

**ROTEAMENTO BASEADO NA INFORMAÇÃO
ESPAÇO-TEMPORAL DE TRAJETÓRIAS EM
REDES VEICULARES AD HOC**

CLAYSON SANDRO FRANCISCO DE SOUSA CELES

**ROTEAMENTO BASEADO NA INFORMAÇÃO
ESPAÇO-TEMPORAL DE TRAJETÓRIAS EM
REDES VEICULARES AD HOC**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação.

**ORIENTADOR: ANTONIO ALFREDO FERREIRA LOUREIRO
COORIENTADORES: ROSSANA MARIA DE CASTRO ANDRADE E
REINALDO BEZERRA BRAGA**

Belo Horizonte
Dezembro de 2013

© 2013, Clayson Sandro Francisco de Sousa Celes.
Todos os direitos reservados.

Celes, Clayson Sandro Francisco de Sousa

C392r Roteamento baseado na Informação
Espaço-Temporal de Trajetórias em Redes Veiculares
Ad hoc / Clayson Sandro Francisco de Sousa Celes. —
Belo Horizonte, 2013
xxviii, 96 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de
Minas Gerais

Orientador: Antonio Alfredo Ferreira Loureiro
Coorientadores: Rossana Maria de Castro Andrade e
Reinaldo Bezerra Braga

1. Computação - Teses. 2. Redes de Computadores -
Administração - Teses. 3. Roteamento (Administração
de redes de computadores) - Teses. I. Orientador.
II.Coorientadores. III. Título.

CDU 519.6*22.3(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Roteamento baseado na informação espaço-temporal de trajetórias em redes
veiculares ad hoc

CLAYSON SANDRO FRANCISCO DE SOUSA CELES

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. ANTONIO ALFREDO FERREIRA LOUREIRO - Orientador
Departamento de Ciência da Computação – UFMG

PROF. REINALDO BEZERRA BRAGA- COORIENTADOR
GREat - UFC

PROFA. ROSSANA MARIA DE CASTRO ANDRADE - COORIENTADORA
Departamento de Computação - UFC

PROFA. FÁTIMA DE LIMA PROCÓPIO D. FIGUEIREDO
Departamento de Ciência da Computação - PUC-MG

PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 09 de dezembro de 2013.

Agradecimentos

Agradeço a Deus por ter me dado saúde para conquistar mais um objetivo. Por abençoar-me de pessoas maravilhosas que sempre me motivaram nas minhas escolhas.

Aos meus familiares, especialmente aos meus pais, que acreditaram no meu potencial e investiram para a minha formação educacional. À minha querida Karine Veríssimo, pelo incentivo constante, pelos momentos de descontração e principalmente pela cumplicidade nas conquistas dos meus objetivos.

Agradeço ao meu orientador Antonio Alfredo Ferreira Loureiro por todo o conhecimento repassado e por todas as experiências acadêmicas compartilhadas. Adicionalmente, gostaria de agradecer a todos os membros do laboratório de redes da UFMG pelo ambiente descontraído de aprendizado e amizade.

Aos meus amigos e companheiros de pesquisa Reinaldo Braga e Carina Oliveira pela motivação constante e todas contribuições para os trabalhos.

Agradeço aos professores Rossana Andrade, Cidcley Teixeira e Elizabeth Furtado, responsáveis pela minha formação acadêmica durante a graduação. Pessoas que prezo muito, devido a todos os ensinamentos durante os anos de iniciação científica.

Aos amigos da Los-Computeros Rodolfo Carneiro, Alan Castro, Felipe Ávila, Rickson Guidolini, Rogério Fonteles e Wallace Favoreto e aos amigos que fiz em Belo Horizonte, por todos os momentos de descontração. Com certeza, vocês foram de grande ajuda durante o mestrado.

À banca pelas contribuições para este trabalho, ao PPGCC e ao CNPq pelo apoio financeiro.

“A smooth sea never made a skillful sailor.”
(Old English proverb)

Resumo

Redes veiculares ad hoc (VANETs - *Vehicular Ad Hoc NETWORKs*) têm sido frequentemente utilizadas para oferecer soluções de troca de dados em sistemas de transportes inteligentes. Devido às características de alta mobilidade dos nós, frequentes mudanças de topologia e rápidas variações de escalabilidade, um dos principais desafios das VANETs é a concepção de esquemas robustos de encaminhamento de mensagens. Neste sentido, estratégias baseadas em posição geográfica têm sido frequentemente adotadas em protocolos de roteamento. Além disso, a incorporação de sistema de navegação (GPS - *Global Positioning Systems*) nos veículos possibilita um avanço no roteamento baseado em informação geográfica, pois as trajetórias dos veículos passam a ser utilizadas no processo de encaminhamento. No entanto, apesar de existirem vários protocolos que exploram o uso de trajetórias de veículos, poucos focam no histórico de trajetórias para auxiliar no processo de encaminhamento de mensagens na rede. Nesta dissertação, são apresentados esquemas de roteamento para redes veiculares, que exploram o uso de informações espaço-temporais de trajetórias de veículos. Primeiramente, o foco é a análise dos dados de trajetórias coletados pelas rotinas dos veículos e a descoberta de padrões de trajetórias, que indicam rotas que são frequentemente seguidas pelos usuários entre uma origem e um destino. Em seguida, esses padrões de trajetórias são aplicados como recursos para auxiliar no processo de roteamento de mensagens em uma VANET. A principal contribuição deste trabalho está relacionada ao aumento da taxa de entrega de pacotes entre usuários pertencentes às VANETs esparsas, que são vistas como cenários que trazem dificuldades para construir rotas fim-a-fim. Além da melhoria da taxa de entrega, os resultados mostram a eficiência da solução proposta na detecção de melhores rotas para o encaminhamento de mensagens, assim como uma nova abordagem de roteamento oportunístico baseado no histórico de deslocamento dos veículos.

Palavras-chave: VANET, Padrões de Trajetórias, Roteamento, Informação Espaço-Temporal.

Abstract

Vehicular ad hoc networks have frequently been used to offer solutions for data exchange in intelligent transport systems. Due to the characteristics of high mobility, frequent topology changes and rapid variations of scalability, one of the main challenges of VANETs is the design of robust schemes for message forwarding. In this sense, strategies based on geographic position have been often adopted by routing protocols. Moreover, the incorporation of in-vehicle navigation systems enabled an improvement in routing based on geographic information because the trajectories of vehicles shall be used for the routing process. However, there are several protocols that exploit the use of trajectories of vehicles, a few of them focus on the historical trajectories to aid in message routing process in the network. In this dissertation, routing schemes for vehicular networks, exploiting the use of information spatiotemporal trajectories of vehicles are presented. First, the focus is on the analysis of data collected from vehicles during their routine trajectories and the discovery of patterns of those trajectories, which denote routes that are often followed by users between a source and a destination. These patterns of trajectories are used to assist in the routing of messages in a VANET. The main contribution of this work is related to the increased rate of packet delivery among users belonging to sparse VANETs, which are seen as scenarios that make it difficult to build end-to-end routes. Besides the improvement of the delivery rate, simulation results show the efficiency of the proposed solution in detecting the best routes for routing messages, as well as a new approach to opportunistic routing based on the historical movement of vehicles.

Keywords: VANET, Trajectories Pattern, Routing, SpatioTemporal Information.

Lista de Figuras

1.1	Visão geral da proposta deste trabalho.	3
2.1	Exemplo de disseminação por <i>unicast</i>	13
2.2	Exemplo de disseminação por <i>broadcast</i>	14
2.3	Exemplo de disseminação por <i>multicast</i>	15
2.4	Exemplo de disseminação por <i>geocast</i>	16
2.5	Exemplo de informação visual de dados de tráfego na cidade de São Francisco (EUA) em um dia da semana, por volta das 7h30 [Li et al., 2007]. . .	17
2.6	Ilustração da detecção de comportamento de mobilidade em um conjunto de dados [Giannotti et al., 2010].	18
2.7	Representação visual do algoritmo <i>Predestination</i> [Krumm & Horvitz, 2007].	18
2.8	Exemplo de uma trajetória. A tabela possui os dados de latitude, longitude e instante de registro de uma trajetória com identificador (id) 19.	20
2.9	Relação entre trajeto e trajetórias de um objeto em movimento.	21
2.10	Exemplo de <i>clustering</i>	23
2.11	Exemplo de detecção de padrões de frequência.	23
2.12	Exemplo de aplicação para predição de trajetória.	24
2.13	Representação da movimentação de três objetos em relação ao espaço e ao tempo durante $t = [1, 4]$ [Jeung et al., 2011].	25
2.14	Padrões de trajetórias espaciais e espaço-temporais [Jeung et al., 2011]. . .	26
2.15	Identificação de trajetórias representativas (Adaptada de Jeung et al. [2011])	26
2.16	Exemplo do processo de roteamento de uma mensagem do ponto P_1 até D utilizando o GeOpps [Leontiadis & Mascolo, 2007].	30
2.17	Ilustração de cenário para o TBD (Adaptada de [Jeong et al., 2011]). . . .	33
2.18	Ilustração de funcionamento do TBD.	34
2.19	Exemplo de cenário para o roteamento de uma mensagem do ponto $AP1$ até o veículo destino, utilizando o TSF (Adaptada de Jeong et al. [2012]). .	36

2.20	Fluxograma de relacionamento entre os protocolos de roteamento geográfico com base em trajetórias de veículos.	39
3.1	Ilustração da diferença entre a exploração da informação espacial e temporal. Considerando a visão espacial, somente um grupo é identificado. No entanto, quando a visão temporal é também utilizada, dois grupos distintos são identificados.	45
3.2	Ilustração da existência de pontos de interesse durante a movimentação diária.	50
3.3	Ilustração das definições de <i>core object</i> e <i>reachability distance</i> no algoritmo OPTICS.	51
3.4	Identificação de <i>clusters</i> no trajeto entre o local de trabalho (<i>work</i>) e casa (<i>home</i>) de acordo com a rota seguida [Andrienko et al., 2007].	52
3.5	Representação de trajetórias sumarizadas a partir dos <i>clusters</i> da figura 3.4 [Andrienko et al., 2007].	53
3.6	Ilustração do processo para obter trajetórias representativas de um usuário.	55
3.7	Taxa de Entrega em função do número de mensagens e TTL.	60
3.8	Número de mensagens descartadas em função do número de mensagens e TTL.	61
3.9	<i>Overhead</i> em função do número de mensagens e TTL.	62
3.10	Latência em função do número de mensagens e TTL.	63
4.1	Representação de uma região de encontro identificada pela mobilidade entre dois veículos.	67
4.2	Principais situações para considerar uma aproximação entre trajetórias representativas de dois veículos.	68
4.3	Ilustração de <i>buffer</i> espacial e temporal.	69
4.4	Ilustração de aplicação do algoritmo de correlação espacial para detecção de regiões de interesse entre duas trajetórias representativas.	70
4.5	Resultado da aplicação do algoritmo de <i>Parzen window</i> para amostras do veículo 9 que passa pela região de interesse <i>A</i> com diferentes valores de h_n .	72
4.6	Resultado da aplicação do algoritmo de <i>Parzen window</i> para amostras do veículo 9 e 37 que passa pela região de interesse <i>A</i> com diferentes valores de h_n	73
4.7	Um exemplo da representação de encontros em um grafo.	75
4.8	Mapa de estradas e movimentações dos veículos na cidade de Berlin.	77
4.9	Trajeto agregadas dos veículos por dia da semana.	78
4.10	Gráficos da relação entre quantidade de trajetórias por marcações temporais.	79

4.11 Taxa de entrega em função do número de veículos.	80
4.12 <i>Overhead</i> em função do número de veículos.	81
4.13 Atraso em função do número de veículos.	82
4.14 Número de mensagens transmitidas em função do número de veículos.	83

Lista de Tabelas

2.1	Comparação qualitativa entre os protocolos de roteamento geográfico com base em trajetórias de veículos.	40
3.1	Características do GPS, GSM e WiFi quando utilizados como técnicas de localização.	47
3.2	Relação entre os problemas e métodos para solucioná-los.	49

Lista de Algoritmos

1	GeoSPIN - Fase 1 - Obtenção de trajetórias representativas	55
2	GeoSPIN - Fase 2 - Encaminhamento de Mensagens	57
3	Correlação Espacial - Identificando regiões de proximidade entre trajetórias de veículos	69
4	Correlação Temporal - algoritmo para efetuar a correlação temporal entre duas trajetórias em uma dada região de interesse.	74

Lista de abreviaturas

AP Access Point

CCT Centro de Controle de Tráfego

CD Core Distance

CVIS Cooperative Vehicle Infrastructure Systems

DSRC Dedicated Short Range Communications

DTN Delay and Disruption Tolerant Network

EDD Expected Delivery Delay

EDR Expected Delay Ratio

ETA Estimate the Time of Arrival

GeOpps Geographical Opportunistic Routing for Vehicular Networks

GeoSPIN Geographic routing based on SPatial INformation

GPS Global Positioning System

GSM Global System for Mobile Communications

GSR Geographic Source Routing

IEEE Institute of Electrical and Electronics Engineers

ITS Intelligent Transport Systems

MANET Mobile Ad hoc Network

METD Minimum Estimated Time of Delivery

NP Nearest Point

OBE OnBoarding Equipment

OPTICS Ordering Points To Identify the Clustering Structure

PDF Probability Density Function

PoI Point of Interest

RD Reachability Distance

RoI Region of Interest

ROSTER ROuting based on Spatio-Temporal Encounter Record

RSU Road Side Unit

SAW Spray And Wait

STDFS Shared Vehicle Trajectories for Data Forwarding Scheme

SUMO Simulation of Urban Mobility

TBD Trajectory-Based Data Forwarding for Light-Traffic Vehicular Networks

TSF Trajectory-based Statistical Forwarding

TTL Time To Live

V2I Vehicle to Infrastructure

V2V Vehicle to Vehicle

VADD Vehicle-Assisted Data Delivery in vehicular ad hoc networks

VANET Vehicular Ad Hoc NETwork

VNI Virtual Navigation Interface

WAVE Wireless Access in Vehicle Environments

ZOF Zone of Forwarding

ZOR Zone of Relevance

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xxi
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	3
1.3 Contribuições	4
1.4 Organização desta dissertação	4
2 Revisão Bibliográfica	7
2.1 Redes Veiculares	7
2.1.1 Conceitos	7
2.1.2 Características	9
2.1.3 Modelos de Disseminação de Dados	11
2.2 Movimentação de Objetos no Espaço Geográfico	16
2.2.1 Representação de Trajetória	19
2.2.2 Processo de Descoberta de Padrões de Trajetórias	22
2.2.3 Aspectos de Caracterização de Padrões de Trajetórias	24
2.3 Roteamento em VANET baseado em Trajetória de Veículos	27
2.3.1 GeOpps	28
2.3.2 GeoSpray	30
2.3.3 GeoDTN+Nav	31

2.3.4	TBD	32
2.3.5	TSF	35
2.3.6	STDFS	37
2.3.7	Discussão	38
2.4	Considerações Finais	42
3	Roteamento Baseado em Informação Espacial	43
3.1	Introdução	43
3.2	Explorando a Informação Espacial	45
3.3	Descrição do Algoritmo de Roteamento	53
3.3.1	Considerações Preliminares	53
3.3.2	Fase 1 - Obtenção de Trajetórias Representativas	54
3.3.3	Fase 2 - Encaminhamento de Mensagens	56
3.4	Análise de Desempenho	58
3.4.1	Descrição do Ambiente e Cenário de Simulação	58
3.4.2	Protocolos <i>Baseline</i>	59
3.4.3	Resultados	60
3.5	Considerações Finais	63
4	Roteamento Baseado em Informação Espaço-Temporal	65
4.1	Introdução	65
4.2	Explorando a Informação Espaço-Temporal	67
4.2.1	Correlação Espacial	68
4.2.2	Correlação Temporal	70
4.3	Descrição do Algoritmo de Roteamento	74
4.3.1	Considerações Preliminares	74
4.3.2	Criação da tabela de roteamento e Encaminhamento	75
4.4	Análise de Desempenho	77
4.4.1	Descrição do Ambiente e Cenário de Simulação	77
4.4.2	Resultados	80
4.5	Considerações Finais	83
5	Conclusão e Trabalhos Futuros	85
	Referências Bibliográficas	87

Capítulo 1

Introdução

Este capítulo apresenta a motivação, os objetivos e as contribuições desta dissertação. A motivação é apresentada na seção 1.1. A seção 1.2 apresenta os objetivos, bem como esboça a abordagem proposta. A seção 1.3 apresenta as contribuições. Por fim, a seção 1.4 apresenta a forma como este trabalho está organizado.

1.1 Motivação

Uma rede móvel *ad hoc* (MANET - *Mobile Ad hoc NETWORK*) é uma rede sem fio composta por nós móveis que cooperam entre si para possibilitar a troca de mensagens independente de qualquer infraestrutura previamente implantada [Tonguz & Ferrari, 2006]. Quando esses nós móveis são veículos, essas redes são chamadas de redes veiculares *ad hoc* (VANET - *Vehicular Ad hoc NETWORK*). Uma VANET pode ser estendida com nós estáticos que são instalados ao longo de rodovias e vias urbanas. Com isso, as redes veiculares fornecem uma estrutura de comunicação de dados para o provimento de serviços típicos de sistemas de transporte inteligente, tais como monitoramento [Gibbons et al., 2003], segurança [Chisalita & Shahmehri, 2002] e acesso à Internet [Gerla & Kleinrock, 2011].

No entanto, em comparação com redes móveis *ad hoc* convencionais, uma rede veicular não é só sujeita a rápidas mudanças de topologia, mas também a variações frequentes da densidade na rede. Por exemplo, geralmente, uma rede veicular em um centro urbano possui uma maior densidade de veículos em determinados momentos do dia, enquanto uma rede veicular em uma região rural ou entre horários de pico tende a ser mais esparsa e, com isso, suscetível a frequentes desconexões. Dessa forma, os desafios de pesquisa em VANETs são normalmente associados à alta mobilidade dos nós, cenários dinâmicos e com variações de escalabilidade em relação ao número de nós.

Portanto, é importante a concepção de protocolos de encaminhamento de mensagem robustos para proporcionar uma comunicação eficaz entre os nós nessas situações.

Um expressivo número de pesquisas em redes veiculares tem focado em protocolos de roteamento que contemplam o tratamento do problema de *broadcast storm* [Ni et al., 1999] em redes com topologia densa, considerando a suposição que uma rede veicular tem a característica de sempre se manter conectada. No entanto, o problema de desconexões da rede causando as fragmentações na topologia é também um crucial desafio de pesquisa em redes veiculares [Wisitpongphan et al., 2007]. Dessa forma, alguns esforços de pesquisa têm sido feitos para a criação de protocolos direcionados para suportar diferentes configurações de topologia.

Nesse sentido, estratégias baseadas em posição geográfica em combinação com a técnica armazena-e-encaminha têm sido utilizadas por um significativo número de protocolos de roteamento [Bernsen & Manivannan, 2009] [Bilal et al., 2013]. Recentemente, alguns estudos tentam melhorar o roteamento geográfico com o uso dos sistemas de navegação (GPS - *Global Positioning System*) embutidos nos veículos, de forma que a trajetória seja utilizada no processo de encaminhamento de mensagens.

A ideia de utilização de sistemas de navegação é bastante promissora e pode fornecer informações valiosas para melhorar o desempenho dos protocolos de roteamento. No entanto, é observado que as pessoas precisam de um sistema de navegação quando elas estão indo para lugares desconhecidos ou seguem rotas não familiares. No dia-a-dia, pessoas seguem rotas familiares entre suas casas e local de trabalho, por exemplo. Em [Froehlich & Krumm, 2008a] [Froehlich & Krumm, 2008b], Froehlich e Krumm mostram que as pessoas seguem um significativo número de trajetórias similares durante o deslocamento diário. Nesse sentido, a descoberta de padrões de trajetórias individuais a partir de dados GPS pode significar uma relevante informação para auxiliar no roteamento em redes veiculares, visto que a real representação de movimentação do usuário¹ é melhor obtida analisando o histórico de movimentação.

Diante disso, o foco desta dissertação é, primeiro, analisar dados de trajetórias coletados pelas rotinas dos veículos para descobrir padrões de trajetórias, que denotam rotas que são frequentemente seguidas pelos usuários entre um local de origem e um destino. Em seguida, esses padrões de trajetórias são avaliados como recursos para auxiliar no processo de roteamento de mensagens em uma rede veicular.

¹Nesta dissertação, usuário, condutor e motorista são utilizados indistintamente como a pessoa que conduz um veículo ao longo dos dias.

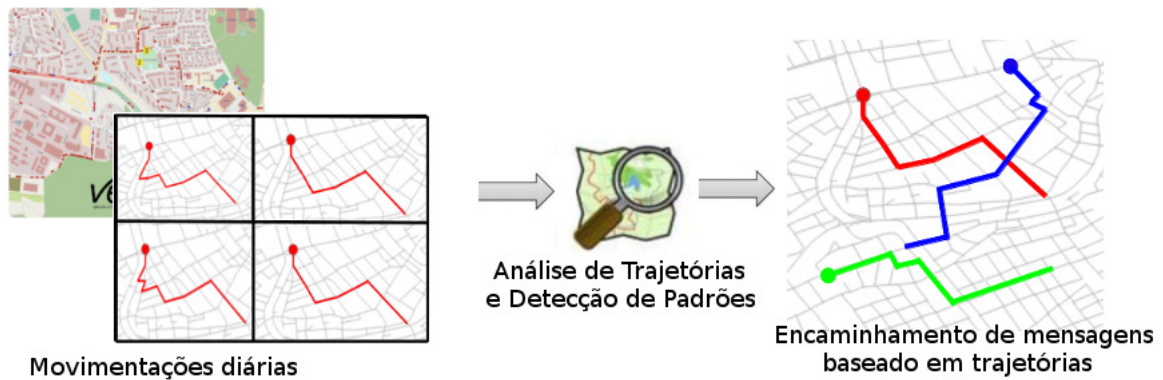


Figura 1.1: Visão geral da proposta deste trabalho.

1.2 Objetivos

Geralmente, as pessoas possuem rotinas com um elevado grau de regularidade espacial e temporal, em que cada indivíduo é caracterizado por frequentar certos lugares, em intervalos de tempo regulares [Gonzalez et al., 2008]. Da mesma forma, as pessoas costumam se deslocar por rotas familiares ao longo dos dias, ou seja, seguem rotas similares no trajeto entre suas residências e o local de trabalho, por exemplo. Assim, a fundamental suposição deste trabalho é que cada veículo possui padrões de deslocamento diário e, com essa informação obtida pelo histórico de movimentações, pode-se descrever comportamentos que representam uma rotina. Com isso, o objetivo geral deste trabalho é apresentar e avaliar esquemas de encaminhamento de mensagens para redes veiculares que utilizam como recurso padrões de trajetórias individuais, visando o aumento da taxa de entrega de mensagens, diminuição do *overhead* e do atraso de entrega.

A figura 1.1 apresenta uma visão geral da proposta deste trabalho. Inicialmente, a partir de dados de movimentações diárias individuais de veículos, é realizada uma análise nos dados e são aplicadas técnicas de mineração de trajetórias a fim de identificar padrões de movimentações. Em seguida, o processo de encaminhamento é feito de acordo com os padrões provenientes dos dados coletados pelos dispositivos GPS. Assim, os objetivos específicos são:

- Apresentar e avaliar um esquema de encaminhamento para redes veiculares que explore informações espaciais de trajetórias;
- Apresentar e avaliar um esquema de encaminhamento para redes veiculares que explore informações espaço-temporais de trajetórias.

1.3 Contribuições

A principal contribuição deste trabalho é a proposição e análise de esquemas de roteamento para redes veiculares que exploram o uso de informações espaciais e espaço-temporais de trajetórias de veículos para encaminhamento de mensagens. Especificamente, as contribuições obtidas deste trabalho de mestrado, relacionadas à contribuição principal, são descritas a seguir:

1. Aplicação da teoria de sistemas de informações geográficas e técnicas de mineração de dados para auxiliar no roteamento de mensagens em redes veiculares;
2. Proposta e avaliação de um esquema que utiliza informação espacial de trajetórias provenientes de dados de GPS no processo de encaminhamento de mensagens em redes veiculares;
3. Proposta e avaliação de um esquema que usa a informação espaço-temporal de trajetórias procedentes de dados de GPS para auxiliar no encaminhamento de mensagens em redes veiculares.

Além disso, fizemos uma revisão bibliográfica de abordagens de roteamento em redes veiculares que se beneficiam de trajetórias de veículos na tarefa de encaminhamento de mensagens e uma análise comparativa das abordagens identificadas na revisão bibliográfica.

1.4 Organização desta dissertação

Esta dissertação está dividida em cinco capítulos. O capítulo 1 apresenta o contexto, os objetivos e as contribuições deste trabalho. No capítulo 2, inicialmente são abordados alguns conceitos pertinentes a redes veiculares, elementos cuja compreensão é relevante para o entendimento das contribuições e resultados obtidos. Em seguida, é exposta uma teoria relacionada à movimentação de objetos no espaço geográfico, destacando a definição de trajetória e a detecção de padrões de movimentação. Ainda no capítulo 2, são analisados os diferentes protocolos de roteamento para redes veiculares que adotam trajetórias de veículos no processo de encaminhamento de mensagens.

O capítulo 3 apresenta um esquema de roteamento de mensagens, denominado GeoSPIN (*Geographic routing based on SPatial INformation*). Inicialmente, é dada uma fundamentação sobre um método para explorar informação espacial de dados GPS. Em seguida, são descritas as duas fases do GeoSPIN: obtenção de trajetórias representativas e encaminhamento de mensagens. Por fim, o esquema é avaliado e

comparado, em diferentes aspectos e cenários, por meio de simulações. Além disso, é mostrada a metodologia utilizada nas simulações e uma análise dos resultados obtidos.

O capítulo 4 apresenta um esquema de roteamento de mensagens, denominado ROSTER (*ROuting based on Spatio-Temporal Encounter Record*). Inicialmente, é apresentada a motivação para exploração de informações espaço-temporais. Em seguida, são apresentados algoritmos para estimar a correlação espacial e temporal de trajetórias de veículos. Após isso, o ROSTER é detalhado e, por fim, avaliado e comparado em diferentes aspectos e cenários por meio de simulações. Além disso, é mostrada a metodologia utilizada nas simulações e uma análise dos resultados obtidos.

Finalmente, no capítulo 5, são apresentadas as considerações finais do trabalho. Além de comentários sobre as propostas e os objetivos alcançados, também são indicados trabalhos futuros da pesquisa apresentada nesta dissertação.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, é apresentada uma revisão bibliográfica sobre os três principais campos de estudo relacionados a este trabalho, são eles: redes veiculares, movimentação de objetos no espaço geográfico e roteamento em redes veiculares baseado em trajetórias de veículos. Os dois primeiros apresentam os principais conceitos aplicados neste trabalho e o último aborda como diferentes protocolos de roteamento em redes veiculares têm utilizado trajetórias de veículos para o encaminhamento de mensagens. Com isso, este capítulo apresenta a seguinte estrutura: na seção 2.1, é apresentada uma visão geral das redes veiculares, na qual são introduzidos os conceitos, características e modelos de disseminação de dados; a seção 2.2 apresenta a teoria relacionada à movimentação de objetos no espaço geográfico, destacando-se a definição de trajetória e a detecção de padrões de movimentação; e finalmente, a seção 2.3 apresenta uma discussão sobre os protocolos de roteamento para redes veiculares que adotam trajetórias de veículos no processo de encaminhamento de mensagens.

2.1 Redes Veiculares

2.1.1 Conceitos

O crescimento populacional associado aos fatores como urbanização e globalização tem impactado negativamente nos diferentes modos de transporte urbano, acarretando problemas como congestionamentos e falta de segurança de tráfego [Cotton, 2013]. Paralelamente, os avanços das áreas de computação móvel e comunicação sem fio surgem como uma alternativa para criação de Sistemas de Transportes Inteligentes (ITS - *Intelligent Transport Systems*) [Qu et al., 2010].

Um ITS consiste de um conjunto de soluções tecnológicas que são empregadas aos

diferentes modos de transporte, visando o fornecimento de serviços de gerenciamento de tráfego, controle de situações e disseminação de informação. Para tanto, um ITS executa tarefas de aquisição de dados, tomada de decisão e comunicação. Particularmente, a comunicação é dada pela formação de uma rede veicular responsável por prover métodos de troca de dados em três tipos de domínios: intraveicular, *ad hoc* e infraestruturado [Al-Sultan et al., 2013], os quais são descritos a seguir.

- i. **Domínio intraveicular (*In-vehicle networking*):** A evolução dos sistemas computacionais em ambientes veiculares para uma arquitetura distribuída traz a necessidade de adoção de uma rede de comunicação de dados, denominada rede intraveicular, entre os equipamentos embutidos nos veículos (OBE - *OnBoarding Equipment*). Essa rede tem o objetivo de prover segurança, comodidade e entretenimento à motoristas e passageiros que utilizam diariamente os seus veículos. Entretanto, requisitos como tempo real, confiabilidade, escalabilidade e tolerância a falhas são determinantes no projeto de redes intraveiculares. Diversos campos de pesquisas fazem parte deste domínio, por exemplo, fusão de dados, inferência, automação e comunicação de dados.
- ii. **Domínio *ad hoc*:** Corresponde ao domínio no qual os nós (i.e., veículos) se comunicam formando uma rede *ad hoc* veicular, chamada VANET (*Vehicular Ad Hoc Network*). Uma VANET utiliza os fundamentos das redes *ad hoc* móveis (MANET - *Mobile Ad hoc Network*) para prover métodos de troca de dados entre veículos. Dessa forma, a comunicação entre os nós de origem e destino pode ser de único salto (i.e., os veículos se comunicam diretamente caso haja uma conexão sem fio disponível entre eles); ou múltiplos saltos (i.e., quando não existe uma conexão sem fio direta entre origem e destino, adota-se um protocolo de roteamento para encaminhar a mensagem por nós intermediários que assumem o papel de roteadores). Assim como as redes *ad hoc* tradicionais, uma VANET apresenta problemas fundamentais relacionados ao controle de topologia e roteamento, principalmente devido à inexistência de um tipo de infraestrutura e, conseqüentemente, de uma topologia conectada entre os nós da rede a qual depende da densidade e do padrão de mobilidade dos veículos.
- iii. **Domínio Infraestruturado:** Neste caso, o domínio *ad hoc* é estendido, visando amenizar os problemas de conectividade por meio da utilização de pontos de acessos instalados ao longo das rodovias, denominados RSU (*Road Side Unit*). Com a utilização de RSU, uma gama de novos serviços pode ser oferecida pelas redes veiculares, por exemplo, disseminação de dados para uma região com baixa den-

sidade de veículos, difusão de informações geograficamente limitadas e acesso à Internet. No entanto, fazer o uso de RSU em larga escala aumenta os custos da rede, comprometendo a viabilidade da solução. No entanto, existem na literatura abordagens híbridas que envolvem uma arquitetura de comunicação em redes veiculares a partir da junção dos domínios *ad hoc* e infraestruturado. Nesse contexto, o termo VANET, típico de rede *ad hoc*, é utilizado por muitos pesquisadores sem distinção entre a presença de infraestrutura ou não.

2.1.2 Características

Uma VANET possui características particulares que a diferencia dos modelos tradicionais de MANET, principalmente devido ao comportamento dos motoristas, às restrições de mobilidade e à velocidade de deslocamento dos veículos. Dessa forma, é importante entender essas características antes de propor qualquer protocolo de comunicação em VANET. A seguir, são apresentadas e discutidas as principais características relacionadas ao projeto e funcionamento dessas redes.

- i. **Mobilidade Restrita:** Em uma VANET, a mobilidade dos nós é restrita à topologia das rodovias e às leis de tráfego impostas pelos órgãos de controle de trânsito. Além disso, os veículos comumente se movem em uma velocidade mais alta que os nós em outros tipos de MANET. Esses fatores influenciam diretamente na conectividade entre os veículos. Por exemplo, veículos se movendo em uma rodovia no mesmo sentido possuem geralmente velocidades similares, podendo assim ter um tempo de contato de conectividade maior que veículos que se movem em sentidos opostos. Com esse exemplo, percebe-se que o estudo da mobilidade e dos padrões de tráfego tem um papel fundamental para o projeto e desempenho de protocolos de roteamento em VANET.

