

Explorando Dados Urbanos: um estudo usando viagens de Táxi da cidade de São Francisco

Karen S. Martins¹, Felipe D. Cunha¹

¹Departamento de Ciência da Computação
Pontifícia Universidade Católica de Minas Gerais
Belo Horizonte – MG – Brasil

karenstemartins@gmail.com, felipe@pucminas.br

Abstract. *Vehicle networks have become a frequently studied topic of the Internet of Things. It allows communication between vehicles and access points along the streets, making the route safer for people. From studies of this network, it is possible to better understand people's behavior in big cities, including their routines, most frequented places for each weekday and mobility models. With the results, it can be created several applications for human benefit and improve network communication protocols. In this work, we analyze taxicabs' traces regarding encounters throughout the day in the city of San Francisco, USA. The goal is to show that there are standards in the number of times these vehicles meet, mainly by comparing weekends with weekdays. For the development, it was used temporal graphs, calibrated traces of the city's taxis and three metrics that are related to the number of encounters. The metrics explored were: repeated encounters throughout the day, repeated encounters throughout the day in the same place, and the ratio of the number of repeated encounters over the total number of encounters for each vehicle. As a result, we find some patterns in the vehicle routines.*

Resumo. *As redes veiculares têm se tornado um tópico da Internet das Coisas bastante estudado. Ela permite a comunicação entre os veículos e pontos de acessos ao longo das ruas, tornando os percursos mais seguros para as pessoas. A partir de estudos dessa rede, é possível entender melhor o comportamento das pessoas nas cidades, incluindo as suas rotinas, lugares mais frequentados em cada dia da semana e modelos de mobilidades. Com os resultados, podemos criar diversas aplicações para o benefício humano e melhorar os protocolos de comunicação da rede. Neste trabalho, analisamos traces de táxis em termos de encontros entre os veículos ao longo do dia na cidade de São Francisco, localizada nos Estados Unidos. O objetivo é mostrar que existem padrões na quantidade de vezes que esses veículos se encontram, principalmente comparando os finais de semanas com os dias úteis da semana. Para o desenvolvimento, foram utilizados grafos temporais, traces calibrados de táxis da cidade e três métricas que relacionam o número de encontros. As métricas exploradas foram: quantidade de encontros repetidos, quantidade de encontros que se repetem na mesma localização e a razão entre a quantidade de encontros que se repetem sobre a quantidade total de encontros para cada veículo. Como resultado, nós encontramos alguns padrões nas rotinas dos veículos.*

1. Introdução

As redes veiculares (VANETs) são um tipo de rede emergente que atraiu o interesse de muitos pesquisadores. Elas são formadas por veículos com capacidade de processamento e comunicação sem fio, trafegando em ruas ou rodovias. Normalmente, os veículos podem estabelecer uma comunicação direta ou pelo uso de um equipamento nas vias. Diferentemente de outras redes, nas VANETs, os nós são veículos como carros, ônibus, táxis e caminhões. Cada tipo de veículo tem comportamento diferente e um padrão de mobilidade [Faezipour et al. 2012, Cunha et al. 2014].

Nas VANETs, a comunicação é altamente influenciada pelos padrões de mobilidade dos veículos. Eles se movem de acordo com as interações que realizam entre os motoristas, os limites de velocidade e direções impostas pelas sinalizações de trânsito. No entanto, as interações entre os diferentes períodos do dia, as rotinas dos motoristas e os veículos também podem influenciar nesta mobilidade. Observando o tráfego durante a semana, podemos notar horas de rush com o tráfego lento em algumas ruas e engarrafamentos. Ao contrário do início da manhã e finais de semana, onde o tráfego é mais silencioso, com baixa densidade de veículos. Essa variação de tráfego em dias e horas diferentes reforça as características da topologia dinâmica das VANETs, tornando a comunicação uma tarefa desafiadora em alguns cenários.

Quando consideramos os táxis, as interações entre pessoas, veículos e lugares têm grande influência na mobilidade. Nesses veículos, o passageiro define o destino final da viagem. Durante as trajetórias, os táxis interagem com outros veículos, atravessam as mesmas ruas e estão sujeitos às mesmas condições de trânsito. Estas características apresentadas sugerem a investigação da mobilidade do táxi, a fim de compreender as interações que realizam durante suas trajetórias. Além disso, para melhorar a conectividade e os serviços das VANETs, seria interessante encontrar veículos que apresentem os mesmos destinos e tenham um comportamento semelhante.

Neste trabalho, será analisado *trace* de táxis na cidade de São Francisco. O objetivo é usar métricas para investigar encontros de táxis ao longo do dia, durante um mês, ou seja, com que frequência dois táxis diferentes conseguiram se comunicar. Em relação à comunicação nas redes veiculares, essa análise pode ajudar a entender melhor as características desses encontros de forma a ajudar o projeto de novos protocolos e serviços. De acordo com Cunha *et al.*, as análises de contatos devem ser realizadas a priori, de forma que o conhecimento que for extraído, possa ajudar no desenvolvimento de novos serviços e aplicativos para os usuários da rede [Cunha et al. 2016].

O restante deste artigo está estruturado da seguinte forma: a Seção 2 é dedicada aos trabalhos relacionados. Na Seção 3, descrevemos a metodologia utilizada para a análise, os conjuntos de dados e métricas que foram avaliadas. Os resultados são apresentados para cada uma das métricas na Seção 4. Na Seção 5, apresentamos algumas aplicações importantes dos estudos de encontros entre veículos para as redes veiculares. Finalmente, na Seção 6, apresentamos as observações finais e trabalhos futuros.

2. Trabalhos Relacionados

Os trabalhos relacionados estão divididos em dois grupos. O primeiro é sobre os trabalhos que usaram *traces* de táxis em diferentes cenários. Está relacionado em

entender a mobilidade dos táxis e a importância desses estudos para as VANETs. O segundo também está relacionado à mobilidade de táxi. Entretanto, os autores exploram características dos contatos entre os veículos, frequência do contato e o tempo entre os contatos para analisar os dados e entender por que essas métricas são importantes neste contexto.

