Capítulo 1

Big Data Analytics no Projeto de Redes Móveis: Modelos, Protocolos e Aplicações

Clayson S. F. de S. Celes (UFMG), Ivan O. Nunes (UFMG), João B. Borges Neto (UFMG), Fabrício A. Silva (UFV), Leonardo Cotta (UFMG), Pedro O. S. Vaz de Melo (UFMG), Heitor S. Ramos Filho (UFAL), Rossana M. C. Andrade (UFC) e Antonio A. F. Loureiro (UFMG)

Abstract

The popularization of smart devices with sensing capabilities has led to a huge volume of spatiotemporal data, obtained from different entities, such as people, vehicles, and objects with computing capabilities. The extraction of knowledge from such data offers unprecedented opportunities for decision making processes in several areas. In the mobile networking domain, communication protocols, infrastructure planning and service delivery are examples of applications that can benefit from mining and analysis of data that is collected from smart devices. For instance, user's historic data contain features that are important for detecting patterns and predicting both mobility and data traffic, in time and spatial domains. The goal of this chapter is to present and discuss the potential of big data analytics in the design of mobile networks. In particular, we aim at showing how 5G cellular networks, vehicular networks, and internet of mobile things (IoMT) can take advantage of the knowledge extracted from their entities' characteristics (e.g., user's mobility). In summary, this chapter presents: (i) the peculiarities and techniques of data analysis for the design of mobile networks; (ii) an overview of the recent contributions in the area; (iii) a framework that covers different aspects, ranging from handling the data to employing the knowledge obtained from such data; and (iv) research challenges and opportunities within the area.

Resumo

A popularização de dispositivos com capacidade de sensoriamento tem permitido a obtenção de um enorme volume de dados com informações espaço-temporais de diferentes entidades, tais como pessoas, veículos e objetos com capacidade de processamento.

A extração de conhecimento a partir desses dados cria oportunidades sem precedentes para processos de tomada de decisão em diversas áreas. No domínio de redes móveis, o projeto de protocolos de comunicação, o planejamento de infraestrutura e o fornecimento de serviços são exemplos de aplicações que podem se beneficiar da mineração e análise dos dados coletados a partir desses dispositivos. Por exemplo, dados históricos de usuários contêm características que são importantes para detectar padrões e realizar predições tanto da mobilidade desses usuários como do tráfego de dados no âmbito espaço-temporal. O objetivo deste capítulo é apresentar e discutir o potencial da análise de grandes volumes de dados (big data analytics) com ênfase no projeto de redes móveis. Mais especificamente, nosso objetivo é mostrar como redes celulares 5G, redes veiculares e internet das coisas móveis podem se beneficiar do conhecimento extraído a partir das características (e.g., mobilidade) de suas entidades. Em resumo, este capítulo apresenta: (i) as peculiaridades e técnicas na análise de dados para o projeto de redes móveis; (ii) uma visão geral das contribuições recentes na área; (iii) um arcabouço que aborda desde a manipulação dos dados até a aplicação do conhecimento obtido a partir desses dados; e (iv) os desafios e oportunidades de pesquisa no tema.

1.1. Introdução

Nos últimos anos, a popularização de dispositivos com capacidade de sensoriamento tem permitido a obtenção de um enorme volume de dados com informações espaço-temporais de diferentes entidades, tais como pessoas, veículos e objetos com capacidade de processamento. A extração de conhecimento a partir desses dados oferece oportunidades sem precedentes na tomada de decisão em diversas áreas incluindo planejamento urbano e sistemas de transportes [Çolak et al. 2016], mobilidade urbana [Silva et al. 2014b], epidemiologia [Tizzoni et al. 2014] e sociologia [Soto et al. 2011].

No domínio de redes móveis, sejam essas redes *ad hoc* ou infraestruturadas, o conhecimento extraído a partir de dados do sistema, dos usuários ou da interação entre eles tem fornecido resultados promissores no projeto e na concepção de soluções para tais redes. No entanto, isso exige o conhecimento multidisciplinar que demanda fundamentos distintos de diversas áreas tais como comunicações móveis, mineração de dados e estatística. Nesse sentido, este capítulo visa apresentar um arcabouço que mostra uma metodologia para aplicar análise de dados no projeto de redes móveis a partir de uma visão geral da área.

A literatura apresenta alguns trabalhos relacionados que mostram uma visão geral de esforços que tratam conjuntamente *big data* e redes de comunicação. Em [Yu et al. 2016], os autores apresentam um tutorial sobre infraestrutura e plataformas de redes para tratar o processamento de grande volume de dados. Em [Wang et al. 2016], os autores fazem uma revisão de literatura sobre a aplicação de análise de dados para entender situações de desastres (e.g., terremoto, tempestades) a fim de criar redes móveis *ad hoc*, de forma que os recursos para composição da rede sejam otimizados considerando os resultados obtidos da análise. Em [Blondel et al. 2015] e [Naboulsi et al. 2016], os autores focam exclusivamente em discutir a análise de dados provenientes das comunicações realizadas por usuários de redes celulares, destacando as perspectivas sociais, de mobilidade e de comunicação. Em [He et al. 2016], os autores aplicam *big data analytics* para melhorar o desempenho das redes celulares tanto na perspectiva dos usuários quanto das operadoras. Em [Bi et al. 2015], os autores discutem os desafios e oportunidades no

projeto de comunicações sem fio (*wireless communications*).

Este capítulo se diferencia dos trabalhos supracitados nos seguintes pontos: (i) destacamos as peculiaridades e técnicas na análise de dados a partir de fontes heterogêneas visando o projeto de redes celulares, veiculares e da *Internet* das coisas; (ii) apresentamos um arcabouço, em formato de tutorial, que aborda desde a manipulação dos dados até a aplicação do conhecimento obtido nas análises; e (iii) fazemos uma revisão de literatura evidenciando os principais exemplos de aplicação com seus desafios e oportunidades de pesquisa.

No contexto do SBRC[1], alguns minicursos apresentados nas edições anteriores podem servir de complementação por tratarem de tópicos relacionados. O minicurso de Computação Urbana [Kamienski et al. 2016] cobriu várias aplicações no contexto de cidades inteligentes que podem ser potencializadas com os conceitos e discussões apresentados neste capítulo. Os trabalhos de [Santos et al. 2016] e [Pires et al. 2015] exploram a *Internet das coisas* (IoT) desde a infraestrutura até o domínio de aplicações. Além de aprofundar os estudos sobre *Internet* das coisas móveis, este capítulo também amplia a discussão sobre como explorar os dados oriundos da IoT. Em [Teles et al. 2013] e [Silva et al. 2015d], os autores abordaram as redes sociais móveis e as redes de sensoriamento participativo, respectivamente. Essas redes, quando aplicadas a cenários de grande escala, podem se beneficiar dos conceitos e técnicas de análise de dados. Vale ressaltar que o conteúdo apresentado neste capítulo, aliado aos desafios discutidos nos minicursos citados, fornecem oportunidades interessantes de pesquisa desde a coleta dos dados até aplicações em cenários de comunicação em cidades inteligentes.

A Figura 1.1 ilustra a estrutura geral deste capítulo, que segue uma abordagem *bottom-up* em camadas, tratando desde a obtenção de dados até a sua aplicação no projeto de redes móveis. O restante deste capítulo está organizado como segue. A Seção 1.2 apresenta alguns conceitos sobre *big data*, os tipos de dados e processamento. Em seguida, a Seção 1.3 apresenta o tópico de caracterização de dados. A Seção 1.4 discute as abordagens e modelos utilizados para representar as informações obtidas dos dados brutos com a finalidade de auxiliar no projeto de redes móveis. A Seção 1.5 apresenta diversos resultados de pesquisas atuais relacionados ao projeto de redes móveis a partir de dados sensoriados. Por fim, a Seção 1.6 apresenta as considerações finais deste capítulo.

1.2. *Big data*, tipos de dados e processamento

1.2.1. *Big data*

Nos últimos anos, o termo *big data* tem se tornado bastante popular e empregado em diferentes situações no contexto de dados massivos. Apesar de existirem diversas e até contraditórias definições para o termo [Ward and Barker 2013], um entendimento comum na literatura é que quando se fala sobre *big data* estamos lidando com um grande volume de dados, gerados e processados em uma velocidade fora do comum e que possuem uma larga variedade de informação por serem obtidos de diferentes fontes [Chen et al. 2014a].

No cenário de redes móveis, infraestruturadas ou *ad hoc*, esse grande volume de dados está sendo gerado tanto no domínio do sistema como no domínio dos usuários da

[1] Simpósio Brasileiro de Redes de Computadores

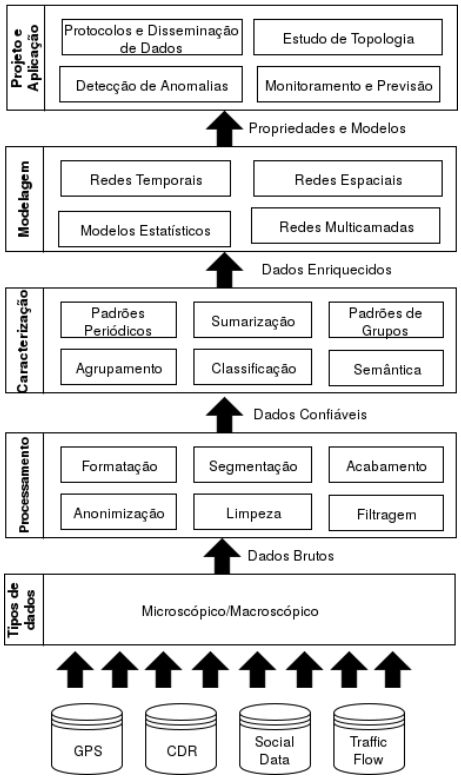


Figura 1.1. Arcabouço geral para aplicar análise de dados no projeto de redes móveis.

rede [Cheng et al. 2017]. Por exemplo, no contexto de redes celulares, do ponto de vista do sistema existem dados sendo gerados pelas interações ou medições entre as comunicações dos dispositivos móveis com as torres celulares, assim como no núcleo da rede. Enquanto no ponto de vista dos usuários, dados estão sendo gerados pela movimentações e registros de chamadas entre usuários. Uma lista bem detalhada dos tipos de dados no contexto de *big data* em redes celulares pode ser encontrada em [Imran et al. 2014].

Explorar a área de *big data* nesse caso é equivalente a agregar informações de diversas fontes a fim de auxiliar a concepção de melhores soluções no projeto de redes móveis. Essa tarefa não é simples, pois, de forma geral, envolve diversos desafios nas seguintes etapas [Wu et al. 2014b]: (i) coleta e tratamento dos dados de forma a não comprometer a segurança do sistema e usuários, além do tratamento dos dados que geralmente possuem várias imperfeições; (ii) armazenamento e manipulação utilizando plataformas computacional de alto desempenho; (iii) caracterização e análise que busca a semântica intrínseca nos dados; (iv) modelagem e aplicação de técnicas relacionadas ao domínio do problema.

1.2.2. Tipos de dados

A ideia de entidades (pessoas, veículos e objetos) como sensores surge como uma oportunidade para coletar e armazenar dados de diferentes dimensões em um cenário urbano. Nesse contexto, a aquisição dos dados pode ser passiva ou ativa [Zheng et al. 2014]. A passiva (ou implícita) consiste em explorar a infraestrutura existente para obter dados gerados pelas entidades. Por exemplo, dados do sistema de navegação de táxi e registros de

Tabela 1.1. Comparação qualitativa dos tipos de fontes de dados.

Tipos de fontes de Dados	Escala (# de entidades)	Granularidade Espacial	Granularidade Temporal	Obtenção dos dados
Registros de detalhamento da chamada (CDR)	Grande	Baixa	Baixa	Fácil
Registros de LBSN	Grande	Alta	Baixa	Fácil
Registros de Transportes	Variante	Alta	Variante	Difícil
Registros de Proximidades	Pequena	N/A	Alta	Difícil
Registros de Posicionamento (HDP)	Pequena	Alta	Alta	Difícil

chamadas nas redes celulares fornecem informações de rastreamento de veículos e pessoas. No entanto, o objetivo primário em cada um desses casos é o fornecimento de rotas e bilhetagem/faturamento para chamadas celulares, respectivamente. Na aquisição ativa (ou explícita) a entidade contribui com o dado intencionalmente para um serviço ou aplicação. Por exemplo, um usuário que compartilha sua localização no *Foursquare* (*check-in*) ou um sensor que fornece leituras de temperatura para um serviço na *Web*.

Nesta seção apresentamos cinco tipos de dados que são largamente usados no projeto de redes móveis: registros de detalhamento da chamada (*Call Detail Record* ou CDR), registros de redes sociais baseadas em localização (*Location based social network* ou LBSN), registros de transportes, registros de proximidades e registros de posicionamento de dispositivos móveis (*Handheld Device Positioning* ou HDP). A Tabela 1.1 apresenta uma comparação qualitativa desses tipos de dados. Em seguida, são apresentados os detalhes sobre cada tipo de dado, destacando suas vantagens, desvantagens, facilidade de obtenção. Além disso, são discutidas possibilidades de aplicação desses tipos de dados.

1.2.2.1. Registros de detalhamento da chamada (CDR)

No âmbito das redes celulares, diversos tipos de dados estão sendo gerados para medições e controle do funcionamento da rede [Cheng et al. 2017]. Particularmente, esta seção foca nos registros de detalhamento da chamada (CDR), que são largamente utilizados na literatura. Registros de detalhamento da chamada são dados dos usuários mantidos pelas companhias telefônicas para fins de controle da rede celular e faturamento dos serviços utilizados. Basicamente, os registros consistem de dados sobre a atividade dos usuários, incluindo realização de chamadas, envio e recebimento de mensagens de texto. Cada registro contém o identificador dos usuários envolvidos na comunicação, data e horário do registro e as estações bases intermediárias usadas na comunicação entre os usuários. Em uma rede celular, cada estação base está associada a uma região geográfica de cobertura (célula), possibilitando, assim, obter uma estimativa da localização dos usuários.

CDRs têm o benefício de cobrir grandes áreas como cidades inteiras, países e até mesmo continentes [Gonzalez et al. 2008]. Além disso, esse tipo de dado permite monitorar um grande volume de usuários, chegando, em alguns lugares, a alcançar um volume na ordem de milhões [Zhang et al. 2014]. Um das principais vantagens de CDRs

é a ausência de custos ou infraestrutura adicionais para a coleta dos dados, já que a sua coleta é realizado pelas empresas de telefonia para fins de cobrança e controle.

A localização de um usuário em CDRs consiste na posição da estação base à qual esse usuário esta associado em um determinado instante, existindo, assim, uma imprecisão espacial no posicionamento dos usuários. Além disso, uma entrada CDR é registrada apenas quando o usuário realiza uma atividade (e.g., efetuar/receber chamadas ou enviar/receber mensagens SMS). Uma vez que os usuários podem não apresentar padrões regulares para a realização dessas atividades, um segundo desafio para a utilização de CDR reside na natureza escassa e não periódica das entradas de posicionamento. Em outras palavras, os dados de CDRs apresentam algumas amostras da trajetória real de um usuário. Apesar dessas restrições, diversos trabalhos na literatura têm adotado dados de CDR para caracterizar e modelar o comportamento dos usuários e do sistema em redes celulares [Gonzalez et al. 2008, Calabrese et al. 2013, Ficek and Kencl 2010, Tanahashi et al. 2012, Wang et al. 2015, Pappalardo et al. 2015].

Geralmente, o acesso a este tipo de dado é restrito às companhias de telefonia, visto que elas o utilizam no serviço de bilhetagem e faturamento. Um desafio consiste em convencer essas companhias a disponibilizar esses dados, pois existe a preocupação com a privacidade dos clientes, além de questões relativas a segredos de negócio [Gramaglia and Fiore 2014]. No entanto, alguns CDRs estão disponíveis publicamente [Blondel et al. 2012, de Montjoye et al. 2014]. Para manter a privacidade do usuários os números de telefone geralmente passam por um processo de anonimização.

1.2.2.2. Registros de LBSN

Redes sociais baseadas em localização, tais como *Foursquare*, *Instagram* e *Twitter*, são um caso particular das redes sociais *online* nas quais os usuários podem informar suas localizações geográficas quando compartilham conteúdos (e.g., *check-in*, foto ou alerta). Com a popularização dos *smartphones*, as LBSNs tornaram-se largamente adotadas e fazem parte da rotina das pessoas que proativamente publicam conteúdos na *Web*. *Web crawlers* podem ser utilizados na extração de conteúdos que possuem dados espaço-temporais dos usuários para compor os registros de LSBN [Silva et al. 2015d].

Uma particularidade quanto à informação de localização nos registros de LBSN é que ela só é disponibilizada se o usuário decidir divulgá-la no momento de compartilhar o conteúdo. Isso introduz imprecisões tanto espaciais quanto temporais para análise de dados. Por exemplo, um usuário pode preferir realizar *check-ins* em localizações específicas ou somente em dias ou horários exclusivos. Portanto, os registros de LSBNs devem ser pre-processados e validados antes de serem submetidos a qualquer tipo de análise e caracterização. Por outro lado, registros de LBSNs têm recebido bastante atenção na literatura devido ao grande volume de check-ins sendo realizados continuamente e à facilidade de coleta desses dados.

1.2.2.3. Registros de Transportes

No contexto deste capítulo, registros de transportes compreendem todos os dados adquiridos de sistemas de transportes que podem ser explorados para obter informações sobre a mobilidade dos usuários. Esses dados vão desde sistemas de GPS (*Global Positioning System*) de veículos até bases de dados de transporte público. Nesta seção, são examinadas algumas das fontes de dados sobre o trânsito e como elas podem ser exploradas para desenvolver estudos no domínio das redes móveis.

Atualmente, uma grande quantidade de veículos estão equipados com sistema de navegação com GPS. Esse sistema fornece aos condutores informações de localização, mapas e rotas. Ao longo da rota de um veículo, o sistema de navegação é capaz de rastrear e armazenar o posicionamento do veículo ao longo do tempo. O conjunto de dados formado por rotas de veículos em um mesmo período e em uma região é tipicamente chamado de *trace*. *Traces* veiculares têm sido frequentemente utilizados para explorar o comportamento urbano e para projetar soluções para redes veiculares. Existem na literatura diversos *traces* de veículos de táxi, por exemplo: São Francisco [Piorkowski et al. 2009a, Piorkowski et al. 2009b], Roma [Amici et al. 2014, Bracciale et al. 2014], Xangai [SUVnet], Shenzhen [Chen et al. 2014b]. Por outro lado, *traces* de veículos de uso pessoal tipicamente não são disponibilizados por questões relativas à privacidade dos usuários.

Além dos *traces* GPS de veículos, com as novas tendências dos sistemas de transporte inteligentes (*Intelligent Transportation Systems* ou ITS), diversos outros tipos de sensores estão sendo instalados no ambiente urbano para a coleta de informações de mobilidade de entidades. Por exemplo, em alguns sistemas públicos de transporte, os usuários de ônibus e metrôs usam cartões inteligentes (*smart cards*) para controle de acesso e saída das estações. Tais tipos de dados podem ser interessantes para o estudo da mobilidade de fluxo de usuários entre regiões, como, por exemplo, determinar a dinâmica da cidade em termos de origem e destino de pessoas durante o horário de pico. Existem alguns conjuntos de dados de *smart cards* disponíveis em *Oyster London Database*[2] e [Zhang et al. 2014].

Os serviços de mapas *online*, tais como *Google Maps*[3], *TomTom*[4], *Here Maps*[5], permitem (por meio de APIs) o acesso a dados de fluxo de tráfego de veículos e condições de trânsito em diversas regiões do mundo quase em tempo real. Por exemplo, o *Google Maps* representa a intensidade de tráfego por uma tabela de cores, enquanto que o *Here Maps* fornece valores numéricos. Esse tipo de informação pode ser aplicada no projeto e validação de modelos e soluções no âmbito das redes veiculares e estudos de sistemas inteligentes de transporte [Tostes et al. 2013].

[2]London OpenData: http://data.london.gov.uk/dataset/oyster-card-journey-information
[3]Google Maps: http://www.google.com.br/maps/
[4]TomTom: http://www.tomtommaps.com/
[5]HereMaps: http://wego.here.com/

1.2.2.4. Registros de Proximidade

Registros de proximidade, ou *traces* de contatos, são bases dados de contatos par-a-par entre entidades. Um contato entre duas entidades consiste em uma interseção espaço-temporal nas trajetórias das duas entidades, i.e., um encontro. Em uma perspectiva de redes oportunistas, um encontro entre duas ou mais entidades é uma oportunidade para disseminação de mensagens/conteúdo entre os usuários/nós da rede. Com uma maior demanda em termos de aplicações para redes oportunistas, tais como o uso de *WiFi Direct* e *Device-to-Device* (D2D) para *offloading* em redes celulares [Rebecchi et al. 2015], os registros de proximidades têm sidos largamente aplicados para simular cenários reais, permitindo avaliação de algoritmos e protocolos.

Existem três formas principais de coleta de dados de registros de proximidades. A primeira é utilizar aplicações móveis de *smartphones* para realizar monitoramento de encontro usando comunicação *Bluetooth* [Eagle and Pentland 2005, Tsai and Chan 2015, Scott et al. 2009]. Alternativamente, uma outra forma utilizada para monitorar encontros é observar a lista de dispositivos móveis conectados aos pontos de acesso de uma rede sem fio [Barbera et al. 2013, Socievole et al. 2014, Henderson et al. 2008, Hsu and Helmy 2005]. A terceira, mais intrusiva, consiste em disponibilizar dispositivos embarcados para um conjunto de pessoas em um ambiente controlado para monitorar seus contatos [Scott et al. 2009, Leguay and Benbadis 2009, Leguay et al. 2006]. Dentre as três formas, a primeira tem se destacado devido ao avanço nas plataformas de desenvolvimento de aplicação para *smartphones*.