No entanto, por limitações logísticas, econômicas e tecnológicas existe uma barreira para o estudo real dessa relação entre protocolos de roteamento e a mobilidade veicular. Portanto, a adoção de um ambiente simulado torna-se a melhor opção para avaliação de protocolos de roteamento [Joerer et al., 2012]. Dois aspectos críticos para criação de simulação em VANET são a definição de um modelo de mobilidade e a configuração de uma topologia, que reflete um real comportamento de tráfego [Sommer & Dressler, 2008]. Para auxiliar no projeto de simulação, Harri et al. [2009] definem uma taxonomia e discutem os principais modelos de mobilidade empregados para simulação em VANET fazendo um paralelo com diferentes tipos de topologia.

ii. **Mobilidade previsível:** As redes veiculares são compostas por diversos tipos de veículos (e.g., públicos e privados). Os veículos públicos, geralmente, seguem rotas fixas em horários de deslocamento predeterminados. Os veículos privados não possuem uma mobilidade tão rígida como os públicos. No entanto, percebe-se que, ao longo dos dias, veículos privados se deslocam com base em uma rotina, ou seja, utilizam as mesmas rotas no mesmo período do dia (e.g., a rota entre casa e o trabalho). Esse conhecimento sobre mobilidade diária aliado às informações de velocidade, localização e trajetória de um veículo pode ser utilizado para estimar movimentos futuros. Por exemplo, Namboodiri & Gao [2007] e Hung et al. [2008] propõem estratégias de roteamento para redes veiculares que consideram padrões de mobilidade com base na velocidade e posição dos veículos.

iii. **Densidade variável da rede:** A densidade da rede em uma VANET varia dependendo da intensidade de tráfego dos veículos, a qual é sensível à região, tempo e situação. Esses fatores impactam na densidade do seguinte modo: a tipologia da região (e.g., comercial, demografia) e o tempo (e.g., feriados, dias úteis, momento do dia), na taxa de concentração de veículos; e as situações do cotidiano (e.g., acidentes, condições climáticas, congestionamentos), no fluxo de tráfego local de uma região.

A comunicação entre os veículos é dada pela conectividade sem fio estabelecida entre eles, sendo assim, dependente diretamente da densidade de veículos. No entanto, devido aos fatores apresentados anteriormente, o estabelecimento da conectividade é um desafio no processo de comunicação em redes veiculares, principalmente em regiões de baixa concentração de veículos, pois a rede é sujeita à desconexões frequentes, ocasionando o problema de conectividade intermitente [Oliveira et al., 2007]. Como solução para esse problema, o uso de pontos de acesso pode auxiliar no processo de comunicação. Entretanto, em caso de não aplicabilidade de pontos de acesso, esquemas de comunicação de dados sem o uso de infraestrutura podem ser utilizados. Tipicamente, esses esquemas são empregados em redes tolerantes a atrasos e desconexões (DTN - Delay and Disruption Tolerant Network) e podem ser adaptados para serem utilizados em uma VANET [Pereira et al., 2012].

iv. **Topologia dinâmica da rede:** Além das características supracitadas, a velocidade de movimentação dos veículos e o comportamento dos motoristas afetam na dinamicidade da rede veicular. Assim, essas características dificultam o processo de definir e manter uma topologia conectada entre os nós da rede. Consequentemente,

impactam negativamente na operação dos protocolos de roteamento sensíveis às rápidas mudanças de topologia.

- v. **Recursos:** Ao contrário dos nós de uma MANET, que possuem limitações de recursos computacionais e de energia, os nós em uma VANET não afetados por tais limitações. Além disso, veículos são equipados com sensores que são utilizados para prover informações aos diversos serviços de uma VANET. Por exemplo, dispositivos GPS (*Global Positioning System*) fornecem informação de posicionamento geográfico dos veículos. Esses recursos potencializam as estratégias para roteamento em uma rede veicular, visto que fornecem tanto novas informações para auxiliar na disseminação de dados como amenizam as preocupações a respeito do consumo computacional e energético.

2.1.3 Modelos de Disseminação de Dados

A disseminação de dados é um dos principais requisitos para a operação de aplicações e fornecimento de serviços em ITS. Com o aumento do número de veículos com suporte às tecnologias de comunicação, pode-se vislumbrar a formação de uma rede *ad hoc* entre veículos, tanto para informar dinamicamente as condições das rodovias de modo que um sistema de transporte seguro e eficiente possa ser alcançado, como para fornecimento de serviços de entretenimento, segurança, saúde, etc.

Existem diversas pesquisas e projetos que objetivam melhorar os sistemas de transportes inteligentes utilizando uma arquitetura de comunicação entre veículos. O projeto CVIS (*Cooperative Vehicle Infrastructure Systems*) utiliza a agregação de informações coletadas por sensores nos veículos e do ambiente para fornecer funcionalidades de controle de tráfego e segurança [CVIS, 2013]. Sinais fisiológicos, tais como eletrocardiograma e pressão sanguínea, podem ser obtidos por uma plataforma de sensoriamento e enviados à um centro de monitoramento de saúde para avaliação do estado do condutor. Em geral, diversos estudos exploram a disseminação de dados entre veículos para expandir a assistência ao condutor e facilitar o desenvolvimento de sistemas de segurança [PATH, 2013] [CEMACS, 2013].

Visando a interoperabilidade e a implantação em uma escala maior desses serviços de ITS, um grupo de trabalho do IEEE tem desenvolvido o padrão 802.11p [Jiang & Delgrossi, 2008]. O atual modelo do IEEE 802.11 para cenários com infraestrutura ou *ad hoc* requer um tempo de contato para autenticação e comunicação não aceitável em uma rede na qual os nós móveis são veículos. O padrão IEEE 802.11p fornece aos dispositivos sem fio a capacidade para desempenhar a comunicação em situações que o tempo de duração de contato entre os veículos seja curto. Esse modo de operação, chamado de

WAVE (Wireless Access in Vehicle Environments) opera em uma frequência de 5.9 GHz e suporta o padrão *DSRC (Dedicated Short Range Communications)* [Uzcategui & Acosta-Marum, 2009].

Os veículos e as RSUs equipados com dispositivos de comunicação WAVE possibilitam a criação de uma rede *ad hoc* veicular. Semelhante a uma MANET, diferentes modelos de disseminação de dados podem ser suportados em uma VANET. A adoção de um determinado modelo está diretamente relacionada aos requisitos da aplicação. Por exemplo, aplicações de segurança, de um modo geral, realizam um *broadcast* local de mensagens, enquanto certas situações se beneficiam da comunicação por múltiplos saltos seguindo um modelo de disseminação *unicast*. O maior contraste entre esses modelos está na forma que as mensagens são disseminadas na rede. A seguir, uma visão geral desses modelos de disseminação no contexto de VANET é apresentada.

- i. **Unicast:** A comunicação em um modelo de disseminação de dados *unicast* consiste em encaminhar uma mensagem enviada de um nó origem e direcioná-la a um nó destino. No âmbito das redes *ad hoc* veiculares, um nó destino pode ser outro veículo ou uma RSU. O mecanismo de comunicação para a entrega de mensagem entre os nós pode ser tanto de único salto (quando o nó origem tem contato direto com o nó destino), como por múltiplos saltos (no qual um esquema de roteamento é utilizado para possibilitar a entrega da mensagem do nó origem ao destino). Nesse último caso, nós intermediários são utilizados como retransmissores para direcionar a mensagem ao destino. Por exemplo, na figura 2.1, o veículo *S* é a origem de uma mensagem e a RSU *D* é o destino. Ao executar o encaminhamento de múltiplos saltos (linha tracejada) a mensagem é direcionada de *S* até *D*.

No entanto, o processo de encaminhamento de mensagens por *unicast* é uma tarefa desafiadora devido à natureza da rede, principalmente devido à alta mobilidade dos nós e à instabilidade dos enlaces sem fio que causam a fragmentação temporária da rede. Por essa razão, diversos algoritmos *unicast* de múltiplos saltos têm adotado uma estratégia de envio de mensagens chamada armazena-e-encaminha (*store-and-forward*). Em outras palavras, a mensagem é armazenada para, em outro momento, ser enviada ao próximo nó, que pode ou não ser o destino.

Com a utilização dessa técnica de armazena-e-encaminha, o processo de encaminhamento de mensagens pode explorar o caráter oportunístico de encontro entre veículos. Dessa forma, a mobilidade dos nós e os encontros possibilitam a entrega de uma mensagem ao destino em aplicações tolerantes à atraso, mesmo sem o estabelecimento de conexão fim-a-fim entre origem e destino. Por exemplo, aplicações de notificações direcionadas de pontos de interesses, serviços de saúde e reportar

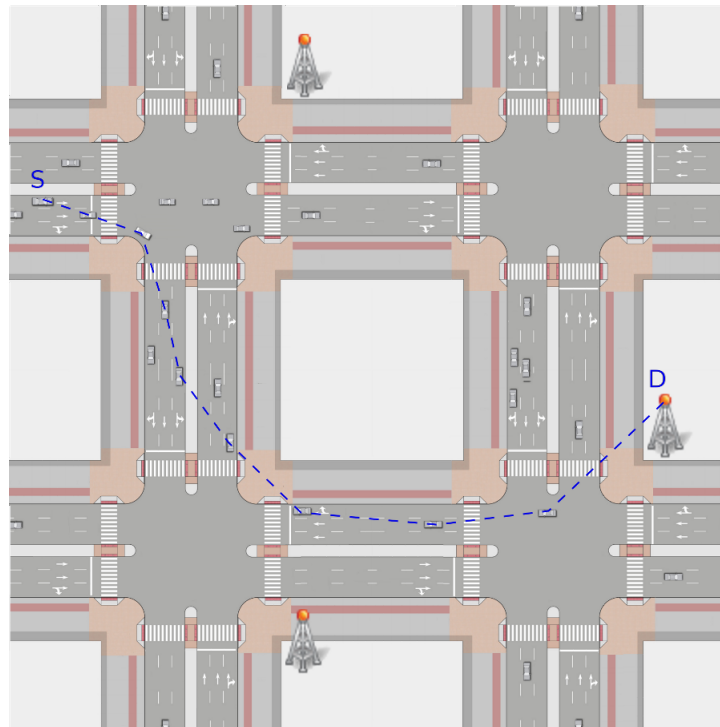


Figura 2.1: Exemplo de disseminação por *unicast*.

colisões ao centro de controle de tráfego podem ser suportadas por *unicast* [Schoch et al., 2008].

- ii. **Broadcast:** Um dos mais importantes benefícios de uma rede veicular refere-se a fornecer aplicações de segurança a partir da comunicação entre os veículos. Algumas aplicações podem não ter restrições de urgência de comunicação, por exemplo, densidade do tráfego e bloqueios na estrada, enquanto outras necessitam que a informação seja propagada de maneira rápida e confiável, como notificações de condições perigosas na rodovia e acidentes. Uma característica comum dessas aplicações é a necessidade de enviar a mensagem a um grupo de veículos, devido a isso o modelo de disseminação de dados *unicast* não é adequado.

O modelo de disseminação de dados por *broadcast* permite que uma mensagem emitida por um nó origem seja propagada pela rede por meio de uma inundação (*flooding*), comunicando um grupo de veículos. A disseminação inicia quando um veículo emite uma mensagem para os demais veículos da vizinhança. Esses veículos servem como retransmissores e, conseqüentemente, disseminam a mensagem aos veículos da vizinhança. A principal vantagem dessa abordagem é que um veículo não precisa saber o identificador de destino e uma rota para encaminhamento das mensagens. Dessa maneira, ele elimina etapas típicas de disseminação de dados em

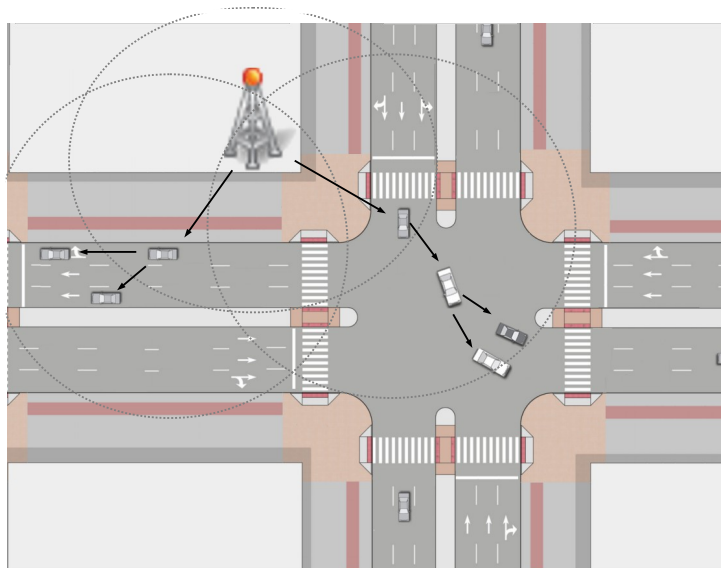


Figura 2.2: Exemplo de disseminação por *broadcast*.

redes *ad hoc*, como descoberta de rota e gerenciamento de topologia. Por exemplo, a figura 2.2 apresenta um exemplo de disseminação por *broadcast* iniciada pela RSU e propagada por meio de *flooding* aos veículos vizinhos.

No entanto, a adoção de um método de *flooding* resulta em problemas de redundância e colisões conhecidos como *Broadcast Storm* [Ni et al., 1999]. Geralmente, isso é contornado aplicando mecanismos eficientes de retransmissão de mensagens para evitar que uma grande quantidade de mensagens sejam enviadas, visto que muitos veículos na mesma vizinhança podem retransmitir as mensagens ao mesmo tempo.

- iii. **Multicast:** Em um modelo de disseminação *Multicast* as mensagens originadas por um nó origem são direcionadas a um grupo específico de vizinhos. A ideia é que as mensagens enviadas por *multicast* só atinjam os veículos relacionados à aplicação que utiliza esse modelo de disseminação. Dessa forma, os recursos são utilizados de maneira otimizada, por exemplo, em termos de largura de banda. Na figura 2.3, um veículo *S* faz o encaminhamento da mensagem aos veículos que se deslocam no sentido contrário, dessa forma, mesmo que existam outros veículos no raio de comunicação (região sombreada), eles não serão destino da mensagem.

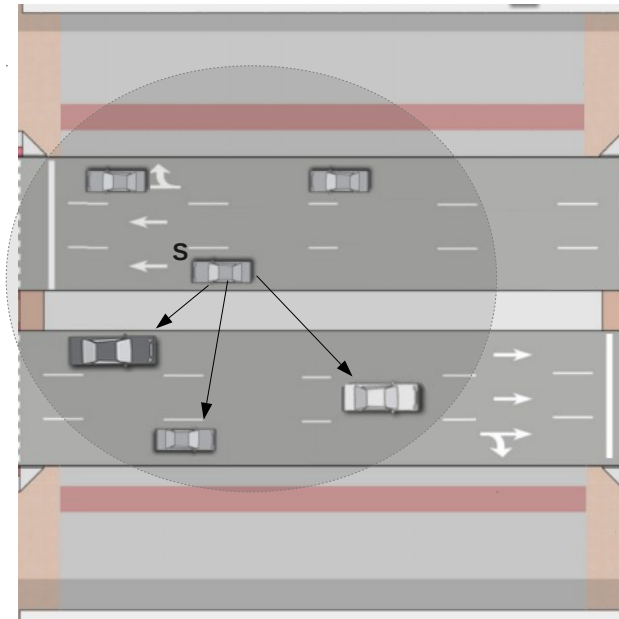


Figura 2.3: Exemplo de disseminação por *multicast*.

- iv. **Geocast:** O modelo de disseminação *geocast* consiste na entrega de mensagens a todos os veículos localizados em uma região geográfica, denominada zona de relevância (ZOR - *Zone Of Relevance*). Dessa forma, o *geocast* é considerado um modelo de *multicast* baseado em localização e possui dois modos de operação. O primeiro refere-se ao fato do remetente da mensagem estar localizado na ZOR, dessa forma, apenas os veículos contidos nessa zona recebem a mensagem. No segundo modo, o remetente está fora da ZOR e a sua operação é dividida em duas fases: na primeira, a mensagem originada pelo remetente é encaminhada salto-a-salto até a ZOR, em seguida, a segunda fase é executada por meio da disseminação da mensagem para todos os nós contidos na ZOR. Como o *geocast* é considerado um *multicast* para uma região específica, normalmente, define-se uma estratégia direcionada de encaminhamento de mensagens até o ZOR, por meio de uma zona de encaminhamento (ZOF - *Zone of Forwarding*), reduzindo a carga de mensagens na rede.

A figura 2.4 apresenta um exemplo de disseminação por *geocast*, no qual o veículo de origem da mensagem está localizado fora da ZOR (região sombreada). Neste caso, imagine uma situação na qual tenha ocorrido uma sequência de colisões que impediram o fluxo em um dos lados da rodovia. Nesse sentido, seria útil enviar mensagens para os veículos que se aproximam do ponto das colisões para que eles possam tomar rotas alternativas. Adotando um modelo de disseminação *geocast*,

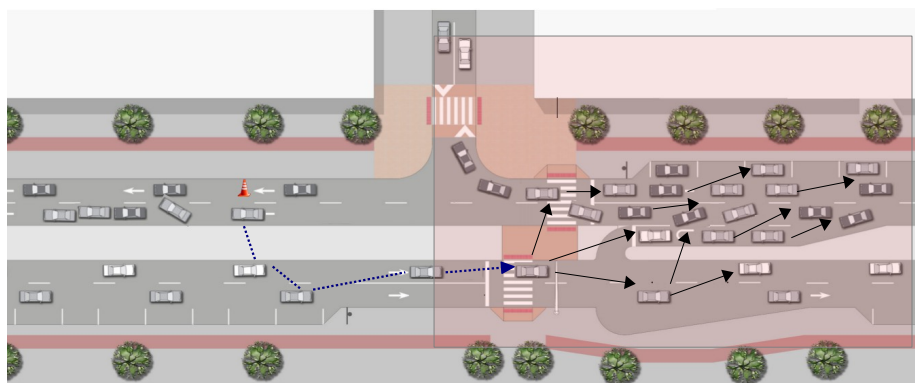


Figura 2.4: Exemplo de disseminação por *geocast*.

pode-se notificar os veículos presentes em uma determinada região (ZOR) para tomarem as rotas alternativas. Nesse caso, eles podem fazer o retorno ou entrar na via à direita. Percebe-se que, neste exemplo, foi utilizado um roteamento *unicast* (linha tracejada) para direcionar a mensagem até a ZOR, formando uma ZOF e, assim, evitando uma indesejável ocupação de largura de banda com os veículos não envolvidos com o evento.

Quando aplicado em uma VANET, o *geocast* tem três problemas relevantes para pesquisa: (1) particionamento de rede, que influencia diretamente o encaminhamento de mensagens, pois ele ocorre devido à inexistência de uma conectividade fim-a-fim entre os veículos origem e destino. Em redes esparsas, é mais comum a ocorrência desse tipo de problema, visto que partições ocorrem na rede e, conseqüentemente, inviabilizam a comunicação direta entre os veículos contidos em partições distintas; (2) *broadcast storm*, que ocorre quando os vários veículos que estão em uma região específica (ZOR) fazem um número alto de transmissões causando congestionamento da rede e colisões de pacotes; (3) peculiaridades da zona de relevância (ZOR), por exemplo, determinar a melhor configuração geométrica da ZOR. Nesse sentido, Jochle et al. [2012] avaliaram os efeitos da qualidade da comunicação e *overhead* de uma ZOR de acordo com a forma geométrica (i.e., círculo, retângulo, polígono).

2.2 Movimentação de Objetos no Espaço Geográfico

Os recentes avanços em tecnologias de localização, impulsionados pelo progresso nas áreas de computação móvel e comunicação sem fio, têm possibilitado a identificação

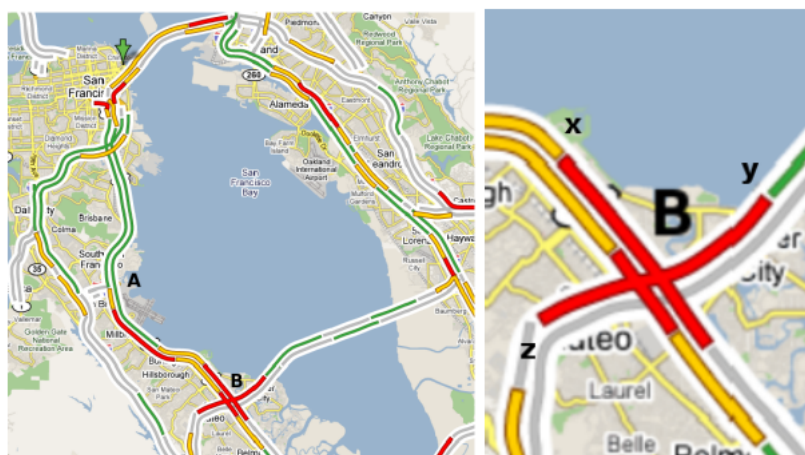


Figura 2.5: Exemplo de informação visual de dados de tráfego na cidade de São Francisco (EUA) em um dia da semana, por volta das 7h30 [Li et al., 2007].

do posicionamento de objetos (e.g., veículos, animais e pessoas) em ambientes internos e externos [Gu et al., 2009] [Boukerche et al., 2008]. O registro da localização desses objetos ao longo do tempo, remete ao conceito de trajetória, que é definida como um conjunto de dados originado pelo movimento de objetos e representada por uma sequência de informações com coordenadas espaciais e marcações temporais [Spaccapietra et al., 2008].

No contexto geográfico, o uso de dispositivos com sensores de posicionamento do tipo GPS (*Global Positioning System*) permite a captura e armazenamento de trajetórias espaciais em formato de coordenadas geográficas durante o movimento de pessoas e veículos. A análise da mobilidade a partir dessas trajetórias tem sido direcionada a aplicações de vários âmbitos, por exemplo:

- **Padrões de fluxo de tráfego:** Li et al. [2007] exploram o problema de descoberta de padrões de fluxo de tráfego em rodovias (*finding hot routes*). Os autores propuseram um algoritmo para detectar trechos de rodovias que possuem tráfego intenso de veículo. A figura 2.5 apresenta um cenário típico de implementação dessa abordagem. As cores representam os diferentes níveis de congestionamento, por exemplo, vermelho é o congestionamento intenso.
- **Comportamentos de Movimentação:** Aplicações deste tipo focam em detectar rotas e locais que os objetos costumam repetir com uma regularidade temporal [Giannotti et al., 2010]. A figura 2.6 apresenta um cenário típico de implementação dessa abordagem. Observa-se que três grupos de trajetórias em direção a região central são encontradas a partir de dados brutos de GPS;

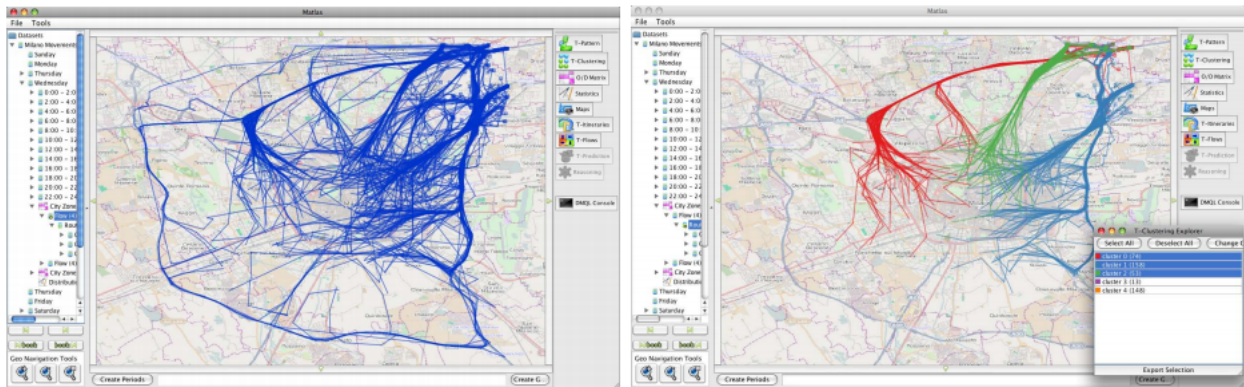


Figura 2.6: Ilustração da detecção de comportamento de mobilidade em um conjunto de dados [Giannotti et al., 2010].

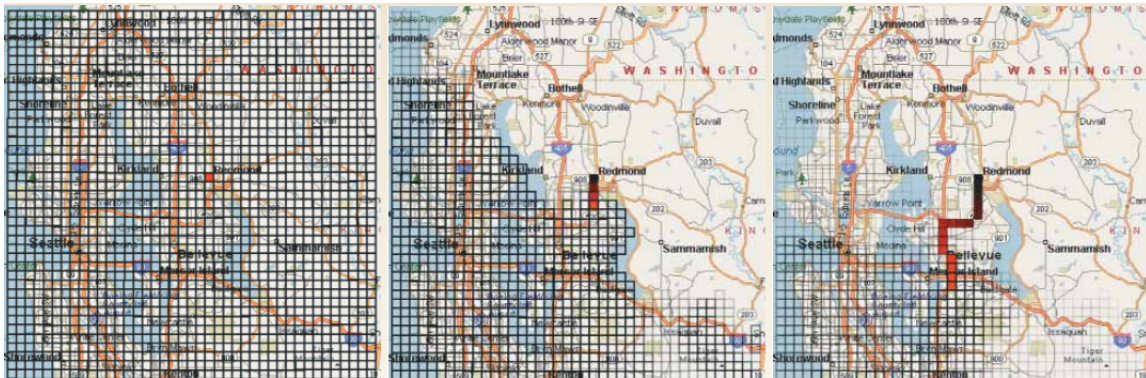


Figura 2.7: Representação visual do algoritmo *Predestination* [Krumm & Horvitz, 2007].

- **Planejamento de Rotas:** A utilização de sistemas de navegação nos veículos cria oportunidades para planejamento de rotas baseado em fatores como interesse ou intensidade de tráfego. Nesse sentido, Krumm & Horvitz [2007] propuseram um algoritmo, chamado *Predestination*, que fornece ao usuário rotas que consideram esses fatores durante o deslocamento. A figura 2.7 ilustra um caso de uso desse algoritmo, na qual o veículo inicia na célula vermelha (figura mais a esquerda) e durante o deslocamento deixa um rastro de células preenchidas. O algoritmo atualiza a probabilidade para cada célula e infere em qual célula o motorista terminará a viagem à medida que o veículo se movimenta.

A análise da mobilidade de pessoas e veículos nessas aplicações é dada pela investigação de registros realizados de forma ativa ou passiva. O registro ativo é aplicado em sistemas de rastreamento de trajetórias para compartilhamento de experiências [Reddy et al., 2010b] [Braga & Martin, 2011] [Ramos et al., 2011] ou durante o uso de redes sociais por meio de marcações de localizações ordenadas cronologicamente (e.g.,

check-ins no aplicativo *Foursquare*) [Noulas et al., 2011] [Cho et al., 2011]. No registro passivo, pessoas carregando *smartphones* capturam involuntariamente trajetórias espaciais por meio dos identificadores de cobertura das torres celulares (*cell tower ID*) [Becker et al., 2013]. Semelhantemente, a mobilidade veicular pode ser capturada por uso de *smartphones* ou dispositivos *GPS*. Por exemplo, os táxis possuem um sensor de localização que reporta as coordenadas geográficas dos trajetos ao longo do tempo. Tal reporte pode ser utilizado para caracterização da mobilidade de veículos no perímetro urbano, similar ao que foi realizado nas aplicações ilustradas nas figuras 2.5, 2.6 e 2.7.

Existem diversas ferramentas e técnicas para auxiliar no processo de captura, modelagem, manipulação, recuperação, análise e apresentação de movimentação de objetos [Lu et al., 2009]. Particularmente, na fase de análise, aplicando métodos estatísticos e mineração de dados, pode-se obter informações semânticas por meio da interpretação dos dados. Para isso, considera-se a premissa que existem informações ocultas presentes em uma dada base de dados, na forma de padrões. Assim, por meio dessa descoberta de padrões, identifica-se um conjunto de atividades, eventos e comportamentos que são desempenhados durante as movimentações diárias de objetos, tais como: pontos de interesse, tempo gasto no deslocamento entre pontos de interesse e reconstrução de rotas.

O restante do texto desta seção concentra nos conceitos relacionados à esses padrões de trajetórias, tais como processo genérico para descoberta e aspectos relevantes para caracterização. Primeiramente, para facilitar no entendimento desses conceitos, uma breve descrição da definição de trajetória é apresentada.

2.2.1 Representação de Trajetória

O estudo de dados de mobilidade baseia-se na movimentação de objetos. A movimentação de objetos pode representar o deslocamento diário de pessoas em veículos indo para o trabalho, pessoas equipadas com um dispositivo GPS enquanto realizam atividades ou migração de animais entre biótopos. Esses movimentos são representados na forma de dados espaço-temporais, denominados trajetórias, nas quais a informação espacial é determinada pelo posicionamento geográfico e a informação temporal é o instante do registro de posição durante o movimento.

A capacidade de capturar e armazenar uma sequência de posições com marcações no tempo é fundamental para o estudo de movimentações de objetos. No entanto, para o projeto de aplicações baseadas em trajetórias, é importante uma representação de dados bem definida. A representação do posicionamento do objeto em movimento é definida pela referência no espaço e no tempo [Parent et al., 1999]. No espaço, a

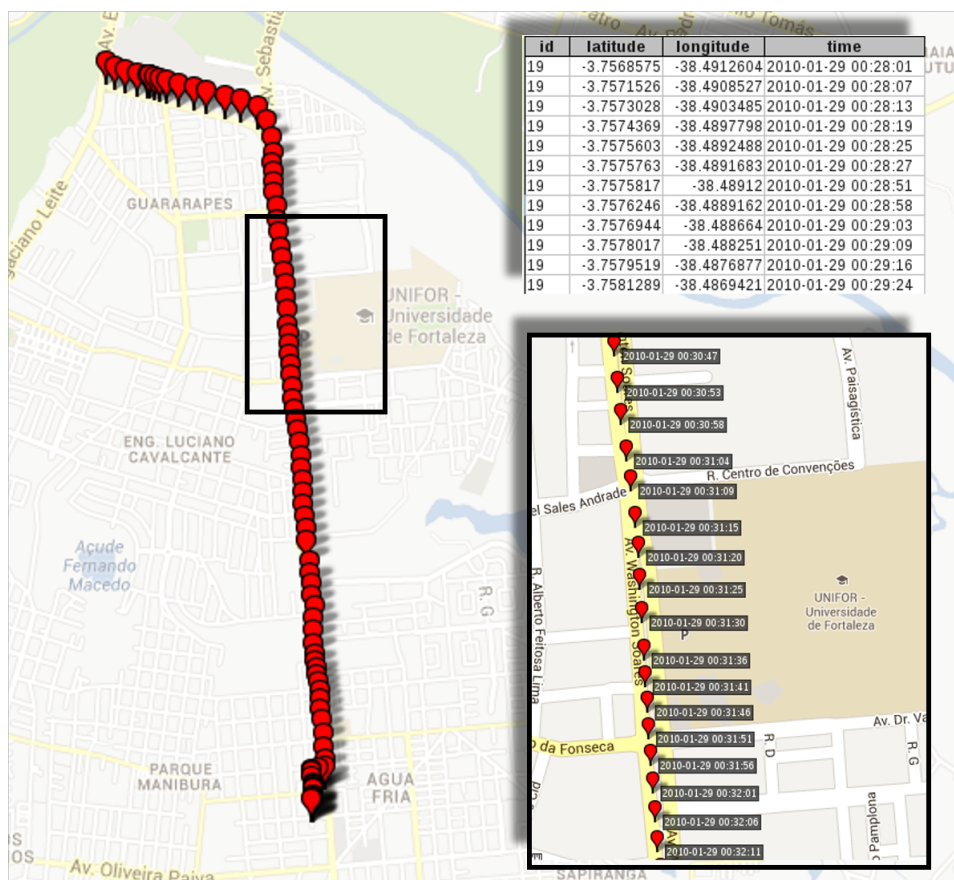


Figura 2.8: Exemplo de uma trajetória. A tabela possui os dados de latitude, longitude e instante de registro de uma trajetória com identificador (id) 19.

movimentação de objetos pode ser georreferenciada ou não. Objetos dinâmicos como pessoas, animais e veículos que se movem sobre o espaço geográfico podem ser geograficamente referenciados e representados no espaço por um sistema de coordenadas (e.g., latitude e longitude) [Dodge et al., 2008]. Similarmente, o tempo é representado por um sistema de referência, considerando um instante (i.e., ponto no tempo) ou um intervalo (i.e., amostra temporal entre instantes) no tempo. A figura 2.8 apresenta um exemplo de trajetória capturada com um dispositivo GPS e representada em um mapa pela sequência registros de posição no espaço geográfico.

Em geral, a trajetória de movimentação dos objetos é representada por uma sequência de pontos ordenada cronologicamente, denotada como $\{\langle x_1, y_1, t_1 \rangle, \langle x_2, y_2, t_2 \rangle, \dots, \langle x_N, y_N, t_N \rangle\}$, na qual x_i, y_i representam as coordenadas geográficas no instante de tempo t_i e N é o número de pontos na sequência. Com base nisso, Spaccapietra et al. [2008] definiram trajetória como o registro da evolução da posição de um objeto em movimento no espaço, durante um intervalo de tempo, a fim de alcançar um dado objetivo.