A mobilidade nas VANETs foi abordada em muitos trabalhos. Os autores no artigo [Al-Mayouf et al. 2016] introduziram um algoritmo de roteamento baseado na mobilidade do usuário para evitar a perda de dados e outros problemas comuns de topologia dinâmica e conectividade intermitente. Ramakrishnan et al. também abordou algoritmos de roteamento [Ramakrishnan et al. 2017]. Além disso, existem muitos trabalhos que se concentram na mobilidade do táxi para melhor compreender a dinâmica da cidade, especialmente porque táxis capturam mais detalhes sobre a cidade e os indivíduos devido à sua alta mobilidade. Gao et al. usou dados de GPS para descobrir padrões de tempo de viagem em Pequim [Gao et al. 2013]. Em [Zhang et al. 2012b], os autores construíram um modelo de mobilidade que refletia comportamentos sociais, como a transição entre regiões e características geográficas como *hotspots*. Um modelo de mobilidade também foi proposto por Huang et al., mas considerando aspectos microscópicos e macroscópicos [Huang et al. 2010]. Seguindo o mesmo tema, os autores usaram três parâmetros para capturar a regularidade da mobilidade dos táxis: “*turn probability, road section speed, and travel pattern*” [Huang et al. 2010]. Esses trabalhos analisaram *trace* de Xangai, na China. Além disso, também foi explorado a conectividade entre os veículos e sua devida importância. [Feng and Xu 2016] investigou uma base de dados de Xangai quanto à conectividade ao longo do tempo, ou seja, a evolução dinâmica das VANETs. Existem algumas propostas que utilizaram diferentes métodos para estudar mobilidade, como em [Ghouti 2016], onde os autores usaram a solução baseada em aprendizagem neural.

Outras propostas investigam a mobilidade em termos de contato entre veículos. Em [Cunha et al. 2016], os autores analisaram *traces* reais de São Francisco, Roma e Xangai para comparar com sua versão calibrada. O objetivo era mostrar que as lacunas existentes afetam as métricas principais, como a duração do contato, o tempo de inter-contato e a capacidade da rede. Quanto aos contatos entre os veículos, a duração do contato mede por quanto tempo dois veículos conseguem se comunicar. O tempo de inter-contato avalia o tempo decorrido entre dois encontros sucessivos de dois veículos. E, a capacidade da rede demonstra como a capacidade da rede é afetada pelas lacunas existentes nos *traces* originais.

Em [Machado et al. 2017], foi estudada a métrica de persistência de aresta em relação porcentagem de vezes que ela acontece. Esta métrica classifica relacionamentos regulares e aleatórios. Os autores também avaliaram a persistência geográfica dos usuários, isto é, a probabilidade de um usuário visitar um local anteriormente visitado. Alvarenga et al. avaliaram o comportamento dos veículos, com foco em Redes Sociais Veiculares (VSNs) [Alvarenga et al. 2016]. Estudando a influência do comportamento humano na mobilidade, identificando padrões e rotinas semelhantes e comuns em Helsinque e Roma. Uma das métricas utilizadas foi o número de nós *versus* vizinhos. A correlação entre o grau do nó e a média do grau dos vizinhos mostra a tendência de um nó se conectar a outros com graus similares. Veículos que compartilham rotinas serão

vinculados entre si e com graus similares. Eles também avaliaram propriedades como mundo pequeno, a centralidade e a persistência da aresta, o que significa o número de vezes que duas arestas estão na mesma janela de tempo ao longo do dia. Em outra análise de contato, os autores investigaram a taxa de contato, o tempo inter-contato e a duração do contato em Dartmouth, San Francisco e Second Life [Zyba et al. 2011].

Ao contrário da maioria dos artigos que consideraram *snapshots* para algumas métricas em determinados intervalos de tempo, Qiao et al. usou dados de Pequim para uma análise temporal das características estruturais [Qiao et al. 2017]. As métricas calculadas foram a acessibilidade do caminho ordenado pelo tempo, ou seja, a conectividade do grafo. Além disso, a eficiência da rede temporal, proximidade da centralidade, conectividade e componente. No entanto, a maioria dos estudos avaliou o tempo entre os contatos. [Xiao and Kui 2015], [Zhu et al. 2015], [Li et al. 2013], [Hu et al. 2012], [Zhang et al. 2012a], e [Lee et al. 2013] estudaram inter-contato por diferentes razões em vários *traces* de táxis.

Diversos trabalhos na literatura apresentam análises temporais de *traces* de veículos. Entretanto, poucos exploram aspectos espaciais destas bases de dados. Assim, neste trabalho pretende-se aproveitar a localização espacial dos encontros entre os veículos de forma a extrair padrões e características que venham ajudar no projeto de novos protocolos e serviços para as redes veiculares.

3. Metodologia

Esta seção apresenta a metodologia utilizada para analisar o *trace* de São Francisco de uma maneira espaço temporal, para mostrar a relação de contato entre carros ao longo do dia. Nesta análise foi usado o *trace* de São Francisco, EUA, modelado como um grafo temporal. As subseções seguintes descrevem mais detalhes sobre o conjunto de dados e métricas usadas.

3.1. Base de dados

Nesta seção, apresentaremos o conjunto de dados e suas especificidades. Incluindo, trabalhos que encontraram problemas em *traces* não calibrados, mostrando suas desvantagens de uso, especialmente no que diz respeito à análise de contato.

