

Análise de Padrões Espaciais e Temporais da Mobilidade de Táxis em San Francisco e Roma

Cristiano Martins Monteiro¹, Fábio Rocha da Silva¹, Cristina Duarte Murta¹

¹Departamento de Computação – CEFET-MG

Resumo. *A disponibilidade de registros de localização e tempo de serviços de táxi possibilita a análise de padrões de mobilidade em centros urbanos e melhor compreensão das rotinas da população local. Este trabalho utiliza um conjunto de dados com mais de trinta milhões de registros de localização e tempo de táxis de San Francisco e Roma para comparar padrões de mobilidade urbana nessas duas metrópoles. As rotas anômalas dos táxis em ambas as cidades foram encontradas e discutidas, assim como as rotas mais frequentes em San Francisco. Os resultados indicam a viabilidade do algoritmo utilizado para corrigir as distâncias das rotas com erros de GPS em que o táxi estava parado.*

Abstract. *The availability of location and time records from taxi services enables the analysis of mobility patterns in urban centers and better understanding of the routines of the local population. This work uses a data set with over thirty million location and time records of taxis of San Francisco and Rome to compare urban mobility patterns in both metropolises. The anomalous routes of taxis in both cities were found and discussed, as well as the most frequent routes in San Francisco. The results show the feasibility of the algorithm used for correct the routes's distances with GPS errors when the taxi was stopped.*

1. Introdução

A proliferação do uso de dispositivos digitais móveis, em particular com habilitação de localização, produz rastros digitais que podem ser agrupados para estudos de padrões de mobilidade urbana. Pesquisas recentes utilizam dados urbanos abertos para entender e observar padrões nas rotinas diárias das pessoas, compondo modelos de mobilidade em grandes centros urbanos. Tais padrões são identificados por meio de análises espaciais e temporais dos dados. Esses dados são normalmente obtidos via dispositivos GPS, redes sociais, histórico de chamadas de celulares, circulação de veículos em geral ou trajetórias de transporte público, dentre outros.

As pegadas digitais dos táxis oferecem uma visão rica e detalhada da dinâmica da movimentação da população de uma cidade. Diferentemente dos ônibus e metrô, os trajetos de táxis não estão limitados a itinerários e horários fixos. Dessa forma, as rotas dos táxis representam bem as dinâmicas e costumes de determinada região [Castro et al. 2013]. Este trabalho utiliza dados de táxi das metrópoles de San Francisco, nos EUA, e Roma, na Itália, com o objetivo de explorar os padrões espaciais e temporais da mobilidade de táxis nessas cidades. As análises nos permitem visualizar as áreas de ambas as cidades com maior operação de táxis; encontrar e analisar rotas anômalas; calcular as distâncias percorridas pelas rotas de táxi; listar as rotas mais frequentes dos táxis com passageiro em San Francisco; em suma, inferir, pelos dados, o

funcionamento dos serviços de táxis nestas cidades ao longo do dia e da semana, inclusive padrões culturais.

Para inferir padrões de mobilidade, foram analisados mais de 30 milhões de registros espaço-temporais coletados durante um mês em cada cidade. Entendemos que se refere a uma análise de *Big Data* pois, além do volume de dados, trata-se de análise tridimensional de dados em espaço (localização georreferenciada) e tempo, os quais são complexos para se trabalhar. Uma das dificuldades de trabalhar com dados de tempo e de localização está na necessidade de se tratar os erros gerados por imprecisões nas coletas de dados e no cálculo das distâncias. Esse desafio é evidenciado ao calcular a distância percorrida por um táxi levando o passageiro ao seu destino, ou circulando até encontrar um novo passageiro. Nesse segundo caso, o taxista pode decidir estacionar o táxi por um determinado período, aguardando no ponto até que um passageiro surja. Uma vez que as localizações de latitude e longitude registradas pelo sistema de GPS apresentam erros [Valero et al. 2014], os quais geram localizações aparentemente instáveis, é importante que as distâncias percorridas pelos táxis sejam corrigidas. Para contornar esse problema, aplicamos um algoritmo para eliminar essas distorções antes de calcular as distâncias das rotas. Os resultados indicam a viabilidade dessa inferência no tratamento dos dados.