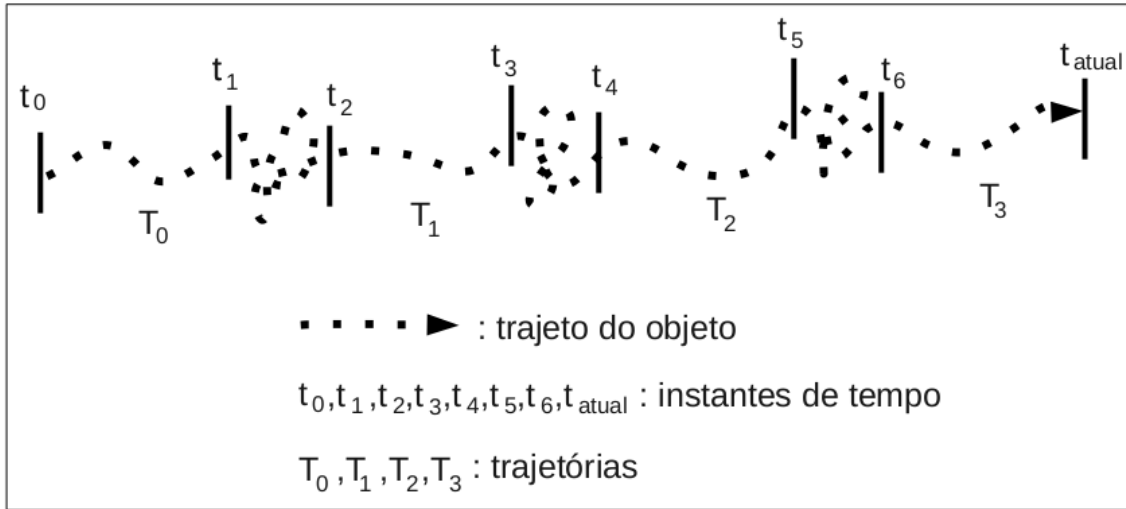


Figura 2.9: Relação entre trajeto e trajetórias de um objeto em movimento.

A figura 2.9 ilustra a movimentação diária de um objeto, que é composto por um conjunto de trajetórias. Por exemplo, considere, nesse caso, um cenário no qual uma pessoa é submetida ao rastreamento de atividades diárias usando um dispositivo GPS. A pessoa inicia a movimentação em sua casa no tempo t_0 e segue pela primeira trajetória em direção ao trabalho até o instante t_1 . No horário de almoço t_2 , ela se desloca até um restaurante seguindo a segunda trajetória. Dessa forma, a pessoa segue a rotina se deslocando entre lugares por outras trajetórias durante o restante do dia. Note que, enquanto a pessoa permanece em um determinado lugar, ela faz pequenas movimentações que podem ser desconsideradas pelas aplicações ou úteis para inferir se a pessoa está em um determinado lugar (ponto de interesse). Generalizando, as seguintes definições são apresentadas:

Definição 1 (Conjunto de Dados de Movimentação): Um conjunto de dados de movimentação de objetos é uma coleção de trajetórias $D = T_1, T_2, \dots, T_m$ em que cada T_i , tal que $1 \leq i \leq m$, é uma trajetória representada por pontos com informações espaço-temporais;

Definição 2 (Trajetória): Uma trajetória é uma sequência de triplas $T = \langle x_1, y_1, t_1 \rangle, \dots, \langle x_n, y_n, t_n \rangle$ onde $t_i (i = 1..n)$ denota um instante de tempo, tal que $\forall 1 \leq i < n, t_i < t_{i+1}$ e (x_i, y_i) são coordenadas no espaço;

Definição 3 (Sub-Trajeto): Seja $T = \langle x_1, y_1, t_1 \rangle, \dots, \langle x_n, y_n, t_n \rangle$ uma trajetória. Uma trajetória $S = \langle x'_1, y'_1, t'_1 \rangle, \dots, \langle x'_n, y'_n, t'_n \rangle$ é uma sub-trajeto de T ou está contida em $T (S \preceq T)$ se existem inteiros $1 \leq i_1 < \dots < i_m \leq n$, tal que $\forall 1 \leq j \leq m, \langle x'_j, y'_j, t'_j \rangle = \langle x_{i_j}, y_{i_j}, t_{i_j} \rangle$.

2.2.2 Processo de Descoberta de Padrões de Trajetórias

O processo de descoberta de padrões de trajetórias refere-se a um conjunto de etapas para extrair e interpretar dados espaço-temporais de um domínio e transformá-los em um conhecimento que descreve propriedades e relacionamentos de objetos nesse domínio [Jeung et al., 2011]. A seguir, cada etapa do processo é descrita.

- i. **Coleta de dados:** Os dados capturados pelos dispositivos de registro de localização são enviados para um servidor ou armazenados temporariamente nos próprios dispositivos. A representação da trajetória é definida por uma tupla contendo os seguintes atributos para cada registro: identificador, localização e tempo. A tupla (ID, Localização, Tempo) contém um ID que se refere ao identificador único para cada usuário e é utilizado para análise da mobilidade individual. O campo Localização da tupla possui as coordenadas de posicionamento e o Tempo armazena a data e hora em que o registro de posicionamento foi realizado. O método de submissão dos dados para o servidor pode ser *online* ou *offline*. No método *online*, os dados coletados pelos dispositivos são prontamente enviados para o servidor. No método *offline*, a submissão dos dados é realizada em um momento distinto do período de coleta, dessa forma a medida que os dados são capturados eles ficam armazenados no dispositivo para serem enviados em um outro instante para o servidor;
- ii. **Pré-processamento:** Nesta etapa ocorre a transformação dos dados de um estado bruto (i.e., como os dados coletados pelos dispositivos) para uma forma apropriada de análise. Tarefas como remoção de falhas de registro, tratamento de imperfeições de posicionamento e conversão para um formato adequado para executar a mineração dos dados são executadas nesta etapa;
- iii. **Mineração de Trajetórias:** Os dados processados e refinados como trajetórias são transformados em padrões por meio da utilização de algoritmos de mineração. Os seguintes exemplos exploram casos de detecção de padrões e possíveis aplicações tanto com dados de mobilidade individual, como de mobilidade de grupos de objetos:
 - Agrupamento (*Clustering*): Método de descoberta de grupo de trajetórias similares e sumarização de cada grupo (figura 2.10). Saber quais são as principais rotas (representado por um grupo de trajetórias) seguidas por pessoas ou veículos ao longo dos dias pode fornecer informações relevantes para a

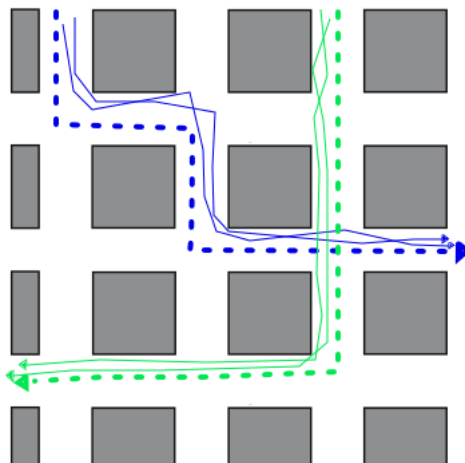


Figura 2.10: Exemplo de *clustering*.

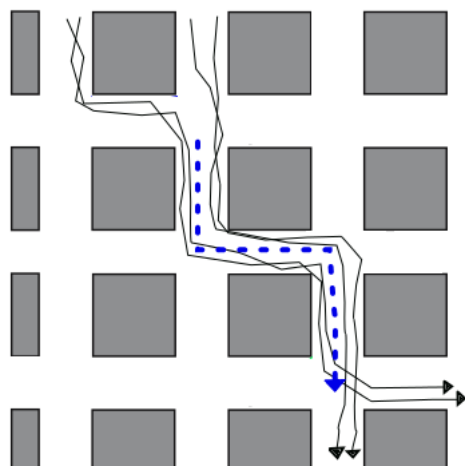


Figura 2.11: Exemplo de detecção de padrões de frequência.

análise de mobilidade. Por exemplo, essa abordagem pode identificar as vias que um determinado veículo costuma se deslocar em uma cidade.

- **Padrões de Frequência:** Identificação de caminhos frequentes de deslocamento, como os ilustrados na figura 2.11. Aplicações de planejamento urbano podem usufruir desta abordagem para detectar padrões de frequências de um caminho de um grupo de veículos. A figura 2.11 apresenta que a partir de sub-trajetórias ocorre a extração de uma trajetória representativa a partir da frequência de deslocamento de diferentes veículos em uma dada região.
- **Classificação:** Relacionada à descoberta de regra de comportamento, visa explicar o comportamento dos usuários e fazer previsão de trajetória. De acordo com a figura 2.12, pode-se determinar a probabilidade de um veículo seguir um determinado caminho em uma trajetória.

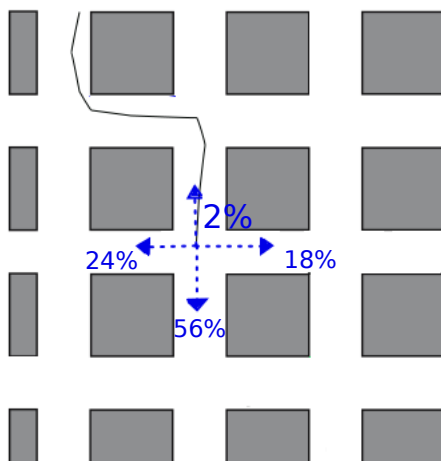


Figura 2.12: Exemplo de aplicação para predição de trajetória.

- iv. **Pós-processamento:** Nessa etapa, ocorre a análise dos padrões extraídos. Por exemplo, os padrões de movimentação de objetos podem ser exibidos em uma ferramenta de visualização para auxiliar um analista de negócio.

Uma vez que é definido um processo adequado para coletar e interpretar o conhecimento a partir de trajetórias, vários cenários de aplicação são possíveis, tais como: planejamento de tráfego e sistemas de transporte urbano, serviços baseados em trajetórias semânticas, previsão de comportamento de tráfego e análise do comportamento de mobilidade para auxiliar em sistemas de apoio à decisão.

2.2.3 Aspectos de Caracterização de Padrões de Trajetórias

O conceito de padrões de trajetórias envolve diferentes tipos que podem ser identificados pela tarefa de mineração de dados [Han et al., 2010]. Cada padrão é caracterizado por uma variedade de aspectos que auxiliam no entendimento das movimentações de objetos no espaço geográfico. A seguir, é apresentada uma categorização, definida por Dodge et al. [2008], dos principais aspectos relacionados à caracterização de padrões de trajetórias.

- **Aspectos de Similaridades Espaciais e Espaço-Temporais:** O aspecto espacial e espaço-temporal define a semântica relativa a um grupo de trajetórias similares. Considere a figura 2.13 que possui a representação da movimentação de três objetos em relação ao espaço e ao tempo. No tempo $t = 1$, os dois objetos O_1 e O_2 iniciam juntos a movimentação, aproximadamente na mesma localização. No tempo $t = 2$, um terceiro objeto, nomeado O_3 , aparece e inicia o movimento próximo à localização de O_1 e O_2 quando estavam no tempo $t = 1$. Enquanto

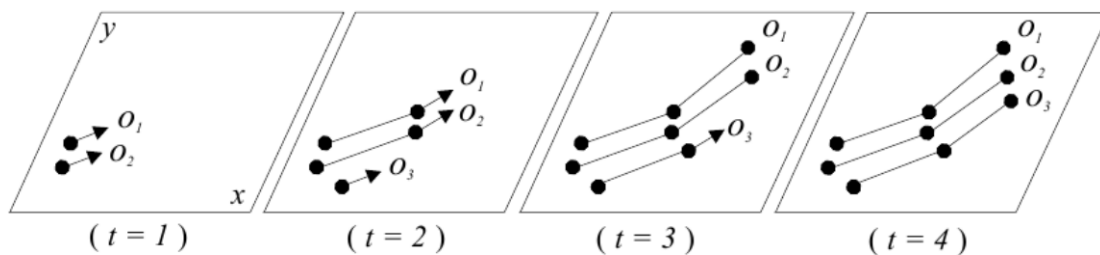


Figura 2.13: Representação da movimentação de três objetos em relação ao espaço e ao tempo durante $t = [1, 4]$ [Jeung et al., 2011].

isso, os objetos O_1 e O_2 continuam se movimentando na região central. No tempo $t = 3$, o objeto O_3 mantém um deslocamento por um caminho semelhante às trajetórias dos outros dois objetos, ao mesmo tempo em que os objetos O_1 e O_2 chegam ao destino e finalizam a movimentação. No entanto, o objeto O_3 continua o deslocamento e para no tempo $t = 4$ em uma posição próxima a localização onde os objetos O_1 e O_2 pararam.

Observando as trajetórias no instante $t = 4$, nota-se que a movimentação dos três objetos no espaço é semelhante. Dessa forma, quando se aplica uma tarefa de agrupamento de trajetórias similares baseada no espaço, essas três trajetórias formam um grupo (*cluster*), embora o objeto O_3 nunca tenha se deslocado junto com O_1 e O_2 . Nesse caso, tem-se um padrão de trajetória espacial, como mostrado na figura 2.14 (a). Apesar de o conceito de trajetória ter relacionamento com grandezas temporais, uma variedade de aplicações não necessita considerar o tempo. Por outro lado, caso o tempo seja relevante para a aplicação, padrões de trajetórias espaço-temporais são considerados em contraste aos espaciais. Por exemplo, as trajetórias do objeto O_3 não pertencem ao *cluster* formado pelas trajetórias dos objetos O_1 e O_2 (figura 2.14 (b)), visto que O_3 nunca se deslocou em conjunto com os outros objetos. Por exemplo, essa abordagem de formação de *cluster* considerando o tempo é aplicável para determinar pontos de congestionamento de veículos.

Um conceito importante relacionado ao *cluster* é o de trajetória representativa. Dado um *cluster*, pode-se obter uma trajetória representativa, que descreve a movimentação das trajetórias pertencentes e, assim, é adotada como um padrão que representa todas as trajetórias do *cluster*. A figura 2.15 ilustra o processo para obtenção de trajetórias representativas para dois *clusters*. Essa trajetória representativa pode ser identificada a partir de *clusters* de trajetórias de vários objetos ou de trajetórias de um mesmo objeto. Por exemplo, uma trajetória

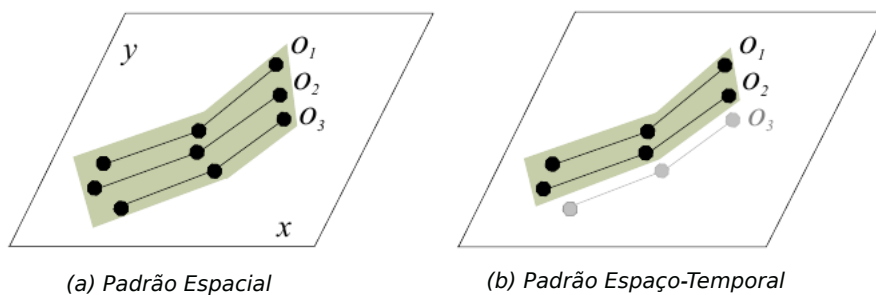


Figura 2.14: Padrões de trajetórias espaciais e espaço-temporais [Jeung et al., 2011].

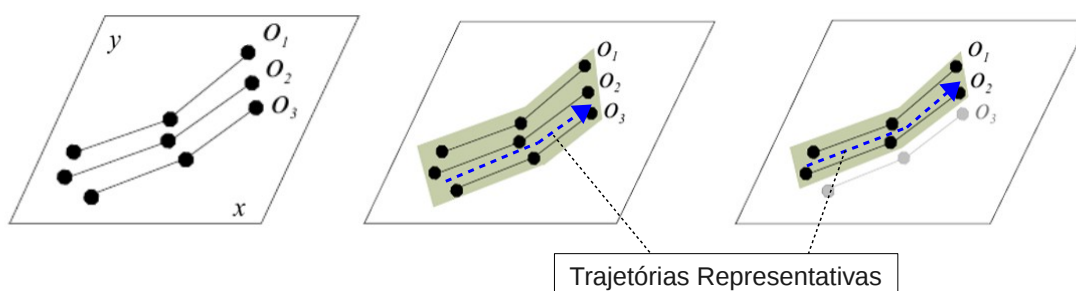


Figura 2.15: Identificação de trajetórias representativas (Adaptada de Jeung et al. [2011])

representativa de um indivíduo que se desloca entre a casa e o trabalho ao longo dos dias.

- **Aspectos de Granularidade:** A granularidade de um padrão de trajetória é caracterizado por dois fatores: fator 1, intervalo de tempo das trajetórias que pertencem ao padrão; fator 2, número de objetos que estão relacionados com o padrão.

Quanto ao fator 1, o padrão pode ser global ou parcial. No padrão global, uma trajetória é tratada como uma entidade indivisível, ou seja, considera-se a trajetória completa como unidade básica para a descoberta de padrões. No padrão parcial, o processo de descoberta de padrões é baseado em sub-trajetórias. A ideia é que mesmo quando trajetórias completas não formem um padrão, partições dessas trajetórias possuem similaridades para obtenção de um padrão.

Quanto ao fator 2, o padrão pode ser individual ou de grupo. No padrão individual, somente trajetórias individuais são consideradas. No padrão de grupo, tem-se a representação de um padrão de trajetórias baseada em um conjunto de trajetórias de diferentes objetos.

- **Aspectos de Restrições Espaciais e Temporais:** Padrões de trajetórias sofrem restrições tanto espaciais quanto temporais. Por exemplo, veículos se movimentam com restrição espacial, pois se deslocam por estradas. Para análise, geralmente, modela-se um mapa de estradas de uma cidade em um formato de grafo no qual os cruzamentos são os nós e os segmentos de rodovias são as arestas. Dessa forma, extrai-se informações relevantes para cada aresta, como sentido de deslocamento e intensidade de tráfego. Quanto ao tempo, padrões de trajetórias consideram que os objetos seguem uma regularidade temporal. Nesse sentido, minerar padrões de trajetória refere-se também à descoberta de comportamentos periódicos dos objetos, ou seja, assume-se que os objetos seguem rotas similares em intervalos de tempos regulares (e.g., diariamente, semanalmente) [Li et al., 2010b]. Por exemplo, muitas pessoas se dirigem de casa para o trabalho em rotas similares ao longo dos dias. Outro exemplo seria o de veículos de transporte público que possuem a mesma rota e possuem uma escala de horário predefinida. O estudo de padrões periódicos revela informações sumarizadas sobre um comportamento, servindo de base para prever futuras movimentações e, em casos de não existência de uma conduta periódica, pode-se inferir que algo anormal ocorreu (i.e., mudança de rotina).

2.3 Roteamento em VANET baseado em Trajetória de Veículos

Em uma rede *ad hoc*, um protocolo de roteamento rege o processo de comunicação entre os nós da rede. O estabelecimento da rota de comunicação, bem como o algoritmo para encaminhamento de mensagens e as ações de manutenção e recuperação de rotas são atividades tratadas pelos protocolos de roteamento. O desafio no projeto de protocolos de roteamento é fornecer soluções para essas atividades visando um baixo atraso (*delay*) de comunicação, baixa sobrecarga (*overhead*) na rede e alta taxa de entrega de dados.

Assim como em uma MANET convencional, um dos maiores desafios em uma VANET é a concepção de protocolos de roteamento, fazendo com que esses possibilitem a comunicação de dados entre os nós da rede. Na literatura, existem diversos algoritmos de roteamento propostos para diferentes cenários de MANET [Liu & Kaiser, 2003] [Maihofer, 2004] [Junhai et al., 2009]. No entanto, as redes veiculares apresentam um conjunto de características que impossibilitam a aplicação direta desses algoritmos em um contexto de comunicação veicular, por exemplo: topologia altamente dinâmica, desconexões frequentes na rede, distribuição irregular dos nós. Diferentes pesquisadores

mostraram que o desempenho dos protocolos de roteamento, típicos de MANET, é comprometido em cenários de VANET, devido a essas características [Santos et al., 2005] [Mateus et al., 2012].

Nesse sentido, novas estratégias de roteamento em redes veiculares têm sido investigadas. Com destaque para estratégias baseadas em posição geográfica, que têm sido utilizadas em um significativo número de protocolos de roteamento [Bernsen & Manivannan, 2009] [Bilal et al., 2013]. Esses protocolos usam informação de coordenadas geográficas para estabelecer rotas de comunicação e, em combinação com uma abordagem gulosa de encaminhamento, realizam um processo de roteamento eficiente na rede [Cadger et al., 2013]. Dessa forma, o roteamento geográfico é composto de duas etapas: i) a obtenção da localização geográfica e ii) o encaminhamento de mensagens, considerando a localização geográfica. Diferente de protocolos baseados em topologia, no roteamento geográfico não há necessidade de criar e manter uma rota entre nós de origem e destino. Portanto, essa estratégia usa informação de posição como potencial solução de roteamento em cenários com alta mobilidade dos nós [Fübler et al., 2002].

A incorporação de sistema de navegação do tipo GPS nos veículos possibilita um avanço no roteamento geográfico, pois a informação de trajetória de veículos passa a ser utilizada no processo de encaminhamento de mensagens. Por exemplo, em um cenário de *geocast* que segue a estratégia de encaminhamento baseado em trajetória, um veículo com uma trajetória em direção ao destino da mensagem pode ser selecionado como um nó intermediário para realizar a entrega. Em um cenário *unicast*, a relação de encontro entre as trajetórias dos veículos também oferece oportunidades de encaminhamento de mensagens.

Na literatura, existem diversos artigos que apresentam taxonomias de esquemas de roteamento para diferentes modelos de disseminação e cenários [Li & Wang, 2007] [Bernsen & Manivannan, 2009] [Lee et al., 2009] [Panichpapiboon & Pattara-Atikom, 2012]. No entanto, por se tratar de uma abordagem recente, verificou-se na literatura a carência de estudos que apresentem uma visão geral do roteamento geográfico com base em trajetória de veículos. Dessa forma, foram desempenhadas duas atividades, uma revisão de literatura nesse campo e uma discussão sobre os protocolos identificados na literatura, ambas apresentadas no restante desta seção.

2.3.1 GeOpps

O GeOpps (*Geographical Opportunistic Routing for Vehicular Networks*) [Leontiadis & Mascolo, 2007] é um protocolo de roteamento que explora tanto o caráter oportunista como as características peculiares de redes veiculares em termos de padrão de

mobilidade, encontro entre veículos e informação geográfica, a partir de sistemas de navegação presentes nos veículos. Esse protocolo opera em circunstâncias nas quais a VANET tem um comportamento de uma rede tolerante a atraso (DTN) [Pereira et al., 2012]. Semelhante aos protocolos típicos de DTN [Jain et al., 2004], o GeOpps adota a técnica de armazenamento e encaminhamento de mensagens entre saltos até atingir o destino. Esse encaminhamento é dado pelo contato oportunístico entre os veículos. Dessa forma, o padrão de mobilidade de cada veículo e a seleção do próximo veículo que deve ser encaminhada a mensagem são responsáveis pela entrega da mensagem ao destino.

A suposição inicial do GeOpps é a existência de um sistema de navegação que sugere uma trajetória entre a posição atual do veículo e o destino. Assim, o GeOpps considera as trajetórias sugeridas pelos sistemas de navegação dos veículos para determinar que o veículo se desloca para um destino próximo ou igual ao destino da mensagem. Dessa forma, pode-se obter o ponto mais próximo (*Nearest Point - NP*) de uma posição na trajetória de um veículo em direção ao destino (*D*) da mensagem. Além disso, quando um veículo possui uma mensagem e encontra um ou mais veículos, ocorre um processo de decisão para determinar se ele mantém a mensagem ou a encaminha para um destes veículos. Para auxiliar nesse processo de decisão, o sistema de navegação é usado para estimar o tempo que um veículo necessita para chegar até o NP. Esse tempo é denominado ETA (*Estimate the Time of Arrival*), sendo determinado a partir do caminho de NP até D. A soma desses valores (ver equação 2.1), chamado METD (*Minimum Estimated Time of Delivery*), é a estimativa de tempo mínima para entregar a mensagem ao destino D e é usada como função utilidade para o processo de decisão de roteamento.

$$METD = (ETA \text{ até } NP) + (ETA \text{ de } NP \text{ até } D) \quad (2.1)$$

Para cada encontro, o veículo que está portando a mensagem verifica se algum dos veículos vizinhos possui um METD menor do que o dele. Em caso positivo, ele encaminha a mensagem. Por exemplo, na figura 2.16, o ponto P_1 é um veículo que deseja encaminhar uma mensagem até o ponto D . Para isso, considerando um contato dele com os veículos a e b , ele verifica qual dos dois possui um menor METD. Esses veículos calculam os seus NP em relação ao ponto D (NP_a e NP_b). Analisando NP_a e NP_b , o veículo b torna-se o portador da mensagem. À medida que o veículo b se desloca até o seu destino, ele pode se encontrar com outros veículos detentores de um menor METD. No ponto P_2 , encontra o veículo c , que se desloca até uma região mais próxima do ponto D . Então, b encaminha a mensagem para c . Dessa forma, como



Figura 2.16: Exemplo do processo de roteamento de uma mensagem do ponto P_1 até D utilizando o GeOpps [Leontiadis & Mascolo, 2007].

mostrado na figura, o veículo b não carregou a mensagem até o ponto NP_b .

Em resumo, o GeOpps é um protocolo de roteamento tolerante a atraso que explora a disponibilidade de trajetória, a partir do sistema de navegação dos veículos, a fim de encaminhar uma mensagem oportunisticamente até uma determinada região geográfica.

Os passos básicos do protocolo são os seguintes:

- **PASSO 1:** Os veículos periodicamente disseminam por *broadcast* o destino das mensagens armazenadas por eles;
- **PASSO 2:** Os vizinhos de um salto calculam o *METD* para entregar cada mensagem ao destino e enviam esse valor para o veículo portador da mensagem;
- **PASSO 3:** O portador da mensagem mantém a mensagem armazenada (se todos os *METD* recebidos dos vizinhos forem maior que o dele) ou a encaminha para o veículo vizinho com menor *METD*;
- **PASSO 4:** O processo é repetido até a mensagem chegar ao destino ou o tempo de vida dela expirar.

2.3.2 GeoSpray

O GeoSpray [Soares et al., 2011] é um protocolo de roteamento baseado no encaminhamento de múltiplas cópias de mensagens para o domínio infraestruturado de redes veiculares. Ele explora a informação geográfica fornecida pelo sistema de navegação

dos veículos em combinação com a técnica armazena-e-encaminha. Este protocolo é direcionado à cenários de redes veiculares esparsos, nos quais ocorre uma comunicação oportunística baseada nos contatos esporádicos e intermitentes entre veículos.

Conceitualmente, o GeoSpray é inspirado nos protocolos GeOpps (descrito na subseção 2.3.1) e Spray-and-Wait (SAW) [Spyropoulos et al., 2005a]. O GeoSpray usa uma abordagem semelhante ao GeOpps pelo fato de relacionar a trajetória de deslocamento dos veículos com a posição de destino da mensagem. No entanto, ao contrário do GeOpps, que mantém no máximo uma cópia de cada mensagem na rede, o GeoSpray permite que múltiplas cópias de uma mensagem sejam replicadas na rede, assim como o SAW. Entretanto, diferente do SAW, o mecanismo de distribuição de cópias é direcionado, ou seja, somente os veículos que se deslocam em direção as proximidades da posição de destino da mensagem recebem uma cópia.

A diferença básica entre o GeoSpray e o GeOpps é uma alteração no PASSO 3 do GeOpps. Nesse passo, no GeoSpray, o portador da mensagem verifica se algum nó vizinho possui um *METD* menor que o dele. Em caso positivo, ao invés da mensagem ser enviada para o nó vizinho, cria-se uma cópia para ser enviada, dando assim um caráter de múltiplas cópias. Para que não haja uma criação de cópias de mensagens de maneira descontrolada, os criadores do protocolo GeoSpray estipularam que o número máximo de cópias de mensagens deve ser definido inicialmente e, a cada encontro, aplica-se o método de *spray* binário do protocolo SAW.

2.3.3 GeoDTN+Nav

O GeoDTN+Nav [Cheng et al., 2010] é um protocolo de roteamento híbrido para VANET. Ele é híbrido no sentido de combinar estratégias de roteamento geográfico com informações de trajetórias de veículos em cenários de rede conectada ou particionada. Para isso, ele tem três modos de encaminhamento, denominados *greedy*, *perimeter* e DTN.

Primeiramente, as mensagens são encaminhadas pelo modo *greedy*. Quando ocorre uma falha (atinge um máximo local¹), o protocolo é alterado para o modo *perimeter*. Se o modo *perimeter* também falhar, muda-se para o modo DTN, dependendo assim da mobilidade dos veículos para entregar a mensagem. Durante as transições entre os modos, duas questões são relevantes: i) Quando o protocolo muda para o modo DTN? ii) Quando o protocolo volta para o modo *greedy*? Para a primeira, utiliza-se uma função de utilidade e um valor limite relacionado com dois fatores (a detecção

¹situação em que o nó não encontra um vizinho mais próximo do destino do que ele mesmo. Isso ocorre devido à distribuição não uniforme da topologia.

de partição e a qualidade de entrega da mensagem por um vizinho). Para a segunda, volta-se para o modo *greedy*, quando um nó intermediário é mais apto a entregar a mensagem do que o nó que iniciou o modo DTN.

A seguir, são descritos os passos básicos do protocolo. No entanto, é necessário assumir que cada veículo é equipado com um dispositivo VNI (*Virtual Navigation Interface*), que fornece a informação da trajetória do veículo.

- **PASSO 1:** Todos os veículos enviam periodicamente por *broadcast* as informações de trajetória aos vizinhos;
- **PASSO 2:** O protocolo opera no modo *greedy* para entregar a mensagem até encontrar um máximo local ou alcançar o destino;
- **PASSO 3:** Troca para o modo *perimeter* e armazena a posição do veículo e o destino na mensagem;
- **PASSO 4:** A cada salto no modo *perimeter*, faz as seguintes tarefas:
 - i. Usa uma função $P(h)$ para calcular a probabilidade de desconexão na rede;
 - ii. Usa $Q(N_i)$ para calcular a qualidade de entrega de cada vizinho, assim como a do portador da mensagem;
 - iii. Usa $Dir(N_i)$ para calcular a qualidade de direção de cada vizinho;
 - iv. Calcula um *global score* para cada nó vizinho. No cálculo, os valores encontrados em $P(h)$, $Q(N_i)$ e $Dir(N_i)$ são considerados;
 - v. Se algum nó vizinho possuir um *global score* maior que o limiar S_{thres} , ele encaminha a mensagem para o respectivo nó vizinho e muda para o modo DTN.
- **PASSO 5:** Cada mensagem possui um tempo de vida na rede, que representa o número de saltos. Se o número de saltos da mensagem for o valor inicialmente estipulado e nenhum nó vizinho possuir um *global score* maior que S_{thres} , a mensagem é descartada.

2.3.4 TBD

O TBD (*Trajectory-Based Data Forwarding for Light-Traffic Vehicular Networks*) [Jeong et al., 2011] realiza o roteamento combinando a informação de trajetória de veículos com estatísticas de tráfego para o domínio infraestruturado de redes veiculares.

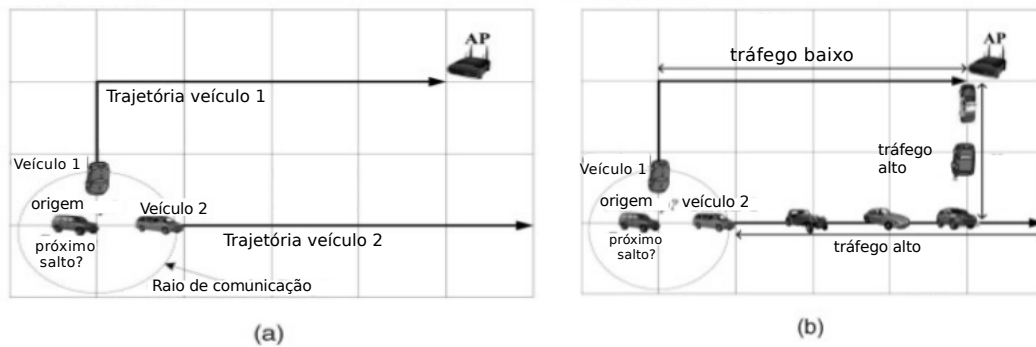


Figura 2.17: Ilustração de cenário para o TBD (Adaptada de [Jeong et al., 2011]).

Como ilustrado na figura 2.17, dada uma região geográfica com um ponto de acesso (AP - *Access Point*) instalado, o problema é minimizar o atraso de entrega de mensagens a partir de um veículo origem até o AP. A figura 2.17 (a) representa um cenário no qual existem três veículos em movimento em uma dada região geográfica. O veículo origem deseja enviar uma mensagem para um dado AP. Além disso, sabe-se que os veículos 1 e 2 estão no raio de comunicação do veículo origem. Se as trajetórias dos veículos são conhecidas, facilmente o veículo origem decidirá para qual veículo encaminhará a mensagem. Nesse caso, o veículo 1 é escolhido, pois possui uma trajetória em direção ao AP. No entanto, para realizar esse procedimento surge o primeiro problema, que é decidir sobre o encaminhamento de mensagens sem que haja o compartilhamento de trajetórias.