O *trace* original contém registros de mobilidade de táxis. Foi coletado na área da Baía de São Francisco, EUA. São coordenadas GPS de cerca de 500 táxis, coletados ao longo de 30 dias. As coordenadas foram coletadas a cada minuto de 17 de maio a 10 de junho de 2008 [Michal Piorkowski and Grossglauser 2009]. Os *traces* de mobilidade de veículos apresentam alguns problemas de qualidade na granularidade espacial e temporal. Uma das razões é a baixa taxa de amostragem devido às grandes distâncias ou tempo entre dois veículos. Em relação as VANETs, as análises das cidades podem ter resultados que não condizem com a realidade, causados pelos *traces* de mobilidade com baixa fidelidade que gera impacto na topologia da rede. Para preencher essas lacunas espaciais e temporais, os autores em [Silva et al. 2015] adotaram uma solução baseada em algoritmos de agrupamento. Eles usaram o *trace* de São Francisco em conjunto com algoritmos de agrupamento para descobrir e preencher as lacunas que existiam entre os veículos, e assim obter uma versão com maior granularidade. A diferença entre esses dois *traces* é que os dados no original foram coletados a cada minuto, enquanto o calibrado possui mais coordenadas de GPS para preencher as lacunas.

Cunha et al. comparou o *trace* original e o calibrado em termos de contatos entre veículos. Eles analisaram métricas importantes de rede, como duração de contato, tempo de inter-contato e capacidade de rede, em Roma, São Francisco e Xangai. Os resultados mostraram que as lacunas afetam essas métricas, especialmente em San Francisco [Cunha et al. 2016]. Por esse motivo, como vamos focar em contato entre veículos, escolhemos a versão calibrada.

3.2. Grafos Temporais

A base de dados foi modelada como um grafo temporal $G(t) = (V, E)$. G representa um grafo não direcionado em um tempo t , V um conjunto de veículos V_i , e E o conjunto de arestas E_{ij} . A aresta E_{ij} existe apenas durante um tempo t , se V_i e V_j são dois veículos diferentes [Alvarenga et al. 2016]. Neste trabalho, analisamos cada dia separadamente, agrupando as arestas a cada 15 minutos e criando grafos separados. Por exemplo, em um determinado dia, existe um grafo que contém todos os veículos e arestas onde t ocorreu entre às 10 : 00 e às 10 : 14 : 59 da manhã. Foi escolhido 15 minutos para poder ter uma amostra significativa de dados. Um valor menor, poderia gerar uma amostra muito pequena de dados com um grafo muito esparso, sem conectividade. E um maior valor, a amostra pode ser muito grande [Alvarenga et al. 2016].

Irá existir uma aresta entre dois carros se eles conseguem estabelecer uma comunicação entre si, levando em conta que ambos os valores de t compartilham a mesma hora e minuto. Adotando o padrão de protocolo IEEE 802.11p [IEEE 2010], a comunicação ocorre se os dois carros estiverem até no máximo 100 metros um do outro. Por exemplo, a Figura 1 mostra o mapa de calor de todos os contatos que ocorreram em uma noite de domingo entre às 21 : 30 e às 21 : 44 : 59. Após a criação de todos os grafos para cada dia existente, as métricas foram computadas.

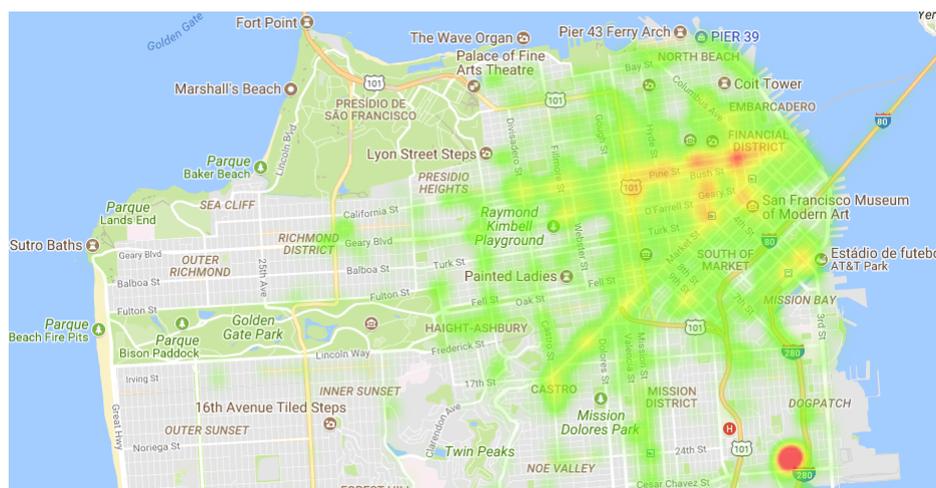


Figura 1. Mapa de calor da cidade de São Francisco em uma noite de domingo entre 21:30 and 21:44:59 horas.

3.3. Métricas

As três métricas foram baseadas na análise espacial temporal de contatos entre veículos: encontros repetidos ao longo do dia, encontros repetidos ao longo do dia na

mesma localização, e, por último, a relação entre esses dois. O objetivo é examinar os padrões da cidade com base em contatos ao longo do dia para cada dia da semana.

Encontros repetidos ao longo do dia: A primeira métrica que será avaliada é a repetição de encontros entre veículos ao longo do dia. Em outras palavras, quantas vezes dois veículos diferentes conseguiram se comunicar um com o outro, em um dia específico. Como estamos trabalhando com *traces* de táxis, a chance de dois táxis se encontrarem mais de uma vez por dia é muito alta. Os táxis, geralmente, possuem uma alta mobilidade, especificamente em São Francisco, que é uma grande cidade com um grande número de habitantes e turistas durante o ano inteiro. É importante descobrir qual dia da semana e horário que esses encontros são mais comuns, encontrando o padrão da cidade. Quanto mais encontros acontecerem entre os carros, melhor será para a rede. Isso significa que compartilhar informações entre carros na rede é mais fácil, em situações graves todos os carros serão informados aproximadamente ao mesmo tempo. Por outro lado, seria necessário ter mais pontos de contato nas ruas para ajudar a espalhar informações. Por fim, para provar que a rede não segue um padrão aleatório, utilizamos os mesmos veículos e pontos de localização, e geramos arestas aleatórias seguindo o método proposto por [Johnson and Kotz 1977]. No final, geramos a mesma métrica de encontros repetidos ao longo do dia para que fosse possível a comparação entre o *trace* original e o aleatório.