A descoberta dos padrões de mobilidade dos táxis e das rotinas diárias dos taxistas e cidadãos que usam o serviço de táxi pode ajudar as próprias agências de táxi ou mesmo serviços concorrentes, tais como o *Uber*. A informação pode beneficiar também comerciantes locais, empresas de turismo, setores de planejamento de transporte urbano, podendo produzir estimativas de fluxo de pessoas e veículos nas regiões.

A seção seguinte discute os trabalhos relacionados; a seção 3 apresenta as bases de dados utilizadas neste trabalho e a metodologia; a seção 4 aborda as análises espaciais, temporais e rotas frequentes de San Francisco e Roma; e a seção 5 finaliza o trabalho.

2. Trabalhos Relacionados

O rápido desenvolvimento das tecnologias da informação tem gerado uma grande massa de dados de trajetórias espaço-temporais, que representam a mobilidade de uma diversidade de objetos como pessoas, veículos e animais [Zheng et al. 2015]. Em anos recentes, o estudo da mobilidade humana foi impulsionado principalmente pelo crescimento expressivo do volume de dados georreferenciados resultante do aumento do uso de sistemas de navegação, como o GPS [Siła-Nowicka et al. 2015, Kellner and Egger 2016].

Vários trabalhos recentes utilizam traços de GPS obtidos por trajetórias de táxis a fim de descobrir padrões de dinâmicas comunitárias [Chen 2014]. Em [Castro et al. 2013] é apresentada uma vasta pesquisa sobre o uso de traços de GPS obtidos principalmente por táxis. Os autores organizaram os trabalhos da área em três categorias principais: dinâmicas sociais, dinâmicas de tráfego e dinâmicas operacionais. A categoria “dinâmicas sociais” lista trabalhos sobre o comportamento da população de uma cidade baseando em sua mobilidade. As “dinâmicas operacionais” apresentam estudos sobre a operação dos taxistas nas ruas. E as “dinâmicas de tráfego” se referem ao fluxo dos veículos nas redes das estradas.

Outro objetivo do estudo da circulação dos táxis nas estradas é avaliar o funcionamento e desempenho de VANETs (*Vehicular Ad hoc Networks*). No trabalho

de [Amici et al. 2014] é avaliado o desempenho de um protocolo epidêmico de roteamento de mensagens em VANETs comparando os resultados de simuladores de mobilidade com as bases de dados de San Francisco, de Roma, e de Shenzhen, na China. Em [Cunha et al. 2015] são analisados padrões de conectividade possíveis entre os carros que compõem as redes veiculares no decorrer do dia e são aplicadas técnicas estatísticas, de teoria dos grafos e de análise de redes nas bases de dados de San Francisco e Roma.

Padrões de comportamento dos taxistas e das rotas de táxis foram estudados por [Chen 2014]. O autor apresenta três principais contribuições: um algoritmo para identificar trajetórias anômalas dentre as demais com o mesmo local de origem e destino; um algoritmo que encontra regiões candidatas a novos pontos de ônibus a partir dos locais de início e fim das rotas de táxi noturnas; e um *framework* para gerar passeios personalizados, interativos e conscientes em relação ao tráfego de veículos de uma cidade.

A procura de passageiros pelos taxistas assim como a relação de táxis ocupados e desocupados e a renda desses taxistas também são temas de pesquisa recentes [Li et al. 2011, Zhang 2014]. O trabalho de [Murashkin and Geroliminis 2015] estima a quantidade de táxis procurando passageiros pelas ruas da cidade de Shenzhen, China. Os autores também avaliam o comportamento dos taxistas nas ruas diferenciando a velocidade e distância média percorrida com os táxis ocupados e procurando passageiros. Em [Liu et al. 2015] é estudada a decisão dos taxistas sobre como conseguir um passageiro. Os resultados demonstram, por exemplo, que taxistas experientes preferem estacionar o veículo e esperar os passageiros aparecerem que rodar pelas ruas procurando-os. No trabalho de [Ding et al. 2015] são analisados padrões de taxistas em Xangai, China, ao procurar passageiros ou parados, esperando clientes. Os autores também estimaram a renda obtida pelas corridas de táxi. Em [Zhu et al. 2012], os autores utilizam padrões históricos das rotas de táxi em uma base de dados de Beijing para inferir o *status* de ocupação dos táxis e momentos em que o veículo ficou estacionado.