Na figura 2.17(b), o veículo 1 possui uma trajetória de movimentação em direção a uma área com baixa densidade de veículos, enquanto o veículo 2 segue uma trajetória em direção à uma área com uma densidade alta de veículos. Destaca-se que em um trajeto com uma alta densidade de veículos se pode realizar um roteamento salto-a-salto. Dessa forma, o veículo origem encaminhará a mensagem para o veículo 2 que, por sua vez, iniciará o roteamento da mensagem em direção ao AP por meio dos outros veículos. O segundo problema é como combinar estatísticas de tráfego (e.g., densidade) com informação de trajetória de veículos para auxiliar no processo de roteamento de mensagens em redes veiculares.

No TBD, cada veículo, considerando sua trajetória, calcula um atraso esperado de entrega (*Expected Delivery Delay - EDD*) da mensagem para o AP. A cada encontro entre veículos verifica-se os atrasos (EDDs) entre eles para determinar o próximo salto. Por exemplo, a figura 2.18 ilustra um cenário simples com apenas três veículos em uma dada região. O veículo origem possui dois veículos vizinhos, nomeados como veículo 1 e veículo 2, sendo estes possíveis nós encaminhadores. Percebe-se que o veículo 1

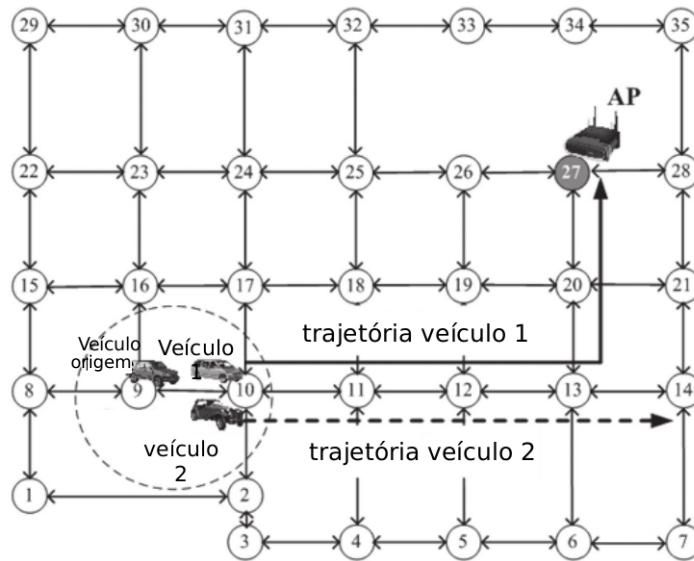


Figura 2.18: Ilustração de funcionamento do TBD.

e o veículo 2 têm inicialmente o mesmo sentido de movimentação, ou seja, ambos se deslocam em direção ao cruzamento 11. Analisando a trajetória de ambos, nota-se que o veículo 2 tem uma trajetória distante do AP, conseqüentemente, fora de alcance da comunicação com o AP. No entanto, o veículo 1 segue uma trajetória que passa em uma posição de comunicação com o AP. Assim, o veículo 1 é selecionado como próximo salto para a entrega da mensagem e utilizando a técnica armazena-e-encaminha ele possui chances de entregar a mensagem ao destino.

Nesse processo de decisão de encaminhamento, cada veículo precisa computar o seu EDD, considerando sua própria trajetória para auxiliar no roteamento. Essa abordagem permite que os veículos não necessitem compartilhar detalhes de suas trajetórias, mas apenas um valor que não compromete a privacidade dos motoristas. O veículo que carrega a mensagem, denominado *portador*, tem uma trajetória T representada por uma sequência de cruzamentos que ele visita, logo: $T = 1 \rightarrow 2 \rightarrow \dots \rightarrow M$. Seja C_{ij} o tempo total que o veículo carrega a mensagem entre os cruzamentos i e j durante a trajetória, sendo que $1 \leq i \leq j \leq M$. Assim, C_{ij} é definido como a parte do atraso esperado de entrega na qual o portador se desloca pelos segmentos da trajetória, entre os cruzamentos i e j . Formalmente, $C_{ij} = \sum_{k=1}^{j-1} l_{k,k+1}/v$, onde $l_{k,k+1}$ é o comprimento do segmento pelo qual o portador se desloca entre os cruzamentos k e $k+1$, e v é a velocidade de deslocamento. O EDD para a trajetória T é calculada como segue:

$$EDD = \sum_{j=1}^M (P[\text{msg carregada do cruzamento 1 até } j] X (C_{1j} + E[\text{atraso até } j])) \quad (2.2)$$

Portanto, o TBD permite que cada veículo calcule seu próprio EDD considerando sua trajetória. Dessa forma, o portador da mensagem seleciona o próximo salto (i.e., próximo portador) entre os veículos vizinhos. No entanto, o TBD foi projetado para cenários nos quais o destino da mensagem é um nó estático, por exemplo, um AP. Assim, nas situações que o destino da mensagem é um nó móvel (i.e., veículo) é necessário considerar a mobilidade desse destino. Nesse sentido, a próxima seção apresenta um protocolo que introduz uma abordagem de múltiplos saltos para comunicação a partir de um nó de infraestrutura com destino um veículo.

2.3.5 TSF

O TSF (*Trajectory-based Statistical Forwarding*) [Jeong et al., 2012] é um esquema de roteamento para encaminhar mensagens de um nó da infraestrutura presente na rodovia (e.g., ponto de acesso) para um nó destino (veículo) em movimento. Dessa forma, o TSF encaminha mensagens utilizando múltiplos saltos para um selecionado ponto alvo, no qual o veículo destino tem ao longo da sua trajetória uma interseção de conectividade. Além de ser um ponto de encontro com a trajetória do veículo destino, esse ponto alvo deve ser definido de forma a minimizar o atraso de entrega da mensagem.

Por exemplo, a figura 2.19 mostra um cenário em que há encaminhamento de mensagens de um veículo para infraestrutura e vice-versa. No primeiro caso, no sentido de encaminhamento veículo para infraestrutura (i.e., veículo origem para ponto de acesso 1 - AP1), pode-se utilizar o protocolo TBD (descrito na subseção 2.3.4). No segundo caso, o encaminhamento da mensagem é do AP1 para o veículo destino. Sabendo que a trajetória do veículo destino consiste de um trajeto que percorre sete interseções (n_2, n_3, \dots, n_{20}) e que esta trajetória é conhecida por um Centro de Controle de Tráfego (CCT), o objetivo do TSF é entregar a mensagem a partir do AP1 para o veículo destino no menor intervalo de tempo. Na figura 2.19, a mensagem chega ao ponto alvo (n_{10}) anteriormente ao veículo destino. Assim, esse ponto alvo é determinado como um ponto altamente esperado para satisfazer o encontro entre veículo destino e a mensagem, no mais curto intervalo de tempo.

O protocolo TSF depende de dois importantes fatores: 1) o caminho de encami-

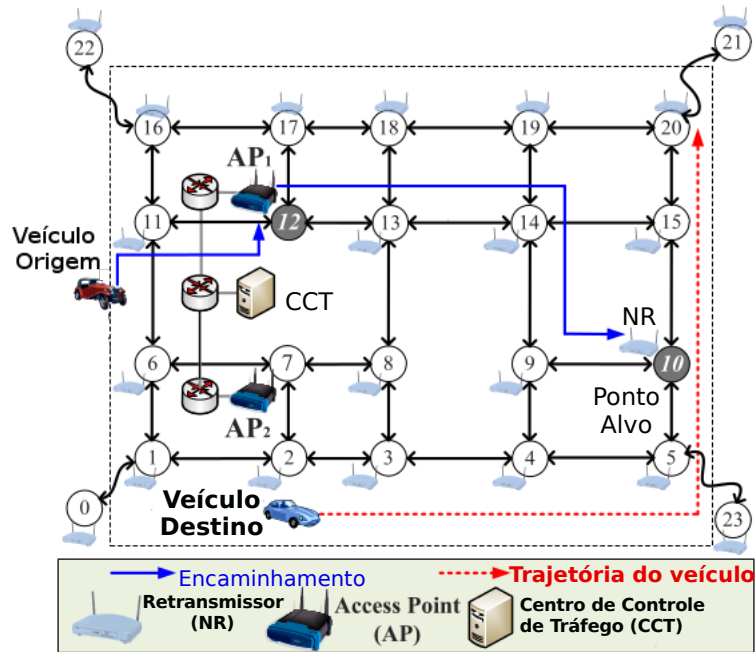


Figura 2.19: Exemplo de cenário para o roteamento de uma mensagem do ponto AP_1 até o veículo destino, utilizando o TSF (Adaptada de Jeong et al. [2012]).

nhamento da mensagem e 2) a trajetória do veículo. O caminho de encaminhamento da mensagem é a lista de cruzamentos no percurso entre o AP e o ponto alvo. A trajetória do veículo é o conjunto de cruzamentos que o veículo destino da mensagem percorre durante o movimento. Uma suposição relevante do TSF é que o veículo destino reporta periodicamente sua posição atual e a trajetória que seguirá para o CCT. Dessa forma, com a trajetória do veículo destino registrada no CCT, o TSF encaminha as mensagens a partir de um AP para o veículo destino em dois passos:

- **PASSO 1:** Encaminhamento da mensagem a partir do AP para o ponto alvo. Quando o CCT tem uma mensagem para ser entregue ao veículo destino, ele computa um caminho de encaminhamento da mensagem a partir de um AP para um ponto alvo. Esse ponto alvo é um NR que tentará encaminhar a mensagem para o veículo destino quando ambos tiverem no mesmo raio de comunicação. Por exemplo, na figura 2.19, o caminho de encaminhamento é $(n_{12}, n_{13}, n_{14}, n_9, n_{10})$. Inicialmente, o AP em n_{12} envia a mensagem para algum veículo em movimento em direção à n_{13} . Esse processo de encaminhar mensagens entre os APs, utilizando veículos entre os cruzamentos selecionados previamente, é mantido até a mensagem chegar no AP n_{10} .
- **PASSO 2:** Encaminhamento da mensagem a partir do ponto alvo para o veí-

culo destino. Como mostrado na figura 2.19, quando a mensagem chega ao nó retransmissor (n_{10}), que corresponde ao ponto alvo, o nó retransmissor mantém armazenada a mensagem até o momento de encaminhar para o veículo destino. Caso exista algum outro veículo que possua uma trajetória de encontro (i.e., se desloque por n_{10} , n_5 , n_4 , n_3) com o veículo destino, pode-se utilizá-lo como portador da mensagem. No entanto, se o portador da mensagem seguir uma nova rota no retransmissor n_5 , pode-se encaminhar e armazenar a mensagem em n_5 e esperar pelo veículo destino ou enviá-la a um novo portador que irá se encontrar com o veículo destino. Esse processo é repetido até a mensagem ser entregue ao veículo destino ou o tempo de vida dela na rede expire.

2.3.6 STDFS

O STDFS (*Shared Vehicle Trajectories for Data Forwarding Scheme*) [Xu et al., 2011] é um esquema de roteamento de múltiplos saltos em redes veiculares por meio do compartilhamento de trajetórias de veículos. Para tanto, o STDFS assume que cada veículo pode obter periodicamente as trajetórias de outros veículos a partir de pontos de acesso distribuídos geograficamente. Com as trajetórias compartilhadas, o STDFS estima encontros entre veículos que se deslocam em direções opostas e representa essa informação em um grafo, denominado grafo de encontros.

O processo de encaminhamento também considera dois fatores: o EDR (*Expected Delivery Ratio*) de um veículo, de forma que EDR_e é a taxa esperada para a entrega de uma mensagem a partir do veículo e ao destino da mensagem; o EDD (*Expected Delivery Delay*) de um veículo, de forma que EDD_e é o atraso previsto de entrega das mensagens enviadas pelo veículo e até o destino.

Dessa forma, é definido o seguinte processo de encaminhamento:

- **PASSO 1:** O veículo que deseja enviar a mensagem computa o grafo de encontros com base nas trajetórias compartilhadas obtidas em um ponto de acesso;
- **PASSO 2:** De acordo com a aplicação ou por fornecimento do usuário são determinados o TTL da mensagem e o limite mínimo R para o EDR;
- **PASSO 3:** Com as informações dos PASSOS 1 e 2 computa-se a sequência de encaminhamento. Assim, o encaminhamento é conduzido pela sequência definida;
- **PASSO 4:** Em caso de um veículo portador da mensagem encontrar outros veículos que não estão na sequência de encaminhamento da mensagem, mas podem

ser úteis no encaminhamento (e.g., devido a informações desatualizadas de trajetórias obtidas dos pontos de acesso), o portador notifica aos vizinhos o destino da mensagem. Com isso, cada veículo da vizinhança calcula seu próprio EDR e EDD ;

- **PASSO 5:** Obtendo os valores de EDR e EDD dos vizinhos, o portador decide encaminhar a mensagem para o vizinho com menor EDD e com $EDR \geq R$;
- **PASSO 6:** O processo é repetido a partir do PASSO 4 até a mensagem chegar ao destino ou o tempo dela expirar.

Como o STDFS requer o compartilhamento de trajetórias entre veículos e pontos de acesso, tem-se uma sobrecarga na rede para encaminhamento de mensagens, visto que os veículos precisam periodicamente atualizar seu grafo de encontros. Além disso, o conhecimento de trajetórias de outros veículos impacta em requisitos de segurança e privacidade.

2.3.7 Discussão

Esta subseção apresenta um fluxograma de relacionamento e uma comparação qualitativa entre os protocolos descritos anteriormente. A figura 2.20 mostra um fluxograma de relacionamento que destaca a influência de outros protocolos na concepção e projeto dos protocolos baseados em trajetórias de veículos descritos neste capítulo. Para facilitar a visualização, foi adotada uma organização de acordo com o ano de publicação de cada protocolo e as setas representam os protocolos precursores que serviram de base. Vale salientar que a reflexão sobre esse relacionamento entre os protocolos foi determinada pelas citações mencionadas pelos autores das publicações e por identificação de semelhança no algoritmo de funcionamento de cada protocolo.

O GeOpps potencialmente sofreu influência do GSR [Lochert et al., 2003] e do MoVe [LeBrun et al., 2005]. O GSR é um protocolo baseado em posição geográfica, que faz uso de um sistema de navegação de veículos para auxiliar na etapa de *greedy* e o MoVe aproveita o conhecimento das velocidades relativas de um veículo e de seus vizinhos para estimar a distância mais próxima que eles podem chegar ao destino. Dessa forma, pode-se relacionar com o fato do GeOpps necessitar de um sistema de navegação para determinar o NP baseado na trajetória dos veículos, fazendo um *greedy* direcionado. Além disso, ele utiliza a velocidade para estimar o tempo de chegada do veículo a esse ponto NP mais próximo do destino D.

O GeoSpray utiliza a mesma abordagem do GeOpps, com exceção de adotar uma técnica de múltiplas cópias de mensagens. Para que a criação das cópias das mensagens

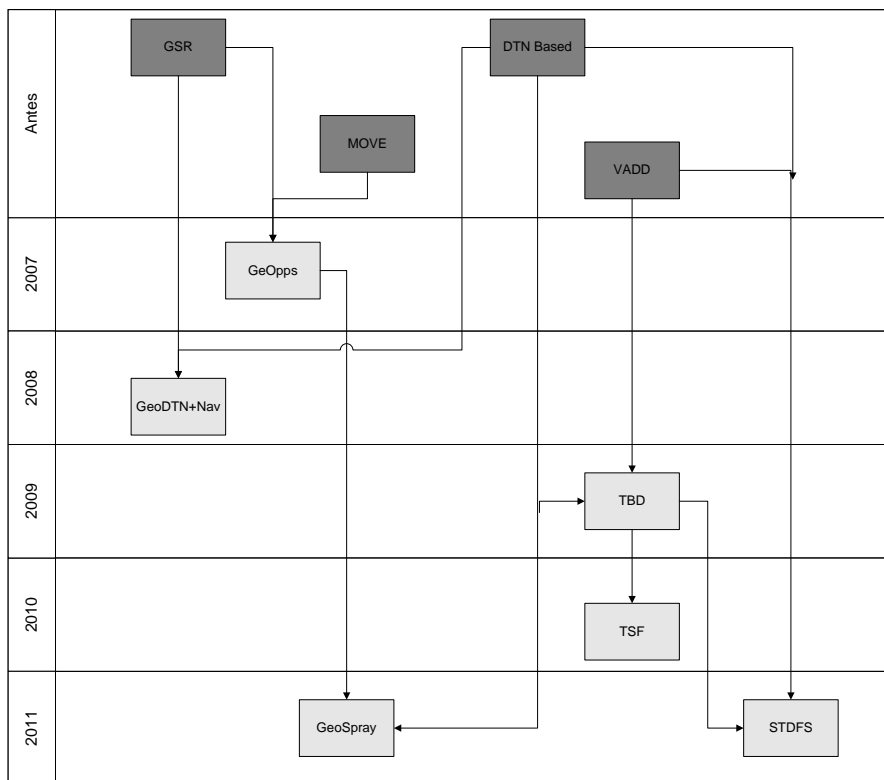


Figura 2.20: Fluxograma de relacionamento entre os protocolos de roteamento geográfico com base em trajetórias de veículos.

seja controlada, o GeoSpray limita o número de cópias semelhante ao protocolo Spray-and-Wait. Na figura 2.20, o GeoSpray herda de um protocolo DTN, nesse caso, o Spray-and-Wait. Em termos de desempenho, quando comparado ao GeOpps, o GeoSpray possui uma taxa de entrega de mensagens e um *overhead* relativamente maiores. Isso se justifica devido ao GeoSpray ser um esquema de múltiplas cópias.

O GeoDTN+Nav engloba o roteamento geográfico, o mecanismo DTN e o sistema de navegação veicular. Ele possui três modos, sendo que dois desses modos são passos do protocolo GSR, i.e., *greedy* e *perimeter*. O terceiro modo, chamado DTN, utiliza o esquema armazena-e-encaminha. Essas abordagens seguidas pelo GeoDTN+Nav agrupam vantagens do roteamento geográfico e do roteamento aplicado em DTN. Além disso, ele utiliza um sistema de navegação para obter a trajetória, auxiliando no processo de decisão de encaminhamento. Em redes esparsas, o GeoDTN+Nav aumenta a taxa de entrega visto que explora essa mobilidade dos veículos em combinação com o sistema de navegação para entregar mensagens entre partições da rede. Em redes densas, o GeoDTN+Nav simplesmente utiliza um roteamento geográfico e faz a entrega de mensagens por meio de múltiplos saltos em direção ao destino da mensagem.

Os protocolos TBD e TSF sofrem influência do VADD [Zhao & Cao, 2006]. O

Protocolo	Trajectoria do Veículo	Tipo de Domínio	Dispositivos de Suporte	Estatística de tráfego	Compromete Privacidade	Rede Esparsa	Número de Cópias
GeOpps	Sim	V2I	Não	Não	Não	Sim	única
GeoSpray	Sim	V2I	Não	Não	Não	Sim	múltiplas
GeoDTN+Nav	Sim	V2I	Não	Não	Não	Sim	única
TBD	Sim	V2I	AP	Sim	Não	Sim	única
TSF	Sim	I2V	AP, CCT, NR	Sim	Não	Sim	múltiplas
STDFS	Sim	V2V, V2I	AP e CCT	Sim	Sim	Sim	única

Tabela 2.1: Comparação qualitativa entre os protocolos de roteamento geográfico com base em trajetórias de veículos.

VADD é um protocolo para roteamento de mensagens em VANET que considera a seguinte estratégia na decisão de encaminhamento: cada veículo possui um mapa digital que fornece informações de topologia geográfica das rodovias e um conjunto de dados estatísticos com densidade de tráfego ao longo do dia. Agregado a essa ideia do VADD, os protocolos TBD e TSF tem o conhecimento das trajetórias dos veículos, que auxilia no roteamento de mensagens.

No projeto do STDFS foram adotadas duas estratégias de semelhança com os demais protocolos: i) o compartilhamento periódico de trajetórias a partir de um ponto de acesso e ii) a modelagem do atraso de entrega esperado para cada mensagem. Vale ressaltar que para a construção da linha do tempo do fluxograma da figura 2.20 foram utilizados os anos de primeira publicação dos protocolos, visto que a maioria teve versões estendidas recentemente.

Além dessas considerações sobre o compartilhamento de estratégias entre os protocolos, é fácil perceber que os seis protocolos utilizam um sistema de navegação para determinação da trajetória em conjunto com o esquema armazena-e-encaminha. Portanto, todos são interligados ao bloco *DTN based*.

A tabela 2.1 apresenta uma comparação qualitativa entre os protocolos descritos nesta seção. Para enfatizar o uso das trajetórias, foi listado que todos os protocolos utilizam a trajetória de deslocamento dos veículos para auxiliar na tomada de decisão durante o encaminhamento das mensagens. Para a obtenção das trajetórias, eles se aproveitam do sistema de navegação embutidos nos veículos, em que a trajetória é

especificada antes do início do trajeto. No entanto, caso os motoristas se desloquem ao longo dos dias por trajetos conhecidos (e.g., entre suas residências e o local de trabalho) e, conseqüentemente, não precisem utilizar tais sistemas de navegação, pode-se assim dizer que existe um comprometimento no funcionamento desses protocolos.

Todos os protocolos, com exceção do TSF e do STDFS, foram projetados para fornecer somente comunicação V2I (veículo-infraestrutura) em uma rede veicular, dessa forma, eles realizam um modelo de disseminação *unicast*, podendo ser utilizados em *geocast* quando o veículo origem estiver externo a zona de relevância. Vale ressaltar que a comunicação V2I refere-se ao encaminhamento de uma mensagem enviada de um veículo origem para uma RSU específica. O TSF, inicialmente, foi proposto para o domínio I2V (infraestrutura-veículo), no entanto, os autores comentam sobre a possibilidade de integração com o TBD. Assim, tem-se um TSF que abrange os outros domínios. O STDFS se aplica nos domínios V2I e V2V (veículo-veículo), mas, para tanto, expõe a privacidade dos motoristas, compartilhando trajetórias durante o roteamento de mensagens.

Os protocolos TSF e STDFS podem ser direcionados a um maior número de aplicações, visto que podem ser utilizados em vários domínios. No entanto, estes protocolos e o TBD precisam de dispositivos auxiliares para o processo de encaminhamento de mensagens. No caso do TBD, somente pontos de acessos (AP) são necessários. Já o TSF necessita adicionalmente de nós retransmissores (NR) e um centro de controle de tráfego (CCT). Enquanto que o STDFS precisa de um CCT e APs para fornecer informações de trajetórias de outros veículos. Esses dispositivos de suporte são necessários nesses três protocolos, pois tanto auxiliam no encaminhamento de mensagens como fornecem informações estatísticas de tráfego.

Todos os protocolos são adequados para o encaminhamento de mensagens quando a rede veicular se torna esparsa. No entanto, o GeoDTN+NAV, por ser um protocolo híbrido, é indicado para cenários de rede veicular densos e esparsos. Quanto ao número de cópias de uma mensagem na rede, a maioria dos protocolos trabalha com uma única cópia de mensagem na rede. Somente o GeoSpray e o TSF permitem uma ou múltiplas cópias da mensagem. Como vantagem do uso controlado de múltiplas cópias, temos um aumento na taxa de entrega de mensagens e diminuição do atraso. No entanto, aumenta-se o *overhead* devido ao número de transmissões.

2.4 Considerações Finais

Este capítulo apresenta uma revisão bibliográfica dos temas relacionados com esta dissertação. Além dos conceitos de redes veiculares e dos modelos de disseminação, ele apresenta uma teoria sobre a movimentação de objetos no espaço geográfico. O objetivo nessa estrutura é mostrar como a pesquisa nas áreas de informação geográfica e mineração de dados pode ser útil como recurso na resolução de problemas e no fornecimento de serviços em redes veiculares. Como o foco desta dissertação é no uso de trajetórias de deslocamento dos veículos, realizou-se uma revisão de literatura dos protocolos que utilizam o conceito de trajetória para auxiliar no roteamento de mensagens.

Apesar de existirem vários protocolos na literatura que exploram o uso de trajetórias de veículos, poucos focam em soluções que observam o histórico de trajetórias de um veículo para auxiliar no processo de encaminhamento de mensagens na rede. As abordagens investigadas são baseadas em sistemas de navegação nos quais o condutor do veículo informa o seu destino e, então, obtém uma trajetória para deslocamento. No entanto, vale ressaltar que os condutores costumam seguir rotas similares (e.g. entre casa e o trabalho) e, geralmente, conhecidas por eles, ao longo dos dias, comprometendo o uso de sistemas de navegação para tal finalidade. Dessa forma, o objetivo é explorar apenas os registros de movimentações diárias dos veículos para identificarmos padrões de trajetórias de cada veículo. Além disso, o conhecimento da rotina de deslocamento dos veículos permite a concepção de protocolos de roteamento voltados para a comunicação veículo-veículo.

Nesse sentido, os próximos capítulos apresentam dois protocolos para redes veiculares que utilizam a descoberta de padrões de trajetórias individuais dos veículos para auxiliar o roteamento. O primeiro foca na informação espacial das movimentações diárias, por isso é direcionado para a comunicação veículo-infraestrutura. O segundo considera tanto a informação espacial como a temporal, dessa forma, pode ser aplicado para comunicação veículo-veículo. Além do funcionamento, a diferença para construção dos dois protocolos é o comportamento do nó destino. No primeiro, o nó destino é fixo, por isso somente a informação espacial é suficiente. No segundo, o nó destino é móvel (veículo). Assim, é necessário estimar em qual momento a mensagem será entregue ao veículo destino.

Capítulo 3

Roteamento Baseado em Informação Espacial

3.1 Introdução

Nos últimos anos, ocorreu um grande crescimento no número de veículos equipados com dispositivos GPS. Estes dispositivos oferecem variados tipos de aplicações, tais como rastreamento de frota de veículos, recomendação de rotas e alerta de congestionamentos. Tudo isso é possível devido à utilização de três informações básicas, a posição geográfica, velocidade e tempo. Embora todas essas informações estejam sendo utilizadas em aplicações de tempo real, existe uma tendência para a utilização de dados históricos gravados a partir desses dispositivos. Isso é devido aos dados conterem informações que podem ser identificadas por meio de técnicas de mineração e análise de dados.

A coleta, o registro e o uso de dados de dispositivos GPS oferecem oportunidades para a construção de uma nova gama de aplicações em redes veiculares, fornecendo informações relevantes de deslocamento e fluxo de tráfego de veículos. Por exemplo, o entendimento da regularidade e da variabilidade das rotinas de deslocamento dos veículos, a partir de dados de movimentações, pode ser usado como uma importante fonte de informação para auxiliar no processo de tomada de decisão e provimento de serviços.

Nesse sentido, um dos objetivos deste trabalho é investir na análise de históricos de movimentações de veículos para adquirir um conhecimento expressivo sobre as rotinas de deslocamento. Estes dados servem para auxiliar no processo de roteamento de mensagens em VANET. Na literatura, as abordagens atuais de roteamento, que utili-

zam informação espacial, podem ser divididas em dois grupos: os protocolos baseados em posição geográfica e os que utilizam sistemas de navegação. O primeiro utiliza a posição geográfica e um método de *greedy* na disseminação de mensagens; o segundo considera a rota de deslocamento provida pelos sistemas de navegação dos veículos para o encaminhamento de mensagem. O segundo grupo tem certa vantagem em relação ao primeiro, pois possui uma informação completa sobre a trajetória dos veículos. No entanto, só é aplicável em redes veiculares nas quais os motoristas informam ao sistema de navegação os destinos antes de iniciarem a viagem e seguem a trajetória recomendada pelo sistema [Leontiadis et al., 2010].

A utilização de sistemas de navegação é comumente aplicada em situações em que os motoristas desconhecem a rota que devem seguir e, por isso, deixam de ser adequados para deslocamento diário, visto que os motoristas costumam seguir rotas conhecidas ao longo dos dias [Froehlich & Krumm, 2008b]. Nesse sentido, acredita-se que a investigação de dados de informação espacial de rotinas de deslocamento fornece uma valiosa informação no processo de encaminhamento de mensagens.

Geralmente, as pessoas possuem rotinas com um elevado grau de regularidade espacial e temporal, em que cada indivíduo é caracterizado por frequentar certos lugares, em intervalos de tempo regulares [Gonzalez et al., 2008]. Da mesma forma, as pessoas costumam se deslocar por rotas familiares ao longo dos dias, ou seja, seguem rotas similares no trajeto entre suas residências e locais de trabalho. Sabendo disso, existem na literatura trabalhos que exploram essa regularidade diária de deslocamento para prover serviços e aplicações. Por exemplo, Trasarti et al. [2011] propõem uma metodologia para a extração de perfis de mobilidade de pessoas a partir de dados GPS, aplicando-a em um serviço que sugere caronas, tendo como base o relacionamento de perfis similares de deslocamento entre os usuários. Palma et al. [2008] apresentam uma solução para a descoberta de pontos de interesse em trajetórias. Dessa forma, eles tentam identificar a semântica da trajetória por meio da detecção de lugares que os usuários estiveram enquanto se deslocavam. Assim, baseados em um histórico, podem fornecer serviços de recomendação de novos lugares similares.

Para cada aplicação, é interessante analisar que tipo de informação é relevante extrair dos dados, pois, em alguns casos, somente o conhecimento de informações espaciais é suficiente, enquanto em outros é necessário combinar a informação espacial com a temporal. A figura 3.1 apresenta uma ilustração que mostra as duas visões (i.e., espacial e temporal) sobre um mesmo conjunto de trajetórias. Considerando a visão temporal, há a identificação de dois grupos de trajetórias que foram formados por um critério de similaridade. No entanto, na visão espacial, há a formação de apenas um grupo, pois o tempo, que era o fator determinante para particionar em dois grupos,

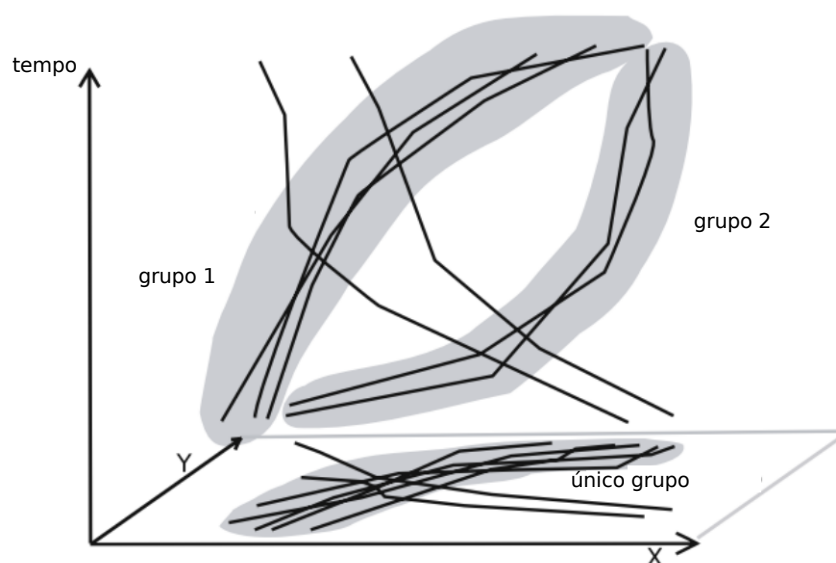


Figura 3.1: Ilustração da diferença entre a exploração da informação espacial e temporal. Considerando a visão espacial, somente um grupo é identificado. No entanto, quando a visão temporal é também utilizada, dois grupos distintos são identificados.

não é considerado neste caso. Um cenário para esse exemplo é o deslocamento de um usuário entre sua residência e o local onde trabalha. Considerando apenas a informação espacial, ocorre a formação de um grupo, no qual o veículo se desloca aproximadamente pela mesma rota ao longo dos dias. No entanto, com a adição da informação temporal, dois grupos são criados, um no sentido residência-trabalho (período da manhã) e o outro no sentido trabalho-residência (período da tarde), visto que cada sentido é realizado em diferentes momentos do dia.

Nesse sentido, este capítulo aborda a exploração da informação espacial para auxiliar no encaminhamento de mensagens em uma VANET. Assim, é apresentado um esquema de roteamento que utiliza a informação espacial no processo de seleção de nós intermediários para direcionar as mensagens para uma localização específica. Para tanto, a seção 3.2 apresenta métodos de análise da informação espacial para favorecer o entendimento do esquema proposto na seção 3.3. Em seguida, a seção 3.4 apresenta uma análise de desempenho do esquema proposto e, por fim, são descritas as considerações finais na seção 3.5.