Encontros repetidos ao longo do dia na mesma localização: Esta métrica é um complemento da anterior. Ela avalia as repetições de encontros ao longo do dia na mesma localização. Ou seja, quantas vezes dois veículos diferentes conseguiram se comunicar um com o outro em um dia específico e na mesma localização. Para mesma localidade, consideramos encontros entre os mesmos dois carros que ocorreram mais de uma vez, e a distância entre os lugares de cada encontro foi até 100 metros de distância. Por exemplo, se os táxis A e B se encontraram às 8 horas no local X, e então, eles se encontraram novamente às 10:30 no lugar Y, para ser uma repetição de localidade, X e Y precisam estar a 100 metros ou menos um do outro. No final, contamos quantas vezes isso ocorreu para cada carro. É importante observar lugares onde a frequência de encontros é maior, podem ser úteis para disseminar informações.

Taxa de repetição de encontros: A última métrica avaliada é uma razão de encontros repetidos. Ela representa a razão do número de encontros que se repetiram sobre o número total de encontros para cada carro durante o dia. Para ser mais específico, primeiro temos quantas vezes ocorre repetição para cada táxi em um determinado dia. Então, nós conseguimos quantas comunicações cada táxi fez no mesmo dia. No final, dividimos o primeiro número pelo segundo. Esta métrica é importante para mostrar a tendência dos encontros acontecerem e se repetirem.

4. Resultados

Nesta seção serão apresentados os resultados de acordo com as métricas. Os resultados serão agrupados de acordo com as métricas definidas anteriormente, eles mostram as métricas para cada dia da semana. A base de dados contém 30 dias, então há aproximadamente quatro dias para cada dia da semana. A Figura 2 mostra o total de encontros para cada dia entre 24 e 30 de maio. A Tabela 1 apresenta os dados para os encontros entre os táxis. É possível notar que esses encontros são mais comuns nos domingos, segundas

feiras e sábados.

Tabela 1. Porcentagem de dados *versus* quantidade de encontros.

Dia	50%	90%
Domingo	690	859
Segunda	589	897
Terça	521	807
Quarta	494	767
Quinta	551	813
Sexta	565	900
Sábado	582	844

É possível notar uma diferença entre dos dias da semana, onde a quantidade de encontros é maior aos finais de semana. Atribui-se essa diferença ao comportamento dos usuários, os quais optam por utilizar mais táxi aos finais de semana. Estes são dias que os mesmos não dirigem e normalmente frequentam lugares para diversão e lazer. Além disso, na região sul de São Francisco, existe o Vale do Silício, onde está localizado muitas das maiores corporações de alta tecnologia do mundo, o que aumenta ainda mais o fluxo de pessoas. Todas estas razões aumentam a necessidade por táxi na cidade.

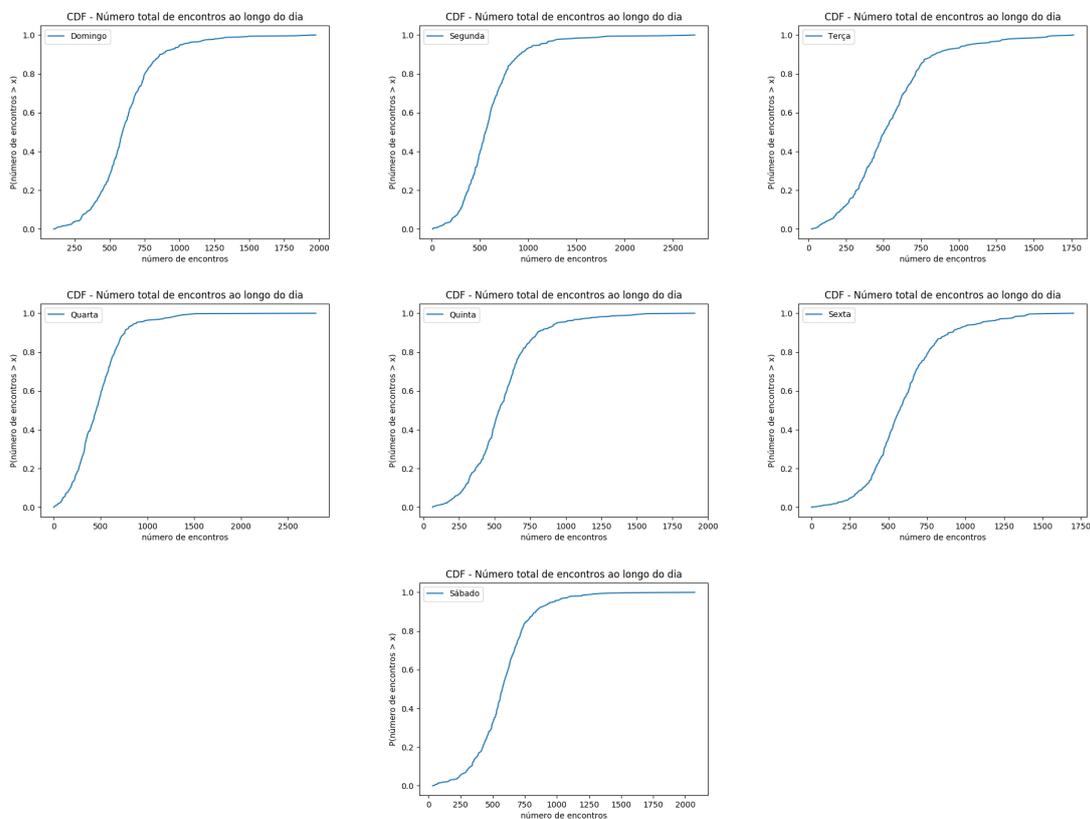


Figura 2. Total de encontros ao longo do dia para cada dia da semana.