Quando *smartphones* são utilizados para coletar dados, primeiramente são recrutados voluntários para instalarem a aplicação de coleta em seus dispositivos. O recrutamento é uma etapa difícil pois envolve convencer os usuários a serem monitorados a partir de seus celulares. Isso se torna um desafio ainda maior em termos de escala (número de usuários), pois poucos usuários aceitam fazer a coleta. Quando se utiliza a coleta por pontos de acessos, deve-se determinar qual será a rede monitorada como, por exemplo, em um ambiente de universidade.

Devidos a esses aspectos, os registros de proximidade sofrem por não possuirem larga escala, em relação ao número de usuários, quando comparados aos registros CDR e LBSN. Por exemplo, o conjunto de dados do MIT Reality [Eagle and Pentland 2006] possui registros de proximidade *Bluetooth* de 80 usuários monitorados durante um período de um ano. O conjunto de dados Dartmouth [Henderson et al. 2008] possui registros de proximidade de 1000 usuários que se conectam aos pontos de acesso de uma universidade durante um período de três meses. Além desses dois conjuntos de dados, outros registros de proximidade foram coletados durante algumas conferências [Pietilainen and Diot 2012, Scott et al. 2009], mas eles apresentam dados de poucos dias de monitoramento.

1.2.2.5. Registros de posicionamento de dispositivos móveis (HDP)

Registros de posicionamento de dispositivos móveis (HDP – Hortonworks Data Platform) são dados obtidos a partir dos receptores GPS nos dispositivos móveis pessoais dos usuários. Normalmente, esses dados são espacialmente precisos e regulares (atualizados

periodicamente), permitindo rastrear as trajetórias dos indivíduos com alta precisão. Vale ressaltar que *traces* de proximidade podem ser obtidos a partir de dados de HDP, visto que um contato pode ser detectado a partir da interseção nas trajetórias de dois indivíduos. Por outro lado, não é possível derivar os registros de HDP a partir apenas de *traces* de contatos.

Como nos registros de proximidade, a coleta desse tipo de dados é feita por aplicações móveis (e.g., MACACO App[6] e *Device Analyzer*[7]) desenvolvidas para essa finalidade. Por isso, *traces* HDP possuem a mesma desvantagem dos *traces* de proximidade: pequena escala. Em geral, *traces* HDP possuem dezenas ou centenas de usuários. Para esse tipo de *trace* é ainda mais difícil convencer os usuários a colaborar porque (i) o monitoramento GPS aumenta consideravelmente o consumo de bateria dos dispositivos, e (ii) os usuários estão preocupados com sua privacidade, ou seja, não desejam divulgar todos os locais que visitam. Por essas razões, existem poucos *traces* desse tipo.

O *GeoLife* [Zheng et al. 2010] é um exemplo de registro HDP disponível publicamente. O conjunto de dados GeoLife possui trajetórias de 182 usuários durante um período de três anos. As entradas do *trace* são uma sequência de posições com registros de latitude, longitude e altitude com granularidade temporal de, em média, 10 segundos.

1.2.2.6. Combinando múltiplas fontes de dados

Todos os tipos de dados apresentados anteriormente possuem algum viés, o qual varia desde a baixa densidade espacial e/ou temporal até a pequena escala em termos de número de usuários. Para contornar tais limitações, uma abordagem interessante consiste em combiná-los. Nesse sentido, alguns esforços na literatura focam em combinar os diferentes tipos de dados para melhorar o entendimento e modelagem das entidades. Por exemplo, em [Silveira et al. 2016], os autores propuseram um modelo que utiliza CDR e LBSN para realizar a predição de mobilidade de pessoas capturando a popularidade de regiões, a frequência de transições entre as regiões e os contatos dos usuários.

1.2.3. Processamento

Um dos principais desafios ao se trabalhar com *big data* consiste em definir como armazenar e processar esses dados. Em geral, os dados brutos capturados provenientes de sensores ou medições possuem redundâncias, inconsistências e vários registros inúteis. Esses dados são, em maioria, não estruturados sendo portanto inadequados de serem armazenados em um banco de dado relacional tradicional [Cattell 2011]. Como alternativa, um banco de dados NoSQL é apto a lidar com tais tipos de dados.

Além disso, os dados brutos, antes de serem armazenados, precisam ser preprocessados, evitando assim o armazenamento de dados com imperfeições. Em [Chen et al. 2014a], os autores destacam as principais tarefas de preprocessando no contexto de *big data*. Limpeza é a tarefa de identificar dados incompletos e imprecisos de forma a removê-los ou modificá-los melhorando assim a qualidade dos dados [Chu et al. 2016]. Eliminação de redundâncias consiste em remover repetições ou dados excedentes, visando reduzir

[6]MACACO App: http://macaco.inria.fr/

[7]Device Analyzer: http://deviceanalyzer.cl.cam.ac.uk/

espaço de armazenamento assim como a tarefa de compressão. Anonimização objetiva ocultar informações prezando a privacidade [Cormode and Srivastava 2009]. Uma visão geral dessas tarefas e técnicas para amenizar os problemas existentes em dados brutos é apresentada em [Karkouch et al. 2016].

Outra questão importante quando se trabalha com um grande volume é o processamento. Nesse caso, o desafio consiste em distribuir as computações dos dados de maneira mais eficiente. O modelo *MapReduce* surgiu como uma excelente contribuição para o processamento paralelo de dados [Dean and Ghemawat 2008]. Atualmente, esse modelo em combinação com o HDFS (*Hadoop Distributed File System*) compõem o núcleo do Apache Hadoop. Hadoop[8] é uma plataforma para processamento distribuído de um grande volume de dados por meio de *clusters* de computadores. Similarmente, Apache Spark[9] é um plataforma para processamento de dados massivos apropriada para aplicações que utilizam múltiplas operações em paralelo, tais como algoritmos de aprendizagem de máquina e mineração de dados [Zaharia et al. 2010].

1.3. Caracterização dos Dados

A caracterização dos dados consiste em uma análise exploratória das principais características de um conjunto de dados [Han et al. 2011]. Com esse objetivo, como descrito em [Tukey 1977], a análise visa maximizar as percepções do conjunto de dados, ao descrever medidas de tendência central e medidas de variabilidade. Além disso, verifica suposições, determina relações entre variáveis, encontra *outliers* e anomalias, entre outros entendimentos implícitos nos dados.

A seguir, vários aspectos identificados na literatura na caracterização dos tipos de dados discutidos na seção anterior são apresentados. Visto que o foco de estudo é de redes móveis, os aspectos tratados aqui são relativos à mobilidade das entidades, atividade temporal, padrões de tráfego de dados e comportamento espaço-temporal de uso dos recursos da rede. Nesse sentido, a caracterização de dados fornece *insights* valiosos para entender o comportamento de usuários e do sistema. Apesar da lista de aspectos apresentada aqui não ser exaustiva, fornece a intuição sobre o propósito de cada um deles.

1.3.1. Estatísticas Básicas dos Dados

Uma necessidade comum quando se trabalha com um grande volume de dados é sumarizá-lo de forma que se possa ter uma visão geral fazendo uso de tabelas, gráficos e valores numéricos. Nesse sentido, adotam-se várias técnicas da estatística descritiva como medidas de tendência central, medidas de dispersão ou variabilidade, tabelas de frequência e gráficos (e.g., diagrama de barras, *box-plot*, histograma) [Freedman et al. 2007]. Por exemplo, em um trabalho recente sobre caracterização do tempo de vida de lugares populares, em [Lu et al. 2016], os autores utilizam dados de localização de embarque e desembarque de táxi na cidade de Nova Iorque[10]. Para sumarizar os dados de mais de um bilhão de viagens dos táxis entre os anos de 2010 e 2015, eles mostraram um conjunto

[8]Apache Hadoop: http://hadoop.apache.org/

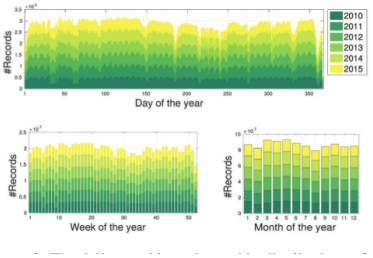
[9]Apache Spark: http://spark.apache.org/

[10]NYC taxicab data from New York City Taxi and Limousine Commission (TLC) official website: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

de estatísticas básicas e gráficos que representam a distribuição de viagens ao longo dos anos. A Figura 1.2 apresenta algumas informações extraídas do conjunto de dados. A Figura 1.2(a) mostra algumas médias em relação ao tempo. A Figura 1.2(b) mostra um gráfico de barras do número de registros para cada ano.

Property	Statistics
#Total records	1,039,821,410
Time Span	01/01/2010 - 12/31/2015
Avg. records / year	173,303,568
Avg. records / month	14,441,964
Avg. records / week	3,266,105
Avg. records / day	473,507

(a)



(b)

Figura 1.2. (a) Informações básicas do número de viagens de táxi. (b) Número de registros de viagens de táxi entre os anos de 2010 e 2015 em Nova Iorque apresentados semanalmente, mensalmente e anualmente.

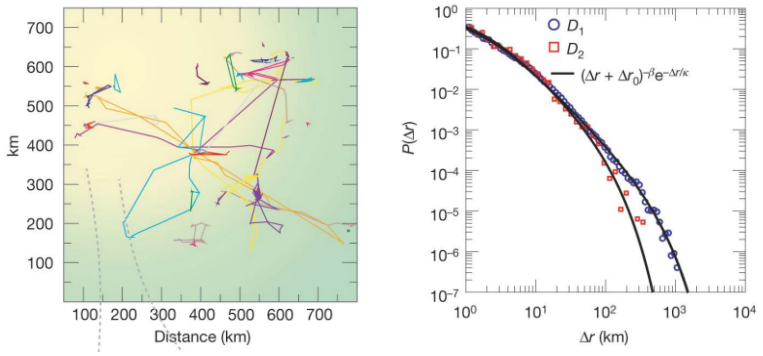
1.3.2. Métricas de Mobilidade

A mobilidade é uma característica fundamental no cotidiano das pessoas e entender os movimentos dos humanos revela muito mais que apenas suas localizações. Dessa maneira, saber como, porque e quando esses movimentos ocorrem, esclarecem vários questionamentos também no projeto de redes móveis tais como sobrecarga na rede, topologia dinâmica, etc. Nesse sentido, esta seção apresenta várias métricas de mobilidade individual e coletiva de entidades presentes no projeto de redes móveis.

Distância de viagem[11] **(*travel distance*).** É uma métrica para descrever o deslocamento de uma entidade [Brockmann et al. 2006]. Ela é computada a partir da interpolação linear entre duas localizações consecutivas em um dado período de tempo, sendo definida como: $\Delta r(u) = \sum_{i=2}^{n} |r_i - r_{i-1}|$ onde $r_u = (r_1, r_2, ..., r_n)$ é a sequência de n posições geográficas registradas no deslocamento de um usuário u durante um período de tempo e $|r_i - r_{i-1}|$ é a distância entre as localizações r_i e r_{i-1}. Em [Gonzalez et al. 2008], os autores analisaram a distância de viagem de pessoas usando dois conjuntos de dados de usuários de rede celular. Eles mediram a distância considerando as posições consecutivas de chamadas telefônicas e encontraram que a distribuição de deslocamento $P(\Delta r)$ em relação a todos os usuários da base de dados segue aproximadamente uma lei de potência truncada (*truncated power-law*). A Figura 1.3 mostra resultados interessantes obtidos nesse trabalho: a maioria dos deslocamentos dos usuários são curtos e as trajetórias humanas mostram um elevado grau de regularidade espacial. Tais resultados serviram como motivação na concepção de vários modelos de mobilidade humana como apresentado em [Hess et al. 2016] .

Raio de giro (*radius of gyration*). Quantifica a dinâmica de mobilidade de uma pessoa em relação ao centro de massa do seu movimento [Gonzalez et al. 2008]. O raio de giro é definido como: $r_g = \sqrt{1/n \sum_{i=1}^{n} (p_i - p_{center})^2}$, onde n é o número de locais visitados por um dado usuário, p_i é a i-ésima posição do usuário e p_{center} é o centro de massa do deslocamento do usuário, obtido como $p_{center} = 1/n \sum_{i=1}^{n} p_i$. O resultado de $p_i - p_{center}$ é a

[11] Também nomeado como deslocamento (*displacement*).



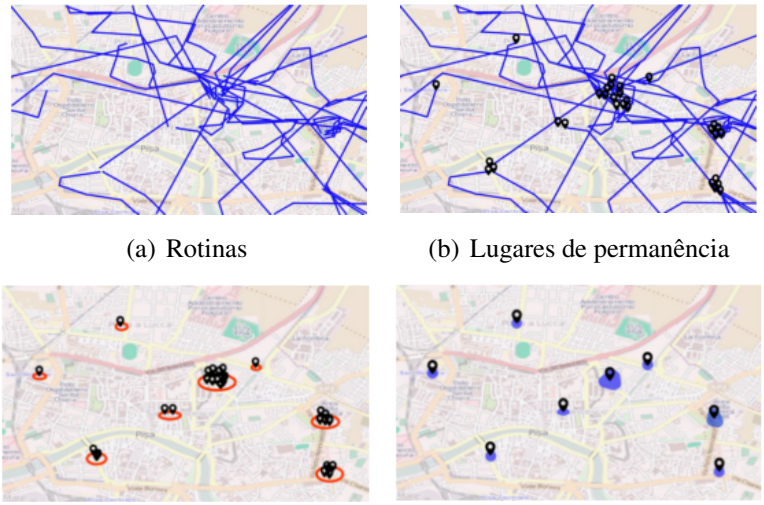
(a) (b)

Figura 1.3. (a) Trajetórias de uma semana de 40 usuários de telefones celulares indicam que a maioria das pessoas deslocam apenas em distâncias curtas, enquanto que algumas deslocam por centenas de quilômetros [Gonzalez et al. 2008]. (b) Função de densidade de probabilidade (P.D.F.) das distâncias de deslocamento obtidas para dois conjuntos de dados estudados (D1 e D2) [Gonzalez et al. 2008].

distância entre um local visitado p_i e o centro de massa da movimentação p_{center}. De uma forma geral, um raio de giro pequeno indica que o usuário se movimenta localmente com viagens curtas, enquanto que um raio de giro grande indica que o usuário se movimenta com viagens mais longas.

Pontos de Interesse (*Points of Interest – PoIs*). Os PoIs de uma cidade são as áreas ou lugares que atraem mais pessoas [Silva et al. 2014a]. A Figura 1.4 ilustra um processo comum para detectar pontos de interesse baseado em dados de mobilidade. Inicialmente, a Figura 1.4(a) mostra um cenário onde diversos indivíduos se locomovem por um certa região. Em seguida, na Figura 1.4(b), observa-se que surgem os lugares de permanência individual (*stay points*). Esses lugares são comumente obtidos pela percepção espaço-temporal entre movimentação dos indivíduos [Jiang et al. 2016, Zheng et al. 2009]. Dado os lugares de permanência, aplicando algoritmos de agrupamento (*clustering*) [Jain 2010] sobre esses lugares é possível obter *clusters* de lugares visitados os quais são potencialmente pontos de interesse, como mostrado nas Figuras 1.4(c) e 1.4(d). Existe uma vasta literatura sobre detecção e recomendação de pontos interesses que podem ser encontrados em [Zhao et al. 2016, Zheng 2015].

Perfis de mobilidade. Analisam o histórico de movimentação das entidades móveis visando estabelecer padrões de comportamentos regulares. Em [Trasarti et al. 2011], os autores exploram o conceito de perfis coletivos de mobilidade em dados de veículos (GPS) para criar um sistema de caronas. Similarmente, o trabalho proposto por [Celes et al. 2013] utiliza os perfis individuais de mobilidade para auxiliar no roteamento de mensagens em redes veiculares. Esses dois trabalhos focam em perfis de mobilidade de trajetórias como podemos ver nas Figuras 1.5 e 1.6. Em [Pappalardo et al. 2015], os autores utilizam tanto CDR como GPS para mostrar a existência de dois perfis de mobilidade baseado na frequência que as entidades visitam determinados lugares e no raio de giro. Eles nomearam de *returners* os indivíduos que vão recorrentemente a poucos lugares e de *explorers* os indivíduos cuja mobilidade não pode



(a) Rotinas (b) Lugares de permanência

(c) *Clustering* (d) Pontos de Interesse

Figura 1.4. Extração de pontos de interesse a partir de dados de mobilidade [Guidotti et al. 2014, Mamei et al. 2016]

ser reduzida a poucos lugares. No domínio de redes móveis, entender as rotinas e perfis de mobilidade das entidades traz benefícios no gerenciamento da mobilidade a partir de padrões preditivos e na melhoria da qualidade de serviços.

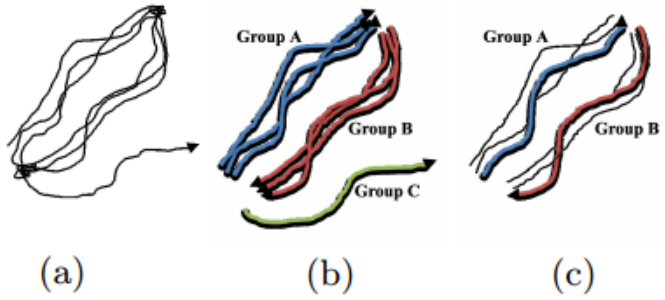


Figura 1.5. Processo para extração do perfil coletivo de mobilidade: (a) identificação das trajetórias dos veículos; (b) detecção de grupos de trajetórias; (c) seleção de perfil de mobilidade mais representativo [Trasarti et al. 2011].

Mobilidade Origem-Destino. Consiste em caracterizar espacialmente pontos de início e fim das movimentações das entidades [Calabrese et al. 2011]. Por exemplo, deslocamento entre cidades, bairros, pontos de interesse. Durante a análise é interessante definir quais as dimensões espaciais e temporais serão utilizadas. A Figura 1.7 mostra os resultados de um estudo que fez a análise de mobilidade origem-destino entre cidades do Reino Unido a partir de dados de LBSN de 2010 a 2013. A Figura 1.7(a) mostra as transições individuais de cada usuário, enquanto a Figura 1.7(b) apresenta uma matriz que mostra a intensidade de transições entre as cidades. Quanto maior a tonalidade de azul, maior é o número de transições entre as cidades. Dessa forma, pode-se perceber um elevado volume de transições entre as cidades de Londres e Edimburgo.

Analisando um outro tipo de dado, trajetórias de veículos, em [Silva et al. 2015b],



Figura 1.6. Processo para extração dos perfis individuais de mobilidade: (a) identificação das trajetórias do veículo; (b) clusterização de trajetórias; (c) sumarização e seleção de perfil de mobilidade mais representativo [Celes et al. 2013].

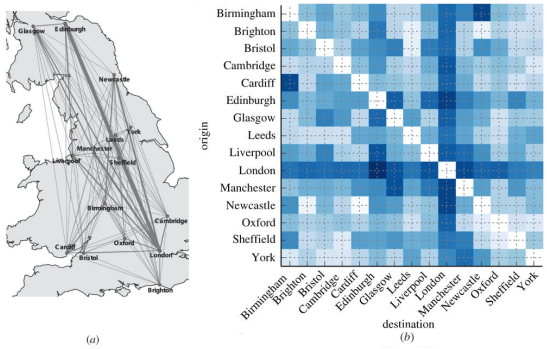


Figura 1.7. Caracterização da mobilidade origem-destino entre cidades do Reino Unido utilizando dados de LBSN [Barchiesi et al. 2015].

os autores caracterizaram a origem e destino dos veículos da cidade de Colônia na Alemanha a partir de um conjunto de dados com trajetórias de mais de 180 000 veículos. Eles particionaram a cidade em células de dimensão 1 km^2 e utilizaram um subconjunto de dados com duas horas de informações. Observando a Figura 1.8(a), percebe-se que as partidas (células de origem dos veículos) tendem a ser igualmente distribuídas por toda cidade, enquanto que observando a Figura 1.8(b), nota-se que as chegadas (células de destino dos veículos) estão concentradas na região central da cidade. Esse comportamento é justificado pelo fato dos autores terem utilizado dados do período de 6:00 às 8:00 da manhã. Nesse período, vários residentes deslocam de casa para o trabalho, ou seja, das regiões de subúrbio para o centro. O estudo realizado nessa caracterização foi aplicado no projeto de replicação de conteúdo em redes veiculares [Silva et al. 2015c].

1.3.3. Métricas de Conectividade

Métricas de conectividade caracterizam as conexões para entender quando elas ocorrem, quão frequentes elas são, quanto duram, entre outras peculiaridades que são importantes para entender a topologia da rede.

