrade Lopes

USP – São Carlos
July 2017

Às pessoas que se sentem realmente vivas com as coisas que fazem.

AGRADECIMENTOS

Antes de tudo, quero agradecer a minha família e gostaria de fazê-lo em minha língua mãe, espanhol. A mis queridos padres, Olga e Ismael, por haberme inculcado desde muy pequeño la costumbre del estudio, los valores de la responsabilidad, respeto por los demás, perseverancia, lucha y un gran deseo de superación. Gracias, también, por ser mi apoyo siempre, en las situaciones más difíciles y también en las fáciles, en todo momento. Gracias por haberme invitado a esta fiesta llamada vida que pienso aprovechar al máximo.

A mi hermano Eduardo, a quien admiro, admiré y seguiré admirando de seguro, por su gran destreza y valentía para afrontar las cosas y por solucionar muchos de los problemas diarios que para mi me resultan complicados. Gracias por ser el hermano mayor que tanto necesité y seguiré necesitando de seguro. Ambos nos motivamos mutuamente y seguro llegaremos más lejos de lo que pensamos.

A mis tíos Armando, Victoria; a mis primos, Fernando, Ayde; a mis sobrinas, Kiara, Valentina, Charbela. También, a mi tía Valeriana, Teresa, a Kelly que es una prima más.

A mi abuela, que ahora es "mamamama", un claro ejemplo viviente de fortaleza que valoro bastante y me sirve de inspiración. A mi tía Rosa, Hilda, tío Alex, Bruno, Atilio, Fernando, José (tío chino), a mis primos Sara, Carla, Angel, Josue, Luis, José, Andrea, Vicky, Angélica, Josue, Rosa, Paola, Silvia, Norma, Norka, Sue. Gracias por la alegría de poder vivir gratos momentos en esa playa grandiosa que es Santa Rosa.

Un saludo especial a mi amigo Marco al que desde que lo conocí sentí que lo conocía desde hace mucho. Gracias por ayudarme desde el primer día en Brasil y por haberme ayudado con la postulación a la maestría.

Agradezco también al profesor Rolando Maguiña de la UNMSM por ser el primero en confiar en mi capacidad e incentivarme a hacer la maestría. De igual forma a la profesora Nora La Serna y al profesor David Mauricio por haberme recomendado y al profesor César Beltrán por haber depositado su confianza en mí a pesar del poco tiempo de conocernos.

Um agradecimiento póstumo a quien en vida fue más que mi jefe un mentor en la vida profesional y personal, Carlos Takano, gracias por haberme enseñado tanto y haberme permitido tomarme el tiempo para estudiar y poder acceder a esta maestría, siempre llevaré presente tus enseñanzas, descansa en paz.

A mis amigos del colegio Víctor y Robert por brindarme una amistad que el tiempo ha dejado intacta y hasta más fuerte. Voluntad, disciplina y acción siempre!

A mis amigos de la FISI-UNMSM Gaby (Gabish), Jhonny (Apo), Alonso (Xavi), Karina, Lucia, Luis, Juan Carlos, Katty, Micaela, Jhussara (Lady Jhu), Giovana, Javier por poder contar con ustedes siempre a pesar de la distancia y la frecuencia.

Agradezco a mis amigas del ICPNA, Melsi y Fiorela por toda esa amistad, compañerismo y fuente inagotable de risas que han oxigenado mi camino en los momentos más tensos durante estos años, gracias.

A mis amigos del trabajo Miguel y David por estar siempre ahí, especialmente en las situaciones difíciles como fue este último año.

A mis amigos hispanos que conocí en São Carlos, Daniel, Paul, Lucho, Edson, Omar, Roque, Sonia, Susan, Mery, Sabrina, Paola, Alfredo, Katterine, Juan Luis, Germain, Erick, David Carbajal, David Saldaña, David Aveiga, Miguel, Rayner, Alex, Ludvin, Omar Chavez, Norman, Evinton, Jarlinton, José Carlos, Alberto, Elvis por todos los momentos vividos y los que vendrán.

Un saludo especial a Jorge, a sin quien seguramente no hubiera conseguido acabar esta maestría, y a quien considero fue mi co-orientador. Muchas gracias por todo!

Agradeço a Dona Nayd por ter me ensinado a apreciar a beleza das coisas mais simples da vida.

Ao professor Alneu, pela grande dedicação e paciência durante todo este tempo. Agradeço sua confiança e direcionamento que me ensinaram a dar o melhor de mim em um projeto de pesquisa.

Aos meus amigos brasileiros, Alan, Thiago, Natália, Giovana, RG, Vinícius, Fabiana, Roberta, Ivone, Celso, Diego pela amizade e gratos momentos vivenciados que guardo comigo.

Agradeço à Universidade de São Paulo (USP) e ao Instituto de Ciências Matemáticas e de Computação (ICMC) pela grande oportunidade de entrar no maravilhoso mundo da ciência e a pesquisa. Também, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo apoio financeiro.

RESUMO

PUMA-ALVAREZ, R. **Mineração de trajetórias em redes sociais geolocalizadas**. 2017. 142 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

O cada vez maior número de tecnologias que fornecem serviços de geolocalização tem possibilitado gerar uma grande quantidade de dados de geolocalização. Estes dados, são armazenados principalmente como pontos de localização com informação temporal. Uma trajetória é definida como uma sequência discreta e finita destes pontos de localização. Nos últimos anos, a recente área de mineração de trajetórias visa aproveitar esta abundância de dados. Nesta área, existem várias técnicas de mineração desenvolvidas, mas todas elas dependem diretamente da qualidade das trajetórias. Assim, o pré-processamento tem um papel primordial na mineração de trajetórias. Entre as tarefas de pré-processamento, um problema relevante é a reconstrução ou inferência de trajetórias. Devido ao alto consumo de energia de dispositivos de localização como o GPS e ao crescente uso de geo-marcações nas redes sociais, que possibilita a construção de trajetórias ordenando temporalmente estas marcações, muitas das trajetórias existentes apresentam taxas de amostragem muito baixas. A maioria das pesquisas nesse problema utilizam, no caso de áreas urbanizadas, informações do grafo formado por ruas e cruzamentos. Porém, elas levam em conta apenas trajetórias de veículos principalmente pelo fato que muitos dos percursos dos pedestres ficam fora das ruas. Atualmente, graças às plataformas livres de mapas colaborativos, é possível incluir estes trajetos como parte das informações de ruas. Assim, este projeto tem o objetivo de investigar o uso das informações das ruas na reconstrução de trajetórias, principalmente de pedestres. O escopo da proposta compreende o desenvolvimento de uma rede social geo-localizada com o intuito de capturar dados de localização. Posteriormente, estes dados serão anonimizados, utilizados na reconstrução de trajetórias de pedestres e disponibilizados para uso em pesquisas futuras.

Palavras-chave: Mineração de Trajetórias, Reconstrução de Trajetórias, Redes Sociais Baseadas em Localização. .

ABSTRACT

PUMA-ALVAREZ, R. **Trajectory mining in location-based social networks**. 2017. 142 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2017.

The ever-greater number of technologies providing location-based services has allowed the generation of big amounts of geolocation data. This data is mainly stored as location points in conjunction with temporal information. A trajectory is defined as a discrete and finite sequence of this kind of points. In recent years, the relatively new field of trajectory data mining aims to leverage this abundance of data. On this field, there are several data mining techniques developed, but all of these depend on trajectory quality. Hence, preprocessing becomes relevant to this field. Among trajectory data mining tasks, one important problem is trajectory reconstruction. Due to the high energy consumption of geolocation devices like GPS and the growing usage of geo-tags in social networks, which can represent trajectories by being sorted chronologically, most of these trajectories are collected at low sampling rates. A majority of research on this problem is focused on using road network information in urbanized areas to reconstruct trajectories. However, these approaches take into account vehicle trajectories only due to fact that most pedestrian paths do not always follow the same road network routes than vehicles. Currently, thanks to open collaborative maps, it is possible to add pedestrian paths to the road network structure. Thereby, this project aims to research the usage of road network information in pedestrian trajectories reconstruction. This project's scope comprises the development of a location-based social network to collect geolocation data. Subsequently, this data will be anonymized, used for pedestrian trajectory reconstruction and, finally, made available for research purposes.

Keywords: Trajectory Data Mining, Trajectory Reconstruction, Trajectory Inference, Location-Based Social Networks.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação de três grafos de uma LBSN: grafo usuário-localização, grafo usuário-usuário e grafo local-local	32
Figura 2 – Estrutura da modelagem do histórico de localização dos usuários em um domínio geoespacial. O GPS Logs de cada usuário contem os pontos geográficos que formam a trajetória percorrida por um usuário	34
Figura 3 – Exemplo de trajetória gerada a partir de um percurso contínuo de um objeto em movimento	36
Figura 4 – Exemplo de trajetória semântica	38
Figura 5 – Exemplo de rede viária e caminhos. Por meio de um processo de correspondência entre pontos de localização e arestas, conseguiu-se gerar um caminho na rede viária. Se a rede viária é denotada como $G = (V, E)$, onde seu conjunto de vértices é $V = A; B; C; D; E; F$ e seu conjunto de arestas dirigidas é $E = AB; BA; BC; CB; BD; DB; DE; ED; DF; FD$, então, o caminho gerado após o processo de correspondência é $\langle AB; BD; DE \rangle$	38
Figura 6 – Framework da Mineração de Trajetórias	39
Figura 7 – Uma consulta baseada em localização que procura trajetórias que atravessem A e B sucessivamente. As duas trajetórias pontilhadas são possíveis trajetórias candidatas. Uma região espacial em uma consulta baseada em região é representada por um retângulo tracejado em vermelho.	43
Figura 8 – Diagrama de implementação do sistema RadrPlus	54
Figura 9 – Modelagem de dados do sistema RadrPlus	56
Figura 10 – Perfil de usuário.	58
Figura 11 – Lista de amigos.	59
Figura 12 – Lista de comunidades.	59
Figura 13 – Página de uma comunidade.	60
Figura 14 – Visualização de trajetória	61
Figura 15 – Tela principal.	61
Figura 16 – Abas da tela principal	62
Figura 17 – Campo de buscas.	62
Figura 18 – Perfil de usuário.	63
Figura 19 – Perfil de usuário.	64
Figura 20 – Visualização de histórico de localização	66
Figura 21 – Comparação de perfil próprio com o de um amigo	66

Figura 22 – Mapa da comunidade.	68
Figura 23 – Menu do proprietário da comunidade	68
Figura 24 – Menu de um membro da comunidade	69
Figura 25 – Visualização do Timelapse	70
Figura 26 – Tag usado para indicar um acontecimento.	71
Figura 27 – Mecanismo de spooling do sistema RadrPlus	77
Figura 28 – Troca dinâmica de sensores segundo a velocidade	78
Figura 29 – Mudança do intervalo de amostragem segundo a velocidade de locomoção	78
Figura 30 – Diagrama de estados da mudança de intervalo de amostragem. Cada estado tem alocado o nome do meio de transporte mais provável segundo a velocidade, e o intervalo correspondente. As transições têm alocadas os valores de velocidade necessários para mudar de estado e, conseqüentemente de intervalo de amostragem. As transições sem valor indicam que se a velocidade for mantida no estado atual, haverá uma mudança sem necessidade de incremento de velocidade. Uma vez atingidos os estados de "Parado" e "Muito Rápido", a única forma de mudar de estado é por meio de um evento detectado pelos sensores acelerômetro e bússola respectivamente.	79
Figura 31 – Distribuição dos dados do RadrPlus	85
Figura 32 – Trajetórias coletadas na rodovia e rua usando o esquema de coleta contínuo.	87
Figura 33 – Trajetória coletada na rodovia usando o esquema de coleta adaptativo.	88
Figura 34 – Comparação de trajetórias coletadas na rodovia usando os esquemas de coleta adaptativo e contínuo.	89
Figura 35 – Comparação de trajetórias coletadas na rua usando esquemas de coleta adaptativo e contínuo.	90
Figura 36 – Exemplos de trajetórias incertas	94
Figura 37 – Processo de reconstrução de trajetórias	97
Figura 38 – Métodos de segmentação de trajetórias sobre uma rede viária	98
Figura 39 – Exemplo de Map matching	100
Figura 40 – Fluxo de trabalho do InferTra	102
Figura 41 – Fluxo do processo de experimentação da reconstrução de trajetórias.	112
Figura 42 – Precision, Recall e F-score de Shortest Path e InferTra para duas trajetórias usadas como ground truth.	114
Figura 43 – Resultados da reconstrução de trajetórias com os dados do RadrPlus utilizando os algoritmos InferTra e Shortest Path. Cada ponto no gráfico indica o valor do F-score obtido com uma dada janela de tempo para as configurações de segmentação (a) Tolerância Alta, (b) Tolerância Média e (c) Tolerância Baixa.	116

Figura 44 – Diferença estatística do desempenho das inferências de trajetórias medido pelo F-score para cada método e nível de tolerância sobre a base de dados RadrPlus.	117
Figura 45 – Resultados da reconstrução de trajetórias com os dados do Geolife utilizando os algoritmos InferTra e Shortest Path. Cada ponto no gráfico indica o valor do F-score obtido com uma dado intervalo de amostragem para as configurações de segmentação (a) Tolerância Alta, (b) Tolerância Média e (c) Tolerância Baixa.	118
Figura 46 – Diferença estatística do desempenho das inferências de trajetórias medido pelo F-score para cada método e nível de tolerância sobre a base de dados Geolife.	118
Figura 47 – Diferença estatística do desempenho das inferências de trajetórias para cada nível de tolerância levando em conta todas as combinações de bases de dados, técnicas de inferência e intervalos de amostragem.	120

LISTA DE QUADROS

Quadro 1 – Velocidade média de meios de locomoção	80
Quadro 2 – Velocidade média e intervalo de amostragem de meios de locomoção	90

LISTA DE ALGORITMOS

Algoritmo 1 – Algoritmo de InferTra	105
---	-----

LISTA DE TABELAS

Tabela 1 – Características da base de dados RadrPlus	85
Tabela 2 – Projeção comparativa de quantidade de leituras usando o esquema de coleta adaptativo em diversos meios de locomoção	91
Tabela 3 – Atributos das bases de dados utilizadas nos experimentos	110
Tabela 4 – Configurações da segmentação de trajetórias	111
Tabela 5 – Atributos dos conjuntos de trajetórias após segmentação.	115
Tabela 6 – Valores de F-score dos resultados da reconstrução de trajetórias nas bases de dados RadrPlus e Geolife sob as configurações de segmentação Tolerância Alta (TA), Tolerância Média (TM) e Tolerância Baixa (TB) para diferentes intervalos de amostragem.	119

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Contexto e motivação	25
1.2	Problema abordado	26
1.3	Objetivos	28
1.4	Contribuições	29
1.5	Organização	29
2	FUNDAMENTOS	31
2.1	Redes Sociais Baseadas em Localização	31
2.1.1	<i>Estrutura e Modelagem das LBSNs</i>	32
2.1.2	<i>Mineração de LBSNs</i>	34
2.2	Mineração de Trajetórias	35
2.2.1	<i>Definições</i>	37
2.2.2	<i>Framework da Mineração de Trajetórias</i>	38
2.3	Considerações finais	49
3	RADRPLUS: UMA REDE SOCIAL GEO-LOCALIZADA BASEADA EM COMUNIDADES	51
3.1	Definição	51
3.2	Arquitetura Tecnológica	53
3.3	Modelagem de dados	55
3.4	Funcionalidades	57
3.4.1	<i>RadrPlus web</i>	57
3.4.2	<i>RadrPlus Móvel</i>	61
3.5	Considerações finais	72
4	COLETA DE DADOS	75
4.1	Processo de coleta	75
4.1.1	<i>Métodos eficientes de coleta de dados geo-localizados</i>	75
4.1.2	<i>SmartTracking</i>	77
4.1.3	<i>Metodologia</i>	83
4.2	Base de dados	84
4.2.1	<i>Descrição</i>	84
4.2.2	<i>Formato</i>	85

4.3	Análise da técnica de captura	86
4.3.1	<i>Análise visual</i>	86
4.3.2	<i>Projeção comparativa</i>	89
4.4	Considerações finais	92
5	RECONSTRUÇÃO DE TRAJETÓRIAS	93
5.1	Reconstrução de trajetórias	93
5.1.1	<i>Trajeto�rias incertas</i>	93
5.1.2	<i>Infer�ncia de trajet�rias</i>	94
5.2	Reconstru��o de trajet�rias de pedestres usando informa��o da rede vi�ria	96
5.2.1	<i>Defini��es preliminares</i>	96
5.2.2	<i>Segmenta��o</i>	98
5.2.3	<i>Map-matching</i>	100
5.2.4	<i>Infer�ncia</i>	101
5.3	Considera��es finais	106
6	AVALIA��O EXPERIMENTAL	109
6.1	Conjuntos de dados	109
6.2	Metodologia experimental	110
6.2.1	<i>Pre-processamento de trajet�rias</i>	110
6.2.2	<i>Experimenta��o</i>	111
6.3	Resultados e an�lise	114
6.3.1	<i>Resultados da Segmenta��o</i>	114
6.3.2	<i>Resultados de Reconstru��o</i>	116
6.4	Considera��es finais	120
7	CONCLUS��ES	121
7.1	Conclus��es e principais contribui��es	121
7.2	Limita��es do trabalho	124
7.3	Trabalhos futuros	125
	REFER�NCIAS	127
	GLOSS�RIO	141

INTRODUÇÃO

1.1 Contexto e motivação

Atualmente, existem muitas tecnologias que fornecem serviços de geolocalização. Algumas delas são o Sistema de Posicionamento Global (GPS), Identificação por radiofrequência (RFID), sensores de celulares, sistemas ultra-sônicos e infravermelhos, etc ([FENG; ZHU, 2016](#)). Dessa forma, é possível gerar dados de trajetórias de objetos em movimento a grande escala e, conseqüentemente, desenvolver diversas tarefas de mineração de dados sobre eles. É neste contexto que a mineração de trajetórias se insere.

Existe um grande número de aplicações da mineração de trajetórias, tais como descoberta de caminhos, predição de localização, análise de comportamento de movimentação de indivíduos e grupos e aplicações no serviço urbano. Não obstante, há muitos desafios a serem endereçados no armazenamento, processo e mineração de trajetórias ([BARANIUK, 2011](#)).

Devido a questões de armazenamento e transmissão, as trajetórias são geralmente coletadas com taxas de amostragem relativamente baixas, o que gera intervalos de tempo longos entre atualizações de localização. Deste modo, estas trajetórias fornecem uma representação muito limitada do verdadeiro percurso. Este tipo de trajetórias são chamadas trajetórias incertas ou incompletas ([ZHENG; ZHOU, 2011](#)).

Em geral, as trajetórias podem ser rastreadas muito acuradamente com dispositivos com GPS como os *smartphones* ou os sistemas de navegação para carros. Porém, um estudo recente tem mostrado que, com intuito de reduzir o consumo de energia, a maioria dos táxis de cidades grandes utilizam uma taxa de amostragem de dois minutos ([WEI; ZHENG; PENG, 2012](#)). O alto consumo de energia do GPS limita, também, seu uso em *smartphones* por períodos de tempo longos. Além disso, a maioria de redes sociais fornecem serviços de *check-in* que permitem compartilhar a localização de um usuário. Ordenando estes *check-ins* temporalmente é possível gerar trajetórias. De forma similar, trajetórias podem ser geradas a partir de fotos geo-localizadas

(*geo-tagged photos*) em sites de compartilhamento de fotos como Flickr¹. Porém, quando as atualizações de localização são feitas através destas aplicações, tais trajetórias são geradas com taxas de amostragem baixas.

Endereçar este problema é de vital importância para as diferentes aplicações de mineração de trajetórias. Por exemplo, as trajetórias geradas por meio das fotos geo-localizadas podem ser reconstruídas ou inferidas para serem usadas em aplicações de recomendação de itinerários. Adicionalmente, a efetividade e eficiência das tarefas de indexação e processamento de consultas poderiam ser afetadas (ZHENG *et al.*, 2012).

1.2 Problema abordado

Motivados por este problema, diversos trabalhos têm sido publicados sobre a tarefa de reconstrução ou inferência de trajetórias. A maioria deles utiliza informação da rede de ruas e rodovias, criando um grafo onde as interseções representam os nós e as ruas as arestas. Por outro lado, tem-se também, alguns trabalhos que não levam em conta este tipo de informação (WEI; ZHENG; PENG, 2012), pois eles tentam atingir os casos de áreas não urbanizadas, trajetórias de animais ou fenômenos naturais como furacões. Porém, neste trabalho estamos interessados na reconstrução de trajetórias de pedestres em áreas urbanizadas.

Um exemplo de um método de inferência que utiliza a rede de rodovias é o *InferTra* (BANERJEE; RANU; RAGHAVAN, 2014). Nesse trabalho, ao invés de mostrar como resultado a rota mais provável, dados um ponto inicial e um final, ele gera um subgrafo com todas as possíveis rotas entre esses pontos. Cada aresta deste grafo tem alocada a probabilidade de um usuário transitar de um extremo da aresta ao outro. Para a inferência das possíveis rotas, o *textitInferTra* utiliza amostragem de Gibb (CASELLA; GEORGE, 1992) para aprender a partir de um banco de dados de trajetórias e gerar um Modelo de Mobilidade de Rede que será usado pelo algoritmo de inferência. Entre os trabalhos que utilizam a informação das ruas com maior destaque, tem-se: Hunter, Abbeel e Bayen (2013), Zheng *et al.* (2012), Li, Ahmed e Smola (2015) e Chiang *et al.* (2013).

Não obstante, todos estes trabalhos utilizam trajetórias de veículos, devido principalmente ao fato que o percurso feito por um pedestre vai além das ruas (e.g., trilhas, pequenos caminhos dentro de campus, parques, etc). Porém, graças a plataformas livres e colaborativas de mapas como OpenStreetMap², os usuários locais podem adicionar ao grafo de ruas aqueles nós e arestas que representam as trilhas e caminhos transitados exclusivamente pelos pedestres. Esta tendência possibilitaria o uso destas informações na inferência de trajetórias de pedestres.

Por outro lado, um desafio importante a ser considerado é a pouca quantidade de bases de dados geo-localizados disponíveis para pesquisa em mineração de trajetórias, principalmente

¹ <<https://www.flickr.com/>>

² <<https://www.openstreetmap.org>>

para a tarefa de reconstrução. A maioria de trabalhos utiliza os *check-ins* das redes sociais mais conhecidas (Facebook, Twitter) para formar as trajetórias, mas, como mencionado na seção anterior, as taxas de amostragem são baixas, o que, em consequência, impossibilita uma correta reconstrução. Por outro lado, atualmente existem redes sociais e serviços geo-localizados (e.g., Swarm³, Facebook Places), mas eles também guardam informação de localização através de *check-ins*. Além disso, muitas destas redes já foram fechadas (e.g, Brightkite (WAUTERS, 2011), Gowalla (CABALONA, 2012), Google Latitude (STEVENS, 2013)).

Além da taxa de amostragem, outro fator importante é encontrar uma base de dados que inclua trajetórias feitas por pedestres. Existem trabalhos que utilizaram dispositivos GPS para a coleta de dados com uma taxa de amostragem maior. Eles apresentam uma grande variedade de trajetórias, porém, a maioria deles não são de pedestres. Temos, por exemplo, ao T-drive⁴ e o Truck⁵, duas bases de dados de trajetórias de táxis e caminhões respectivamente. Por outro lado, temos a Bikely⁶, uma rede social de compartilhamento de trajetórias de ciclistas e, inclusive, bases de dados de trajetórias de animais como a Wild-Baboon⁷.

No entanto, a dificuldade é ainda maior quando se leva em conta questões como a privacidade. Um exemplo disto é Macaco⁸, uma base de dados de trajetórias de pessoas da Europa e América do Sul cujo acesso é restrito ao público devido às regulações de privacidade locais. Em contraste, temos o OpenStreetMap, que além de fornecer livre acesso ao mapa do mundo, permite, também, compartilhar e acessar as trajetórias disponibilizadas publicamente pelos usuários. Porém, uma limitação importante é o desconhecimento da forma de locomoção (e.g., carro, bicicleta, a pé) e o fato da taxa de amostragem não ser controlada.

Por fim, uma das poucas bases de dados de acesso público, com uma alta taxa de amostragem e que contem e permite distinguir trajetórias de pedestres, é a base do projeto Geolife⁹. Esta é uma rede social geo-localizada entre cujas principais funcionalidades destacam-se a captura e visualização de trajetórias de usuários, assim como a recomendação de amizades, itinerários e locais, baseado nos históricos de localização. Apesar do fato desta rede não estar disponibilizada como um serviço público, uma grande quantidade de dados foi obtida e disponibilizada publicamente após um processo de coleta de quatro anos, feita por um grupo de usuários com celulares habilitados com GPS. Porém, esta base de dados fornece apenas informações das trajetórias. Assim, informações dos usuários (e.g., idade, sexo, profissão) não são levadas em conta. Estas informações poderiam ser muito úteis para pesquisas na área, por exemplo, fornecendo contextos e enriquecendo os critérios de recomendação de amizades e locais. Adicionalmente, seria muito interessante acrescentar informações de comunidade, isto é, grupos criados por usuários da rede

³ <<https://www.swarmapp.com>>

⁴ <<https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/>>

⁵ <<http://chorochronos.datastories.org/>>

⁶ <<http://www.bikely.com>>

⁷ <<https://www.datarepository.movebank.org/handle/10255/move.405>>

⁸ <<https://macaco.inria.fr/macacoapp/>>

⁹ <<http://research.microsoft.com/en-us/projects/geolife/default.aspx>>

baseados em interesses em comum. Isto permitiria um estudo mais aprofundado do comportamento dos usuários, assim como abrir a porta para o estudo de comunidades em redes sociais geo-localizadas.

Finalmente, as informações apresentadas anteriormente convergem à ideia de que a reconstrução de trajetórias de pedestres é ainda uma área pouco explorada. Dentre as principais razões disto acontecer, têm-se a falta de informação nos mapas dos caminhos exclusivamente transitados por pedestres e as poucas bases de dados com as características necessárias para esta tarefa.

Contudo, o uso de plataformas como Openstreetmap e bases de dados como Geolife permitem superar essas dificuldades, possibilitando assim, a aplicação de métodos de reconstrução neste tipo de trajetórias. Por outro lado, mostrou-se a necessidade da existência de uma rede social geo-localizada que permita capturar dados de usuários, principalmente pedestres, com uma taxa de amostragem alta e que permita acessar os dados livremente. Além disso, a fim de garantir uma coleta de dados sustentável e um correto desenvolvimento e crescimento da área de pesquisa, mostra-se relevante que tal rede seja independente de terceiros. Por fim, comentou-se a relevância de incluir informações adicionais, de usuários e comunidades, neste tipo de redes com intuito de enriquecer as aplicações atuais e de abrir um leque de novas possibilidades de pesquisa.

1.3 Objetivos

O presente trabalho tem como objetivo geral investigar o uso de informações da rede viária na reconstrução de trajetórias de pedestres a partir de dados de localização coletados por uma plataforma baseada em dispositivos móveis como celulares com GPS. O objetivo geral pode ser subdividido em outros quatro objetivos específicos:

- A partir de um trabalho inicial de um *framework* para coleta de dados de trajetórias via celular, este trabalho desenvolveu novas funcionalidades para o desenvolvimento de uma rede social geo-localizada que permita capturar dados de localização com taxas de amostragem alta e que forneça informações adicionais de usuários e comunidades que possam ser usadas para outras pesquisas;
- desenhar um método de captura eficiente de dados de localização e utilizá-lo para coletar dados de usuários da rede social desenvolvida. Posteriormente, estes dados serão anonimizados e disponibilizados para uso em pesquisas futuras;
- investigar os métodos de reconstrução de trajetórias mais relevantes do estado da arte e estudar sua aplicabilidade nas trajetórias de pedestres;

- avaliar o desempenho da reconstrução de trajetórias de pedestres com os dados da rede social desenvolvida e de outras bases de dados utilizando diversos métodos do estado da arte.

1.4 Contribuições

Este trabalho foi iniciado com o desenvolvimento da rede social geo-localizada *Radrplus*. Esta é uma rede social geo-localizada, porém, diferentemente de outras redes sociais, com foco nas comunidades. Como toda rede social, permite fazer amizade com outros usuários e compartilhar informação pessoal. A rede é geo-localizada, pois coleta e usa as localizações geográficas dos usuários para fornecer funcionalidades interessantes dentro da rede. Tais funcionalidades levam em conta o grupo ou comunidade de usuários e não apenas o indivíduo como outros serviços disponíveis atualmente.

O sistema está disponibilizado tanto para plataformas web quanto móvel. Além de fornecer os dados de geolocalização, o aplicativo móvel tem muitas outras funcionalidades incluindo aquelas presentes na versão web. Entre as principais funcionalidades desta rede tem-se: gerenciamento de contas, gerenciamento de comunidades, visualização de trajetórias, *Smart Tracking* (mecanismo de *spooling* para situações sem conexão com a Internet). Por outro lado, um método de captura eficiente de dados foi desenvolvido e avaliado com os dados de usuários da rede desenvolvida. Este método utiliza os sensores do *smartphone* (bússola e acelerômetro) para detectar o movimento do usuário e calcular a sua velocidade a fim de variar a taxa de amostragem segundo um esquema pré-estabelecido com o intuito de poupar energia e perder o mínimo possível de informação. Este método destaca-se pela sua simplicidade e bom desempenho frente a outros métodos da literatura. Adicionalmente, foi criada uma base de dados de trajetórias de usuários através de um processo de coleta de nove meses. Estes dados, junto com outros de outra base, foram utilizados nos experimentos de inferência de trajetórias de pedestres, possibilitando, assim, a obtenção de importantes descobertas.

Finalmente, os resultados científicos obtidos foram utilizados na elaboração de um artigo submetido e aceito em uma conferência internacional importante da área: Puma-Alvarez, Ricardo e de Andrade Lopes, Alneu (2017). *Reconstructing Uncertain Pedestrian Trajectories From Low-Sampling-Rate Observations*. Em *4th Annual International Symposium on Information Management and Big Data*.

1.5 Organização

O restante deste documento está organizado da seguinte maneira:

- **Capítulo 2:** Apresenta os principais conceitos relacionados às redes sociais geo-localizadas e a área de mineração de trajetórias.

- **Capítulo 3:** Descreve as funcionalidades, filosofia, arquitetura, limitações e o processo de desenvolvimento do sistema *RadrPlus*.
- **Capítulo 4:** Expõe como foi realizado o processo de coleta de dados e descreve a base de dados obtida bem como o método de captura desenvolvido e seu desempenho.
- **Capítulo 5:** Apresenta a fundamentação, técnicas, aplicações e tendências futuras do problema de reconstrução de trajetórias e descreve o *framework* utilizado para o caso das trajetórias de pedestres.
- **Capítulo 6:** Exibe os experimentos e resultados da reconstrução das trajetórias obtidas do *RadrPlus* e de outra base de dados.
- **Capítulo 7:** apresenta as considerações finais do trabalho, conclusões, contribuições da pesquisa e sugestões de trabalhos futuros.

FUNDAMENTOS

Neste capítulo, são apresentados conceitos de Redes Sociais Baseadas em Localização e Mineração de trajetórias.

2.1 Redes Sociais Baseadas em Localização

Segundo (ZHENG; ZHOU, 2011), uma rede social baseada em localização (LBSN - Location-Based Social Network) não só compartilha informação de localização entre os membros de uma rede social, mas também fornecem uma nova estrutura social composta de indivíduos conectados entre si pela interdependência derivada de suas localizações no mundo real, bem como pelo conteúdo da mídia social tal como texto, vídeos e fotos.

Em síntese, a localização de um indivíduo em um instante de tempo e as anteriores localizações que um indivíduo teve em um dado período de tempo dão origem à localização física de tal indivíduo. Assim, a interdependência em uma LBSN implica a co-ocorrência de duas pessoas no mesmo local ou que estas compartilhem um histórico de localização similar, mas também que compartilhem interesses, atividades ou comportamentos comuns.

Existem três maneiras de representar as LBSNs, uma é de maneira absoluta (usando coordenadas de longitude e latitude), outra é relativa (tomando pontos de referência) e finalmente a simbólica (considerando lugares específicos inclusive de objetos com localizações fixas). Esta classificação dá lugar a seguinte categorização das LBSNs.

- a) As que usam geo-marcações na mídia, nestas redes os usuários adicionam rótulos de localização a suas fotos, vídeos ou texto. Por exemplo, Panoramio¹ e Flickr² são exemplos deste tipo, mas também qualquer outra rede que permita indicar a localização poder ser considerada, como no caso de Facebook ou Twitter.

¹ <<http://www.panoramio.com/>>

² <<http://www.flickr.com/>>

- b) As que usam localização pontual, nestas redes os usuários compartilham suas localizações atuais, a mais conhecida deste tipo é sem sombra de dúvida Foursquare, mas também podemos considerar a Google Latitude nesta categoria.
- c) As baseadas na trajetória, nesta categoria existem alguns serviços como Bikely³ que oferecem detalhes sobre as localizações pontuais e a rota que as conecta, estes detalhes podem ser, por exemplo, velocidade, distância percorrida e tempo de duração.

2.1.1 Estrutura e Modelagem das LBSNs

Devido aos diferentes atores ou elementos existentes nas LBSNs, a representação da sua estrutura é diferente das redes sociais tradicionais. Assim, como pode-se observar na Figura 1, usuários visitam alguns locais no mundo real e compartilham suas experiências gerando conteúdo com geo-marcações. Se todos os locais visitados por um usuário são conectados em termos de tempo, pode-se obter uma trajetória. Com base principalmente nas trajetórias dos usuários, podem ser gerados três grafos que facilitam a análise de uma LBSN: o grafo local-local, o grafo usuário-localização, e o grafo usuário-usuário.

Figura 1 – Representação de três grafos de uma LBSN: grafo usuário-localização, grafo usuário-usuário e grafo local-local

Fonte: Adaptada de Zheng e Zhou (2011).

No grafo usuário-localização existem dois tipos de vértices: usuários e locais. Uma aresta vai de um usuário até um local indicando a presença física do usuário nesse local. O peso da aresta neste grafo pode indicar o número de vezes que o usuário esteve nesse local em um determinado tempo. No grafo local-local, um vértice é um local e uma aresta direcionada indica que alguns usuários passaram por esses dois locais. O peso associado a uma aresta neste grafo

³ <<http://www.bikely.com/>>

representa a correlação entre os dois locais que formam tal aresta. No grafo usuário-usuário, um vértice é um usuário e uma aresta representa tanto a interdependência social entre usuários quanto a interdependência derivada a partir dos locais visitados por eles.