O presente trabalho se difere dos trabalhos relacionados nos seguintes aspectos: inferimos os momentos em que os táxis estavam parados para reduzir erros no cálculo das distâncias; discutimos estatísticas das distâncias das rotas de San Francisco; avaliamos diferenças quantitativas entre rotas com o táxi ocupado e com o táxi desocupado; e contextualizamos os resultados com o fato dos dados de San Francisco serem sobre táxis e dos dados de Roma serem sobre taxistas.

3. Bases de Dados e Metodologia

Neste trabalho foram utilizadas duas bases de dados de movimentação de táxis, uma da cidade de San Francisco¹, EUA, abordada em [Piorkowski et al. 2009]; e outra de Roma², Itália, apresentada em [Bracciale et al. 2014]. Tais bases foram escolhidas por terem conteúdos semelhantes e estarem disponíveis para *download*. Não foram encontradas bases de dados similares de cidades do Brasil.

A base de San Francisco é composta por 536 arquivos, um para cada táxi, somando 11.219.955 linhas. Cada linha contém um registro com os seguintes dados: a identificação do táxi; a localização georreferenciada (latitude e longitude); o *status* de ocupação (1 para

¹<http://crawdad.org/epfl/mobility/20090224/>

²<http://crawdad.org/roma/taxi/20140717/>

táxi com passageiro e 0 para táxi sem passageiro); e o *timestamp* no formato *Unix Epoch* do momento de aquisição desses dados. A base de dados de Roma possui 21.817.851 registros em um único arquivo sobre 316 taxistas diferentes contendo, para cada registro, a identificação do taxista; a localização georreferenciada; a data e a hora de aquisição.

Utilizamos o termo “ocupado” para representar um táxi com passageiro; “desocupado” para representar um táxi sem passageiro; “registro” ao se referir a cada linha das bases de dados; “trecho” para definir o movimento do táxi a cada dois registros consecutivos da base de dados; e “rota” para representar uma sequência de registros com o mesmo *status* de ocupação. É importante ressaltar que na base de dados de San Francisco há a identificação dos táxis e não dos taxistas. Já em relação à Roma, há a identificação dos taxistas mas não há a identificação dos táxis nem um registro informando quando o táxi estava ocupado ou desocupado. Os trechos de San Francisco são registrados a cada 60 segundos em média e o intervalo médio entre os registros de Roma é de 7 segundos.

Ambas as bases de dados foram filtradas para eliminar registros que aparentemente apresentam erros de localização e para tratar possíveis erros das distâncias das rotas. As rotas de táxi com medidas ou localizações muito distintas das demais são denominadas como “rotas anômalas”. Foram retirados todos os registros posicionados fora de uma região com estradas (localizados no mar, por exemplo); e registros fora da região das cidades de Roma ou San Francisco e que não formam uma trajetória.

Assim, foram excluídos dez registros da base de dados de San Francisco localizados em um mesmo ponto da montanha canadense chamada “Stein Valley Nlaka’pamux Heritage Park”. Esses registros foram obtidos em momentos e veículos diferentes e se distanciam num raio de aproximadamente 1.400 km do próximo registro não anômalo. Também foi excluído um registro de uma rota anômala com o táxi ocupado dentro da cidade de San Francisco. Neste caso, o táxi teria se deslocado num raio de 600 km em 75 segundos, alcançando o estado de Nevada, e retornado outros 600 km até a cidade de San Francisco nos próximos 49 segundos. Outros 585 registros foram retirados somente da análise das distâncias das rotas devido a API do Google Maps não conseguir aproximar o ponto para uma estrada ou encontrar uma rota de carro possível para o trecho. Após esse tratamento, a base de dados de San Francisco perdeu 897 registros, o equivalente a menos de 0,008%.