4.1. Encontros repetidos ao longo do dia

Durante a análise dos resultados das repetições de encontros ao longo do dia, percebemos que, para cada um dos dias repetidos, não houve grandes mudanças (ou seja, as quatro segundas-feiras têm um padrão muito similar). Por essa razão, optamos por focar apenas em uma semana do conjunto de dados em vez de todo o mês. A semana começa em 24 de maio (sábado) e termina em 30 de maio (sexta-feira). A Figura 3 contém o gráfico da função de distribuição cumulativa (CDF) para cada um desses dias gerados a partir dos dados reais e também da comparação com o grafo aleatório.

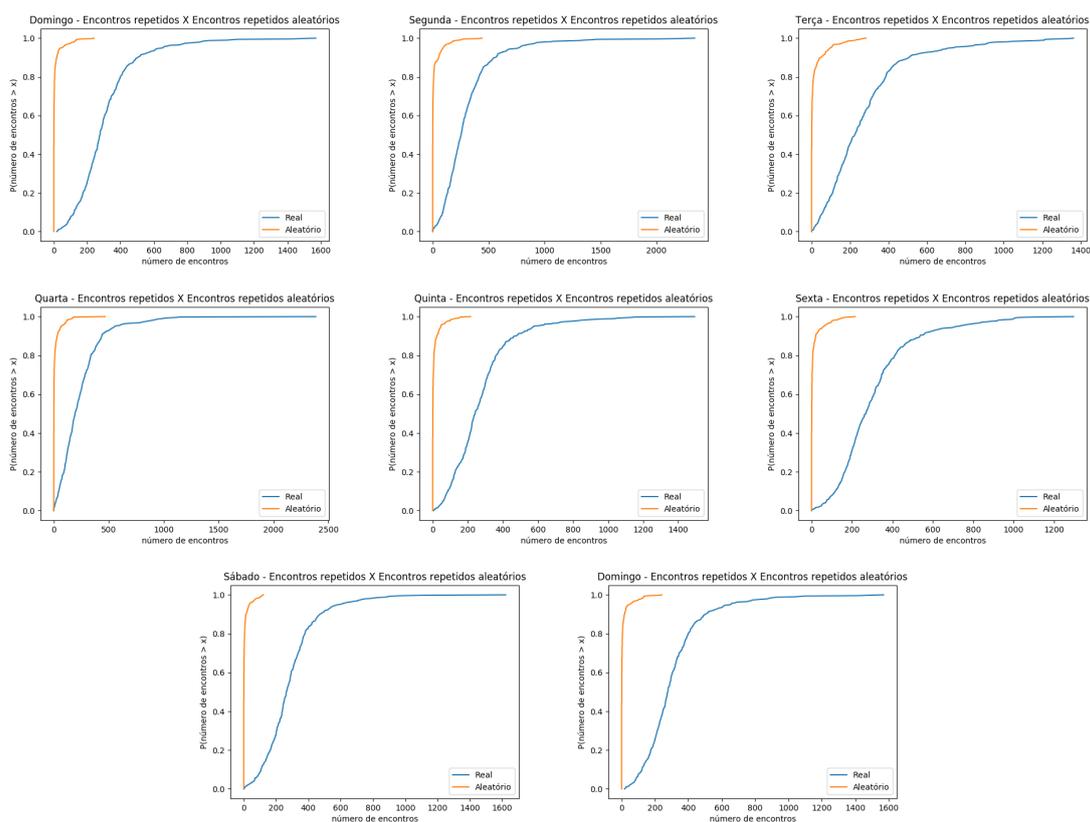


Figura 3. Encontros repetidos ao longo do dia para cada dia da semana.

Comparando os encontros repetidos dos dados reais (linha azul) e do grafo aleatório (linha vermelha), observa-se que para todos os dias da semana a quantidade de encontros repetidos do real é muito maior, tanto para 50% quanto 90%. Isso significa que a rede original não segue um padrão aleatório e os encontros realmente acontecem, i.e., suas características favorecem a ocorrência de encontros. Analisando somente os dados reais, é possível observar que, durante toda a semana o número de encontros repetidos é bem similar. Esse comportamento pode ser explicado pelo fato de táxis representarem o desejo de muitos e também destes veículos estarem em constante circulação. A tabela 2 mostra com mais detalhes essa relação, comparando o número de encontros que 50% e 90% dos dados representam.

Embora a diferença entre os números não seja muito grande, o dia da semana, principalmente segunda e sexta-feira, tem um maior número de encontros repetidos entre os táxis. Precisamente por causa da rotina dos residentes da cidade. Eles frequentam

Tabela 2. Porcentagem de dados versus quantidade de encontros.

Dia	50%	90%
Domingo	274	497
Segunda	261	557
Terça	225	472
Quarta	245	446
Quinta	244	448
Sexta	280	512
Sábado	255	467

os mesmos lugares durante a semana. Consequentemente, a chance de dois táxis se encontrarem mais de uma vez é maior, principalmente em regiões com maior número de empresas e escolas. Na Figura 4, pode-se observar que os encontros durante os dias da semana estão mais concentrados em uma região, e no fim de semana esses encontros se espalham por toda a cidade.



Figura 4. Encontros durante a semana.

4.2. Encontros repetidos ao longo do dia na mesma localização

Como na métrica anterior, os resultados foram muito semelhantes entre os dias da semana, por isso escolhemos a mesma semana de maio, assim facilitando a comparação entre as duas métricas. A Figura 5 e Tabela 3 representam os resultados do número de encontros na mesma localização. Assim como nas métricas anteriores, os maiores resultados ocorreram durante a semana e não nos finais de semana.