Com o uso da estrutura de três grafos de uma LBSN, os relacionamentos entre usuários e locais podem ser melhor estudados. Tal entendimento, facilita a modelagem do histórico de localização de um usuário. Assim, considerando que cada ponto de localização foi obtido por um sensor (*e.g.*, um GPS), (ASHBROOK; STARNER, 2003) e (HARIHARAN; TOYAMA, 2004) apresentaram as primeiras propostas para modelagem do histórico de localização. Tais propostas focaram principalmente na detecção de locais importantes considerando apenas as informações do sensor, isto é, não consideraram a importância dos locais para os usuários.

Para enfrentar esse problema, diferentes autores (LI *et al.*, 2008a; ZHENG *et al.*, 2009a; XIAO *et al.*, 2010; ZHENG *et al.*, 2011; ZHENG; XIE, 2011) apresentaram uma solução sistemática baseada no uso de uma estrutura hierárquica para a modelagem dos históricos de localização de usuários em um domínio geoespacial. Na Figura 2 é apresentado tal modelo, o qual é construído seguindo três passos descritos a seguir.

1. Detecção dos pontos de permanência: determinar todos os pontos de permanência, isto é, cada local ou região geográfica onde um usuário ficou por um dado período de tempo, em uma trajetória percorrida.
2. Construção da estrutura de hierarquia compartilhada: todos os pontos de permanência são armazenados em um banco de dados. Usando algum algoritmo de agrupamento hierárquico, deve-se particionar os elementos do banco gerando diferentes grupos de pontos de permanência. Dessa maneira, pontos de permanência comuns são agrupados no mesmo grupo. A estrutura de agrupamentos formada a partir dos pontos de permanência é chamada também de estrutura de hierarquia compartilhada.
3. Construção do grafo de localização pessoal: mediante a projeção do histórico de localização de um dado usuário sobre a estrutura de hierarquia compartilhada, pode ser construído um grafo direcionado, no qual um vértice representa um agrupamento contendo os pontos de permanência do usuário e as arestas representam a sequência de locais pertencente à trajetória percorrida por um usuário.

Quando a informação da permanência de um usuário em um local do mundo real está acompanhada de um significado semântico, *i.e.*, pontos de permanência nomeados como “museu”, “cinema”, “restaurante”, entre outros, é possível utilizar também o modelo de representação de históricos de localização apresentado anteriormente. Porém, nesse caso, uma das principais dificuldades encontra-se em identificar com exatidão os pontos de interesse (*point of interest* - POI) de um usuário, devido ao fato que sensores como o GPS terem margens de erro, bem como pela densidade da distribuição de POIs em uma região geográfica dada (*e.g.*, uma cidade).

Figura 2 – Estrutura da modelagem do histórico de localização dos usuários em um domínio geoespacial. O GPS Logs de cada usuário contêm os pontos geográficos que formam a trajetória percorrida por um usuário

Fonte: Adaptada de Zheng e Zhou (2011).

O uso do modelo de representação de históricos de localização, seja em um domínio geoespacial ou semântico, tem as seguintes vantagens: i) torna possível a modelagem do histórico de localização de usuários facilitando a realização de operações, tal como a comparação, entre dois ou mais históricos de localização; ii) torna possível a modelagem das rotas seguidas pelos usuários, facilitando o entendimento de seus comportamentos e interesses. É importante destacar que, a modelagem dos históricos de localização de usuários pode variar dependendo do problema abordado.

2.1.2 Mineração de LBSNs

A mineração dos dados de usuários das LBSNs leva ao entendimento do comportamento dos próprios usuários e dos locais, bem como da interdependência existente entre eles. Porém, é importante considerar que devido à estrutura própria das LBSNs, as tarefas de mineração de dados aplicadas nas redes sociais tradicionais têm que ser adaptadas. Assim, a seguir são apresentados algumas tarefas de mineração de LBSNs, as quais são direcionadas ao entendimento dos usuários e dos locais. Essas tarefas se reforçam mutuamente e não são abordadas isoladamente.

Tarefas direcionadas ao entendimento dos usuários: analisam o comportamento dos usuários com base nos relacionamentos existentes entre eles e nas informações dos locais que eles visitam.

- Estimação da similaridade entre usuários. A ocorrência de históricos de localização similares implica que usuários compartilham interesses e comportamentos similares

(LI *et al.*, 2008a; LI *et al.*, 2013a; LEE; CHUNG, 2011).

- Detecção de comunidades. Com base na similaridade estimada entre os históricos de localização dos usuários, é possível a detecção de grupos de usuários que mostrem determinados comportamentos em locais comuns (BACKSTROM *et al.*, 2006; MORIANO; FINKE, 2013).
- Predição de *links*. Com base na similaridade estimada entre os históricos de localização dos usuários, bem como nos relacionamentos existentes entre eles, é possível prever os futuros relacionamentos a serem estabelecidos, bem como os locais que poderiam ser visitados por um dado usuário (WANG *et al.*, 2011; SCELLATO; NOULAS; MASCOLO, 2011; NOULAS *et al.*, 2012a; STEURER; TRATTNER, 2013; ZHANG; KONG; YU, 2014).

Tarefas direcionadas ao entendimento dos locais: analisam a importância dos locais com base no comportamento dos usuários.

- Detecção dos locais mais interessantes. Minerando os históricos de localização dos usuários é possível detectar os locais mais interessantes ou mais visitados em uma determinada região geográfica (*e.g.*, uma cidade), e, assim, sugerir a visita desses locais a potenciais turistas (ZHENG *et al.*, 2009a).
- Planejamento de itinerários. Criação de itinerários personalizados em base ao conhecimento aprendido a partir dos históricos de localização de outros usuários com comportamentos similares (YOON *et al.*, 2010; YOON *et al.*, 2012).
- Recomendação de Locais-Atividades. É possível fazer dois tipos de recomendação: i) as atividades mais populares que são realizadas em um determinado local e ii) os locais mais populares para realizar uma determinada atividade. Esses dois tipos de recomendação usam, além dos históricos de localização, informações obtidas do conteúdo da mídia (*e.g.*, comentários em geo-marcações) (ZHENG *et al.*, 2010).
- Descoberta de eventos. Usando informações obtidas do conteúdo da mídia também é possível detectar, em uma determinada região geográfica, a ocorrência de eventos anômalos, tais como shows, acidentes de trânsito, festivais, entre outros (LEE; SUMIYA, 2010; LIU *et al.*, 2011).

2.2 Mineração de Trajetórias

Atualmente, a grande quantidade de tecnologias disponíveis que disponibilizam serviços de localização permitem gerar grandes quantidades de dados de trajetórias gerados a partir dos deslocamentos dos diferentes objetos em movimento.

Um percurso de um objeto em movimento em um espaço geográfico é contínuo, enquanto uma trajetória é apenas um conjunto discreto de amostras de pontos de localização pelas quais

o objeto se desloca. Comumente, uma trajetória espacial ou geográfica é representada por um sequência de pontos de localização acompanhados de uma estampa de tempo (*timestamp*), e.g., $(p_0, t_0), (p_1, t_1), \dots, (p_n, t_n)$. Em geral, a taxa de amostragem e o comprimento das trajetórias dependem da aplicação final. A Figura 3 ilustra a geração de uma trajetória a partir de um percurso contínuo de um objeto em movimento.

Figura 3 – Exemplo de trajetória gerada a partir de um percurso contínuo de um objeto em movimento

Fonte: Adaptada de Feng e Zhu (2016).

As fontes das quais são coletados os dados de trajetórias são muito variadas. Comumente, os veículos equipados com dispositivos GPS são as principais fontes deste tipo de dados. Porém, atualmente, têm-se muitas trajetórias geradas por *smartphones*, *check-ins*, conteúdo geo-localizado compartilhado na mídia, sensores RFID, etc. Desta forma, os objetos em movimento podem ser além de humanos, animais, veículos e até inclusive fenômenos naturais como furacões.

O número de aplicações que podem ser beneficiadas pela mineração de trajetórias é bem amplo. Dentre estas aplicações, têm-se a descoberta de rotas, predição de locais e destinos, análise de comportamento de indivíduos e grupos de pessoas, interpretação de trajetórias com aplicações em serviços urbanos, etc. Assim, estas aplicações não têm apenas o potencial de beneficiar usuários individuais, mas instituições governamentais, organizações comerciais, ONGs, etc.

Contudo, o processamento e a mineração de dados de trajetórias torna-se uma tarefa desafiante (BARANIUK, 2011; LIU *et al.*, 2013). Dentre os principais desafios a serem enfrentados, tem-se o armazenamento do grande volume de dados das trajetórias, o qual tem um crescimento cada vez mais acelerado. Por outro lado, é praticamente intratável definir uma medida de similaridade para comparar trajetórias (a qual é uma tarefa relevante para a mineração de trajetórias), dado que as trajetórias são geradas com diferentes estratégias e taxas de amostragem. Por fim, o processamento de consultas sobre as ingentes quantidades de dados de trajetórias torna-se uma tarefa altamente difícil em termos de complexidades espacial e temporal. Desta forma, com intuito de endereçar estes desafios, vários esforços têm sido investidos no desenvolvimento de diversas abordagens, as quais são apresentadas mediante um *framework* a fim de classificar todas estas abordagens de forma mais ordenada (FENG; ZHU, 2016). Neste *framework*, distinguem-se

basicamente três camadas: coleta de dados, técnicas de mineração de dados, e aplicações. A camada de técnicas de mineração, especificamente, está constituída por cinco componentes principais: Pré-processamento, Gerenciamento de dados, Processamento de consultas, Tarefas de mineração de trajetórias, e Proteção da privacidade. Cada um destes componentes junto com os das outras camadas já mencionadas são detalhados nesta seção, porém, primeiro, com intuito de garantir um melhor entendimento do *framework*, são apresentados as definições dos conceitos mais importantes em quanto à mineração de trajetórias.

2.2.1 Definições

Definição 1 (Trajetória). Uma trajetória de um objeto em movimento é uma representação discreta de o percurso contínuo realizado por tal objeto em movimento em um espaço geográfico. Em geral, uma trajetória é uma sequência de pontos geolocalizados com uma estampa de tempo correspondente. Ela pode ser representada da seguinte forma: $T = \langle p_1, t_1 \rangle, \langle p_2, t_2 \rangle, \dots, \langle p_n, t_n \rangle$, onde cada elemento $\langle p_i, t_i \rangle$ indica que um objeto em movimento esteve na localização p_i no tempo t_i . Adicionalmente, estes pontos estão ordenados ascendentemente segundo o valor da estampa de tempo, i.e., $t_j < t_k$ se $1 \leq j < k \leq n$. Uma trajetória poderia representar o percurso feito por um pessoa da sua casa ao trabalho, o percurso feito por um animal durante sua busca de comida ou o trajeto feito por um veículo durante uma viagem. Cada ponto de localização é comumente expressado como uma tupla de $\langle longitude, latitude \rangle$ indicado um único ponto no espaço geográfico. Por outro lado, as taxas de amostragem das trajetórias variam bastante dependendo da fonte dos dados. Principalmente, o consumo de energia e as limitações de armazenamento são fatores que influenciam nesta variedade de taxas de amostragem. Assim, por exemplo, as trajetórias de veículos usualmente têm taxas de amostragem mais altas que as dos dispositivos móveis como *smartphones*, devido a sua maior capacidade de energia e armazenamento. A trajetória definida pode ser chamada trajetória geográfica, pois está composta por pontos de geolocalização, assim, em seguida apresentamos um outro tipo de trajetórias, as trajetórias semânticas.

Definição 2 (Trajetórias semânticas). As trajetórias semânticas são geralmente geradas através de etiquetas geolocalizadas que contém além da posição geográfica, alguma referência significativa de um lugar no mundo real. Estas etiquetas geolocalizadas são comumente chamadas de *check-ins*. Adicionalmente, as trajetórias que são associadas com alguma descrição que expressa algum tipo de emoção ou sentimento também são chamadas de trajetórias semânticas. Na Figura 4, pode-se observar um exemplo de trajetórias semântica.

Definição 3 (Rede viária). A rede viária é um grafo dirigido $G = (V, E)$, onde V e E são um conjunto de vértices e arestas respectivamente. Um vértice $v_j \in V$ é uma interseção de segmentos de rua ou um ponto terminal. Uma aresta $e_k = v_p v_q \in E$ denota um segmento de rua dirigido, sobre o qual o objeto em movimento viaja de v_p a v_q .

Figura 4 – Exemplo de trajetória semântica

Fonte: Elaborada pelo autor.

Definição 4 (Caminho). Um caminho $P = \langle e_1, e_2, \dots, e_{|P|} \rangle$, representa uma sequência de arestas na rede viária, onde $e_i \in E$ e $e_i \neq e_j$ se $i \neq j$. Assim, arestas consecutivas e_i e e_{i+1} devem compartilhar um vértice o qual é o vértice destino de e_i e o vértice origem e_{i+1} .

A Figura 5, mostra um exemplo de rede viária e um caminho nela.

Figura 5 – Exemplo de rede viária e caminhos. Por meio de um processo de correspondência entre pontos de localização e arestas, conseguiu-se gerar um caminho na rede viária. Se a rede viária é denotada como $G = (V, E)$, onde seu conjunto de vértices é $V = A; B; C; D; E; F$ e seu conjunto de arestas dirigidas é $E = AB; BA; BC; CB; BD; DB; DE; ED; DF; FD$, então, o caminho gerado após o processo de correspondência é $\langle AB; BD; DE \rangle$.

Fonte: Adaptada de Feng e Zhu (2016).

2.2.2 Framework da Mineração de Trajetórias

Como mencionado no começo da seção, Feng e Zhu (2016), propõem um *framework* que resume o processo inteiro da mineração de trajetórias. Tal *framework* pode-se ser observado na Figura 6. É importante mencionar que nem todos os passos na camada de técnicas de mineração de dados são necessários. Apenas alguns deles podem ser aplicáveis dependendo da aplicação e a data coletada. O *framework* começa com a geração e posterior coleta de dados de trajetórias a partir de diferentes objetos em movimento. Depois disso, são aplicadas as técnicas de

mineração de dados estruturadas nos seus cinco componentes: pré-processamento, gerenciamento de dados, processamento de consultas, tarefas de mineração de dados, e proteção da privacidade. Finalmente, na camada de aplicações, o conhecimento obtido na anterior camada é utilizado em algum dos seis tipos diferentes de aplicações.

Figura 6 – Framework da Mineração de Trajetórias

Fonte: Adaptada de [Feng e Zhu \(2016\)](#).

• Pré-processamento.

O pré-processamento é uma tarefa básica que é realizada no começo do processo de mineração e visa melhorar a qualidade dos dados de trajetórias, bem como gerar sub-trajetórias. Existem cinco tipos comuns de operações de pré-processamento nesta etapa.

- *Limpeza.* A presença de *outliers* nas trajetórias reduz drasticamente o desempenho das técnicas de mineração de dados. Existem trabalhos que têm endereçado o problema de detectar objetos suspeitos em movimento ou capturar as características de trajetórias com padrões de movimentação anômalos ([GE et al., 2010](#)). Por outro lado, a tecnologia RFID apresenta problemas de ambiguidade dos dados em certos casos como, por exemplo, a incapacidade de definir uma localização determinística dada a detecção de um objeto por múltiplos leitores. Desta forma, a limpeza de trajetórias tenta descartar localizações impossíveis ou trajetórias que estejam tirando partido de restrições específicas como velocidade máxima ou restrições inatingíveis ([FAZZINGA et al., 2014](#)).
- *Segmentação.* Em muitas aplicações, uma trajetória é particionada em sub-trajetórias ([PELEKIS et al., 2010](#)), cada uma das quais é chamada de segmento, partição ou *frame*. A geração de sub-trajetórias é uma ideia racional devido a que, naturalmente, as trajetórias estão compostas de unidades menores bem definidas, por exemplo,

os segmentos de rua em uma trajetória restrita a rede viária. Uma abordagem de partição e sumarização pode ser encontrado em [Su et al. \(2015\)](#), onde tenta-se gerar uma descrição compreensível pelas pessoas dos dados de trajetórias, e, além disso, dividir uma trajetória em várias partições segundo os comportamentos dos objetos em movimento. Aliás, em [Wang et al. \(2014\)](#), [Wang et al. \(2015\)](#), as trajetórias são segmentadas em *frames* a fim de armazenar eficientemente pontos de localização de um objeto em movimento, os quais estão alinhados em intervalos de tempo com intuito de aproveitar os sistemas de armazenamento orientados a colunas do estado da arte.

- *Reconstrução.* Devido as limitações de armazenamento e transmissão de dados, as trajetórias são majoritariamente coletadas sob baixas taxas de amostragem, desta forma, fornecendo observações parciais das rotas reais. Este tipo de trajetórias com baixas taxas de amostragem são chamadas de trajetórias incertas. A inferência destas trajetórias levando em conta as restrições da realidade é um problema muito relevante devido a sua implicância em diversas tarefas de mineração de trajetórias. Assim, existem vários trabalhos que têm endereçado esta questão ([BANERJEE; RANU; RAGHAVAN, 2014](#); [LI; AHMED; SMOLA, 2015](#); [CHIANG et al., 2013](#); [ZHENG et al., 2012](#); [LARUSSO; SINGH, 2013](#)).
- *Calibração.* As trajetórias representam uma aproximação discreta das rotas originais sob distintas estratégias e taxas de amostragem. Esta heterogeneidade tem um efeito negativo sobre a estimação da similaridade entre trajetórias. Por exemplo, a comparação de duas trajetórias geradas sob diferentes estratégias de amostragem por meio de medidas de aproximação espacial como a distância Euclideana, torna-se uma tarefa complexa. No obstante, têm-se trabalhos que visam transformar estas trajetórias heterogêneas em trajetórias com estratégias de amostragem unificadas ([SU et al., 2013](#); [SU et al., 2015](#)).
- *Amostragem.* As diferentes abordagens na área de mineração de trajetórias comumente realizam operações sobre bancos de dados de trajetórias de grande escala, assim, estas operações resultam complicadas e custosas em termos de tempo. Assim, as abordagens de amostragem de trajetórias ([PELEKIS et al., 2010](#); [PANAGIOTAKIS et al., 2012](#); [LI et al., 2015a](#)) visam reduzir um grande banco de dados de trajetórias, tomando apenas as amostras mais importantes e representativas do banco de dados. Nesse processo, é importante manter nestas amostras os padrões de mobilidade ocultos no banco de dados original. Um exemplo desta tarefa pode ser encontrado em [Pelekis et al. \(2010\)](#), onde uma trajetória é representada como um vetor simbólico que quantifica a representatividade de cada trajetória, e, então, propõe um método não supervisionado para amostrar as trajetórias representativas. Por outro lado, [Panagiotakis et al. \(2012\)](#) realiza a amostragem de trajetórias no nível de sub-trajetórias, derivando descritores locais das trajetórias que representam

segmentos de linha. Além disso, um estudo recente cujo objetivo é suportar as consultas agregadas de trajetórias, adicionalmente, endereça o problema de amostragem focando na aproximação do processamento de consultas com restrições de tempo de resposta.

- **Gerenciamento de dados.**

O armazenamento de grandes quantidades de dados de trajetórias é sem sombra de dúvida um problema crucial na mineração de trajetórias. Assim, este problema é abordado a partir de dois aspectos relevantes (FENG; ZHU, 2016).

- *Compressão.* Devido a que alguns pontos de localização são frequentemente redundantes, o armazenamento e transmissão de ingentes quantidades de dados de trajetórias torna-se um processo ainda mais problemático. Desta forma, têm surgido algoritmos de compressão de trajetórias (NIBALI; HE, 2015; SONG *et al.*, 2014a; LIU *et al.*, 2015) que visam reduzir os requerimentos de armazenamento e as cargas de comunicação. Sobretudo, uma abordagem de compressão realiza um *tradeoff* entre a razão de compressão e o erro máximo. Em geral, quanto maior a razão de compressão, pior a qualidade das trajetórias comprimidas. Nibali e He (2015), por exemplo, consegue uma razão de compressão relativamente boa, desde que os usuários estejam dispostos a tolerar uma pequena quantidade de erro. Esta abordagem possui um preditor que prediz o próximo ponto de localização, e um esquema de codificação residual que gera pequenos resíduos com intuito de compensar a diferença entre o valor predito e o valor real. Por outro lado, (SONG *et al.*, 2014a) apresenta um *framework* que separa a representação espacial de uma trajetória da representação temporal e propõe um algoritmo de compressão para cada caso. Desta forma, a compressão resulta eficiente devido principalmente ao paralelismo dos procedimentos de compressão temporal e espacial. Assim, as consultas espaço-temporais podem ser executadas sem ter que descomprimir os dados das trajetórias. Por outra parte, Liu *et al.* (2015) propõe um sistema de compressão online limitado por erro. Esta abordagem cria um sistema de coordenadas virtual que é centrado no ponto inicial e constrói envoltórios convexos para limitar os pontos. Aliás, outros trabalhos argumentam que os dados de trajetória originais são significativamente grandes e deveriam ser simplificados (LONG; WONG; JAGADISH, 2013; LONG; WONG; JAGADISH, 2014). Porém, tal simplificação poderia gerar um tipo de compressão sujeito a perdas de dados.
- *Sistemas de armazenamento e estruturas de índices.* Como mencionado anteriormente, o armazenamento de enormes quantidade de dados de trajetórias é um problema de extrema relevância para a área de mineração de trajetórias (WANG *et al.*, 2014; WANG *et al.*, 2015; CUDRE-MAUROUX; WU; MADDEN, 2010; RANU *et al.*, 2015; POPA *et al.*, 2011; NI; RAVISHANKAR, 2007). A fim de melhorar o desempenho do processamento de consultas de dados de trajetórias, Wang *et al.*

(2014), Wang *et al.* (2015) propõem o SharkDB, um sistema de armazenamento de trajetórias baseado em memória. Para utilizar o armazenamento orientado em coluna, as trajetórias são particionadas em *frames*, logo, estes *frames* são posteriormente comprimidos e estruturados corretamente a fim de suportar a mineração de trajetórias. Outro sistema similar é o TrajStore (CUDRE-MAUROUX; WU; MADDEN, 2010), o qual é um sistema de armazenamento dinâmico que é capaz de recuperar todos os dados de uma região específica eficientemente. Ao invés de simplesmente indexar dados geo-espaciais, o TrajStore particiona as trajetórias em sub-trajetórias segundo as regiões espaço-temporais, e armazena dados compactados em cada região em conjunto. Uma estrutura de índice chamada *TrajTree* (RANU *et al.*, 2015) é desenvolvida para gerenciar dados de trajetórias especialmente para tarefas de recuperação como consultas *k-NN*. Assim mesmo, outro estudo levado a cabo em Ni e Ravishankar (2007) apresenta um método de indexação de espaço paramétrico que utiliza aproximações polinomiais para indexar segmentos de linha em dados de trajetórias.

- **Processamento de consultas.**

A recuperação de dados de um sistema de armazenamento subjacente é uma tarefa de grande importância. O objetivo da recuperação é encontrar os dados procurados de forma eficiente. Assim, em seguida são apresentados diversos tipos de consultas que são projetados para recuperar dados de trajetórias.

- *Consultas baseadas em localização.* O objetivo das consultas baseadas em localização é encontrar trajetórias que estejam próximas às localizações presentes na consulta, sendo estas um pequeno conjunto de localizações com ou sem uma restrição de ordem específica. Uma aplicação comum é a recomendação de rotas para uma viagem a muitos destinos diferentes. Na Figura 7, por exemplo, um viajante na localização origem *S* deseja encontrar uma rota válida que passe pelas localizações *A* e *B*, presentes em uma consulta, e atinja finalmente o destino *D*. Este tipo de problema foi endereçado em Chen *et al.* (2010), onde consegue-se encontrar as *k* trajetórias melhor conectadas (*k-BCT*) cada uma das quais conecta todas as localizações presentes na consulta. Por outro lado, um problema de busca baseada em localização onde os pontos de localização têm importâncias diferenciadas é abordado em Yan *et al.* (2015). Neste trabalho, uma localização com uma fotografia geo-etiquetada é mais importante do que outras localizações sem conteúdo semântico. A consulta proposta deve encontrar as *k* trajetórias conectadas mais importantes (*k-ICT*). Adicionalmente, Yan *et al.* (2015) propõe algoritmos eficientes para responder este tipo de consultas, aproveitando tanto a proximidade espacial quanto a duração temporal em cada lugar importante.

- *Consultas baseadas em regiões.* As consultas baseadas em regiões sobre as trajetórias são importantes em um amplo leque de aplicações de mineração de trajetórias. Uma consulta que especifica se um valor encontra-se dentro de uma região delimitada por um limite inferior e outro superior é considerada uma consulta baseado em regiões. Um exemplo deste tipo de consultas é encontrar todas as trajetórias de um viajante em específico entre 11 e 14h. Na Figura 7, pode-se observar uma consulta baseada em regiões espaciais que procura todas as trajetórias que atravessem um região espacial denotada por um retângulo tracejado. Têm-se trabalhos que estudam este tipo de consultas no caso de trajetórias incertas. [Zheng et al. \(2011\)](#) criaram pela primeira vez um modelo probabilístico para representar as possíveis localizações de um objeto em movimento em um tempo específico. Adicionalmente, [Zheng et al. \(2011\)](#) projetaram uma estrutura de índice efetiva para processar consultas baseadas em regiões probabilísticas. Por fim, outro trabalho ([ZHAN et al., 2015](#)), utiliza consultas baseadas em regiões para recuperar trajetórias que estão consistentemente cobertas por uma dada área com uma alta probabilidade durante certo período de tempo.

Figura 7 – Uma consulta baseada em localização que procura trajetórias que atravessem A e B sucessivamente. As duas trajetórias pontilhadas são possíveis trajetórias candidatas. Uma região espacial em uma consulta baseada em região é representada por um retângulo tracejado em vermelho.

Fonte: Adaptada de [Feng e Zhu \(2016\)](#).

- *Consultas dos vizinhos mais próximos (NN).* A busca dos vizinhos mais próximos é uma tarefa muito comum é importante na mineração de trajetórias espaço-temporais ([GÜTING; BEHR; XU, 2010](#); [NIEDERMAYER et al., 2013](#)). Em [Güting, Behr e Xu \(2010\)](#), é endereçado o problema de encontrar os k vizinhos mais próximos a uma dada trajetória em um banco de dados de trajetórias para um intervalo de tempo específico. Alias, [Niedermayer et al. \(2013\)](#) estende as consultas de vizinhos mais próximos a um caso mais realista, as consultas probabilísticas de vizinhos mais

próximos, onde as trajetórias são incertas. Tomando como base uma representação de trajetórias incertas como processos estocásticos, os autores estudam vários tipos de consultas probabilísticas de vizinhos mais próximos cujas entradas são uma dada trajetória e um intervalo de tempo.

- *Consultas Top-k*. O problema de encontrar as k trajetórias mais similares para uma dada trajetória em um conjunto de trajetórias incertas é endereçado por [Ma et al. \(2013\)](#). O solução a este problema repousa sobre a ideia de quantificar adequadamente as dissimilaridades entre duas trajetórias incertas. Uma nova métrica de distância é proposta e uma estrutura de índice é construída posteriormente a fim de suportar tarefas de mineração de trajetórias.
- *Consultas baseadas em padrões*. Em [Vieira, Bakalov e Tsotras \(2010\)](#), é projetada uma consulta que seleciona trajetórias baseado em um padrão específico de movimentação. Este tipo de consulta oferece um conjunto de restrições que devem ser satisfeitas em uma determinada ordem. Um predicado ou restrição pode ser uma condição do tipo baseado em região ou uma condição de vizinho mais próximo. O principal problema presente no processo deste tipo de consultas é criar uma descrição suficientemente robusta para poder representar padrões em uma consulta como expressões regulares.
- *Consultas agregadas*. [Li et al. \(2015a\)](#) apresenta outro tipo de consulta chamada de consultas agregadas de trajetórias cujos resultados não são trajetórias mas medições agregadas. Este tipo de consultas tenta recuperar estatísticas das trajetórias que atravessam uma dada região em um intervalo de tempo específico. Um exemplo de uma consulta agregada no caso de trajetórias de veículos é recuperar a velocidade média atingida para um segmento de rua especificado pelo usuário.
- *Consultas ad-hoc*. Têm-se muitas consultas específicas para distintas aplicações. Por exemplo, em [Zheng et al. \(2015\)](#), um consulta de palavras-chave para trajetórias semânticas é explorada. Este tipo de consultas resulta muito útil para diversas aplicações baseadas em localização. Pode-se imaginar um turista que gostaria de almoçar em um *restaurante* e assistir um filme no *cinema*. Assim, estas duas palavras-chave seriam extraídas para procurar trajetórias relacionadas. A maior diferença entre este tipo de consulta e as tradicionais consultas geoespaciais é que não existem geolocalizações neste tipo de consulta. O trabalho visa suportar a busca eficiente de trajetórias para palavras-chave aproximadas em dados de trajetórias semânticas. Por outro lado, em [Zheng et al. \(2013\)](#), um caso especial de trajetórias semânticas chamadas de trajetórias de atividades é apresentado; adicionalmente, o problema da busca eficiente deste tipo de trajetória é endereçado. Nas trajetórias de atividades, os conteúdos semânticos que são alocados nas localizações são informações de atividades em lugares particulares como esporte, compras, almoço e entretenimento. Este trabalho

visa, também, identificar as trajetórias com atividades similares, o qual seria muito útil para aplicações de recomendação de rotas e planejamento de viagens.

- **Tarefas de mineração de dados.**

As tarefas de mineração de dados são classificadas em várias categorias segundo o tipo de tarefa.

- *Mineração de padrões.* A mineração de padrões visa analisar os padrões de mobilidade de um objeto em movimento ou de múltiplos objetos se movimentando juntos. Existem diversos tipos de padrões, tais como padrões de reuniões ou grupos (ZHENG *et al.*, 2013; ZHENG *et al.*, 2014; LI *et al.*, 2013b), padrões sequenciais (ZHANG *et al.*, 2014) e padrões periódicos (LI *et al.*, 2010; LI *et al.*, 2012). Se observarmos cada trajetória com uma sequência, um padrão sequencial é frequentemente definido como uma subsequência que pelo menos é compartilhada por δ trajetórias, onde δ é um limiar. O problema de minerar padrões sequenciais em trajetórias semânticas é abordado em Zheng *et al.* (2014). Neste trabalho, é proposto um procedimento de dois passos chamado SPLITTER, o qual visa descobrir padrões sequenciais ocultos. SPLITTER primeiro recupera uma coleção de padrões pouco refinados agrupando lugares similares juntos e, em seguida, gera padrões mais refinados segmentando um padrão pouco refinado de cima para baixo. Um padrão periódico é um padrão de trajetória muito comum, o qual é relevante no entendimento do comportamento dos objetos em movimento. O problema de mineração da periodicidade é estudado por Li *et al.* (2010), Li *et al.* (2012). Este trabalho resolve dois subproblemas cruciais, a detecção de períodos baseado em locais de referência, e a mineração do comportamento de movimento periódico baseado em modelos probabilísticos.
- *Agrupamento.* Uma linha de pesquisa (VRIES; SOMEREN, 2010; LIU *et al.*, 2013; LIU *et al.*, 2010) é dedicada ao estudo do problema de agrupamento de trajetórias devido à utilidade de agrupar trajetórias com padrões de movimento similares. Liu *et al.* (2013) apresenta um *framework* chamado TODMIS para a mineração de comunidades a partir de diversas fontes de trajetórias. Grupos de objetos em movimento são identificados baseado em informações relacionadas com as trajetórias (e.g., dispersão espacial, duração temporal, velocidade de deslocamento), bem como conteúdos semânticos das localizações. Um problema interessante e útil é detectar os *hot spots* ou pontos de acesso de veículos em movimento. Dado que os pontos de acesso podem ser interpretados como áreas de alta densidade de veículos, o agrupamento é um método promissório para resolver este problema. Diferentemente dos agrupamentos tradicionais baseados em densidade, este trabalho usa um agrupamento baseado em mobilidade, cuja ideia principal é que um veículo com alta mobilidade (velocidade) provavelmente implique uma baixa aglomeração e vice-versa. Este agrupamento

baseado em mobilidade é menos sensível ao tamanho da base de trajetórias do que o agrupamento baseado na densidade.

- *Classificação.* A classificação de trajetórias é a tarefa de construir um modelo a partir de um conjunto de dados de treinamento e aplicá-lo na predição das etiquetas das trajetórias de teste. [Patel et al. \(2012\)](#) abordam um problema de classificação e propõem o uso da informação da duração para melhorar a acurácia da predição. Este método leva em conta não apenas informação espacial, mas informação da duração como características para a classificação, dado que a duração tem um papel relevante na diferenciação de objetos em movimento a diferentes velocidades.
- *Descoberta de conhecimento.* O conhecimento subjacente nas trajetórias pode ser muito útil para muitas aplicações. Por exemplo, [Yuan et al. \(2015\)](#), [Yuan, Zheng e Xie \(2012\)](#) estudam a descoberta de regiões de diferentes funções em uma cidade. Este conhecimento poderia ajudar aos cidadãos a tomar melhores decisões como se investir em imóveis ou não. Outros estudos ([WANG et al., 2013](#); [AU et al., 2010](#)) focam no problema de detecção de eventos. Os eventos detectados representam um tipo de conhecimento valioso.

- **Proteção da privacidade.**

Fornecer suporte às tarefas de mineração de trajetórias e proteger a privacidade dos usuários é uma problema desafiante. Contudo, existem muitas abordagens para a proteção da privacidade ([ANDRIENKO et al., 2015](#); [COSTA et al., 2011](#); [PELEKIS et al., 2011](#); [KONG et al., 2015](#)) nas técnicas de mineração de trajetórias. Em [Andrienko et al. \(2015\)](#), é utilizada uma técnica de preservação da privacidade que transforma dados geo-referenciados originais em trajetórias em um espaço semântico abstrato sobre o qual os dados das trajetórias são posteriormente processados. Por outra parte, o SmartTrace ([COSTA et al., 2011](#)) encontra as trajetórias mais similares a partir de uma dada trajetória usada como consulta. Este método tenta realizar uma medição da similaridade distribuída onde os usuários não precisam subir seus dados e, desse modo, proteger sua informação de geolocalização sensível. O trabalho de [Pelekis et al. \(2011\)](#) facilita o compartilhamento consciente da privacidade de dados de mobilidade e desenvolve um motor de trajetórias para fornecer acesso restrito ao banco de dados de trajetórias. Finalmente, [Kong et al. \(2015\)](#) utiliza um esquema de encriptação homomórfico para garantir a preservação da privacidade na recuperação de trajetórias.

- **Aplicações da mineração de trajetórias.**

Existem uma grande quantidade de aplicações baseadas na mineração de trajetórias. Em seguida, são apresentados seis tipos de aplicações.