Duração de contatos e Tempo entre contatos. São duas características consideradas no domínio de redes *ad hoc*, principalmente nas redes oportunistas. A duração do contato é o intervalo de tempo no qual dois nós da rede estão aptos a se comunicarem por estarem dentro do raio de comunicação. O tempo entre contatos refere-se ao intervalo de

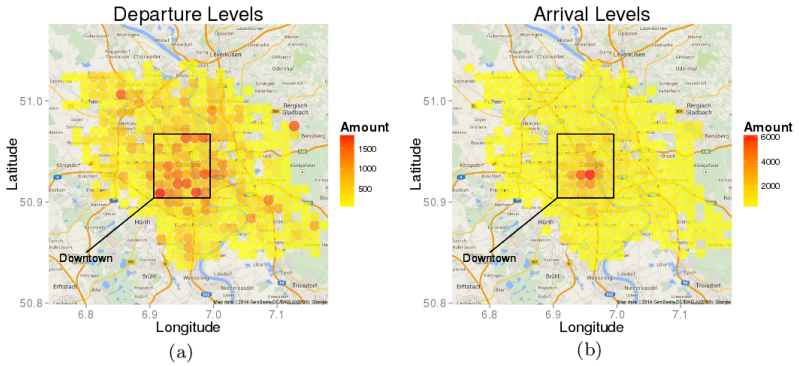


Figura 1.8. Caracterização da mobilidade origem-destino entre regiões da cidade de Colônia na Alemanha utilizando dados de trajetórias de veículos [Silva et al. 2015b].

tempo entre dois contatos. Essas duas características influenciam diretamente no projeto de redes oportunistas em termos da capacidade de dados que podem ser transmitidos em cada contato e maximizando o sucesso de transmissão de mensagens no menor tempo possível, respectivamente.

Topologia. Consiste na disposição dos elementos da rede (e.g., nós, conexões). Explorar os dados gerados pelas entidades da rede visando entender a dinâmica da topologia é extremamente útil na concepção de novos protocolos e serviços. Por exemplo, em [Naboulsi and Fiore 2016], os autores caracterizaram a topologia de uma rede veicular composta de mais de 180 000 veículos em Colônia na Alemanha. Por meio da análise de topologia feita pelos autores é possível responder se a rede formada ao longo do dia é densa ou esparsa, como a conectividade da rede varia ao longo do tempo, como a conectividade é relacionada com as regiões da cidade. A Figura 1.9 mostra como componentes[12] de uma rede veicular variam em relação ao tempo e espaço. Os autores observaram que antes das 6:00 horas, a rede é bastante esparsa e componentes pequenos de no máximo 40 veículos são formados. Entre 7:00 e 8:00 horas ocorre um impacto na topologia de surgimento de componentes gigantes formados por milhares de veículos e outros componentes médios formados por dezenas de veículos, devido ao horário de pico. Esse efeito desaparece e volta acontecer no horário de pico da tarde aproximadamente às 18:00 horas. Vale salientar que os componentes maiores surgem no centro da cidade, onde o tráfego de veículos é denso. Existem outros trabalhos que analisam a topologia de redes veiculares em termos de disponibilidade, conectividade e confiabilidade [Cunha et al. 2016b, Hou et al. 2016, Zhang et al. 2016].

1.3.4. Aspectos Sociais

É típico do comportamento humano estabelecer laços sociais que expressam afinidade e relacionamentos entre indivíduos no dia-a-dia. Esses laços sociais podem ser inferidos a partir dos dados de comunicação entre os dispositivos pessoais de cada indivíduo, já que esses dispositivos, hoje em dia, se tornaram ubíquos. Nesse sentido, observar

[12]Componente é um subgrafo dos nós da rede em que um veículo pode alcançar outro por múltiplos saltos em um intervalo definido t. O tamanho do componente é o número de veículos que pertence a ele.

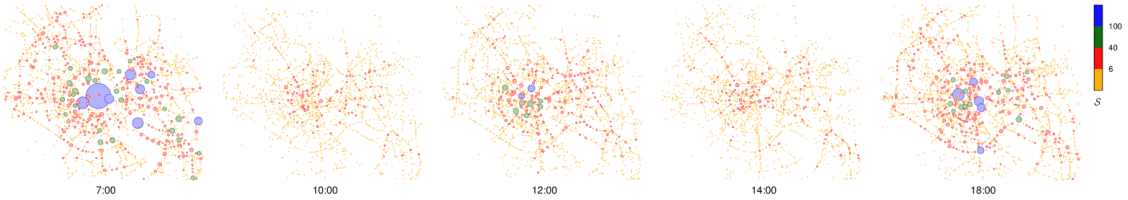
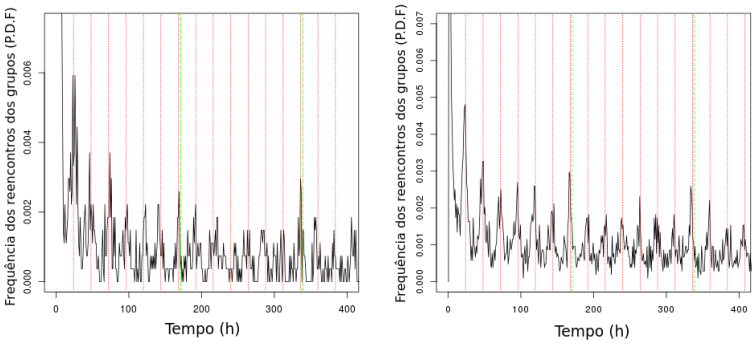


Figura 1.9. Visualização da formação dos componentes de uma rede veicular ao longo do dia em Colônia na Alemanha [Naboulsi and Fiore 2016].

aspectos sociais permite um melhor entendimento da dinâmica e do cotidiano social dos indivíduos. O contexto social, por sua vez pode ser utilizado para aprimorar o desempenho de protocolos de comunicação em redes móveis [Li et al. 2014]. Exemplos de métodos para observar o contexto social a partir de *traces* de contatos são a detecção de comunidades e a detecção de encontros de grupos.

Comunidades. Em diversos tipos de redes, sejam elas sociais, complexas ou móveis, existem nós que são mais interligados entre si formando um aglomerado de nós chamado de comunidades [Palla et al. 2005]. Por exemplo, uma rede social pode ser composta por interações entre pessoas a partir de dados de chamadas telefônicas entre usuários [Hidalgo and Rodriguez-Sickert 2008] ou encontros [Gao et al. 2009]. Os algoritmos propostos por [Palla et al. 2005] e [Gregory 2010] são os mais efetivos e conhecidos na literatura para detecção de comunidades quando tais redes são representadas como grafos estáticos. Em [Nguyen et al. 2011], os autores propuseram uma abordagem semelhante para detecção de comunidades em grafos dinâmicos, i.e., quando nós e arestas surgem e somem ao longo do tempo.

Encontros de Grupos. Em [Nunes et al. 2016b], os autores propõem uma metodologia de detecção de encontros coletivos, i.e., encontros de grupos de três ou mais pessoas. Encontros são definidos como eventos em que duas ou mais entidades (e.g., pessoas, veículos) ocupam a mesma região em um mesmo intervalo de tempo, ou seja, essas entidades encontram-se fisicamente próximas. Ao realizar uma caracterização temporal dos encontros de grupos, percebeu-se que esses encontros apresentam alta periodicidade. Como mostrado na Figura 1.10, em diferentes *traces* de contatos, temos que os encontros de grupos se repetem frequentemente, apresentando principalmente periodicidade diária e semanal. Mais especificamente, a Figura 1.10 mostra a função de densidade de probabilidade (PDF) para os reencontros de grupos de pessoas nos *traces* de contatos do MIT e de Dartmouth. Pode-se notar que as PDFs apresentam picos de maior densidade em torno de períodos de 24 horas (representados pelas linhas pontilhadas vermelhas). Além disso, são observados picos maiores em períodos de sete dias (indicados pelas linhas tracejadas verdes). Esse resultado evidencia a existência de regularidade temporal nos encontros de grupos. Na seção 1.5.1, será discutida a utilização do conhecimento sobre encontros de grupos para o projeto de um protocolo de encaminhamento oportunístico em redes D2D.



(a) *Trace* do MIT (b) *Trace* de Dartmouth

Figura 1.10. Função de densidade de probabilidade (PDF) do tempo entre reencontros de grupos de pessoas após o primeiro encontro em $t = 0$.

1.3.5. Aspectos de Uso da Rede

Com o objetivo de atender a crescente demanda do uso de recursos da rede celular, é importante entender a dinâmica do tráfego e seu impacto na alocação de recursos na rede da operadora de serviços. A partir dessa compreensão, pode-se tanto definir um melhor planejamento de recursos como auxiliar o projeto de rede, beneficiando de alguma forma os usuários de telefonia.

Distribuição espaço-temporal do tráfego. Uma investigação interessante consiste em saber como as pessoas costumam utilizar o tráfego de dados no domínio do tempo e em que região da cidade. Em [Xu et al. 2016], os autores analisaram dados da cidade de Xangai para verificar a distribuição temporal de uso de dados. A Figura 1.11 mostra a distribuição de uso de dados em diferentes escalas de tempo[13]. Na Figura 1.11(a), os dados agregados do tráfego são exibidos e percebe-se que a distribuição segue aproximadamente o comportamento da rotina humana, tráfego intenso durante o dia e tráfego reduzido à noite. Note que existem dois picos durante o dia, um em torno das 12:00 horas e outro em torno das 22:00 horas, possivelmente pelo consumo de dados após o almoço e antes de dormir. As Figuras 1.11(b) e 1.11(c) mostram o comportamento periódico existente ao longo dos dias. A primeira mostra uma semana de dados, enquanto a segunda mostra um mês. Para ambos os casos, percebe-se a regularidade quanto aos uso do dados, tendo no finais de semana uma diminuição do consumo. Além de verificar o comportamento temporal, os mesmos autores analisaram em que regiões da cidade tinham consumo (*bytes* transmitidos por hora por km^2) maior em diversos momentos do dia como mostra a Figura 1.12. Nota-se que independente da hora do dia, as torres localizadas no centro da cidade (região com coloração vermelha) apresentam uma intensidade alta de dados em relação às demais torres de outras regiões. Um trabalho semelhante é apresentado em [Nika et al. 2016], onde os autores propuseram uma metodologia para identificar *hotspots* (lugares com tráfego intenso na rede) com a finalidade de fornecer qualidade de serviço aos usuários mesmo em

[13]Em [Xu et al. 2016], os autores utilizaram dados CDR de 150 000 usuários de um mês da cidade de Xangai na China. As informações analisadas contém o identificador (ID) dos dispositivos, início e fim da conexão, estação base que o ID está associado e quantidade de dados 3G e LTE usados na conexão. Os dados contêm 2.4 petabytes registros, 77 terabytes por dia e 8 gigabytes, em média, por estação base.

situações de sobrecarga.

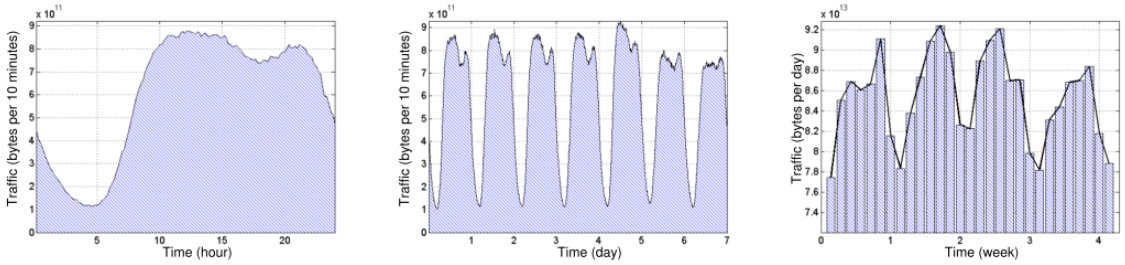


Figura 1.11. Distribuição temporal do tráfego em rede celular para diferentes escalas de tempo [Xu et al. 2016]. (a) Hora em hora. (b) Diariamente. (c) Semanalmente.

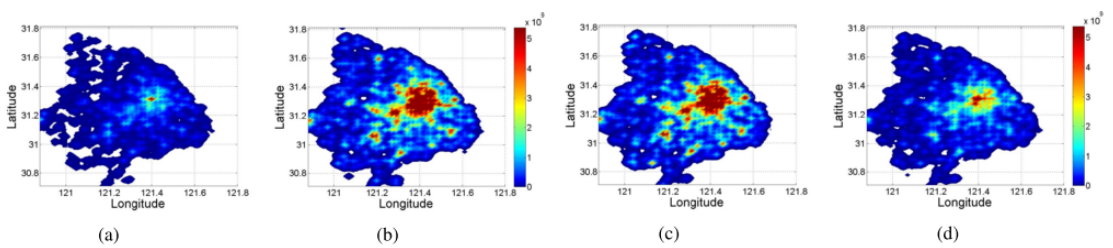


Figura 1.12. Distribuição espacial do tráfego em rede celular em diferentes momentos [Xu et al. 2016]. (a) 04:00. (b) 10:00. (c) 16:00. (d) 22:00.

Distribuição espaço-temporal de tipos de conteúdo. Um outro aspecto importante consiste em caracterizar que tipo de conteúdo os usuários da rede celular utilizam em relação ao tempo e região geográfica. Em [Shafiq et al. 2015], os autores caracterizaram a relação entre volume de tráfego e tipo de conteúdo. A Figura 1.13(a) ilustra a utilização do tráfego de quatro diferentes tipos de conteúdos (*dating*, *maps*, *social network*, *web*) em função do tempo. Pode-se observar que o volume de tráfego para *dating* e *social network* é maior durante o período da noite. Enquanto que o volume de tráfego para navegação *web* e *maps* concentra-se desde o meio-dia até o resto da tarde. A Figura 1.13(b) mostra que para determinadas regiões a intensidade de tráfego para um tipo de conteúdo predomina em relação aos demais. Em outro trabalho semelhante ao descrito anteriormente [Trestian et al. 2009], os autores investigaram a correlação dos interesses dos usuários com localizações geográficas a partir de dados da rede celular.

1.3.6. Considerações

A capacidade para identificar os aspectos descritos anteriormente está fortemente relacionada à aplicação de técnicas e algoritmos do domínio da estatística, mineração de dados e aprendizagem de máquina. Por isso, o restante desta seção apresenta uma breve introdução a essas técnicas e algoritmos.

Agrupamento. É um método não supervisionado (não requer processamento de treinamento de aprendizagem) para agrupar as amostras de dados e, principalmente, classificá-las de acordo com suas características. Os algoritmos de agrupamento buscam

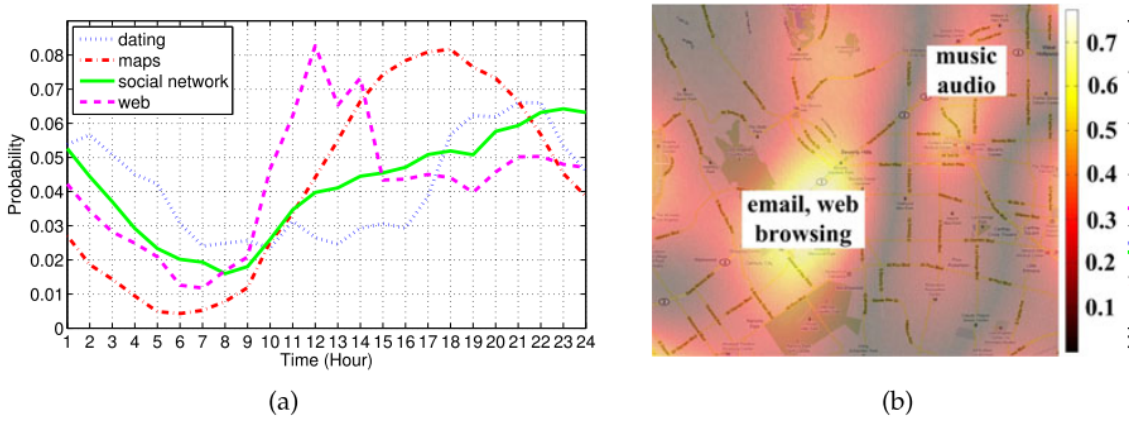


Figura 1.13. Distribuição espaço-temporal de diferentes tipos de conteúdos em redes celulares [Shafiq et al. 2015]. (a) Distribuição temporal de hora em hora. (b) Tipos de conteúdos mais consumidos em regiões distintas.

particionar as amostras em grupos (ou *clusters*) nos quais as amostras pertencentes a cada grupo possuem uma maior similaridade. Existem diferentes tipos de algoritmos de agrupamento que dependendo dos dados e aplicação podem ser mais adequados em uma dada situação tais como agrupamento baseado em densidade, agrupamento baseado em particionamento, agrupamento hierárquico e agrupamento espectral [Hand et al. 2001].

Correlação. É uma técnica para investigar a relação entre duas variáveis contínuas. As medidas comumente aplicadas para medir a correlação são as correlações de Pearson e Spearman. Por exemplo, em [Centellegher et al. 2016], os autores investigaram a correlação entre o número de mensagens SMS e chamadas telefônicas realizadas por usuário e verificaram que existe uma forte relação do uso do telefone celular para essas duas variáveis.

Regressões. A regressão linear permite explorar e estimar um valor esperado quantitativo de uma variável (Y) a partir dos valores de outras variáveis (X) [James et al. 2014]. Ela é dita linear pois existe uma relação linear entre Y e X_1, X_2, ...,X_p. Existe outro tipo de regressão conhecida como regressão logística que se diferencia da linear, principalmente pelo fato da variável resposta ser categórica.

Padrões frequentes. Consiste em detectar itens, subsequências ou estruturas que são recorrentes em um conjunto de dados [Aggarwal and Han 2014]. Encontrar padrões frequentes nos dados pode ser útil tanto para verificar associações e relacionamentos como para auxiliar nas tarefas de indexação, classificação e agrupamento dos dados.

Séries Temporais. De forma geral, trata da representação dos dados em função do tempo [Shumway and Stoffer 2010]. A análise de séries temporais compreende métodos para analisar dados a fim de extrair estatísticas significativas e outras características dos dados. Por exemplo, por meio de séries temporais é possível verificar tendências, sazonalidades e *outliers* nos dados.

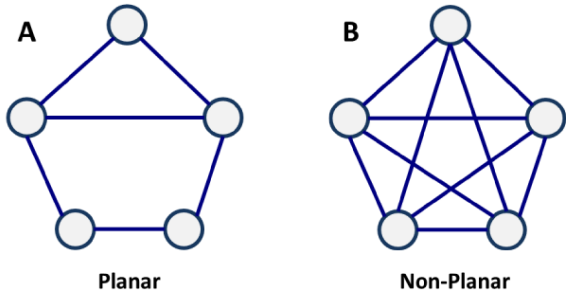


Figura 1.14. Exemplos de grafos planar (A) e não-planar (B)

1.4. Modelagem

A Seção 1.3 discutiu como caracterizar alguns exemplos de dados provenientes de redes móveis. Essa etapa é importante no entendimento de peculiaridades, comportamento macro e microscópio dos aspectos que serão explorados. Em seguida, busca-se um modelo que represente o comportamento real desses aspectos no sistema observado. Por fim, a partir de um modelo descritivo dos dados, cria-se um modelo generativo do comportamento observado.

A seguir, apresentamos como podemos modelar os dados observados usando redes espaciais, redes temporais, redes sociais, redes multicamadas e, por fim, como criar modelos generativos. Em cada seção, apresentamos os tipos de dados apropriados em cada modelo, suas vantagens e desvantagens, exemplos e estratégias de uso.

1.4.1. Redes Espaciais

Em uma modelagem tradicional de redes, ou grafos, temos um conjunto de nós N e um conjunto de arestas E conectando os nós. Usualmente, nós e arestas podem possuir atributos, caracterizando assim a geometria da rede. De maneira geral, uma rede espacial é uma rede onde os nós estão posicionados em um espaço equipado de uma métrica [Barthélemy 2011]. Para a maior parte das aplicações, esse espaço tem duas dimensões e a métrica é a distância Euclidiana. Nessa definição de rede espacial, a probabilidade de dois nós estarem conectados é inversamente proporcional à distância entre eles. Entretanto, essa definição não garante a planaridade da rede. Uma rede planar é uma rede que pode ser desenhada no plano sem que haja intersecção de arestas, como mostra a Figura 1.14. Um exemplo tipicamente encontrado na prática de uma rede espacial não planar é a de aeroportos. Nessa rede, aeroportos formam o conjunto de nós e vôos entre dois aeroportos compõem o conjunto de arestas. Nesse caso, as arestas são direcionadas no sentido do destino do vôo. Apesar da probabilidade de existir um vôo entre dois aeroportos muito distantes ser pequena, há fatores como tamanho das cidades que influenciam mais o surgimento de arestas, tornando a rede não planar. Outras redes, como a social, podem ser vistas como uma rede espacial, porém não planar. Duas pessoas localizadas longe espacialmente são menos propícias a serem amigas, entretanto, esse tipo de amizade existe por conta de diversos outros motivos, como a própria *Web*. Todavia, como veremos a seguir, a maioria das modelagens de mobilidade resultam em redes planares.