Em relação à base de Roma, foram retirados 18 registros de uma rota anômala com o taxista parado num ponto localizado em Londres. Nesse caso, passou tempo o suficiente para terem deslocado o localizador GPS da Itália até a Inglaterra, porém, essa rota foi excluída por não haver uma trajetória ligando-a às demais rotas desse taxista. Foram retirados outros 3 registros de um taxista parado num raio de 400 km de distância do registro não anômalo mais próximo. Outros 2 registros foram excluídos por indicarem que um taxista movimentou num raio de 53 km em somente 15 segundos. Não foram calculadas distâncias das rotas utilizando a API do Google Maps porque a base de dados de Roma não possui o *status* de ocupação informando quando as rotas iniciaram ou terminaram. Portanto, não há registros retirados devido à API. A base de dados de Roma perdeu 23 registros, o equivalente a cerca de 0,0001%.

Utilizamos a *Google Maps Distance Matrix API* para estimar as distâncias trafeadas pelos táxis por considerar as ruas e o sentido das vias para carros, e porque essa

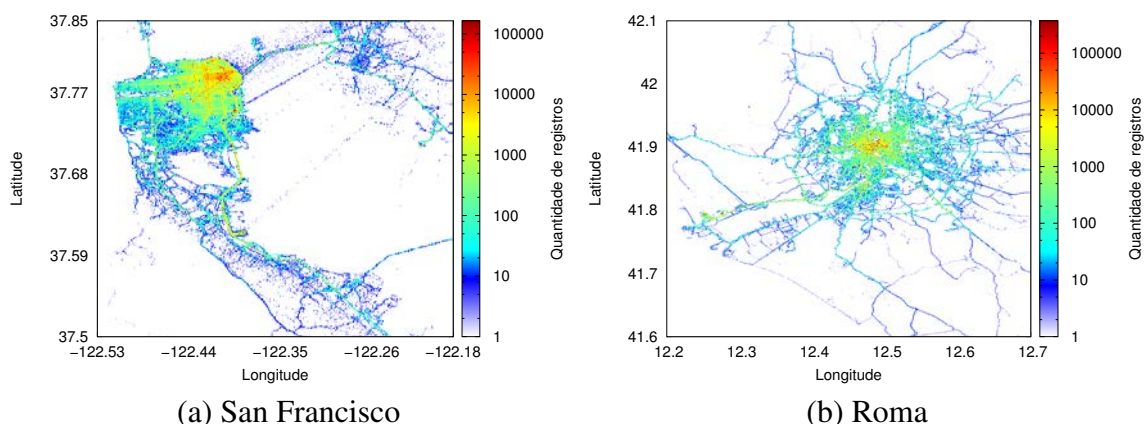


Figura 1. Mapas de calor das cidades considerando toda a base de dados

ferramenta usa os mesmos mapas do *My Maps*, utilizado para apresentar as rotas mais frequentes. Assumimos que o táxi percorreu a menor distância entre cada um dos trechos registrados na base de dados de San Francisco. No entanto, o limite de uso gratuito é de 2.500 requisições diárias, um valor bastante pequeno para a quantidade de dados em questão. As latitudes e longitudes da base de dados de San Francisco possuem 5 casas decimais, o que significa uma precisão de 1,132 metro [Rossi et al. 2015, Wikipedia 2015]. Acreditamos que tamanha precisão não é necessária para este trabalho. Para reduzir a quantidade de buscas na API do Google Maps, arredondamos as coordenadas geográficas de San Francisco de 5 para 4 casas decimais, reduzindo a precisão para 11,132 metros [Wikipedia 2015], o que entendemos ser razoável para um automóvel.