- *Descoberta de caminhos* A descoberta de caminhos é uma das aplicações mais comuns da mineração de trajetórias, devido a que é muito importante encontrar o caminho mais adequado em muitos cenários. A adequação de um caminho depende da aplicação, e.g., o mais rápido, o mais curto, o mais popular, etc. Um grande número de pesquisas têm sido realizadas a respeito (LU; LEE; TSENG, 2011; CHEN; SHEN; ZHOU, 2011; WEI; ZHENG; PENG, 2012; LUO *et al.*, 2013; LIU *et al.*, 2014; DAI *et al.*, 2015) A descoberta de caminhos é também chamada de descoberta de rotas, e seu objetivo é encontrar pelo menos um caminho que satisfaça um objetivo predefinido dados uma origem e um destino. As rota ou caminhos devem ser derivados levando em conta uma rede viária específica. Assim, as trajetórias com formatos de localizações geográficas devem ser transformadas em um formato composto por uma sequência de segmentos de rua derivados da rede viária escolhida. Neste tipo de aplicações, o histórico de trajetórias fornece um valioso conhecimento para estimar, comparar e construir caminhos candidatos.
- *Predição de destinos ou locais*. Os serviços baseados em localização fornecem importantes benefícios as pessoas em áreas urbanas. Muitas aplicações baseadas em localização precisam da predição de locais ou destinos para enviar publicidade a um público objetivo, a fim de recomendar lugares turísticos, ou para alocar destinos nos sistemas de navegação. A predição do destino está estreitamente relacionada com a descoberta de caminhos. Se uma viagem em andamento coincidissem com uma parte de uma rota frequente na base de dados de trajetórias históricas, o destino da rota frequente poderia ser o destino da viagem em andamento. Porém, existem algumas restrições no mundo real. Por exemplo, (XUE *et al.*, 2013b; XUE *et al.*, 2013a) remarcam a relevância do problema da escassez dos dados, o qual indica que as trajetórias disponíveis são muito poucas para cobrir todas as possíveis trajetórias. Para enfrentar o problema da escassez de dados, todas as trajetórias são particionadas em sub-trajetórias, e, assim trajetórias sintéticas são geradas conectando sub-trajetórias. Desta forma, um conjunto de trajetórias que pode suportar a predição de destinos é exponencialmente incrementado com este método. Por outro lado, (NOULAS *et al.*, 2012b) endereçam o problema de prever o próximo lugar que um usuário visitará, explorando os padrões de mobilidades humanos. Uma grande quantidade de dados de *check-ins* são usados para estudar a movimentação das pessoas com uma representação qualitativa.
- *Análise de comportamento de indivíduos*. Os dados de trajetórias possibilitam o análise do comportamento locomotor de objetos em movimento (RENZO *et al.*, 2013; SONG *et al.*, 2014b; ANDO; SUZUKI, 2011; GAO *et al.*, 2013; LIU; QU; WANG, 2015; GONZALEZ; HIDALGO; BARABASI, 2008). A descoberta de padrões de movimentação é de grande importância no entendimento do comportamento humano. Um desafio importante neste campo é extrair informação semântico de

alto nível do comportamento dos indivíduos como, por exemplo, inferir os fins ou os papéis de um objeto em movimento. Em (RENZO *et al.*, 2013), é proposta uma abordagem para entender o comportamento de indivíduos que se movimentam em um contexto geográfico, extraíndo padrões de comportamento locomotor. Dessa forma, o comportamento dos indivíduos é inferido a partir desses padrões os quais são minerados dos dados das trajetórias. Aliás, predizer o comportamento dos indivíduos acuradamente em situações de emergência é muito importante antes, durante e depois do desastre. Song *et al.* (2014b) analisam o comportamento de indivíduos em situação de emergência e seus padrões de mobilidade depois de um grande acidente nuclear no Japão. A análise demonstrou que o comportamento emergencial depois do desastre algumas vezes está correlacionado com os padrões de mobilidade normais. Adicionalmente, vários fatores como relações sociais, intensidade de desastre, nível de dano, mídia, fluxo de população, são estudados e, logo, um modelo preditivo é elaborado.

- *Análise de comportamento de grupos.* Os objetos em movimento, especialmente as pessoas e animais, frequentemente forma grupos devido ao seu comportamento social. Por exemplo, a movimentação de uma pessoa não é afetada apenas por atividades pessoais, mas laços sociais com os membros do grupo ao que pertence. Além disso, um padrão de ajuntamento descreve o padrão de movimentação de um grupo de objetos em movimento. Alguns exemplos são comemorações, desfiles, engarrafamento, protestos, etc. O tema de mineração de padrões de ajuntamento ou padrões grupais tem recebido especial atenção da comunidade científica (ZHENG *et al.*, 2013; ZHENG *et al.*, 2014; LI *et al.*, 2013b; GUPTA *et al.*, 2013; MCGUIRE; JANEJA; GANGOPADHYAY, 2014). Um padrão de ajuntamento gerado por um grupo denso e contínuo de objetos em movimento é estudado em (ZHENG *et al.*, 2013). O ajuntamento remove os requerimentos de "membrosia" coerente em padrões grupais tradicionais como comboios e enxames, possibilitando, assim, uma membrosia geral que permite aos objetos em movimento entrar e sair do grupo em qualquer momento.
- *Serviços urbanos.*

O conhecimento descoberto com as técnicas de mineração de trajetórias ajuda a melhorar a qualidade de vida em áreas urbanas em vários aspectos (LI *et al.*, 2015b; LIU *et al.*, 2012; LIEBIG *et al.*, 2012; YANG; FANTINI; JENSEN, 2013; YUAN *et al.*, 2015; YUAN; ZHENG; XIE, 2012). Através da análise de uma grande quantidade de dados de trajetória coletados de veículos eletrônicos, Li *et al.* (2015b) resolve o desafiante problema de implantar estações de carga e pontos de carga, assim, minimizando o tempo médio até a estação de carga mais próximo e o tempo de espera médio para um ponto de carga disponível. A inferência de mapas de redes viárias a partir de percursos GPS a grande escala é uma promissora e atrativa aplicação, dado que construir mapas de estudos geográficos são custosos e pouco frequentes. Desta

forma, [Liu et al. \(2012\)](#) aborda o problema da inferência de mapas de redes viárias em uma configuração prática. Os percursos GPS têm uma resolução e frequência de amostragem muito baixa. Assim, diversas técnicas de inferência de mapas a partir de dados esparsos são investigados e avaliados. Por outra parte, a estimação do volume de tráfego é uma tarefa crucial em muitas aplicações como análise de risco, qualidade de serviço, ranking de locais, etc. Um estudo recente [Liebig et al. \(2012\)](#) tenta estimar o volume de tráfego de pedestres dentro de ambientes fechados. Este conhecimento da presença das pessoas fornece uma valiosa oportunidade para melhorar distintas estruturas como postos de atendimento, lojas, banheiros, largura de de caminhos de corredores em um estádio, etc.

- *Interpretação de trajetórias.* Os dados de trajetórias sem processar, as quais estão em um formato de sequência de localizações geográficas e estampas de tempo, não conseguem ser entendidas pelas pessoas sem uma descrição semântica. Assim, existem muitos trabalhos que facilitam a interpretação dos dados sem processar ([SU et al., 2015](#); [SU et al., 2014](#); [KIM; HAN; YUAN, 2015](#); [LV; CHEN; CHEN, 2012](#); [LU et al., 2011](#)). Diferentemente das trajetórias semânticas que não expressam propriedades de locomoção dos objetos em movimento (excesso de velocidade, paragem), [Su et al. \(2015\)](#) propõe uma abordagem de partição e sumarização que automaticamente gera um breve texto legível pelas pessoas descrevendo a trajetória. A abordagem estende a expressividade das trajetórias semânticas e evita o problema de armazenamento, processamento e transmissão de grandes volumes de trajetórias semânticas. Os dados de uma trajetória são segmentados segundo o comportamento do objeto em movimento e, em seguida, as características de cada segmentação são resumidas em forma de uma breve descrição textual. Adicionalmente, um protótipo de um sistema chamado STMaker ([SU et al., 2014](#)) baseado nesta ideia é desenvolvido.

2.3 Considerações finais

Neste capítulo, apresentaram-se os principais conceitos relacionados com as redes sociais geo-localizadas (LBSN) e de mineração de trajetórias. Foram explicados os tipos de LBSN e como elas são formadas, assim como, a modelagem e estrutura delas segundo a bibliografia respectiva. Adicionalmente foram detalhadas as principais tarefas de mineração de LBSN existentes, orientadas ao entendimento de usuários e locais. Por outro lado, foram dadas as definições mais relevantes na área de mineração de trajetórias. Adicionalmente, apresentou-se um *framework* proposto em [Feng e Zhu \(2016\)](#), que descreve de forma estruturada os diferentes aspectos da área de mineração de trajetórias. Este *framework* leva em conta a geração e coleta de dados geo-localizados, as técnicas de mineração de trajetórias, e, finalmente, as aplicações. Cada um destes componentes é descrito detalhadamente, definindo com clareza os conceitos envolvidos e mostrando os trabalhos do estado de arte para cada caso.

Este capítulo apenas apresentou uma pequena introdução do que é a reconstrução de trajetórias, porém, mostrou como esta tarefa insere-se dentro da área de mineração de trajetórias. No Capítulo 5, esta tarefa é abordada de forma mais aprofundada.

RADRPLUS: UMA REDE SOCIAL GEO-LOCALIZADA BASEADA EM COMUNIDADES

Neste capítulo é apresentado o sistema RadrPlus. Uma rede social geo-localizada com foco nas comunidades. O capítulo começa definindo o sistema, descrevendo seu conceito e estrutura, após isso, são apresentadas a arquitetura tecnológica, a modelagem de dados, uma descrição detalhada das distintas funcionalidades do sistema e, por fim, as considerações finais do capítulo.

A arquitetura tecnológica mostra a forma de interação entre o servidor e os clientes, assim como as tecnologias em cada lado mediante as quais tal interação é possível.

A modelagem, por sua vez, mostra como as tecnologias de persistência de dados são utilizadas e, adicionalmente, detalha as características das estruturas de dados especificamente desenhadas para o desenvolvimento do sistema, assim como o propósito de cada uma delas.

As funcionalidades do sistema são explicadas mostrando as interfaces gráficas e o fluxo de atividade correspondente.

Finalmente, nas considerações finais, incluímos importantes aplicações potenciais e possíveis extensões funcionais do sistema, assim como melhoras técnicas.

3.1 Definição

Nos últimos anos houve um crescimento generalizado de aplicações georeferenciadas (*Global Positioning System* - GPS, Google Maps , Google Latitude , etc), bem como aplicações inovadoras orientadas às redes sociais (Facebook , Twitter , etc). Um outro aspecto recente e importante nesse contexto é o fenômeno chamado *Volunteered Geographic Information* (VGI),

que é a permissão ou disponibilização voluntária de informações geográficas pessoais (SUI; ELWOOD; GOODCHILD, 2012). Tais informações podem ser disponibilizadas automaticamente, por exemplo, pela localização do celular ou IP (*Internet Protocol*) do usuário, via Google APIs, ou diretamente pelo usuário. Nesse cenário, é possível imaginar grupos de amigos ou comunidades que possam ter, em tempo real, seus membros georeferenciados compartilhando essas informações entre si. A disponibilização da informação de localização de pessoas que pertençam a uma rede social ampla ou mesmo a um pequeno grupo, abre um leque enorme de possibilidades de interação, serviços e lazer. Desta forma, potenciais usuários deste tipo de serviços poderiam ser os mesmos usuários de redes sociais que têm interesses em compartilhar informações de localização com membros de sua comunidade. Por exemplo, pode-se imaginar uma comunidade formada por colegas de uma mesma turma em uma universidade. Porém, o principal uso de uma rede social geo-localizada orientada a comunidades poderia se dar em situações de catástrofes (naturais ou não) ou grandes eventos como meio de coordenação e visualização de uma rede de indivíduos em campo. Finalmente, como mencionado no Capítulo 1, a conjunção das informações de localização com as informações da rede, especialmente as de comunidades, possibilitariam um estudo mais aprofundado do comportamento dos usuários e abririam a porta para o estudo de comunidades em redes sociais geo-localizadas. Compreendida a relevância dos dados enriquecidos com informações da rede e levando em conta a falta de bases de dados com estas características, a importância do desenvolvimento do RadrPlus faz-se evidente.

Assim, é nesse contexto onde insere-se o RadrPlus, uma rede social geo-localizada que, diferentemente da maioria das redes online cujo foco é no usuário, tem foco nas comunidades.

Uma comunidade, no contexto do RadrPlus, é definida como um grupo de pessoas que tem interesses/atividades afins num determinado período de tempo ou por tempo indeterminado. O relacionamento pode ser pautado por relações de amizade ou apenas por um período e/ou atividade momentâneos. Toda comunidade tem um *owner* que é quem a cria. Essa pessoa é encarregada de enviar os convites para participação na comunidade, bem como de aceitar os convites dos interessados em se tornar membros da comunidade. Por *default*, os membros são visíveis apenas para os outros membros daquela comunidade. Como toda rede social, ela permite fazer amizade com outros usuários e compartilhar informação pessoal. Ela é geo-localizada, pois coleta e usa as localizações geográficas dos usuários para fornecer funcionalidades interessantes dentro da rede. Tais funcionalidades levam em conta o grupo ou comunidade de usuários e não apenas o indivíduo, diferentemente de outros serviços disponíveis atualmente.

O projeto para um sistema de gerenciamento de uma rede social geo-localizada começou como um Trabalho de Conclusão de Curso (TCC) no ano 2012 desenvolvido pelo aluno de graduação em Engenharia da Computação, Thiago Flora, sob a orientação do Prof.Dr. Alneu de Andrade Lopes. O intuito deste trabalho era desenvolver uma ferramenta de rastreamento em tempo real que mostrasse os deslocamentos de um grupo de usuários. Uma vez atingido tal obje-

tivo, este projeto foi continuado e estendido neste trabalho de mestrado a fim de desenvolver uma rede social geo-localizada orientada a comunidades com funcionalidades que a fizessem atrativa e que permitissem coletar dados de forma segura e eficiente para utilizá-los posteriormente com fins de pesquisa. Um dos primeiros desafios enfrentado foi a adoção da tecnologia de banco de dados não relacional (Cassandra), a qual necessitou uma mudança drástica de paradigma quanto à modelagem de dados e no processo de gerenciamento dos mesmos. O segundo desafio mais relevante foi a implementação de um sistema eficiente de coleta de dados. O GPS é um sistema de localização com uma grande precisão, porém, possui um alto consumo de energia que limita drasticamente o tempo de coleta. Desta forma, implementou-se o *SmartTracking*, um sistema eficiente de coleta de dados, que reduziu o consumo de energia, possibilitando, assim, a coleta dos dados necessários para os posteriores experimentos.

O sistema está disponível tanto para plataformas web quanto móvel. Além das funcionalidades relacionadas ao consumo de serviços do *backend*¹, o aplicativo móvel apresenta um sistema de coleta de dados chamado *SmartTracking*, o qual permite a captura *offline* de dados de geolocalização mediante um sistema de *spooling*. Adicionalmente, o RadrPlus possui um sistema de rastreamento adaptativo que muda a taxa de amostragem de localização de acordo com a velocidade e uma heurística projetada para encontrar um balanço entre economia de energia e a qualidade da trajetória. Tal sistema de rastreamento é explicado detalhadamente no Capítulo 4.

Dessa forma, o RadrPlus permite uma captura eficiente de dados geo-localizados e fornece uma série de funcionalidades geo-sociais baseada neles, dentro de um entorno de rede social focado nas comunidades, acessível tanto desde a web quanto de dispositivos móveis.

3.2 Arquitetura Tecnológica

Como mencionado na introdução, o RadrPlus é uma plataforma que compreende um sistema web e um aplicativo móvel. Assim, os clientes podem acessar ao servidor a partir navegadores web e também de dispositivos móveis que tenham o aplicativo instalado.

Em síntese, o sistema disponibiliza um conjunto de serviços que utilizam bancos de dados como suporte operacional e que são consumidos pelo *frontend*, em forma de páginas web, e pelo aplicativo móvel associado ao sistema. Desta forma, utilizando-se de tecnologias para a criação de serviços, gerenciamento de bancos de dados, *frontend* e aplicativo móvel. Adicionalmente, alguns serviços externos são requeridos como suporte aos próprios.

A Figura 8 sumariza graficamente as tecnologias escolhidas para a arquitetura mencionada.

Nesta figura, pode-se observar que os serviços foram criados utilizando o *framework*

¹ O termo "backend" junto com outros utilizados no texto encontram-se no glossário no final do documento

Figura 8 – Diagrama de implementação do sistema RadrPlus

Fonte: Elaborada pelo autor.

Ruby on Rails (versão 3.2)². Além disso, é importante mencionar que o consumo destes serviços é feito mediante o envio de requisições *REST*.

Por outro lado, observamos que o serviço de persistência de dados utiliza um esquema misto de bancos de dados relacional e não relacional. *PostgreSQL*³ é a tecnologia escolhida para o banco relacional enquanto o *Cassandra*⁴ foi escolhida para o não relacional.

No lado do *frontend*, observamos que ele é desenvolvido usando as tecnologias tradicionais *HTML*, *CSS* e *JavaScript*; porém, devemos detalhar que, buscando acelerar o processo de desenvolvimento do *frontend* e aprimorar a qualidade estética da aplicação, adotou-se o pacote *Bootstrap*, do Twitter. Este pacote traz um conjunto de arquivos *HTML*, *CSS* e *JavaScript* voltados para o desenvolvimento rápido de aplicativos web. São fornecidos controles, formulários, elementos tipográficos e outros componentes pré-estilizados e prontos para uso. Por outro lado, a escolha do servidor web pode ser feita independentemente do sistema, assim o servidor *Puma* é mostrado apenas como sugestão.

² <<http://rubyonrails.org>>

³ <<https://www.postgresql.org>>

⁴ <<http://cassandra.apache.org>>

No lado do aplicativo, a única versão móvel disponível por enquanto é para *Android*⁵. A escolha desta plataforma foi devido principalmente ao fato dela ser atualmente uma das mais populares (GRONLI *et al.*, 2014; LOMAS, 2016; VINCENT, 2017).

Aliás, o aplicativo está compilado para funcionar com versões de *SDK* (*Software Development Kit*) 21 ou superior, mas também possui compatibilidade mínima com *SDK* 11.

Finalmente, o serviço externo *Google Maps* é utilizado para a criação e gerenciamento de mapas no *frontend* e no aplicativo móvel através das *APIs* (*Application Programming Interface*) oferecidas para cada uma destas plataformas.

3.3 Modelagem de dados

Como mencionado na seção anterior, a arquitetura dos dados utiliza tecnologias de bases de dados relacionais e não relacionais. No primeiro caso, utilizou-se *PostgreSQL* como sistema de gerenciamento de banco de dados. Nele, foi apenas criado uma tabela cuja função é armazenar os dados das contas dos usuários. Esta decisão foi tomada devido ao fato que os principais *frameworks* de gerenciamento de contas de usuários para *Ruby on Rails* trabalham apenas com bancos de dados relacionais. Além disso, os dados de contas de usuário são muito reduzidos comparados com os outros dados do sistema e, portanto, não necessitam de um alto desempenho. Assim, um esquema relacional é suficiente.

No segundo caso, utilizou-se o *Cassandra*. Neste banco de dados não relacional, foram armazenados os dados de comunidades, amizades e dados de geolocalização. Dentro do paradigma do *Cassandra*, temos as famílias de colunas, que podem ser entendidas como o equivalente das tabelas nos bancos de dados relacionais. Assim, na Figura 9, observam-se as famílias de colunas e tabela utilizadas na modelagem dos dados.

Note que, as famílias de colunas estão em cinza e a tabela “Members” está em creme. Adicionalmente, é importante mencionar que nas famílias de colunas, o número, nome e tipo de dado das colunas podem ser variáveis, consequentemente, cada instância de uma família de colunas pode ter uma configuração de colunas totalmente distinta. Por isso, a maioria das famílias de colunas na Figura 9 não apresentam atributos. Eles são criados dinamicamente no momento da inserção dos registros. Apenas duas famílias, “Communities” e “UserInfo”, apresentam colunas fixas. Outra informação importante a mencionar é que não há relações entre as famílias de colunas, diferentemente das tabelas. Adicionalmente, no lado do *backend*, o código em comum, relativo à persistência no banco *Cassandra*, das classes correspondentes às famílias de colunas “UserInfo” e “Communities”, foi extraído para uma biblioteca interna, utilizando meta-programação, a fim de reduzir o tamanho do código.

Por fim, em seguida, são descritas brevemente cada uma das famílias de colunas. Como

⁵ <<https://www.android.com>>

Figura 9 – Modelagem de dados do sistema RadrPlus

Fonte: Elaborada pelo autor.

mencionado, a tabela “Members” guarda apenas informação das contas de usuário, logo, não precisa maior explicação.

- **UserInfo.** Armazena dados como o email, nome completo, nome de usuário e outras informações do usuário.
- **Communities.** Armazena dados como nome da comunidade, descrição e tipo de comunidade.
- **UserCommunities.** Representa as comunidades das quais o usuário participa. São divididas em comunidades pessoais (de amigos do usuário) e públicas.
- **RecentActivity.** Representa as atualizações recentes de uma comunidade, permitindo rápida inicialização de sua visualização.
- **LocationUpdates.** Representa todas as atualizações de localização dos participantes de uma comunidade.
- **UserLog.** Registra todas as posições de um usuário, permitindo a reconstrução de sua trajetória.
- **UserFriends.** Registra os amigos de cada usuário.
- **CommDescIndex.** Armazena o índice textual do nome e descrição das comunidades.
- **UserNameIndex.** Armazena o índice textual do nome dos usuários.
- **CommunityUsers.** Representa os usuários pertencentes a cada comunidade. É o contrário da família de colunas “UserCommunities”.

- **Tags.** Armazena dados como nome, mensagem, data de vencimento, localização, criador e outras informações de *tags*. Este conceito é explicado na próxima seção.
- **CommunityTags.** Representa os *tags* pertencentes a cada comunidade. Utilizado para suportar os distintos níveis de visibilidade dos *tags*.

Cada comunidade possui duas famílias de colunas dedicadas ao armazenamento das atividades de seus integrantes. A primeira, “LocationUpdates”, armazena todas as atividades ocorridas em uma comunidade indexadas por tempo. A segunda, “RecentActivity”, armazena as atividades ocorridas em uma comunidade, nos últimos 15 minutos, indexadas por usuário. A expiração de dados é realizada automaticamente pelo Cassandra, por meio da especificação de um TTL (*Time To Live*) no momento de inserção das colunas.

Por outro lado, o envio de novas posições ao sistema pode ser feito individualmente (uma atualização por requisição) ou em conjunto (várias atualizações por vez), mediante um sistema de *spooling*. Este sistema é explicado na seção posterior detalhadamente.

Para a realização de buscas por comunidades, foi necessária a implementação de um índice textual. O índice textual é armazenado na família de colunas “CommDescIndex”, que atua como índice invertido em nível de documento. Analogamente, a busca por usuário foi implementada, utilizando, neste caso, a família de colunas “UserNameIndex”. A operação de busca para ambos os casos é efetuada da maneira usual, utilizando o modelo booleano de recuperação de informações.

3.4 Funcionalidades

Como visto na seção anterior, o RadrPlus apresenta um sistema web e um aplicativo móvel. Além disso, dentro do aplicativo móvel, encontra-se um serviço de rastreamento chamado *SmartTracking*, cuja função é coletar dados de geolocalização dos usuários. Devido a sua complexidade, decidiu-se explicar este serviço no capítulo posterior.

Por fim, a continuação, são apresentados O RadrPlus web e RadrPlus móvel.

3.4.1 RadrPlus web

Nesta subseção, descrevem-se as funcionalidades web do sistema, devidamente acompanhadas das suas respectivas interfaces.

- **Login.** O mecanismo de autenticação fornecido permite *login* pelo modelo convencional (mediante fornecimento de usuário e senha). Para isso, foi adotado o *gem Devise*, que fornece toda a infraestrutura necessária para a autenticação de usuários em um aplicativo *Rails*. Também possibilita a realização de *login* por provedores de terceiros, como

Facebook, Google, Twitter, etc. Por fim, fornece mecanismos de autenticação por meio de *tokens*, funcionalidade adequada para a implementação serviço web responsável por tratar as requisições do aplicativo Android. Como o *gem Devise* requer um *backend* de persistência que seja suportado pelo sistema *ActiveRecord*, foi adicionado um banco de dados relacional – *PostgreSQL* – de forma a possibilitar seu funcionamento. Este SGBD (Sistema de Gerenciamento Banco de Dados) é responsável apenas pela manutenção do cadastro e *login* dos usuários.

- **Perfil de Usuário.** A primeira tela que é apresentada depois do *login* é a tela do perfil. Ela mostra uma barra de tarefas, a imagem do usuário e à direita um painel com informações do usuário. Na barra de tarefas tem-se as opções de ver o perfil, a tela atual, tela de amigos, comunidades, trajetórias e sair. Do outro lado da barra, tem-se dois buscadores, um para usuários e outro para comunidades. No caso do primeiro o valor de busca pode ser qualquer palavra que seja parte do nome real da pessoa (não o *username*) enquanto que para o segundo, tal valor é uma parte da descrição da comunidade. Por fim, no painel pode-se observar as informações de nome de usuário, suas estatísticas, uma lista reduzida de amigos e outra de comunidades, como mostrado na Figura 10.

Figura 10 – Perfil de usuário.

Fonte: Elaborada pelo autor.

Por outro lado, as imagens de exibição dos usuários – conhecidas como avatares – foram implementadas por meio do serviço *Gravatar*⁶. Este serviço provê hospedagem e exibição de avatares gratuitamente, atrelando as imagens de exibição ao *email* dos usuários. São dois os benefícios atingidos com a utilização deste sistema: i) redução do consumo de banda do aplicativo web, pois a transferência das imagens fica terceirizado; ii) os usuários

⁶ <<http://en.gravatar.com>>

finais do sistema podem atualizar seus avatares em um lugar centralizado, uma vez que a alteração afeta todos os sites que fazem uso do *Gravatar*.

- **Lista de Amigos.** A opção Friends mostra uma tela com uma lista dos amigos do usuário, como mostra a Figura 11. Ao acessar ao perfil de um dos nossos amigos, a tela mostrada será a mesma da Figura 10 exceto por um botão que aparecerá embaixo da imagem do usuário. Tal botão permite iniciar ou desfazer a amizade com tal usuário.

É possível encontrar um amigo no RadrPlus através do buscador de usuários, para o qual precisamos pelo menos uma parte do nome da pessoa.

Figura 11 – Lista de amigos.

Fonte: Elaborada pelo autor.

- **Lista de Comunidades.** Um caso parecido ao de Friends é o de Communities. Esta opção mostra uma tela com a lista de comunidades as quais o usuário pertence, mas o mais importante é que nessa tela encontra-se a opção de criar comunidades. O único requisito necessário para criar uma comunidade é colocar um nome e uma descrição. A Figura 12 mostra a lista de comunidades de um usuário. Como no caso da busca de usuários, as comunidades possuem um campo de pesquisa próprio.

Figura 12 – Lista de comunidades.

Fonte: Elaborada pelo autor.

- **Atualizações de Comunidade.** Uma vez selecionada uma comunidade através do buscador ou de nossa própria lista, uma tela com um mapa e uma tabela à direita aparecerá. Nesse mapa, são mostradas em forma de marcadores as atualizações dos usuários membros

da comunidade. Essas atualizações são em tempo real, assim as posições dos marcadores no mapa acompanham as mudanças de localização. Na Figura 13 observa-se a tela mencionada com as localizações de dois usuários membros da comunidade.

Figura 13 – Página de uma comunidade.

Fonte: Elaborada pelo autor.

O acompanhamento da localização dos membros de comunidade é realizado em forma de *polling*, por meio de requisições AJAX (*Asynchronous JavaScript and XML*). A posição de cada usuário é apresentada sobre um *widget* do *Google Maps*.

O código responsável pelo funcionamento do mapa, do lado do cliente, foi desenvolvido em *CoffeeScript* – compilado para *Javascript* pelo *asset pipeline* do *Rails*. Além disso, o código foi implementado como um *plugin* da biblioteca *jQuery*.

A inicialização da visualização do mapa é realizada consultando-se a atividade recente dos integrantes da comunidade. Estes dados são retornados em formato *JSON* e são utilizados para a criação de marcadores no mapa.

- **Trajeto**rias. A opção Trajectories apresenta uma tela simples que permite mostrar o percurso que o usuário fez, dados uma data inicial e final como entradas. O caminho percorrido é mostrado com marcadores no lugar de cada atualização e uma linha que os une. Pode-se observar tal tela na Figura 14.

Figura 14 – Visualização de trajetória

Fonte: Elaborada pelo autor.

3.4.2 RadrPlus Móvel

Nesta subseção, veremos as funcionalidades presentes no aplicativo do RadrPlus para Android, seu funcionamento e as interfaces correspondentes.

- **Tela Principal.** Após entrar em sua conta no aplicativo RadrPlus você verá a tela principal do aplicativo. Essa tela possui um campo de buscas na parte superior e 3 abas logo abaixo, como mostra a Figura 15.

Figura 15 – Tela principal.

Fonte: Elaborada pelo autor.

As Figuras 16 (a), 16 (b), 16 (c), mostram o conteúdo de cada aba citada anteriormente.

Na aba de amigos, Figura 16 (a), é exibida uma lista com todos os seus amigos. Na aba de comunidades, Figura 16 (b), é exibida uma lista com todas as comunidades das quais você faz parte e na aba de menu, mostrada na Figura 16 (c), são listadas algumas opções de usabilidade, como ver o histórico de localização e criar uma nova comunidade. Todas essas opções serão detalhadas mais adiante.

- **Buscas.** Como vimos na tela principal, o campo de busca encontra-se na parte superior da tela principal do aplicativo. Na Figura 17 é mostrado o campo de buscas para usuários,

Figura 16 – Abas da tela principal

(a) Aba de amigos

(b) Aba de comunidades

(c) Menu

Fonte: Elaborada pelo autor.

como a busca pode ser tanto por usuário ou comunidade é necessário escolher o tipo de busca que será realizada.

Figura 17 – Campo de buscas.

Fonte: Elaborada pelo autor.

Como podemos ver no lado direito da barra de buscas, existe um pequeno ícone que é usado para a seleção do tipo de busca que desejamos realizar, por padrão veremos um único avatar indicando que a busca a ser realizada será por usuários, tocando sobre esse avatar veremos a figura mudar para um par de avatares, indicando que a busca a partir de agora será por comunidades. Tocando sobre esse botão troca-se o critério da busca (usuários/comunidades).

Buscar usuários.

Buscar comunidades.

Além de selecionar o tipo de busca é necessário inserir uma palavra chave que será usada para encontrar a pessoa ou comunidade desejada, essa palavra chave pode ser o nome ou parte do nome do que desejamos encontrar. Para inserir a palavra chave toque sobre o ícone de uma lupa, localizado no lado direito da barra de buscas, e o campo de texto abrirá automaticamente. Digite a palavra chave e confirme, logo abaixo do campo de buscas aparecerá uma lista com todos os resultados para a busca realizada.

Inserir palavra chave.

O código correspondente aos serviços de busca por usuários e comunidade foram apresentado na última parte da seção de Modelagem de dados.

- **Perfil de usuário.**

Botão “Profile”: Pressione para visualizar o seu perfil de usuário.

Figura 18 – Perfil de usuário.

Fonte: Elaborada pelo autor.

Uma vez pressionada a opção “Profile” na aba de menu, é apresentada a tela mostrada na Figura 18.

Nesta tela podemos apreciar a imagem do usuário ou avatar que também é nosso marcador personalizado na hora de mostrar nossa localização sobre o mapa. Adicionalmente, são mostradas as estatísticas de usuário, lista de amigos e comunidades como no site web. Cabe mencionar que tais listas são listas deslizantes, assim você pode passar o dedo de baixo para cima para ver os demais itens (se houverem). Nessas listas serão mostrados no máximo cinco itens, mas o último item em cada lista é a opção “View all” que ao ser pressionado mostra uma tela inteira com todos os itens da lista.

A imagem do avatar do usuário é na verdade mais que isso, é um botão gráfico que permite acessar a tela de edição de perfil apresentada em seguida.

- **Edição de Perfil.**

Pressione seu avatar na tela de perfil para editar.

Figura 19 – Perfil de usuário.

Fonte: Elaborada pelo autor.

A Figura 19 mostra a tela de edição de perfil, onde podemos escolher uma imagem para usar como avatar e também o desenho de nossa “pegada”.

Uma funcionalidade que será apresentada posteriormente é o Mapa de atualização de comunidades. Tal funcionalidade é análoga àquela no site web. As mudanças de localização

são mostradas no mapa em tempo real, assim uma pegada é uma anterior localização do usuário representada no mapa com o desenho de pegada escolhido. Dessa forma é possível ter uma trilha composta de todas as pegadas; porém, para evitarmos lotar o mapa com pegadas de todos os membros da comunidade, o “Trail size” ou tamanho de trilha irá definir o máximo número de pegadas a serem mostradas. Assim, se tal número for “n”, apenas serão mostradas as últimas “n” pegadas.

Pressione o avatar na tela de edição para selecionar um novo.

Pressione a pegada na tela de edição para selecionar uma nova.

Após selecionar o avatar e a pegada desejada basta clicar sobre “Save” para salvar as alterações ou “Discard” para descartar as mudanças feitas.

- **Histórico de localizações.**

Botão “Location history”: Histórico de localização.

A ideia de histórico de localizações é poder visualizar o percurso feito pelo usuário em um período de tempo dado. Assim, as únicas informações necessárias são as datas e horas inicial e final. A tela correspondente é mostrada na Figura 20 (a). Uma vez informados os valores das datas inicial e final, pressionamos o botão “Run” e uma nova tela com um mapa aparecerá, Figura 20 (b). Nessa tela, como no caso do site web, são mostradas todas as localizações do usuário representadas como marcadores. O marcador verde indica o primeiro ponto (localização) e o vermelho o ponto final. Aliás, cada marcador mostra ao ser pressionado a data e hora em que o usuário estava naquele ponto.

- **Amigos.** O perfil dos outros usuários é idêntico ao nosso exceto por o botão que permite iniciar ou desfazer nossa amizade com ele. Tal detalhe pode ser apreciado na Figura 21. O botão “Add as friend” com ícone de pessoa e símbolo “mais” significa que o usuário ainda não é nosso amigo e podemos adicioná-lo a nossa lista de amizade. Por outro lado, se aparecer o botão “Unfriend” com apenas o ícone de pessoa, isso quer dizer que já é nosso amigo e pressionando o botão iríamos desfazer nossa amizade.

Figura 20 – Visualização de histórico de localização

(a) Seleção de intervalo

(b) Histórico de localizações

Fonte: Elaborada pelo autor.

Figura 21 – Comparação de perfil próprio com o de um amigo

(a) Amigo

(b) Usuário

Fonte: Elaborada pelo autor.

- **Solicitações de amizade.** Na tela principal podemos ver a opção “Friends request” no menu, ao tocar sobre essa opção é possível visualizar todas as solicitações de amizade

aguardando aprovação.