Por exemplo, se considerarmos o aspecto de mobilidade dos nós. Modelar mo-

bilidade humana com redes espaciais é bastante direto e intuitivo. Na maior parte das abordagens, um nó representa uma seção do espaço como, por exemplo, o subespaço que envolve um quarteirão ou um aeroporto de uma cidade. Uma aresta representa a possibilidade de movimentação direta entre dois espaços. Essa abordagem pode ser vista como uma maneira de discretização do espaço. Da mesma forma, a mobilidade de um entidade pode ser discretizada em uma sequência de espaços no tempo. Essa sequência, chamada de trajetória, define um caminho que a entidade percorre na rede. Os nós e arestas da rede podem receber atributos para representar a mobilidade das entidades observada nos dados. A maneira mais comum de fazer isso é atribuir às arestas pesos proporcionais ao fluxo de entidades entre os dois nós ligados pela aresta. Observe como essa modelagem permite apenas a análise estática da mobilidade. Isto é, como o componente de tempo não é considerado na rede, ela representa uma fração do tempo, no qual trajetórias são agregadas na mesma rede. Essa abordagem normalmente nos permite fazer análises mais globais sobre mobilidade. Observar eventos que dependem do tempo, como o movimento pendular em grandes cidades, se torna inviável. Todavia, podemos fazer análises macroscópicas de movimento. Em [Barthélemy 2011], usa-se o exemplo da lei de gravidade em mobilidade utilizando essa modelagem. Nesse caso, são adicionados os atributos de distância física entre nós nas arestas e de população do espaço nos nós. Os nós representam cidades ligadas por vôos existentes entre elas. É mostrado como o fluxo entre elas é diretamente proporcional às suas populações e à distância entre elas.

No contexto de redes móveis, uma análise estática, isto é, apenas espacial, de mobilidade pode ser útil em diversas aplicações. Um exemplo seria analisar quais regiões do espaço possuem menor fluxo, e portanto, deveriam receber mais atenção no envio de mensagens. Por exemplo, caso se saiba que uma entidade contém uma aresta (a,b) de peso pequeno em sua trajetória futura, pode-se encaminhar para ela o máximo de mensagens possível que tem como destino b. Em um estudo utilizando dados de registro de chamadas e SMS [Lambiotte et al. 2008], os autores investigaram a relação entre comunicações e posicionamento geográfico dos usuários. Eles modelaram os nós como as regiões de coberturas das estações bases que os usuários estão associados e as arestas como as ligações de comunicações. Essa rede possui 2,5 milhões de usuários monitorados em um período de seis meses. Um resultado interessante observado por eles foi que a duração média das chamadas aumenta com a distância até um limite de 40 km, como pode ser observado na Figura 1.15. Os autores justificam esse fato pelo fato de quando as pessoas vivem próximas umas das outras tendem a se encontrar e comunicar pessoalmente.

1.4.2. Redes Temporais

De maneira geral, uma rede temporal é uma rede, ou grafo, em que seu conjunto de arestas é uma função do tempo. Isto é, uma aresta existe ou não de acordo com uma dada janela de tempo em que observamos a rede. Existem diversas formas de modelarmos essa estrutura e a melhor escolha depende da análise que pretendemos fazer [Holme and Saramäki 2012].

A modelagem mais usual de uma rede temporal é via a análise de redes agregadas [Holme and Saramäki 2012]. Definimos uma janela de tempo e agregamos as arestas contidas nela. Assim, analisamos separadamente cada rede como uma rede estática, podendo, por exemplo, empregar algoritmos da teoria clássica de grafos. Na Figura 1.16, observamos um exemplo de redes agregadas e suas sequências. Uma prática muito comum,

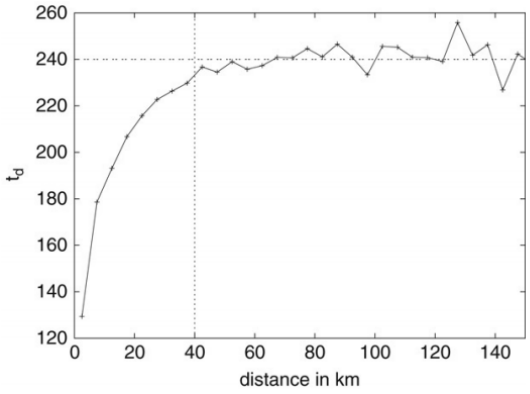


Figura 1.15. Duração média t_d de chamadas telefônicas (em segundos) em função da distância [Lambiotte et al. 2008].

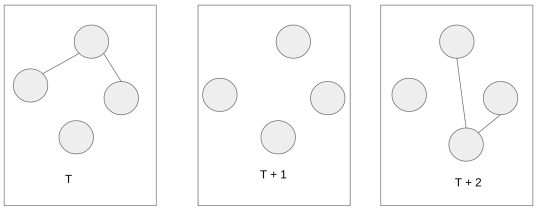


Figura 1.16. Exemplo de redes agregadas por janela de tempo

porém perigosa, é a de empregar algoritmos de caminhamento nas redes agregadas. Nada impede que o algoritmo considere um caminho impossível na prática. Considere um caminho de tamanho k na rede agregada como a sequência de arestas $a_1, a_2, ..., a_k$, o tempo de aparecimento de uma aresta a_i como $ti(a_i)$ e o tempo de desaparecimento de uma aresta a_i como $tf(a_i)$. O caminho é impossível caso exista $tf(a_{i+1}) < ti(a_i)$. Esse é um dos problemas fundamentais de redes temporais e diversos algoritmos foram propostos para lidar com ele [Holme and Saramäki 2012]. Devido ao grande uso de redes temporais para modelagem de redes de contatos e, consequentemente, encaminhamento de mensagens em redes móveis, recomendamos a leitura de tais algoritmos, suas aplicações e desafios.

No contexto de redes móveis, as redes temporais podem ser utilizadas para modelar tanto os contatos entre entidades quanto as suas mobilidades. Pelo lado da mobilidade, podemos agregar o fator tempo às redes espaciais apresentadas anteriormente, por exemplo. Criamos, então, redes espaço-temporais [Williams and Musolesi 2016], capazes de representar o fluxo de entidades no espaço condicionado ao tempo. Isto é, análises de movimento pendular podem ser feitas com essas estruturas. Em [Williams and Musolesi 2016], os autores estudam propriedades clássicas de redes, como vulnerabilidade nas redes de mobilidade propostas.

Além de estudar a mobilidade propriamente dita, podemos usar redes temporais para estudar os contatos gerados pela mobilidade das entidades. Esse é o uso mais comum e, talvez, mais importante de redes temporais em redes móveis. Um contato entre duas entidades pode ser definido como uma interseção de tempo e espaço. Um contato ocorre

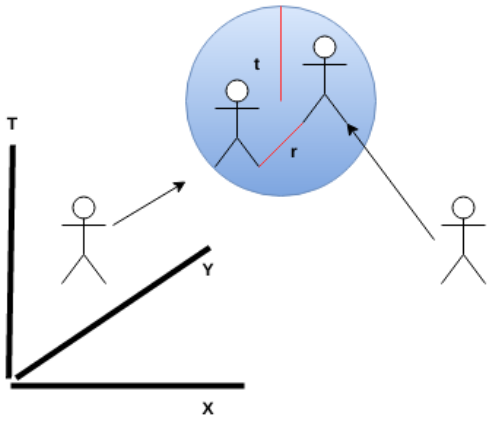


Figura 1.17. Duas entidades geram um contato na rede caso ocupem o mesmo subconjunto do espaço tempo definido por (r,t)

quando duas entidades estão a uma distância máxima r no mesmo instante de tempo. A distância r é definida de acordo com a aplicação analisada. Por exemplo, se estudamos redes D2D podemos definir r como o raio de comunicação da tecnologia Bluetooth. A duração do contato é o tempo no qual as duas entidades permanecem a uma distância de no máximo r. Caso o contato esteja sendo extraído computacionalmente de dados de trajetórias, temos de colocar uma tolerância t na intersecção de tempo. Lembrando que uma trajetória é uma amostra discreta de tempo e espaço. Assim, observamos que um contato é definido dentro de um subconjunto de tempo e espaço, como mostrado na Figura 1.17. Dessa maneira, podemos montar uma rede temporal representando os contatos entre as entidades. Nela, cada nó representa uma entidade e cada aresta representa um contato. Os intervalos de tempo de existência de cada aresta são iguais às durações dos contatos entre as entidades. Essa modelagem é bastante comum em redes móveis pois possibilita o estudo do encaminhamento de mensagens entre as entidades.

A maior parte das aplicações de redes móveis, seja de mobilidade veicular ou humana, utiliza a modelagem de redes temporais para estudar e tirar vantagem de propriedades como periodicidade, centralidade e outras métricas sociais que explicamos a seguir [Mota et al. 2014]. Um exemplo do uso de redes temporais em redes veiculares é dado em [Wu et al. 2011]. Na publicação, os autores prevêem trajetórias dos veículos podendo montar uma rede temporal futura. Nela, eles analisam o problema de roteamento de pacotes e mostram que esse é um problema NP-difícil. Os autores propõem heurísticas locais e globais para esse problema.

Em [Michel and Julien 2016], os autores modelaram um cenário de comunicação entre dispositivos em uma rede móvel *ad hoc* como uma rede temporal. A Figura 1.18 mostra uma visão geral do funcionamento da transmissão de duas mensagens em um grafo temporal. Primeiramente, na Figura 1.18(a), os autores modelaram com um tipo específico de representação de grafo temporal. Os vértices são os nós da rede e as arestas indicam os contatos, sendo que os rótulos dos contatos representam o tempo de contato entre os nós. A Figura 1.18(b) mostra os intervalos de interseção entre os contatos e a Figura 1.18(c)

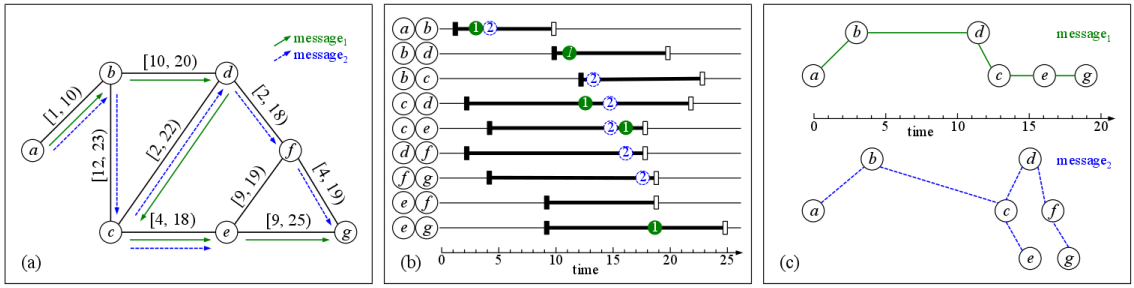


Figura 1.18. Processo de transmissão de mensagem em um grafo temporal [Michel and Julien 2016]. (a) uma representação de uma rede móvel oportunista. (b) intervalos de encontros para pares de nós. (c) identificação do tempo de recebimento de mensagens pelos nós.

exibe o tempo de recebimento das mensagens por cada nó. Um ponto interessante aqui é a possibilidade de explorar a caracterização para entender como os encontros acontecem com objetivo estimar os tempos e durações de encontros que serão utilizados como rótulos das arestas.

1.4.3. Redes Sociais

Redes Sociais Online surgiram (RSO) recentemente e já se tornaram uma vasta área de pesquisa em ciência da computação [Garton et al. 1997]. A premissa básica no estudo de RSO é a de que cada nó representa uma pessoa, ou entidade, e cada aresta representa um laço social, ou amizade, entre elas. Diversos estudos vêm abordando as estruturas formadas pelos laços sociais das pessoas e o que elas significam na prática [Leskovec et al. 2009]. Muitos outros usaram tais estruturas para projetar melhores sistemas [Benevenuto et al. 2009].

É intuitivo pensar que as redes de contatos apresentadas anteriormente podem ser vistas como redes sociais, já que o contato físico é uma forma de interação social. Todavia, muitos desafios são apresentados nessa modelagem. O primeiro é o fato de RSOs serem redes estáticas. Como vimos anteriormente, as redes de contatos são redes temporais (ou espaço-temporais). O primeiro desafio seria converter a rede temporal em uma rede estática. Em uma RSO, as pessoas declaram explicitamente que são amigas, gerando assim o laço entre elas. O problema na conversão da rede de contatos para uma RSO seria identificar quais contatos são frutos de uma amizade e quais são aleatórios. Podemos ir além e classificar o grau de socialização entre as pessoas ou entidades como, por exemplo, amigos e conhecidos. Esse desafio de pesquisa foi explorado em [de Melo et al. 2015]. Os autores utilizam a periodicidade e a estrutura social dos contatos para caracterizá-los como amigos, conhecidos ou aleatórios. Em [Hui et al. 2011], os autores consideram todos os contatos em uma janela de tempo como laços sociais e exploram as estruturas sociais, como comunidade e centralidade para criar um protocolo de roteamento. Em [Cunha et al. 2016a], os autores exploram interações sociais entre veículos, extraídas da mesma forma que em [Hui et al. 2011], para criar um novo protocolo de encaminhamento de mensagens. Por outro lado, em [Nunes et al. 2016c], os autores apresentam um novo conceito social: o encontro de grupos em redes de contatos. Com essa nova medição social, os autores projetaram um novo protocolo de encaminhamento de mensagens D2D.

A extração de propriedades sociais das redes de contatos já se mostrou muito

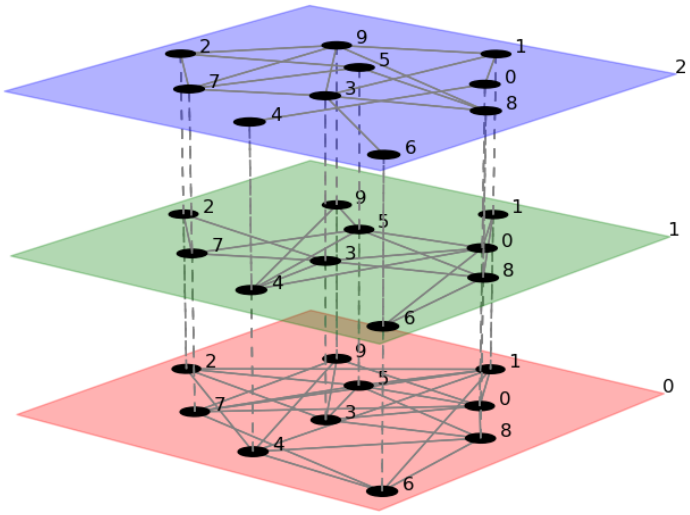


Figura 1.19. Exemplo de rede multicamada com três domínios, ou camadas, diferentes.

eficiente para aplicações de redes D2D e, mais recentemente, de redes V2V. Os principais problemas atrelados a elas são a distributividade das computações e a limitação e dependência de dados para validação. Isto é, as computações envolvidas normalmente são centralizadas, pois precisam de informações de todos os nós. Na prática, os protocolos de redes móveis são distribuídos e o acesso e computação dessas informações pelos nós se torna inviável [Hui et al. 2011]. Além disso, propriedades sociais podem ser bastante dependentes da fonte de dados que usamos. Isto é, analisar propriedades sociais de uma base de dados de um campus universitário naturalmente vai ser diferente de fazer a mesma tarefa em uma base com pessoas aleatórias de uma cidade. Na última, a rede de contatos e, consequentemente a rede social, vai ser muito esparsa. Assim, observamos que como toda solução orientada a dados, as soluções que usam propriedades sociais devem ser ponderadas pelo contexto em que foram avaliadas.

1.4.4. Redes Multicamadas

Um paradigma recente em teoria de redes é o de redes multicamadas [Kivelä et al. 2014]. Essas redes abordam o problema de entidades possuirem relações de diferentes naturezas. Isto é, para cada par de nós na rede, é definido um conjunto de domínios. Em cada domínio, ou camada, existe um conjunto de arestas, que definem as relações entre as entidades naquele domínio. A teoria de redes multicamadas propõe novos algoritmos que exploram eficientemente as estruturas definidas [Kivelä et al. 2014]. O exemplo mais intuitivo talvez seja o de relações humanas, que possuem diversas naturezas. Por exemplo, duas pessoas possuem uma relação definida no domínio do trabalho, empregado e empregador, e uma no domínio social, amigos. A Figura 1.19 ilustra uma rede com o mesmo conjunto de nós em três camadas, sendo que, para cada camada, existem diferentes relações entre os nós.

Depois de apresentarmos modelagens de redes móveis considerando domínios diferentes, a modelagem de redes multicamadas se torna bastante intuitiva nesse contexto. Como vimos, uma rede de contatos possui aspectos temporais, espaciais e sociais. Assim, podemos considerá-los como domínios de uma rede multicamada. Existem diversas outras possibilidades de domínios para redes de contatos como, por exemplo, a separação de contatos entre aparelhos móveis e veículos. Nesse último caso, podemos analisar em conjunto redes D2D e V2V, o que ainda não foi explorado na literatura.

Apesar da modelagem de redes móveis via redes multicamadas ser intuitiva, ainda não temos muitas abordagens. Isso deve a recente evolução da teoria de redes multicamadas [Kivelä et al. 2014]. Alguns trabalhos exploram como estruturar dados heterogêneos em domínios diferentes [Machado et al. 2015, Kivelä et al. 2014]. Em [Asgari et al. 2016], os autores consideram diferentes tipos de mobilidade, via diferentes meios de transporte, como os domínios de uma rede multicamadas. A partir dessa modelagem, eles projetaram um algoritmo para recuperar a trajetória completa de cada usuário, mesmo com amostras esparsas. O uso das redes multicamadas em aplicações móveis ainda é bastante restrito. Assim, uma direção importante de pesquisa em redes móveis é a modelagem de redes multicamadas para o projeto de aplicações.

1.4.5. Modelos Generativos

Até o momento, apresentamos modelos que apenas representam os dados observados. Como citado anteriormente, é de suma importância avaliar aplicações de redes móveis em larga escala. Apesar da era do *Big Data*, ainda encontramos dificuldade em trabalhar com amostras grandes de mobilidade [Mota et al. 2014]. Normalmente, os dados usados no projeto e avaliação de protocolos são limitados e, portanto, seu funcionamento pode ser questionado. Independente das tentativas de conseguir dados em maior escala, em geral, estamos limitados, o que pode dificultar o avanço da pesquisa na área. Todavia, um recurso muito usado em diversos contextos de simulações de sistemas é o uso de modelos estatísticos generativos. Com eles, podemos, a partir do comportamento dos dados observados, gerar mais amostras probabilisticamente semelhantes às observadas.

Um modelo estatístico generativo é um modelo para gerar aleatoriamente dados observáveis, tipicamente com parâmetros escondidos. Essa tarefa é fundamental nas áreas de Aprendizado de Máquina e Mineração de Dados, por exemplo. Existem diversas abordagens e, justamente por isso, existem diversos modelos generativos de mobilidade disponíveis na literatura [Mota et al. 2014]. De maneira simplificada, a tarefa é, a partir das observações disponíveis, gerar um modelo estatístico capaz de gerar amostras da mesma distribuição que gerou os dados observados. Além de agregar os dados observados, grande parte dos modelos utilizam premissas, buscando uma abordagem Bayesiana do assunto.

Há diversos modelos de mobilidade para diferentes contextos na literatura, porém nenhum modelo é considerado estado da arte. Em grande parte isso ocorre pelo fato de não existirem aspectos de mobilidade observados neles, mas observados em dados reais [Mota et al. 2014]. Os modelos mais famosos de mobilidade individual e veicular são apresentados em [Ekman et al. 2008, Lee et al. 2009, Mei and Stefa 2009, Rhee et al. 2011].

Para ilustrar essa difícil tarefa, apresentamos um modelo simples de mobilidade extraído apenas dos dados observados. Consideramos a rede espacial de mobilidade montada anteriormente. Isto é, construímos uma rede em que cada nó n_i representa um local no espaço e cada aresta direcionada (n_i, n_j) conecta dois nós alcançáveis espacialmente. As arestas possuem pesos $w(n_i, n_j)$ proporcionais ao fluxo de entidades que transitaram entre os dois espaços representados pelos nós. Podemos montar uma cadeia de Markov [Kemeny et al. 1960] usando a rede apresentada. Nela, cada estado da cadeia $X_1, ...X_n$ corresponde a um nó da rede. Cada probabilidade $P(X_j|X_i)$ é definida como $\frac{w(n_i,n_j)}{\sum_{k=1}^{n} w(n_i,n_k)}$. Assim, obtemos uma cadeia de Markov e podemos simular transições entre estados ou, na prática, entre espaços, levando a uma simulação de mobilidade. Na prática, essa simulação não é eficiente por diversas premissas tomadas. A principal é a de que apenas o espaço influencia o movimento das entidades. A probabilidade de movimentação não está condicionada ao tempo neste caso, o que não é verdade na prática. Por exemplo, uma pessoa tem mais probabilidade de se movimentar do centro para a periferia de uma cidade no final do dia do que no início dele. Outra premissa duvidosa é a de que o próximo lugar para o qual nos movimentamos depende apenas de onde estamos. A nossa trajetória por completo define de maneira muito mais precisa a nossa movimentação [Wu et al. 2011].

1.5. Projeto e Aplicações

Esta seção apresenta diversos estudos de casos de pesquisa atuais relacionados ao projeto de redes móveis a partir de dados sensoriados. Para cada um deles, as oportunidades de pesquisas e os desafios são também discutidos.

1.5.1. Redes Celulares e Comunicação Dispositivo-Dispositivo (D2D)

Atualmente, os telefones celulares *smartphones* passaram a ser o dispositivo computacional mais popular entre as pessoas em todos os lugares do mundo. Atualmente, eles produzem uma quantidade de dados imensa todos os dias. Devido a isso, as redes celulares se tornaram uma das principais fontes de dados e as operadoras se tornaram as detentoras desse grande base de dados. A partir da análise desses dados, pode-se melhorar significantemente o desempenho das redes celulares em termos de qualidade de serviço e maximização de receita, por meio de diferentes aplicações tais como balanceamento de carga, planejamento da rede e comunicação D2D [He et al. 2016].