A razão desses resultados é a mesma que a anterior, durante a semana, a necessidade por táxis é muito menor, então eles tendem a ter um fluxo menor e trabalhar na mesma localidade ao longo do dia, fazendo viagens mais curtas. Aos finais de semana, a chance de dois táxis se encontrar é menor porque eles trafegam mais em toda a cidade, e esses encontros não acontecem nos mesmos lugares, são mais espalhados. Além disso, quando a demanda por táxis é menor, eles precisam ficar em pontos aguardando passageiros. Consequentemente, eles se encontram mais vezes um com o outro, e geralmente estão nos mesmos locais. A Figura 6 mostra locais onde os encontros repetidos com a mesma localização ocorreram com mais frequência. O primeiro lugar destacado, Figura 6-(a), é

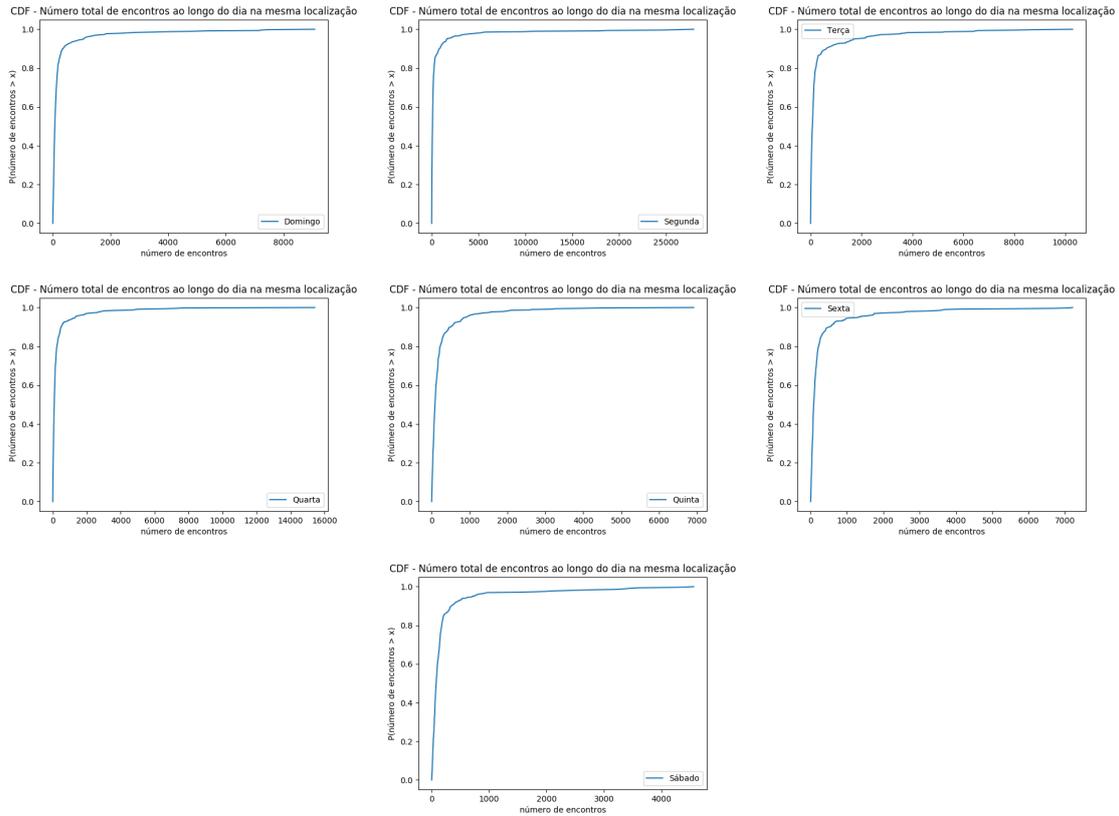


Figura 5. Encontros repetidos ao longo do dia na mesma localização para cada dia da semana.

o distrito financeiro de São Francisco, onde está localizado o distrito central de negócios com o maior número de sedes corporativas da cidade. É uma região bastante agitada da cidade durante a semana; muitas pessoas trabalham nessas grandes empresas. O segundo, Figura 6-(b), é uma empresa de táxis. A demanda de táxis é menor neste período; eles ficam muito tempo parados na empresa.

4.3. Taxa de repetição de encontros

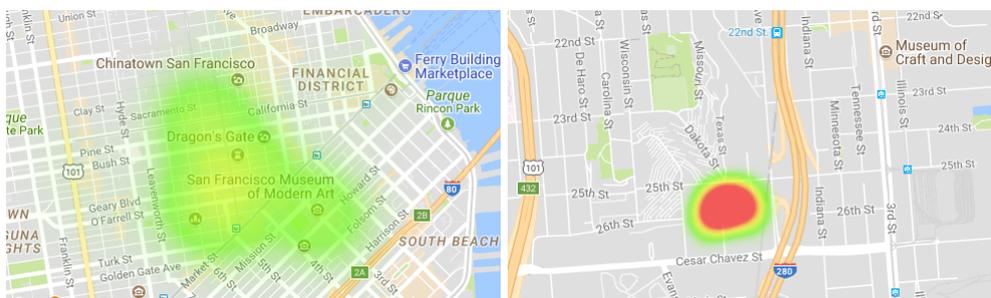
Esta métrica calcula a relação entre o número total de encontros de um veículo e aqueles que foram repetidos para cada carro. Como pode ser visto na Figura 7, a taxa de encontros repetidos é muito maior durante a semana. O mesmo pode ser verificado na Tabela 4. Isso significa que os táxis são mais propícios de se encontrarem um com o outro. Eles circulam mais nas mesmas regiões, repetindo sua rotina ao longo da semana. Segunda e sexta-feira são os dias com o maior número de encontros. Sexta-feira por ser o início do fim de semana, onde as pessoas costumam sair muito para bares e boates. Segunda-feira por ser o primeiro dia da semana, onde as pessoas ainda estão voltando para a suas rotinas e têm uma maior demanda por transporte.

5. Aplicações

Estudos sobre o comportamento e os padrões das pessoas nas grandes cidades geram muitos benefícios para a criação de protocolos de rede, desenvolvimento de aplica-

Tabela 3. Porcentagem de dados versus quantidade de encontros repetidos na mesma localização.