Essa opção também garante maior privacidade, pois o usuário só tem acesso às trajetórias de membros de suas comunidades. E para convidar alguém para sua comunidade ou pertencer à comunidade de alguém, é necessário serem amigos.

Botão “Friend request”: Solicitações de amizade.

- **Nova comunidade.** As comunidades são grupos de pessoas que possuem alguma relação ou interesse em comum, por exemplo, você pode criar uma comunidade chamada “Família” e adicionar seus familiares a ela, ou entrar em uma comunidade de fotógrafos que une pessoas que se interessam por fotografia.

Botão “New community”: Permite criar uma nova comunidade.

Após pressionar o botão “New community” na aba de menu, será exibida a tela de criação de comunidade. No campo “Name” digite um nome apropriado para a comunidade e no campo “Description” faça uma breve descrição da mesma. Em seguida pressione o botão “Save community” para finalizar a criação da comunidade. Após esse procedimento é possível ver sua nova comunidade na aba de comunidades, também é possível adicionar membros e visualizar o mapa da mesma.

- **Mapa de comunidades.**

A funcionalidade mapa de comunidades pode ser acessada pressionando sobre um item da lista de comunidades ou utilizando a busca de comunidades. A tela correspondente é mostrada na Figura 22. A imagem escolhida pelo usuário no seu perfil é mostrada como marcador, enquanto o desenho de pegada é utilizado com marcador das últimas localizações. Ao pressionar sobre uma pegada ou marcador, aparecerá uma janela mostrando a data e hora correspondente, assim como o nome do usuário. Adicionalmente, se pressionarmos sobre essa janela, iremos visitar o perfil do usuário. Observa-se, também, que o nome da comunidade é mostrado na barra de tarefas junto com um ícone de pontos suspensivos que ao ser pressionado irá mostrar uma série de opções, dependendo se somos proprietários ou apenas membros da comunidade.

A Figura 23, mostra as opções do proprietário da comunidade. O proprietário pode adicionar amigos (mandar convites), eliminar membros, visualizar um *timelapse* e deixar a comunidade (eliminar a comunidade). O conceito de *timelapse* é explicado mais adiante.

Figura 22 – Mapa da comunidade.

Fonte: Elaborada pelo autor.

Figura 23 – Menu do proprietário da comunidade

Fonte: Elaborada pelo autor.

Figura 24 – Menu de um membro da comunidade

Fonte: Elaborada pelo autor.

Por outro lado, na Figura 24, pode-se observar as opções de um membro de uma comunidade, quem pode apenas visualizar o *timelapse* ou deixar a comunidade.

- **Notícias de comunidade.** Na tela principal podemos ver a opção “Community news” no menu, ao tocar sobre essa opção é possível visualizar todas as solicitações de outros usuários pedindo pertencer a uma de nossas comunidades, assim também como convites para pertencermos às comunidades deles.

Botão “Community news”: Notícias de comunidade.

- **Timelapse.** A ideia de *Timelapse* é a de mostrar uma animação de todos os usuários membros de uma comunidade se locomovendo no mapa ao mesmo tempo. Importantes padrões podem ser apreciados dessa forma, por exemplo, saber se em certo período de tempo a comunidade teve muita atividade, que lugares foram os mais visitados, que caminhos os mais transitados, que usuários se movimentam mais e quais ficam quase sempre nos mesmos lugares. Para isso é preciso indicar o período de análise fornecendo a data inicial e final. Uma vez definido o período, este é dividido em faixas de igual tamanho ou também chamados “frames”. O tamanho do *frame* define a quantidade total deles no período, então, quanto maior tamanho menor quantidade. Dado que o tamanho de

um *frame* pode ser uma quantidade de segundos, dias ou semanas e as atualizações de localização podem ser a cada dez segundos, cada *frame* pode conter muitas atualizações de localização; porém, também cabe a possibilidade que não contenha nenhuma, devido a que em tal período de tempo não houveram atualizações. No caso de haver muitas atualizações por *frame*, para dar ao usuário a flexibilidade de ter tanto uma visão geral como em detalhe da informação, cada *frame* leva em conta a última atualização de localização de cada usuário para posicionar o marcador correspondente no mapa, enquanto o resto de atualizações anteriores são mostrados como uma trajetória multilinear de uma cor distinta para cada um deles.

Desse jeito, o *frame* torna-se uma foto da comunidade com todos seus membros em um dado instante de tempo. Quanto menor tamanho de *frame* maior detalhe (mais *frames* e mais atualizações) e vice-versa. Por fim, outro parâmetro a ser definido é a velocidade na qual o *frame* é mostrado no mapa. Para isso, só é preciso definir a unidade de tempo, pois o valor unitário dele será o tempo usado por cada *frame*.

Figura 25 – Visualização do Timelapse

(a) Configuração do timelapse

(b) Timelapse

Fonte: Elaborada pelo autor.

A Figura 25 (a) mostra a tela com todos os parâmetros antes mencionados. Uma vez pressionado o botão “Run”, outra tela com um mapa aparecerá e a animação começará a ser executada. Nessa tela, temos botões para repetir a animação ou para-la. Além disso, no canto superior direito pode-se observar o avanço da animação que não é outra coisa que a porcentagem de frames que já foram mostrados. A Figura 25 (b) ilustra esta última ideia.

- **Tags.** Em uma comunidade, cada membro pode criar e eliminar *tags* em qualquer parte do mapa. Um *tag* contém informações de algum evento, lugar ou qualquer coisa que o usuário ache relevante (e.g., problemas urbanos, tais com buracos no pavimento, foco de dengue, risco de assalto, entre outros). Um exemplo de *tag* pode ser visto na Figura 26.

Cada *tag* pode ser visualizado dentro da comunidade em que foi criado ou em todas elas dependendo da sua configuração de privacidade.

Figura 26 – Tag usado para indicar um acontecimento.

Fonte: Elaborada pelo autor.

- **Tracking.**

Botão “Turn on tracking”: Inicia o serviço de rastreamento.

Botão “Turn off tracking”: Finaliza o serviço de rastreamento.

Para garantir a segurança e privacidade dos seus usuários, o aplicativo do RadrPlus permite que os mesmos escolham quando o aplicativo pode coletar dados de localização. Para iniciar o serviço de rastreamento do RadrPlus, basta ir a aba de menu e pressionar a opção “Turn on tracking”. Ao fazer isso o serviço inicia a coleta de dados de localização. Para finalizá-lo, basta pressionar o mesmo botão, que enquanto o serviço estiver rodando exibirá a mensagem “Turn off tracking”.

Ao iniciar o serviço de rastreamento, o RadrPlus passa a coletar dados de localização e enviá-los ao servidor. Se não houver conexão enquanto o serviço estiver rodando, os dados

são guardados na memória do aparelho e enviados ao servidor assim que uma conexão for estabelecida. Esta funcionalidade junto com muitas outras formam um serviço de rastreamento chamado *SmartTracking*, explicado detalhadamente no capítulo posterior.

3.5 Considerações finais

Neste capítulo apresentou-se o sistema RadrPlus, uma rede social geo-localizada com foco nas comunidades. Este sistema fornece muitas funcionalidades novas e atrativas que a diferenciam de outras redes sociais. Consequentemente, este sistema tem um grande potencial de uso tanto para o desenvolvimento de novas funcionalidades úteis no contexto social quanto como para pesquisas relacionadas às áreas de mineração de trajetórias, visualização de dados e mineração de redes sociais geo-localizadas. Assim também, neste capítulo, foram apresentados a arquitetura geral do sistema, as tecnologias utilizadas no seu desenvolvimento e o modelo de dados. Adicionalmente, as funcionalidades do sistema junto com as suas interfaces gráficas e correspondentes fluxos de atividade foram descritos e explicados detalhadamente.

O RadrPlus permite gerenciar as contas de usuários e perfis, fazer e desfazer amizades com outros usuários, realizar buscas de usuários e comunidades, criar comunidades e formar parte de comunidades de terceiros, assim como enviar convites para outros pertencerem às nossas comunidades ou rejeitar convites de outros. Adicionalmente, o RadrPlus permite a visualização das localizações dos membros de uma comunidade em tempo real e também em um período determinado mediante uma animação do tipo *time-lapse*; além disso, permite visualizar a trajetória individual de um usuário. O RadrPlus permite, também, a criação de *Tags*, que disponibilizam em um mapa algum evento ou lugar interessante junto com informação pertinente e, finalmente, fornece um método robusto de rastreamento que funciona tanto *online* como *offline* e que economiza espaço e energia adaptando a taxa de amostragem à velocidade do usuário.

Por outro lado, o RadrPlus pavimenta o caminho para o desenvolvimento de funcionalidades de exploração de rede social mais avançadas como :

- Criação de *check-ins*;
- recomendação de amizades, comunidades e locais;
- prover alertas de amigos próximos.

Adicionalmente, uma outra aplicação potencial do RadrPlus poderia ser visualizar e coordenar pessoas em trabalhos conjuntos, onde a localização de seus integrantes é importante. Por exemplo: mutirões de resgate em caso de desastres naturais, coordenação de unidades policiais ou corpo de bombeiros para o atendimento de emergências, acompanhamento de colheitas e atividades rurais, etc. Assim, as principais funcionalidades futuras a serem implementadas seriam:

- Acompanhar a posição de todas as unidades disponíveis (ou aquelas que atendam alguns requisitos) em tempo real num mapa;
- encontrar a unidade mais próxima a uma emergência;
- possibilitar a comunicação entre estas unidades por meio de mensagens.

A questão relativa a "que atendam alguns requisitos" poderia envolver, por exemplo, pessoas com capacitação médica (relevante em desastres naturais), unidades do corpo de bombeiros com capacidade de atender determinados tipos de emergências (incêndios, resgates etc), unidades fortemente blindadas de polícia, etc.

Este foco também pode ser estendido e incorporar aspectos sociais. Por exemplo: convidados de um evento podem divulgar suas posições pela duração do mesmo e trocar mensagens, permitindo a coordenação de caronas e evitando perguntas como "onde você está", "por que está demorando", etc. Pode-se, inclusive, integrar a ferramenta a outras ferramentas de redes sociais, como o Facebook – que já possui eventos, com as informações de localização e horário.

Finalmente, como melhoras técnicas futuras, planeja-se a implementação do modelo de programação *Map Reduce* no sistema, mediante a integração do *Cassandra* com a plataforma *Hadoop*, a fim de melhorar o desempenho do sistema e garantir uma escalabilidade ainda maior. Outra tarefa futura seria o desenvolvimento de versões móveis do sistema para *iOS* e *Windows Phone*.

COLETA DE DADOS

O presente capítulo é iniciado descrevendo o processo de coleta efetuado para a obtenção da base de dados geo-localizados que será utilizada nos experimentos de reconstrução de trajetórias. Posteriormente, as características da base de dados obtida são descritas detalhadamente juntamente com o formato dos dados. Dessa forma, a própria base produzida, e disponibilizada posteriormente, já se constitui em uma contribuição deste mestrado.

Finalmente, a fim de avaliar o desempenho do sistema de rastreamento, são comparadas as trajetórias obtidas mediante o algoritmo adaptativo do *SmartTracking* com as obtidas através de um método simples de captura com taxa de amostragem fixa chamado de esquema contínuo.

4.1 Processo de coleta

Nesta seção, são descritos primeiramente alguns dos métodos e abordagens mais relevantes quanto à captura eficiente de dados geo-localizados, em seguida, é apresentado detalhadamente o sistema de rastreamento do RadrPlus, chamado *SmartTracking*, e, finalmente, é descrita a metodologia utilizada para executar a coleta de dados de geolocalização.

4.1.1 Métodos eficientes de coleta de dados geo-localizados

Devido ao cada vez maior número de serviços de geolocalização oferecidos para dispositivos móveis, muitos dos problemas relacionados com estes serviços como a gestão da mobilidade, qualidade de serviço, eficiência energética, segurança e privacidade, têm-se tornado temas relevantes de estudo. Um dos problemas mais críticos é a eficiência energética em dispositivos móveis. Dado que a indústria de fabricação de baterias avança lentamente quanto à capacidade energética da bateria (apenas 5% anualmente (ROBINSON, 2009)), e a demanda de computação e armazenamento aumenta rapidamente, prover uma melhor experiência de usuário com uma capacidade energética limitada da bateria tem-se tornado um assunto urgente nestes

últimos anos.

O GPS é o mecanismo de localização mais preciso, porém, mais custoso em termos de energia dispendida. Portanto, a granularidade das atualizações ou taxa de amostragem na qual são geradas as atualizações de localização tem grande influência na duração da bateria (ZHUANG; KIM; SINGH, 2010).

Dessa forma, o desenvolvimento de métodos de captura eficiente de localização tem atraído a atenção da comunidade científica e, como resultado, uma série de abordagens e métodos têm sido propostos (WANG *et al.*, 2014).

Uma destas abordagens é simplesmente a substituição do GPS (*GPS Substitution*) por outros meios menos custosos energeticamente como os *cell-ids* do GSM ou redes UMTS (*Universal Mobile Telecommunications System*), Wifi, ou até as balizas de *bluetooth* para a obtenção da localização (FENG; ZHU, 2016). Por outro lado, apesar que a substituição poderia ser útil dentro de edifícios, onde o GPS fica inutilizado, a baixa precisão de localização destes métodos constitui uma desvantagem muito grande. Uma outra abordagem é a geo-localização híbrida (*Hybrid Positioning*). Os métodos desta classe de otimização energética tentam reduzir o consumo de energia utilizando apenas o meio de localização mais energeticamente eficiente sempre que a sua precisão seja suficiente (BARETH, 2012). Um exemplo deste tipo de técnica poder ser encontrado em Bareth e Kupper (2011). Esse trabalho desenvolve um algoritmo que dinamicamente desativa as diferentes tecnologias de localização e apenas ativa a tecnologia com o menor consumo energético e que forneça uma precisão suficiente para determinar corretamente relações topológicas.

Adicionalmente, uma das abordagens de otimização energética mais utilizadas no contexto de geo-localização em dispositivos móveis é a redução do uso do GPS ou mais conhecido em Inglês como *Duty Cycling*. Este tipo de método desativa temporariamente o GPS e o substitui por outro método de localização. Comumente, este tipo de método é dividido em *Static Duty Cycle* (SDC) e *Dynamic Duty Cycle* (DDC). O SDC ativa e desativa o GPS em intervalos regulares de tempo, enquanto o DDC regula os ciclos de leitura baseado em modelos analíticos ou eventos gerados por outros sensores (WU *et al.*, 2011). O SDC opera de maneira cega, sem considerar nenhum tipo de informação extra que possa melhorar o *trade-off* entre economia de energia e precisão. Por outro lado, os métodos do tipo DDC, levam em conta distintas informações adicionais, algumas mais complexas que outras. Por exemplo, em Lu *et al.* (2010), descreve-se um algoritmo que leva em conta os padrões de mobilidade e comportamento dos usuários para regular os ciclos de leitura. Para isso, cria um modelo de cadeias de Markov utilizando os dados de geo-localização históricos. Um exemplo mais simples pode ser encontrado em Wu *et al.* (2011), onde descreve-se um modelo adaptativo baseado em DDC, o ADC (*Adaptive Duty Cycle*), dentro do qual os ciclos de leituras são regulados levando em conta sua correlação espacial e temporal com a recepção do sinal GPS.

Por fim, a heurística do modelo de captura proposto no *SmartTracking* é um tipo de ADC

baseado em dois critérios não utilizados previamente: o tamanho de quarteirão e a velocidade média dos distintos meios de locomoção.

Assim, com este modelo simples (pois não utiliza modelos analíticos), espera-se testar a efetividade do uso dos dois critérios mencionados anteriormente em uma abordagem do tipo *Duty Cycling*, para uma coleta eficiente de dados de localização em dispositivos móveis.

4.1.2 SmartTracking

Devido à frequente perda de conexão com a Internet nos dispositivos móveis, faz-se necessário um mecanismo assíncrono de coleta de dados de localização para evitar perder dados enquanto um usuário está se movimentando. Assim, RadrPlus tem um mecanismo de *spooling* embutido no aplicativo móvel que armazena os dados gerados pelo GPS enquanto não há conexão à Internet e, assim que detectar uma conexão válida, envia todos os dados guardados ao servidor. Este mecanismo é mostrado na Figura 27.

Figura 27 – Mecanismo de spooling do sistema RadrPlus

Fonte: Elaborada pelo autor.

Por outro lado, o sistema também tem implementado um seguimento ou *tracking* adaptativo, que graças aos sensores acelerômetro e bússola, como mostrado na Figura 28, muda a taxa de amostragem de localização de acordo com a velocidade e uma heurística projetada para encontrar um compromisso entre economia de energia e a qualidade da trajetória. Dessa maneira, RadrPlus reduz o tempo de uso do GPS e economiza energia e espaço de armazenamento no sistema ao gerar menos atualizações. A velocidade é calculada a cada três atualizações de localização, levando em conta a distância percorrida e o tempo transcorrido desde a primeira até a terceira atualização.

Na Figura 29, observa-se o critério da mudança de intervalo de amostragem baseado na velocidade. Para exemplificar este conceito, se dissermos que temos um intervalo de leitura de 600s significa que será realizada uma leitura do GPS a cada 600 segundos. O serviço de rastreamento transita entre distintos estados, os quais, em alguns casos, têm o mesmo valor de intervalo. Além disso, para atingir um certo estado, tem-se que passar necessariamente pelos estados anteriores. O intervalo inicial de amostragem (no momento de ligar o serviço) é de 10

Figura 28 – Troca dinâmica de sensores segundo a velocidade

Fonte: Elaborada pelo autor.

segundos. À medida que a velocidade muda, o valor do intervalo vai se deslocando para estados com um valor menor. Tomando como referência a Figura 29, o estado atual do serviço irá se deslocando à direita se a velocidade aumentar ou à esquerda se ela diminuir. Dessa maneira, o intervalo mínimo será sempre o inicial.

Figura 29 – Mudança do intervalo de amostragem segundo a velocidade de locomoção

Fonte: Elaborada pelo autor.

A ideia subjacente deste procedimento é que um usuário que está se locomovendo lentamente não vai gerar mudanças significativas de localização em períodos curtos de tempo, então seria conveniente reduzir o intervalo de amostragem. Adicionalmente, se ele apresentar uma velocidade muito pequena ou um período de inatividade prolongado, pode significar que ele está parado e, muito possivelmente, irá permanecer nesse estado por um tempo, assim, o serviço adota o intervalo de amostragem máximo de leitura do GPS, 10 minutos, e começa a ler os sinais do acelerômetro. Assim que detectar que o usuário voltou a se locomover, o serviço parará as leituras do acelerômetro e voltará a ler o GPS com a taxa do estado inicial.

Por outro lado, se o usuário se locomover muito rápido, presumivelmente de carro ou ônibus, é muito provável que ele esteja seguindo uma trajetória reta, então não é preciso guardar suas localizações em períodos curtos de tempo. Além disso, como a maioria das interseções de

ruas são quase perpendiculares, o usuário deve diminuir a velocidade para dobrar e tomar outra rua. É nesse momento que a taxa deveria aumentar. Para isso, quando o usuário for muito rápido, o serviço adota o intervalo de amostragem máximo e começa a ler a bússola. Quando a bússola detectar que o usuário mudou de direção, um giro perto de 90 graus, parará a leitura da bússola e voltará a ler o GPS com o intervalo de amostragem do estado inicial. A Figura 30 ilustra as ideias comentadas nestes dois últimos parágrafos.

Figura 30 – Diagrama de estados da mudança de intervalo de amostragem. Cada estado tem alocado o nome do meio de transporte mais provável segundo a velocidade, e o intervalo correspondente. As transições têm alocadas os valores de velocidade necessários para mudar de estado e, consequentemente de intervalo de amostragem. As transições sem valor indicam que se a velocidade for mantida no estado atual, haverá uma mudança sem necessidade de incremento de velocidade. Uma vez atingidos os estados de "Parado" e "Muito Rápido", a única forma de mudar de estado é por meio de um evento detectado pelos sensores acelerômetro e bússola respectivamente.

Fonte: Elaborada pelo autor.

Além dos critérios explicados anteriormente respeito da redução e incremento dos intervalos de amostragem segundo a velocidade, os quais explicam o comportamento parabólico da Figura 29, um outro assunto importante a ser explicado é o critério de seleção dos valores dos limiares de velocidade e os intervalos de amostragem. Para isso, teve-se em consideração os dois aspectos mencionados na subseção anterior: o tamanho de quarteirão e a velocidade média dos distintos meios de locomoção.

Considerando o primeiro critério, em algumas cidades dos Estados Unidos, recomenda-se um tamanho de quarteirão quadrado (*city block*) entre 330' (100.854m) e 550' (167.64m) (HARRIS; AECOM; HINTZ, 2008). Adicionalmente, no Reino Unido, a regra do polegar para definir o tamanho de quarteirão é que um tamanho entre 80 e 90 metros fornece um bom *trade-off*

entre os diferentes critérios urbanísticos correspondentes em diversas localidades urbanas e sob variadas circunstâncias (YEANG *et al.*, 2000). Contudo, em geral, na Europa e nos Estados Unidos, um tamanho de quarteirão de 80 a 100 metros fornece uma rede viária ótima para veículos motorizados, pedestres e ciclistas (CENTER *et al.*, 2012). Dessa forma, finalmente, adotou-se um tamanho de quarteirão de 100 metros para a elaboração da heurística.

A respeito do segundo critério, a velocidade de caminhada mais comumente escolhida pelas pessoas é aproximadamente 1.4 m/s (MOHLER *et al.*, 2007; BROWNING *et al.*, 2006; LEVINE; NORENZAYAN, 1999). Apesar que muitas pessoas são capazes de caminhar a velocidades maiores de 2.5 m/s , especialmente em distâncias curtas, elas comumente preferem não fazê-lo. Velocidades mais rápidas ou mais lentas são consideradas desconfortáveis pela maioria das pessoas. (MINETTI, 2000).

Por outro lado, a velocidade média de um carro durante seu trajeto nas ruas varia de 18.35 a 32.23 km/h (5.09 a 8.95 m/s), enquanto a velocidade média nas rodovias pode variar de 50 a 120 km/h (13.88 a 33.33 m/s) (MONTAZERI-GH; NAGHIZADEH, 2003; VÁRHELYI; MÄKINEN, 2001). Contudo, é possível atingir velocidades ainda maiores em algumas rodovias. Por exemplo, em alguns estados dos Estados Unidos, a velocidade pode variar entre 121 a 129 km/h (33.6 a 35.8 m/s) (INSTITUTE, 2017). Porém, poderia ser possível superar os 129 km/h devido à falta de controle dos limites de velocidade nas zonas rurais. Finalmente, a velocidade média de um ciclista calcula-se em 14.5 km/h (4.02 m/s) (JENSEN *et al.*, 2010).

Por fim, levando em conta todas essas informações, elaboramos o Quadro 1, o qual mostra as velocidades médias adotadas para cada um dos meios de locomoção terrestre mais comuns.

Quadro 1 – Velocidade média de meios de locomoção

Meio de locomoção	Velocidade média (m/s)
Pedestre	1.40
Bicicleta	4.02
Carro em ruas (cidade)	5.09 - 8.95
Carro em avenidas (cidade)	13.88 - 33.33
Carro em rodovias	33.6 - 35.8

Utilizando o Quadro 1 e assumindo um tamanho de quarteirão de 100 metros, projetamos os valores dos limiares de velocidade e os intervalos de amostragem.

Para definir os limiares de velocidade, primeiro, arredondamos para a cima o valor da velocidade média de pedestre até o inteiro mais próximo e calculamos a diferença entre ambos os valores. Depois disso, utilizamos essa diferença para definir o limite inferior do intervalo, restando tal diferença do valor da velocidade média de pedestre. Dessa forma, obtemos o primeiro intervalo (0.8-2).

O segundo intervalo é calculado tomando o limite superior do anterior intervalo e

alocando como limite superior o menor valor da velocidade média de carro na rua. Assim, obtemos o segundo intervalo (2-5).

Os três últimos intervalos, são definidos calculando a diferença entre o maior valor da velocidade média de carro na rodovia e o menor valor da velocidade média de carro na rua. Uma vez feito isso, tal diferença é dividida entre três e o quociente arredondando é utilizado como largura para a criação dos três últimos intervalos: 5-15, 15-25, 25-35. O intuito desta divisão é tentar distribuir equitativamente as possibilidades de cada intervalo conter os valores do espectro definido pelas velocidades médias de carro em diferentes contextos. Este procedimento diferencia-se do efetuado na criação dos dois primeiros intervalos, devido ao maior margem de valores existente nas velocidades de carro quando comparado com as velocidades de pedestre e bicicleta, as quais são muito próximas entre si.

Cabe mencionar que uma velocidade menor do limite inferior do primeiro intervalo (0.8) indica que o usuário fez uma parada, portanto, conseqüentemente, o GPS é desligado e o sistema começa a ler o acelerômetro. Aliás, uma velocidade maior do limite superior do último intervalo (35), fará, também, que o GPS seja desligado e, em seguida, o sistema começará a ler a bússola.

Por outro lado, para definir os intervalos de amostragem, o critério geral foi reduzir ao máximo o número de leituras que garantisse uma boa qualidade de reconstrução de trajetória. Para cumprir com tal objetivo, escolhemos um intervalo de amostragem que, segundo a velocidade do usuário em cada caso, permitisse realizar uma leitura no final de cada quarteirão. Desta forma, dado que o tamanho de quarteirão escolhido foi de 100 metros, a intervalo de amostragem para cada um dos intervalos de velocidade anteriormente definidos foi calculada de tal forma que uma leitura tenha sido feita aproximadamente a cada 100 metros nos casos em que o usuário está em velocidade compatível com ruas de centros urbanos.

Assim, no caso do segundo intervalo, levando em conta a velocidade média de ciclista de 4.02 m/s , calculou-se uma taxa de 25 segundos. De forma similar, no terceiro intervalo (5-15), tomou-se o valor na metade do intervalo como velocidade média para fazer o cálculo do intervalo de amostragem. O valor do intervalo resultante foi de 10 segundos.

Não obstante, diferentemente dos casos anteriores, no cálculo do intervalo de amostragem no primeiro intervalo de velocidade, seguiu-se um procedimento distinto. Sob a hipótese que o pedestre é mais suscetível de realizar paradas e entrar em alguns lugares durante seu percurso, devido principalmente a sua menor velocidade e maior liberdade de locomoção, detectou-se um maior risco de perda de sinal do *GPS* nos trajetos de pedestres. Tal risco é devido ao fato que o sinal de *GPS* precisa de uma vista desobstruída do céu para gerar uma leitura com uma qualidade suficientemente boa. Desta forma, com intuito de diminuir a possibilidade de perder a localização de um usuário dentro de um quarteirão devido à perda do sinal, optou-se por realizar a leitura do *GPS* antes do final do quarteirão. Porém, para isso, foi importante levar em conta que se tal leitura fosse realizada na metade do quarteirão ou antes, teríamos três leituras ou mais por quarteirão, o que infringiria o critério geral de redução de leituras. Assim, finalmente, decidiu-se

que a leitura seja feita aproximadamente após percorridos dois terços do quarteirão, a fim de reduzir tanto o número de leituras (no máximo duas por quarteirão) quanto a possibilidade de perda de sinal. Desta forma, finalmente, levando em conta a velocidade média de pedestre (1.4 m/s), a taxa para o primeiro intervalo (0.8-2) foi fixada em 45 segundos.

Finalmente, devido a que as rodovias, diferentemente das ruas, carecem de alguma unidade padrão de medida útil que indique a distância até uma próxima conexão com outro caminho, como os quarteirões, decidiu-se adotar algumas das taxas anteriores para os dois últimos intervalos (15-25 e 25-35). O número de leituras necessárias para poder reconstruir uma trajetória nas avenidas e rodovias é menor devido às maiores extensões delas, assim, a adoção destas taxas resulta uma decisão razoável. Dessa forma, por fim, um intervalo de amostragem de 25 e outro de 45 segundos foram alocados nos intervalos de velocidade quarto (15-25) e quinto (25-35) respectivamente.

Cabe mencionar, também, que os valores dos intervalos de amostragem consecutivos de 5 e 10 minutos para os casos onde o usuário fica parado ou alcança uma velocidade maior do último intervalo de velocidade definido, foram selecionados com intuito de evitar deixar passar muito tempo sem uma atualização e também para manter atualizadas as informações dos satélites, as quais têm um tempo de validade e cujas atualizações poderiam atrasar as atualizações de localização se o *GPS* ficar desligado por muito tempo (LEHTINEN; HAPPONEN; IKONEN, 2008). Aliás, para transitar do intervalo de amostragem de 5 a de 10 minutos não é preciso nenhum limiar de velocidade. Quando o usuário ultrapassar o último intervalo de velocidade, terá um intervalo de amostragem de 5 minutos a primeira vez. Se, uma vez transcorridos os 5 minutos, o usuário mantém uma velocidade maior do último intervalo, imediatamente adotará um intervalo de amostragem de 10 minutos e não voltará ao de 5 minutos a menos que a bússola detecte um giro significativo que indique uma mudança de direção. Nesse caso, o intervalo de amostragem será reiniciado e adotará o valor inicial (10 segundos). De forma similar, se o usuário continuar sem se movimentar, uma vez transcorridos os primeiros 5 minutos, o intervalo de amostragem mudará para 10 minutos.

Por outro lado, um outro procedimento importante a descrever é a detecção de movimento, a qual tem como base a análise das leituras do acelerômetro. A ideia é contar o número de picos que aparecem na série temporal formada pelas leituras do acelerômetro dentro de um período de tempo dado. Se o número de picos ultrapassar certo limiar, indicaremos que o usuário começou a se movimentar. Apesar da simplicidade desta abordagem, sua eficácia é comprovada na prática.

Uma dificuldade importante com esta ideia é que se o valor do limiar for muito baixo, atividades simples como empurrar o celular ou guardar ele no bolso poderiam gerar uma falso positivo. Por outro lado, se o valor for muito alto, corremos o risco de perder informação de localização importante no período prévio à ativação do GPS, especialmente se o usuário iniciar o movimento com uma alta velocidade. Adicionalmente, é importante notar que quando o usuário começar a se locomover e o celular estiver dentro de um bolso apertado como no caso de uma

calça jeans, o celular vai se movimentar menos devido às restrições de espaço do que em um bolso mais folgado como o de uma mochila. Dessa forma, o número de picos será maior no segundo caso.

Por fim, através de um processo de experimentação levando em conta todos estes critérios, adotou-se um limiar de 20 picos em um período de um minuto. Se o número de picos não ultrapassar o limiar nesse período, a contagem é reiniciada. Além disso, estabeleceram-se validações a cada 5 segundos, assim, se o valor do limiar for atingido antes de um minuto, não precisaremos esperar até o final do período para iniciar o rastreamento. É importante mencionar que a validação não reinicia a contagem a menos que o limiar seja ultrapassado.

Por outro lado, uma tarefa importante é determinar os picos que são válidos para a contagem. Alguns picos podem ser produto do ruído do acelerômetro, assim, é importante selecionar aqueles picos significativos dentre as leituras. Para isso, estabeleceu-se uma altura mínima para cada pico e períodos máximo e mínimo entre picos. Mediante experimentação, foram determinados um valor de 1 m/s^2 para a altura mínima de pico, um período mínimo entre picos de 1 segundo e um máximo de 10 segundos.

Para o caso da bússola, apenas foi necessário transformar alguns valores de orientação obtidos através da matriz de rotação para calcular o azimuth ou distância angular medida sobre o horizonte. Uma vez definido um azimuth inicial, a diferença entre ele e os subsequentes azimuths derivados de cada leitura da bússola são avaliados de tal forma que se o valor da diferença estiver próximo de 90 graus sexagesimais, indicaremos que o usuário mudou de direção e o GPS será ligado novamente.

4.1.3 Metodologia

Uma vez explicado o sistema de rastreamento, detalhamos como foi o processo de coleta.

Estabeleceram-se dois períodos de coleta. O primeiro deles teve uma duração de 5 meses, durante o qual utilizou-se um intervalo de amostragem fixo de 10 segundos. O intuito deste esquema de coleta foi elaborar um *benchmark* contra o qual poder comparar o método adaptativo do *SmartTracking*. Por outro lado, devido a que o intervalo de amostragem neste período de coleta ser o mais baixo dos utilizados nos dois períodos (10 segundos), este esquema de coleta foi chamado de esquema contínuo.

O segundo período de coleta teve uma duração de 4 meses e utilizou-se nele o rastreamento adaptativo apresentado anteriormente. Assim, este esquema de coleta foi chamado de esquema adaptativo. Aliás, é importante lembrar que um intervalo de amostragem variável, em geral, permite obter uma economia de energia e espaço de armazenamento ao efetuar apenas as leituras consideradas necessárias segundo as informações do contexto. Isto é, minimizando a necessidade de leitura sem prejuízo da qualidade da informação sobre a trajetória do usuário. O experimento nos dois períodos foi feito na cidade de São Carlos no estado de São Paulo e

contou com a participação de um grupo de 15 voluntários equipados com um celular com GPS. Estes voluntários foram instruídos para ativarem o serviço de rastreamento o máximo de tempo possível durante o dia e realizarem suas atividades cotidianas de forma normal deixando o GPS ligado. No caso do primeiro período, o tempo de coleta total foi maior, uma vez que o alto custo energético de manter o GPS ativo e de realizar leituras com uma alta frequência limitou de forma significativa as horas efetivas de coleta por dia. Ou seja devido a estas restrições energéticas, o aplicativo não podia ser utilizado por um dia inteiro.

Em contraste, o segundo período teve um tempo de coleta menor, devido principalmente ao uso do esquema de coleta adaptativo, o qual permitiu incrementar as horas efetivas de coleta diárias ao reduzir o consumo de energia nos celulares dos usuários. Este esquema atingiu tal objetivo mediante a regulação do intervalo de amostragem, o qual permitiu reduzir a quantidade de leituras. Adicionalmente, outro fator relevante que contribuiu com a economia de energia foi a desativação automática do GPS durante períodos de movimentação nula ou reduzida, tais como: durante uma aula, trabalho, estudo ou refeição.

Um detalhe importante a mencionar é que, em ambos os períodos, foi utilizado o *spooling*, o qual contribuiu com a economia de bateria ao não ter que estar conectado à Internet para gerar atualizações de localização.

Finalmente, um outro detalhe a considerar é que apenas dois voluntários registraram trajetórias de carro, enquanto a maioria registrou trajetórias de pedestre. Por outro lado, embora o RadrPlus permita obter dados de vários meios de transporte, do total de trajetórias da base de dados resultante, apenas foram levadas em conta nos experimentos aquelas trajetórias feitas por pedestres.

4.2 Base de dados

Nesta seção, descrevemos as características e distribuição, bem como o formato final da base de dados obtida através do processo de coleta durante os dois períodos.