Como discutido em [Zheng et al. 2016], a Figura 1.20 ilustra uma visão geral de como *Big data analytics* pode ser explorada no contexto de redes móveis celulares de comunicação e de acesso à Internet. Os dados podem ser obtidos de diversas fontes como mostra essa figura. São elas: equipamentos dos usuários (*User Equipment* – UE), rede de acesso (*Radio Access Network* – RAN), núcleo da rede (*Core Network* – CN) e provedores (*Internet Service Providers* – ISPs). O volume, a velocidade e variedade desse dados coletados é tão grande que surgem os primeiros desafios para manipulação e armazenamento desses dados. Depois de coletar e armazenar, o desafio seguinte consiste em processar e extrair o conhecimento dos dados, visando otimizar os recursos da rede de forma que os problemas sejam identificados e que seja possível decidir o quê fazer para contorná-los. A seguir, apresentamos uma lista (não exaustiva) de exemplos que envolvem o projeto de redes celulares e a área de análise de dados.

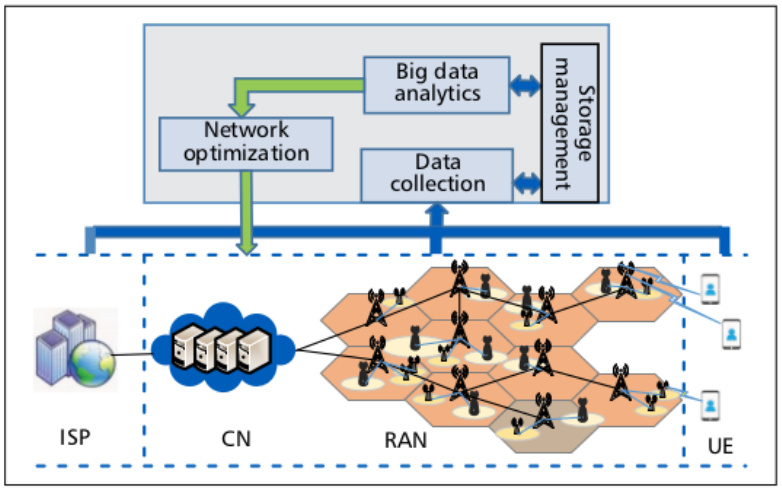


Figura 1.20. Visão geral de uma arquitetura para exploração de dados e sua aplicação em redes celulares [Zheng et al. 2016].

1.5.1.1. Alocação de Recursos e Cache de Conteúdos

Como discutido na Seção 1.3, os recursos de uma rede celular podem variar bastante em termos do domínio espaço-temporal. Além disso, anomalias como eventos sociais de larga escala (e.g., protestos, *shows*) podem ser motivos de desbalanceamento de recursos da rede causando congestionamento de tráfego na rede. Por meio da análise de dados históricos e da predição do tráfego, é possível contornar esse problema. Para tanto, técnicas relacionadas à detecção de anomalias [Chandola et al. 2009] e *Cloud-RAN* [Checko et al. 2015] podem ser combinadas e aplicadas.

Similarmente, o estudo de dados históricos pode ser aplicado a outros cenários tratando-se de redes celulares como, por exemplo, na distribuição de conteúdos. As CDNs (*Content Delivery Networks*) e ICN (*Information Centric Networks*) são redes relacionadas ao armazenamento em *cache* de conteúdos, tendo como um dos principais objetivos distribuí-los reduzindo o tempo de acesso. Uma estratégia interessante, a partir da análise de dados, é otimizar a implantação de servidores *cache* visando diminuir o tempo de resposta e reduzir custos. Por exemplo, a Figura 1.21 ilustra um caso em que servidores *cache* distribuídos podem ser implantados na infraestrutura, visando uma melhor distribuição de conteúdo na rede. Essa implantação é feita a partir da observação de padrões no uso de determinados conteúdos ao longo tempo em determinadas regiões, como foi discutido na etapa de caracterização na Figura 1.13 da Seção 1.3. Nesse sentido, conteúdos mais populares em uma dada região podem ser alocados aos servidores, sabendo que fatores associados à mobilidade, como discutidos anteriormente, também impactam essa implantação e alocação de conteúdos.

1.5.1.2. Análise de *Handovers* e Planejamento de Redes

Existem diferentes motivos para a ocorrência de *handovers* em redes celulares. No entanto, de maneira geral, a causa mais comum de *handovers* consiste da transição de uma chamada

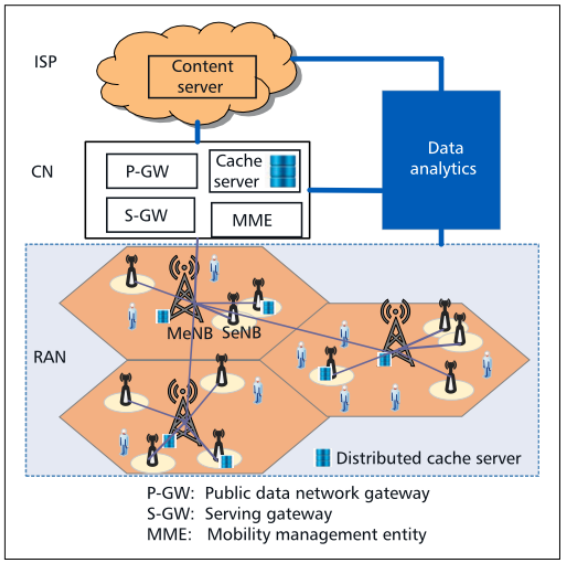


Figura 1.21. Ilustração de implantação de servidores *cache* [Zheng et al. 2016].

em curso ou sessão de dados de uma célula para outra. Em um estudo sobre mobilidade de usuários de telefonia celular com base em *handovers* [Demissie et al. 2013], os autores analisaram um conjunto de dados de Lisboa em Portugal com média de 2.5 milhões de *handovers* de um total de 7.2 milhões de chamadas.

A Figura 1.22(a) mostra o volume de chamadas em cada estação base entre 8:00 e 9:00 da manhã do dia 12 de abril de 2010. A Figura 1.22(b) mostra o fluxo de *handovers* entre as estações da rede. As setas representam as transições e a intensidade de cor representa o volume de *handovers*. Pode-se perceber que, apesar de existir uma maior quantidade de antenas na região central, a maioria das movimentações no período analisado concentra-se na região leste e parte da região norte. Do ponto de vista da operadora, é interessante observar que padrões de movimentações existem para otimizar os canais de comunicação.

A Figura 1.22(c) mostra, para cada torre, a relação de dois tipos de *handovers*: *outgoing handovers*, quando o usuário efetua uma chamada e sai da célula; e *incoming handovers*, quando o usuário entra em uma célula e finaliza a chamada. O tamanho do círculo é definido em relação ao número total de *handovers* na célula. Os autores fornecem alguns resultados que são pertinentes para o planejamento de redes celulares: torres celulares caracterizadas por alto e balanceado número de *incoming handovers* e *outgoing handovers* estão localizadas próximas às principais rodovias, visto que essas possuem um maior fluxo de pessoas. Existe uma relação entre a presença de pessoas e o número de *incoming handovers*, justificando torres próximas aos pontos de chegadas ou interesse.

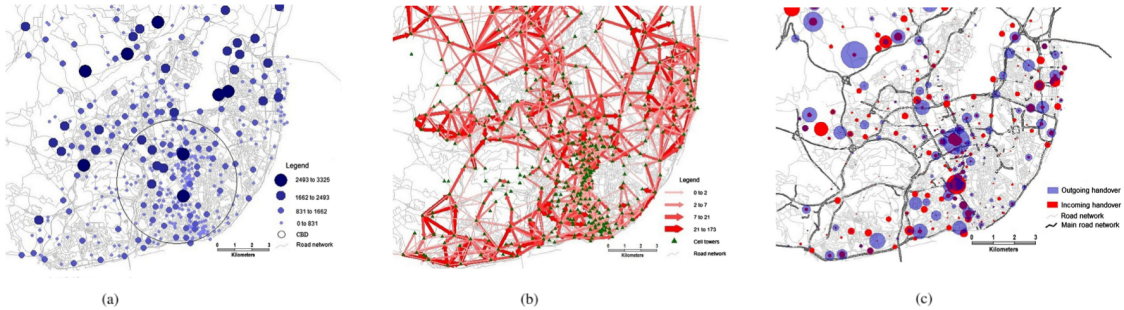


Figura 1.22. Exploração de informações de *handovers* de rede celular em um ambiente urbano. (a) Volume de chamadas por torre celular em Lisboa. (b) *Handovers* entre as torres celulares. (c) *Incoming handovers* e *Outgoing handovers* na cidade de Lisboa [Demissie et al. 2013].

1.5.1.3. Comunicação D2D

O aumento na demanda por dados nas redes celulares traz cada vez mais desafios para as operadoras de telefonia móvel. De forma a atender essa crescente demanda, as redes celulares de quinta geração (5G) propõem mudanças significativas na arquitetura de comunicação [Akyildiz et al. 2016]. Entre as novas características presentes no padrão 5G destaca-se a comunicação Dispositivo-Dispositivo ou *Device-to-Device* (D2D).

O termo D2D refere-se à transmissão direta de dados entre os dispositivos móveis da rede celular (e.g., *smartphones* e *tablets*) que estejam suficientemente próximos uns dos outros. Esse tipo de comunicação tem o potencial de descarregar a demanda das estações bases que, nas redes celulares tradicionais, são responsáveis por toda a transmissão de dados. Nas redes celulares atuais (por exemplo, 3G e 4G), quando dois dispositivos usam a Internet móvel para a compartilhar um arquivo entre si, o arquivo é primeiramente transmitido do dispositivo origem para a estação base e, posteriormente, transmitido da estação base para o dispositivo destino. Isso ocorre mesmo que os dispositivos origem e destino se encontrem lado a lado. Esse modelo, além de aumentar o atraso na comunicação (considerando transmissões D2D de apenas um salto), gera uma alta sobrecarga na estação base, que deve atender a um grande número de dispositivos ao mesmo tempo. Nesse contexto, a comunicação D2D permite o descarregamento da alta demanda por dados na estação base, possibilitando a melhoria dos serviços oferecidos aos clientes, incluindo melhores taxas de transmissão.

Como discutido em [Laya et al. 2014], o suporte à comunicação D2D modifica os planos de dados e controle das redes celulares, como explicado a seguir:

- **Plano de dados:** A comunicação D2D provê um novo caminho para os dados, que pode envolver um ou mais dispositivos móveis na rede, para enviar mensagens do dispositivo de origem para o dispositivo destino sem a necessidade dessas mensagens trafegarem pela estação base.

- **Plano de Controle:** Nas redes D2D, as mensagens de controle, que definem as políticas de encaminhamento D2D dos dados, podem ser transmitidas de forma distribuída, utilizando a própria conectividade D2D, ou de forma centralizada, com

o auxílio da estação base, que é capaz de se comunicar diretamente com todos os dispositivos móveis de sua célula. Vale ressaltar que a combinação de ambas as formas de sinalização também é possível.

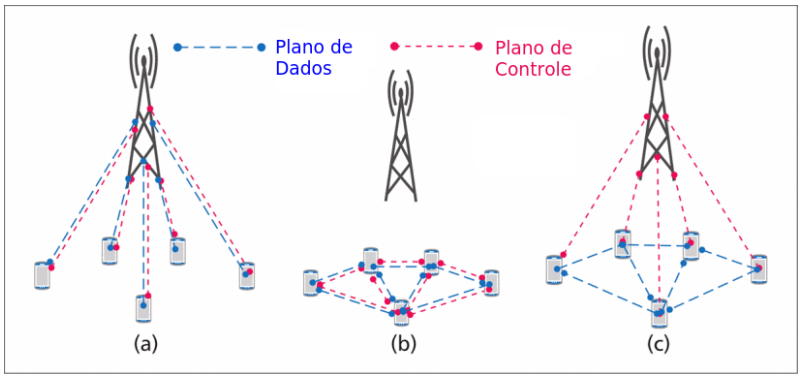


Figura 1.23. Possibilidades de arquiteturas em redes móveis e celulares: (a) Arquitetura tradicional centralizada; (b) Arquitetura *Ad Hoc* distribuída; (c) arquitetura D2D híbrida (figura adaptada de [Nunes et al. 2016c])

A Figura 1.23 mostra os tipos de arquiteturas possíveis para os planos de dados e controle. Em contraste com as redes celulares tradicionais e com as redes puramente *ad hoc* mostradas nas Figuras 1.23(a) e 1.23(b), respectivamente, a nova geração de comunicação celular oferece suporte para um plano de controle centralizado e um plano de dados distribuído, que juntos possibilitam a redução da demanda das estações bases por meio das transmissões de dados D2D (Figura 1.23(c)). A possibilidade de um plano de controle centralizado é uma diferença fundamental das redes D2D para as redes puramente *ad hoc*, em que os algoritmos de encaminhamento de dados devem ser completamente distribuídos, como mostrado na Figura 1.23(b). Vale ressaltar que as mensagens de controle são, em geral, muito menores que os dados sendo transmitidos. Por isso, mesmo que as mensagens de controle permaneçam centralizadas na estação base, a carga gerada por essas mensagens é praticamente desprezível quando comparada às taxas de transmissão requeridas por dados como vídeos, multimídia, jogos e outras aplicações.

Outro aspecto crucial nas redes D2D é a mobilidade, já que os elementos que provêm a comunicação D2D, ou seja, os próprios dispositivos móveis, são carregados por pessoas. Por sua vez, as pessoas possuem rotinas e trajetórias diárias que muitas vezes são previsíveis. Portanto, um melhor entendimento dos padrões de mobilidade das pessoas é fundamental para o projeto eficaz de protocolos para redes D2D. Nesse contexto, a análise de dados, em especial dados de mobilidade, é uma ferramenta imprescindível para extrair informações relevantes para o projeto das redes D2D, possibilitando o aprimoramento dos protocolos e serviços oferecidos aos usuários da rede.

Entre os desafios de pesquisa em redes D2D que podem se beneficiar da análise de dados de mobilidade, ressaltam-se os seguintes:

- **Alocação e gerenciamento de recursos:** Recursos em redes D2D são divididos em duas classes: recursos da rede e recursos dos dispositivos. Associados com a primeira classe estão desafios como a alocação dinâmica de canais de frequência

para a comunicação direta entre os dispositivos [Song et al. 2014, Yin et al. 2013]. Problemas de otimização como (i) alcance da transmissão versus interferência entre dispositivos; e (ii) taxas de transmissão versus largura do espectro alocado para cada par de dispositivos estão associados à primeira classe. A segunda classe trata do gerenciamento dos recursos dos dispositivos [Silva et al. 2017a, Silva et al. 2017b], como memória e energia, de forma que esses dispositivos possam colaborar com a rede possibilitando transmissões diretas de dados, mas ao mesmo tempo não sejam prejudicados pela utilização exagerada de seus recursos pela rede.

- **Comunicação D2D de múltiplos saltos e oportunística:** Nas redes D2D, a comunicação de múltiplos saltos envolve múltiplos dispositivos, formando um caminho D2D da origem de um determinado dado até o destino. Esse caminho pode ser definido oportunisticamente, utilizando os encontros entre pares de dispositivos em tempo real, de acordo com políticas de controle definidas pela estação base. Esse tipo de comunicação, em geral, envolve atrasos elevados e é adequado para alguns tipos específicos de dados, em que a entrega mais lenta dos conteúdos pode ser tolerada [Nunes et al. 2016c]. Exemplos de aplicações são a atualização de aplicações e transmissões de propagandas em vídeos. Esses tipos de conteúdos podem ser descarregados da estação base por meio da comunicação D2D oportunística de múltiplos saltos.

- **Dispositivos heterogêneos:** As redes D2D são vistas não só como uma arquitetura de comunicação entre *smartphones*, mas também como uma rede que inclui os mais diversos sistemas computacionais móveis para possibilitar uma comunicação eficiente e inteligente. Entre esses dispositivos, podemos considerar carros com capacidade de comunicação (VANETs), infraestrutura de cidades inteligentes e objetos inteligentes na Internet das Coisas.

- **Contexto social em redes D2D:** O contexto social, i.e., relações sociais entre usuários da rede podem ser utilizados, por exemplo, para prever subconjuntos de usuários da rede que têm maior probabilidade de estarem próximos uns aos outros frequentemente. Essa informação, por sua vez, pode ser utilizada para definir políticas eficientes de alocação de recursos para usuários que têm mais chance de se encontrarem e compartilharem arquivos entre si [Li et al. 2014].

No restante desta seção será apresentado um estudo de caso que exemplifica alguns dos desafios de pesquisa discutidos acima. Nesses exemplo, padrões extraídos por meio da análise de dados de mobilidade são utilizados para projetar protocolos e soluções para redes móveis D2D.

Mobilidade, Encontros e seus Papéis na Comunicação D2D

Nas redes móveis, ***encontros*** são definidos como eventos em que duas ou mais entidades (e.g., pessoas, veículos) ocupam a mesma região em um mesmo intervalo de tempo, ou seja, essas entidades encontram-se fisicamente próximas. Em outras palavras, os encontros são interseções espaço-temporais nas trajetórias móveis das entidades (veja duração de contatos e tempo entre contatos na Seção 1.3.3). Encontros são particularmente

importantes em redes D2D pois eles são as oportunidades para transmitir dados diretamente entre os dispositivos, permitindo o descarregamento da estação base. Portanto, um melhor entendimento das leis que governam os encontros entre entidades em redes móveis podem favorecer significativamente o projeto eficaz de protocolos para redes D2D.

Como discutido na Seção 1.2, um tipo de *dataset* tipicamente utilizado para esse tipo de estudo são os *traces* de contatos. Alguns exemplos de *traces* de contatos públicos são o *traces* do MIT [Eagle and Pentland 2006], de Dartmouth [Henderson et al. 2008], da USC [Hsu and Helmy 2005] e de NCCU [Tsai and Chan 2015]. Após um preprocessamento que assegure a qualidade e confiabilidade dos dados coletados, esses *traces* podem ser utilizados na análise e caracterização das propriedades dos encontros nessas redes.

Diversos trabalhos na literatura buscam caracterizar as propriedades dos encontros visando, entre outros aspectos, a aplicação dessas propriedades no projeto de redes *ad hoc* e mais recentemente de redes D2D.

Encontros de Grupos e Comunicação D2D Oportunística

A seguir, é apresentado o protocolo Groups-Net [Nunes et al. 2016d, Nunes et al. 2016a] como um estudo de caso que mostra como os padrões de mobilidade, identificados e caracterizados a partir *traces* de contatos, podem ser utilizados na modelagem e no projeto de soluções para redes D2D.

O Groups-Net é um protocolo de encaminhamento oportunístico para redes D2D de múltiplos saltos. Ele foi projetado com o objetivo de maximizar a quantidade de mensagens entregues utilizando a comunicação D2D e ao mesmo tempo minimizar as retransmissões de mensagens entre os dispositivo móveis da rede D2D. Para alcançar esse objetivo, as mensagens devem ser encaminhadas entre a origem e o destino utilizando rotas de múltiplos saltos que envolvam o menor número possível de dispositivos móveis intermediários, ou seja, rotas mais curtas. Uma alta taxa de entrega permite que mais dados sejam transmitidos usando a comunicação D2D, resultando em uma maior descarga da demanda da estação base. Por outro lado, uma grande quantidade de retransmissões impacta negativamente os dispositivos da rede, por exemplo, reduzindo a duração média de suas baterias.

O Groups-Net explora a regularidade dos encontros de grupos para realizar o roteamento das mensagens por meio de um modelo que segue as duas premissas abaixo:

- **Regularidade dos encontros de grupos:** Assume-se que grupos que se encontraram mais frequentemente no passado recente (por exemplo, cinco vezes nas últimas semanas) têm uma chance maior de se encontrarem novamente no futuro próximo do que grupos que se encontraram com uma frequência menor (por exemplo, uma vez no último mês). Para modelar esse comportamento, uma probabilidade de reencontro é assinalada a cada grupo de acordo com um processo de Poisson onde a taxa de reencontros é estimada a partir do número médio de reencontros por unidade de tempo no passado recente. A utilidade de saber quais grupos têm maior probabilidade de se reencontrarem reside no fato de que em uma reunião de um grupo, um determinada mensagem pode ser transmitida para todos os nós envolvidos.

- **Interseção nos membros de dois grupos:** Embora uma mensagem possa ser transmitida para todos os membros de um determinado grupo durante um encontro do grupo, muitas vezes a origem e o destinatário de uma determinada mensagem não pertencem a nenhum grupo em comum. Dessa forma, é necessário que a mensagem seja propaganda entre diferentes grupos. É natural que uma pessoa se encontre regularmente com mais de um grupo de pessoas (e.g., colegas de trabalho, familiares, pessoas que utilizam o mesmo ônibus no mesmo horário). Essas pessoas podem sevir para propagar as mensagens entre diferentes grupos. Com esse objetivo, a probabilidade de uma mensagem ser transmitida entre dois determinados grupos ($G1$ e $G2$) é calculada com base no número de membros que estão regularmente presentes nos encontros de ambos os grupos $G1$ e $G2$.