Dia	50%	90%
Domingo	50	312
Segunda	215	772
Terça	79	696
Quarta	119	734
Quinta	99	667
Sexta	88	503
Sábado	76	319



(a) Distrito Financeiro.

(b) Empresa de táxis.

Figura 6. Lugares mais frequentes onde os encontros ocorreram na mesma localização.

tivos e serviços. Neste trabalho, analisamos encontros de táxis, e foi possível identificar várias características em São Francisco.

Os resultados podem ser aplicados em várias soluções para projetos de redes veiculares. Os veículos que são mais populares, encontram mais frequentemente com outros veículos, são excelentes opções para ajudar na transmissão da rede. Uma informação disseminada por esses veículos atingiria um número maior de carros rapidamente. Também é importante no projeto de protocolos de comunicação da rede. É possível identificar, através do número de encontros, quais são os dias e os horários em que a rede precisará suportar uma maior quantidade de dados e conexões. Em regiões com maior frequência, uma maior quantidade de pontos de acesso será necessária para garantir uma boa cobertura da rede.

Outra aplicação para os resultados é no desenvolvimento de serviços e aplicativos. As regiões mais visitadas em determinados horários do dia podem ser estudadas com mais detalhes para descobrir o padrão de visitação dos usuários. Assim, com esse conhecimento, podem ser criadas aplicações que ajudem comerciantes e empresários. Por outro lado, lugares mais visitados à noite e ao final de semana podem ser úteis em aplicações turísticas, para aplicativos com sugestões de lugares para visitação e diversão. Além disso, serviços que ajudem os motoristas com rotas alternativas para evitar algumas regiões em um determinado momento do dia, evitando os congestionamentos, podem ser projetados. Ao mesmo tempo que, o conhecimento dos locais mais frequentados viabiliza o desenvolvimento de serviços para compartilhamentos de veículos, o que pode contribuir para melhorias no tráfego diário.

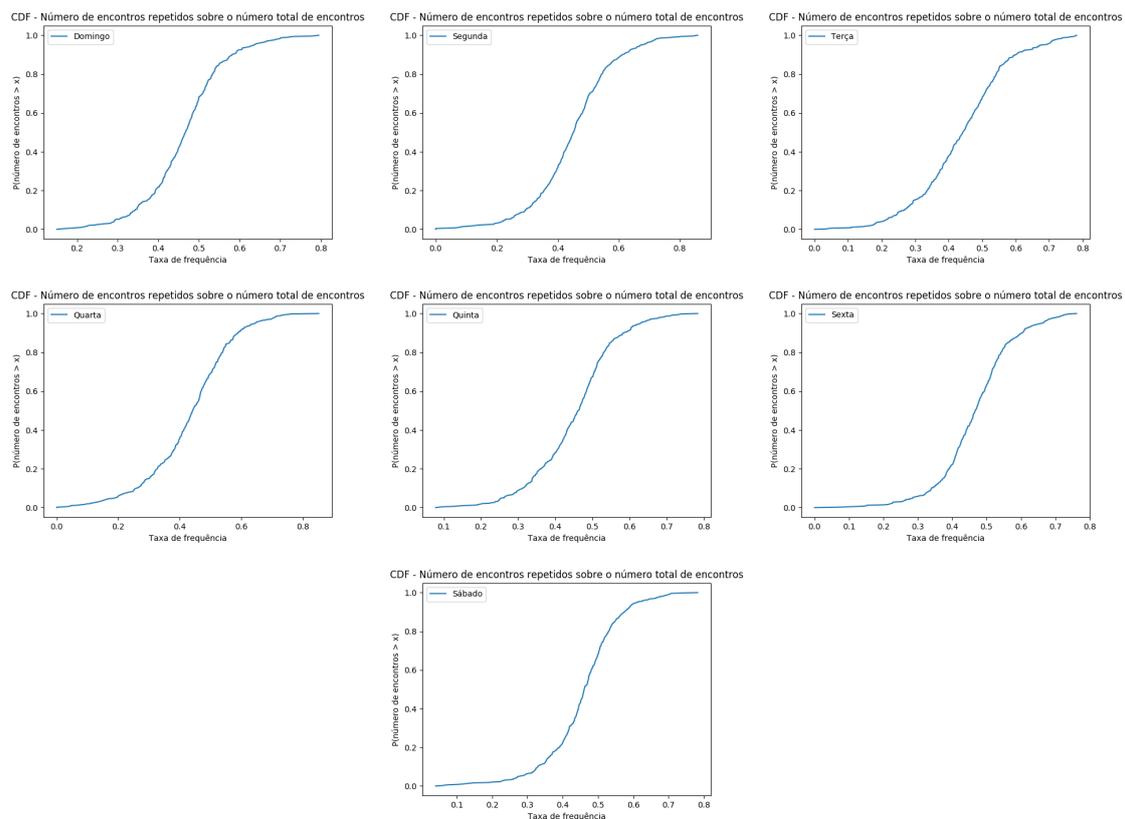


Figura 7. Número de encontros repetidos sobre o número total de encontros para cada dia da semana.

6. Conclusão

Neste trabalho, desenvolveu-se uma análise espaço temporal de encontros de táxis em toda a cidade de São Francisco – EUA. Para o desenvolvimento, utilizou-se grafos temporais, *traces* de táxis calibrados, como o conjunto de dados, e três métricas principais: encontros repetidos ao longo do dia, encontros repetidos ao longo do dia na mesma localização e taxa frequência. Além disso, demonstramos a quantidade total de encontros. Cada dia da semana foi analisado separadamente para uma possível comparação entre eles.