4.2.1 Descrição

Na Tabela 1, pode-se observar o período de coleta de ambos os esquemas, o número de usuários e a quantidade de pontos ou atualizações de localização geradas tanto por usuários pedestres quanto motoristas. O número total de pontos considerando os dois períodos é 10150, 7417 pontos capturados com o esquema adaptativo e 2733 pontos com o esquema contínuo. Por outro lado, o número total de pontos gerados por usuários pedestres e motoristas foi 5167 e 4983 respectivamente.

Na Figura 31, podemos observar a distribuição dos dados em ambos os períodos mediante o uso de mapas de calor. A Figura 31a mostra os dados junto com as informações de rodovias,

Tabela 1 – Características da base de dados RadrPlus

	Esquema contínuo	Esquema adaptativo
Período de coleta	02/2016 - 06/2016	07/2016 - 10/2016
Num. de usuários	15	15
Num. de pontos feitos de carro	1562	3421
Num. de pontos feitos a pé	1171	3996

Fonte: Dados da pesquisa.

Figura 31 – Distribuição dos dados do RadrPlus

Fonte: Elaborada pelo autor.

enquanto a Figura 31b mostra uma imagem satelital. Em ambas as imagens, notamos que a área de coleta foi principalmente o campus da USP de São Carlos e seus redores; porém, algumas atualizações de localização foram feitas fora dessa área, mas por serem bem poucas em comparação ao resto, resultam quase invisíveis neste tipo de gráfico. Por outro lado, o gráfico permite enxergar algumas informações interessantes como as trajetórias mais comuns e os lugares mais concorridos, graças ao padrão de cores. Os pontos vermelhos nestes gráficos indicam os pontos mais "quentes" ou os lugares onde houve maior quantidade de atualizações. Tais locais correspondem a pontos de permanências (residência ou trabalho) de usuários que mantiveram o RadrPlus ativo por mais tempo. À medida que tal quantidade diminui, a cor muda a amarelo e finalmente a verde. Além disso, a intensidade da cor verde varia em proporção direta com a quantidade de atualizações.

4.2.2 Formato

As bases de dados tanto do primeiro período quanto do segundo foram armazenadas em um arquivo plano com um formato de valores separados por vírgulas (CSV). Tal formato é mostrado em seguida:

- Campo 1: Identificador de usuário.
- Campo 2: Estampa de tempo ou *timestamp* em formato horário Unix¹.
- Campo 3: Latitude em graus decimais.
- Campo 4: Longitude em graus decimais.
- Campo 5: Acurácia em metros.

Cabe mencionar que a acurácia representa a diferença máxima entre o valor medido e o valor real, assim, quanto menor seu valor, maior a qualidade da leitura.

Com intuito de ilustrar melhor o formato mencionado, apresentamos, em seguida, um exemplo das entradas da base de dados.

```
0001, 1470965613, 65.9667, 0.0, 3.9
0015, 1470175672, -22.0065377, -47.895631, 4
0001, 11472126814, -22.0069667, -47.894055, 6.5
```

4.3 Análise da técnica de captura

Nesta seção, comparamos o desempenho do esquema de captura de dados adaptativo contra o contínuo. Para isso, primeiro mostramos graficamente as diferenças entre as trajetórias coletadas com ambos os esquemas em diferentes cenários, e analisamos a quantidade de leituras efetuadas e tempo transcorrido entre elas. Após isso, complementamos a análise prática com a teórica, fazendo uma projeção comparativa da quantidade de leituras feitas utilizando ambos os métodos em diferentes cenários, nos quais existe uma velocidade fixa representativa para cada um deles.

4.3.1 Análise visual

Na Figura 32, pode-se apreciar duas trajetórias cujos dados de localização foram coletados mediante o esquema contínuo. A trajetória da esquerda representa um percurso feito na rodovia, enquanto a trajetória da direita descreve um percurso feito na rua. Observa-se que, em ambos os casos, o tempo entre leituras (o intervalo de amostragem) é, como esperado, sempre o mesmo (10 segundos). O esquema contínuo garante que, sem importar a situação, o valor da taxa de amostragem será sempre a mesma. Esta característica é refletida graficamente na maior quantidade de leituras da trajetória e no estreito espaço entre elas.

¹ <<http://www.unixtimestamp.com>>

Figura 32 – Trajetórias coletadas na rodovia e rua usando o esquema de coleta contínuo.

Fonte: Elaborada pelo autor.

Por outro lado, na Figura 33, pode-se observar uma trajetória obtida na rodovia com o esquema de coleta adaptativo, bem como o detalhe das informações temporais das leituras que a compõem.

A Figura 33a mostra claramente uma trajetória com um comportamento diferenciado. No começo da trajetória, o canto inferior direito, as leituras estão muito próximas uma de outra; porém, à medida que o usuário aproxima-se à rodovia, o espaço entre leituras vai aumentando gradualmente até atingir um espaço constante máximo no trecho final da trajetória. É importante mencionar que as leituras no começo da trajetória foram feitas em uma avenida ou rua, enquanto o resto de leituras foi feito na rodovia mesmo.

Com intuito de entender melhor o comportamento desta trajetória, a Figura 33b mostra o *timestamp* de cada leitura do trecho de trajetória onde ocorre a mudança de comportamento. Assim, nesta figura, podemos notar que as leituras feitas ainda na rua têm um intervalo de amostragem de 25 segundos, que segundo a heurística do esquema adaptativo mostrado na Figura 29, corresponde a um usuário com uma velocidade entre 15 e 24 m/s, o que poderia indicar um usuário indo de carro por uma avenida. Por outro lado, nas primeiras leituras feitas na rodovia, observa-se que o intervalo de amostragem ainda continua sendo de 25 segundos, porém, esse intervalo muda a 45 segundos na leitura final. Este comportamento reflete claramente o funcionamento da heurística mencionada. Ao entrar na rodovia, o usuário incrementa sua velocidade e atinge e/ou supera os 25 m/s, mudando automaticamente o intervalo de amostragem a 45 segundos. Desta forma, fica evidenciada a fidelidade entre a heurística projetada no esquema de coleta adaptativo e o processo de captura finalmente executado.

Por outro lado, na Figura 34, pode-se observar as atualizações de localização de dois usuários fazendo aproximadamente o mesmo percurso na rodovia. No lado esquerdo da figura,

Figura 33 – Trajetória coletada na rodovia usando o esquema de coleta adaptativo.

Fonte: Elaborada pelo autor.

mostra-se o percurso feito utilizando o esquema adaptativo, enquanto, no lado direito, utilizou-se o esquema contínuo.

À simples vista, observa-se uma grande diferença na quantidade de pontos ou atualizações entre ambos os percursos, no entanto, torna-se possível fazer a reconstrução da trajetória com uma alta precisão. Apesar do tempo total do trajeto ser aproximadamente o mesmo (5 minutos), o percurso feito com o esquema adaptativo contém muito menos pontos do que o percurso feito com o esquema contínuo. Este resultado estava previsto, pois como comentado anteriormente, o esquema contínuo mantém um intervalo de amostragem fixo e, consequentemente, não tem a possibilidade de aproveitar as condições do entorno, como, por exemplo, viajar pela rodovia, onde tem-se apenas uma via, comumente reta, por um período longo de tempo. Dessa forma, no caso dos percursos feitos na rodovia, não é preciso uma frequência alta de atualizações para guardar tal percurso com uma boa qualidade.

No caso dos percursos feitos nas ruas, tem-se, também, um comportamento diferenciado entre as trajetórias capturadas com os dois esquemas de coleta. A Figura 35 permite apreciar esta diferença por meio de uma comparação entre percursos feitos na rua usando ambos os esquemas.

Na Figura 35a, pode-se observar um percurso feito na rua usando o esquema adaptativo. Nesta figura, o intervalo de amostragem no começo do percurso, canto inferior esquerdo, é de 10 segundos, em seguida, muda seu valor a 25 segundos e, finalmente, retorna ao valor inicial de 10 segundos. Este comportamento parece indicar uma locomoção descontínua ou interrompida, *i.e.*, um percurso com algumas paradas intermediárias. Dado que ao iniciar o movimento o intervalo de amostragem é de 10 segundos, o retorno a este valor só pode indicar um reinício de movimento após uma parada. Uma possível explicação para este comportamento seria o caso de

Figura 34 – Comparação de trajetórias coletadas na rodovia usando os esquemas de coleta adaptativo e contínuo.

Fonte: Elaborada pelo autor.

um usuário pedestre com um itinerário de atividades, o qual requer visitar vários lugares. Por outro lado, uma outra possibilidade seria um usuário motorista frente à presença de um semáforo ou de uma situação de engarrafamento veicular. Esta última ideia poderia ser muito útil para alguma aplicação de monitoramento de tráfego. A mudança do intervalo de amostragem junto com a velocidade poderiam servir para detectar um motorista dentro de um engarrafamento, enquanto os dados de localização indicariam o lugar exato do acontecimento.

Em contraste com a anterior figura, a Figura 35b mostra um percurso feito na rua usando o esquema de coleta contínuo. Nesta figura, observa-se que a diferença na quantidade de leituras entre ambos os esquemas não é tão marcada quando comparada com a diferença entre os percursos feitos na rodovia. Este comportamento poderia dever-se à maior quantidade de conexões e vias, menores limites de velocidade, e maior quantidade de possíveis paradas no contexto de ruas. Contudo, existe uma diminuição de leituras e consequente economia de energia consideráveis no caso do esquema adaptativo, dado que, além da possibilidade de utilizar intervalos de amostragem maiores (25 ou 45 segundos), este esquema detém as leituras do *GPS* quando o usuário encontra-se parado, diferentemente do esquema contínuo, que continua fazendo leituras e com o mesmo intervalo baixo.

4.3.2 Projeção comparativa

Com intuito de avaliar de forma mais exaustiva o desempenho do esquema de captura adaptativo, foi feita uma projeção das quantidades de leituras geradas utilizando os diferentes

Figura 35 – Comparação de trajetórias coletadas na rua usando esquemas de coleta adaptativo e contínuo.

Fonte: Elaborada pelo autor.

intervalos de amostragem calculados segundo a velocidade média dos meios de locomoção terrestre mais comuns (pedestre, bicicleta e carro). Posteriormente, comparamos tais resultados contra a quantidade de leituras projetada utilizando o esquema de coleta contínuo. Utilizamos o número de leituras como métrica porque permite medir tanto a quantidade de espaço de armazenamento quanto o gasto de energia. A quantidade de leituras é diretamente proporcional ao gasto de energia incorrido.

Levando em conta o Quadro 1 apresentado previamente, acrescentamos nele o intervalo de amostragem e, assim, elaboramos o Quadro 2, onde mostramos os diferentes meios de locomoção e suas respectivas velocidades médias, bem como o intervalo de amostragem correspondente a cada velocidade, calculado utilizando a heurística projetada no esquema de captura adaptativo.

Quadro 2 – Velocidade média e intervalo de amostragem de meios de locomoção

Meio de locomoção	Velocidade média (m/s)	Intervalo de amostragem (s)
Pedestre	1.40	45
Bicicleta	4.02	25
Carro em ruas (cidade)	5.09 - 8.95	10
Carro em avenidas (cidade)	13.88 - 33.33	25 - 45
Carro em rodovias	33.6 - 35.8	300 - 600

Desta forma, finalmente, na Tabela 2, observa-se o número de leituras projetado teoricamente pelo esquema de coleta contínuo e pelos diferentes meios de locomoção utilizando o esquema adaptativo, durante diferentes períodos de tempo. Adicionalmente, calculou-se a taxa de redução média das quantidades de leitura geradas com os diferentes meios de locomoção

respeito das quantidades de leituras geradas com o esquema contínuo.

Tabela 2 – Projeção comparativa de quantidade de leituras usando o esquema de coleta adaptativo em diversos meios de locomoção

Tempo (mins)	Contínuo	Pedestre	Bicicleta	Carro-ruas	Carro-avenidas	Carro-rodovias
1	6	1	2	6	2	0
2	12	2	4	12	4	0
3	18	4	7	18	7	0
4	24	5	9	24	9	0
5	30	6	12	30	12	1
6	36	8	14	36	14	1
7	42	9	16	42	16	1
8	48	10	19	48	19	1
9	54	12	21	54	21	1
10	60	13	24	60	24	1
% de redução	-	80	62	0	62	97

Fonte: Dados da pesquisa.

Nesta projeção, assumimos que não há paradas e que a velocidade varia o mínimo para que o intervalo de amostragem se mantenha estável. Outra observação importante é que, no caso do carro em avenidas, levou-se em conta o valor médio do intervalo respectivo (23 m/s); assim, consequentemente, seguindo a heurística do esquema adaptativo, utilizou-se um intervalo de amostragem de 25 segundos. Por outro lado, cabe mencionar que, depois dos primeiros 5 minutos, o intervalo de amostragem do carro em rodovias muda para 10 minutos. Dessa forma, à medida que o tempo passar, apenas haverá uma leitura a cada 10 minutos, reduzindo, assim, ainda mais o número de leituras. Esta última observação explica a razão pela qual, no minuto 10, contabilizamos apenas uma leitura, pois a segunda leitura seria feita depois de transcorridos 10 minutos, o que aconteceria no minuto 15.

Por outro lado, na Tabela 2, pode-se observar, também, que a maior redução foi alcançada no caso do carro em rodovias com um 97%. Este resultado indica que, a maior economia de energia e espaço de armazenamento obtém-se quando o usuário está se locomovendo a altas velocidades. Em contraste, no caso do carro em ruas, não houve redução nenhuma. Este resultado deve-se ao fato que a velocidade média dos carros na rua tem alocada, segundo o esquema de coleta adaptativo, um intervalo de amostragem igual ao intervalo do esquema contínuo (10 segundos). Dessa forma, o número de leituras do carro na rua usando o esquema de coleta adaptativo é idêntico ao do esquema contínuo.

Além disso, pode-se apreciar que tanto no caso da bicicleta quanto do carro na rodovia, a redução é de 62%, o que representa uma redução muito considerável e significativa, especialmente levando em conta que cabe a possibilidade que alguns carros se movimentem pelas ruas com uma velocidade similar à das bicicletas, devido ao trânsito. Adicionalmente, alguns carros nas ruas poderiam atingir uma velocidade similar aos carros nas avenidas ao encontrar-se viajando

por uma rua muito longa e com pouco trânsito. Finalmente, no caso dos pedestres, a redução foi de 80%, o que é o resultado mais importante, pois os pedestres representam a maior quantidade de possíveis usuários do sistema proposto neste trabalho. Assim, uma redução considerável de energia no serviço de rastreamento permitirá ampliar o tempo de coleta e a quantidade de dados, bem como reduzir o desconforto do usuário ao ter um gasto adicional de energia por manter o *GPS* ligado.

4.4 Considerações finais

Neste capítulo, apresentamos uma revisão das principais abordagens de coleta efetiva de dados de geolocalização em dispositivos móveis, e descrevemos detalhadamente o serviço de coleta *SmartTracking*, explicando seu funcionamento, bem como os critérios para a elaboração da heurística na qual se baseia. Adicionalmente, descrevemos a metodologia de coleta utilizada.

Por outro lado, a base de dados coletada com o RadrPlus foi descrita detalhadamente, mostrando suas características, distribuição geográfica dos dados e o formato final dos mesmos.

Finalmente, realizou-se uma comparação entre o esquema de coleta adaptativo utilizado no *SmartTracking* e um esquema contínuo com um intervalo de amostragem fixo. Tal comparação foi feita mediante uma análise gráfica de trajetórias reais coletadas com ambos os esquemas, assim como através de uma projeção comparativa do número de leituras geradas por cada esquema usando diferentes meios de locomoção e com diversos tempos de viagem. Os resultados mostraram a grande superioridade do esquema adaptativo frente ao contínuo, conseguindo reduzir o número de leituras em um 80%, 62% e 97% para os pedestres, ciclistas e motoristas respectivamente. Cabe mencionar que, para o caso de motoristas na rua, não teve redução devido ao intervalo de amostragem ser o mesmo do esquema contínuo.

Uma possível melhora do método de coleta adaptativo seria alterar dinamicamente os valores dos intervalos de amostragem e limiares de velocidade dependendo da região (país, estado ou cidade) onde o usuário encontra-se. Assim, por meio de uma leitura do *GPS*, o sistema poderia saber em que região encontra-se o usuário e adaptar sua configuração segundo as informações locais de tamanho médio dos bairros e velocidade média dos veículos nas ruas, avenidas e rodovias. O tamanho médio dos bairros poderia ser obtido do regulamento vigente ou através de estudos de campo. Por outro lado, a velocidade média dos veículos poderiam ser calculadas levando em conta os limites de velocidade ou estatísticas de velocidade locais. Aliás, as velocidades dos pedestres e ciclistas poderiam permanecer com os mesmos valores a menos que os estudos locais evidenciem uma diferença significativa entre as velocidades médias locais e as predefinidas na heurística.

RECONSTRUÇÃO DE TRAJETÓRIAS

Neste capítulo, são apresentados os fundamentos da tarefa de reconstrução de trajetórias, o objetivo dessa tarefa, a origem e as diversas categorias e principais técnicas que têm surgido até o momento. Adicionalmente, apresenta-se um *framework* projetado para aplicar as técnicas de reconstrução de trajetórias descritas, no contexto de pedestres.

5.1 Reconstrução de trajetórias

À medida que as atualizações de localização de um objeto são armazenadas sob um certo intervalo de amostragem, a trajetória obtida é comumente apenas uma amostra do verdadeiro movimento de tal objeto. Consequentemente, o movimento de um objeto entre dois pontos consecutivos torna-se desconhecido ou *incerto*. Desta forma, a primeira intuição seria a de realizar esforços por reduzir essa incerteza nas trajetórias, porém, esta tarefa depende muito da aplicação. Por exemplo, se o objetivo for proteger a privacidade do usuário, deve-se, então, fazer a trajetória ainda mais incerta. No obstante, existem muitas outras aplicações que visam estudar o comportamento do objeto e precisam reduzir tal incerteza o máximo possível; assim, a tarefa de reconstrução de trajetórias surge como resposta a tal necessidade. Os termos reconstrução e inferência de trajetórias são intercambiáveis, porém, neste trabalho, consideramos a reconstrução como o processo todo, o qual inclui o pré-processamento e também a inferência, enquanto considera-se inferência apenas como a aplicação final de um algoritmo sobre os dados pré-processados a fim de reconstruir a trajetória. Esta ideia é descrita mais claramente na Figura 37 e sua correspondente seção.

5.1.1 Trajetórias incertas

Muitas trajetórias são coletadas com uma taxa de amostragem baixa, o que gera trajetórias com espaços entre atualizações de localização incertos. Assim, por exemplo, na Figura 36.a, pode-

se observar que as coordenadas GPS de um taxi (p_1, p_2, p_3) foram coletadas com um intervalo de amostragem longo com intuito de reduzir o consumo de energia; desta forma, permitindo considerar muitos possíveis caminhos entre cada par de pontos consecutivos. De forma similar, na Figura 36.b, observa-se como é possível formar trajetórias ordenando cronologicamente *check-ins* gerados com algum serviço de geolocalização. Contudo, estes *check-ins* apresentam intervalos de tempo longos, uma vez que os usuários não compartilham sua localização frequentemente, inclusive, podem transcorrer horas até um novo *check-in* ser gerado. Por fim, na Figura 36.c, mostra-se o caso de uma trajetória gerada por uma ave migratória com um dispositivo GPS instalado. Com intuito de economizar energia o dispositivo apenas envia uma atualização de localização por dia, gerando, assim, uma trajetória com elevada incerteza.

Figura 36 – Exemplos de trajetórias incertas

Fonte: Adaptada de Zheng (2015).

5.1.2 Inferência de trajetórias

Devido as incertezas comentadas anteriormente, têm surgido recentemente alguns métodos para lidar com ela em questões importantes como a aplicação de consultas em bancos de dados de objetos em movimento. Uma destas abordagens, por exemplo, é estabelecer limites geométricos em forma de corpos cilíndricos tridimensionais (TRAJCEVSKI *et al.*, 2004; TRAJCEVSKI *et al.*, 2009) ou colares (TRAJCEVSKI *et al.*, 2010) para aproximar uma trajetória. Por outro lado, existem outras abordagens que ao invés de modelar a incerteza, tentam reduzi-la o máximo possível, estas são as técnicas de inferência de trajetórias. A ideia deste tipo de técnicas é inferir as k rotas ou caminhos mais prováveis que um objeto poderia ter seguido entre alguns pontos amostrados (observações) baseado em um grupo de trajetórias incertas. O raciocínio por trás dessa ideia é que as trajetórias que compartilham alguma rota similar poderiam se completar mutuamente. Dessa forma, poderíamos dizer que, "incerto + incerto $\rightarrow$ certo" (ZHENG, 2015). Tomando como exemplo a Figura 36.a, onde os pontos de diversas cores representam as atualizações de localização de diferentes taxis, pode-se inferir que a rota em azul é a mais provável rota que atravessa (p_1, p_2, p_3). De forma similar, na Figura 36.b, poderíamos dizer que a curva em azul é a rota mais provável entre os três *check-ins* em azul. Por fim, na Figura 36.c, levando em

conta as trajetórias de várias aves, poderíamos inferir o caminho seguido por outras aves entre alguns pontos de amostragem.

Por outro lado, como mencionado anteriormente, a reconstrução das trajetórias ou a redução de incerteza das trajetórias (ambos os termos resultam equivalentes) possibilitam um grande número de aplicações. Desta forma, existem diversos métodos para endereçar este problema, os quais podem ser divididos em duas categorias (ZHENG, 2015) comentadas a seguir.

A primeira delas compreende aqueles métodos projetados para a reconstrução de trajetórias sobre uma rede viária (ZHENG *et al.*, 2012; BANERJEE; RANU; RAGHAVAN, 2014; HUNTER; ABBEEL; BAYEN, 2013; ZHENG *et al.*, 2012; LI; AHMED; SMOLA, 2015; CHIANG *et al.*, 2013). Poderia resultar natural encontrar muita similaridade com os métodos de *Map matching*, porém, existem diferenças importantes a serem consideradas. Primeiro, os métodos de reconstrução levam em conta informação de outras trajetórias, enquanto os de *Map matching* apenas usam a informação geográfica de uma trajetória isolada e a informação topológica da rede viária. Em segundo lugar, o intervalo de amostragem, no caso dos métodos de reconstrução, podem ser maiores, de até 10 minutos, o que resulta quase impossível para um método de *Map matching*. Outro método de reconstrução de trajetória geradas na rede viária é apresentado em Banerjee, Ranu e Raghavan (2014). Esta técnica, ao invés de prever a rota mais provável, retorna um grafo ponderado que sumariza todas as possíveis rotas. A reconstrução da trajetória utiliza a técnica de amostragem de Gibbs (CASELLA; GEORGE, 1992) para aprender um Modelo de Mobilidade na Rede a partir de uma base de dados histórica de trajetórias.

A segunda categoria abrange os métodos projetados para os espaços livres, onde os objetos não seguem caminhos nas redes viárias (WEI; ZHENG; PENG, 2012). As Figuras 36.b e 36.c permitem entender melhor a ideia. Este tipo de métodos enfrenta o problema de encontrar aquelas trajetórias que poderiam resultar relevantes para um conjunto de observações, bem como construir uma rota que aproxime tais trajetórias relevantes. Por exemplo, Wei, Zheng e Peng (2012) apresentam um *framework* de inferência de trajetórias baseado em conhecimento coletivo chamado RICK (*Route Inference based on Collective Knowledge*). Dada uma sequência de pontos (localizações) e uma janela de tempo, RICK constrói as *top-k* rotas que atravessam os pontos dentro da janela de tempo. Para isso, ele utiliza os históricos de trajetória para dividir o espaço geográfico em regiões e, finalmente, criar um grafo com as possíveis rotas entre as regiões. Cada nó é uma região e uma aresta é o caminho entre elas. Por outro lado, Su *et al.* (2013) propuseram um sistema de calibração baseado em âncoras que alinha as trajetórias a um conjunto fixo de pontos âncoras. Para isso, esta abordagem leva em conta a relação espacial entre os pontos âncora e as trajetórias. Além disso, esta técnica treina modelos de inferência a partir de um histórico de trajetórias para melhorar a calibração.

5.2 Reconstrução de trajetórias de pedestres usando informação da rede viária

Apesar dos vários trabalhos que visam endereçar o problema de reconstrução de trajetórias, a grande maioria levam em conta apenas trajetórias de veículos, principalmente pelo fato que muitos dos percursos dos pedestres ficam fora das ruas (e.g., trilhas, pequenos caminhos dentro de campus, parques, becos, etc). Porém, atualmente, com o uso das plataformas livres de mapas colaborativos como OpenStreetMap¹, os usuários locais podem adicionar ao grafo da rede viária aqueles nós e arestas que representam as trilhas e caminhos transitados exclusivamente pelos pedestres, possibilitando, assim, o uso destas informações na reconstrução de trajetórias de pedestres.

Na Figura 37, pode-se observar o processo geral da reconstrução de trajetórias utilizado. Este processo está composto de três etapas: segmentação, map matching e inferência. Durante a etapa da segmentação, os *logs* dos usuários, i.e., os históricos de atualizações de localização (normalmente um arquivo por usuário) são transformados em trajetórias, levando em conta certos critérios específicos. Uma vez transformados os *logs*, as trajetórias resultantes são agrupadas e a informação relacionada aos usuários é descartada. Após a segmentação, na etapa de *Map matching*, cada uma destas trajetórias trocam seu formato de sequência de pontos espaço-temporais (pontos GPS), e são representados em termos das arestas de uma rede viária. Finalmente, estas trajetórias restritas à rede junto com a rede viária são utilizadas na etapa de Inferência para reconstruir a trajetória associada a um conjunto de observações recebida como entrada. Todos estes conceitos são explicados detalhadamente em seguida.

5.2.1 Definições preliminares

Nesta subseção, apresentam-se algumas definições necessárias para o correto entendimento do processo de reconstrução de trajetórias. Tais definições são baseadas em Banerjee, Ranu e Raghavan (2014).

Definição 5 (Rede viária). A rede viária é um grafo $G(V, E)$ dirigido, onde V é um conjunto de vértices representando as interseções e pontos terminais, e E é o conjunto de arestas $e = (v_i, v_j)$, conectando $v_i, v_j \in V$, representando os segmentos de rua. A posição de um nó é representada pela sua latitude e longitude.

A notação $e.p_1$ e $e.p_2$ denota os pontos terminais de uma aresta e , onde $p_1, p_2 \in V$. Em geral, uma trajetória $T = \{s_1, \dots, s_n\}$, também chamada de trajetória GPS, é uma sequência ordenada temporalmente de pontos espaço-temporais, também chamados de pontos GPS. Um ponto GPS $s = (l, t)$ é uma tupla que contém uma localização espacial l e uma estampa de tempo t que codifica a hora do dia na qual l foi transitada. Adicionalmente, $T.s_i$ denota o i_{th} ponto

¹ <<https://www.openstreetmap.org>>

Figura 37 – Processo de reconstrução de trajetórias

Fonte: Elaborada pelo autor.

espaço temporal em T , e $s.l$ e $s.t$ denotam a localização e estampa de tempo em s respectivamente. Assim, $T.s_i.t < T.s_{i+1}.t$.

Definição 6 (*log GPS*). Um *log GPS* é uma coleção de pontos GPS, $P = p_1, p_2, \dots, p_n$. Em geral, tem-se um *log GPS* para cada usuário de um serviço de geolocalização.

Definição 7 (*Trajetoória restrita à rede*). T é restrita à rede viária $G(V, E)$, se $\forall T.s_i, T.s_i.v \in V$, e $\forall T.s_i, T.s_{i+1}, (T.s_i.v, T.s_{i+1}.v) \in E$. Assim, em outras palavras, uma trajetória restrita à rede é uma rota associada a uma rede viária.

Definição 8 (*Observações parciais*). Uma observação parcial $O = s_1, \dots, s_n$, é uma sequência de pontos GPS ordenados temporalmente, onde $\forall O.s_i, O.s_i.v \in V$. Geralmente, estas observações apresentam lacunas espaço-temporais, i.e., entre um ponto $O.s_k$ e um outro $O.s_{k+1}$, não é possível definir com exatidão qual foi a rota percorrida. Isto acontece, principalmente, devido ao uso de um intervalo de amostragem muito longo no processo de captura dos dados de geolocalização. Por outro lado, em uma tarefa de reconstrução de trajetórias, dado um conjunto de observações parciais O , o objetivo é inferir a trajetória mais provável da qual foram tomadas tais observações. Desta forma, procura-se reconstruir a trajetória original a partir destas observações através de um processo de inferência.

Definição 9 (*Trajetoória incerta*). Dada uma rede viária $G(V, E)$, um histórico de trajetórias $\mathbb{H}$, e um conjunto de observações parciais O , uma trajetória incerta é um grafo ponderado $U(V', E') \subseteq G(V, E)$, tal que $E' = \{e \in E | p(e|O) > 0\}$, onde $p(e|O)$ denota a probabilidade da

aresta e ser percorrida dado O . Por outro lado, V' é definido analogamente levando em conta os pontos terminais das arestas em E' . Desta forma, uma trajetória incerta U contém todas as arestas que possuem uma probabilidade de serem transitadas maior que zero, desta forma, representando todas as possíveis trajetórias.

5.2.2 Segmentação

Muitas tarefas de mineração de trajetórias, tais como agrupamento e classificação, precisam dividir uma trajetória em segmentos previamente. A segmentação não só reduz a complexidade computacional, mas, além disso, possibilita minerar conhecimentos mais ricos como padrões em sub-trajetórias (ZHENG, 2015). No contexto da reconstrução de trajetórias, tem-se os *logs* GPS, os quais, em geral, armazenam os deslocamentos das pessoas por longos períodos de tempo (e.g., dias, semanas, meses e anos), adicionalmente, considera-se cada *log* como se fosse uma única e muito longa trajetória. Porém, durante este período, a pessoa poderia ter feito diversas viagens, as quais estariam representadas como sub-trajetórias, que acabariam se perdendo se tomássemos em conta cada *log* como uma trajetória única. Desta forma, torna-se relevante para a nossa tarefa a segmentação de trajetórias.

Em geral, têm-se três tipos de métodos de segmentação. A primeira categoria é baseada no intervalo de tempo. Se o intervalo de tempo entre dois pontos consecutivos é maior que um certo limiar, a trajetória é dividida em duas partes, como pode-se observar na Figura 38.a. Adicionalmente, outra abordagem deste tipo é dividir a trajetória em segmentos de igual tempo de duração.

Figura 38 – Métodos de segmentação de trajetórias sobre uma rede viária

Fonte: Adaptada de Zheng (2015).

A segunda categoria é baseada na forma da trajetória. É possível dividir uma trajetória levando em conta os pontos de inflexão, aqueles onde tem-se uma mudança de direção que supere certo limiar, como mostrado na Figura 38.b. Além disso, uma outra abordagem do mesmo tipo seria a identificação de pontos chaves que mantêm a forma de uma trajetória, mediante o uso de algoritmos de simplificação linear como o algoritmo de Douglas-Peucker (SAALFELD, 1999). Tal abordagem é ilustrada na Figura 38.c. Por outro lado, de forma similar, Lee, Han e

Whang (2007) propuseram uma segmentação de trajetórias utilizando o conceito de Descrição de Comprimento Mínimo (DCM), o qual compreende dois componentes: $L(H)$ e $L(D|H)$. $L(H)$ é o comprimento em bits da descrição dos dados quando codificados com ajuda da hipótese. A melhor hipótese H que explique D é aquela que minimize a soma de $L(H)$ e $L(D|H)$. Especificamente, eles utilizam $L(H)$ para representar o comprimento total dos segmentos particionados, enquanto $L(D|H)$ representa a distância total entre a trajetória original e os novos segmentos particionados. Por meio de um algoritmo de aproximação, eles encontram uma lista de pontos característicos que minimizam $L(H) + L(D|H)$ de uma trajetória. A trajetória é, então, particionada em segmentos levando em conta tais pontos característicos.

A terceira categoria de métodos são os baseados na semântica dos pontos nas trajetórias. Como mostrado na Figura 38.d, uma trajetória pode ser dividida em segmentos, $p_1 \rightarrow p_2 \rightarrow p_3$ e $p_8 \rightarrow p_9$, baseado nos pontos de permanência que ela contém. A remoção dos pontos de permanência depende da aplicação. Por exemplo, no caso de estimação da velocidade, os pontos de permanência gerados pelos taxis, quando eles estiveram parados esperando por passageiros, deveriam ser removidos. Por outro lado, se o nosso intuito é estimar a similaridade entre usuários (LI *et al.*, 2008b), poderíamos focar exclusivamente na sequência de pontos de permanência e descartar o resto de pontos. Outra forma de segmentação baseada no conteúdo semântico é dividir a trajetória em segmentos de diferentes modos de transporte, como de carro, bicicleta, e a pé. Um exemplo deste tipo de critério pode ser encontrado em Zheng *et al.* (2008a), Zheng *et al.* (2008b), Zheng *et al.* (2010). A principal suposição deste método é que as pessoas trocam regularmente seu modo de transporte. Desta forma, é possível distinguir entre pontos de movimentação de pedestre (*walk point*) e pontos de movimentação de não pedestres *non-walk points* baseado na velocidade e aceleração em um ponto. Assim, a trajetória é dividida alternadamente entre segmentos de pedestres *walk segments* e segmentos de não pedestres *non-walk segments*.

Por fim, quando um pedestre permanece dentro de uma certa área por um relativamente longo período de tempo, este acontecimento poderia indicar o fim de uma trajetória e o começo da próxima. Porém, o critério do uso exclusivo de intervalos de tempo mostra-se insuficiente para a nossa tarefa de reconstrução, devido à possibilidade de um usuário não se movimentar e gerar uma nova atualização depois de um tempo longo (e.g., após perder o sinal do GPS), o qual segmentaria erradamente a trajetória. Por outro lado, as trajetórias de pedestres nas áreas urbanas estão restritas à rede viária e, consequentemente, os possíveis pontos de inflexão não teriam algum significado relevante, pois as trajetórias nas ruas, becos e outros caminhos de pedestres, dependeriam da arquitetura urbanística e não da vontade do pedestre. Portanto, com intuito de refletir a real intenção do pedestre o melhor possível, segmentamos os *logs* GPS em trajetórias efetivas utilizando o critério de pontos de permanência. Assim, para cada trajetória formada pelos pontos dos *logs* GPS, detectamos seus pontos de permanência e os removemos. Dessa forma, finalmente, a trajetória original é naturalmente dividida em várias viagens ou rotas representadas como sub-trajetórias.

5.2.3 Map-matching

Map matching é o processo de transformar uma sequência de pontos GPS (latitude, longitude) em uma sequência de segmentos de rua. Este tipo de tarefas foi focado principalmente nas trajetórias de veículos e tem aplicações importantes na avaliação do fluxo de trânsito, navegação do veículo, predição do destino do veículo e detecção da rota mais frequente entre uma origem e um destino. No obstante, devido à inclusão de informações de caminhos exclusivamente transitados por pedestres, como becos e trilhas, nas redes viárias, por meio de mapas colaborativos, é possível transformar as sequências de pontos GPS de pedestres em sequências de becos, aleias, e, também, segmentos de rua, dado que o pedestre transita também por elas caminhando pela calçada. Por outro lado, *Map matching* não é um problema fácil, dadas as estradas paralelas, viadutos, etc (KRUMM, 2011). Com intuito de ilustrar esta ideia, a Figura 39 permite apreciar o uso da técnica de *Map matching* sobre uma trajetória real. A sequência de pontos pretos descreve a trajetória original, enquanto a sequência de pontos vermelhos representa a trajetória resultante do processo de *Map matching*.