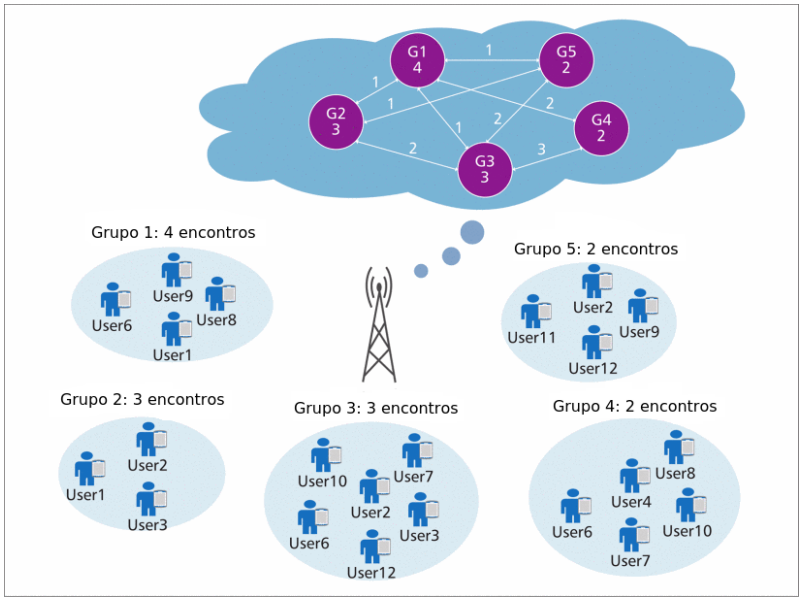


Figura 1.24. Operação do Groups-Net: A estação base mantém um grafo onde o peso dos nós representa o número de encontros de cada grupo e o peso das arestas o número de membros em comum entre dois grupos. As probabilidades de cada nó e aresta são derivadas de seus pesos e a mensagem é transmitida entre a origem e o destino pela rota grupo-a-grupo mais provável (figura adaptada de [Nunes et al. 2016c]).

Utilizando as duas probabilidades descritas acima, o Groups-Net encaminha as mensagens por meio da rota grupo-a-grupo mais provável entre a origem da mensagem e seu destinatário. Os detalhes do algoritmo de seleção da rota grupo-agrupo mais provável e a definição das fórmulas para obtenção das probabilidades descritas acima podem ser encontrados em [Nunes et al. 2016a]. Vale mencionar que a detecção dos encontros dos grupos pode ser realizada de forma distribuída, pelos próprios nós envolvidos no encontro do grupo.

Como ilustrado na Figura 1.24, o número de encontros de cada grupo é reportado (pelos próprios nós envolvidos nos encontros) para a estação base, utilizando mensagens de controle. A estação base, por sua vez, mantém um grafo em que cada nó representa um grupo. O peso das arestas nesse grafo representa o número de usuários que pertencem a

ambos os grupos interligados por uma aresta. Por exemplo, na Figura 1.24, os usuários 6, 7 e 10 pertencem a ambos os Grupos 3 e 4. Portanto a aresta interligando esses grupos tem peso 3. Por fim, o peso dos nós no grafo é definido pelo número de vezes que cada grupo se encontrou no passado recente (e.g., no último mês). Quando um dado nó origem deseja transmitir uma mensagem para um determinado destino, a estação base utiliza o grafo mostrado na Figura 1.24 para computar a rota grupo-a-grupo mais provável e envia essa rota para a origem da mensagem, que procede com o encaminhamento da mensagem de acordo com a rota recebida da estação base.

Para validar o modelo proposto em cenários reais, os autores aplicaram o Groups-Net nos *traces* de contatos do MIT e de Dartmouth. Nesses cenários, a política de encaminhamento proposta demonstrou um desempenho melhor que soluções anteriores, considerando o número de mensagens entregues com sucesso utilizando a comunicação D2D (ou seja, número de mensagens descarregadas da estação base) *versus* número de retransmissões D2D necessárias para a entrega das mensagens.

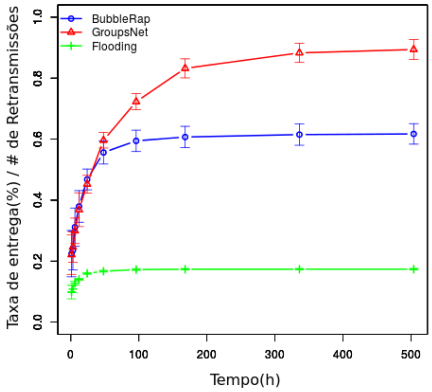


Figura 1.25. Comparação da relação benefício-custo do Groups-Net com a de protocolos de encaminhamento anteriores.

A Figura 1.25 mostra a comparação do Groups-Net com outros protocolos anteriores. Nela é apresentada a relação benefício-custo, definida como o número de mensagens entregues dividido pelo número de retransmissões necessárias para a entrega, ao longo do tempo de operação da rede. Esses resultados mostram um exemplo de como a análise e a caracterização de dados, combinadas com uma modelagem consistente do problema, podem ser aplicados para propor soluções que melhoram a eficiência de redes móveis *ad hoc* e, neste exemplo, redes D2D oportunísticas.

1.5.2. Redes Veiculares

As redes veiculares são compostas por veículos providos de capacidade de comunicação sem fio entre eles (*vehicle-to-vehicle*, ou apenas V2V), ou com estações infraestruturadas (*vehicle-to-infrastructure*, ou apenas V2I). O objetivo de uma rede veicular é prover serviços antes não viáveis para os veículos e seus usuários. As aplicações para esse tipo de rede vão desde notificações de alertas de trânsito até a entrega de conteúdo de entretenimento.

As características das redes veiculares fazem com o que o projeto da rede seja uma tarefa não-trivial. Em geral, as redes veiculares devem estar preparadas para funcionarem em larga escala, dada a enorme quantidade de veículos existentes em alguns centros urbanos. Além disso, a densidade da rede é variada, indo de topologias muito esparsas a muito densas, dependendo da região e do horário. A topologia da rede também é extremamente dinâmica devido à mobilidade dos veículos, tornando os contatos altamente intermitentes. Por fim, as cidades possuem características distintas, o que dificulta ainda mais a definição de padrões de projetos de redes.

Com todas essas dificuldades, os projetistas de redes veiculares podem se beneficiar significativamente de conhecimentos extraídos de grandes volumes de dados sobre a mobilidade urbana. No restante desta seção, são apresentados os trabalhos da literatura que exploram esses dados para obterem conhecimento útil que pode ser utilizado em diferentes aspectos das redes veiculares.

1.5.2.1. Roteamento e Disseminação de Dados

Um dos grandes desafios para os pesquisadores na área de redes veiculares está no roteamento de pacotes e disseminação de dados. As características peculiares desse tipo de rede, como larga escala, mobilidade constante, diferentes densidades ao longo do tempo e espaço, e demanda por entrega eficiente e rápida em alguns casos, fazem com que essas tarefas não sejam triviais. Portanto, nos últimos anos, muitos pesquisadores têm explorado dados de mobilidade veicular existentes para conhecer melhor a conectividade das redes veiculares e, com isso, elaborar soluções mais apropriadas. A seguir, são descritos os trabalhos que fazem uso de dados existentes para propor soluções de roteamento e disseminação em redes veiculares.

Em [Celes et al. 2013], os autores projetaram um protocolo de roteamento ciente da mobilidade diária dos veículos para realizar disseminação orientada à geo-localização (*geocast*). Para isso, primeiramente foi feita a caracterização das trajetórias de veículos de um conjunto de dados da cidade de Borlange na Suécia, com o objetivo de se conhecer os padrões de mobilidade. O protocolo realiza o *geocast* a partir das informações dos padrões de mobilidade dos veículos da rede, reduzindo o atraso do tempo de entrega e aumentando a taxa de entrega de mensagens.

Em [Monteiro et al. 2012], os autores estudaram métricas de redes complexas em redes veiculares e, com base nos resultados, sugeriram melhorias para protocolos existentes. Esse estudo avalia características dos grafos de contatos em redes veiculares urbanas e rodoviárias em termos de grau dos veículos, coeficiente de agrupamento e caminhos mínimos. Os dados utilizados são do modelo de *Automata Cellular* [Tonguz et al. 2009] e da rodovia I-80 fornecido pelo *Berkeley Highway Laboratory* [14]. Os resultados obtidos desses cenários levaram a conclusões que ajudaram a melhorar um protocolo de disseminação existente, o UV-CAST. No caso, os autores propuseram dois mecanismos: um para aumentar o alcance da rede em ambientes esparsos e outro para diminuir o impacto de mensagens *broadcast* em ambientes densos. Para isso, cada veículo avalia a quantidade de vizinhos por meio de mensagens de *beacons*. Com base neste valor, ele é capaz de inferir se está em uma região

[14] https://www.fhwa.dot.gov/publications/research/operations/06137/

com pouca densidade ou densa, e assim tomar a melhor decisão. Mais especificamente, em regiões esparsas, o veículo retransmite uma mensagem com probabilidade maior, para tentar aumentar o alcance. Por outro lado, em cenários densos, essa probabilidade diminui.

Em [Cornejo et al. 2013], os autores utilizaram a base real de táxis de São Francisco para criar um modelo de rede. Com base nesse modelo, foi proposto um algoritmo distribuído chamado *CabChat* que permite priorizar mensagens *broadcast*. O modelo e o algoritmo foram validados por meio de experimentos, que demostraram serem eficientes, e o *CabChat* foi capaz de entregar mensagens com prioridade dentro de um limite de tempo estabelecido.

O trabalho desenvolvido por [Amici et al. 2014] utilizou uma base real de mobilidade da cidade de Roma para avaliar o desempenho de um protocolo de disseminação epidêmica para redes veiculares. Os dados foram coletados de *smartphones* instalados em táxis e representam a posição de cada veículo ao longo do tempo durante um mês. A principal contribuição pode ser vista como uma nova base de dados reais que permitirá a avaliação e validação de soluções para redes veiculares.

Outro conjunto de dados de mobilidade em larga escala é utilizado por [Trullols-Cruces et al. 2015] para avaliar a disseminação epidêmica em redes veiculares. Nesse trabalho, os dados realistas da cidade de Zurique, na Suíça, são explorados para se criar um modelo de disseminação epidêmica. O modelo é então validado em outro conjunto de dados da cidade de Madri, na Espanha. Como principal conclusão, os autores observaram que mensagens maliciosas podem infectar muitos milhares de veículos em grandes áreas, em poucos minutos. Outro ponto importante está relacionado à localidade geográfica de partida da mensagem maliciosa, que pode fazer com que a infestação seja ainda mais rápida. Com base nos resultados obtidos, os autores propuseram alterações no modelo para prevenir a contaminação. A solução proposta utiliza a rede celular para atualizar veículos escolhidos de acordo com suas localizações geográficas e, assim, esses veículos seriam responsáveis por interromper a contaminação.

Em geral, a maioria dos trabalhos estudam soluções para disseminação de dados partindo da rede infraestruturada (e.g., Internet) em direção aos veículos. Por outro lado, em [Stanica et al. 2013], os autores utilizam uma base de mobilidade realista e de larga escala para propor uma solução de entrega de dados coletados pelos veículos para servidores localizados na rede infraestruturada. A solução proposta pelos autores visa escolher determinados veículos, sendo o menor número possível, com base nos resultados de contatos obtidos da análise dos dados de mobilidade. A escolha desses veículos foi modelada como o problema de *Conjunto Mínimo Dominante* em grafos, e os resultados demonstraram que é possível reduzir significativamente os dados enviados por meio de infraestrutura.

A proposta de [Liu et al. 2011] visa utilizar veículos estacionados como unidades estáticas de comunicação, com o objetivo de aumentar a conectividade da rede veicular sem a necessidade de implantação de infraestrutura que cobre 100% da área. Para validar essa ideia, os autores primeiramente avaliaram dados estatísticos sobre estacionamentos em três locais diferentes: Montreal (Canadá), Ann Harbor e Hattiesburg (EUA). Esses dados revelam que uma grande quantidade de veículos fica estacionada por um tempo significativo. Além disso, eles caracterizaram dados reais de tráfego e observaram que os

veículos parados superam significativamente os em movimento durante qualquer hora do dia. Para concluir, os autores realizaram simulações com base em dados de seis semanas da área de Chengdu, China. Os resultados mostraram que, ao se utilizar veículos estacionados como parte da rede veicular, todas as métricas de desempenho avaliadas alcançaram valores melhores.

A conectividade dos táxis de São Francisco, EUA, é explorada por [Hoque et al. 2014] para se criar um novo modelo de propagação de mensagens em redes veiculares. Para isso, os autores primeiramente definiram algoritmos para análises desses tipos de dados em larga escala. Em seguida, os algoritmos foram utilizados para analisar a conectividade dos veículos, principalmente em termos de alcance em múltiplos passos. A principal contribuição desse trabalho é a solução de análise dos dados, que permite processar grandes volumes de dados de forma eficiente em termos de espaço de armazenamento e tempo de processamento. Com isso, novas soluções que utilizam grandes volumes de dados podem ser propostas com mais facilidade.

Em [Chen et al. 2013], os autores exploram informações sobre componentes conectados de dados reais de mobilidade no projeto de algoritmos de roteamento em redes veiculares. Os dados utilizados representam as localizações de táxis das cidades de São Francisco (EUA) e Shenzhen (China). Os resultados revelaram informações importantes relativas à estabilidade e localização dos componentes maiores, que podem ser considerados os mais relevantes para a disseminação de dados. Os conhecimentos alcançados com os resultados do trabalho podem ser utilizados para a definição de algoritmos de roteamento.

1.5.2.2. Replicação e Entrega de Conteúdo

Além da disseminação de dados, outra área que vem se destacando recentemente é a entrega de conteúdo em redes veiculares [Silva et al. 2016b]. A entrega envolve, além de algoritmos para disseminar, a escolha de servidores replicadores para se formar uma rede de entrega de conteúdo (CDN) que irá atender aos clientes da maneira mais adequada. Alguns trabalhos da literatura exploram dados de mobilidade veicular existentes para a escolha de nós replicadores na rede veicular, como discutido a seguir.

Em [Gossa et al. 2008], os autores descrevem um dos primeiros trabalhos nesta área. O objetivo é propor um sistema que escolhe nós móveis replicadores visando manter um número pequeno deles sem impactar na qualidade do serviço. Para isso, os autores propuseram um modelo de predição de mobilidade que utiliza dados de mobilidade como entrada. O modelo foi avaliado em um cenário de dados reais coletados da cidade de Viena, Áustria. Os resultados revelaram que a predição da mobilidade com base em dados históricos pode melhorar a escolha de replicadores.

Em [Silva et al. 2015c], os autores exploram o conhecimento de mobilidade veicular para escolher os melhores replicadores de conteúdo em uma rede veicular. Para isso, são utilizadas informações de um modelo de mobilidade [Silva et al. 2015b] que define como os veículos se movem em termos de regiões origem-destino. Esse modelo foi desenvolvido com base em dados realistas de mobilidade veicular da cidade de Colônia, na Alemanha. Com base nesse modelo, é possível estimar os contatos dos veículos ao longo de toda a área da rede em determinados instantes de tempo e, assim, escolher bons replicadores que

sejam capazes de entregar o conteúdo a todos os veículos de maneira eficiente e com baixa redundância. Os resultados mostraram que foi possível cobrir praticamente toda a rede de maneira mais eficiente, em termos de recursos de rede, que a disseminação epidêmica.

O mesmo modelo [Silva et al. 2015b] é utilizado por [Silva et al. 2016c] para se formar uma rede de entrega de conteúdo, porém considerando conteúdos geo-referenciados, ou seja, que devem ser entregues somente dentro de uma região de interesse. Para isso, os autores propuseram um algoritmo que escolhe veículos replicadores de forma a cobrir a região de interesse sem muita redundância. Os resultados mostraram que foi possível alcançar uma boa taxa de entrega a um baixo custo de rede.

1.5.2.3. Planejamento de Infraestrutura

Uma rede veicular é composta por veículos e estações de acostamento (RSUs – *Roadside Units*). As RSUs representam pontos de acesso que os veículos podem ter com a infra-estrutura e são extremamente relevantes para se manter a cobertura da rede e garantir a qualidade dos serviços. Com isso, surge uma importante questão: onde instalar as RSUs? Considerando que o custo de instalação e manutenção é alto, pode não ser viável implantar RSUs em 100% da área de cobertura da rede. Por outro lado, regiões importantes, que não sejam cobertas por RSUs, podem enfrentar momentos de desconexão, dependendo da densidade da rede. Portanto, é importante que sejam escolhidos locais estratégicos para se implantar as RSUs, de forma a equilibrar o custo e o benefício. Esse problema também vem sendo abordado por pesquisadores nos últimos anos, sendo que alguns trabalhos utilizam dados existentes sobre as redes veiculares em suas propostas, como apresentado a seguir.

Os dados de mobilidade da cidade de Colônia, Alemanha, são explorados por [Silva et al. 2015a] para a decisão sobre onde implantar RSUs. A solução propõe a divisão da rede viária em partições similares, e considera a taxa de mobilidade entre as partições para escolher aquelas que tendem a ser mais impactadas positivamente com a implantação de RSUs. Em comparação com outros estudos, o trabalho conseguiu alcançar melhores taxas de coberturas.

O mesmo conjunto de dados é explorado por [Silva et al. 2016a], também para a implantação de RSUs em uma rede veicular. Porém, o objetivo agora é que a escolha seja baseada na capacidade da rede em entregar conteúdos de diferentes tamanhos a uma maior quantidade de veículos. Em outras palavras, o tempo de conexão dos veículos com as RSUs é fundamental para que a entrega de um conteúdo seja efetivada. Para isso, os autores propuseram uma estratégia que considera o percentual de veículos cobertos por um determinado tempo. Os resultados também foram satisfatórios, indicando que o conhecimento do cenário é fundamental para as tomadas de decisões em termos de implantação de RSUs.

1.5.2.4. Fluxo de Mobilidade

O estudo do fluxo de veículos também é uma área recente que vem sendo explorada. Ao se conhecer padrões relacionados ao fluxo da mobilidade veicular, é possível tomar decisões em termos de rotas de deslocamento, roteamento de dados e planejamento de infraestrutura, por exemplo. Os trabalhos descritos abaixo utilizam dados em larga escala para tentar identificar e prever padrões de mobilidade.

Em [Wang et al. 2014], os autores modelaram a regularidade do fluxo de tráfego de veículos nas rodovias e interseções a partir da mineração de trajetórias. A partir disso, foi elaborado um método para calcular o atraso esperado de entrega de mensagens de uma origem a um destino.

Em [Zhang et al. 2013], os autores exploram os dados reais de táxis da cidade de Shanghai para criar um modelo que identifica vias importantes para o fluxo de veículos. Com base nos pontos visitados pelos táxis em cada via, os autores observaram que aproximadamente 10% das vias são responsáveis por 90% dos pontos. Enquanto isso, 33% das vias não foi visitada nenhuma vez por algum táxi monitorado. Com o conhecimento das vias relevantes, os autores avaliaram o intervalo entre contatos dos veículos par-a-par e definiram um modelo de mobilidade que leva em consideração o fluxo de veículos observado.

O conjunto de dados de Shanghai também é estudado por [Zhu et al. 2011], principalmente em termos de intervalo entre contatos. Os autores mostraram que a chance de táxis se encontrarem é alta, diferentemente de outros estudos que fazem essa análise com base em dados de contato entre pessoas. Além disso, foram descobertas áreas populares, em que táxis visitam com frequência. Com base nessas observações, foi proposto um modelo que considera essas regiões populares e o intervalo entre contatos dos veículos.

Em [Xiao et al. 2014], os autores estudaram a regularidade na mobilidade veicular com base nos dados de táxis das cidades de Shanghai e Beijing, na China. Eles dividiram as cidades em regiões, e avaliaram a frequência de visitas dos veículos em cada região. Foram encontrados padrões de visitas em termos espaciais e temporais, que podem ser previstos com uma boa precisão, conforme a avaliação apresentada.

1.5.2.5. Considerações

Nos últimos anos, vários conjuntos de dados, reais e sintéticos, referentes à mobilidade veicular, se tornaram disponíveis publicamente. Com isso, pesquisadores começaram a explorar esses dados com o intuito de conhecer melhor a mobilidade veicular e, assim, oferecerem propostas adequadas de protocolos, aplicações e serviços. Surgiram, então, propostas que utilizam esse conhecimento para projetarem soluções de roteamento e disseminação de dados, replicação e entrega de conteúdo, implantação de infraestrutura, e de alternativas relativas ao fluxo de veículos.

No entanto, ainda há muito espaço para melhorias. Primeiramente, é importante que novos conjuntos de dados atuais sejam disponibilizados por iniciativas públicas e privadas, para que possam ser explorados por pesquisadores. Em seguida, é preciso caracterizar

esses dados para que novos conhecimentos sejam obtidos. Por fim, esses conhecimentos devem ser utilizados para que novas soluções, em diferentes aspectos, sejam propostas.

1.5.3. Internet das Coisas Móveis (IoMT)

A Internet das Coisas [Atzori et al. 2010] vem recebendo uma atenção especial tanto da academia quanto da indústria nos últimos anos e tem um papel fundamental para a convergência entre os mundos físico e informacional [Gubbi et al. 2013]. Tendo como principal característica a utilização de sensores capazes de interagir com o mundo físico, por meio de monitoramento e atuação, e a sua comunicação através da Internet, estes sensores são os "olhos" e "ouvidos" para as aplicações de IoT.