3.2 Explorando a Informação Espacial

Considerando que a informação espacial é relacionada com a localização de uma entidade em um espaço geográfico, o significado semântico da informação presente em uma

base de dados de movimentações de um objeto depende da técnica de localização e, conseqüentemente, da granularidade dos dados. No contexto de redes veiculares, Boukerche et al. [2008] fizeram um levantamento de técnicas de localização que podem ser utilizadas para determinar o posicionamento de veículos. Entretanto, mesmo com uma variedade de técnicas, as mais populares e comumente usadas pelas aplicações para registro de dados de movimentações são GPS, WiFi e GSM. A tabela 3.1 apresenta uma comparação dessas três técnicas, destacando o tipo de dado registrado e outras quatro propriedades [Hightower & Borriello, 2001]:

- **Dado:** É uma leitura, observação ou registro a partir dos quais as informações são derivadas. Como se trata de um conceito abstrato, há várias formas para capturar e armazenar dados, dependendo da aplicação. Neste trabalho, as técnicas GPS, WiFi e GSM fornecem diferentes tipos de dados que podem ser utilizados na extração de informações. Por exemplo, o GPS provê dados em formatos de coordenadas geográficas (e.g., latitude e longitude);
- **Tipo de Informação:** Essas técnicas de localização podem fornecer dois tipos de informação: física ou simbólica. O GPS provê uma posição física. Por exemplo, o *Shopping Center* está localizado em $3^{\circ} 45' 24'' S$ e $38^{\circ} 29' 17'' W$. Por outro lado, a informação simbólica engloba uma ideia abstrata de localização de um objeto. Por exemplo, na sala de estar. Uma informação física pode ser enriquecida para fornecer uma informação de localização simbólica. Um exemplo canônico dessa abordagem é a utilização da técnica de *reverse geocoding* para obter um nome de um ponto de interesse a partir de coordenadas geográfica (i.e., latitude e longitude). Além disso, as aplicações podem utilizar a informação física para obter um conjunto de simbólica. Por exemplo, dada uma localização física, obter todos os restaurantes japoneses no raio de três quilômetros;
- **Sistema de Referência:** Quanto ao sistema de referência, a técnica de localização pode ser classificada em absoluta ou relativa. Uma técnica de localização absoluta usa um sistema de referência compartilhado entre todos os dispositivos. Por exemplo, todos os dispositivos GPS usam latitude, longitude e altitude para reportar a localização de um objeto. Assim, dois dispositivos GPS que realizam uma leitura na mesma posição, reportam resultados equivalentes. Por outro lado, a técnica de localização relativa possui seu próprio sistema de referência. Por exemplo, bombeiros podem buscar vítimas de um desastre usando *smartphones* para obter a posição relativa entre eles e as vítimas;

Técnicas	Dado	Tipo de Informação	Sistema de Referência	Acurácia	Cobertura
GPS	coordenadas geográficas	Física	Absoluto	1 - 5 metros	<i>Outdoors</i>
Rede Celular (e.g, GSM)	ID da torre celular + nível de sinal ou coordenadas geográficas	Física/Simbólica	Absoluto/Relativo	50 - 200 metros nas cidades	Cobertura da célula da rede (5 - 30 km de uma torre)
WiFi	ID do ponto de acesso + nível do sinal	Física/Simbólica	Relativo	1 - 20 metros	cerca de 100 m de um ponto de acesso

Tabela 3.1: Características do GPS, GSM e WiFi quando utilizados como técnicas de localização.

- **Acurácia:** refere-se a margem de segurança entre a localização registrada como dado em relação à posição real. Dependendo dos requisitos da aplicação é necessário uma precisão maior em relação ao dado que foi registrado;
- **Cobertura:** área de cobertura da técnica de localização de acordo com a infraestrutura.

Dentre as técnicas apresentadas na tabela 3.1, o GPS é a técnica mais adequada para capturar informações de posicionamento em redes veiculares. Primeiro, os sistemas de navegação e localização embutidos nos veículos são baseados em posicionamento por GPS. Segundo, comparando as características apresentadas, os dados de GPS possuem uma maior consistência, grau de certeza e confiabilidade em relação ao posicionamento dos objetos do que os dados típicos das outras técnicas. Nesta dissertação, o foco é na utilização de dados de dispositivos GPS. Então, a partir de uma base com dados coletados por dispositivos GPS é possível explorar a informação espacial utilizando diversos métodos de mineração de dados dependendo do problema em questão.

Nesse sentido, a tabela 3.2 apresenta um conjunto de problemas e os métodos mais utilizados para solucioná-los com base em uma revisão de literatura. A seguir, é introduzida uma breve conceituação desses métodos, destacando suas relevâncias na exploração da informação espacial, a partir de dados de localização geográfica. Uma revisão mais abrangente desses métodos pode ser encontrada em [Pan et al., 2013b].

- **Ranking:** O método de *ranking* é utilizado para encontrar as rotas e regiões mais desejadas. Por exemplo, recomendação de locais mais populares. O funcionamento visa obter um modelo que mapeie as amostras para intervalos de

relevância baseados em um ordenamento parcial, dado um conjunto de amostras e suas propriedades. *Ranking* tem sido amplamente utilizado para encontrar pessoas relevantes em redes sociais baseadas em localização para realizar recomendação de rotas [Kurashima et al., 2010].

- **Classificação:** É um método de aprendizagem supervisionada que identifica a que classe pertence cada amostra. Para tanto, as classes devem ser definidas a priori. Assim, um modelo é obtido a partir de um conjunto de treinamento que possui os rótulos de cada classe. Métodos de classificação têm sido aplicados para determinar modos de transporte [Reddy et al., 2010a], eventos sociais [Calabrese et al., 2010] e regiões semânticas [Pan et al., 2013a].
- **Regressão:** É utilizada para estimar a relação entre variáveis. Dessa forma, constrói-se uma função contínua entre uma variável dependente e uma ou mais variáveis independentes, sendo que a variável dependente é definida de acordo com o problema específico. Por exemplo, o grau de afinidade entre indivíduos para construir um relacionamento social. Eagle et al. [2009] propõem um método para medir a correlação entre amizade e localização entre indivíduos baseado em dados de mobilidade.
- **Clustering:** É um método de aprendizagem não supervisionada para agrupar amostras com base na similaridade existente entre elas, sem qualquer conhecimento a priori do grupo alvo. Considerando a informação espacial, *clustering* pode ser aplicado para amostras sendo pontos ou trajetórias. Métodos de *clustering* têm sido aplicados para encontrar pontos de interesse [Andrienko et al., 2011] e padrões de movimentação [Giannotti et al., 2007].

Com base nessa revisão exposta sobre a informação espacial, é adotado neste trabalho um método de *clustering*, denominado OPTICS (*Ordering Points To Identify the Clustering Structure*) [Ankerst et al., 1999], a fim de alcançar o objetivo de encontrar padrões de trajetórias similares de um indivíduo, a partir de dados brutos, coletados pelas movimentações diárias. Dessa forma, esses padrões são utilizados para auxiliar no encaminhamento de mensagens em uma rede veicular. Portanto, no restante desta seção, é apresentado o processo utilizado para obtenção de trajetórias representativas individuais destacando o algoritmo utilizado para detectar padrões de trajetórias. Vale ressaltar que o processo utilizado segue o que foi apresentado na seção 2.2.2.

Para o início do processo é considerada a existência de uma base de dados¹ com informações de trajetórias, coletada por um dispositivo GPS contemplando o que foi

¹Para este caso, foi definido que cada veículo possui uma base de dados de movimentação.

	<i>Ranking</i>	Classificação	Regressão	<i>Clustering</i>
Recomendação de rotas	X			
Reconhecimento de atividades		X		
Inferir tipo de eventos		X		
Correlação de rede sociais			X	
Reconhecimento de regiões semânticas		X		
Comportamento de movimentação				X

Tabela 3.2: Relação entre os problemas e métodos para solucioná-los.

descrito sobre dados no início desta seção. Mesmo sabendo que são dados de trajetórias, estes não possuem informações explícitas sobre origem e destino das trajetórias. Além disso, não há lugares semanticamente definidos, apenas coordenadas geográficas. Assim, informações relevantes, como locais e trajetórias, são extraídas a partir de uma análise de dados.

Uma importante tarefa para a análise de movimento é a detecção de lugares significantes, também conhecidos como pontos de interesse (PoI). Por exemplo, para um certo conjunto de dados, é possível identificar pontos de interesse que incluem as residências dos usuários, local de trabalho, lugares que são frequentados regularmente, tais como *shopping* e supermercados. A descoberta dos pontos de interesse auxilia na extração de trajetórias, visto que os usuários se deslocam regularmente entre lugares nos quais permanecem por um determinado tempo.

Como dito anteriormente, a movimentação dos objetos é representada por dados sequenciais de registro de posição. Essa movimentação necessita ser particionada em entidades menores, correspondendo a trajetória entre dois pontos de interesse. Por exemplo, nota-se na figura 3.2 a presença de pontos de interesses (e.g., casa, local de trabalho) e trajetórias entre eles (e.g, T_0, T_1, T_2, T_3). A noção de trajetória e pontos de interesses pode ser explorada de diferentes maneiras, dependendo dos objetivos e da aplicação. A figura mostra uma possível forma de dividir os dados de movimentação em trajetórias, que é adequada para deslocamento utilizando veículos. Ela é baseada em um valor máximo de tempo de permanência em um determinado lugar. Por exemplo, dependendo da aplicação, se um veículo permanece em um mesmo local por um tempo

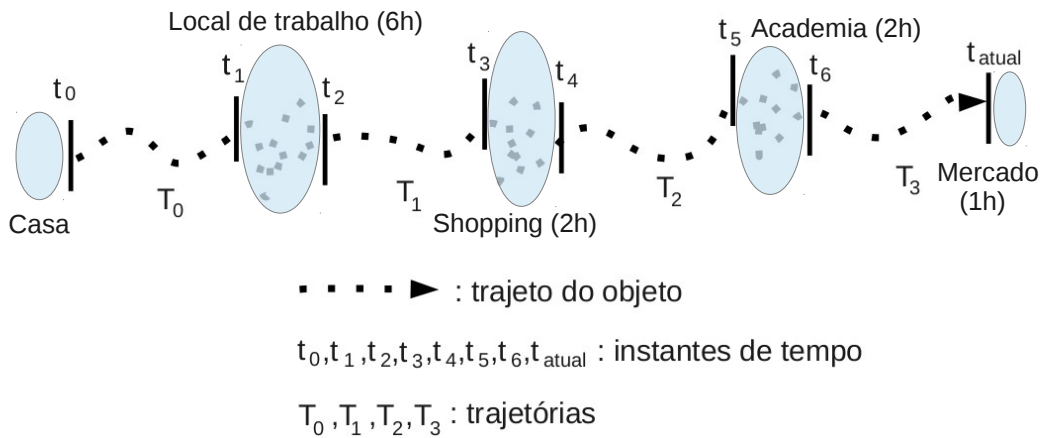


Figura 3.2: Ilustração da existência de pontos de interesse durante a movimentação diária.

estimado de 6 horas, provavelmente o usuário desse veículo está no local de trabalho². Diversos trabalhos na literatura têm investigado a busca semântica de movimentação. Como nesta dissertação o foco não é a questão semântica de PoI, considera-se que os veículos param em pontos de interesse, na maioria das vezes em casa e no local de trabalho e, a partir disso, as trajetórias são identificadas.

Supondo que as pessoas possuam rotinas para frequentar pontos de interesse, o objetivo é buscar diferentes aspectos que caracterizam a similaridade de trajetórias de deslocamento entre esses PoI, tais como locais de início e fim, sobreposição total ou parcial de trajetos, entre outros. A escolha desses aspectos depende dos requisitos da aplicação e dos objetivos de análise. Para este trabalho, a ideia é realizar *clustering* de trajetórias. Para tanto, é adotada uma implementação do algoritmo OPTICS para o agrupamento de trajetórias [Andrienko et al., 2007]. A ideia básica é separar o conjunto de dados com informações de trajetórias em grupos, tendo como base a similaridade existente entre elas.

Para tanto, o OPTICS baseia-se na ordenação do conjunto de dados, considerando dois parâmetros, denominados *core distance* e *reachability distance*, para cada amostra. Uma breve explicação do OPTICS é necessária para auxiliar no entendimento do algoritmo. Seja ρ uma amostra de um conjunto de dados D , ε uma distância máxima, $N_\varepsilon(\rho)$ um subconjunto de amostras de D que possuem distância menor que ε em rela-

²Outras informações podem ser adicionadas para enriquecer os dados de trajetórias.

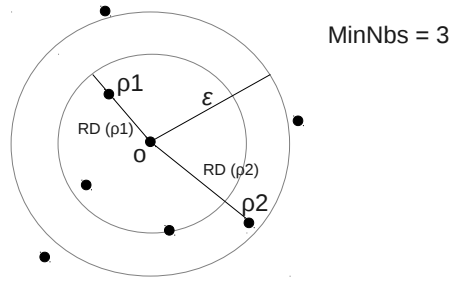


Figura 3.3: Ilustração das definições de *core object* e *reachability distance* no algoritmo OPTICS.

ção à ρ (i.e., $x \in D | d(\rho, x) < \varepsilon$), $minNbs$ um número natural, $minNbs - distance(\rho)$ uma distância de ρ ao seu $minNbs$ -ésimo vizinho em ordem de proximidade. A *Core Distance* (CD) é definida como:

$$CD_{\varepsilon, minNbs}(\rho) = \begin{cases} INDEFINIDA & , |N_{\varepsilon}(\rho)| < minNbs \\ minNbs - distance(\rho) & , \text{ caso contrário} \end{cases}$$

Com base nisso, a *core distance* representa a menor distância ε entre ρ e uma amostra na sua vizinhança. Caso contrário, a *core distance* é indefinida. A *reachability distance* assume que ρ e o são amostras de D , $N_{\varepsilon}(o)$ é a vizinhança de o limitada pelo valor ε , $minNbs$ um número natural. A *Reachability Distance* (RD) de ρ em relação à o é definida como:

$$RD_{\varepsilon, minNbs}(\rho, o) = \begin{cases} INDEFINIDA & , |N_{\varepsilon}(o)| < minNbs \\ max(core - distance(o), distance(o, \rho)) & , \text{ caso contrário} \end{cases}$$

A figura 3.3 ilustra um exemplo que envolve as definições de *core object* e *reachability distance*. A amostra o é um *core object* desde que na vizinhança de raio ε exista pelo menos outras 3 amostras ($minNbs = 3$). A *core distance* da amostra o é dada pela distância dela até o terceiro vizinho mais próximo (círculo de raio menor). A *reachability distance* da amostra $\rho1$ denotada por $RD(\rho1)$ é igual a *core distance*. A *reachability distance* da amostra $\rho2$, denotada por $RD(\rho2)$, é igual a distância de o até $\rho2$.

O algoritmo OPTICS atribui uma *reachability distance* para cada amostra de D . Inicialmente, uma amostra ρ_0 é escolhida aleatoriamente. Então, em cada iteração i , a próxima amostra ρ_i escolhida deve ter a menor *reachability distance* em relação aos *core*

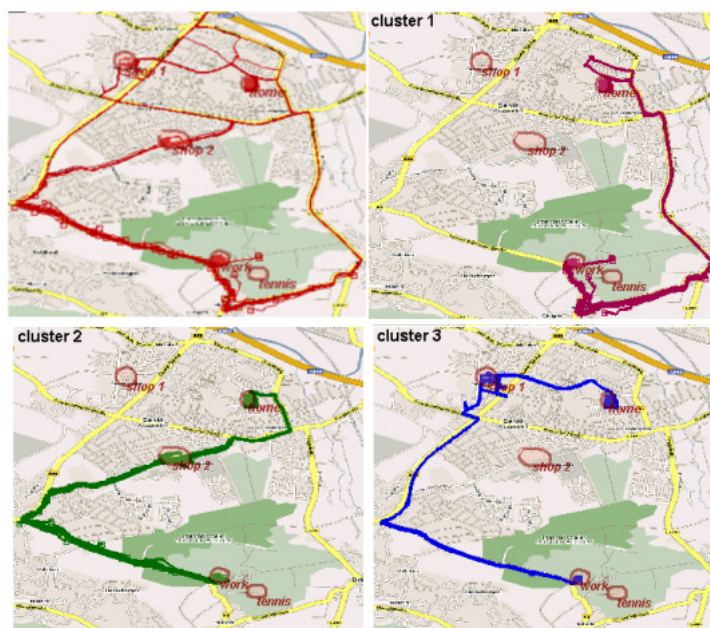


Figura 3.4: Identificação de *clusters* no trajeto entre o local de trabalho (*work*) e casa (*home*) de acordo com a rota seguida [Andrienko et al., 2007].

objects já visitados. Como saída, o OPTICS produz um gráfico chamado *reachability plot*, que possui no eixo das abscissas as amostras ordenadas por visitação e no eixo das ordenadas o valor das *reachability distance*. Assim, os *clusters* são obtidos pela escolha de um valor limiar de *reachability distance*.

Dada essa explanação sobre o OPTICS e realizando uma adaptação para o contexto deste trabalho, as trajetórias são as amostras e a distância entre as trajetórias é dada por uma função de distância. A função de distância usada neste trabalho foi a proposta por Andrienko et al. [2007] denominada *route similarity*. A figura 3.4 apresenta um exemplo de identificação de *clusters* de trajetórias de acordo com as rotas seguidas entre dois locais. Vale salientar que a função de similaridade de rotas utilizada desconsidera a informação temporal, dessa forma, somente a informação espacial é relevante para detecção dos *clusters*.

Dados os *clusters* obtidos, o próximo objetivo é sumarizar esses *clusters*. Nesse sentido, foi adotado o algoritmo proposto por Andrienko et al. [2007] para obtenção de trajetórias sumarizadas. A figura 3.5 mostra o produto da sumarização procedente dos *clusters* da figura 3.4. As setas representam o sentido das trajetórias e a largura dessas setas são proporcionais ao número de trajetórias existentes no *cluster*, ou seja, da frequência que a rota foi seguida. Com a sumarização realizada, identifica-se no conjunto de dados a trajetória que reflete um *cluster*, sendo a mais frequente classificada como trajetória representativa.

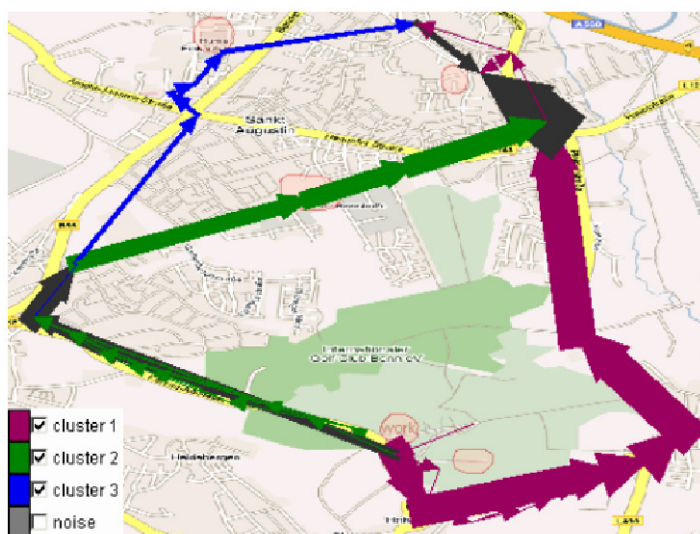


Figura 3.5: Representação de trajetórias sumarizadas a partir dos *clusters* da figura 3.4 [Andrienko et al., 2007].

3.3 Descrição do Algoritmo de Roteamento

Nesta seção, é proposta uma abordagem para roteamento geográfico que explora informações espaciais de deslocamento diário dos veículos para encaminhar mensagens para um local específico. Essa abordagem, denominada GeoSPIN³ (*Geographic routing based on SPatial INformation*) [Celes et al., 2013], é dividida em duas fases. A primeira fase extrai os padrões de trajetórias de cada veículo a partir de dados GPS coletados ao longo dos dias. A segunda fase aplica um algoritmo de roteamento distribuído para encaminhamento das mensagens baseado nos padrões de trajetórias individuais. As duas fases são descritas em detalhes a seguir, após a exposição de algumas considerações preliminares.

3.3.1 Considerações Preliminares

Uma trajetória espacial pode ser definida como um rastro gerado pela movimentação de um objeto em um espaço geográfico, contendo uma sequência de coordenadas geoespaciais e momentos de registro. Por exemplo, a movimentação de objetos é uma representação de um deslocamento migratório de animais ou uma pessoa que se conduz de casa para o local de trabalho todos os dias.

³Esta seção é baseado no artigo *GeoSPIN: An approach for Geocast routing based on SPatial INformation in VANETs* apresentado no IEEE 78th Vehicular Technology Conference (VTC2013-Fall)

Enquanto as capacidades de capturar e armazenar uma sequência de posições são atividades básicas para o estudo de movimentações, o projeto de aplicações baseadas em trajetórias requer uma representação de dados bem definidas. Adicionalmente, muitas aplicações necessitam de dados estruturados, como uma sequência temporal de jornadas, cada uma tendo o ponto de partida para um ponto de destino de deslocamento. Para auxiliar na explanação do protocolo, é considerada uma representação estruturada dos dados e algumas definições de representação são adicionadas aos conceitos discutidos na seção 2.2.

Nesse sentido, as definições a seguir são úteis para sustentar a discussão. **Ponto** (P) é uma tupla de informação (i.e., latitude, longitude, altitude, *timestamp*) coletados por um dispositivo GPS com um específico *id*. Nesta dissertação, o foco é em um espaço de duas dimensões para representar a posição corrente de um veículo. Então, uma amostra do conjunto de dados no tempo t é $P = (id, x_t, y_t, t)$. **Trajétoria** (T) é um conjunto de k consecutivos pontos com informação espaço-temporal capturado por um dispositivo GPS durante um trajeto de um usuário no seu veículo: $T = \{P_1, P_2, \dots, P_k\}$. **Histórico** é um conjunto de m trajetórias diárias e que pode ser definido como $Histórico = \{T_1, T_2, \dots, T_m\}$. **Padrão** é um conjunto de c trajetórias representativas (*Traj*), que são descobertas como resultados dos algoritmos de mineração de dados no conjunto *Histórico*, então $Padrão = \{Traj_1, Traj_2, \dots, Traj_c\}$. Para concepção do protocolo, foi considerado que as trajetórias não são decompostas, ou seja, um padrão é uma trajetória inteira entre duas localidades específicas. Além disso, é assumido que cada veículo possui um dispositivo capaz de capturar e armazenar o seu atual posicionamento geográfico e com capacidade computacional para executar os algoritmos descritos nas subseções seguintes.

3.3.2 Fase 1 - Obtenção de Trajetórias Representativas

Pessoas seguem trajetos similares ao longo dos dias, que podem ser considerados rotinas diárias de deslocamento (e.g., trajetórias similares para ir de casa ao local de trabalho). Nesta fase, é aplicado um processo de *clustering* a fim de obter informações relevantes das trajetórias diárias dos motoristas. Como explanado anteriormente, *Clustering* é uma técnica que permite identificar similaridades e dissimilaridades de objetos e agrupá-los em *clusters*. Para tanto, foi adotado o OPTICS. Esse algoritmo identifica *clusters* de acordo com a densidade de dados GPS registrados em um certo intervalo de tempo, ou seja, detecta trajetórias similares de um veículo no intervalo de tempo, por exemplo, entre [7:00 am - 8:00 am].

O algoritmo 1 mostra os principais passos desta primeira fase. Sabe-se que cada



Figura 3.6: Ilustração do processo para obter trajetórias representativas de um usuário.

Algoritmo 1: GeoSPIN - Fase 1 - Obtenção de trajetórias representativas

Entrada: Um conjunto de trajetórias de um veículo

Saída: Cada trajetória representativa com sua respectiva probabilidade

- 1 procedimento `GeoSPINRoutingPhase1()`
 - 2 $Histórico \leftarrow obterTodasTrajetoriasIndividuais()$
 - 3 $Clusters \leftarrow aplicarOpticsClustering(Histórico)$
 - 4 $Padrão \leftarrow sumarizarClusters(Clusters)$
 - 5 $Prob_Traj \leftarrow obterProbabilidades(Padrão)$
 - 6 fim procedimento
-

veículo executa o algoritmo individualmente. O primeiro passo é a obtenção de todos os dados de deslocamento diário (armazenados pelo dispositivo GPS como trajetórias) e atribuí-los a uma estrutura chamada *Histórico* (algoritmo 1, Linha 2). Em seguida, aplica-se *Histórico* na atividade de *clustering* usando o algoritmo do OPTICS (algoritmo 1, Linha 3). Basicamente, nesse passo, o algoritmo remove ruídos e gera *clusters* densos de formas arbitrárias enquanto a função distância do OPTICS compara duas trajetórias para a distância Euclidiana entre elas. O algoritmo de *clustering* OPTICS necessita de dois parâmetros para executar a descoberta de *clusters*: o número mínimo de amostras similares (amostras próximas) e a distância máxima entre vizinhos. A figura 3.6(b) mostra um exemplo de *textitclusters* detectados pelo OPTICS (limiar de distância de 500m) usando uma função de similaridade de rotas. Esses *clusters* (figura 3.6(b)) são gerados a partir de um conjunto de trajetórias armazenadas no formato de dados brutos de GPS (figura 3.6(a)) de um mesmo veículo. A frequência usada para executar o algoritmo da fase 1 é determinada de acordo com um período predefinido (e.g., semanalmente, mensalmente).

Posteriormente, o algoritmo identifica as trajetórias representativas por meio de uma tarefa de sumarização (algoritmo 1, Linha 4). Essa tarefa considera pontos de interesse nas trajetórias (i.e., pontos de início de trajetória, pontos finais e paradas

realizadas) e segmentos de trajetórias que pertencem ao *cluster*. A figura 3.6(c) mostra um exemplo do produto da tarefa de sumarização. As setas representam o sentido do movimento e a espessura de cada seta é proporcional à frequência de ocorrência que um veículo segue em certo sentido. Portanto, essa informação foi utilizada para computar a probabilidade empírica de um veículo tomar cada trajetória representativa, baseando-se no número total de trajetórias que um veículo realizou (algoritmo 1, Linha 5). Assim, a trajetória representativa mais adequada é selecionada pela computação da soma das probabilidades, como segue: $\sum_{i=1}^c Prob(X = Traj_i) = 1$, onde c é o número de *clusters* encontrado pelo algoritmo OPTICS. Na figura 3.6(c) a trajetória representativa em cor verde tem probabilidade igual a 0.6, enquanto que a de cor vermelha tem probabilidade igual a 0.4.

3.3.3 Fase 2 - Encaminhamento de Mensagens

Esta segunda fase é responsável pelo processo de encaminhamento da mensagem a partir do veículo origem ao destino. Por exemplo, considerando um cenário de *geocast* no qual o veículo remetente está fora da zona de relevância, realiza-se um *unicast* até a mensagem atingir a ZOR (*Zone of Relevance*). Se o destino for uma RSU, um *unicast* é realizado até a mensagem alcançar o raio de comunicação da RSU. Para ambos os casos, existem restrições de topologia, pois, devido às características da rede veicular, é possível se deparar com situações que impossibilitam a conexão fim-a-fim entre os nós origem e destino ou circunstâncias em que não existe nenhum veículo próximo a ZOR para fazer o papel de nó intermediário. Para superar esses problemas, foi adotada a técnica de armazena-e-encaminha para possibilitar a comunicação em cenários sujeitos a desconexões. Conseqüentemente, o GeoSPIN atinge seu objetivo, apoiando redes veiculares passíveis de conexões intermitentes e esparsas.

Baseado nas considerações apresentadas na seção 3.3.1 e nos padrões de mobilidade de cada veículo (obtidos por meio da Fase 1), o GeoSPIN inicia a fase de encaminhamento de mensagens tendo duas suposições: encontro e convergência. *Encontro*(n, r) define um grupo de, pelo menos, n nós que estão oportunisticamente dentro de um raio de comunicação r em um dado instante de tempo. *Convergência*(n, r) assume que, pelo menos, n nós irão passar por uma região circular (chamada ZOR ou raio de comunicação da RSU) com raio r , não necessariamente no mesmo tempo.

Nesse contexto, foi introduzido outro importante ponto da abordagem proposta, que é relacionada à seleção do nó intermediário (portador temporário da mensagem) para encaminhamento. Esta estratégia de roteamento considera duas restrições: (i) a

Algoritmo 2: GeoSPIN - Fase 2 - Encaminhamento de Mensagens

Entrada: Um vetor resumo e a posição corrente do veículo(x, y)

```

1 procedimento GeoSpinRoutingPhase2()
2   se  $isZOR(x,y) == false$  então ▷ neste caso para geocast
3     para cada veiculo in Vizinhança faça
4        $Info \leftarrow trocarSummaryVector()$ 
5       para cada msg in Info faça
6         se  $N_c > 1$  então ▷ número de cópias ( $N_c$ ) de msg
7            $MyD_{destino} \leftarrow getD_{destino}(Padrão, msg.Q)$ 
8            $MyTraj \leftarrow getTraj(Padrão, MyD_{Destino})$ 
9           se  $MyD_{Destino} \leq no.msg.D_{Destino}$  e
              $Prob(X = MyTraj) \geq no.msg.getProb_Traj$  então
10             $ArmazenaMensagem \leftarrow requisitaMensagem(msg)$ 
11             $Atualiza o número de cópias (N_c) de msg$ 
12          fim se
13        fim se
14      fim para cada
15    fim para cada
16  fim se
17 fim procedimento

```

distância entre o padrão de mobilidade individual e a região circular; (ii) a frequência que cada veículo segue cada padrão.

O algoritmo 2 descreve o processo de encaminhamento de mensagens. Inicialmente (algoritmo 2, Linha 2), é verificado se o veículo está fora da região circular (neste caso, foi adotada a ZOR como exemplo, mas poderia ser o raio de comunicação da RSU), caso não esteja, é verificado se possui uma mensagem para o destino e envia. Se está, executa os passos restantes do algoritmo. Dessa forma, quando dois nós se encontram, eles trocam um vetor resumo (*summary vector*) contendo uma lista de mensagens com seus respectivos identificadores (algoritmo 2, Linhas 3 - 4). O *summary vector* contém também a referência da região de destino para cada mensagem (Q é a referência ou ponto central), a distância para a região destino ($D_{Destino}$) em relação ao padrão de mobilidade individual, e a probabilidade que um veículo possui para seguir um trajeto até o destino. Depois disso, cada veículo verifica se os valores de $D_{Destino}$ e a probabilidade são menor e maior, respectivamente, do que os valores do veículo vizinho. Se ambas as condições são verdadeiras (algoritmo 2, Linha 9), o veículo requisita a mensagem ao vizinho (algoritmo 2, Linha 10).

A distância entre a trajetória representativa $Traj$ e um ponto $Q_{(x,y)}$ é definida como o valor mínimo entre um ponto geográfico pertencente a $Traj$ e o ponto Q . Então,

$D_{Destino}(Traj, Q) = \min(\|P_t - Q\|)$, onde $P \in Traj$ e a distância Euclidiana ($\|P_t - Q\|$) entre dois pontos é definida como $\|P_t - Q\| = \sqrt{(P_{t.x} - Q_x)^2 + (P_{t.y} - Q_y)^2}$.

A fim de reduzir o *overhead* na rede, foi estipulado um parâmetro chamado *Número de Cópias* (N_c) para cada mensagem criada (algoritmo 2, Linhas 6 e 11). Foi adotada uma estratégia sugerida pelos autores de [Small & Haas, 2005]. Por exemplo, no instante de tempo $t = 0$ (momento de criação da mensagem), o nó 1 carrega a mensagem e marca como $N_c - 1$ o número de replicações necessárias. Quando a primeira replicação ocorre entre o nó 1 e o nó 2, ambos os nós assumem a mesma probabilidade de encontrar outro nó. Então, $(N_c - 1)/2$ é o número de replicação restantes que é atribuído para nó 1 e suas futuras replicações. As outras $(N_c - 1)/2$ replicações são atribuídas ao nó 2 e suas possíveis replicações.

3.4 Análise de Desempenho

Esta seção apresenta uma análise de desempenho do GeoSPIN por meio da comparação com outras abordagens que são introduzidas logo na sequência. Em seguida, é feito um relato sobre a configuração de simulação e uma discussão sobre os resultados obtidos.

3.4.1 Descrição do Ambiente e Cenário de Simulação

As simulações foram realizadas com o uso de um *framework* de simulação de redes veiculares (Veins)⁴ utilizado em conjunto com o simulador de eventos discretos OMNeT++ e um simulador de mobilidade e tráfego chamado SUMO [Sommer et al., 2011]. O Veins provê um modelo realístico para 802.11p DSRC e para as camadas PHY e MAC. Os parâmetros PHY e MAC são definidos de acordo com o padrão 802.11p [Jiang & Delgrossi, 2008].

Para a configuração do cenário de simulação, é adotado um conjunto de dados real chamado *Borlange* [Axhausen et al., 2013]. Esse conjunto de dados foi coletado durante dois anos (1999-2001) na cidade de Borlange, na Suécia. O conjunto de dados *Borlange* foi criado no contexto do projeto *Ratt fart* com o objetivo de auxiliar os experimentos para controle de tráfego e aplicações de segurança. Em torno de 200 veículos foram utilizados para registro de trajetórias. Segundo os autores, os dados foram manualmente checados e corrigidos, resultando em um subconjunto de dados de 24 veículos com um total de 420.814 trajetórias em uma região de $15Km \times 15Km$. Embora esse cenário seja suficiente para validar este trabalho, como trabalho futuro é

⁴Vehicles in Network Simulation - <http://veins.car2x.org/>

necessário melhorar a precisão dos resultados em cenários com uma densidade maior de veículos.