Avaliamos se há diferença significativa entre as distâncias convencionais e após a inferência de táxis parados, utilizando o teste estatístico de Kolmogorov-Smirnov para duas amostras. Esse teste foi escolhido por ser não paramétrico, sendo então mais robusto [Gibbons and Chakraborti 2003].

4. Resultados

Esta seção aborda os resultados das análises das bases de dados de San Francisco e Roma. A seção 4.1 trata da distribuição espacial dos registros em ambas as cidades e o cálculo das distâncias das rotas de San Francisco. A seção 4.2 discute a quantidade de táxis em atividade em cada base de dados no decorrer dos dias e da semana.

4.1. Análise Espacial

Frequentemente, os veículos não se deslocam aleatoriamente, e sim seguem padrões em suas trajetórias. Um dos objetivos do monitoramento e rastreamento de veículos é possibilitar a análise de mobilidade humana, para que seja possível detectar anomalias e prever comportamentos futuros. Uma vez obtidas, essas informações poderão ser utilizadas por gestores responsáveis em ações de planejamento do sistema de transporte com intuito de melhorar os deslocamentos dentro das cidades.

A Figura 1 apresenta a região abrangida pelas bases de dados de San Francisco e Roma utilizando mapas de calor em escala de intensidade logarítmica para ambas as cidades. O mapa de calor é uma representação gráfica de pontos em uma matriz colorida. No caso da Figura 1, é representada a quantidade de registros da base de dados para cada

posição geográfica (latitude e longitude). Quanto menor for a quantidade de registros em uma posição do mapa, mais próximo de branco e azul será a cor. No caso inverso, quanto maior for a quantidade de registros em uma posição, mais próximo de vermelho será a cor registrada no mapa.

Na Figura 1(a), nota-se maior intensidade de registros no nordeste da cidade de San Francisco, incluindo principalmente as rodovias 101 e 280 que conectam a região central ao “Aeroporto Internacional de San Francisco” (latitude 37.62 e longitude -122.38); a famosa ponte “San Francisco-Oakland Bay Bridge” (latitude 37.8 e longitude -122.38); e a rodovia 80 em direção a Oakland (latitude 37.8 e longitude -122.3). A maior quantidade de registros da base de dados está centralizada na rua “*Yellow Cab Access Rd.*” (latitude 37.751 e longitude -122.412), onde está localizada a “*San Francisco Yellow Cab Co-op*”, companhia de táxis amarelos e seu local de estacionamento.

Em relação ao mapa de calor de Roma apresentado na Figura 1(b), pode-se notar uma concentração maior no centro da cidade. Uma hipótese para essa maior concentração de registros de táxi está na mobilidade urbana da população local e de turistas para pontos famosos na região, tais como o “Coliseu”, o “Vaticano”, o “Pantheon” e as basílicas. A concentração de registros em Roma diminui à medida em que se distancia do centro. Porém, essa concentração se mantém em auto estradas importantes como a “Autostrada Roma - Aeroporto de Fiumicino” que conecta Roma ao “Aeroporto Internacional Leonardo da Vinci” localizado na coordenada latitude 41.8 e longitude 12.24.

A análise visual dos mapas de calor evidencia a maior demanda dos serviços de táxi para as regiões centrais e aeroportos internacionais de ambas as metrópoles. Esses padrões espaciais também retratam a importância das vias que ligam o centro das cidades aos aeroportos para os serviços de transporte público. Além disso, percebe-se que pode haver uma relação entre o número dos táxis e os pontos turísticos de ambas as cidades.

4.1.1. Inferência de Táxis Parados e Rotas Anômalas

A inferência de táxis parados é importante para o cálculo das distâncias percorridas pelas rotas porque erros de GPS podem alterar as coordenadas geográficas dos táxis mesmo que o veículo não se mova. Considerar as distâncias estimadas devido a variações de GPS quando o táxi estava parado pode resultar em valores anômalos.