Figura 39 – Exemplo de Map matching

Fonte: Adaptada de Ochieng, Quddus e Noland (2003).

Têm-se principalmente duas abordagens para classificar os métodos de *Map matching*, segundo a informação adicional usada e, segundo a extensão dos pontos de amostragem considerados nas trajetórias. O primeiro tipo, segundo a informação adicional usada, compreende basicamente quatro grupos: geométricos (GREENFELD, 2002), topológicos (CHEN *et al.*, 2003; YIN; WOLFSON, 2004), probabilísticos (OCHIENG; QUDDUS; NOLAND, 2003; PINK; HUMMEL, 2008; QUDDUS; NOLAND; OCHIENG, 2006), e outras técnicas mais avançadas (LOU *et al.*, 2009; NEWSON; KRUMM, 2009; YUAN *et al.*, 2010). Os algoritmos geométricos de *Map matching* consideram a forma das conexões individuais na rede viária. Por exemplo, fazendo uma correspondência entre um ponto GPS e rua mais próxima. Por outro lado, os algoritmos topológicos focam na conectividade da rede viária. Os algoritmos deste tipo com maior destaque são aqueles que usam a distância *Fréchet* para medir o ajuste entre uma sequência de pontos GPS e uma sequência de segmentos de ruas (BRAKATSOULAS *et al.*, 2005). Um

problema frequente a enfrentar é o ruído e a baixa taxa de amostragem das trajetórias. Perante esse problema, os algoritmos probabilísticos (OCHIENG; QUDDUS; NOLAND, 2003; PINK; HUMMEL, 2008; QUDDUS; NOLAND; OCHIENG, 2006) consideram múltiplos caminhos possíveis através da rede viária para encontrar o melhor. Finalmente, os algoritmos mais avançados de *Map matching* têm surgido recentemente, e levam em conta tanto a topologia da rede viária quanto o ruído nos dados da trajetória (LOU *et al.*, 2009; NEWSON; KRUMM, 2009; YUAN *et al.*, 2010). Estes algoritmos encontram uma sequência de segmentos de rua que, simultaneamente, aproximam-se aos dados ruidosos da trajetória, e formam uma rota razoável através da rede viária. Por fim, segundo a extensão dos pontos de amostragem considerados, os métodos de *Map matching* podem ser classificados em duas categorias: métodos locais ou incrementais, e métodos globais. Os métodos locais ou incrementais (CIVILIS; JENSEN; PAKALNIS, 2005; CHAWATHE, 2007) seguem a estratégia gulosa de estender sequencialmente a solução de uma porção já resolvida, i.e., uma subsequência de pontos GPS cuja sequência de segmentos de rua correspondente foi encontrada. Estes métodos tentam encontrar um ótimo local baseado na similaridade de distância e orientação. Adicionalmente, os métodos locais ou incrementais são executados eficientemente e são frequentemente adotados em aplicações online. Porém, quando a taxa de amostragem de uma trajetória é baixa, a acurácia do *matching* decresce. Em contraste, os algoritmos globais (ALT *et al.*, 2003; BRAKATSOULAS *et al.*, 2005) buscam fazer uma correspondência completa com uma rede viária. Uma forma de fazer isto é levando em conta os predecessores e sucessores de um ponto. Além disso, os algoritmos globais são mais acurados, porém, menos eficientes do que os métodos locais. Assim, estes métodos são usualmente utilizados em tarefas *offline* como mineração de padrões frequentes em trajetórias, onde as trajetórias completas já têm sido geradas com antecedência.

Finalmente, no presente trabalho, utilizamos a técnica avançada apresentada em Newson e Krumm (2009), a qual leva em conta tanto a topologia da rede viária quanto o ruído presente nas trajetórias. Este algoritmo, basicamente, calcula um número de candidatos para um possível *matching* dentro de um certo raio ao redor de cada ponto GPS. Em seguida, o algoritmo de Viterbi (FORNEY, 1973) é utilizado para calcular a sequência de candidatos a *matching* mais provável. Desta forma, tanto as distâncias entre pontos GPS e os candidatos a *matching*, quanto as distâncias através da rede viária entre candidatos de *matching* consecutivos são levados em conta.

5.2.4 Inferência

Como mostrado na Seção 5.1.2, existem muitas técnicas de reconstrução de trajetórias que utilizam diversas estratégias e informações. Dentre este leque de possíveis técnicas, escolhemos aquelas que utilizam informação da rede viária. Esta decisão deve-se principalmente ao fato que, como mencionado já anteriormente, as tecnologias de mapas colaborativos, ao adicionar os caminhos de pedestres dentro da rede viária, possibilitam o uso deste tipo de técnicas no contexto

de pedestres. Este fato fornece uma possibilidade muito interessante e ainda não explorada de reconstrução de trajetórias no caso de pedestres.

Assim, dentre técnicas de reconstrução de trajetórias já mencionadas que utilizam as informações da rede viária, tem-se o InferTra (BANERJEE; RANU; RAGHAVAN, 2014). Esta técnica é a de maior destaque dentre seu tipo, segundo a bibliografia. Desta forma, neste trabalho, utiliza-se esta técnica na etapa de inferência do processo de reconstrução de trajetórias de pedestres, a qual é descrita detalhadamente em seguida. O InferTra possui duas etapas, uma etapa de aprendizado *offline* para construir o Modelo de Mobilidade na Rede (MMR), e um algoritmo de inferência online utilizando tal MMR.

Na Figura 40, pode-se observar o fluxo do processo de inferência levado a cabo pelo InferTra.

Figura 40 – Fluxo de trabalho do InferTra

Fonte: Adaptada de Banerjee, Ranu e Raghavan (2014).

A. Aprendizado do Modelo de Mobilidade na Rede

O MMR é um modelo generativo para trajetórias cuja tarefa é conectar um conjunto de observações (ou nós) sem comprometer os padrões de mobilidade das trajetórias históricas. O conjunto de todas as possíveis rotas entre um nó origem e o destino em uma rede viária pode ser muito grande. Porém, na maioria das trajetórias, pode ser identificado um conjunto limitado de ruas e caminhos para os quais parece existir certa afinidade. Estes padrões espaciais e temporais fornecem uma rica caracterização dos movimentos nas trajetórias, a qual é aprendida pelo MMR através do uso de um Modelo de Markov de ordem superior (RAFTERY, 1985). O MMR é aprendido de duas fontes de dados de entrada: uma rede viária $G(V, E)$, e uma base de dados de trajetórias históricas $\mathbb{H}$. Em um processo de Markov, os elementos do sistema transitam de um estado ao outro segundo a história precedente. O tamanho da história é conhecida como a ordem do modelo (RABINER; JUANG, 1986). Por exemplo, no processo de Markov de primeira ordem, a transição de um estado de um elemento é influenciado unicamente pelo seu estado atual. No MMR, o espaço de estados é definido pelos nós V de G , e as transições são feitas apenas

através das arestas. As probabilidades associadas às transições são aprendidas da base de dados de trajetórias históricas.

O objetivo do MMR é entender a probabilidade de que um segmento de rua seja transitado segundo a história recente de um objeto e a hora do dia. Estas probabilidades são modeladas como transições de estado. Desta forma, o conceito de *m-history sequence* de um nó $v \in V$ é definido, onde m é a ordem do processo Markoviano no MMR.

Definição 10 (*m-history sequence*). A *m-history sequence* de um nó v é um caminho $H = \{v_1, \dots, v_p\}$ através de p nós conectados de G tal que $p \leq m$ e $v_p = v$.

Definição 11 (Contenção espaço-temporal). H é contido na trajetória T no tempo t , se $\forall i, 1 \leq i \leq |H|, H.v_i = T.s_{i+a}.v$, onde $\exists a, 0 \leq a \leq (|T| - |H|)$. Além disso, $|t - T.s_{|H|+a}.t| \leq \delta$. Esta relação é representada usando $H \in_t T$.

De forma simples, $H \in_t T$ se H for uma subsequência de T e o destino final em H for atingido dentro de um tempo δ desde t . Incluir a hora do dia no modelo permite capturar as periodicidades inerentes no padrões de mobilidade. δ é um parâmetro cujo valor é fornecido pelo usuário.

Definição 12 (Afinidade de arestas). A afinidade $\alpha(e, H, t)$ de uma aresta e no tempo t com respeito a uma *m-history sequence* H de $e.p_1$ é a probabilidade de e ser transitada por uma trajetória T , dado que $H \in_t T$. Formalmente:

$$\begin{aligned} \alpha(e, H, t) &= \max\{P(H \cup \{e.p_2\} \in_t T | H \in_t T), \varepsilon\} \\ &= \max\left\{ \frac{|\{H \cup \{e.p_2\} \in_t T | T \in \mathbb{H}\}|}{|\{\forall e' \in E, e'.p_1 = e.p_1, H \cup \{e'\} \in T | T \in \mathbb{H}\}|}, \varepsilon \right\} \quad \text{onde} \quad \varepsilon \approx 0. \end{aligned}$$

Em síntese, $\alpha(e, H, t)$ quantifica a probabilidade de transição de $e.p_1$ a $e.p_2$ baseado na *m-history sequence* H e a estampa de tempo t . O MMR é, então, construído seguindo esse procedimento.

B. Inferência utilizando o Modelo de Mobilidade na Rede

Como mencionado anteriormente, o objetivo do Infertra é gerar uma trajetória incerta U que permita modelar a incerteza envolvida no processo de reconstrução a partir de um conjunto de observações O . Desta forma, esta trajetória incerta permite representar todas as possíveis trajetórias que têm a primeira e última observação do O como origem e destino respectivamente.

No InferTra, U é modelada como uma distribuição multivariada de suas arestas, i.e., $p(U|O) = p(e_1, \dots, e_m|O)$, onde $\{e_1, \dots, e_m\} \in U$. Assim, o peso das arestas são as distribuições

marginais, as quais são necessárias para aprender das observações conjuntamente com os dados do histórico de trajetórias. Desta forma, para cada aresta $e \in U$,

$$P(e|O) = \sum_{\forall T \in \mathbb{T}} P(T|O) \quad (5.1)$$

onde $\mathbb{T}$ é o conjunto de todas as trajetórias restritas à rede conectando a sequência de nós observados em O , e que, além disso, contem a aresta e . Assim, se calcularmos $p(T|O)$ para todas as trajetórias em U , então, podemos identificar todas as arestas e com $p(e|O) > 0$, e, portanto, calcular a trajetória incerta correspondente a O .

Por outro lado, o MMR modela a probabilidade de um segmento de rua ser transitado baseado na sua *m-history sequence* e a hora do dia. Executando *caminhadas aleatórias semi-supervisadas com reinício* (*Randow Walk with Restart*) (TONG; FALOUTSOS; PAN, 2008), as quais são definidas em seguida, no MMR, calcula-se a condicional de uma aresta ser transitada sem ter que realizar uma inspeção de trajetórias. Por meio do Gibbs Sampling (GS), repetidamente realiza-se amostragem destas condicionais, e computa-se a distribuição condicional conjunta $p(U|O)$.

Caminhadas Aleatórias Semi-supervisadas com Reinício: Uma caminhada aleatória com reinício (CAR) simula a trajetória de um caminhante aleatório que começa em um nó origem e , iterativamente, salta de um nó até um vizinho. A probabilidade de saltar a um nó vizinho é proporcional ao peso da aresta correspondente. De forma similar, com uma probabilidade de retorno ou reinício r , o caminhante salta de volta ao nó origem. Quando o caminhante retornar ao nó origem, uma nova caminhada é iniciada.

Geração de trajetórias: Geram-se trajetórias do MMR através da CAR semi-supervisada. Cada trajetória gerada representa uma amostra do espaço da distribuição conjunta de $p(U|O)$. Gerando estas trajetórias repetidamente, a trajetória incerta U é inferida.

Dado um conjunto de observações, cada par de observações consecutivas, s_i e s_{i+1} , é tomado e uma CAR é iniciada a partir de $s_i.v$. Para cada salto, como com uma RWR normal, o novo destino é selecionado baseado nas afinidades de todas arestas com origem em $s_i.v$. No entanto, na modelagem dos padrões de mobilidade de uma trajetória, unicamente consideram-se as arestas que não geram um ciclo. A ideia subjacente é que cada transição em uma trajetória aproxima-se ao destino. Com o intuito de evitar ciclos, é mantido um conjunto S que armazena cada nó visitado durante a caminhada. Assim, dados um nó atual v e sua *m-history sequence* em um tempo t , as chances de selecionar uma aresta incidente e é proporcional à sua afinidade. A probabilidade de transição através de uma aresta e é representada matematicamente como:

$$p(e) = \begin{cases} (1-r) \frac{\alpha(e,H,t)}{\sum_{e' \in \bar{E}} \alpha(e',H,t)} & \text{se } e \in \bar{E} \\ 0 & \text{caso contrário} \end{cases}$$

$\bar{E} = \{e.p_1 = v, e.p_2 \notin \{e'.p_1 \cup e'.p_2 | \forall e' \in S\} | e \in E\}$ é o conjunto de arestas incidentes que não geram um ciclo. Adicionalmente, v é o nó atual, e t é o tempo atual.

Algoritmo 1 – Algoritmo de InferTra

```

1: procedimento AMOSTRATRAJETORIA( $s_1, s_2$ )
2:    $atual \leftarrow s_1.v$ 
3:    $t \leftarrow s_1.t$ 
4:    $r \leftarrow$  reiniciar probabilidade selecionada segundo a Equação 5.2
5:    $S \leftarrow \emptyset$ 
6:   enquanto  $atual \neq s_2.v$  faça
7:      $p \leftarrow$  tomar amostra uniformemente de  $[0,1]$ 
8:     se  $p \leq r$  então
9:        $atual \leftarrow s_1.v$ 
10:       $S \leftarrow \emptyset$ 
11:     senão
12:        $\bar{E} \leftarrow \{e.p_1 = atual, e \in E \text{ não gera um ciclo na caminhada atual}\}$ 
13:        $H \leftarrow$  obter  $m$ -history sequence de  $atual$  de  $S$ 
14:        $e \leftarrow$  selecionar aresta de  $\bar{E}$  proporcional a  $\frac{\alpha(e,H,t)}{\sum_{e' \in \bar{E}} \alpha(e',H,t)}$ 
15:        $t \leftarrow t + velocidade(e,t) * comprimento(e)$ 
16:        $atual \leftarrow e.p_2$ 
17:       se  $atual = s_2.v$  e  $p \leq \tau$  então
18:          $atual \leftarrow s_1.v$ 
19:          $S \leftarrow \emptyset$ 
20:       senão
21:          $S \leftarrow S \cup \{e\}$ 
22:       fim se
23:     fim se
24:   fim enquanto
25:   retorna  $S$ 
26: fim procedimento

```

Captura do viés dos destinos: Uma transição através de uma aresta até e na caminhada aleatória não calcula a probabilidade condicional de e' . A probabilidade condicional é equivalente a calcular a probabilidade de transição no subconjunto de trajetórias históricas que passam pelo destino s_2 . Porém, este processo requer uma inspeção exaustiva das trajetórias, o que implica um alto custo computacional. Assim, utiliza-se uma aproximação para alocar o viés dos destinos na CAR mediante o uso da probabilidade de reinício r . Na [Equação 5.2](#), apenas $(1 - r)$ da massa de probabilidade é distribuída entre as transições das arestas. Como em uma CAR normal, existe a chance de saltar de volta ao nó origem com uma probabilidade r . Se o reinício ocorrer, semanticamente, considera-se que a caminhada foi feita em uma direção errada e, então, é descartada reinicializando S com um conjunto vazio. Uma nova caminhada é inicializada com o intuito de atingir o destino $s_{i+1}.v$ e, finalmente, apenas aquelas caminhadas que atinjam $s_{i+1}.v$ com sucesso são armazenadas. Consequentemente, o viés dos destinos é alocado em todas as transições das arestas. Em uma CAR tradicional, a probabilidade de reinício é estática dado que o objetivo é calcular a proximidade na rede ao nó origem. Neste caso, o objetivo é atingir o nó destino $s_{i+1}.v$ dentro da duração esperada $X_t = s_{i+1}.t - s_i.t$. Caso contrário, é provável que o caminhante esteja no caminho errado. Portanto, deve-se dar tempo suficiente ao caminhante

para atingir o objetivo satisfatoriamente. Porém, se o caminhante não tiver sucesso atingindo seu destino dentro de um tempo X_t , então, as chances de reiniciar o caminho deveriam ser levadas em conta. Com intuito de modelar estes requerimentos, define-se a probabilidade de reinício r da seguinte forma:

$$r = 1 - \frac{1}{e^{\frac{\max\{0, t - X_t\}}{X_t}}} \quad (5.2)$$

onde t é o tempo utilizado na caminhada atual. Desta forma, enquanto o tempo transcorrido na caminhada for menor do que o tempo esperado X_t , a probabilidade de reinício é 0. Quando t ultrapassar X_t , a probabilidade de reinício começa a crescer exponencialmente. Em síntese, geram-se trajetórias amostrando repetidamente das condicionais expressadas por meio das probabilidades de transição das arestas e o viés dos destinos. Cada trajetória gerada é um ponto no espaço da distribuição conjunta $p(U|O)$. Dessa forma, U é definido sobre todas as arestas que são amostradas pelo menos uma vez. Assim, finalmente, a CAR termina quando a distribuição conjunta de U converge.

Todo o processo de inferência explicado encontra-se sumarizado no Algoritmo 1. Ele ilustra de forma concisa e eficiente o funcionamento desta técnica.

5.3 Considerações finais

Devido à natureza continua do movimento, as trajetórias de qualquer objeto sempre têm certo grau de incerteza nas suas representações. A taxa de amostragem é o que define o quão incerto é uma trajetória, i.e., quão longo será o espaço entre uma atualização de localização e a próxima. Apesar que algumas aplicações precisam da existência dessa incerteza para seus fins (e.g., aplicações de privacidade), outras, que visam estudar o comportamento e encontrar os padrões subjacentes nas trajetórias, precisam reduzir tal incerteza o máximo possível. Desta forma, esta ingente necessidade deu origem ao surgimento de diversas técnicas cujo objetivo é lidar com a incerteza. Algumas destas técnicas tentam modelar a incerteza e adotá-la como uma variável a mais do sistema, enquanto outras visam reduzi-la. Estas últimas técnicas são as de inferência de trajetórias. Dentre estas técnicas encontram-se aquelas que se focam nas trajetórias geradas na rede viária. Estas diferenciam-se das técnicas de *Map matching*, pois as primeiras utilizam outras trajetórias para a inferência e podem lidar com maiores intervalos de amostragem, enquanto as últimas apenas levam em conta a geometria da trajetória e a topologia da rede viária, e não são capazes de lidar com intervalos de amostragem muito longos. Por outro lado, tem-se o outro tipo de técnicas de inferência que trabalham com trajetórias não restritas às redes viárias. Estas técnicas geram um espaço comumente representado por uma grade, onde as trajetórias ficam modeladas como uma sequência de blocos espaciais e estabelecem-se certos critérios para encontrar as trajetórias mais relevantes para certos pontos a fim de realizar a inferência. A utilidade deste tipo de técnicas tem lugar nas áreas não urbanizadas, nas trajetórias de animais e

de alguns fenômenos naturais como furacões. Por outro lado, explicou-se que a tendência da adição de informações de caminhos exclusivamente transitados por pedestres como becos, aleais e trilhas à rede viária por meio dos mapas colaborativos, possibilitam a aplicação da inferência de trajetórias restritas à rede viária no contexto de pedestres, um cenário muito pouco explorado.

Adicionalmente, foi apresentado um *framework* explicando o processo de reconstrução de trajetórias, onde teve-se em conta as etapas de segmentação, *Map-matching* e a de inferência propriamente dita. Previamente foram explicados conceitos comuns a estas etapas. Além disso, a definição e estado da arte de cada tarefa presente nas diferentes etapas do *framework* foram explicadas.

Finalmente, definiu-se a técnica a utilizar em cada etapa: segmentação baseado em pontos de permanência, *Map-matching* avançado proposto por [Newson e Krumm \(2009\)](#), e inferência de trajetórias restritas à rede viária usando o Infertra ([BANERJEE; RANU; RAGHAVAN, 2014](#)). Esta técnica de inferência foi explicada detalhadamente na subseção correspondente.

AVALIAÇÃO EXPERIMENTAL

Neste capítulo, são apresentados os experimentos realizados para a avaliação da reconstrução de trajetórias de pedestres, bem como a discussão dos resultados obtidos. Dessa maneira, são apresentados os conjuntos de dados utilizados, a metodologia de avaliação e a análise comparativa dos resultados obtidos pelos diferentes métodos e configurações.

6.1 Conjuntos de dados

Para validar o desempenho das propostas apresentadas neste trabalho, foram realizados experimentos com as bases de dados de duas redes sociais: o RadrPlus, criada dentro do projeto, e o Geolife (ZHENG *et al.*, 2009b; ZHENG *et al.*, 2008a; ZHENG; XIE; MA, 2010).

Como mencionado no Capítulo 1, o Geolife é uma das poucas bases de dados disponíveis que satisfazem os requerimentos necessários para nossos experimentos. Por outro lado, no Capítulo 4, a base de dados do RadrPlus foi apresentada e detalhada. Porém, é importante mencionar que, nos experimentos realizados, os dados utilizados do RadrPlus abrangem apenas os percursos feitos por pedestres e dentro da cidade de São Carlos, sobretudo dentro do campus da USP e arredores. Além disso, devido à grande quantidade de dados necessários para o correto funcionamento dos métodos de reconstrução, tanto os dados capturados com o esquema de coleta contínuo quanto com o adaptativo foram consolidados em uma base, e utilizados nos experimentos.

O projeto *Geolife* é uma rede social geo-localizada desenvolvida por *Microsoft Research Asia*¹ cuja base de dados compreende as trajetórias de um total de 182 usuários equipados com celulares com *GPS* durante um período de quatro anos. Os dados foram coletados em 30 cidades da China e algumas cidades dos Estados Unidos e Europa; porém, a maioria deles foram obtidos na cidade de Beijing, China. Além disso, uma parte dos usuários etiquetaram as suas trajetórias

¹ <<https://www.microsoft.com/en-us/download/details.aspx>>

com a forma de locomoção (bicicleta, carro, avião, ônibus, trem e a pé), possibilitando, assim, uma maior variedade de análises a serem feitas. Do total de usuários, apenas 73 etiquetaram as suas trajetórias. Assim, de forma similar ao RadrPlus, dessas trajetórias, foram selecionadas apenas aquelas feitas por pedestres.

Aliás, como explicado no Capítulo 5, a etapa de *Map-matching* da reconstrução de trajetórias precisa do grafo da rede viária das cidades onde as atualizações de localização foram capturadas. Dessa forma, na base de dados do Geolife, com intuito de evitar o excessivo custo computacional de trabalhar com a informação da rede viária de todos os países envolvidos, e aproveitando que o 95% das trajetórias foram geradas na cidade de Beijing, apenas consideramos para nossos experimentos as trajetórias feitas dentro desta cidade.

Assim, uma vez aplicadas as restrições anteriormente mencionadas em ambas as bases de dados, na Tabela 3, são apresentadas as principais características de cada uma delas. Maiores detalhes da base de dados do Geolife podem ser encontrados no guia de usuário (ZHENG *et al.*, 2011).

Tabela 3 – Atributos das bases de dados utilizadas nos experimentos

	Geolife	RadrPlus
Número de usuários	73	15
Meses de coleta	48	9
Número de trajetórias	3601	116
Comprimento médio de trajetória (pp)	13.78	17.31
Duração média (mm)	13.65	10.06
Intervalo de amostragem (ss)	1-5	10-45

Note que, o comprimento de uma trajetória é calculado pelo número de pontos (pp) que a conformam. Lembramos também que, cada ponto contém uma estampa temporal ou *timestamp* e as coordenadas de localização (latitude, longitude).

6.2 Metodologia experimental

O objetivo principal dos experimentos realizados no presente capítulo é estudar o uso das informações da rede viária na reconstrução das trajetórias de pedestres dado seu desempenho sobre os dados das redes geo-localizadas utilizadas. Para esse fim, foram adotadas uma série de estratégias, métricas e técnicas.

6.2.1 Pre-processamento de trajetórias

Como mencionado na Capítulo 5, tem-se duas etapas prévias à etapa de inferência: a segmentação e o *Map-matching*.

Na etapa de segmentação, como explicado, também, no Capítulo 5, decidiu-se utilizar o critério de pontos de permanência (*stay points*) para dividir as trajetórias originais em sub-trajetórias significativas. Para isso, foram testadas três diferentes configurações de pontos de permanência com o intuito de estudar a relação da forma de segmentação com os resultados da reconstrução. Os parâmetros utilizados nesta etapa alteram o número, comprimento e duração das trajetórias, como pode-se observar na Tabela 4.

Tabela 4 – Configurações da segmentação de trajetórias

	Tolerância Média	Tolerância Alta	Tolerância Baixa
Distância máxima (m)	300	600	150
Tempo máximo (mm)	10	20	5

Fonte: Dados da pesquisa.

Cada uma destas configurações de segmentação baseados em pontos de permanência foram denominados como tolerâncias. Assim, a configuração de Tolerância Média (TM) apresenta uma distância mínima de aproximadamente o tamanho de três quarterões, 300 metros (YEANG *et al.*, 2000; CENTER *et al.*, 2012; HARRIS; AECOM; HINTZ, 2008). Por outro lado, o valor do tempo entre trajetórias foi concebido sob a ideia do pedestre poder fazer pequenas paradas de até 10 minutos (e.g., uma conversa, esperar a mudança do sinal do semáforo) e continuar com seu percurso inicial. Finalmente, a configuração de Tolerância Alta (TA) duplica estes valores de distância e tempo máximos, enquanto a de Tolerância Baixa (TB) os reduz à metade.

Aliás, cabe mencionar que os valores das três últimas variáveis na Tabela 3 (número de trajetórias, comprimento médio e duração média) foram obtidos utilizando a configuração de tolerância média na etapa de segmentação.

Por outro lado, na etapa do *Map matching*, utilizou-se o código fonte da ferramenta livre Graphhopper² para aplicar o algoritmo de *Map matching* avançado descrito em Newson e Krumm (2009), como mencionado no Capítulo 5. Além disso, para fornecer o grafo da rede viária das cidades de Beijing e São Carlos, no processo de *Map matching* das trajetórias do Geolife e RadrPlus respectivamente, foi utilizada a plataforma OpenStreetMap³. Desta plataforma, baixou-se os grafos das redes viárias de ambas as cidades, e, finalmente, utilizou-se o Graphhopper para executar a tarefa de *Map matching* sobre as trajetórias resultantes do processo de segmentação.

6.2.2 Experimentação

No presente trabalho, foram testados dois métodos, o InferTra (BANERJEE; RANU; RAGHAVAN, 2014) e o *Shortest Path*. O primeiro deles utiliza as informações da rede viária e o histórico de localização das pessoas como entrada dentro do processo de reconstrução de trajetórias. Além disso, segundo experimentos realizados pelos autores, este método ultrapassa os

² <<https://www.graphhopper.com.org>>

³ <<https://www.openstreetmap.org>>

outros métodos desta categoria. Assim, consideramos apenas este método dentro desta categoria. Por outro lado, o *Shortest Path* leva em conta apenas a estrutura da rede viária e não o histórico de localizações das pessoas. Dessa maneira, este método limita-se a encontrar o caminho mais curto entre os pontos inicial e final da lacuna (intervalo de amostragem) que deseja-se preencher ou inferir.

Desta forma, a seguir, descrevemos o processo de experimentação tomando como referência a Figura 41.

Figura 41 – Fluxo do processo de experimentação da reconstrução de trajetórias.

Fonte: Elaborada pelo autor.

Para cada conjunto de trajetórias da Tabela 5 e para valores de intervalo de amostragem de 1 até 10 minutos é utilizado um esquema de validação cruzada em 10 *folds* como observado na Figura 41. Uma vez escolhido o intervalo de amostragem, procede-se a trabalhar com cada uma das iterações da validação cruzada. Para isso, a base de trajetórias foi previamente dividida em 10 conjuntos disjuntos de trajetórias restritas à rede viária após o processo de *Map-matching*, como descrito no Capítulo 5. Assim, para cada iteração, têm-se o conjunto de teste e o conjunto de treinamento. O conjunto de teste é utilizado apenas pelo InferTra para criar o Modelo de Mobilidade na Rede (MMR) que será usado no seu algoritmo de inferência. Por outro lado, no conjunto de teste, é calculado o comprimento médio do conjunto, entendendo-se como comprimento de uma trajetória o número de pontos que a compõem. Tal comprimento médio $\bar{c}$ é calculado da seguinte forma:

$$\bar{c} = \frac{\sum_{i=1}^n Ti.c}{n}$$

Uma vez calculado o comprimento médio $\bar{c}$ das trajetórias do conjunto de teste, para cada uma destas trajetórias são criadas um número de réplicas igual a este comprimento médio. A Figura 41 permite observar que para a primeira trajetória do conjunto de teste, foram criadas $\bar{c}$ trajetórias. Desta forma, após isso, cada uma destas trajetórias é modificada, criando um segmento incerto ou lacuna dentro de cada uma delas cujo espaço de tempo seja igual ao intervalo de amostragem (IA) com o qual se está trabalhando nesse momento. A escolha da posição do ponto inicial dessa lacuna é feita aleatoriamente. Na Figura 41, pode-se observar que os índices na trajetória correspondentes ao ponto inicial e final da lacuna são denotados como i e j respectivamente. Assim, teríamos que $i_{T11} = \text{random}(1, T11.c - IA)$, e $j_{T11} = i_{T11} + IA$. Onde $T11.c$ é o comprimento ou índice da posição do ponto final na trajetória. Finalmente, para cada uma das $\bar{c}$ trajetórias, é realizado a inferência a fim de reconstruir ou preencher a lacuna criada. O algoritmo do InferTra é o único que utiliza o MMR, pois o Shortest Path utiliza apenas o grafo da rede viária. Desta forma, o InferTra é o único que faz uso do conjunto de treinamento. Voltando à Figura 41, observamos que a trajetória reconstruída é $T11^*$. Assim, em seguida, procede-se a calcular o valor do F-score, $F_{T11} = F - \text{score}(T11^*, T1)$. O *ground truth*, neste caso, seria a trajetória T1, a qual foi tomada do conjunto de teste e deu origem à lista de trajetórias incertas. Uma vez feito isso, vamos explicar de forma geral o cálculo do F-score final para cada valor do intervalo de amostragem F_{IA} .

$$F_{Tpq} = F - \text{score}(Tpq^*, Tp)$$

$$F_{Tp} = \frac{\sum_{i=1}^{\bar{c}} F_{Tpi}}{\bar{c}}$$

$$F_{iter} = \frac{\sum_{i=1}^n F_{Ti}}{n}$$

$$F_{IA} = \frac{\sum_{i=1}^{10} F_{iter_i}}{10}$$

Por outro lado, uma questão importante a mencionar é que, dado que o InferTra gera como resultado um grafo ponderado, então a medida de F-score neste caso foi adaptada como indicado em Banerjee, Ranu e Raghavan (2014). No caso do Shortest Path o F-score é a medida comum, que para o caso de trajetórias, seria a comparação das arestas presentes na trajetória original e na reconstruída.

Como sabemos, o F-score é calculado como uma média ponderada entre o *precision* e o *recall*, onde o melhor desempenho corresponde a um valor de 1 e o pior corresponde a 0. Porém, em uma trajetória incerta, como no caso do InferTra, as arestas tem alocadas probabilidades, assim, é preciso modificar as definições de *precision* e *recall* para este caso.

Assim, seja $U_p = (V_p, E_p)$ o grafo ponderado resultante da reconstrução de uma trajetória usando o InferTra, e $T = (V, E)$, a trajetória *ground truth*. Adicionalmente, seja $E_c = E \cap E_p$ o

conjunto de arestas comuns entre T e U_p . Então, tem-se:

$$recall = \frac{\sum_{e \in E_c} peso_aresta(e)}{|E|}$$

$$precision = \frac{\sum_{e \in E_c} peso_aresta(e)}{\sum_{e \in E_p} peso_aresta(e)}$$

$$F - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (6.1)$$

Como observa-se nas anteriores equações, a definição do F-score é a única que mantém-se sem alteração.

A Figura 42 ilustra o cálculo destas medidas levando em contas as anteriores equações definidas, e, também, para o caso do Shortest Path, onde as medidas não foram alteradas.

Figura 42 – Precision, Recall e F-score de Shortest Path e InferTra para duas trajetórias usadas como ground truth.

Fonte: Adaptada de Banerjee, Ranu e Raghavan (2014).

6.3 Resultados e análise

Na presente seção, apresentam-se os resultados dos experimentos de reconstrução de trajetórias realizados levando em conta as diferentes configurações de tolerâncias na fase de segmentação nas duas bases de dados do RadrPlus e Geolife. Adicionalmente, analisam-se estes resultados e estabelecem-se comparações entre as técnicas de reconstrução e entre as configurações de tolerâncias.

6.3.1 Resultados da Segmentação

Na Tabela 5, pode-se observar as características dos conjuntos de trajetórias após a fase segmentação, utilizando as três configurações propostas: Tolerância Baixa (TB), Tolerância Média (TM) e Tolerância Alta (TA).

É importante lembrar que as técnicas de reconstrução precisam de um conjunto de trajetórias para serem aplicadas. Desta forma, as atualizações de localização das bases de dados (*logs*) são agrupadas para formar distintas trajetórias. Uma trajetória é composta por uma sequência de atualizações de localização ordenadas cronologicamente. O começo e o fim de uma trajetória são definidos pela configuração de tolerância na fase de segmentação, como explicado no Capítulo 5.

Tabela 5 – Atributos dos conjuntos de trajetórias após segmentação.

	Nº trajetórias	Comprimento médio (pp)	Duração média (s)
RadrPlus (TB)	76	14.12	487.48
RadrPlus (TM)	116	17.31	604.15
RadrPlus (TA)	155	21.06	1045.61
Geolife (TB)	3895	13.06	721.74
Geolife (TM)	3601	13.71	819.57
Geolife (TA)	3134	13.79	877.2

Fonte: Dados da pesquisa.

Um processo de coleta com uma alta e pouco variável taxa de amostragem, e com pouca presença de interrupções ou lacunas, deveria, teoricamente, apresentar uma maior quantidade de trajetórias, de menor comprimento e duração, perante uma baixa tolerância. Este comportamento é evidenciado no caso do Geolife, como pode ser visto na Tabela 5. Em contraste, como esperado, uma alta tolerância produz uma menor quantidade de trajetórias de menor comprimento e duração.