O ambiente de simulação é composto de 24 veículos com espaço de armazenamento *buffer* de mensagens ilimitado. O tempo de simulação foi de 6 horas, de forma que os veículos se movimentam nas rodovias seguindo as informações disponíveis no conjunto de dados *Borlange*. As mensagens são originadas aleatoriamente e têm como destino uma região circular aleatória. Para representar um cenário com diferentes aplicações de redes veiculares, dois fatores são variados durante as simulações: número de mensagens, o qual assume os valores 100, 325, 550, 775 e 1000; e o *Time-To-Live (TTL)*, o qual assume os valores 1800, 5400 e 9000 segundos. O TTL corresponde a um valor limite de tempo que uma mensagem pode permanecer no *buffer* até ser descartada. Por exemplo, o tempo máximo para entregar uma mensagem de uma aplicação de saúde pode ser definido como uma hora. Se esta exigência não for atendida, a mensagem pode ser descartada. Portanto, o TTL, neste caso, está diretamente relacionado com os requisitos da aplicação.

Ademais, para todos os cenários, cada configuração da simulação foi repetida 30 vezes com diferentes sementes de aleatoriedade. Os resultados apresentados possuem um intervalo de confiança de 95%.

3.4.2 Protocolos Baseline

Pelas características do cenário de simulação, é adequado uma comparação entre protocolos de roteamento tipicamente utilizados na literatura como *baseline* para redes esparsas e tolerantes a desconexões, tais como Epidêmico, Spray-And-Wait e Prophet. A seguir, é exposta uma breve introdução sobre cada um desses protocolos.

- *Epidêmico* [Vahdat & Becker, 2000] é um esquema de encaminhamento de mensagens não controlado. Quando dois nós se encontram, eles trocam um vetor com informações e requisitam uma cópia das mensagens que eles não possuem. O roteamento epidêmico é interessante como comparação, pois fornece um limite superior em termos de taxa de entrega e *overhead*, e um limite inferior com relação ao atraso de entrega da mensagem.
- *Spray-And-Wait (SAW)* [Spyropoulos et al., 2005b] segue o conceito de comunicação oportunística do protocolo Epidêmico. No entanto, ele tenta reduzir o *overhead na rede*. No início (fase chamada de Spray), o nó fonte da mensagem define um número de cópias ($L = \{10, 15\}$ nas simulações deste trabalho) para cada mensagem. Então, quando essas cópias não podem ser mais replicadas, o

protocolo inicia a segunda fase, chamada *wait*. Cada nó carrega uma cópia da mensagem e espera um contato oportunista para direcionar a mensagem em direção ao destino.

- *Prophet* [Lindgren et al., 2003] é um protocolo de roteamento probabilístico baseado na frequência de contatos entre pares de nós. Ele define um método para calcular a *probabilidade de entrega* de todos os nós (a), para cada destino (b) conhecido. Quando dois nós se encontram, eles atualizam suas *probabilidades de entrega*. Então, um nó encaminha uma mensagem para outro nó de acordo com esta probabilidade e um limiar específico. Para as simulações, foram adotados os valores dos parâmetros sugeridos pelos autores do protocolo no artigo original.

3.4.3 Resultados

Este trabalho foi avaliado em termos de taxa de entrega, número de mensagens descartadas, *overhead* e o atraso de entrega das mensagens (latência).

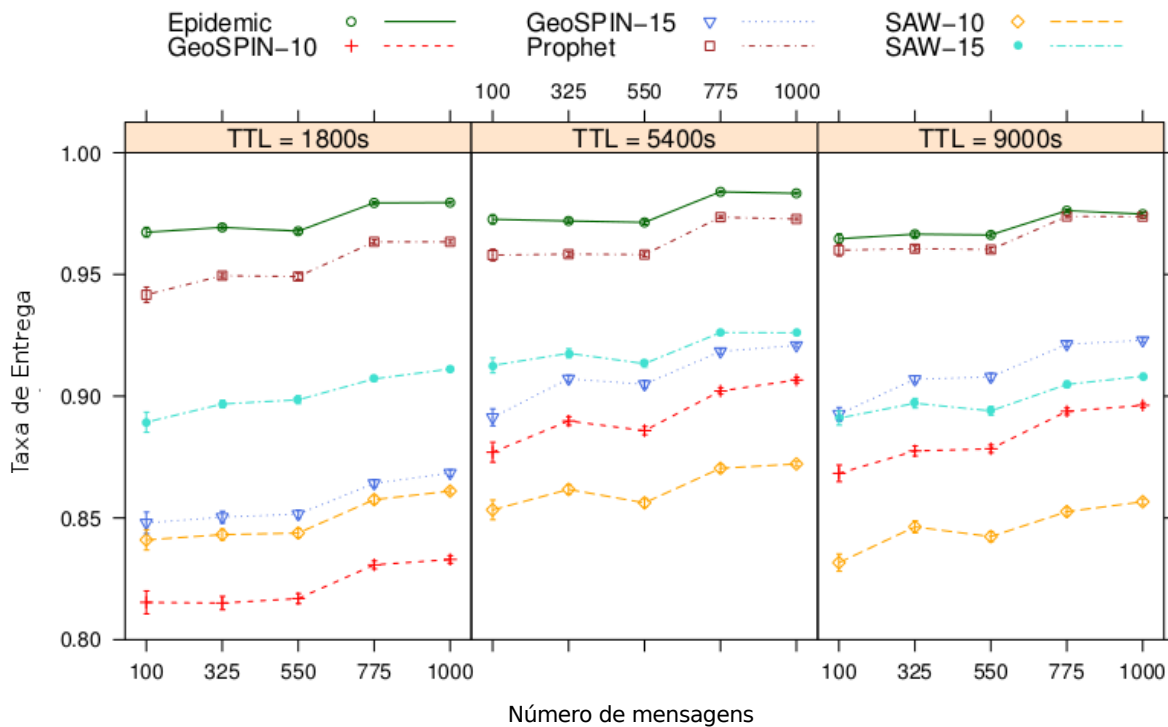


Figura 3.7: Taxa de Entrega em função do número de mensagens e TTL.

A **taxa de entrega** é dada pela relação entre o número de mensagens recebidas na região circular e o número total de mensagens enviadas pelos veículos fonte. A figura 3.7 mostra o comportamento dos protocolos quanto a taxa de entrega. Como esperado, o protocolo Epidêmico obtém altos valores. O Prophet tem uma alta taxa de entrega,

principalmente, devido a regularidade de movimentação dos veículos e, além disso, possui uma propriedade de transitividade para encontro dos nós. Com o aumento do número do TTL para cada mensagem, pode-se observar que o GeoSPIN aumenta sua taxa de entrega, superando o SAW, especialmente quando o número de cópias inicial é igual a 15. Isso ocorre porque o tempo que um veículo carrega a mensagem é suficiente para atingir a região de destino usando a estratégia do GeoSPIN. Comportamento semelhante é notado quando o número de mensagem na rede aumenta.

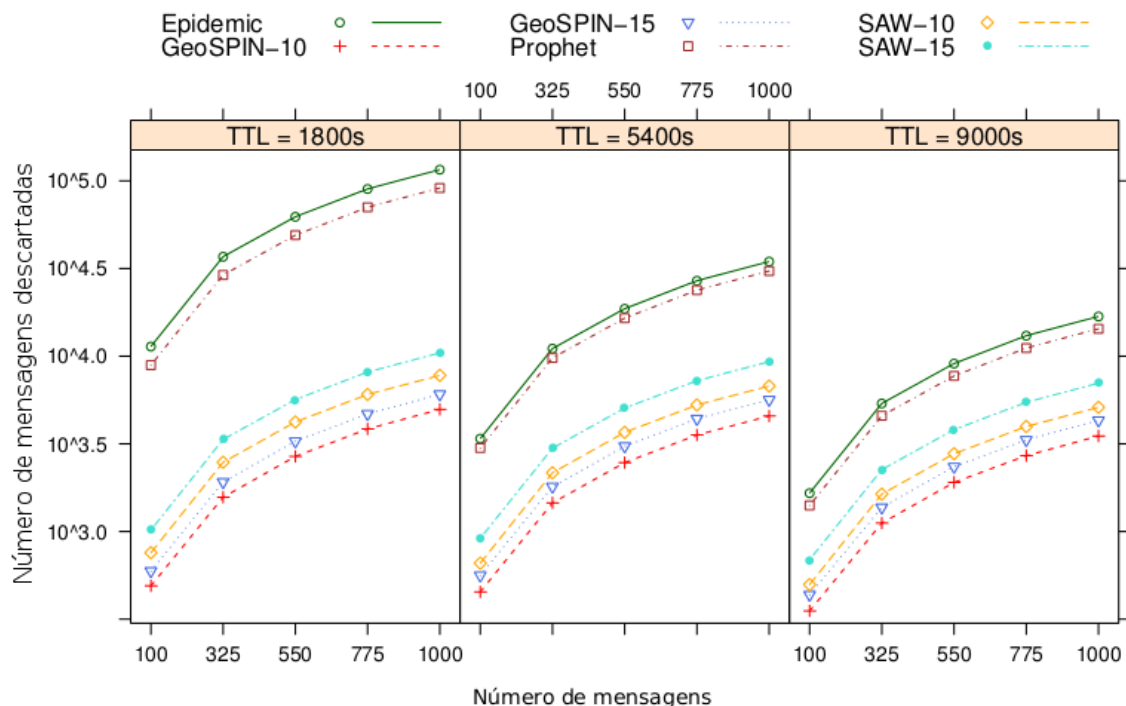


Figura 3.8: Número de mensagens descartadas em função do número de mensagens e TTL.

O número de **mensagens descartadas** corresponde ao número de mensagens descartadas a partir do *buffer* dos veículos devido ao limite do TTL. Como podemos observar na figura 3.8, o valor do TTL influencia diretamente no número de mensagens descartadas. O Epidêmico e o Prophet têm um alto número de mensagens descartadas, pois geram um número grande de cópias de mensagens na rede. Tanto o GeoSPIN como o SAW criam um pequeno número de cópias na rede. Em termos relativos entre taxa de entrega e número de cópias, o GeoSPIN obtém bons resultados. Isso é justificado pelo número de cópias geradas pelo GeoSPIN, pois ele tem uma atividade que verifica quais entre os veículos vizinhos possuem uma trajetória representativa que passa próxima a região de destino e a cópia é gerada somente em situações que a probabilidade de atingir a região de destino seja alta.

O **overhead** representa a relação entre o número de mensagens transmitidas e o

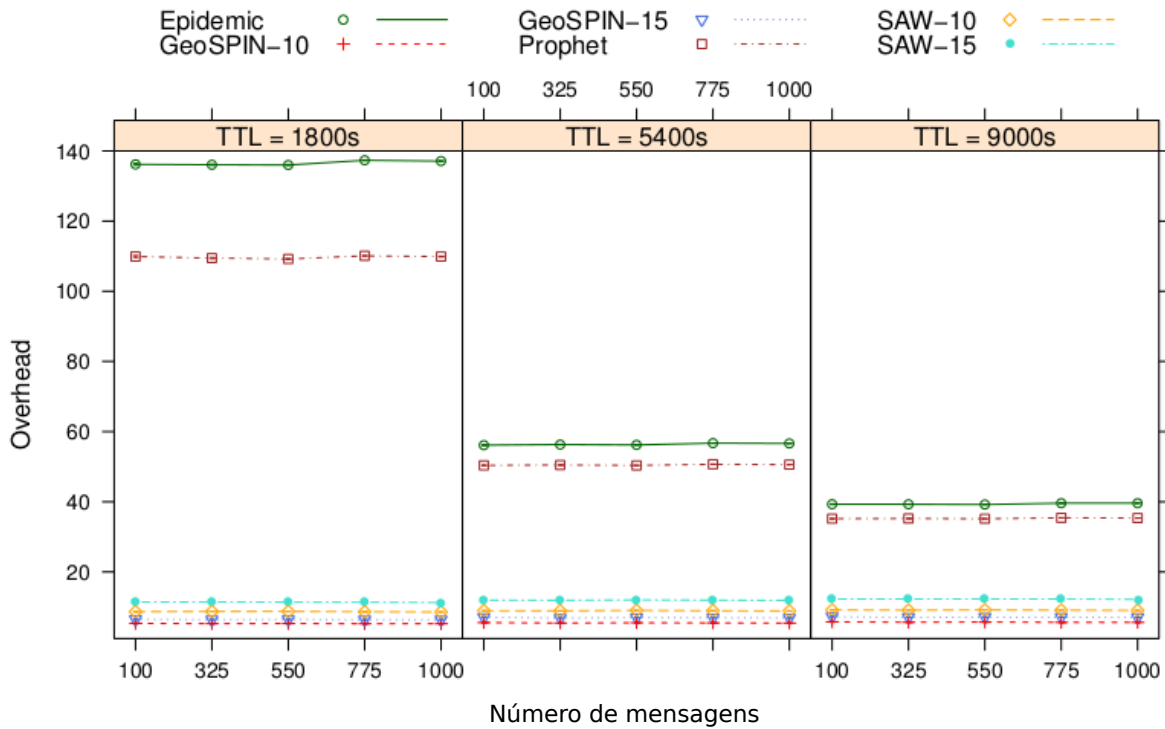


Figura 3.9: *Overhead* em função do número de mensagens e TTL.

número de mensagens entregues ao destino. O GeoSPIN com ambas as configurações atingiu bons desempenhos em termos de *overhead* (figura 3.9). No algoritmo do GeoSPIN, quando dois veículos se encontram, eles realizam uma decisão distribuída sobre que veículo deve carregar a mensagem baseado em sua rotina, isso reduz o número de transmissões na rede. O Epidêmico e Prophet reduzem o *overhead* quando o TTL é maior porque os veículos possuem um tempo maior para entregar a mensagem. No caso do GeoSPIN e do SAW, o *overhead* se mantém praticamente o mesmo quando o TTL varia. Como o número de cópias na rede é restrito, o número de transmissões é também limitado. No entanto, o GeoSPIN apresenta resultado ainda menor que o SAW devido as condições impostas na linha 9 do algoritmo 2.

A **latência** refere-se ao período entre o momento de geração da mensagem pelo veículo fonte até a entrega na região destino. Como mostrado na figura 3.10, o Epidêmico e Prophet apresentam os menores atrasos entre os protocolos. No entanto, eles possuem um alto número de mensagens duplicadas na rede. O GeoSPIN obtém um atraso maior a medida que o TTL aumenta. No entanto, ele apresenta uma pequena diferença de atraso quando o valor do TTL é pequeno (i.e., para o limite inferior do protocolo Epidêmico).

Considerando os resultados obtidos e a configuração de simulação, o protocolo

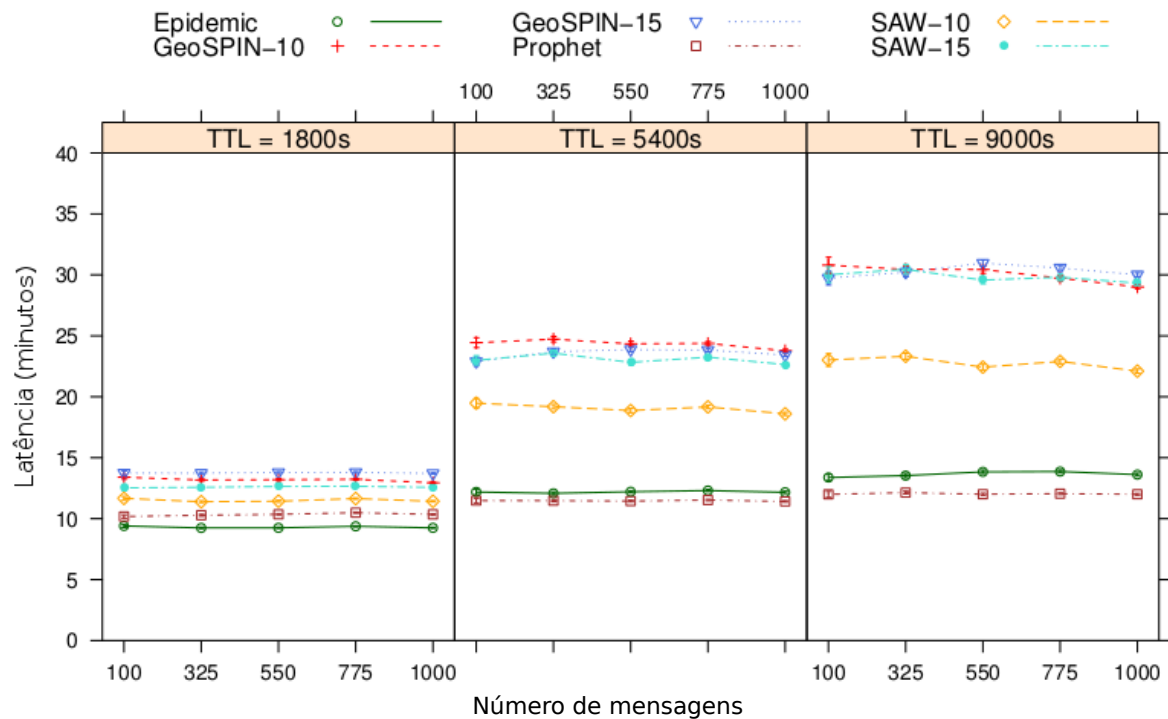


Figura 3.10: Latência em função do número de mensagens e TTL.

GeoSPIN obteve um bom desempenho em relação aos demais protocolos comparados. Em resumo, GeoSPIN apresentou uma taxa de entrega aceitável e um pequeno número de mensagens descartadas, produzindo um pequeno valor de *overhead* e latência em determinadas configurações.

3.5 Considerações Finais

Neste capítulo, é apresentado um esquema de roteamento para redes veiculares esparsas, que considera a informação espacial do histórico de trajetórias de movimentações dos veículos. É mostrado que as trajetórias dos veículos podem ser combinadas com a técnica armazena-e-encaminha para melhorar a taxa de entrega de dados em VANETs esparsas com um cenário de *Geocast*. Além disso, demonstrou-se que a adoção de um método de agrupamento (*clustering*) para detectar padrão de mobilidade nas trajetórias individuais pode ser eficiente para auxiliar na tomada de decisão durante o processo de encaminhamento de mensagens. As simulações demonstram que o GeoSPIN é capaz de alcançar um bom desempenho em termos de taxa de entrega, mantendo valores aceitáveis de *overhead* e de atraso em relação à outros três protocolos bem conhecidos na literatura.

No contexto de trajetórias, o GeoSPIN utiliza somente a informação espacial e, a partir disso, explora o contato oportunístico [Conti et al., 2010] entre os nós da rede. Nesse caso, fatores como a mobilidade dos veículos e o número de veículos podem favorecer o desempenho do GeoSPIN. No entanto, deve-se considerar que a mobilidade é baseada em padrões regulares de comportamento.

Como direcionamento futuro, objetiva-se analisar o impacto de tráfego real no GeoSPIN e realizar simulações com variações de escalabilidade na rede. Além disso, visando a cobertura de outros cenários de VANETs, a exploração de informações espaço-temporais de trajetórias são apresentadas no próximo capítulo. Nesse sentido, o intuito é explorar a informação espaço-temporal para identificar contatos previsíveis entre os nós de acordo como o histórico de movimentação.

Capítulo 4

Roteamento Baseado em Informação Espaço-Temporal

4.1 Introdução

Considerando o que foi exposto sobre o uso da informação espacial, este capítulo aborda uma ampliação na exploração de dados de trajetórias para um domínio espaço-temporal. Como comentado anteriormente, é pertinente analisar que tipo de informação é relevante extrair dos dados, pois determinadas aplicações requerem um maior grau de conhecimento sobre o comportamento de objetos (i.e., pessoas, animais).

Em muitas aplicações reais, a análise de dados espaço-temporais se refere a identificar padrões periódicos de movimentações, pois muitos objetos seguem as mesmas rotas em intervalos regulares de tempo. Por exemplo, muitas pessoas vão ao trabalho todos os dias seguindo rotas similares nos mesmos horários, o transporte público (e.g., ônibus, metrô, trens) é regido por horários e os animais migram regularmente para se reproduzir ou procurar por condições climáticas mais favoráveis para sobrevivência. Esses padrões de movimentações fornecem informações para o entendimento sobre o comportamento periódico (e.g., diário, semanal, mensal) desses objetos.

Padrões periódicos são úteis tanto para representação de movimentos passados [Cao et al., 2007] como para predição de movimentos futuros [Jeung et al., 2008a]. No entanto, é importante considerar um determinado grau de incerteza na determinação de padrões periódicos de movimentações, pois apesar de ser razoável esperar que um objeto repita um comportamento regular a cada período de tempo, as movimentações podem diferir ligeiramente de um período para o outro, mas ainda assim, resultam em um padrão [Jeung et al., 2011].

No contexto de redes veiculares, os padrões periódicos de movimentações de cada veículo podem ser considerados uma importante informação para auxiliar no roteamento, principalmente, quando a rede é esparsa. Esta rede, um tipo de DTN (*Delay Tolerant Network*), é caracterizada por atrasos longos e variáveis, dominância de desconexões, quebras frequentes de conexões e conectividade intermitente [Oliveira & Duarte, 2007]. De acordo com os autores de [Jain et al., 2004], os protocolos de roteamento que priorizam o conhecimento das condições da rede (e.g., topologia, mobilidade dos nós) alcançam melhor desempenho. Este conhecimento pode ser obtido também por meio da identificação de padrões.

Em DTNs, o conceito de encontro deve ser considerado, pois em combinação com estratégias de armazenamento persistente e ao método de armazena-e-encaminha, ele dá suporte ao roteamento de mensagens entre os nós. A arquitetura DTN define vários tipos de encontros, também chamado de contatos, são eles: programados, oportunistas e previsíveis [Oliveira et al., 2007]. No encontro programado, os nós envolvidos possuem uma agenda preestabelecida com os momentos para ocorrer a troca de informação. No oportunista, o encontro entre os nós não é previamente programado, sendo totalmente ao acaso. O previsível é baseado em informações históricas de encontros entre os nós. Ao contrário dos encontros programados, os encontros previsíveis possuem uma incerteza sobre ocorrência do encontro.

O estudo realizado neste trabalho baseia-se nos encontros que podem ser caracterizados como previsíveis, levando em consideração os padrões periódicos de movimentações para o roteamento de mensagens em uma rede veicular esparsa. Para tanto, são aplicados algoritmos para extrair informações espaço-temporais das movimentações diárias dos veículos. Primeiramente, a partir de um conjunto de trajetórias de cada veículo, identificam-se as trajetórias representativas (similar ao que foi apresentado no capítulo 3). Em seguida, de acordo com as trajetórias representativas de cada veículo, são identificadas as regiões de encontros entre cada par de veículos. Além de descobrir as regiões de encontros entre os veículos, é necessário observar em quais instantes esses encontros ocorrem. Feito isso, é possível modelar a rede com contatos previsíveis entre os veículos e, assim, estabelecer rotas para o encaminhamento de mensagens.

Um encontro é definido quando dois veículos estabelecem um histórico de contatos de comunicação em uma região de área A , pelo menos em k marcações consecutivas no tempo [Benkert et al., 2008]. A figura 4.1 apresenta o conceito de encontro utilizado neste trabalho. Os veículos O_1 e O_2 possuem ambos uma trajetória representativa e, observando-se o histórico de deslocamento, constrói-se uma região de encontro em um determinado ponto de suas trajetórias, devido a interseção tanto espacial como temporal durante o deslocamento. Então, o encontro é formado pela região de interesse

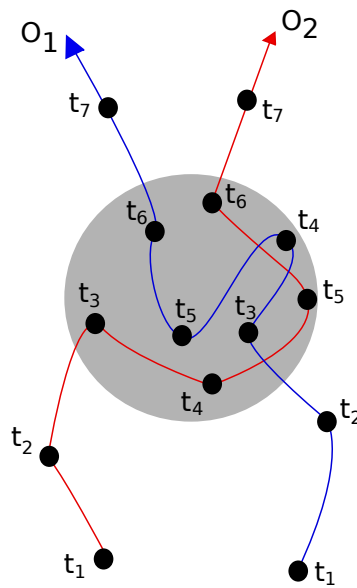


Figura 4.1: Representação de uma região de encontro identificada pela mobilidade entre dois veículos.

(RoI) definida espacialmente, durante os instantes de tempo t_3, \dots, t_6 .

Nesse sentido, este capítulo aborda a exploração da informação espaço-temporal para auxiliar no processo de encaminhamento de mensagens em uma VANET. Assim, é apresentado um esquema de roteamento que utiliza a informação espaço-temporal para detectar regiões de encontros entre veículos e, assim, baseado no histórico de deslocamento, selecionar os nós candidatos para entregarem a mensagem ao destino. Para tanto, a seção 4.2 apresenta métodos de análise da informação espaço-temporal empregados neste trabalho para favorecer o entendimento do esquema proposto na seção 4.3. Em seguida, a seção 4.4 apresenta uma análise de desempenho do esquema proposto e, por fim, são descritas as considerações finais na seção 4.5.

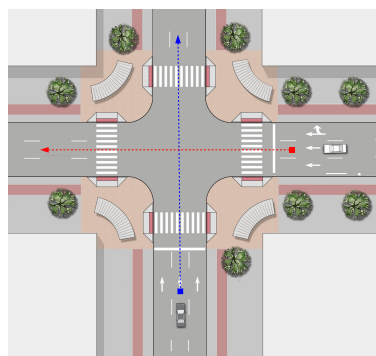
4.2 Explorando a Informação Espaço-Temporal

Apesar de as pessoas possuírem dinâmicas de mobilidade distintas, pois moram ou trabalham em locais diferentes, é possível identificar correlações de trajetórias nas suas rotinas diárias. Portanto, por meio dessas correlações entre as rotinas diárias de deslocamento se pode extrair informações para auxiliar no roteamento de mensagens em redes veiculares. Considerando o conceito de correlação como uma relação mútua entre as rotinas de dois veículos, esta seção apresenta algoritmos de correlação para a extração de similaridades tanto espaciais como temporais. Primeiramente, é descrito um algoritmo para a determinação da correlação espacial. Em seguida, a correlação

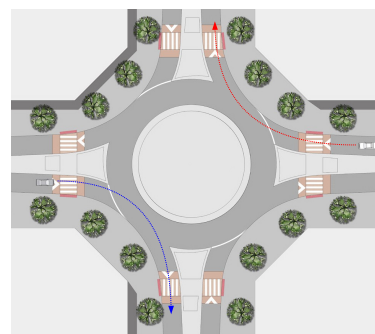
temporal é apresentada com base no algoritmo de *Parzen-Window*. A composição dessas correlações fornece informações relevantes sobre a mobilidade dos nós da rede e, conseqüentemente, dos momentos e locais de encontros entre eles.

4.2.1 Correlação Espacial

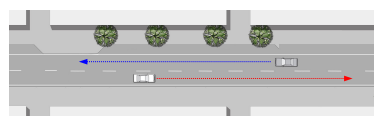
Levando-se em conta a ideia de analisar as rotinas diárias de deslocamento, a fim de identificar relações de proximidade espacial entre os veículos, é proposto um algoritmo que visa identificar pontos de proximidades entre as trajetórias representativas dos veículos. O algoritmo visa dar suporte à identificação de relações espaciais nas três situações definidas por Braga et al. [2011], descritas a seguir. Na primeira situação, ilustrada na figura 4.2(a), dois veículos tem um ponto de encontro exatamente no cruzamento entre duas rodovias. A figura 4.2(b) ilustra uma situação que não há uma interseção entre as trajetórias representativas dos veículos. A figura 4.2(c) representa uma situação entre as trajetórias representativas de dois veículos que compartilham a mesma rodovia.



(a) Situação de cruzamento.



(b) Situação de proximidade.



(c) Situação de compartilhamento de rodovia.

Figura 4.2: Principais situações para considerar uma aproximação entre trajetórias representativas de dois veículos.

O principal fundamento do algoritmo é o conceito de *buffering*. *Buffering* é um método para definir uma área de interesse, chamada de *buffer*, em torno de uma localização baseada na distância ou no tempo. A figura 4.8(a) ilustra um exemplo de *buffer* espacial. Neste caso, a partir de um ponto C estabelece-se uma circunferência de raio

R . O *buffer* temporal é representado pela região que tem alcance ao ponto C em um tempo máximo M , como ilustrado na figura 4.3(b).

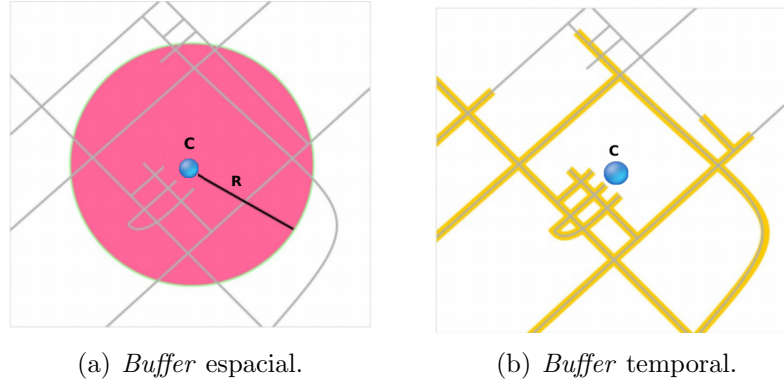


Figura 4.3: Ilustração de *buffer* espacial e temporal.

A versão implementada nesta dissertação segue uma abordagem baseada no *buffer* espacial, no entanto, tal algoritmo pode ser adaptado para utilização de um *buffer* temporal. Dessa forma, considerando as situações descritas no início desta seção e o método de *buffering*, o seguinte algoritmo identifica as regiões de proximidade entre trajetórias de veículos, denominadas regiões de interesse (RoI).

Algoritmo 3: Correlação Espacial - Identificando regiões de proximidade entre trajetórias de veículos

Entrada: Trajetórias representativas (T) de dois veículos e um raio (R) para o *buffer* espacial

Saída: Um conjunto de regiões de interesse (RoI) que representam áreas de proximidade

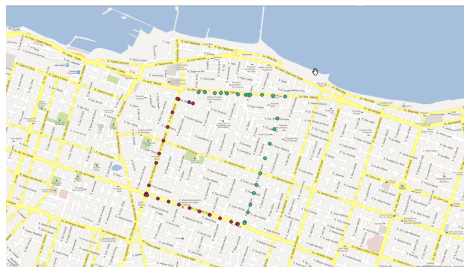
```

1 procedimento CorrelacaoEspacial( $T$ ,  $R$ )
2    $RoIs \leftarrow []$ 
3   para cada  $traj$  in  $T$  faça
4     para cada  $p$  in  $traj$  faça
5        $p.buffer(R)$ 
6     fim para cada
7     realizarUniaoEntreBuffers( $traj$ )
8   fim para cada
9    $RoIs \leftarrow identificarSobreposicaoEntreTrajetorias(T)$ 
10  return  $RoIs$ 
11 fim procedimento

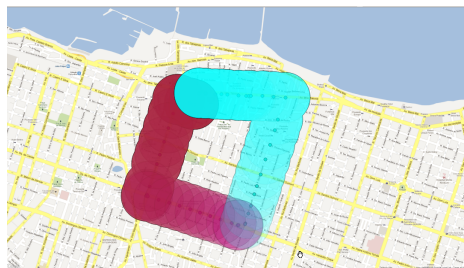
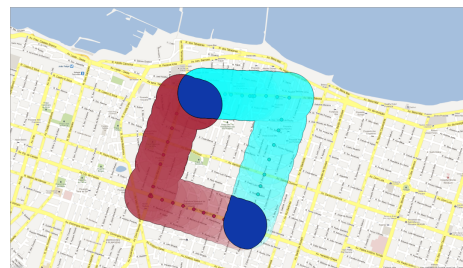
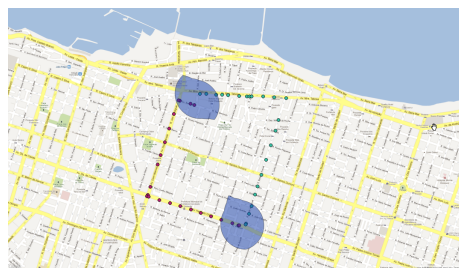
```

O algoritmo 3, uma contribuição desta dissertação, apresenta a sequência de tarefas necessárias para obtenção das regiões de interesse baseada na proximidade entre trajetórias representativas. Para auxiliar no entendimento do algoritmo, a figura 4.4 ilustra visualmente todas as tarefas descritas a seguir. Inicialmente, dado um conjunto de trajetórias e um raio de *buffer* predefinido, aplica-se um procedimento de *buffer*

para cada ponto das trajetórias (algoritmo 3, Linhas 4 - 6), como pode ser visualizado na figura 4.4(b). Em seguida, realiza-se a união entre os *buffers* espaciais de cada trajetória (algoritmo 3, Linhas 7), como pode ser visualizado na figura 4.4(c). Finalmente, identificam-se as regiões de interesse com base na sobreposição espacial existente entre as trajetórias após a aplicação do método de *buffering* (algoritmo 3, Linhas 9), como exibido na figura 4.4(d). A figura 4.4(e) mostra a representação espacial das regiões de interesse armazenadas no vetor *RoIs*.



(a) Trajetórias de dois veículos.

(b) Aplicação do método de *buffering* em cada ponto.(c) União entre os *buffers* da mesma trajetória.(d) Interseção entre as trajetórias após o *buffering*.

(e) Regiões de interesse.

Figura 4.4: Ilustração de aplicação do algoritmo de correlação espacial para detecção de regiões de interesse entre duas trajetórias representativas.