Os resultados foram muito próximos para cada dia repetido da semana, então escolhemos focar em apenas uma semana. Foi possível identificar que durante os dias da semana, principalmente na segunda e sexta-feira, houve um maior número de encontros repetidos entre veículos. Embora, nos finais de semana, o número de encontros seja maior. Esta diferença pode ser explicada pela necessidade das pessoas, especialmente dos turistas, pessoas que frequentam bares e boates, e conseqüentemente, não dirigem. Além disso, o número de encontros entre os táxis nos mesmos lugares é maior durante a semana, precisamente porque eles se locomovem a distancias mais curtas e frequentam as mesmas regiões. Ao contrário do fim de semana, em que circulam pela cidade.

Como trabalhos futuros, pretende-se replicar esses estudos para outras cidades, como Nova York, Roma e Xangai, podendo comparar as diferenças entre as rotinas nessas cidades. Além disso, pode-se aplicar mais métricas que analisam encontros entre veículos

Tabela 4. Porcentagem de dados versus quantidade de encontros repetidos sobre o total de encontros.

Dia	50%	90%
Domingo	0.469	0.585
Segunda	0.454	0.613
Terça	0.439	0.612
Quarta	0.441	0.586
Quinta	0.457	0.585
Sexta	0.465	0.605
Sábado	0.458	0.580

e também analisar dados pela hora do dia, em vez de todo o dia, como foi feito neste trabalho.

Referências

- Al-Mayouf, Y. R. B., Ismail, M., Abdullah, N. F., Wahab, A. W. A., Mahdi, O. A., Khan, S., and Choo, K. K. R. (2016). Efficient and stable routing algorithm based on user mobility and node density in urban vehicular network. *PLoS ONE*, 11(11).
- Alvarenga, D. A., Cunha, F. D., Viana, A. C., Mini, R. A. F., and Loureiro, A. A. F. (2016). Classificando Comportamentos Sociais em Redes Veiculares. *SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.
- Cunha, F., Carneiro Viana, A., Mini, R. A. F., and A.F. Loureiro, A. (2014). Is it possible to find social properties in vehicular networks? In *IEEE Symposium on Comp.and Com.(ISCC '14)*.
- 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. (2016). Communication analysis of real vehicular calibrated traces. *2016 IEEE International Conference on Communications, ICC 2016*.
- Faezipour, M., Nourani, M., Saeed, A., and Addepalli, S. (2012). Progress and challenges in intelligent vehicle area networks. *Communications ACM*, 55(2):90–100.
- Feng, H. and Xu, Y. (2016). An Empirical Study on Evolution of the Connectivity for VANETs Based on Taxi GPS Traces. *International Journal of Distributed Sensor Networks*, 2016.
- Gao, M., Zhu, T., Wan, X., and Wang, Q. (2013). Analysis of travel time patterns in urban using taxi GPS data. *Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCoM 2013*, pages 512–517.
- Ghouti, L. (2016). Mobility prediction in mobile ad hoc networks using neural learning machines. *Simulation Modelling Practice and Theory*, 66:104–121.
- Hu, Y., Wang, H., Xia, C., Li, W., and Yang, Y. (2012). On the distribution of inter contact time for DTNs. *Proceedings - Conference on Local Computer Networks, LCN*, pages 152–155.
- Huang, H., Zhu, Y., Li, X., Li, M., and Wu, M. Y. (2010). META: A mobility model of Metropolitan Taxis extracted from GPS traces. *IEEE Wireless Communications and Networking Conference, WCNC*.

- IEEE (2010). Ieee 802.11p/d10.0.
- Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory*. Wiley, New York.
- Lee, C. H., Kwak, J., and Eun, D. Y. (2013). Characterizing link connectivity for opportunistic mobile networking: Does mobility suffice? *Proceedings - IEEE INFOCOM*, pages 2076–2084.
- Li, Y., Jin, D., Wang, Z., Zeng, L., and Chen, S. (2013). Exponential and power law distribution of contact duration in urban vehicular ad hoc networks. *IEEE Signal Processing Letters*, 20(1):110–113.
- Machado, K., Boukerche, A., Cerqueira, E., and Loureiro, A. A. F. (2017). A socially-aware in-network caching framework for the next generation of wireless networks. *IEEE Communications Magazine*, 55(12):38–43.
- Michal Piorkowski, N. S.-D. and Grossglauser, M. (2009). The epfl/mobility dataset (v. 2009-02-24). <https://crawdad.cs.dartmouth.edu/epfl/mobility/20090224/>. Accessed: October 2017.
- Qiao, L., Shi, Y., and Chen, S. (2017). An empirical study on the temporal structural characteristics of VANETs on a Taxi GPS Dataset. *IEEE Access*, 5:722–731.
- Ramakrishnan, B., Selvi, M., and Nishanth, R. B. (2017). Efficiency measure of routing protocols in vehicular ad hoc network using freeway mobility model. *Wireless Networks*, 23(2):323–333.
- Silva, F. A., Celes, C., Boukerche, A., Ruiz, L. B., and Loureiro, A. A. (2015). Filling the gaps of vehicular mobility traces. In *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, MSWiM '15*, pages 47–54, New York, NY, USA. ACM.
- Xiao, X. F. and Kui, X. (2015). The Characterizes of Communication Contacts Between Vehicles and Intersections for Software-Defined Vehicular Networks. *Mobile Networks and Applications*, 20(1):98–104.
- Zhang, D., Huang, H., Chen, M., and Liao, X. (2012a). Empirical study on taxi GPS traces for Vehicular Ad Hoc Networks. *IEEE International Conference on Communications*, pages 581–585.
- Zhang, L., Ahmadi, M., Pan, J., and Chang, L. (2012b). Metropolitan-scale taxicab mobility modeling. *GLOBECOM - IEEE Global Telecommunications Conference*, pages 5404–5409.
- Zhu, Y., Zhao, Q., and Zhang, Q. (2015). Delay-constrained data aggregation in VANETs. *IEEE Transactions on Vehicular Technology*, 64(5):2097–2107.
- Zyba, G., Voelker, G. M., Ioannidis, S., and Diot, C. (2011). Dissemination in opportunistic mobile ad-hoc networks: The power of the crowd. *Proceedings - IEEE INFOCOM*, pages 1179–1187.