Como exemplo de rota com valores anômalos está uma feita pelo táxi “new_ejesbay”. Essa rota ocorreu com o táxi sem passageiro, durou 15 horas, 28 minutos e 14 segundos e mediu 111,263 km. Todos os registros dessa rota são localizados na quadra onde está sediada a “*San Francisco Yellow Cab Co-op.*”. Conjecturamos que o comprimento da rota chegou a 111,263 km porque, mesmo com o táxi estacionado, as localizações do GPS variavam ao redor daquele mesmo local dando a impressão de que o táxi estava se movendo aleatoriamente pela região. Esse problema é ainda intensificado por possíveis imprecisões da API utilizada e pela distância ser calculada seguindo as vias e sentido permitidos. Nesse caso, uma variação de GPS que em linha reta mediria apenas 12 metros, teve a distância estimada em 958 metros pela API do Google Maps. Para contornar esse problema, aplicamos um processo de inferência para identificar se o táxi estava estacionado ou não, antes de somar as distâncias dos trechos de uma rota.

Os trabalhos [Zhu et al. 2012, Ding et al. 2015] propõem métodos de clusterização baseados em densidade para inferir se o táxi estava ou não estacionado. Uma vez que para este trabalho não é necessário identificar o local onde os táxis estavam estacionados, mas apenas se o táxi estava parado ou não, adaptamos a heurística proposta em [Zhu et al. 2012] ao cálculo das distâncias percorridas pelos táxis. Os autores definem um táxi como parado se após τ minutos (no mínimo) o veículo movimentou menos de δ metros. Os parâmetros utilizados no artigo citado e também neste trabalho foram $\tau = 3$ minutos e $\delta = 50$ metros. A medida de distância empregada nessa heurística foi a distância de Haversine [Robusto 1957] considerando o raio médio da Terra igual a 6.371 km, conforme também implementado em [Thang et al. 2012].

Uma vez que as bases de dados não possuem a informação de quando o táxi estava parado, não foi possível obter as quantidades de elementos relevantes, verdadeiros positivos e falsos positivos para calcular as medidas de *precision* e *recall* da inferência. Portanto, após estimar as distâncias corrigidas para as rotas, avaliamos se houve mudança significativa entre as distribuições de distâncias antes e após a inferência de táxis parados.

Considerando que é provável que um táxi estacionado esteja sem passageiro, e que é pouco provável que o táxi fique estacionado durante uma rota com passageiro, espera-se que haja diferença maior para as rotas com o táxi desocupado e que não haja diferença significativa para as rotas com o táxi ocupado. Avaliamos a hipótese nula de igualdade entre as distâncias das rotas antes e após a inferência de táxis parados. O teste foi realizado de três formas: entre todas as distâncias de rotas da base de dados; somente entre rotas com o táxi ocupado; e somente entre rotas com o táxi desocupado. Para o conjunto de rotas e para as rotas com o táxi desocupado o p-valor retornado com o teste de Kolmogorov-Smirnov no software R foi menor que $2,2 \times 10^{-16}$. Então, podemos dizer, com 5% de significância, que há evidências estatísticas suficientes para se rejeitar a hipótese nula de igualdade entre as distribuições. Porém, avaliando somente as distâncias com o táxi ocupado, o p-valor foi de 0,4296. Portanto, com 5% de significância, não temos evidências estatísticas para rejeitar a hipótese nula de igualdade entre as distribuições.

A técnica implementada tem efeito estatisticamente significativo nas distâncias dos trajetos e nas distâncias das rotas com o táxi desocupado, no sentido que as distribuições das distâncias antes e após a inferência de táxis parados são diferentes. Porém a técnica não causa impacto significativo nas distâncias das rotas com o táxi ocupado, uma vez que o teste utilizado não rejeitou a hipótese de igualdade entre as distâncias das rotas antes e após a inferência.