4.2.2 Correlação Temporal

Além de descobrir as regiões de interesse entre trajetórias representativas, é importante o conhecimento temporal dos momentos do dia que os veículos passam por cada região.

Com esse objetivo foi adotada a técnica *Parzen-window* [Parzen, 1962]. Para este trabalho, a técnica Parzen-window permite estimar a densidade de ocorrência temporal de cada veículo passar pelas regiões de interesse. Após isso, aplica-se um algoritmo baseado nessa densidade a fim de realizar uma correlação temporal entre os momentos em que os veículos passam pelas regiões de interesse descobertas usando o algoritmo 3.

Por definição, a Parzen-window é uma técnica não paramétrica para estimar a função densidade de probabilidade de uma variável aleatória, considerando uma interpolação dos dados [Wand & Jones, 1995]. Considerando que uma função de densidade de probabilidade (PDF - *Probability Density Function*) é uma ferramenta estatística utilizada para representar o comportamento de uma variável aleatória, pode-se utilizá-la em uma abordagem de mineração de dados para investigar as propriedades em um conjunto de dados. Dessa forma, assumindo um conjunto de n amostras de dados x_1, x_2, \dots, x_n , a Parzen-window estima a função densidade de probabilidade $p(x)$ se baseando inteiramente nas amostras. Assim, a PDF é estimada pela sobreposição de uma função *kernel* para cada amostra x_i . A $p(x)$ é dada por:

$$p(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n^{dim}} K\left(\frac{x - x_i}{h_n}\right), \quad (4.1)$$

onde $K()$ é a função *kernel*, dim é o espaço dimensional e h_n é o comprimento da janela. Com base na equação 4.1, é possível obter o valor de $p(x)$ para qualquer amostra x . Para este trabalho, uma função *kernel* Gaussiana foi adotada, de forma que a PDF $p(x)$ torna-se:

$$p(x) = \begin{cases} \frac{1}{n} \sum_{i=1}^n \frac{1}{(h_n \sqrt{2\pi})^{dim}} e^{-\frac{1}{2} \left(\frac{x - x_k}{h_n}\right)^2} & , \text{ se } t_i \leq x \leq t_f \\ 0 & , \text{ caso contrário} \end{cases} \quad (4.2)$$

sendo que t_i é o instante de tempo inicial e t_f é o instante de tempo final de um intervalo para cada região de interesse. O valor de $dim = 1$, pois é analisado um ponto em relação aos outros pontos em um dado intervalo.

O parâmetro h_n , também chamando de *smoothing parameter* ou *bandwidth*, determina o valor da janela de particionamento dos dados. Este parâmetro afeta tanto a amplitude como a largura da função, isso pode ser observado na figura 4.5. Valores para h_n são definidos empiricamente como $h_n = \{0.10, 0.25, 0.50, 0.75\}$ sobre um conjunto de dados de um veículo 9 que passa entre as 6 e 12 horas por uma determinada região de interesse A obtida pelo algoritmo de correlação espacial. Os registros temporais de cada amostra podem ser visualizados no eixo horizontal (abscissa). Para essa região de interesse se pode concluir que este veículo 9 passa, na maioria das vezes, em torno das

8 horas, isso é definido pela alta concentração de amostras por volta desse horário.

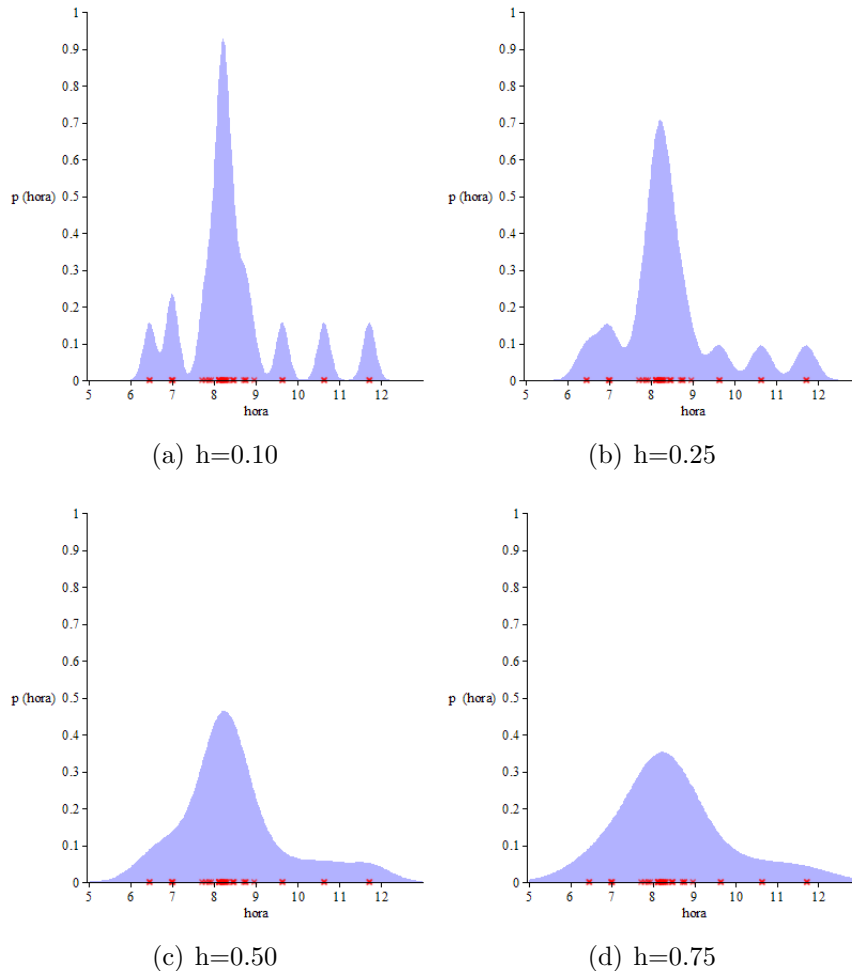


Figura 4.5: Resultado da aplicação do algoritmo de *Parzen window* para amostras do veículo 9 que passa pela região de interesse A com diferentes valores de h_n .

Para evitar a determinação do valor de h_n empiricamente, de acordo com [Silverman, 1986] e [Martinez & Martinez, 2001], quando o *Kernel* Gaussiano é utilizado, o valor ótimo de h_n é definido por:

$$h_n = \left(\frac{4\sigma^5}{3n} \right)^{\frac{1}{5}} \quad (4.3)$$

em que n é o número de amostras no intervalo de tempo e σ é o desvio padrão das amostras.

Enquanto que a figura 4.5 apresenta a modelagem do comportamento do veículo 9 para uma dada RoI A , a figura 4.6 mostra a correlação entre o veículo 9 e outro veículo, o 37, na RoI A . Dessa forma, é possível estimar quais as chances dos dois veículos se encontram na RoI A . Para tanto, visando quantificar a interseção entre as curvas, foram utilizados os valores de h_n das amostras de cada veículo e a frequência

relativa de cada veículo está na RoI A de acordo com o algoritmo 4.

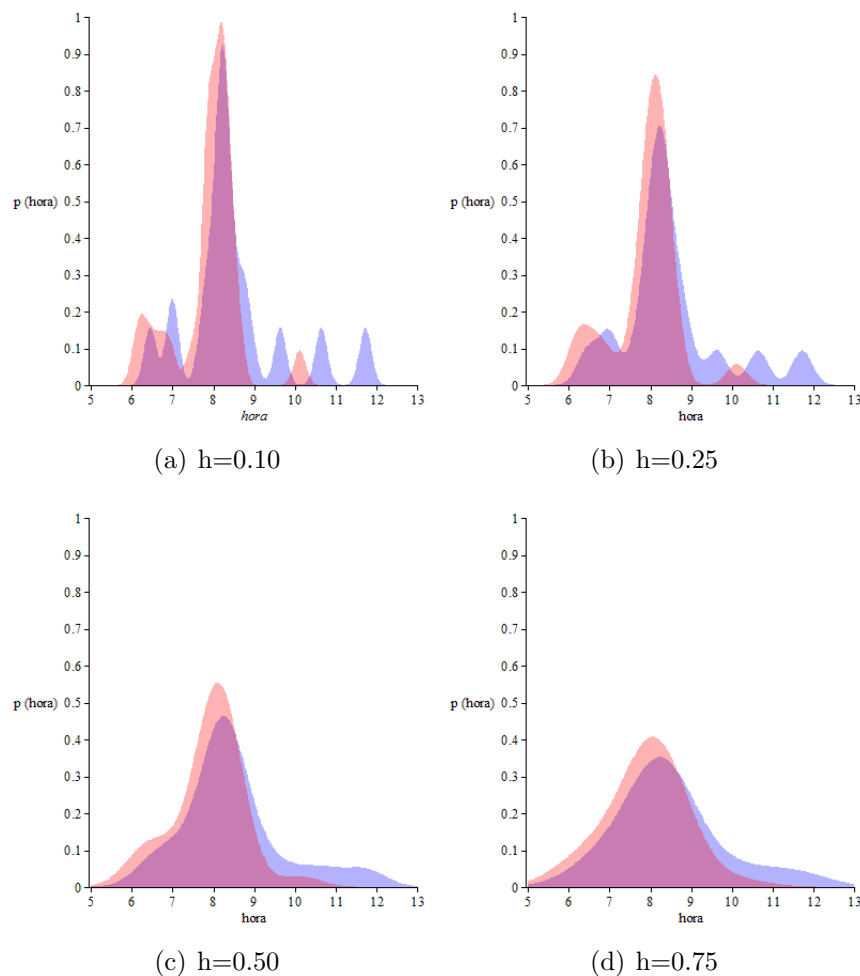


Figura 4.6: Resultado da aplicação do algoritmo de *Parzen window* para amostras do veículo 9 e 37 que passa pela região de interesse A com diferentes valores de h_n .

O algoritmo de correlação temporal (algoritmo 4) inicia calculando o valor de h_n para as trajetórias representativas de cada veículo. Então, é obtido o valor de h_n menor entre os dois encontrados, chamado de $h_n Best$ (algoritmo 4, Linha 2). O fato de ser menor é devido a busca pela precisão do intervalo de ocorrência, como mostrado nas figuras 4.5 e 4.6. Em seguida, as amostras de cada trajetória são ordenadas temporalmente e particionadas de acordo com o tamanho da janela (algoritmo 4, Linha 3). Depois disso, calcula-se a frequência relativa de ocorrência de amostras em cada janela para cada trajetória (algoritmo 4, Linha 4). Em seguida, computa-se a probabilidade independente¹, entre os valores obtidos no passo anterior, para cada janela (algoritmo 4, Linha 5). Finalmente, retorna o valor de probabilidade maior e o intervalo de ocorrência dessa probabilidade (algoritmo 4, Linha 6).

¹ $P(A \cap B) = P(A)P(B)$, onde A e B são os eventos dos veículos passarem por uma dada região.

Algoritmo 4: Correlação Temporal - algoritmo para efetuar a correlação temporal entre duas trajetórias em uma dada região de interesse.

Entrada: Trajetórias representativas (T) de dois veículos e uma Região de Interesse

Saída: A probabilidade dos dois veículos se encontrarem na região de interesse dada e o intervalo temporal identificado.

```

1 procedimento CorrelacaoTemporal( $T, A$ )
2   identifica quando cada veículo passa por A
    $h_nBest \leftarrow \text{calculaTamanoDaJanela}(T)$ 
3   particiona trajetórias em janelas de tamanho  $h_nBest$ 
4   calcula a frequência relativa por janela em cada trajetória
5   computa a probabilidade de interseção por janela
6   return retorna a maior probabilidade e a janela de ocorrência
7 fim procedimento
```

4.3 Descrição do Algoritmo de Roteamento

Nesta seção, é proposta uma abordagem, denominada ROSTER (*R*outing based on *S*patio-*T*emporal *E*ncounter *R*ecord), para roteamento que explora informações espaço-temporais de deslocamento diário dos veículos para encaminhar mensagens de um veículo origem para um veículo destino. O algoritmo consiste na extração de padrões periódicos de trajetórias de cada veículo a partir de dados GPS coletados ao longo dos dias e, com base nessa informação, cria uma tabela de roteamento. Em seguida, aplica-se o encaminhamento das mensagens utilizando a tabela de roteamento. Os detalhes são descritos a seguir, após a exposição de algumas considerações preliminares.

4.3.1 Considerações Preliminares

A relação de encontro entre trajetórias dos veículos oferece oportunidades de encaminhamento de mensagens. No entanto, descobrir o momento exato do estabelecimento de cada encontro entre dois veículos é uma tarefa complexa. Contudo, aplicando os algoritmos de correlação espacial e temporal descritos nas seções anteriores, é possível estimar qual o intervalo de tempo dentro do qual um encontro poderá ocorrer. Dessa forma, busca-se tratar a incerteza do horário de estabelecimento dos encontros, mas a incerteza quanto a ocorrência dos mesmos ainda persiste. Oliveira [2008] trata essa incerteza de ocorrência dos encontros utilizando pontos de acesso. De forma que um veículo origem pode enviar a mensagem para o ponto de acesso e quando o veículo destino tiver contato com o ponto de acesso recebe a mensagem. No entanto, para a abordagem atual deste trabalho é inviável distribuir pontos de acesso para cada região

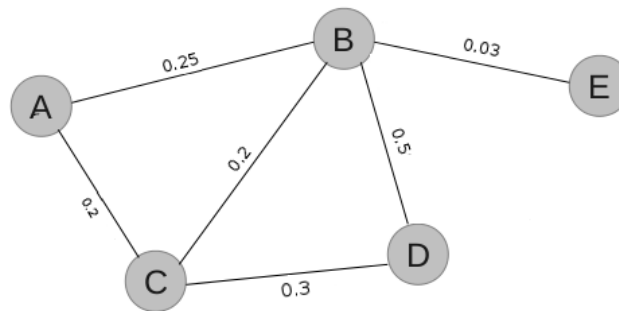


Figura 4.7: Um exemplo da representação de encontros em um grafo.

de interesse. Logo, cada intervalo de tempo associado a um encontro entre dois veículos é considerado que um encontro poderá ocorrer.

Para a representação dos encontros entre veículos é adotado um modelo de grafos evolutivos para o domínio no tempo proposto por [Oliveira, 2008]. Dessa forma, cada aresta possui um valor que representa o intervalo de tempo dentro do qual um encontro poderá ocorrer com uma certa probabilidade, como ilustrado na figura 4.7.

Para concepção do protocolo, foi considerado que as trajetórias não são decompostas, ou seja, um padrão é uma trajetória inteira entre duas localidades específicas. Além disso, é assumido que cada veículo possui um dispositivo capaz de capturar e armazenar o seu atual posicionamento geográfico e com capacidade computacional para executar os algoritmos descritos na subseção seguinte.

4.3.2 Criação da tabela de roteamento e Encaminhamento

O encaminhamento consiste em direcionar uma mensagem de um veículo origem para um veículo destino. A entrega de mensagem entre os veículos pode ser tanto de único salto, como por múltiplos saltos. Nesse último caso, veículos intermediários são utilizados como retransmissores para direcionar a mensagem ao destino. A escolha da rota de encaminhamento mais adequada se dar por algum conhecimento prévio definido em uma tabela de roteamento. A tabela de roteamento do ROSTER é criada com base nos encontros de trajetórias representativas dos veículos. Para tanto, inicialmente é necessário aplicar para cada veículo o processo de clusterização usando o OPTICS e obter trajetórias representativas para cada *cluster* individual. Em seguida, faz-se a correlação espacial e temporal entre cada par de veículos. Feito isso, é construído um grafo com os nós sendo os identificadores dos veículos e as arestas sendo rotuladas com os valores de intervalo de tempo e a frequência relativa estimada pela correlação

temporal.

Neste trabalho, o custo da rota entre dois nós da rede é calculado de acordo com o método definido por [Burgess et al., 2006]. Um veículo calcula localmente, sabendo o valor dos outros veículos, $c(i, i + 1, \dots, d)$, para cada rota possível para o destino d , até n saltos. O custo de um rota, definido na equação 4.4, usando os nós $i, i + 1, \dots, d$ é a soma das probabilidades de uma rota não ocorrer, estimada como sendo um menos a probabilidade de um encontro ocorrer.

$$c(i, i + 1, \dots, d) = \sum_{i=1}^{d-1} (1 - p_{x+1}^x) \quad (4.4)$$

O custo para um destino é o valor de menor custo entre todas as rotas. Como cenário de ilustração, seguindo a figura 4.7 pode-se calcular o custo da rota entre os veículos. Por exemplo, considerando que o objetivo é saber o custo da rota entre os veículos A e D, as seguintes computações são necessárias:

$$ABD = (1 - 0.25) + (1 - 0.5) = 1.25$$

$$ABCD = (1 - 0.25) + (1 - 0.2) + (1 - 0.3) = 2.25$$

$$ACBD = (1 - 0.2) + (1 - 0.2) + (1 - 0.5) = 2.1$$

$$ACD = (1 - 0.2) + (1 - 0.3) = 1.5$$

Assim, a rota entre os veículos é a ABD, pois essa possui um custo mínimo. Dessa forma, o veículo A computa todos os valores de sua tabela de roteamento. Em termos de implementação, aplica-se um algoritmo para busca de menor caminho entre dois veículos (i.e., origem e destino).

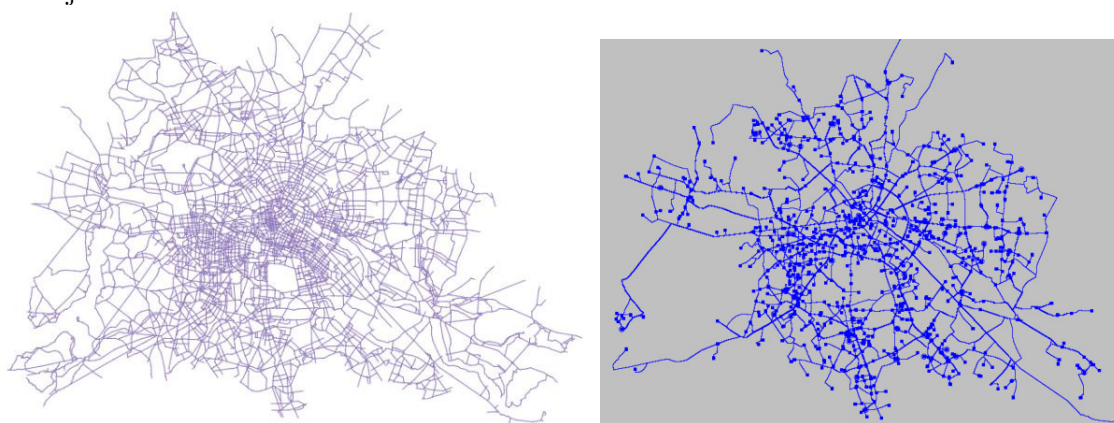
Feito isso, a métrica de roteamento e, conseqüentemente, de seleção de rota para uma mensagem é o custo de menor rota entre origem e destino. Baseando-se nisso, os veículos são capazes de decidir qual a melhor rota para encaminhar uma mensagem. Logo, uma dada mensagem é encaminhada por múltiplos saltos até alcançar o seu destino.

4.4 Análise de Desempenho

4.4.1 Descrição do Ambiente e Cenário de Simulação

As simulações foram realizadas com o uso de um *framework* de simulação de redes veiculares (Veins)² utilizado em conjunto com o simulador de eventos discretos OMNeT++ e um simulador de mobilidade e tráfego chamado SUMO [Sommer et al., 2011]. O Veins provê um modelo realístico para 802.11p DSRC e para as camadas PHY e MAC. Os parâmetros PHY e MAC são definidos de acordo com o padrão 802.11p [Jiang & Delgrossi, 2008].

Para a configuração do cenário de simulação, foi adotado um conjunto de dados de mobilidade de veículos chamado *BerlinMOD*³ [Axhausen et al., 2013]. Esse conjunto de dados representa o comportamento diário de motoristas na cidade Berlin, na Alemanha, possuindo trajetórias de ida e volta entre suas casas e o local de trabalho, bem como trajetórias adicionais realizadas no tempo de lazer. O conjunto original possui registros de 2000 veículos durante um período de 28 dias. No entanto, devido a necessidade de redução do tempo de simulação, 50 veículos foram escolhidos aleatoriamente para compor o conjunto de dados utilizados nos experimentos deste trabalho. Para mostrar que esse subconjunto de veículos mantém as características originais da base, é apresentada uma caracterização da base de dados que verifica os mesmos atributos do conjunto de dados de 2000 veículos.



(a) Mapa de estradas da cidade de Berlin, na Alemanha.

(b) Representação espacial formada a partir das movimentações dos veículos na cidade de Berlin.

Figura 4.8: Mapa de estradas e movimentações dos veículos na cidade de Berlin.

²Vehicles in Network Simulation - <http://veins.car2x.org/>

³<http://dna.fernuni-hagen.de/secondo/BerlinMOD/BerlinMOD.html>

4.4.1.1 Caracterização Espacial da Base de Dados

A figura 4.8 (a) mostra o mapa de estradas da cidade de Berlin, na Alemanha. A figura 4.8 (b) apresenta todas as movimentações dos veículos durante um período de 28 dias. Comparando as duas figuras, percebe-se que algumas regiões da cidade não são alcançadas durante as movimentações dos veículos. Além disso, é notável a intensidade de movimentações na região central do mapa.

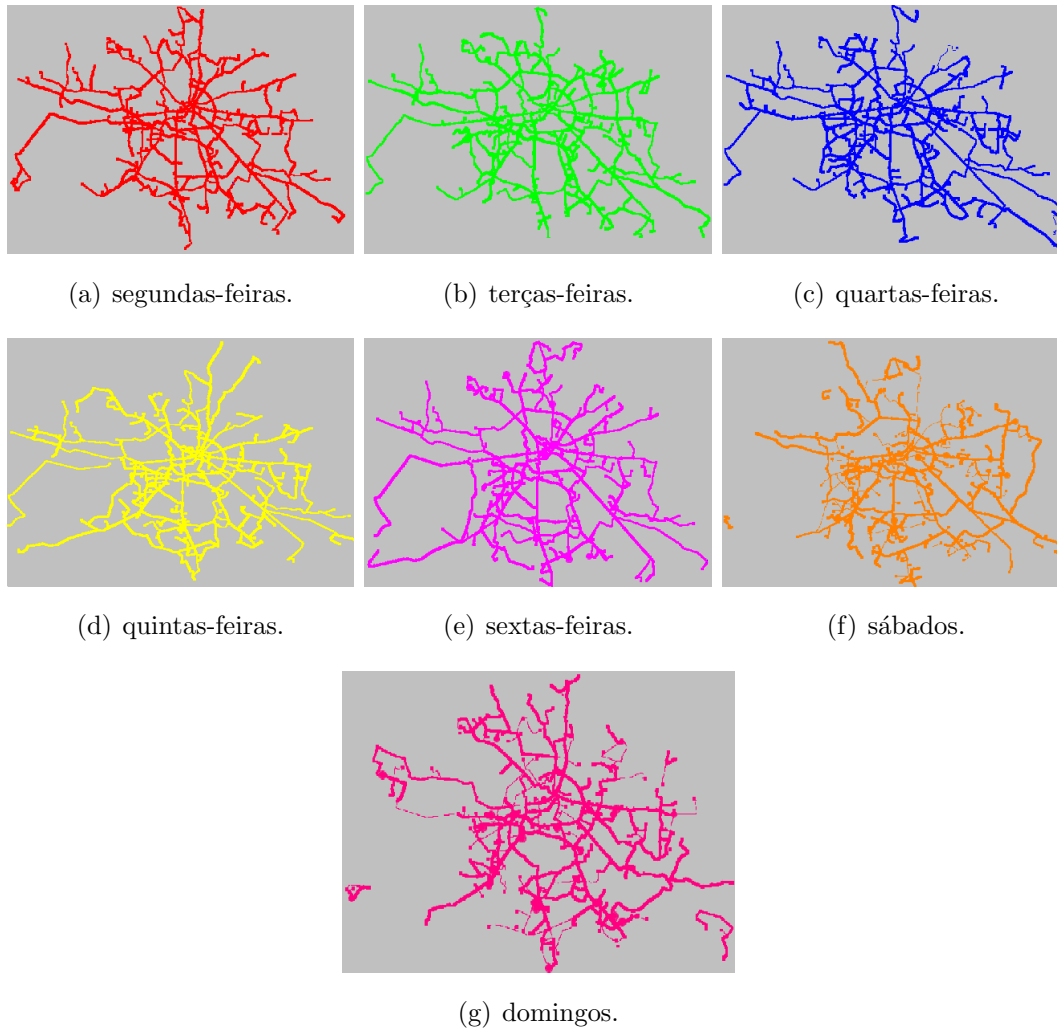


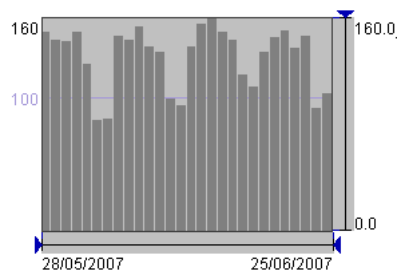
Figura 4.9: Trajetórias agregadas dos veículos por dia da semana.

A figura 4.9 apresenta uma análise agregada das movimentações dos veículos. De forma que para cada dia da semana, é exposto uma regularidade espacial de deslocamento. Pode-se perceber que todos os dias, com exceção do sábado e domingo, apresentam uma similaridade espacial de deslocamento. No entanto, no final de semana, existe uma distorção em relação ao comportamento dos demais dias, visto que são notadas certas aberturas, bem como uma menor intensidade de movimentações. Essa análise do comportamento espacial é interessante para verificar se os dados de

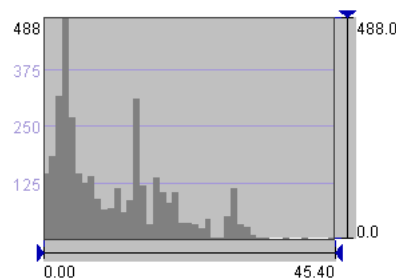
trabalho realmente representam uma regularidade espacial de movimentações ao longo dos dias.

4.4.1.2 Caracterização Temporal da Base de Dados

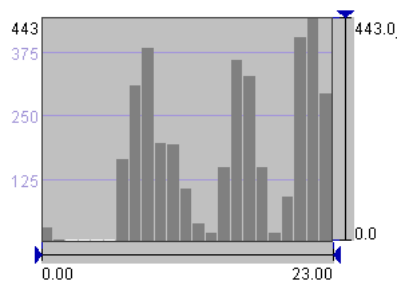
Além da regularidade espacial, existe a premissa que esses dados possuam uma regularidade temporal de deslocamento. Para isso, a seguir, é mostrada uma caracterização temporal agregada dos veículos. A figura 4.10 (a) mostra a quantidade de trajetórias por dia de registro. Nota-se que como comentado na caracterização espacial, existe um número menor de trajetórias nos finais de semana. A figura 4.10 (b) apresenta a quantidade de trajetórias em relação a duração das mesmas. Grande parte das trajetórias possuem duração máxima de 15 minutos. A figura 4.10 (c) mostra a quantidade de trajetórias iniciadas por hora. Nota-se nos três turnos que existem horários no qual o número de trajetórias iniciadas é maior, são eles: 8 horas, 16 horas e 22 horas. O período entre 0 e 5 horas possui um número de trajetórias iniciadas próximo a zero. A figura 4.10 (d) apresenta a quantidade de trajetórias por dia da semana. Nesta figura, sábado e domingo possuem um número menor de trajetórias iniciadas em relação aos demais dias.



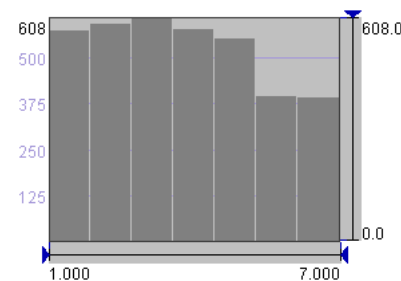
(a) Relação entre a quantidade de trajetórias e o dia de registro.



(b) Relação entre a quantidade de trajetórias por duração.



(c) Relação entre a quantidade de trajetórias iniciadas a cada hora



(d) Relação entre a quantidade de trajetórias por dia da semana.

Figura 4.10: Gráficos da relação entre quantidade de trajetórias por marcações temporais.

A caracterização realizada nesta seção é dada pelo agregado de todos os veículos.

No entanto, o ROSTER utiliza informações espaço-temporais individuais dos veículos. Por isso, que algoritmos de clusterização e correlações são aplicados separadamente para cada veículo.

4.4.2 Resultados

Para obtenção dos resultados descritos a seguir foram realizadas simulações com duração de 4 dias, de forma que cada veículo inicialmente gera uma mensagem para um destino aleatório na primeira hora do primeiro dia de simulação. No final do quarto dia é feita a análise dos resultados. Para verificar o impacto do número de veículos na rede, foram criados cenários com 20, 35 e 50 veículos. Ademais, para todos os cenários, cada configuração da simulação foi repetida 30 vezes com diferentes sementes de aleatoriedade. Os resultados apresentados possuem um intervalo de confiança de 95%.

Pelas características dos cenários de simulação, é adequado uma comparação entre protocolos de roteamento tipicamente utilizados na literatura como *baseline* para redes esparsas e tolerantes a desconexões, tais como Epidêmico e Spray-And-Wait. Uma visão geral sobre o funcionamento desses protocolos pode ser lida no capítulo 3. A avaliação deste trabalho foi feita em termos de taxa de entrega, *overhead*, atraso de entrega das mensagens (latência) e o número de mensagens transmitidas.

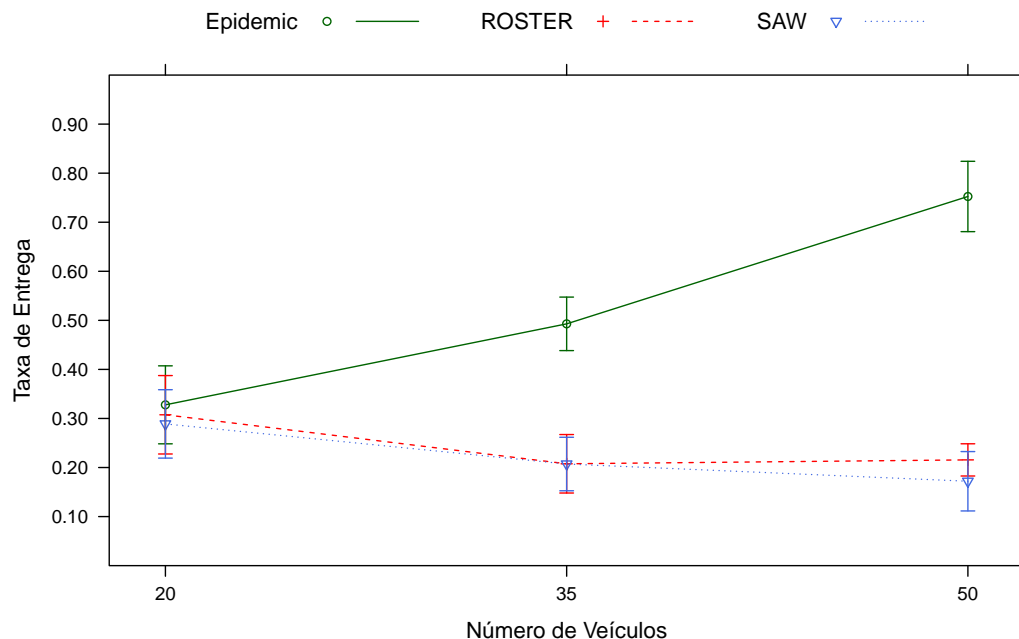


Figura 4.11: Taxa de entrega em função do número de veículos.

A **taxa de entrega** é dada pela relação entre o número de mensagens recebidas pelos veículos destino e o número total de mensagens enviadas pelos veículos fonte. A figura 4.11 mostra o comportamento dos protocolos quanto a taxa de entrega. Como

esperado, o protocolo Epidêmico obtém a maior taxa de entrega para os diferentes cenários. A taxa de entrega do Epidêmico aumenta a medida que o número de veículos também aumenta. Apesar de o SAW e o ROSTER apresentarem uma taxa de entrega similar ao Epidêmico em um cenário com 20 veículos, eles diminuem a taxa de entrega para cenários com mais veículos. Para o SAW, o fato da rede ser muito esparsa e o número de replicações ($L = 10$) de uma mensagem ser limitado, as replicações acabam sendo realizadas para um número limitado de veículos que não possuem nenhum contato com o destino. Para o ROSTER, o fato da rede ser muito esparsa aliada aos casos que veículos podem originar mensagens para destino que sejam não alcançáveis por eles, prejudicam o desempenho do protocolo.



Figura 4.12: *Overhead* em função do número de veículos.

O **overhead** representa a relação entre o número de mensagens transmitidas e o número de mensagens entregues ao destino. O ROSTER para os três cenários apresenta valores baixos de *overhead* (figura 4.12). Isso mostra um comportamento interessante do ROSTER, que a maioria das mensagens que são transmitidas pela rede atinge o destino, visto que a mensagem sendo transmitida segue uma rota de encaminhamento e, baseado nos encontros previsíveis, as mensagens são entregues. Enquanto que o Epidêmico e SAW apresentam valores altos de *overhead*, pois além de serem protocolos de múltiplas cópias, grande parte das mensagens que são transmitidas pela rede não alcançam o destino.