Por outro lado, diferentemente do caso dos dados do Geolife, os dados do RadrPlus apresentam uma menor quantidade de trajetórias, de menores comprimento e duração, perante uma baixa tolerância. Adicionalmente, uma alta tolerância produz uma maior quantidade de trajetórias com maiores comprimento e duração.

Este último comportamento pode-se dever principalmente ao fato que a base de dados do RadrPlus está composta maioritariamente pelos dados obtidos com o esquema de coleta adaptativo, como visto no Capítulo 4. Tal esquema tem taxas maiores e mais variadas das usadas no Geolife. Além disso, um outro possível fator é que o comprimento mínimo que uma trajetória deve ter para ser considerada na base de trajetórias é de 10 pontos ou atualizações de localização. Assim, devido a que muitos dos dados de localização do RadrPlus foram capturados dentro do campus da USP e arredores, é possível que estes percursos tenham sido muito curtos para serem levados em conta (*e.g.*, *um percurso feito de um edifício a um outro*). Por fim, outra diferença importante entre ambas as bases de dados que poderia ter influenciado neste comportamento é que a presença de edifícios poderia ter obstruído o sinal do *GPS* em algumas partes dos percursos, criando, assim, lacunas que tenham segmentado os percursos em outros mais curtos. Desta forma, os verdadeiros percursos poderiam ser detectados apenas perante uma maior tolerância. Um outro detalhe a considerar é que, diferentemente do RadrPlus, a maioria dos dados do Geolife

foram capturados nas ruas e avenidas, onde o sinal do GPS é muito melhor.

6.3.2 Resultados de Reconstrução

Na Figura 43, pode-se observar os valores do F-score obtidos pelos algoritmos InferTra e Shortest Path para o conjunto de trajetórias geradas com os dados do RadrPlus, utilizando as três diferentes configurações de segmentação ou tolerâncias. Nesta figura, pode-se apreciar claramente o melhor desempenho do Shortest Path em todas as tolerâncias e para todos os intervalos de amostragem. Adicionalmente, pode-se observar um amplo margem de diferença entre ambas as técnicas, mas não é possível distinguir com clareza alguma diferença entre os três tipos de tolerância.

Figura 43 – Resultados da reconstrução de trajetórias com os dados do RadrPlus utilizando os algoritmos InferTra e Shortest Path. Cada ponto no gráfico indica o valor do F-score obtido com uma dada janela de tempo para as configurações de segmentação (a) Tolerância Alta, (b) Tolerância Média e (c) Tolerância Baixa.

Fonte: Elaborada pelo autor.

Assim, com intuito de analisar estes resultados com maior detalhe, agrupamos as técnicas de inferência com as distintas tolerâncias a fim de estudar a diferença estatística entre estas combinações levando em conta os distintos intervalos de tempo. Para isso, é computado o teste de Friedman e, posteriormente, o valor da diferença crítica como ilustrado em Demšar (2006). Assim, finalmente, temos os grupos: InferTra com Tolerância Alta (IT-TA), Média (IT-TM) e Baixa (IT-TB), bem como as respectivas combinações com o Shortest Path, SP-TA, SP-TM e SP-TB.

A diferença crítica é apresentada mediante um diagrama no qual o valor de tal diferença é visualizada mediante uma linha cinza acima de um eixo onde são indicados os valores resumidos dos resultados para as diferentes abordagens ordenados mediante um ranking elaborado segundo o valor do desempenho, neste caso o F-score. Adicionalmente, umas linhas em negrito indicam aquelas abordagens que *não* são estatisticamente diferentes, i.e, aquelas cuja diferença seja menor que o valor da diferença crítica.

Assim, a Figura 44 permite observar, em primeiro lugar, que, com uma diferença crítica de 2.38, como esperado, os resultados do InferTra e o Shortest Path são estatisticamente diferentes quando comparados com a mesma tolerância, i.e., $IT-TA \neq SP-TA$, $IT-TM \neq SP-TM$, e $IT-TB \neq SP-TB$. Desta forma, podemos afirmar que existe uma diferença estatisticamente significativa entre ambos os métodos de inferência favorecendo ao Shortest Path por sobre o InferTra. Em segundo lugar, pode-se observar, também, que as distintas tolerâncias não são estatisticamente diferentes em cada método. Porém, observa-se que uma alta tolerância no InferTra faz com que ele tenha um resultado estatisticamente similar com as configurações de tolerâncias média e baixa no Shortest Path. Este último resultado, poderia indicar uma possível relação entre uma alta tolerância e um melhor resultado de inferência.

Figura 44 – Diferença estatística do desempenho das inferências de trajetórias medido pelo F-score para cada método e nível de tolerância sobre a base de dados RadrPlus.

Fonte: Elaborada pelo autor.

Por outro lado, na Figura 45, que mostra os resultados das inferências de trajetórias na base de dados do Geolife, pode-se observar um resultado similar ao do RadrPlus. O Shortest Path ultrapassa em todos os intervalos de amostragem e em todas as configurações de tolerância ao InferTra. De forma similar a como efetuado na anterior análise, estudamos a significância estatística da diferença destes resultados procedendo de forma análoga do que com os dados do RadrPlus.

Pode-se observar na Figura 46, que, o valor da diferença crítica neste caso é de 2.38, igual que o do RadrPlus. Adicionalmente, também ao igual que com a base de dados do RadrPlus, existe uma diferença estatística entre os métodos InferTra e Shortest Path com as mesmas configurações de tolerância, favorecendo ao segundo sobre o primeiro. Por outra parte, estas configurações de tolerância, novamente, não resultam estatisticamente diferentes dentro de cada método, porém, diferentemente da anterior análise com os dados do RadrPlus, o InferTra com

Figura 45 – Resultados da reconstrução de trajetórias com os dados do Geolife utilizando os algoritmos InferTra e Shortest Path. Cada ponto no gráfico indica o valor do F-score obtido com uma dado intervalo de amostragem para as configurações de segmentação (a) Tolerância Alta, (b) Tolerância Média e (c) Tolerância Baixa.

Fonte: Elaborada pelo autor.

Figura 46 – Diferença estatística do desempenho das inferências de trajetórias medido pelo F-score para cada método e nível de tolerância sobre a base de dados Geolife.

Fonte: Elaborada pelo autor.

tolerâncias alta e média apresenta uma similaridade estatística com o Shortest Path com uma tolerância baixa. Novamente, este último resultado parece indicar a possibilidade que quanto maior a tolerância melhor o resultado.

Desta forma, a fim de estudar de forma mais rigorosa a hipótese que uma tolerância alta na segmentação favorece o resultado da reconstrução, na Tabela 6, mostramos o detalhe dos resultados dos experimentos, apresentados de tal forma que seja possível estabelecer uma melhor comparação entre as distintas configurações de tolerância.

A simples vista, a Tabela 6 mostra que a configuração de Tolerância Alta apresenta um

Tabela 6 – Valores de F-score dos resultados da reconstrução de trajetórias nas bases de dados RadrPlus e Geolife sob as configurações de segmentação Tolerância Alta (TA), Tolerância Média (TM) e Tolerância Baixa (TB) para diferentes intervalos de amostragem.

	Intervalo de amos- tragem	TB	TM	TA
RadrPlus (ShortestPath)	1	0.944	0.930	0.963
	2	0.846	0.852	0.903
	3	0.770	0.790	0.857
	4	0.796	0.771	0.800
	5	0.793	0.748	0.762
	6	0.746	0.689	0.732
	7	0.718	0.680	0.717
	8	0.639	0.636	0.700
	9	0.558	0.567	0.676
	10	0.578	0.558	0.673
Geolife (ShortestPath)	1	0.959	0.960	0.962
	2	0.902	0.909	0.907
	3	0.862	0.868	0.870
	4	0.818	0.835	0.837
	5	0.775	0.789	0.797
	6	0.754	0.782	0.783
	7	0.746	0.766	0.771
	8	0.721	0.751	0.761
	9	0.694	0.717	0.732
	10	0.671	0.705	0.727
RadrPlus (InferTra)	1	0.502	0.608	0.654
	2	0.510	0.481	0.593
	3	0.446	0.430	0.539
	4	0.467	0.404	0.507
	5	0.404	0.350	0.434
	6	0.298	0.323	0.411
	7	0.205	0.303	0.353
	8	0.197	0.237	0.315
	9	0.018	0.230	0.273
	10	0.005	0.154	0.231
Geolife (InferTra)	1	0.558	0.569	0.558
	2	0.484	0.494	0.490
	3	0.437	0.455	0.450
	4	0.395	0.421	0.424
	5	0.363	0.382	0.401
	6	0.342	0.371	0.386
	7	0.314	0.344	0.369
	8	0.281	0.326	0.346
	9	0.264	0.300	0.326
	10	0.250	0.285	0.309

melhor resultado para os diferentes valores dos intervalos de amostragem, técnicas de inferência e bases de dados. No entanto, com intuito de validar mais formalmente esta intuição, levamos em conta por cada configuração de tolerância os resultados de F-score de todos intervalos de amostragem, técnicas de inferência e bases de dados a fim de conhecer se existe alguma diferença estatística entre estas configurações.

Desta forma, na Figura 47, podemos confirmar nossa intuição ao verificar que a configuração de Tolerância Alta (TA) é estatisticamente diferente e superior das outras configurações levando em conta os resultados de todas as outras variáveis mencionadas no anterior parágrafo. A diferença crítica é de 0.5241 e tanto as configurações de baixa e média tolerância resultam não sendo estatisticamente diferentes. Assim, a tolerância alta é a única configuração que parece ter certa influência nos resultados da reconstrução de trajetórias.

Figura 47 – Diferença estatística do desempenho das inferências de trajetórias para cada nível de tolerância levando em conta todas as combinações de bases de dados, técnicas de inferência e intervalos de amostragem.

Fonte: Elaborada pelo autor.

6.4 Considerações finais

Neste capítulo, realizamos a reconstrução de trajetórias de pedestres utilizando a informação da rede viária. Três diferentes configurações de segmentação de trajetórias foram propostas a fim de analisar sua influência na reconstrução. Estas configurações foram estabelecidas tomando como base o conceito de tolerância, o qual faz referência aos pontos de permanência, cujo critério leva em conta tanto distância quanto o tempo transcorrido entre pontos de uma trajetória. Adicionalmente, duas técnicas do estado da arte foram testadas. Uma destas técnicas leva em conta a informação da rede viária e o histórico de trajetórias, enquanto a outra técnica, mais simples, unicamente leva em conta a rede viária como informação para fazer a reconstrução de trajetórias.

A análise empírica destas duas técnicas sobre duas bases de dados mostra que a técnica mais simples apresenta um melhor desempenho sob condições limitadas de dados do que o método mais complexo no contexto de reconstrução de trajetórias de pedestres. Além disso, a configuração de segmentação de alta tolerância proposta obteve melhores resultados de reconstrução de forma geral, levando em conta as distintas bases de dados, técnicas de inferência e intervalos de amostragem.

CONCLUSÕES

Este capítulo apresenta as conclusões e contribuições obtidas pela pesquisa. Adicionalmente, são discutidas as limitações do trabalho, assim como possíveis trabalhos futuros.

Observa-se que este trabalho teve dois desdobramentos fundamentais: um, de natureza tecnológica, cujo objetivo principal foi prover o grupo de pesquisa com uma plataforma de gerenciamento de redes sociais geo-localizadas (RadarPlus), cujas informações fossem disponibilizadas voluntariamente para fins de pesquisa. Tal processo conhecido como VGI (Volunteered Geographic Information) em poucos dias de uso rendeu uma pequena base de trajetórias que foi usada neste projeto e está disponível para outros pesquisadores. O segundo desdobramento, de natureza científica, teve como objetivo investigar uma das inúmeras possibilidades de pesquisa permitidas no contexto da plataforma RadarPlus. Neste caso a análise de técnicas de reconstrução de trajetórias de pedestre.

7.1 Conclusões e principais contribuições

A revisão bibliográfica permitiu perceber a pouca quantidade de trabalhos focados no estudo da reconstrução de trajetórias de pedestres. Apesar das dificuldades como a escassez de dados etiquetados e ainda mais de pedestres, muito do esforço na reconstrução de trajetórias tinha sido investido nos dados de trajetórias com uso de veículos principalmente. Além disso, o levantamento bibliográfico permitiu encontrar que a maioria de redes sociais geo-localizadas tinham foco no indivíduo, mas não existiam redes sociais geo-localizadas que focassem nas comunidades de usuários e muito menos que disponibilizaram informações diferentes das trajetórias dos mesmos, sem nenhum tipo de informação adicional como dados pessoais ou de relação com o resto de usuários.

Desta forma, foi implementado o RadarPlus, uma rede social geo-localizada com foco nas comunidades, a qual permite obter além das informações de trajetórias, informações pessoais

dos usuários e das comunidades a que pertencem. Este sistema permite uma nova experiência ao fornecer funcionalidades únicas orientadas às comunidades, assim como possibilitar o acesso a informação mais rica que pode ser utilizada nas pesquisas de mineração de redes sociais geo-localizadas e especialmente na mineração de trajetórias. Por outro lado, com intuito de endereçar o problema de mineração de trajetórias de pedestres, aproveitou-se a tendência nos mapas colaborativos de adicionar caminhos exclusivamente transitados por pedestres nos grafos das redes viárias para, desta forma, poder utilizar técnicas mais avançadas de reconstrução de trajetórias que precisam de tal informação.

Assim, depois de um processo de coleta de dados utilizando o RadrPlus e um serviço de coleta eficiente de dados de geolocalização, projetado como parte deste projeto, levou-se a cabo uma série de experimentos de reconstrução de trajetórias de pedestres. Cabe mencionar que, também, foi considerada uma das poucas bases de dados disponíveis que contém trajetórias de pedestres, o Geolife, a fim de obter resultados mais conclusivos. Desta forma, sobre os dados de ambas as bases de dados, RadrPlus e Geolife, aplicaram-se duas técnicas de reconstrução de trajetórias distintas, Infertra e o Shortest Path, o primeiro mais complexo que o segundo, levando em conta três configurações diferentes de segmentação, tolerância alta, média e baixa, a fim de estudar seus desempenhos e influencia nos resultados finais dos experimentos. Adicionalmente, cada técnica e configuração de segmentação foi testada usando intervalos de amostragem de 1 até 10 minutos.

Por fim, a partir dos resultados obtidos nos experimentos e apresentados no Capítulo 6, foi possível extrair as seguintes conclusões.

- Um método simples como o Shortest Path, que apenas leva em conta a estrutura da rede viária, obtém melhores resultados de reconstrução que um método mais complexo como o Infertra, o qual leva em conta os históricos de trajetórias dos usuários e a estrutura da rede viária, sob diversas configurações de segmentação e para intervalos de amostragem de 1 a 10 minutos. Esta descoberta parece indicar que os dados de trajetórias pedestres disponíveis atualmente não são suficientes para que um método que aprende um modelo de inferência, como o Infertra, possa ter um desempenho ótimo a tal ponto de ser ultrapassado por métodos muito mais simples. Por outro lado, constatou-se que, como esperado, quanto maior é o intervalo de amostragem, menor é o desempenho das técnicas de reconstrução;
- em bases de dados de trajetórias coletadas com uma alta e pouco variável taxa de amostragem, e com pouca presença de interrupções ou lacunas, quanto menor é a tolerância (valores dos limiares de tempo e distância nos pontos de permanência), maior é a quantidade de trajetórias, e menor o comprimento e duração delas. Por outro lado, nas bases de dados com taxas de amostragem variáveis, altas e baixas, acontece o efeito contrário, quanto menor a tolerância, menor a quantidade de trajetórias, e menores, também, o comprimento e duração delas;

- uma segmentação de trajetórias com uma tolerância alta, i.e., pontos de permanência com limiares de distância e tempo de 600 metros e 20 minutos respectivamente, permite obter resultados de reconstrução de trajetórias de pedestres estatisticamente diferenciados e superiores do que as configurações com tolerâncias menores. Este resultado é válido para bases de dados de trajetórias com taxas de amostragem diferentes, e aplicando métodos simples, como o Shortest Path, e complexos, como o Infertra;
- o esquema adaptativo de coleta de dados baseado nos critérios de velocidade de locomoção e tamanho médio de quarteirões, reduz o número de leituras GPS dos pedestres, ciclistas e motoristas de veículos particulares nas rodovias e estradas em 80, 62 e 97% respectivamente, quando comparados com uma leitura contínua com um intervalo de amostragem de 10 segundos. Por outro lado, no caso dos motoristas na rua, não tem redução devido a que o intervalo de amostragem é o mesmo que o esquema contínuo. Além disso, a pesar da redução, as trajetórias coletadas a partir do esquema adaptativo preservam as características suficientes das originais para, assim, poder ser reconstruídas;
- mais dados de pedestres precisam ser coletados para que técnicas mais complexas como o InferTra possam mostrar toda sua capacidade. Porém, nos casos atuais onde tem-se poucos dados de trajetórias de pedestres, uma técnica simples como o Shortest Path é uma melhor alternativa;

Levando isso em conta, as principais contribuições são:

- Estudo do processo de reconstrução de trajetórias de pedestres. Poucas pesquisas realizaram uma investigação de reconstrução de trajetórias exclusivamente de pedestres. Assim, este trabalho utilizou os dados de uma das poucas bases de dados disponíveis com este tipo de dados, o Geolife, juntamente com os dados iniciais coletados com o próprio sistema RadrPlus, e aproveitou a inclusão das informações de caminhos exclusivamente transitados por pedestres nos grafos das redes viárias dos mapas colaborativos existentes, a fim de poder utilizar as técnicas de reconstrução de trajetórias em áreas urbanas do estado da arte, e avaliar seus resultados.
- Avaliação experimental de dois diferentes algoritmos de reconstrução de trajetórias sobre duas bases de dados coletadas em diferentes cidades, considerando diferentes configurações de segmentação que levam em conta o critério de pontos de permanência. Para avaliar estas configurações e algoritmos (os quais possuem abordagens muito diferenciadas), foram aplicadas uma análise visual complementada de uma análise de diferença estatística levando em conta os resultados destes algoritmos com diferentes intervalos de amostragem. O primeiro destes algoritmos é o método considerado o principal neste tipo de problemas segundo a bibliografia, e leva em conta tanto a estrutura da rede viária quanto o histórico de trajetórias dos usuários. Em contraste, o segundo algoritmo apenas leva em conta a

estrutura da rede viária, onde utiliza a heurística do caminho mais curto para inferir os dados faltantes nas trajetórias. Com essa avaliação, foi possível determinar o algoritmo que melhor desempenho tem sobre as trajetórias de pedestres, considerando a escassez dos mesmos, assim como, as condições nas quais estas técnicas de reconstrução alcançam seu melhor desempenho. Desta forma, faz-se possível o aproveitamento deste conhecimento para ser aplicado na reconstrução de trajetórias de pedestres.

- Criação de uma base de dados de trajetórias de usuários pedestres e motoristas (de veículos particulares) compilada por um grupo de 15 usuários por um período de 9 meses. Esta base de dados está disponibilizada e pode ser aproveitada para projetos de pesquisa, especialmente na área de mineração de trajetórias.
- Criação da rede social geo-localizada RadrPlus, a qual é uma plataforma totalmente escalável e, conseqüentemente, garante o suporte para o adequado gerenciamento dos dados de geolocalização dos usuários. Além disso, as características únicas do sistema, como a orientação a comunidades, permite o acesso a valiosas informações de grupos de usuários que não estão disponíveis em outras plataformas. Adicionalmente, o RadrPlus possibilita o uso de informações pessoais de usuários, desta forma, enriquecendo ainda mais a base de dados. Finalmente, o fato do sistema ser propriedade do grupo de pesquisa proporciona uma independência das plataformas de terceiros, muitas das quais fecharam, limitaram ou deixaram de disponibilizar seus dados. Este sistema encontra-se em processo de registro.
- Projeto e avaliação teórico-experimental de uma técnica de coleta eficiente de dados, a qual permite economizar energia e reduzir a perda de dados relevantes. Esta técnica foi implementada no sistema RadrPlus como um serviço de rastreamento chamado de Smart-Tracking. Descreveram-se a heurística e os critérios seguidos nesta técnica e propuseram-se algumas sugestões para sua melhora. Por outro lado, devido à modularidade e o baixo acoplamento deste serviço, ele pode ser utilizado em qualquer outro sistema que precise de um serviço de rastreamento e coleta eficiente de dados de geolocalização.
- Artigo submetido na conferência internacional 4th Annual International Symposium on Information Management and Big Data, track SNMAM, Special Track on Social Network and Media Analysis and Mining, Lima, Peru.

7.2 Limitações do trabalho

Neste trabalho, não foram abordados todos os aspectos da reconstrução de trajetórias, mas apenas como uma técnica complexa e uma simples juntamente com diferentes configurações de segmentação podem influenciar na reconstrução de trajetórias de pedestres.

Dentre as principais restrições, têm-se:

- Não foram implementadas as outras técnicas de reconstrução de trajetórias restritas à rede viária que aprendem modelos de inferência como HRIS (ZHENG *et al.*, 2012). Porém, a técnica utilizada, segundo a bibliografia, ultrapassa ao resto de técnicas do seu tipo;
- não foram testadas técnicas de reconstrução não restritas à rede viária como o RICK (WEI; ZHENG; PENG, 2012), devido principalmente a que o intuito do projeto é aproveitar a inclusão dos trajetos de pedestres na rede viária, através dos mapas colaborativos, para a reconstrução das trajetórias;
- não foram testadas as diferentes técnicas de *Map matching*. Foi tomada uma técnica considerada avançada pela bibliografia, cuja efetividade é refletida na criação do serviço Graphhopper e a sua crescente comunidade de usuários;
- devido às restrições de tempo no mestrado, não se conseguiu implementar diversas técnicas de coleta eficiente de dados geo-localizados do estado da arte, a fim de poder estabelecer comparações com a técnica projetada. Porém, projetou-se uma técnica pertencente a um tipo de abordagem considerado entre os de melhor desempenho, o qual foi avaliado empiricamente e comparado contra uma abordagem muito comum neste tipo de problema;
- considerou-se apenas o critério de pontos de permanência no processo de segmentação de trajetórias. Dentre o grupo de possíveis abordagens, dadas as características do problema, este critério foi o que melhor se ajustou aos nossos requerimentos.
- os intervalos de amostragem avaliados foram até 10 minutos. Intervalos mais longos produziam taxas de acurácia muito baixas, pelo qual o estudo dos resultados foi limitado até esse intervalo.

7.3 Trabalhos futuros

As limitações apresentadas anteriormente, denotam a existência de diversas questões que devem ser estudadas para o melhoramento dos trabalhos na reconstrução de trajetórias de pedestres. Assim, tem-se as seguintes melhorias:

- Realizar novas coletas de dados utilizando o RadrPlus com mais usuários e durante um tempo maior, assim como, implementar novas funcionalidades no sistema a fim de torná-lo mais atrativo para os usuários.
- implementar diversos métodos de coleta eficiente de dados geo-localizados do estado da arte e comparar os seus resultados contra o método projetado neste trabalho;
- aplicar outras técnicas de reconstrução de trajetórias tanto restritas à rede viária quanto as que não o são a fim de compará-las contra as técnicas já testadas;

- avaliar o desempenho de outros métodos de segmentação do estado da arte e testar novas configurações neles nas tarefas de reconstrução de dados de pedestres, assim como, realizar um estudo mais aprofundado dos parâmetros das configurações já utilizadas;
- com intuito de endereçar o problema da insuficiência de dados de trajetórias de pedestres, utilizar bases de dados de trajetórias que não estão etiquetadas com os meios de transporte e inferir esta informação mediante as diversas técnicas existentes a fim de contar com mais dados de pedestres nas pesquisas.
- melhorar a técnica de coleta adaptativa de dados geo-localizados, alterando dinamicamente os valores das taxas de amostragem e limiares de velocidade dependendo da região (país, estado ou cidade) onde o usuário encontra-se. Assim, por meio de uma leitura do GPS, o sistema poderia saber em que região encontra-se o usuário e adaptar sua configuração segundo as informações locais de tamanho médio dos quarteirões e velocidade média dos veículos nas ruas, rodovias e autoestradas.

REFERÊNCIAS

ALT, H.; EFRAT, A.; ROTE, G.; WENK, C. Matching planar maps. In: SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS. **Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms**. [S.l.], 2003. p. 589–598. Citado na página 101.

ANDO, S.; SUZUKI, E. Role-behavior analysis from trajectory data by cross-domain learning. In: IEEE. **Data Mining (ICDM), 2011 IEEE 11th International Conference on**. [S.l.], 2011. p. 21–30. Citado na página 47.

ANDRIENKO, N.; ANDRIENKO, G.; FUCHS, G.; JANKOWSKI, P. Visual analytics methodology for scalable and privacy-respectful discovery of place semantics from episodic mobility data. In: SPRINGER. **Joint European Conference on Machine Learning and Knowledge Discovery in Databases**. [S.l.], 2015. p. 254–258. Citado na página 46.

ASHBROOK, D.; STARNER, T. Using GPS to learn significant locations and predict movement across multiple users. **Personal Ubiquitous Comput.**, Springer-Verlag, London, UK, UK, v. 7, n. 5, p. 275–286, out. 2003. ISSN 1617-4909. Disponível em: <<http://dx.doi.org/10.1007/s00779-003-0240-0>>. Citado na página 33.

AU, T. S.; DUAN, R.; KIM, H.; MA, G.-Q. Spatiotemporal event detection in mobility network. In: IEEE. **Data Mining (ICDM), 2010 IEEE 10th International Conference on**. [S.l.], 2010. p. 28–37. Citado na página 46.

BACKSTROM, L.; HUTTENLOCHER, D.; KLEINBERG, J.; LAN, X. Group formation in large social networks: membership, growth, and evolution. In: **Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining**. New York, NY, USA: ACM, 2006. (KDD '06), p. 44–54. ISBN 1-59593-339-5. Disponível em: <<http://doi.acm.org/10.1145/1150402.1150412>>. Citado na página 35.

BANERJEE, P.; RANU, S.; RAGHAVAN, S. Inferring uncertain trajectories from partial observations. In: IEEE. **2014 IEEE International Conference on Data Mining**. [S.l.], 2014. p. 30–39. Citado nas páginas 26, 40, 95, 96, 102, 107, 111, 113 e 114.

BARANIUK, R. G. More is less: signal processing and the data deluge. **Science**, American Association for the Advancement of Science, v. 331, n. 6018, p. 717–719, 2011. Citado nas páginas 25 e 36.

BARETH, U. Simulating power consumption of location tracking algorithms to improve energy-efficiency of smartphones. In: IEEE. **Computer Software and Applications Conference (COMPSAC), 2012 IEEE 36th Annual**. [S.l.], 2012. p. 613–622. Citado na página 76.

BARETH, U.; KUPPER, A. Energy-efficient position tracking in proactive location-based services for smartphone environments. In: IEEE. **Computer Software and Applications Conference (COMPSAC), 2011 IEEE 35th Annual**. [S.l.], 2011. p. 516–521. Citado na página 76.

- BRAKATSOULAS, S.; PFOSE, D.; SALAS, R.; WENK, C. On map-matching vehicle tracking data. In: VLDB ENDOWMENT. **Proceedings of the 31st international conference on Very large data bases**. [S.l.], 2005. p. 853–864. Citado nas páginas 100 e 101.
- BROWNING, R. C.; BAKER, E. A.; HERRON, J. A.; KRAM, R. Effects of obesity and sex on the energetic cost and preferred speed of walking. **Journal of Applied Physiology**, Am Physiological Soc, v. 100, n. 2, p. 390–398, 2006. Citado na página 80.
- CABALONA, J. **Gowalla Is Officially Shut Down**. 2012. Disponível em: <<http://mashable.com/2012/03/11/gowalla-shuts-down/>>. Acesso em: 26/02/2017. Citado na página 27.
- CASELLA, G.; GEORGE, E. I. Explaining the gibbs sampler. **The American Statistician**, Taylor & Francis, v. 46, n. 3, p. 167–174, 1992. Citado nas páginas 26 e 95.
- CENTER, T. S. G. of the G. A. *et al.* **Urban Block Design guideline / manual to best practice - Project METRASYS**. [S.l.], 2012. Citado nas páginas 80 e 111.
- CHAWATHE, S. S. Segment-based map matching. In: IEEE. **Intelligent Vehicles Symposium, 2007 IEEE**. [S.l.], 2007. p. 1190–1197. Citado na página 101.
- CHEN, W.; YU, M.; LI, Z.; CHEN, Y. Integrated vehicle navigation system for urban applications. 2003. Citado na página 100.
- CHEN, Z.; SHEN, H. T.; ZHOU, X. Discovering popular routes from trajectories. In: IEEE. **Data Engineering (ICDE), 2011 IEEE 27th International Conference on**. [S.l.], 2011. p. 900–911. Citado na página 47.
- CHEN, Z.; SHEN, H. T.; ZHOU, X.; ZHENG, Y.; XIE, X. Searching trajectories by locations: an efficiency study. In: ACM. **Proceedings of the 2010 ACM SIGMOD International Conference on Management of data**. [S.l.], 2010. p. 255–266. Citado na página 42.
- CHIANG, M.-F.; LIN, Y.-H.; PENG, W.-C.; YU, P. S. Inferring distant-time location in low-sampling-rate trajectories. In: ACM. **Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2013. p. 1454–1457. Citado nas páginas 26, 40 e 95.
- CIVILIS, A.; JENSEN, C. S.; PAKALNIS, S. Techniques for efficient road-network-based tracking of moving objects. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 17, n. 5, p. 698–712, 2005. Citado na página 101.
- COSTA, C.; LAOUDIAS, C.; ZEINALIPOUR-YAZTI, D.; GUNOPULOS, D. Smarttrace: Finding similar trajectories in smartphone networks without disclosing the traces. In: IEEE. **Data Engineering (ICDE), 2011 IEEE 27th International Conference on**. [S.l.], 2011. p. 1288–1291. Citado na página 46.
- CUDRE-MAUROUX, P.; WU, E.; MADDEN, S. Trajstore: An adaptive storage system for very large trajectory data sets. In: IEEE. **Data Engineering (ICDE), 2010 IEEE 26th International Conference on**. [S.l.], 2010. p. 109–120. Citado nas páginas 41 e 42.
- DAI, J.; YANG, B.; GUO, C.; DING, Z. Personalized route recommendation using big trajectory data. In: IEEE. **Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 543–554. Citado na página 47.

- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine learning research**, v. 7, n. Jan, p. 1–30, 2006. Citado na página 116.
- FAZZINGA, B.; FLESCA, S.; FURFARO, F.; PARISI, F. Cleaning trajectory data of rfid-monitored objects through conditioning under integrity constraints. In: **EDBT**. [S.l.: s.n.], 2014. p. 379–390. Citado na página 39.
- FENG, Z.; ZHU, Y. A survey on trajectory data mining: Techniques and applications. **IEEE Access**, IEEE, v. 4, p. 2056–2067, 2016. Citado nas páginas 25, 36, 38, 39, 41, 43, 49 e 76.
- FORNEY, G. D. The viterbi algorithm. **Proceedings of the IEEE**, IEEE, v. 61, n. 3, p. 268–278, 1973. Citado na página 101.
- GAO, H.; TANG, J.; HU, X.; LIU, H. Modeling temporal effects of human mobile behavior on location-based social networks. In: ACM. **Proceedings of the 22nd ACM international conference on Conference on information & knowledge management**. [S.l.], 2013. p. 1673–1678. Citado na página 47.
- GE, Y.; XIONG, H.; ZHOU, Z.-h.; OZDEMIR, H.; YU, J.; LEE, K. C. Top-eye: Top-k evolving trajectory outlier detection. In: ACM. **Proceedings of the 19th ACM international conference on Information and knowledge management**. [S.l.], 2010. p. 1733–1736. Citado na página 39.
- GONZALEZ, M. C.; HIDALGO, C. A.; BARABASI, A.-L. Understanding individual human mobility patterns. **Nature**, Nature Publishing Group, v. 453, n. 7196, p. 779–782, 2008. Citado na página 47.
- GREENFELD, J. S. Matching gps observations to locations on a digital map. In: **Transportation Research Board 81st Annual Meeting**. [S.l.: s.n.], 2002. Citado na página 100.
- GRONLI, T.-M.; HANSEN, J.; GHINEA, G.; YOUNAS, M. Mobile application platform heterogeneity: Android vs windows phone vs ios vs firefox os. In: IEEE. **Advanced Information Networking and Applications (AINA), 2014 IEEE 28th International Conference on**. [S.l.], 2014. p. 635–641. Citado na página 55.
- GUPTA, A.; MISHRA, A.; VADLAMUDI, S. G.; CHAKRABARTI, P.; SARKAR, S.; MUKHERJEE, T.; GNANASAMBANDAM, N. A mobility simulation framework of humans with group behavior modeling. In: IEEE. **Data Mining (ICDM), 2013 IEEE 13th International Conference on**. [S.l.], 2013. p. 1067–1072. Citado na página 48.
- GÜTING, R. H.; BEHR, T.; XU, J. Efficient k-nearest neighbor search on moving object trajectories. **The VLDB Journal—The International Journal on Very Large Data Bases**, Springer-Verlag New York, Inc., v. 19, n. 5, p. 687–714, 2010. Citado na página 43.
- HARIHARAN, R.; TOYAMA, K. Project Lachesis: Parsing and Modeling Location Histories. In: **GIScience**. [S.l.: s.n.], 2004. p. 106–124. Citado na página 33.
- HARRIS, D.; AECOM; HINTZ, C. C. **New Jersey Long Range Transportation Plan 2030. Technical Memorandum. Task 11: Local Street Connectivity Redefined**. New Jersey, 2008. Citado nas páginas 79 e 111.
- HUNTER, T.; ABBEEL, P.; BAYEN, A. M. The path inference filter: model-based low-latency map matching of probe vehicle data. In: **Algorithmic Foundations of Robotics X**. [S.l.]: Springer, 2013. p. 591–607. Citado nas páginas 26 e 95.