A gama de aplicações baseadas nos sensores da IoT só cresce em todo o mundo, com o potencial de geração de dados das mais variadas fontes e tipos sem precedentes. Com aplicações que vão desde automação residencial [Coronado and Iglesias 2016] e de agricultura [Wu et al. 2014a] a cidades inteligentes [Zanella et al. 2014] e monitoramento climático [Greengard 2014], a quantidade de dados geradas por estes cenários é impressionante. A Internet das Coisas Móveis (IoMT, *Internet of Mobile Things*) surge como uma expansão do conceito de IoT, ampliando ainda mais o espectro de aplicações neste novo cenário [Nahrstedt et al. 2016]. Nesta seção, discutimos as principais características e desafios que essa ampliação introduz aos desafios da IoT estática, enfatizando aspectos relacionados à mobilidade dos dispositivos e o impacto nos dados gerados.

1.5.3.1. Definição e exemplos

Podemos compreender a IoMT como uma rede composta de dispositivos móveis capazes de interagir com o mundo físico na qual estão inseridos, seja através de sensores capazes de realizar monitoramento do estado de uma dada entidade física e/ou atuadores capazes de alterar o estado de uma entidade. Este conceito deriva-se dos paradigmas que a compõem: IoT e redes móveis. Desta forma, as aplicações de IoMT podem ser consideradas como um subconjunto das aplicações de IoT. Sua principal restrição diz respeito à capacidade que seus dispositivos possuem de alterar sua localização durante sua operação.

Para exemplificar esta restrição mais formalmente, seja s um dispositivo sensor da IoMT coletando dados de uma determinada entidade, para que cada uma de suas leituras s_i seja válida para uma aplicação, é necessário considerar tanto os valores monitorados quanto as suas informações espaço-temporais. Em outras palavras, $s_i = (v,t,l)$, onde i representa a i-ésima leitura do sensor, v corresponde ao valor coletado pelo sensor, no instante de tempo t e na localização demarcada por l. Diversas aplicações que seguem estas restrições compreendem o escopo da IoMT. A seguir, apresentamos alguns exemplos para ilustrar seu potencial.

Internet dos veículos. A Internet dos Veículos (IoV – *Internet of Vehicles*) surge como uma evolução do conceito de VANETs por considerar um veículo (e.g., carros, ônibus ou trens) como uma entidade inteligente conectada à IoT [Yang et al. 2014]. Neste sentido, os veículos passam a ser vistos como sensores capazes de monitorar o ambiente no qual estão inseridos (e.g., tráfego, condições climáticas, poluição) e, além disso, capazes de

prover novos serviços inteligentes por meio da integração entre humanos, veículos, coisas e o ambiente.

Para isto, além das tecnologias de comunicação em rede, já propostas pelas VANETs, agora dispomos de novas tecnologias (e.g., *big data analytics*, *deep learning*, computação cognitiva, inteligência artificial) para a operação coordenada e inteligente dos veículos. Por meio da aquisição da grande quantidade de dados gerados por todos esses veículos, será possível a criação de serviços que visam aprimorar desde a operação da própria rede dos veículos e os serviços oferecidos aos seus passageiros, até aplicações para cidades inteligentes, como estacionamentos inteligentes, redução de tráfego e poluição.

Robôs móveis. Outra possibilidade que surge a partir da inserção de entidades inteligentes à IoT, com capacidade de mobilidade no âmbito espaço-temporal, é a utilização de robôs móveis autônomos, terrestres ou aéreos (veículos aéreos não tripulados, UAV – *Unmanned Aerial Vehicles*) para auxiliar nas tarefas cotidianas dos humanos [Chen and Hu 2012, Grieco et al. 2014].

Por meio da cooperação entre os humanos, as entidades de IoT, os robôs móveis e o ambiente, será possível a criação de aplicações nas mais diversas áreas, como saúde, agricultura, militares e de resgate em locais de risco. Neste último caso, podemos destacar a utilização de UAVs para o reconhecimento e detecção de sobreviventes em ambientes de desastres naturais. Para isto, será necessária a constante coleta e processamento de dados sobre o ambiente a fim de otimizar a coordenação dos robôs como, por exemplo, o reconhecimento e identificação de vítimas, bem como a definição de prioridades e melhores rotas [Grieco et al. 2014].

Por se tratarem de entidades potencialmente dotadas de maior capacidade computacional, os robôs móveis podem operar de forma autônoma e coordenada, tanto a partir do processamento remoto e centralizado dos dados coletados quanto pelo processamento local e comunicação entre os demais robôs móveis [Stojmenovic 2014]. Esta capacidade integra um subconjunto da IoT também denominado de *Internet of Intelligent Things* (IoIT), onde as suas entidades são autônomas, móveis, capazes de sensoriar e de atuar, indo além de simples sensores coletando dados.

Dispositivos vestíveis. Do outro lado do espectro de aplicações da IoMT encontra-se a miniaturização dos dispositivos computacionais, de forma que eles possam ser "vestidos" por seus usuários. Dispositivos como relógios inteligentes, pulseiras, ou até *smartphones*, capazes de realizar o monitoramento tanto das condições físicas de seus usuários quanto do ambiente ao seu redor, são uma fonte de dados fundamentais para aplicações que vão desde cuidados médicos até comportamento social de seus indivíduos [Silva et al. 2015d].

Uma das principais características desses dispositivos é que sua mobilidade, ao contrário do caso dos robôs móveis, não é autônoma. A movimentação desses dispositivos é dependente da movimentação de seus usuários. Desta forma, vários estudos realizados sobre a mobilidade e comportamento dos usuários podem ser aplicados [Nunes et al. 2016b], de forma que modelos de mobilidade e comunicação possam auxiliar no aperfeiçoamento das aplicações neste contexto. Neste sentido, há a chamada IoT

oportunística [Guo et al. 2013], que consiste, por exemplo, na disseminação e compartilhamento de informações de forma oportunística entre comunidades, que são formadas de acordo com a mobilidade natural dos indivíduos.

1.5.3.2. Desafios e oportunidades de *Big data analytics*

Os principais desafios encontrados pelas aplicações da IoMT são uma composição dos desafios originalmente encontrados pelas aplicações de IoT e redes móveis. Essa interseção de paradigmas impacta de diferentes formas a concepção da IoMT. A seguir destacamos alguns destes desafios e oportunidades de aplicação de estratégias de *big data analytics* em sua solução.

Heterogeneidade. A grande variedade de dispositivos, tecnologias, aplicações e seus domínios, bem como a diversidade de ambientes nos quais as entidades podem estar inseridas são aspectos fundamentais que precisam ser considerados no projeto de IoMT [Sowe et al. 2014]. Por exemplo, aplicações que precisam integrar dados coletados por sensores pode ter que lidar com diferentes padrões e implementações. Isso pode gerar dados de diferentes tipos, escalas e granularidades, o que requer um maior esforço na etapa de pré-processamento [Wu et al. 2014b], de forma a abstrair tais diferenças e sincronizar os dados de forma confiável.

Escalabilidade. Da mesma forma como ocorre com a IoT, desenvolvedores de aplicações para a IoMT devem estar preparados para lidar com a impressionante quantidade de dispositivos e dados gerados pela IoMT [Ma et al. 2013]. Por exemplo, simples algoritmos de agregação de dados podem estar sujeitos a problemas de escalabilidade se não considerarem que estes números podem crescer de forma mais acentuada que um crescimento linear. Estratégias que precisem armazenar os dados gerados para mineração e processamento histórico, assim como aquelas que efetuem processamento *online* e em tempo real de *streamings* de dados, precisarão ser capazes de garantir algum nível de escalabilidade [Tsai et al. 2014].

Restrições computacionais. Apesar de não ser uma regra para todas as aplicações de IoT, e agora IoMT, a probabilidade de se deparar com dispositivos com limitadas capacidades computacionais não pode ser descartada. Tanto para sensores quanto atuadores, uma característica herdada das redes de sensores sem fio [Loureiro et al. 2003], são as limitações de processamento, memória, comunicação e energia. Isso ocorre pois, ao contrário dos cenários mais comumente encontrados no mundo informacional, os dispositivos que interagem com o mundo físico geralmente são pequenos e, consequentemente, com limitados recursos.

Conectividade. A possibilidade de mobilidade dos dispositivos, ao contrário da maioria das aplicações em IoT, traz à IoMT um novo aspecto de dinamicidade necessário de ser considerado. Muitas aplicações podem se beneficiar deste novo aspecto de mobilidade,

com o potencial de dados gerados em larga escala, tanto em quantidade quanto cobertura [Musolesi 2014]. Mas, devido à movimentação dos dispositivos, que geralmente são dotados de comunicação sem fio de curto alcance, sua conectividade está diretamente relacionada com a manutenção de um raio de distância suficiente para a comunicação com outros dispositivos. Assim, a conectividade desses dispositivos passa a ser intermitente, fazendo com que dados possam ser entregues com variações de atraso ou até mesmo perdidos. O estudo destes padrões de mobilidade e a sua aplicação na predição e otimização da utilização dos dados [Nunes et al. 2016b] pode ser de grande valia para a IoMT.

Disponibilidade. Como decorrência da intermitente conectividade dos dispositivos da IoMT, se um determinado sensor está sendo utilizado como fonte de dados para uma aplicação, não é garantido que este sensor esteja disponível durante toda a sua operação. Isto pode afetar diretamente a qualidade das aplicações baseadas nos dados da IoMT [Borges Neto et al. 2015]. Isto pode desencadear a constante necessidade de execução de processos de descoberta de novos sensores, agregando um custo maior de tempo e recursos para as suas aplicações [Perera et al. 2014].

Localidade espaço-temporal. Por se tratar de dispositivos móveis gerando dados, tão importante quanto o dado gerado são as informações de tempo e localização, que informe quando e onde o dado foi gerado. Contudo, por se tratarem de dispositivos pequenos com restrições computacionais, nem sempre estarão dotados de *hardware* específicos para detectar sua localização (e.g., GPS) [Loureiro et al. 2003]. Assim, a informação de localidade pode ter que ser obtida de forma indireta, seja pelas interações desses dispositivos com outras entidades, como pontos de acesso ou outros dispositivos que possuam tal capacidade. Assim, registros de proximidade que podem ser utilizados na inferência da localização destes dispositivos, enquanto se movimentam e interagem com outros dispositivos, pode se aplicar [Cunha et al. 2016a].

1.6. Considerações Finais

Este capítulo apresentou os principais conceitos, técnicas e aplicações relacionadas ao projeto de redes móveis, infraestruturadas ou *ad hoc*, dirigido a dados. Para tanto, foi proposto e discutido um arcabouço que engloba diferentes etapas do processamento do dado, desde a sua obtenção até a sua aplicação no projeto de redes. Esse arcabouço foi definido com base em diversos exemplos de propostas discutidas na literatura e tem como objetivo definir um fluxo que permite aplicar os conhecimentos extraídos dos dados no projeto de redes móveis. Dessa forma, este capítulo fornece tanto os primeiros direcionamentos para introduzir novos interessados na área, assim como inspirá-los em pesquisas futuras.

Nesse sentido, diversas oportunidades e desafios de pesquisa foram discutidos ao longo do texto visto que essa área tem recebido bastante atenção nas principais conferências. Por exemplo, a conferência SIGCOMM possui um *workshop* chamado BIG-DAMA[15] que trata exclusivamente do tópico sobre *big data analytics* e aprendizagem de máquina para redes de comunicação. A conferência PERCOM possui um *workshop* chamado DAMN![16]

[15]Big-DAMA Workshop: http://conferences.sigcomm.org/sigcomm/2017/workshop-big-dama.html
[16]DAMN Workshop: http://damn2017.conf.citi-lab.fr/

que trata puramente de tópicos relacionados a *data analytics* para redes móveis. Além desses *workshops*, as trilhas principais nessas e outras conferências da área têm apresentado seções relevantes no âmbito do tópico deste capítulo.

Para novas pesquisas, é importante termos novas fontes de dados que estejam disponíveis publicamente, para podermos aplicar algoritmos de aprendizagem de máquina e mineração de dados considerando as peculiaridades das redes móveis. Finalmente, questões de segurança dos dados e aspectos práticos que considerem a importância da análise de dados na concepção de redes móveis também devem ser tratadas.

Referências

[Aggarwal and Han 2014] Aggarwal, C. C. and Han, J. (2014). *Frequent pattern mining*. Springer.

[Akyildiz et al. 2016] Akyildiz, I. F., Nie, S., Lin, S.-C., and Chandrasekaran, M. (2016). 5g roadmap: 10 key enabling technologies. *Computer Networks*, 106:17–48.

[Amici et al. 2014] Amici, R., Bonola, M., Bracciale, L., Rabuffi, A., Loreti, P., and Bianchi, G. (2014). Performance assessment of an epidemic protocol in vanet using real traces. *Procedia Computer Science*, 40:92–99.

[Asgari et al. 2016] Asgari, F., Sultan, A., Xiong, H., Gauthier, V., and El-Yacoubi, M. A. (2016). Ct-mapper: mapping sparse multimodal cellular trajectories using a multilayer transportation network. *Computer Communications*, 95:69–81.

[Atzori et al. 2010] Atzori, L., Iera, A., and Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805.

[Barbera et al. 2013] Barbera, M. V., Epasto, A., Mei, A., Perta, V. C., and Stefa, J. (2013). Signals from the crowd: uncovering social relationships through smartphone probes. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 265–276. ACM.

[Barchiesi et al. 2015] Barchiesi, D., Preis, T., Bishop, S., and Moat, H. S. (2015). Modelling human mobility patterns using photographic data shared online. *Royal Society open science*, 2(8):150046.

[Barthélemy 2011] Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1):1–101.

[Benevenuto et al. 2009] Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. (2009). Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 49–62. ACM.

[Bi et al. 2015] Bi, S., Zhang, R., Ding, Z., and Cui, S. (2015). Wireless communications in the era of big data. *IEEE Communications Magazine*, 53(10):190–199.

[Blondel et al. 2015] Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10.

[Blondel et al. 2012] Blondel, V. D., Esch, M., Chan, C., Clérot, F., Deville, P., Huens, E., Morlot, F., Smoreda, Z., and Ziemlicki, C. (2012). Data for development: the d4d challenge on mobile phone data. *arXiv preprint arXiv:1210.0137*.

[Borges Neto et al. 2015] Borges Neto, J., Silva, T., Assunção, R., Mini, R., and Loureiro, A. (2015). Sensing in the Collaborative Internet of Things. *Sensors*, 15(3):6607–6632.

[Bracciale et al. 2014] Bracciale, L., Bonola, M., Loreti, P., Bianchi, G., Amici, R., and Rabuffi, A. (2014). CRAWDAD data set roma/taxi (v. 2014-07-17). Downloaded from http://crawdad.org/roma/taxi/.

[Brockmann et al. 2006] Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465.

[Calabrese et al. 2011] Calabrese, F., Di Lorenzo, G., Liu, L., and Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4):36–44.

[Calabrese et al. 2013] Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., and Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313.

[Cattell 2011] Cattell, R. (2011). Scalable sql and nosql data stores. *SIGMOD Rec.*, 39(4):12–27.

[Celes et al. 2013] Celes, C., Braga, R. B., Oliveira, C. T. D., Andrade, R. M. C., and Loureiro, A. A. F. (2013). Geospin: An approach for geocast routing based on spatial information in vanets. In *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pages 1–6.

[Centellegher et al. 2016] Centellegher, S., De Nadai, M., Caraviello, M., Leonardi, C., Vescovi, M., Ramadian, Y., Oliver, N., Pianesi, F., Pentland, A., Antonelli, F., et al. (2016). The mobile territorial lab: a multilayered and dynamic view on parents' daily lives. *EPJ Data Science*, 5(1):3.

[Chandola et al. 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15.

[Checko et al. 2015] Checko, A., Christiansen, H. L., Yan, Y., Scolari, L., Kardaras, G., Berger, M. S., and Dittmann, L. (2015). Cloud ran for mobile networks—a technology overview. *IEEE Communications surveys & tutorials*, 17(1):405–426.

[Chen et al. 2014a] Chen, M., Mao, S., and Liu, Y. (2014a). Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209.

[Chen and Hu 2012] Chen, Y. and Hu, H. (2012). Internet of intelligent things and robot as a service. *Simulation Modelling Practice and Theory*, 34:159–171.

[Chen et al. 2013] Chen, Y., Xu, M., Gu, Y., Li, P., and Cheng, X. (2013). Understanding topology evolving of vanets from taxi traces. *Adv. Sci. Technol. Lett*, 42(Mobile and Wireless):13–17.

[Chen et al. 2014b] Chen, Y., Xu, M., Gu, Y., Li, P., Shi, L., and Xiao, X. (2014b). Empirical study on spatial and temporal features for vehicular wireless communications. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):1–12.

[Cheng et al. 2017] Cheng, X., Fang, L., Hong, X., and Yang, L. (2017). Exploiting mobile big data: Sources, features, and applications. *IEEE Network*, 31(1):72–79.

[Chu et al. 2016] Chu, X., Ilyas, I. F., Krishnan, S., and Wang, J. (2016). Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2201–2206. ACM.

[Çolak et al. 2016] Çolak, S., Lima, A., and González, M. C. (2016). Understanding congested travel in urban areas. *Nature communications*, 7.

[Cormode and Srivastava 2009] Cormode, G. and Srivastava, D. (2009). Anonymized data: generation, models, usage. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pages 1015–1018. ACM.

[Cornejo et al. 2013] Cornejo, A., Newport, C., Gollakota, S., Rao, J., and Giuli, T. J. (2013). Prioritized gossip in vehicular networks. *Ad Hoc Networks*, 11(1):397–409.

[Coronado and Iglesias 2016] Coronado, M. and Iglesias, C. A. (2016). Task automation services: Automation for the masses. *IEEE Internet Computing*, 20(1):52–58.

[Cunha et al. 2016a] Cunha, F. D., Alvarenga, D. A., Maia, G., Viana, A. C., Mini, R. A., and Loureiro, A. A. (2016a). Exploring interactions in vehicular networks. In *Proceedings of the 14th ACM International Symposium on Mobility Management and Wireless Access*, pages 131–138. ACM.

[Cunha et al. 2016b] Cunha, F. D., Silva, F. A., Celes, C., Maia, G., Ruiz, L. B., Andrade, R. M., Mini, R. A., Boukerche, A., and Loureiro, A. A. (2016b). Communication analysis of real vehicular calibrated traces. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–6. IEEE.

[de Melo et al. 2015] de Melo, P. O. V., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Le Mouël, F., Loureiro, A. A., Addepalli, L., and Guangshuo, C. (2015). Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19–36.

[de Montjoye et al. 2014] de Montjoye, Y.-A., Smoreda, Z., Trinquart, R., Ziemlicki, C., and Blondel, V. D. (2014). D4d-senegal: the second mobile phone data for development challenge. *arXiv preprint arXiv:1407.4885*.

[Dean and Ghemawat 2008] Dean, J. and Ghemawat, S. (2008). Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113.

[Demissie et al. 2013] Demissie, M. G., de Almeida Correia, G. H., and Bento, C. (2013). Exploring cellular network handover information for urban mobility analysis. *Journal of Transport Geography*, 31:164–170.

[Eagle and Pentland 2006] Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268.

[Eagle and Pentland 2005] Eagle, N. and Pentland, A. S. (2005). CRAWDAD dataset mit/reality (v. 2005-07-01). Downloaded from http://crawdad.org/mit/reality/20050701.

[Ekman et al. 2008] Ekman, F., Keränen, A., Karvo, J., and Ott, J. (2008). Working day movement model. In *Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 33–40. ACM.

[Ficek and Kencl 2010] Ficek, M. and Kencl, L. (2010). Spatial extension of the reality mining dataset. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on*, pages 666–673. IEEE.

[Freedman et al. 2007] Freedman, D., Pisani, R., and Purves, R. (2007). *Statistics*. W. W. Norton and Co, 4th edition.

[Gao et al. 2009] Gao, W., Li, Q., Zhao, B., and Cao, G. (2009). Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, pages 299–308. ACM.

[Garton et al. 1997] Garton, L., Haythornthwaite, C., and Wellman, B. (1997). Studying online social networks. *Journal of Computer-Mediated Communication*, 3(1):0–0.

[Gonzalez et al. 2008] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

[Gossa et al. 2008] Gossa, J., Janecek, A. G., Hummel, K. A., Gansterer, W. N., and Pierson, J.-M. (2008). Proactive replica placement using mobility prediction. In *Mobile Data Management Workshops, 2008. MDMW 2008. Ninth International Conference on*, pages 182–189. IEEE.

[Gramaglia and Fiore 2014] Gramaglia, M. and Fiore, M. (2014). On the anonymizability of mobile traffic datasets. *arXiv preprint arXiv:1501.00100*.

[Greengard 2014] Greengard, S. (2014). Weathering a New Era of Big Data. *Association for Computing Machinery. Communications of the ACM*, 57(9):12.

[Gregory 2010] Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10):103018.

[Grieco et al. 2014] Grieco, L., Rizzo, A., Colucci, S., Sicari, S., Piro, G., Di Paola, D., and Boggia, G. (2014). IoT-aided robotics applications: Technological implications, target domains and open issues. *Computer Communications*, 54:32–47.

[Gubbi et al. 2013] Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660.

[Guidotti et al. 2014] Guidotti, R., Monreale, A., Rinzivillo, S., Pedreschi, D., and Giannotti, F. (2014). Retrieving points of interest from human systematic movements. In *International Conference on Software Engineering and Formal Methods*, pages 294–308. Springer.

[Guo et al. 2013] Guo, B., Zhang, D., Wang, Z., Yu, Z., and Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, 36(6):1531–1539.

[Han et al. 2011] Han, J., Kamber, M., and Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.

[Hand et al. 2001] Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. MIT press.

[He et al. 2016] He, Y., Yu, F. R., Zhao, N., Yin, H., Yao, H., and Qiu, R. C. (2016). Big data analytics in mobile cellular networks. *IEEE Access*, 4:1985–1996.