O **atraso** refere-se ao período entre o momento de geração da mensagem pelo veículo fonte até a entrega para o veículo destino. Como pode ser observado na fi-

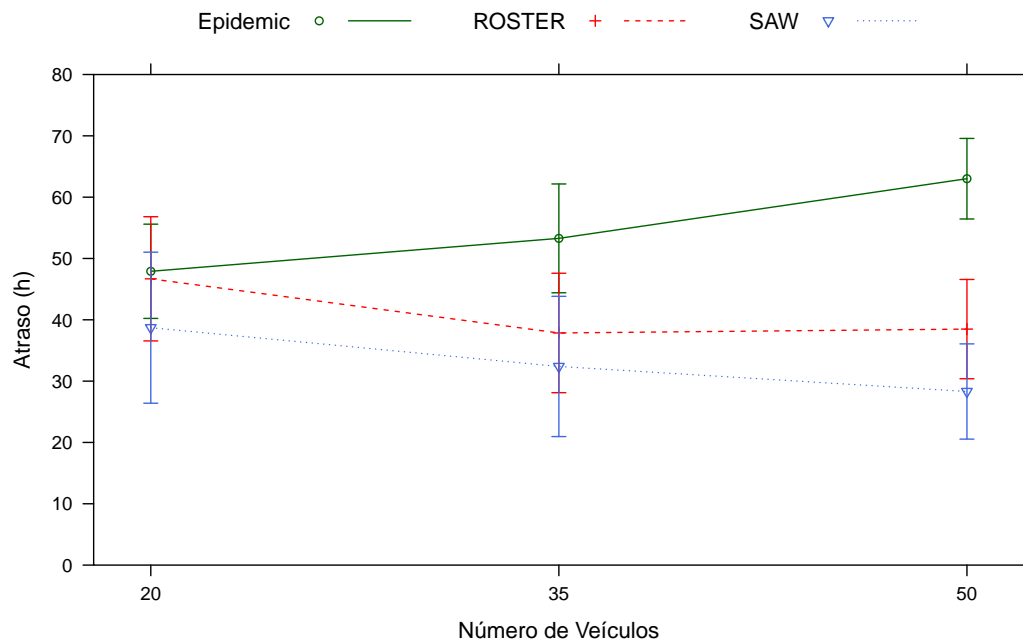


Figura 4.13: Atraso em função do número de veículos.

gura 4.13, todos os protocolos apresentam valores elevados de atrasos. Na realidade, o desempenho de cada protocolo é prejudicado pelas características dos cenários de simulação. Dado que um veículo recebeu uma mensagem para retransmissão, ele poderá esperar até o dia seguinte para encaminhá-la novamente, ocasionando atrasos altos na entrega da mensagem. Para contornar essa situação, fica como trabalho futuro aumentar o número de veículos e tornar a rede mais densa. Na figura 4.13, apesar de o SAW ter atrasos menores que os demais, vale ressaltar que dentre os protocolos, ele possui a menor taxa de entrega. Por outro lado, o Epidêmico apresenta valores elevados de atraso, pois algumas mensagens que são entregues ao destino gastam mais tempo para chegar ao destino. Esse comportamento também é válido para o ROSTER.

O **número de mensagens transmitidas** refere-se ao número de retransmissões que são contabilizadas durante as simulações. Como pode ser visualizado na figura 4.14, o ROSTER apresenta valores baixos de retransmissões. Isso é devido ao protocolo estimar a rota de encaminhamento da mensagem. Dessa forma, uma única cópia de cada mensagem é transmitida pela rede até alcançar o veículo destino. Enquanto o Epidêmico e o SAW são protocolos de múltiplas cópias, ocasionando um alto número de replicações de mensagens. O SAW apresenta valores menores que o Epidêmico, pois possui um parâmetro para limitar o número de cópias de uma mensagem.

Considerando os resultados obtidos e as configurações de simulação, o protocolo ROSTER obteve um comportamento razoável, de acordo com as avaliações realizadas.

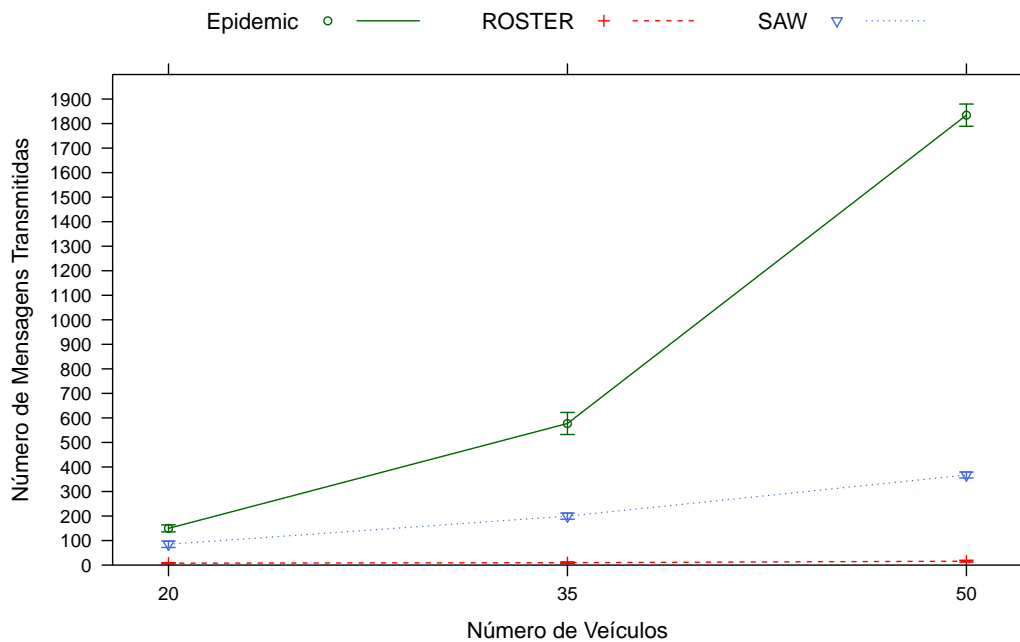


Figura 4.14: Número de mensagens transmitidas em função do número de veículos.

No entanto, apesar do protocolo apresentar bons valores quanto ao número de mensagens transmitidas e baixo *overhead*, refinamentos adicionais deverão ser realizados visando uma melhoria na taxa de entrega e diminuição do atraso de entrega das mensagens. Como uma pesquisa imediata, fica como trabalho futuro uma avaliação em cenários com um maior número de veículos que formem uma rede mais densa.

4.5 Considerações Finais

Neste capítulo, é apresentado um esquema de roteamento para redes veiculares esparsas, que considera a informação espaço-temporal do histórico de trajetórias de movimentações dos veículos. A proposta descrita neste capítulo explora a relação de encontro entre trajetórias dos veículos como informação auxiliar no processo de encaminhamento de mensagens. Para tanto, algoritmos de correlação tanto espacial quanto temporal foram criados para detectar os encontros entre cada par de veículos. Com base nessa informação de encontro, o encaminhamento das mensagens é dado por múltiplos saltos na rede de acordo com uma tabela de roteamento.

O ROSTER representa uma proposta de roteamento de mensagens entre um veículo origem e um veículo destino. Diferentemente dos trabalhos relacionados apresentados no capítulo 2, o ROSTER objetiva o encaminhamento de mensagens em um domínio V2V. Pesquisas adicionais são necessárias visando o aprimoramento do proto-

colo, bem como uma análise para diferentes cenários de simulação, tais como esparso e denso.

Além disso, este capítulo, apresentou outras contribuições. Algoritmos de correlação espacial e temporal foram criados visando a detecção de encontros entre objetos móveis (e.g., veículos), baseando-se em trajetórias representativas. Quanto a isso, como direcionamento futuro, objetiva-se analisar a aplicabilidade desses algoritmos em outros contextos, diferentes de redes veiculares.

Capítulo 5

Conclusão e Trabalhos Futuros

As redes veiculares ad hoc fornecem aos usuários uma variedade de serviços tais como informações de trânsito em tempo real, aplicações de entretenimento e notificações de segurança. A evolução dessas redes tem apresentado novos desafios de pesquisa que são distintos das MANETs convencionais. Esses desafios são motivados pelas suas características peculiares como alta mobilidade dos veículos, frequentes mudanças de topologia e rápidas variações de escalabilidade.

Um dos principais desafios das VANETs é o estabelecimento de rotas de comunicação entre veículos. Para tanto, diversos esforços têm sido feitos para a concepção de protocolos de roteamento de mensagens. Inicialmente, abordagens provenientes das MANETs foram aplicadas, mas devido às características descritas acima desse tipo de rede, apresentaram resultados insatisfatórios em vários cenários de VANETs. No entanto, estratégias baseadas em posição geográfica em combinação com técnicas de DTN (*Delay and Disruption Tolerant Networks*) se mostram promissoras ao serem aplicadas a protocolos de roteamento para VANETs, pois permitem a comunicação mesmo em situações de rede esparsa e conexão intermitente. Além disso, a incorporação de um sistema de navegação (e.g., GPS - *Global Positioning Systems*) nos veículos possibilita um avanço no roteamento baseado em informação geográfica, pois as trajetórias dos veículos passam a ser utilizadas no processo de encaminhamento. No entanto, apesar de existirem vários protocolos que exploram o uso de trajetórias de veículos, poucos focam no histórico de trajetórias para auxiliar no processo de encaminhamento de mensagens na rede.

Nesse sentido, esta dissertação apresenta esquemas de roteamento de mensagens que exploram o histórico de trajetórias de veículos. Para tanto, é realizada uma análise de trajetórias reais para identificar quais as rotinas diárias de cada veículo e, assim, obter informações de regularidade tanto espacial como temporal. A partir dessas in-

formações é possível detectar padrões de encontro e comportamentais dos nós da rede.

Considerando esses aspectos, a principal contribuição deste trabalho é a proposição e análise de esquemas de roteamento para redes veiculares que exploram o uso de informações espaciais e espaço-temporais de trajetórias de veículos para encaminhamento de mensagens. Além disso, são dadas contribuições adicionais relacionadas à teoria de sistemas de informações geográficas e técnicas de mineração de dados para aplicações que utilizam trajetórias de objetos móveis.

O primeiro esquema de roteamento proposto, chamado GeoSPIN, considera a informação espacial do histórico de trajetórias de movimentações dos veículos. É mostrado que as trajetórias dos veículos podem ser combinadas com a técnica armazena-e-encaminha para melhorar a taxa de entrega de dados em VANETs com um cenário de *Geocast*. Além disso, demonstrou-se que a adoção de um método de agrupamento (*clustering*) para detectar padrão de mobilidade nas trajetórias individuais pode ser eficiente para auxiliar na tomada de decisão durante o processo de encaminhamento de mensagens. Como este esquema utiliza apenas a informação espacial de trajetórias, um novo esquema que contempla também a informação temporal foi proposto.

O segundo esquema de roteamento proposto, chamado ROSTER, considera tanto a informação espacial como a temporal para estimar o intervalo de encontros entre veículos. A partir disso, é criada uma tabela de roteamento para encaminhamento das mensagens procedentes de um veículo origem para um veículo destino. Para tanto, algoritmos de correlação espacial e temporal foram criados visando identificar quais as regiões de encontro entre cada par de veículos.

Como trabalho futuro, objetiva-se o aprimoramento dos esquemas e aplicação deles em diferentes cenários de simulação. Além disso, as propostas apresentadas nesta dissertação consideram somente o comportamento individual dos veículos. No entanto, há a possibilidade de exploração de informações coletivas dos veículos. Dessa forma, técnicas como *flock* [Gudmundsson & van Kreveld, 2006], *convoy* [Jeung et al., 2008b] e *swarm* [Li et al., 2010a] podem ser aplicadas, abrindo novas possibilidades de investigação. Além disso, outras abordagens, tal como padrão de frequência, podem ser aplicadas no contexto de redes veiculares para detecção de estradas que possuam maior ocorrência de tráfego e, assim, auxiliem a tomada de decisão para o encaminhamento de mensagens. Dificilmente, uma única solução de encaminhamento de mensagens será eficiente nos mais diversos cenários de redes veiculares. Assim, uma outra oportunidade de pesquisa é investigar como determinar dinamicamente o melhor esquema de encaminhamento de mensagens.

Referências Bibliográficas

- Al-Sultan, S.; Al-Doori, M. M.; Al-Bayatti, A. H. & Zedan, H. (2013). A comprehensive survey on vehicular ad hoc network. *Journal of Network and Computer Applications*, (0):- . ISSN 1084-8045.
- Andrienko, G.; Andrienko, N.; Hurter, C.; Rinzivillo, S. & Wrobel, S. (2011). From movement tracks through events to places: Extracting and characterizing significant places from mobility data. Em *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pp. 161–170.
- Andrienko, G.; Andrienko, N. & Wrobel, S. (2007). Visual analytics tools for analysis of movement data. *SIGKDD Explor. Newsl.*, 9(2):38–46. ISSN 1931-0145.
- Ankerst, M.; Breunig, M. M.; Kriegel, H.-P. & Sander, J. (1999). Optics: ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28(2):49–60.
- Axhausen, K. W.; Schöfeller, S.; Wolf, J.; Oliveira, M. & Samaga, U. (2013). Eighty weeks of gps traces: Approaches to enriching trip information.
- Becker, R.; Cáceres, R.; Hanson, K.; Isaacman, S.; Loh, J. M.; Martonosi, M.; Rowland, J.; Urbanek, S.; Varshavsky, A. & Volinsky, C. (2013). Human mobility characterization from cellular network data. *Commun. ACM*, 56(1):74–82. ISSN 0001-0782.
- Benkert, M.; Gudmundsson, J.; Hübner, F. & Wolle, T. (2008). Reporting flock patterns. *Computational Geometry*, 41(3):111–125.
- Bernsen, J. & Manivannan, D. (2009). Unicast routing protocols for vehicular ad hoc networks: A critical comparison and classification. *Pervasive and Mobile Computing*, 5(1):1 – 18. ISSN 1574-1192.
- Bilal, S. M.; Bernardos, C. J. & Guerrero, C. (2013). Position-based routing in vehicular networks: A survey. *Journal of Network and Computer Applications*, 36(2):685–697. ISSN 1084-8045.

- Boukerche, A.; Oliveira, H. A.; Nakamura, E. F. & Loureiro, A. A. (2008). Vehicular ad hoc networks: A new challenge for localization-based systems. *Computer Communications*, 31(12):2838 – 2849. ISSN 0140-3664.
- Braga, R. B. & Martin, H. (2011). Captain: A context-aware system based on personal tracking. Em *DMS*, pp. 130–133.
- Braga, R. B.; Medeiros da Costa, S. d. M. & Martin, H. (2011). A trajectory correlation algorithm based on users' daily routines. Em *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 501–504. ACM.
- Burgess, J.; Gallagher, B.; Jensen, D. & Levine, B. N. (2006). Maxprop: Routing for vehicle-based disruption-tolerant networks. Em *INFOCOM*, volume 6, pp. 1–11.
- Cadger, F.; Curran, K.; Santos, J. & Moffett, S. (2013). A survey of geographical routing in wireless ad-hoc networks. *Communications Surveys Tutorials, IEEE*, 15(2):621–653. ISSN 1553-877X.
- Calabrese, F.; Pereira, F. C.; Di Lorenzo, G.; Liu, L. & Ratti, C. (2010). The geography of taste: analyzing cell-phone mobility and social events. Em *Proceedings of the 8th international conference on Pervasive Computing, Pervasive'10*, pp. 22–37, Berlin, Heidelberg. Springer-Verlag.
- Cao, H.; Mamoulis, N. & Cheung, D. W. (2007). Discovery of periodic patterns in spatiotemporal sequences. *Knowledge and Data Engineering, IEEE Transactions on*, 19(4):453–467.
- Celes, C.; Braga, R.; Oliveira, C.; Andrade, R. & Loureiro, A. A. F. (2013). Geospin: An approach for geocast routing based on spatial information in vanets. Em *IEEE 78th Vehicular Technology Conference, 2013. VTC Fall 2013. IEEE*, pp. 1926–1930.
- CEMACS (2013). Cemacs: Complex embedded automotive control systems.
- Cheng, P.-C.; Lee, K.; Gerla, M. & Härrri, J. (2010). Geodtn+nav: Geographic dtn routing with navigator prediction for urban vehicular environments. *Mobile Networks and Applications*, 15(1):61–82. ISSN 1383-469X.
- Chisalita, L. & Shahmehri, N. (2002). A peer-to-peer approach to vehicular communication for the support of traffic safety applications. Em *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, pp. 336–341.

- Cho, E.; Myers, S. A. & Leskovec, J. (2011). Friendship and mobility: user movement in location-based social networks. Em *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pp. 1082--1090, New York, NY, USA. ACM.
- Conti, M.; Giordano, S.; May, M. & Passarella, A. (2010). From opportunistic networks to opportunistic computing. *Communications Magazine, IEEE*, 48(9):126--139.
- Cotton, B. (2013). Intelligent urban transportation predicting, managing, and integrating traffic operations in smarter cities.
- CVIS (2013). Cvis: Cooperative vehicle infrastructure systems.
- Dodge, S.; Weibel, R. & Lautenschitz, A.-K. (2008). Towards a taxonomy of movement patterns. *Information Visualization*, 7(3):240--252. ISSN 1473-8716.
- Eagle, N.; Pentland, A. S. & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274--15278.
- Froehlich, J. & Krumm, J. (2008a). A markov model for driver turn prediction. *World Congress in Society of Automotive Engineers (SAE)*, 2193:53 – 53.
- Froehlich, J. & Krumm, J. (2008b). Route prediction from trip observations. *SAE SP*, 2193:53 – 53.
- Fübler, H.; Mauve, M.; Hartenstein, H.; Käsemann, M. & Vollmer, D. (2002). A comparison of routing strategies for vehicular ad-hoc networks.
- Gerla, M. & Kleinrock, L. (2011). Vehicular networks and the future of the mobile internet. *Computer Networks*, 55(2):457 – 469. ISSN 1389-1286. <ce:title>Wireless for the Future Internet</ce:title>.
- Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F.; Renso, C.; Rinzivillo, S. & Trastanti, R. (2010). Mobility data mining: discovering movement patterns from trajectory data. Em *Proceedings of the Second International Workshop on Computational Transportation Science*, IWCTS '10, pp. 7--10, New York, NY, USA. ACM.
- Giannotti, F.; Nanni, M.; Pinelli, F. & Pedreschi, D. (2007). Trajectory pattern mining. Em *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '07, pp. 330--339, New York, NY, USA. ACM.

- Gibbons, P.; Karp, B.; Ke, Y.; Nath, S. & Seshan, S. (2003). Irisnet: an architecture for a worldwide sensor web. *Pervasive Computing, IEEE*, 2(4):22–33. ISSN 1536-1268.
- Gonzalez, M. C.; Hidalgo, C. A. & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- Gu, Y.; Lo, A. & Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *Communications Surveys Tutorials, IEEE*, 11(1):13–32. ISSN 1553-877X.
- Gudmundsson, J. & van Kreveld, M. (2006). Computing longest duration flocks in trajectory data. Em *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, pp. 35–42. ACM.
- Han, J.; Li, Z. & Tang, L. (2010). Mining moving object, trajectory and traffic data. Em Kitagawa, H.; Ishikawa, Y.; Li, Q. & Watanabe, C., editores, *Database Systems for Advanced Applications*, volume 5982 of *Lecture Notes in Computer Science*, pp. 485–486. Springer Berlin Heidelberg.
- Harri, J.; Filali, F. & Bonnet, C. (2009). Mobility models for vehicular ad hoc networks: a survey and taxonomy. *Communications Surveys Tutorials, IEEE*, 11(4):19–41. ISSN 1553-877X.
- Hightower, J. & Borriello, G. (2001). Location systems for ubiquitous computing. *Computer*, 34(8):57–66. ISSN 0018-9162.
- Hung, C.-C.; Chan, H. & Wu, E.-K. (2008). Mobility pattern aware routing for heterogeneous vehicular networks. Em *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, pp. 2200–2205. ISSN 1525-3511.
- Jain, S.; Fall, K. & Patra, R. (2004). Routing in a delay tolerant network. Em *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, SIGCOMM '04, pp. 145–158, New York, NY, USA. ACM.
- Jeong, J.; Guo, S.; Gu, Y.; He, T. & Du, D.-C. (2011). Trajectory-based data forwarding for light-traffic vehicular ad hoc networks. *Parallel and Distributed Systems, IEEE Transactions on*, 22(5):743–757. ISSN 1045-9219.
- Jeong, J.; Guo, S.; Gu, Y.; He, T. & Du, D.-C. (2012). Trajectory-based statistical forwarding for multihop infrastructure-to-vehicle data delivery. *Mobile Computing, IEEE Transactions on*, 11(10):1523–1537. ISSN 1536-1233.

- Jeung, H.; Liu, Q.; Shen, H. T. & Zhou, X. (2008a). A hybrid prediction model for moving objects. Em *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pp. 70--79. IEEE.
- Jeung, H.; Yiu, M. L. & Jensen, C. S. (2011). Trajectory pattern mining. Em Zheng, Y. & Zhou, X., editores, *Computing with Spatial Trajectories*, pp. 143--177. Springer.
- Jeung, H.; Yiu, M. L.; Zhou, X.; Jensen, C. S. & Shen, H. T. (2008b). Discovery of convoys in trajectory databases. *Proceedings of the VLDB Endowment*, 1(1):1068--1080.
- Jiang, D. & Delgrossi, L. (2008). Ieee 802.11 p: Towards an international standard for wireless access in vehicular environments. Em *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 2036--2040. IEEE.
- Jochle, T.; Wiedersheim, B.; Schaub, F. & Weber, M. (2012). Efficiency analysis of geocast target region specifications for vanet applications. Em *Vehicular Networking Conference (VNC), 2012 IEEE*, pp. 250--257. ISSN 2157-9857.
- Joerer, S.; Sommer, C. & Dressler, F. (2012). Toward Reproducibility and Comparability of IVC Simulation Studies: A Literature Survey. *IEEE Communications Magazine*, 50(10):82--88.
- Junhai, L.; Danxia, Y.; Liu, X. & Mingyu, F. (2009). A survey of multicast routing protocols for mobile ad-hoc networks. *Communications Surveys Tutorials, IEEE*, 11(1):78--91. ISSN 1553-877X.
- Krumm, J. & Horvitz, E. (2007). Predestination: Where do you want to go today? *Computer*, 40(4):105--107. ISSN 0018-9162.
- Kurashima, T.; Iwata, T.; Irie, G. & Fujimura, K. (2010). Travel route recommendation using geotags in photo sharing sites. Em *Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10*, pp. 579--588, New York, NY, USA. ACM.
- LeBrun, J.; Chuah, C.-N.; Ghosal, D. & Zhang, M. (2005). Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. Em *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, volume 4, pp. 2289--2293 Vol. 4. ISSN 1550-2252.

- Lee, K. C.; Lee, U. & Gerla, M. (2009). Survey of routing protocols in vehicular ad hoc networks. *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges*, IGI Global, 21.
- Leontiadis, I.; Costa, P. & Mascolo, C. (2010). Extending access point connectivity through opportunistic routing in vehicular networks. Em *Proceedings of the 29th conference on Information communications*, INFOCOM'10, pp. 486--490, Piscataway, NJ, USA. IEEE Press.
- Leontiadis, I. & Mascolo, C. (2007). Geopps: Geographical opportunistic routing for vehicular networks. Em *World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a*, pp. 1--6.
- Li, F. & Wang, Y. (2007). Routing in vehicular ad hoc networks: A survey. *Vehicular Technology Magazine, IEEE*, 2(2):12--22. ISSN 1556-6072.
- Li, X.; Han, J.; Lee, J.-G. & Gonzalez, H. (2007). Traffic density-based discovery of hot routes in road networks. Em *Proceedings of the 10th international conference on Advances in spatial and temporal databases*, SSTD'07, pp. 441--459, Berlin, Heidelberg. Springer-Verlag.
- Li, Z.; Ding, B.; Han, J. & Kays, R. (2010a). Swarm: Mining relaxed temporal moving object clusters. *Proceedings of the VLDB Endowment*, 3(1-2):723--734.
- Li, Z.; Ding, B.; Han, J.; Kays, R. & Nye, P. (2010b). Mining periodic behaviors for moving objects. Em *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pp. 1099--1108, New York, NY, USA. ACM.
- Lindgren, A.; Doria, A. & Schelén, O. (2003). Probabilistic routing in intermittently connected networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(3):19--20.
- Liu, C. & Kaiser, J. (2003). *A survey of mobile ad hoc network routing protocols*. Universität Ulm, Fakultät für Informatik.
- Lochert, C.; Hartenstein, H.; Tian, J.; Fussler, H.; Hermann, D. & Mauve, M. (2003). A routing strategy for vehicular ad hoc networks in city environments. Em *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pp. 156--161.

- Lu, C.-T.; Boedihardjo, A. P. & Shekhar, S. (2009). Analysis of spatial data with map cubes: highway traffic data. *Geographic Data Mining and Knowledge Discovery, 2nd edition edn*, pp. 69--97.
- Maihofer, C. (2004). A survey of geocast routing protocols. *Communications Surveys Tutorials, IEEE*, 6(2):32--42. ISSN 1553-877X.
- Martinez, W. L. & Martinez, A. R. (2001). *Computational statistics handbook with MATLAB*. CRC press.
- Mateus, B.; De Oliveira, C.; Callado, A.; Fernandes, S. & Andrade, R. (2012). Impact of density, load, and mobility on the performance of routing protocols in vehicular networks. Em *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, pp. 1--5. ISSN 1090-3038.
- Namboodiri, V. & Gao, L. (2007). Prediction-based routing for vehicular ad hoc networks. *Vehicular Technology, IEEE Transactions on*, 56(4):2332--2345. ISSN 0018-9545.
- Ni, S.-Y.; Tseng, Y.-C.; Chen, Y.-S. & Sheu, J.-P. (1999). The broadcast storm problem in a mobile ad hoc network. Em *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, MobiCom '99*, pp. 151--162, New York, NY, USA. ACM.
- Noulas, A.; Scellato, S.; Mascolo, C. & Pontil, M. (2011). An empirical study of geographic user activity patterns in foursquare. *ICWSM*.
- Oliveira, C. T. (2008). *Uma Proposta de Roteamento Probabilístico para Redes Tolerantes a Atrasos e Desconexões*. Tese de doutorado, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO.
- Oliveira, C. T. & Duarte, O. C. M. B. (2007). Uma análise da probabilidade de entrega de mensagens em redes tolerantes a atrasos e desconexoes. Em *Computer Networks and Distributed Systems. SBRC 2007 . 25th Brazilian Symposium on*, pp. 293--305. SBC.
- Oliveira, C. T.; Moreira, M. D.; Rubinstein, M. G.; Costa, L. H. M. & Duarte, O. C. M. (2007). Redes tolerantes a atrasos e desconexoes. *SBRC Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.
- Palma, A. T.; Bogorny, V.; Kuijpers, B. & Alvares, L. O. (2008). A clustering-based approach for discovering interesting places in trajectories. Em *Proceedings of the*

- 2008 ACM symposium on Applied computing, SAC '08, pp. 863--868, New York, NY, USA. ACM.
- Pan, G.; Qi, G.; Wu, Z.; Zhang, D. & Li, S. (2013a). Land-use classification using taxi gps traces. *Intelligent Transportation Systems, IEEE Transactions on*, 14(1):113--123. ISSN 1524-9050.
- Pan, G.; Qi, G.; Zhang, W.; Li, S.; Wu, Z. & Yang, L. (2013b). Trace analysis and mining for smart cities: issues, methods, and applications. *Communications Magazine, IEEE*, 51(6):-. ISSN 0163-6804.
- Panichpapiboon, S. & Pattara-Atikom, W. (2012). A review of information dissemination protocols for vehicular ad hoc networks. *Communications Surveys Tutorials, IEEE*, 14(3):784--798. ISSN 1553-877X.
- Parent, C.; Spaccapietra, S. & Zimányi, E. (1999). Spatio-temporal conceptual models: data structures + space + time. Em *Proceedings of the 7th ACM international symposium on Advances in geographic information systems*, GIS '99, pp. 26--33, New York, NY, USA. ACM.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065--1076.
- PATH (2013). Path: Partners for advanced transportation technology.
- Pereira, P. R.; Casaca, A.; Rodrigues, J. J. P. C.; Soares, V. N. G. J.; Triay, J. & Cervello-Pastor, C. (2012). From delay-tolerant networks to vehicular delay-tolerant networks. *Communications Surveys Tutorials, IEEE*, 14(4):1166--1182. ISSN 1553-877X.
- Qu, F.; Wang, F.-Y. & Yang, L. (2010). Intelligent transportation spaces: vehicles, traffic, communications, and beyond. *Communications Magazine, IEEE*, 48(11):136--142. ISSN 0163-6804.
- Ramos, H. S.; Zhang, T.; Liu, J.; Priyantha, N. B. & Kansal, A. (2011). Leap: a low energy assisted gps for trajectory-based services. Em *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pp. 335--344, New York, NY, USA. ACM.
- Reddy, S.; Mun, M.; Burke, J.; Estrin, D.; Hansen, M. & Srivastava, M. (2010a). Using mobile phones to determine transportation modes. *ACM Transaction Sensor Network*, 6(2):13:1--13:27. ISSN 1550-4859.

- Reddy, S.; Shilton, K.; Denisov, G.; Cenizal, C.; Estrin, D. & Srivastava, M. (2010b). Biketastic: sensing and mapping for better biking. Em *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pp. 1817--1820, New York, NY, USA. ACM.
- Santos, R. A.; Edwards, A.; Edwards, R. M. & Seed, N. L. (2005). Performance evaluation of routing protocols in vehicular ad-hoc networks. *Int. J. Ad Hoc Ubiquitous Comput.*, 1(1/2):80--91. ISSN 1743-8225.
- Schoch, E.; Kargl, F.; Weber, M. & Leinmuller, T. (2008). Communication patterns in vanets. *Communications Magazine, IEEE*, 46(11):119--125. ISSN 0163-6804.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.
- Small, T. & Haas, Z. J. (2005). Resource and performance tradeoffs in delay-tolerant wireless networks. Em *ACM SIGCOMM Workshop on Delay-tolerant Networking*.
- Soares, V. N.; Rodrigues, J. J. & Farahmand, F. (2011). Geospray: A geographic routing protocol for vehicular delay-tolerant networks. *Information Fusion*, (0):-. ISSN 1566-2535.
- Sommer, C. & Dressler, F. (2008). Progressing Toward Realistic Mobility Models in VANET Simulations. *IEEE Communications Magazine*, 46(11):132--137.
- Sommer, C.; German, R. & Dressler, F. (2011). Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis. *IEEE Transactions on Mobile Computing*, 10(1):3--15.
- Spaccapietra, S.; Parent, C.; Damiani, M. L.; de Macedo, J. A.; Porto, F. & Vangenot, C. (2008). A conceptual view on trajectories. *Data and Knowledge Engineering*, 65(1):126--146. ISSN 0169-023X.
- Spyropoulos, T.; Psounis, K. & Raghavendra, C. S. (2005a). Spray and wait: an efficient routing scheme for intermittently connected mobile networks. Em *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pp. 252--259, New York, NY, USA. ACM.
- Spyropoulos, T.; Psounis, K. & Raghavendra, C. S. (2005b). Spray and wait: an efficient routing scheme for intermittently connected mobile networks. Em *ACM SIGCOMM Workshop on Delay-tolerant Networking*.

- Tonguz, O. K. & Ferrari, G. (2006). Ad hoc wireless networks: A communication-theoretic perspective.
- Trasarti, R.; Pinelli, F.; Nanni, M. & Giannotti, F. (2011). Mining mobility user profiles for car pooling. Em *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pp. 1190--1198, New York, NY, USA. ACM.
- Uzcategui, R. & Acosta-Marum, G. (2009). Wave: A tutorial. *Communications Magazine, IEEE*, 47(5):126–133. ISSN 0163-6804.
- Vahdat, A. & Becker, D. (2000). Epidemic routing for partially-connected ad hoc networks. Relatório técnico.
- Wand, M. & Jones, M. (1995). Kernel smoothing. vol. 60 of monographs on statistics and applied probability.
- Wisitpongphan, N.; Bai, F.; Mudalige, P.; Sadekar, V. & Tonguz, O. (2007). Routing in sparse vehicular ad hoc wireless networks. *Selected Areas in Communications, IEEE Journal on*, 25(8):1538–1556. ISSN 0733-8716.
- Xu, F.; Guo, S.; Jeong, J.; Gu, Y.; Cao, Q.; Liu, M. & He, T. (2011). Utilizing shared vehicle trajectories for data forwarding in vehicular networks. Em *INFOCOM, 2011 Proceedings IEEE*, pp. 441–445. ISSN 0743-166X.
- Zhao, J. & Cao, G. (2006). Vadd: Vehicle-assisted data delivery in vehicular ad hoc networks. Em *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pp. 1–12. ISSN 0743-166X.