- INSTITUTE, H. L. D. **Speed limits**. 2017. Disponível em: <<http://www.iihs.org/iihs/topics/laws/speedlimits?topicName=speed>>. Acesso em: 30/04/2017. Citado na página 80.
- JENSEN, P.; ROUQUIER, J.-B.; OVTRACHT, N.; ROBARDET, C. Characterizing the speed and paths of shared bicycle use in lyon. **Transportation research part D: transport and environment**, Elsevier, v. 15, n. 8, p. 522–524, 2010. Citado na página 80.
- KIM, Y.; HAN, J.; YUAN, C. Toptrac: Topical trajectory pattern mining. In: ACM. **Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.], 2015. p. 587–596. Citado na página 49.
- KONG, L.; HE, L.; LIU, X.-Y.; GU, Y.; WU, M.-Y.; LIU, X. Privacy-preserving compressive sensing for crowdsensing based trajectory recovery. In: IEEE. **Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on**. [S.l.], 2015. p. 31–40. Citado na página 46.
- KRUMM, J. Trajectory analysis for driving. In: **Computing with Spatial Trajectories**. [S.l.]: Springer, 2011. p. 213–241. Citado na página 100.
- LARUSSO, N. D.; SINGH, A. Efficient tracking and querying for coordinated uncertain mobile objects. In: IEEE. **Data Engineering (ICDE), 2013 IEEE 29th International Conference on**. [S.l.], 2013. p. 182–193. Citado na página 40.
- LEE, J.-G.; HAN, J.; WHANG, K.-Y. Trajectory clustering: a partition-and-group framework. In: ACM. **Proceedings of the 2007 ACM SIGMOD international conference on Management of data**. [S.l.], 2007. p. 593–604. Citado na página 99.
- LEE, M.-J.; CHUNG, C.-W. A user similarity calculation based on the location for social network services. In: **Proceedings of the 16th international conference on Database systems for advanced applications - Volume Part I**. Berlin, Heidelberg: Springer-Verlag, 2011. (DAS-FAA'11), p. 38–52. ISBN 978-3-642-20148-6. Disponível em: <<http://dl.acm.org/citation.cfm?id=1997305.1997313>>. Citado na página 35.
- LEE, R.; SUMIYA, K. Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection. In: **Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks**. New York, NY, USA: ACM, 2010. (LBSN '10), p. 1–10. ISBN 978-1-4503-0434-4. Disponível em: <<http://doi.acm.org/10.1145/1867699.1867701>>. Citado na página 35.
- LEHTINEN, M.; HAPPONEN, A.; IKONEN, J. Accuracy and time to first fix using consumer-grade gps receivers. In: IEEE. **Software, Telecommunications and Computer Networks, 2008. SoftCOM 2008. 16th International Conference on**. [S.l.], 2008. p. 334–340. Citado na página 82.
- LEVINE, R. V.; NORENZAYAN, A. The pace of life in 31 countries. **Journal of cross-cultural psychology**, Sage Publications, v. 30, n. 2, p. 178–205, 1999. Citado na página 80.
- LI, M.; AHMED, A.; SMOLA, A. J. Inferring movement trajectories from gps snippets. In: ACM. **Proceedings of the Eighth ACM International Conference on Web Search and Data Mining**. [S.l.], 2015. p. 325–334. Citado nas páginas 26, 40 e 95.

LI, Q.; ZHENG, Y.; XIE, X.; CHEN, Y.; LIU, W.; MA, W.-Y. Mining user similarity based on location history. In: **Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems**. New York, NY, USA: ACM, 2008. (GIS '08), p. 34:1–34:10. ISBN 978-1-60558-323-5. Disponível em: <<http://doi.acm.org/10.1145/1463434.1463477>>. Citado nas páginas 33 e 35.

_____. Mining user similarity based on location history. In: ACM. **Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems**. [S.l.], 2008. p. 34. Citado na página 99.

LI, R.-H.; LIU, J.; YU, J. X.; CHEN, H.; KITAGAWA, H. Co-occurrence prediction in a large location-based social network. **Front. Comput. Sci.**, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 7, n. 2, p. 185–194, abr. 2013. ISSN 2095-2228. Disponível em: <<http://dx.doi.org/10.1007/s11704-013-3902-8>>. Citado na página 35.

LI, X.; CEIKUTE, V.; JENSEN, C. S.; TAN, K.-L. Effective online group discovery in trajectory databases. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 25, n. 12, p. 2752–2766, 2013. Citado nas páginas 45 e 48.

LI, Y.; CHOW, C.-Y.; DENG, K.; YUAN, M.; ZENG, J.; ZHANG, J.-D.; YANG, Q.; ZHANG, Z.-L. Sampling big trajectory data. In: ACM. **Proceedings of the 24th ACM International on Conference on Information and Knowledge Management**. [S.l.], 2015. p. 941–950. Citado nas páginas 40 e 44.

LI, Y.; LUO, J.; CHOW, C.-Y.; CHAN, K.-L.; DING, Y.; ZHANG, F. Growing the charging station network for electric vehicles with trajectory data analytics. In: IEEE. **Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 1376–1387. Citado na página 48.

LI, Z.; DING, B.; HAN, J.; KAYS, R.; NYE, P. Mining periodic behaviors for moving objects. In: ACM. **Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2010. p. 1099–1108. Citado na página 45.

LI, Z.; HAN, J.; DING, B.; KAYS, R. Mining periodic behaviors of object movements for animal and biological sustainability studies. **Data Mining and Knowledge Discovery**, Springer, v. 24, n. 2, p. 355–386, 2012. Citado na página 45.

LIEBIG, T.; XU, Z.; MAY, M.; WROBEL, S. Pedestrian quantity estimation with trajectory patterns. In: SPRINGER. **Joint European Conference on Machine Learning and Knowledge Discovery in Databases**. [S.l.], 2012. p. 629–643. Citado nas páginas 48 e 49.

LIU, J.; ZHAO, K.; SOMMER, P.; SHANG, S.; KUSY, B.; JURDAK, R. Bounded quadrant system: Error-bounded trajectory compression on the go. In: IEEE. **Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 987–998. Citado na página 41.

LIU, S.; LIU, Y.; NI, L. M.; FAN, J.; LI, M. Towards mobility-based clustering. In: ACM. **Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2010. p. 919–928. Citado na página 45.

LIU, S.; PU, J.; LUO, Q.; QU, H.; NI, L. M.; KRISHNAN, R. Vait: A visual analytics system for metropolitan transportation. **IEEE Transactions on Intelligent Transportation Systems**, IEEE, v. 14, n. 4, p. 1586–1596, 2013. Citado na página 36.

LIU, S.; QU, Q.; WANG, S. Rationality analytics from trajectories. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, ACM, v. 10, n. 1, p. 10, 2015. Citado na página 47.

LIU, S.; WANG, S.; JAYARAJAH, K.; MISRA, A.; KRISHNAN, R. Todmis: Mining communities from trajectories. In: ACM. **Proceedings of the 22nd ACM international conference on Conference on information & knowledge management**. [S.l.], 2013. p. 2109–2118. Citado na página 45.

LIU, W.; ZHENG, Y.; CHAWLA, S.; YUAN, J.; XING, X. Discovering spatio-temporal causal interactions in traffic data streams. In: **Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**. New York, NY, USA: ACM, 2011. (KDD '11), p. 1010–1018. ISBN 978-1-4503-0813-7. Disponível em: <<http://doi.acm.org/10.1145/2020408.2020571>>. Citado na página 35.

LIU, X.; BIAGIONI, J.; ERIKSSON, J.; WANG, Y.; FORMAN, G.; ZHU, Y. Mining large-scale, sparse gps traces for map inference: comparison of approaches. In: ACM. **Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2012. p. 669–677. Citado nas páginas 48 e 49.

LIU, Y.; LIU, C.; YUAN, N. J.; DUAN, L.; FU, Y.; XIONG, H.; XU, S.; WU, J. Exploiting heterogeneous human mobility patterns for intelligent bus routing. In: IEEE. **Data Mining (ICDM), 2014 IEEE International Conference on**. [S.l.], 2014. p. 360–369. Citado na página 47.

LOMAS, N. **Gartner: Android's smartphone marketshare hit 86.2% in Q2**. 2016. Disponível em: <<https://techcrunch.com/2016/08/18/gartner-androids-smartphone-marketshare-hit-86-2-in-q2/>>. Acesso em: 18/07/2017. Citado na página 55.

LONG, C.; WONG, R. C.-W.; JAGADISH, H. Direction-preserving trajectory simplification. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 6, n. 10, p. 949–960, 2013. Citado na página 41.

_____. Trajectory simplification: on minimizing the direction-based error. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 8, n. 1, p. 49–60, 2014. Citado na página 41.

LOU, Y.; ZHANG, C.; ZHENG, Y.; XIE, X.; WANG, W.; HUANG, Y. Map-matching for low-sampling-rate gps trajectories. In: ACM. **Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems**. [S.l.], 2009. p. 352–361. Citado nas páginas 100 e 101.

LU, C.-T.; LEI, P.-R.; PENG, W.-C.; SU, I.-J. A framework of mining semantic regions from trajectories. In: SPRINGER. **Database Systems for Advanced Applications**. [S.l.], 2011. p. 193–207. Citado na página 49.

LU, E. H.-C.; LEE, W.-C.; TSENG, V. S. Mining fastest path from trajectories with multiple destinations in road networks. **Knowledge and information systems**, Springer, v. 29, n. 1, p. 25–53, 2011. Citado na página 47.

LU, H.; YANG, J.; LIU, Z.; LANE, N. D.; CHOUDHURY, T.; CAMPBELL, A. T. The jigsaw continuous sensing engine for mobile phone applications. In: ACM. **Proceedings of the 8th**

ACM conference on embedded networked sensor systems. [S.l.], 2010. p. 71–84. Citado na página 76.

LUO, W.; TAN, H.; CHEN, L.; NI, L. M. Finding time period-based most frequent path in big trajectory data. In: **ACM. Proceedings of the 2013 ACM SIGMOD international conference on management of data**. [S.l.], 2013. p. 713–724. Citado na página 47.

LV, M.; CHEN, L.; CHEN, G. Discovering personally semantic places from gps trajectories. In: **ACM. Proceedings of the 21st ACM international conference on Information and knowledge management**. [S.l.], 2012. p. 1552–1556. Citado na página 49.

MA, C.; LU, H.; SHOU, L.; CHEN, G. Ksq: Top-k similarity query on uncertain trajectories. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 25, n. 9, p. 2049–2062, 2013. Citado na página 44.

MCGUIRE, M. P.; JANEJA, V. P.; GANGOPADHYAY, A. Mining trajectories of moving dynamic spatio-temporal regions in sensor datasets. **Data Mining and Knowledge Discovery**, Springer, v. 28, n. 4, p. 961–1003, 2014. Citado na página 48.

MINETTI, A. The three modes of terrestrial locomotion. **Biomechanics and biology of movement**, Human Kinetics, Champaign, IL, p. 67–78, 2000. Citado na página 80.

MOHLER, B. J.; THOMPSON, W. B.; CREEM-REGEHR, S. H.; PICK, H. L.; WARREN, W. H. Visual flow influences gait transition speed and preferred walking speed. **Experimental brain research**, Springer, v. 181, n. 2, p. 221–228, 2007. Citado na página 80.

MONTAZERI-GH, M.; NAGHIZADEH, M. Development of car drive cycle for simulation of emissions and fuel economy. In: **Proceedings of 15th European simulation symposium**. [S.l.: s.n.], 2003. Citado na página 80.

MORIANO, P.; FINKE, J. On the formation of structure in growing networks. **Journal of Statistical Mechanics: Theory and Experiment**, v. 2013, n. 06, p. P06010, 2013. Disponível em: <<http://stacks.iop.org/1742-5468/2013/i=06/a=P06010>>. Citado na página 35.

NEWSON, P.; KRUMM, J. Hidden markov map matching through noise and sparseness. In: **ACM. Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems**. [S.l.], 2009. p. 336–343. Citado nas páginas 100, 101, 107 e 111.

NI, J.; RAVISHANKAR, C. V. Indexing spatio-temporal trajectories with efficient polynomial approximations. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 19, n. 5, 2007. Citado nas páginas 41 e 42.

NIBALI, A.; HE, Z. Trajic: an effective compression system for trajectory data. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 27, n. 11, p. 3138–3151, 2015. Citado na página 41.

NIEDERMAYER, J.; ZÜFLE, A.; EMRICH, T.; RENZ, M.; MAMOULIS, N.; CHEN, L.; KRIEGEL, H.-P. Probabilistic nearest neighbor queries on uncertain moving object trajectories. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 7, n. 3, p. 205–216, 2013. Citado na página 43.

NOULAS, A.; SCELLATO, S.; LATHIA, N.; MASCOLO, C. Mining user mobility features for next place prediction in location-based services. In: **Proceedings of the 2012 IEEE 12th International Conference on Data Mining**. Washington, DC, USA: IEEE Computer Society, 2012. (ICDM '12), p. 1038–1043. ISBN 978-0-7695-4905-7. Disponível em: <<http://dx.doi.org/10.1109/ICDM.2012.113>>. Citado na página 35.

_____. Mining user mobility features for next place prediction in location-based services. In: IEEE. **Data mining (ICDM), 2012 IEEE 12th international conference on**. [S.l.], 2012. p. 1038–1043. Citado na página 47.

OCHIENG, W. Y.; QUDDUS, M. A.; NOLAND, R. B. Map-matching in complex urban road networks. © Brazilian Society of Cartography, Geodesy, Photogrammetry and Remote Sensing (SBC), 2003. Citado nas páginas 100 e 101.

PANAGIOTAKIS, C.; PELEKIS, N.; KOPANAKIS, I.; RAMASSO, E.; THEODORIDIS, Y. Segmentation and sampling of moving object trajectories based on representativeness. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 24, n. 7, p. 1328–1343, 2012. Citado na página 40.

PATEL, D.; SHENG, C.; HSU, W.; LEE, M. L. Incorporating duration information for trajectory classification. In: IEEE. **Data Engineering (ICDE), 2012 IEEE 28th International Conference on**. [S.l.], 2012. p. 1132–1143. Citado na página 46.

PELEKIS, N.; GKOUALALAS-DIVANIS, A.; VODAS, M.; KOPANAKI, D.; THEODORIDIS, Y. Privacy-aware querying over sensitive trajectory data. In: ACM. **Proceedings of the 20th ACM international conference on Information and knowledge management**. [S.l.], 2011. p. 895–904. Citado na página 46.

PELEKIS, N.; KOPANAKIS, I.; PANAGIOTAKIS, C.; THEODORIDIS, Y. Unsupervised trajectory sampling. **Machine learning and knowledge discovery in databases**, Springer, p. 17–33, 2010. Citado nas páginas 39 e 40.

PINK, O.; HUMMEL, B. A statistical approach to map matching using road network geometry, topology and vehicular motion constraints. In: IEEE. **Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on**. [S.l.], 2008. p. 862–867. Citado nas páginas 100 e 101.

POPA, I. S.; ZEITOUNI, K.; ORIA, V.; BARTH, D.; VIAL, S. Indexing in-network trajectory flows. **The VLDB Journal—The International Journal on Very Large Data Bases**, Springer-Verlag New York, Inc., v. 20, n. 5, p. 643–669, 2011. Citado na página 41.

QUDDUS, M. A.; NOLAND, R. B.; OCHIENG, W. Y. A high accuracy fuzzy logic based map matching algorithm for road transport. **Journal of Intelligent Transportation Systems**, Taylor & Francis, v. 10, n. 3, p. 103–115, 2006. Citado nas páginas 100 e 101.

RABINER, L.; JUANG, B. An introduction to hidden markov models. **ieee assp magazine**, IEEE, v. 3, n. 1, p. 4–16, 1986. Citado na página 102.

RAFTERY, A. E. A model for high-order markov chains. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 528–539, 1985. Citado na página 102.

- RANU, S.; DEEPAK, P.; TELANG, A. D.; DESHPANDE, P.; RAGHAVAN, S. Indexing and matching trajectories under inconsistent sampling rates. In: IEEE. **Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 999–1010. Citado nas páginas 41 e 42.
- RENZO, C.; BAGLIONI, M.; MACEDO, J. A. F. de; TRASARTI, R.; WACHOWICZ, M. How you move reveals who you are: understanding human behavior by analyzing trajectory data. **Knowledge and information systems**, Springer, p. 1–32, 2013. Citado nas páginas 47 e 48.
- ROBINSON, S. Cellphone energy gap: Desperately seeking solutions. **Strategy Analytics**, 2009. Citado na página 75.
- SAALFELD, A. Topologically consistent line simplification with the douglas-peucker algorithm. **Cartography and Geographic Information Science**, Taylor & Francis, v. 26, n. 1, p. 7–18, 1999. Citado na página 98.
- SCELLATO, S.; NOULAS, A.; MASCOLO, C. Exploiting place features in link prediction on location-based social networks. In: **Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining**. New York, NY, USA: ACM, 2011. (KDD '11), p. 1046–1054. ISBN 978-1-4503-0813-7. Disponível em: <<http://doi.acm.org/10.1145/2020408.2020575>>. Citado na página 35.
- SONG, R.; SUN, W.; ZHENG, B.; ZHENG, Y. Press: A novel framework of trajectory compression in road networks. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 7, n. 9, p. 661–672, 2014. Citado na página 41.
- SONG, X.; ZHANG, Q.; SEKIMOTO, Y.; SHIBASAKI, R. Prediction of human emergency behavior and their mobility following large-scale disaster. In: ACM. **Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2014. p. 5–14. Citado nas páginas 47 e 48.
- STEURER, M.; TRATTNER, C. Acquaintance or partner? predicting partnership in online and location-based social networks. In: **IEEE/ACM International Conference on Advances in Social Network Analysis and Mining**. [S.l.: s.n.], 2013. (ASONAM 2013). Citado na página 35.
- STEVENS, T. **Google Latitude shuts down August 9, but Google+ location sharing will go on (and on)**. 2013. Disponível em: <<https://www.engadget.com/2013/07/10/google-latitude-shuts-down/>>. Acesso em: 26/02/2017. Citado na página 27.
- SU, H.; ZHENG, K.; HUANG, J.; WANG, H.; ZHOU, X. Calibrating trajectory data for spatio-temporal similarity analysis. **The VLDB Journal**, Springer, v. 24, n. 1, p. 93–116, 2015. Citado na página 40.
- SU, H.; ZHENG, K.; WANG, H.; HUANG, J.; ZHOU, X. Calibrating trajectory data for similarity-based analysis. In: ACM. **Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data**. [S.l.], 2013. p. 833–844. Citado nas páginas 40 e 95.
- SU, H.; ZHENG, K.; ZENG, K.; HUANG, J.; ZHOU, X. Stmaker: a system to make sense of trajectory data. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 7, n. 13, p. 1701–1704, 2014. Citado na página 49.

- SU, H.; ZHENG, K.; ZENG, K.; HUANG, J.; SADIQ, S.; YUAN, N. J.; ZHOU, X. Making sense of trajectory data: A partition-and-summarization approach. In: IEEE. **Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 963–974. Citado nas páginas 40 e 49.
- SUI, D. Z.; ELWOOD, S.; GOODCHILD, M. **Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 52.
- TONG, H.; FALOUTSOS, C.; PAN, J.-Y. Random walk with restart: fast solutions and applications. **Knowledge and Information Systems**, Springer, v. 14, n. 3, p. 327–346, 2008. Citado na página 104.
- TRAJCEVSKI, G.; CHOUDHARY, A.; WOLFSON, O.; YE, L.; LI, G. Uncertain range queries for necklaces. In: IEEE. **Mobile Data Management (MDM), 2010 Eleventh International Conference on**. [S.l.], 2010. p. 199–208. Citado na página 94.
- TRAJCEVSKI, G.; TAMASSIA, R.; DING, H.; SCHEUERMANN, P.; CRUZ, I. F. Continuous probabilistic nearest-neighbor queries for uncertain trajectories. In: ACM. **Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology**. [S.l.], 2009. p. 874–885. Citado na página 94.
- TRAJCEVSKI, G.; WOLFSON, O.; HINRICHS, K.; CHAMBERLAIN, S. Managing uncertainty in moving objects databases. **ACM Transactions on Database Systems (TODS)**, ACM, v. 29, n. 3, p. 463–507, 2004. Citado na página 94.
- VÁRHELYI, A.; MÄKINEN, T. The effects of in-car speed limiters: field studies. **Transportation Research Part C: Emerging Technologies**, Elsevier, v. 9, n. 3, p. 191–211, 2001. Citado na página 80.
- VIEIRA, M. R.; BAKALOV, P.; TSOTRAS, V. J. Querying trajectories using flexible patterns. In: ACM. **Proceedings of the 13th International Conference on Extending Database Technology**. [S.l.], 2010. p. 406–417. Citado na página 44.
- VINCENT, J. **99.6 percent of new smartphones run Android or iOS**. 2017. Disponível em: <<https://www.theverge.com/2017/2/16/14634656/android-ios-market-share-blackberry-2016>>. Acesso em: 18/07/2017. Citado na página 55.
- VRIES, G. de; SOMEREN, M. van. Clustering vessel trajectories with alignment kernels under trajectory compression. **Machine Learning and Knowledge Discovery in Databases**, Springer, p. 296–311, 2010. Citado na página 45.
- WANG, D.; PEDRESCHI, D.; SONG, C.; GIANNOTTI, F.; BARABASI, A.-L. Human mobility, social ties, and link prediction. In: **Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 2011. (KDD '11), p. 1100–1108. ISBN 978-1-4503-0813-7. Disponível em: <<http://doi.acm.org/10.1145/2020408.2020581>>. Citado na página 35.
- WANG, H.; ZHENG, K.; XU, J.; ZHENG, B.; ZHOU, X.; SADIQ, S. Sharkdb: An in-memory column-oriented trajectory storage. In: ACM. **Proceedings of the 23rd ACM international conference on conference on information and knowledge management**. [S.l.], 2014. p. 1409–1418. Citado nas páginas 40, 41 e 42.

WANG, H.; ZHENG, K.; ZHOU, X.; SADIQ, S. Sharkdb: An in-memory storage system for massive trajectory data. In: **ACM. Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data**. [S.l.], 2015. p. 1099–1104. Citado nas páginas 40, 41 e 42.

WANG, L.; CUI, Y.; STOJMENOVIC, I.; MA, X.; SONG, J. Energy efficiency on location based applications in mobile cloud computing: a survey. **Computing**, Springer, v. 96, n. 7, p. 569–585, 2014. Citado na página 76.

WANG, X.; LI, G.; JIANG, G.; SHI, Z. Semantic trajectory-based event detection and event pattern mining. **Knowledge and information systems**, Springer, p. 1–25, 2013. Citado na página 46.

WAUTERS, R. **Brightkite Winds Down, Says It Will Come Back With ‘Something Better’ (Again)**. 2011. Disponível em: <<https://techcrunch.com/2011/12/20/brightkite-winds-down-says-it-will-come-back-with-something-better-again/>>. Acesso em: 26/02/2017. Citado na página 27.

WEI, L.-Y.; ZHENG, Y.; PENG, W.-C. Constructing popular routes from uncertain trajectories. In: **ACM. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2012. p. 195–203. Citado nas páginas 25, 26, 47, 95 e 125.

WU, C.-L.; HUANG, Y.-T.; WU, C.-L.; CHU, H.-h.; HUANG, P.; CHEN, L.-J. An adaptive duty-cycle scheme for gps scheduling in mobile location sensing applications. **Proc. of PhoneSense**, 2011. Citado na página 76.

XIAO, X.; ZHENG, Y.; LUO, Q.; XIE, X. Finding similar users using category-based location history. In: **Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems**. New York, NY, USA: ACM, 2010. (GIS '10), p. 442–445. ISBN 978-1-4503-0428-3. Disponível em: <<http://doi.acm.org/10.1145/1869790.1869857>>. Citado na página 33.

XUE, A. Y.; ZHANG, R.; ZHENG, Y.; XIE, X.; YU, J.; TANG, Y. Desteller: A system for destination prediction based on trajectories with privacy protection. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 6, n. 12, p. 1198–1201, 2013. Citado na página 47.

XUE, A. Y.; ZHANG, R.; ZHENG, Y.; XIE, X.; HUANG, J.; XU, Z. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In: **IEEE. Data Engineering (ICDE), 2013 IEEE 29th International Conference on**. [S.l.], 2013. p. 254–265. Citado na página 47.

YAN, D.; CHENG, J.; ZHAO, Z.; NG, W. Efficient location-based search of trajectories with location importance. **Knowledge and Information Systems**, Springer, v. 45, n. 1, p. 215–245, 2015. Citado na página 42.

YANG, B.; FANTINI, N.; JENSEN, C. S. ipark: Identifying parking spaces from trajectories. In: **ACM. Proceedings of the 16th International Conference on Extending Database Technology**. [S.l.], 2013. p. 705–708. Citado na página 48.

YEANG, L. D. *et al.* Urban design compendium. **English Partnerships/Housing Corporation, London**, 2000. Citado nas páginas 80 e 111.

YIN, H.; WOLFSON, O. A weight-based map matching method in moving objects databases. In: **IEEE. Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on**. [S.l.], 2004. p. 437–438. Citado na página 100.

YOON, H.; ZHENG, Y.; XIE, X.; WOO, W. Smart itinerary recommendation based on user-generated GPS trajectories. In: **Proceedings of the 7th international conference on Ubiquitous intelligence and computing**. Berlin, Heidelberg: Springer-Verlag, 2010. (UIC'10), p. 19–34. ISBN 3-642-16354-8, 978-3-642-16354-8. Disponível em: <<http://dl.acm.org/citation.cfm?id=1929661.1929669>>. Citado na página 35.

_____. Social itinerary recommendation from user-generated digital trails. **Personal Ubiquitous Comput.**, Springer-Verlag, London, UK, UK, v. 16, n. 5, p. 469–484, jun. 2012. ISSN 1617-4909. Disponível em: <<http://dx.doi.org/10.1007/s00779-011-0419-8>>. Citado na página 35.

YUAN, J.; ZHENG, Y.; XIE, X. Discovering regions of different functions in a city using human mobility and pois. In: **ACM. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.], 2012. p. 186–194. Citado nas páginas 46 e 48.

YUAN, J.; ZHENG, Y.; ZHANG, C.; XIE, X.; SUN, G.-Z. An interactive-voting based map matching algorithm. In: **IEEE. Mobile Data Management (MDM), 2010 Eleventh International Conference on**. [S.l.], 2010. p. 43–52. Citado nas páginas 100 e 101.

YUAN, N. J.; ZHENG, Y.; XIE, X.; WANG, Y.; ZHENG, K.; XIONG, H. Discovering urban functional zones using latent activity trajectories. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 27, n. 3, p. 712–725, 2015. Citado nas páginas 46 e 48.

ZHAN, L.; ZHANG, Y.; ZHANG, W.; WANG, X.; LIN, X. Range search on uncertain trajectories. In: **ACM. Proceedings of the 24th ACM International Conference on Information and Knowledge Management**. [S.l.], 2015. p. 921–930. Citado na página 43.

ZHANG, C.; HAN, J.; SHOU, L.; LU, J.; PORTA, T. L. Splitter: Mining fine-grained sequential patterns in semantic trajectories. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 7, n. 9, p. 769–780, 2014. Citado na página 45.

ZHANG, J.; KONG, X.; YU, P. S. Transferring heterogeneous links across location-based social networks. In: **Proceedings of the 7th ACM International Conference on Web Search and Data Mining**. New York, NY, USA: ACM, 2014. (WSDM '14), p. 303–312. ISBN 978-1-4503-2351-2. Disponível em: <<http://doi.acm.org/10.1145/2556195.2559894>>. Citado na página 35.

ZHENG, B.; YUAN, N. J.; ZHENG, K.; XIE, X.; SADIQ, S.; ZHOU, X. Approximate keyword search in semantic trajectory database. In: **IEEE. Data Engineering (ICDE), 2015 IEEE 31st International Conference on**. [S.l.], 2015. p. 975–986. Citado na página 44.

ZHENG, K.; SHANG, S.; YUAN, N. J.; YANG, Y. Towards efficient search for activity trajectories. In: **IEEE. Data Engineering (ICDE), 2013 IEEE 29th International Conference on**. [S.l.], 2013. p. 230–241. Citado na página 44.

ZHENG, K.; TRAJCEVSKI, G.; ZHOU, X.; SCHEUERMANN, P. Probabilistic range queries for uncertain trajectories on road networks. In: **ACM. Proceedings of the 14th International Conference on Extending Database Technology**. [S.l.], 2011. p. 283–294. Citado na página 43.

ZHENG, K.; ZHENG, Y.; XIE, X.; ZHOU, X. Reducing uncertainty of low-sampling-rate trajectories. In: IEEE. **2012 IEEE 28th International Conference on Data Engineering**. [S.l.], 2012. p. 1144–1155. Citado nas páginas 26, 40, 95 e 125.

ZHENG, K.; ZHENG, Y.; YUAN, N. J.; SHANG, S. On discovery of gathering patterns from trajectories. In: IEEE. **Data Engineering (ICDE), 2013 IEEE 29th International Conference on**. [S.l.], 2013. p. 242–253. Citado nas páginas 45 e 48.

ZHENG, K.; ZHENG, Y.; YUAN, N. J.; SHANG, S.; ZHOU, X. Online discovery of gathering patterns over trajectories. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 26, n. 8, p. 1974–1988, 2014. Citado nas páginas 45 e 48.

ZHENG, V. W.; ZHENG, Y.; XIE, X.; YANG, Q. Collaborative location and activity recommendations with GPS history data. In: **Proceedings of the 19th international conference on World wide web**. ACM, 2010. (WWW '10), p. 1029–1038. ISBN 978-1-60558-799-8. Disponível em: <http://doi.acm.org/10.1145/1772690.1772795>. Citado na página 35.

ZHENG, Y. Trajectory data mining: an overview. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM, v. 6, n. 3, p. 29, 2015. Citado nas páginas 94, 95 e 98.

ZHENG, Y.; CHEN, Y.; LI, Q.; XIE, X.; MA, W.-Y. Understanding transportation modes based on gps data for web applications. **ACM Transactions on the Web (TWEB)**, ACM, v. 4, n. 1, p. 1, 2010. Citado na página 99.

ZHENG, Y.; FU, H.; XIE, X.; MA, W.-Y.; LI, Q. **Geolife GPS trajectory dataset - User Guide**. [S.l.], 2011. Disponível em: <https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/>. Citado na página 110.

ZHENG, Y.; LI, Q.; CHEN, Y.; XIE, X.; MA, W.-Y. Understanding mobility based on gps data. In: ACM. **Proceedings of the 10th international conference on Ubiquitous computing**. [S.l.], 2008. p. 312–321. Citado nas páginas 99 e 109.

ZHENG, Y.; LIU, L.; WANG, L.; XIE, X. Learning transportation mode from raw gps data for geographic applications on the web. In: ACM. **Proceedings of the 17th international conference on World Wide Web**. [S.l.], 2008. p. 247–256. Citado na página 99.

ZHENG, Y.; XIE, X. Learning travel recommendations from user-generated gps traces. **ACM Trans. Intell. Syst. Technol.**, ACM, New York, NY, USA, v. 2, n. 1, p. 2:1–2:29, jan. 2011. ISSN 2157-6904. Disponível em: <http://doi.acm.org/10.1145/1889681.1889683>. Citado na página 33.

ZHENG, Y.; XIE, X.; MA, W.-Y. Geolife: A collaborative social networking service among user, location and trajectory. **IEEE Data Eng. Bull.**, Citeseer, v. 33, n. 2, p. 32–39, 2010. Citado na página 109.

ZHENG, Y.; ZHANG, L.; MA, Z.; XIE, X.; MA, W.-Y. Recommending friends and locations based on individual location history. **ACM Trans. Web**, ACM, New York, NY, USA, v. 5, n. 1, p. 5:1–5:44, fev. 2011. ISSN 1559-1131. Disponível em: <http://doi.acm.org/10.1145/1921591.1921596>. Citado na página 33.

ZHENG, Y.; ZHANG, L.; XIE, X.; MA, W.-Y. Mining interesting locations and travel sequences from GPS trajectories. In: **Proceedings of the 18th international conference on World wide web**. New York, NY, USA: ACM, 2009. (WWW '09), p. 791–800. ISBN 978-1-60558-487-4. Disponível em: <http://doi.acm.org/10.1145/1526709.1526816>. Citado nas páginas 33 e 35.

_____. Mining interesting locations and travel sequences from gps trajectories. In: ACM. **Proceedings of the 18th international conference on World wide web**. [S.l.], 2009. p. 791–800. Citado na página 109.

ZHENG, Y.; ZHOU, X. **Computing with Spatial Trajectories**. [S.l.]: Springer New York, 2011. Citado nas páginas 25, 31, 32 e 34.

ZHUANG, Z.; KIM, K.-H.; SINGH, J. P. Improving energy efficiency of location sensing on smartphones. In: ACM. **Proceedings of the 8th international conference on Mobile systems, applications, and services**. [S.l.], 2010. p. 315–330. Citado na página 76.

GLOSSÁRIO

Backend: é a parte do *software* que processa a entrada do *frontend*. Diferentemente do código do *frontend* que encontra-se no lado cliente, este código reside no servidor.

CSS: abreviação para a expressão inglesa *Cascading Style Sheets* (CSS), é um simples mecanismo para adicionar estilo (cores, fontes, espaçamento etc) a um documento web. Em vez de colocar a formatação dentro do documento, o CSS cria um link (ligação) para uma página que contém os estilos. Quando quiser alterar a aparência do portal basta portanto modificar apenas um arquivo.

Framework: é uma abstração que une códigos comuns entre vários projetos de *software* provendo uma funcionalidade genérica. *Frameworks* são projetados com a intenção de facilitar o desenvolvimento de *software*, habilitando designers e programadores a gastarem mais tempo determinando as exigências do *software* do que com detalhes de baixo nível do sistema.

Frontend: em projeto de *software*, o *frontend* é a parte do *software* que interage com os usuários. Geralmente fazem referência às interfaces gráficas, mas também as linhas de comando são levadas em conta.

HTML: abreviação para a expressão inglesa *HyperText Markup Language*, que significa Linguagem de Marcação de Hipertexto. É uma linguagem de marcação utilizada na construção de páginas na Web. Documentos HTML podem ser interpretados por navegadores.

JavaScript: é uma linguagem de programação interpretada. Foi originalmente implementada como parte dos navegadores web para que *scripts* pudessem ser executados do lado do cliente e interagissem com o usuário sem a necessidade deste *script* passar pelo servidor, controlando o navegador, realizando comunicação assíncrona e alterando o conteúdo do documento exibido.

Padrões de projeto: ou *Design Pattern*, descreve uma solução geral reutilizável para um problema recorrente no desenvolvimento de sistemas de *software* orientados a objetos. Não é um código final, é uma descrição ou modelo de como resolver o problema do qual trata, que pode ser usada em muitas situações diferentes.

REST: a *Representational State Transfer* (REST), em português Transferência de Estado Representacional, é uma abstração da arquitetura da *World Wide Web* (Web), um estilo arquitetural que consiste de um conjunto coordenado de restrições arquiteturais aplicadas a componentes, conectores e elementos de dados dentro de um sistema de hipermídia distribuído. O REST ignora os detalhes da implementação de componente e a sintaxe de protocolo com o objetivo de focar nos papéis dos componentes, nas restrições sobre sua interação com outros componentes e na sua interpretação de elementos de dados significantes.

Template: é um documento sem conteúdo, com apenas a apresentação visual (apenas cabeçalhos por exemplo) e instruções sobre onde e qual tipo de conteúdo deve entrar a cada parcela da apresentação.

Web: sinônimo mais conhecido de *World Wide Web* (WWW). É a interface gráfica da Internet que torna os serviços disponíveis totalmente transparentes para o usuário e ainda possibilita a manipulação multimídia da informação.