[Henderson et al. 2008] Henderson, T., Kotz, D., and Abyzov, I. (2008). The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712.

[Hess et al. 2016] Hess, A., Hummel, K. A., Gansterer, W. N., and Haring, G. (2016). Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. *ACM Computing Surveys (CSUR)*, 48(3):38.

[Hidalgo and Rodriguez-Sickert 2008] Hidalgo, C. A. and Rodriguez-Sickert, C. (2008). The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024.

[Holme and Saramäki 2012] Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.

[Hoque et al. 2014] Hoque, M. A., Hong, X., and Dixon, B. (2014). Efficient multihop connectivity analysis in urban vehicular networks. *Vehicular Communications*, 1(2):78–90.

[Hou et al. 2016] Hou, X., Li, Y., Jin, D., Wu, D. O., and Chen, S. (2016). Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments. *IEEE Transactions on Vehicular Technology*, 65(4):2753–2758.

[Hsu and Helmy 2005] Hsu, W.-j. and Helmy, A. (2005). Impact: Investigation of mobile-user patterns across university campuses using wlan trace analysis. *arXiv preprint cs/0508009*.

[Hui et al. 2011] Hui, P., Crowcroft, J., and Yoneki, E. (2011). Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11):1576–1589.

[Imran et al. 2014] Imran, A., Zoha, A., and Abu-Dayya, A. (2014). Challenges in 5g: how to empower son with big data for enabling 5g. *IEEE Network*, 28(6):27–33.

[Jain 2010] Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.*, 31(8):651–666.

[James et al. 2014] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

[Jiang et al. 2016] Jiang, S., Ferreira, J., and Gonzales, M. C. (2016). Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*.

[Kamienski et al. 2016] Kamienski, C., Biondi, G. O., Borelli, F. F., Heideker, A., Ratusznei, J., and Kleinschmidt, J. H. (2016). Computação urbana: Tecnologias e aplicações para cidades inteligentes. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Karkouch et al. 2016] Karkouch, A., Mousannif, H., Al Moatassime, H., and Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73:57–81.

[Kemeny et al. 1960] Kemeny, J. G., Snell, J. L., et al. (1960). *Finite markov chains*, volume 356. van Nostrand Princeton, NJ.

[Kivelä et al. 2014] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, 2(3):203–271.

[Lambiotte et al. 2008] Lambiotte, R., Blondel, V. D., De Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., and Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325.

[Laya et al. 2014] Laya, A., Wang, K., Widaa, A. A., Alonso-Zarate, J., Markendahl, J., and Alonso, L. (2014). Device-to-device communications and small cells: enabling spectrum reuse for dense networks. *IEEE Wireless Communications*, 21(4):98–105.

[Lee et al. 2009] Lee, K., Hong, S., Kim, S. J., Rhee, I., and Chong, S. (2009). Slaw: A new mobility model for human walks. In *INFOCOM 2009, IEEE*, pages 855–863. IEEE.

[Leguay and Benbadis 2009] Leguay, J. and Benbadis, F. (2009). Crawdad data set upmc/rollernet (v. 2009-02-02).

[Leguay et al. 2006] Leguay, J., Lindgren, A., Scott, J., Friedman, T., and Crowcroft, J. (2006). Opportunistic content distribution in an urban setting. In *Proceedings of the 2006 SIGCOMM workshop on Challenged networks*, pages 205–212. ACM.

[Leskovec et al. 2009] Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123.

[Li et al. 2014] Li, Y., Wu, T., Hui, P., Jin, D., and Chen, S. (2014). Social-aware d2d communications: qualitative insights and quantitative analysis. *IEEE Communications Magazine*, 52(6):150–158.

[Liu et al. 2011] Liu, N., Liu, M., Lou, W., Chen, G., and Cao, J. (2011). Pva in vanets: Stopped cars are not silent. In *INFOCOM, 2011 Proceedings IEEE*, pages 431–435. IEEE.

[Loureiro et al. 2003] Loureiro, A. A. F., Nogueira, J. M. S., Ruiz, L. B., Mini, R. A. d. F., Nakamura, E. F., and Figueiredo, C. M. S. (2003). Redes de Sensores Sem Fio. *Simpósio Brasileiro de Redes de Computadores (SBRC)*, pages 179–226.

[Lu et al. 2016] Lu, X., Yu, Z., Sun, L., Liu, C., Xiong, H., and Guan, C. (2016). Characterizing the life cycle of point of interests using human mobility patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1052–1063. ACM.

[Ma et al. 2013] Ma, M., Wang, P., and Chu, C.-H. (2013). Data Management for Internet of Things: Challenges, Approaches and Opportunities. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pages 1144–1151. IEEE.

[Machado et al. 2015] Machado, K., Silva, T. H., de Melo, P. O. V., Cerqueira, E., and Loureiro, A. A. (2015). Urban mobility sensing analysis through a layered sensing approach. In *Mobile Services (MS), 2015 IEEE International Conference on*, pages 306–312. IEEE.

[Mamei et al. 2016] Mamei, M., Colonna, M., and Galassi, M. (2016). Automatic identification of relevant places from cellular network data. *Pervasive and Mobile Computing*, 31:147–158.

[Mei and Stefa 2009] Mei, A. and Stefa, J. (2009). Swim: A simple model to generate small mobile worlds. In *INFOCOM 2009, IEEE*, pages 2106–2113. IEEE.

[Michel and Julien 2016] Michel, J. and Julien, C. (2016). From human mobility to data mobility: Leveraging spatiotemporal history in device-to-device information diffusion. In *Mobile Data Management (MDM), 2016 17th IEEE International Conference on*, volume 1, pages 198–207. IEEE.

[Monteiro et al. 2012] Monteiro, R., Sargento, S., Viriyasitavat, W., and Tonguz, O. K. (2012). Improving vanet protocols via network science. In *Vehicular Networking Conference (VNC), 2012 IEEE*, pages 17–24. IEEE.

[Mota et al. 2014] Mota, V. F., Cunha, F. D., Macedo, D. F., Nogueira, J. M., and Loureiro, A. A. (2014). Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, 48:5–19.

[Musolesi 2014] Musolesi, M. (2014). Big Mobile Data Mining: Good or Evil? *IEEE Internet Computing*, 18(1):78–81.

[Naboulsi and Fiore 2016] Naboulsi, D. and Fiore, M. (2016). Characterizing the instantaneous connectivity of large-scale urban vehicular networks. *IEEE Transactions on Mobile Computing*.

[Naboulsi et al. 2016] Naboulsi, D., Fiore, M., Ribot, S., and Stanica, R. (2016). Large-scale mobile traffic analysis: a survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161.

[Nahrstedt et al. 2016] Nahrstedt, K., Li, H., Nguyen, P., Chang, S., and Vu, L. (2016). Internet of Mobile Things: Mobility-Driven Challenges, Designs and Implementations. In *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 25–36. IEEE.

[Nguyen et al. 2011] Nguyen, N. P., Dinh, T. N., Tokala, S., and Thai, M. T. (2011). Overlapping communities in dynamic networks: their detection and mobile applications. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 85–96. ACM.

[Nika et al. 2016] Nika, A., Ismail, A., Zhao, B. Y., Gaito, S., Rossi, G. P., and Zheng, H. (2016). Understanding and predicting data hotspots in cellular networks. *Mobile Networks and Applications*, 21(3):402–413.

[Nunes et al. 2016a] Nunes, I. O., Celes, C., Vaz de Melo, P. O., and Loureiro, A. A. (2016a). Groups-net: Group meetings aware routing in multi-hop d2d networks. *arXiv preprint arXiv:1605.07692*.

[Nunes et al. 2016b] Nunes, I. O., de Melo, P. O. S. V., and Loureiro, A. A. F. (2016b). Group mobility: Detection, tracking and characterization. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6.

[Nunes et al. 2016c] Nunes, I. O., de Melo, P. O. V., and Loureiro, A. A. (2016c). Leveraging d2d multihop communication through social group meeting awareness. *IEEE Wireless Communications*, 23(4):12–19.

[Nunes et al. 2016d] Nunes, I. O., de Melo, P. O. V., and Loureiro, A. A. F. (2016d). Groups-net: Roteamento ciente de encontros de grupos em redes móveis d2d. In *Proceedings of the Brazilian Symposium on Computer Networks and Distributed Systems*, Salvador, Bahia.

[Palla et al. 2005] Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.

[Pappalardo et al. 2015] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6.

[Perera et al. 2014] Perera, C., Zaslavsky, A., Liu, C. H., Compton, M., Christen, P., and Georgakopoulos, D. (2014). Sensor Search Techniques for Sensing as a Service Architecture for the Internet of Things. *IEEE Sensors Journal*, 14(2):406–420.

[Pietilainen and Diot 2012] Pietilainen, A.-K. and Diot, C. (2012). CRAWDAD dataset thlab/sigcomm2009 (v. 2012-07-15). Downloaded from http://crawdad.org/thlab/sigcomm2009/20120715.

[Piorkowski et al. 2009a] Piorkowski, M., Sarafijanovic-Djukic, N., and Grossglauser, M. (2009a). CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.org/epfl/mobility/.

[Piorkowski et al. 2009b] Piorkowski, M., Sarafijanovoc-Djukic, N., and Grossglauser, M. (2009b). A Parsimonious Model of Mobile Partitioned Networks with Clustering. In *The First International Conference on COMmunication Systems and NETworkS (COMSNETS)*.

[Pires et al. 2015] Pires, P. F., Delicato, F., Batista, T., Barros, T., Cavalcante, E., and Pitanga, M. (2015). Plataformas para a internet das coisas. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Rebecchi et al. 2015] Rebecchi, F., De Amorim, M. D., Conan, V., Passarella, A., Bruno, R., and Conti, M. (2015). Data offloading techniques in cellular networks: a survey. *IEEE Communications Surveys & Tutorials*, 17(2):580–603.

[Rhee et al. 2011] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., and Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking (TON)*, 19(3):630–643.

[Santos et al. 2016] Santos, B. P., Silva, L. A., Celes, C. S., Borges, J. B., Neto, B. S. P., Vieira, M. A. M., Vieira, L. F. M., Goussevskaia, O. N., and Loureiro, A. A. (2016). Internet das coisas: da teoria à prática. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Scott et al. 2009] Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C., and Chaintreau, A. (2009). CRAWDAD dataset cambridge/haggle (v. 2009-05-29). Downloaded from http://crawdad.org/cambridge/haggle/20090529.

[Shafiq et al. 2015] Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., and Wang, J. (2015). Geospatial and temporal dynamics of application usage in cellular data networks. *IEEE Transactions on Mobile Computing*, 14(7):1369–1381.

[Shumway and Stoffer 2010] Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

[Silva et al. 2015a] Silva, C. M., Aquino, A. L., and Meira, W. (2015a). Deployment of roadside units based on partial mobility information. *Computer Communications*, 60:28–39.

[Silva et al. 2016a] Silva, C. M., Silva, F. A., Sarubbi, J. F., Oliveira, T. R., Meira, W., and Nogueira, J. M. S. (2016a). Designing mobile content delivery networks for the internet of vehicles. *Vehicular Communications*.

[Silva et al. 2016b] Silva, F. A., Boukerche, A., Silva, T. R., Ruiz, L. B., Cerqueira, E., and Loureiro, A. A. (2016b). Vehicular networks: A new challenge for content-delivery-based applications. *ACM Computing Surveys (CSUR)*, 49(1):11.

[Silva et al. 2015b] Silva, F. A., Boukerche, A., Silva, T. R., Ruiz, L. B., and Loureiro, A. A. (2015b). A novel macroscopic mobility model for vehicular networks. *Computer Networks*, 79:188–202.

[Silva et al. 2015c] Silva, F. A., Boukerche, A., Silva, T. R. B., Benevenuto, F., Ruiz, L. B., and Loureiro, A. A. (2015c). Odcrep: Origin–destination-based content replication for vehicular networks. *IEEE Transactions on Vehicular Technology*, 64(12):5563–5574.

[Silva et al. 2016c] Silva, F. A., Boukerche, A., Silva, T. R. B., Ruiz, L. B., and Loureiro, A. A. (2016c). Geo-localized content availability in vanets. *Ad Hoc Networks*, 36:425–434.

[Silva et al. 2017a] Silva, M., de Oliveira Nunes, I., Loureiro, A. A. F., and Mini, R. (2017a). St-drop: Uma nova estratégia de gerenciamento de buffer em redes d2d oportunistas. In *SBRC 2017*.

[Silva et al. 2017b] Silva, M., Nunes, I., Mini, R. A. F., and A.F. Loureiro, A. (2017b). ST-Drop: a novel buffer management strategy for D2D opportunistic networks. In *22nd IEEE Symposium on Computers and Communication (ISCC 2017) (ISCC 2017)*, Heraklion, Greece.

[Silva et al. 2014a] Silva, T. H., De Melo, P. O. V., Almeida, J. M., and Loureiro, A. A. (2014a). Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51.

[Silva et al. 2015d] Silva, T. H., de Melo, P. O. V., Neto, J. B., IJT, A., Ribeiro, C. S. d. S., Mota, V. F., da Cunha, F. D., Ferreira, A. P., Machado, K. L. d. S., Mini, R. A. d. F., et al. (2015d). Redes de sensoriamento participativo: Desafios e oportunidades. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Silva et al. 2014b] Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., and Loureiro, A. A. (2014b). Revealing the city that we cannot see. *ACM Transactions on Internet Technology (TOIT)*, 14(4):26.

[Silveira et al. 2016] Silveira, L. M., de Almeida, J. M., Marques-Neto, H. T., Sarraute, C., and Ziviani, A. (2016). Mobhet: Predicting human mobility using heterogeneous data sources. *Computer Communications*.

[Socievole et al. 2014] Socievole, A., De Rango, F., and Caputo, A. (2014). Wireless contacts, facebook friendships and interests: Analysis of a multi-layer social network in an academic environment. In *Wireless Days (WD), 2014 IFIP*, pages 1–7. IEEE.

[Song et al. 2014] Song, L., Niyato, D., Han, Z., and Hossain, E. (2014). Game-theoretic resource allocation methods for device-to-device communication. *IEEE Wireless Communications*, 21(3):136–144.

[Soto et al. 2011] Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 377–388. Springer.

[Sowe et al. 2014] Sowe, S. K., Kimata, T., Dong, M., and Zettsu, K. (2014). Managing Heterogeneous Sensor Data on a Big Data Platform: IoT Services for Data-Intensive Science. In *2014 IEEE 38th International Computer Software and Applications Conference Workshops*, pages 295–300. IEEE.

[Stanica et al. 2013] Stanica, R., Fiore, M., and Malandrino, F. (2013). Offloading floating car data. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, pages 1–9. IEEE.

[Stojmenovic 2014] Stojmenovic, I. (2014). Machine-to-Machine Communications With In-Network Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical Systems. *IEEE Internet of Things Journal*, 1(2):122–128.

[SUVnet] SUVnet. Shanghai data trace. Online (available at http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html).

[Tanahashi et al. 2012] Tanahashi, Y., Rowland, J. R., North, S., and Ma, K.-L. (2012). Inferring human mobility patterns from anonymized mobile communication usage. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*, pages 151–160. ACM.

[Teles et al. 2013] Teles, A., Pinheiro, D., Gonçalves, J., Batista, R., Almeida, V., Endler, M., and Silva, F. (2013). Redes sociais móveis: conceitos, aplicações e aspectos de segurança e privacidade. *31º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*.

[Tizzoni et al. 2014] Tizzoni, M., Bajardi, P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol*, 10(7):e1003716.

[Tonguz et al. 2009] Tonguz, O. K., Viriyasitavat, W., and Bai, F. (2009). Modeling urban traffic: a cellular automata approach. *IEEE Communications Magazine*, 47(5).

[Tostes et al. 2013] Tostes, A. I. J., de LP Duarte-Figueiredo, F., Assunção, R., Salles, J., and Loureiro, A. A. (2013). From data to knowledge: City-wide traffic flows analysis and prediction using bing maps. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 12. ACM.

[Trasarti et al. 2011] Trasarti, R., Pinelli, F., Nanni, M., and Giannotti, F. (2011). Mining mobility user profiles for car pooling. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1190–1198. ACM.

[Trestian et al. 2009] Trestian, I., Ranjan, S., Kuzmanovic, A., and Nucci, A. (2009). Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 267–279. ACM.

[Trullols-Cruces et al. 2015] Trullols-Cruces, O., Fiore, M., and Barcelo-Ordinas, J. M. (2015). Worm epidemics in vehicular networks. *IEEE Transactions on Mobile Computing*, 14(10):2173–2187.

[Tsai et al. 2014] Tsai, C.-W., Lai, C.-F., Chiang, M.-C., and Yang, L. T. (2014). Data Mining for Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, 16(1):77–97.

[Tsai and Chan 2015] Tsai, T.-C. and Chan, H.-H. (2015). Nccu trace: social-network-aware mobility trace. *Communications Magazine, IEEE*, 53(10):144–149.

[Tukey 1977] Tukey, J. W. (1977). *Exploratory Data Analysis*. Behavioral Science: Quantitative Methods. Addison-Wesley, Reading, Mass.

[Wang et al. 2016] Wang, J., Wu, Y., Yen, N., Guo, S., and Cheng, Z. (2016). Big data analytics for emergency communication networks: A survey. *IEEE Communications Surveys Tutorials*, 18(3):1758–1778.

[Wang et al. 2015] Wang, W., Yuan, N., Pan, L., Jiao, P., Dai, W., Xue, G., and Liu, D. (2015). Temporal patterns of emergency calls of a metropolitan city in china. *Physica A: Statistical Mechanics and its Applications*.

[Wang et al. 2014] Wang, Y., Huang, L., Gu, T., Wei, H., Xing, K., and Zhang, J. (2014). Data-driven traffic flow analysis for vehicular communications. In *INFOCOM, 2014 Proceedings IEEE*, pages 1977–1985. IEEE.

[Ward and Barker 2013] Ward, J. S. and Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*.

[Williams and Musolesi 2016] Williams, M. J. and Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society Open Science*, 3(6):160196.

[Wu et al. 2014a] Wu, M., Wang, Y., and Liao, Z. (2014a). A new clustering algorithm for sensor data streams in an agricultural IoT. *Proceedings - 2013 IEEE International Conference on High Performance Computing and Communications, HPCC 2013 and 2013 IEEE International Conference on Embedded and Ubiquitous Computing, EUC 2013*, pages 2373–2378.

[Wu et al. 2014b] Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014b). Data mining with big data. *ieee transactions on knowledge and data engineering*, 26(1):97–107.

[Wu et al. 2011] Wu, Y., Zhu, Y., and Li, B. (2011). Trajectory improves data delivery in vehicular networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 2183–2191. IEEE.

[Xiao et al. 2014] Xiao, X., Li, Y., and Kui, X. (2014). Location patterns and predictability of large scale urban vehicular mobility. In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pages 2705–2709. IEEE.

[Xu et al. 2016] Xu, F., Li, Y., Wang, H., Zhang, P., and Jin, D. (2016). Understanding mobile traffic patterns of large scale cellular towers in urban environment. *IEEE/ACM Transactions on Networking*, PP(99):1–15.

[Yang et al. 2014] Yang, F., Wang, S., Li, J., Liu, Z., and Sun, Q. (2014). An overview of Internet of Vehicles. *China Communications*, 11(10):1–15.

[Yin et al. 2013] Yin, R., Yu, G., Zhong, C., and Zhang, Z. (2013). Distributed resource allocation for d2d communication underlaying cellular networks. In *Communications Workshops (ICC), 2013 IEEE International Conference on*, pages 138–143. IEEE.

[Yu et al. 2016] Yu, S., Liu, M., Dou, W., Liu, X., and Zhou, S. (2016). Networking for big data: A survey. *IEEE Communications Surveys & Tutorials*.

[Zaharia et al. 2010] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., and Stoica, I. (2010). Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95.

[Zanella et al. 2014] Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32.

[Zhang et al. 2013] Zhang, D., Huang, H., Zhou, J., Xia, F., and Chen, Z. (2013). Detecting hot road mobility of vehicular ad hoc networks. *Mobile Networks and Applications*, 18(6):803–813.

[Zhang et al. 2014] Zhang, D., Huang, J., Li, Y., Zhang, F., Xu, C., and He, T. (2014). Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 201–212. ACM.

[Zhang et al. 2016] Zhang, F., Jin, B., Wang, Z., Liu, H., Hu, J., and Zhang, L. (2016). On geocasting over urban bus-based networks by mining trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 17(6):1734–1747.

[Zhao et al. 2016] Zhao, S., King, I., and Lyu, M. R. (2016). A survey of point-of-interest recommendation in location-based social networks. *arXiv preprint arXiv:1607.00647*.

[Zheng et al. 2016] Zheng, K., Yang, Z., Zhang, K., Chatzimisios, P., Yang, K., and Xiang, W. (2016). Big data-driven optimization for mobile networks toward 5g. *IEEE Network*, 30(1):44–51.

[Zheng 2015] Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29.

[Zheng et al. 2014] Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38.

[Zheng et al. 2010] Zheng, Y., Xie, X., and Ma, W.-Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39.

[Zheng et al. 2009] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM.

[Zhu et al. 2011] Zhu, H., Li, M., Fu, L., Xue, G., Zhu, Y., and Ni, L. M. (2011). Impact of traffic influxes: Revealing exponential intercontact time in urban vanets. *IEEE Transactions on Parallel and Distributed Systems*, 22(8):1258–1266.