Em relação ao exemplo da rota anômala com o táxi desocupado, mencionada anteriormente, a distância caiu de 111,263 km para 1,696 km após a aplicação da técnica. Porém, não houve alterações relevantes nos quartis, médias, moda e mediana das distâncias convencionais e das distâncias após a inferência de táxis parados. Esses resultados indicam que a técnica implementada é válida para o tratamento das distâncias percorridas pelos táxis. No geral, não há impactos nas estatísticas das distâncias das rotas, com exceção das rotas anômalas obtidas com táxis estacionados.

A menor distância e também a moda das distâncias das rotas com o táxi ocupado e desocupado, antes e após a inferência foram iguais a zero. Rotas medindo zero metro mesmo antes da inferência de táxis parados podem ter ocorrido por falhas no dispositivo

que registra o *status* de ocupação das rotas, equívocos do taxista ao operar esse dispositivo, ou em situações quando o taxista termina uma rota com passageiro e logo após inicia outra rota com outro passageiro [Castro et al. 2013]. Nesse último caso, pode ser que não tenha passado tempo o suficiente para obter ao menos um registro com o táxi desocupado, e assim, separar essas duas rotas.

A situação em que não houve ao menos um registro para separar as duas rotas pode explicar a existência de rotas com distâncias enormes mesmo após a inferência de táxis parados. Embora a média das distâncias com o táxi ocupado seja 6,305 km antes da inferência e 6,283 km após a inferência, o percentil 99 para essas mesmas medidas foi de 50,723 km e 50,579 km, respectivamente.

Em relação às distâncias com o táxi desocupado, a média antes da inferência de táxis parados foi de 6,935 km e após a inferência foi de 6,684 km. Os percentis 99 foram de 77,646 km e 76,162 km para antes e após a inferência, respectivamente. A maior rota com o táxi desocupado teve 1.391.593 km antes da inferência e 1.327.492 km após a inferência de táxis parados. Esses resultados indicam que mesmo tratando a questão dos táxis parados, rotas com distâncias anômalas ainda são presentes nos dados de San Francisco. As hipóteses para a causa dessas medidas anômalas são a falta de uma marcação do fim das rotas de táxi e equívocos dos taxistas ou erros do dispositivo ao registrar o *status* das rotas.

4.1.2. Análise das Rotas Mais Frequentes de San Francisco

A análise das rotas mais frequentes dos táxis ocupados em uma cidade é importante para sugerir novos pontos de táxi, orientar melhor os taxistas na busca do próximo passageiro, propor melhorias no transporte público e para planejamento das vias urbanas. Como a base de dados de Roma não possui alguma informação indicando o início e final das rotas, foram analisadas as rotas mais frequentes somente da cidade de San Francisco.

A Figura 2 apresenta as dez rotas com maior número de ocorrências. As linhas em azul representam as rotas de menor percurso de carro entre os pontos de origem e fim segundo o Google Maps. Os locais em verde são o início e os locais em vermelho são os destinos das rotas. O local em amarelo representa a rota, dentre as mais frequentes, que iniciou e terminou na mesma região. Na Figura 2(a) são mostradas todas as dez rotas identificadas. Há basicamente dois grupos de rotas: um na região central da cidade, mostrado na Figura 2(b), e outro na parte inferior do gráfico localizado no “Aeroporto Internacional de San Francisco”. A rota em amarelo ocorreu na “*Yellow Cab Acces Rd.*”, rua que dá acesso ao “San Francisco Yellow Cab Co-op” mencionado anteriormente.

Alguns pontos da cidade alternaram-se como locais de início e fim das dez rotas mais frequentes. Na Figura 2(b), os pontos que aparecem mais de uma vez dentre as rotas mais frequentes são: a praça “Union Square” (ponto em verde no centro horizontal do mapa), o *shopping* “Pier 39” (na parte superior), e o hotel “San Francisco Marriot Marquis” (na parte inferior do mapa). Em relação ao hotel San Francisco Marriot Marquis, as duas rotas mais frequentes ligavam o hotel ao Aeroporto Internacional da cidade, tendo o início e fim dessas rotas situados em saídas diferentes do hotel e do aeroporto. Outros dois locais apareceram entre as dez rotas mais frequentes somente como terminos de

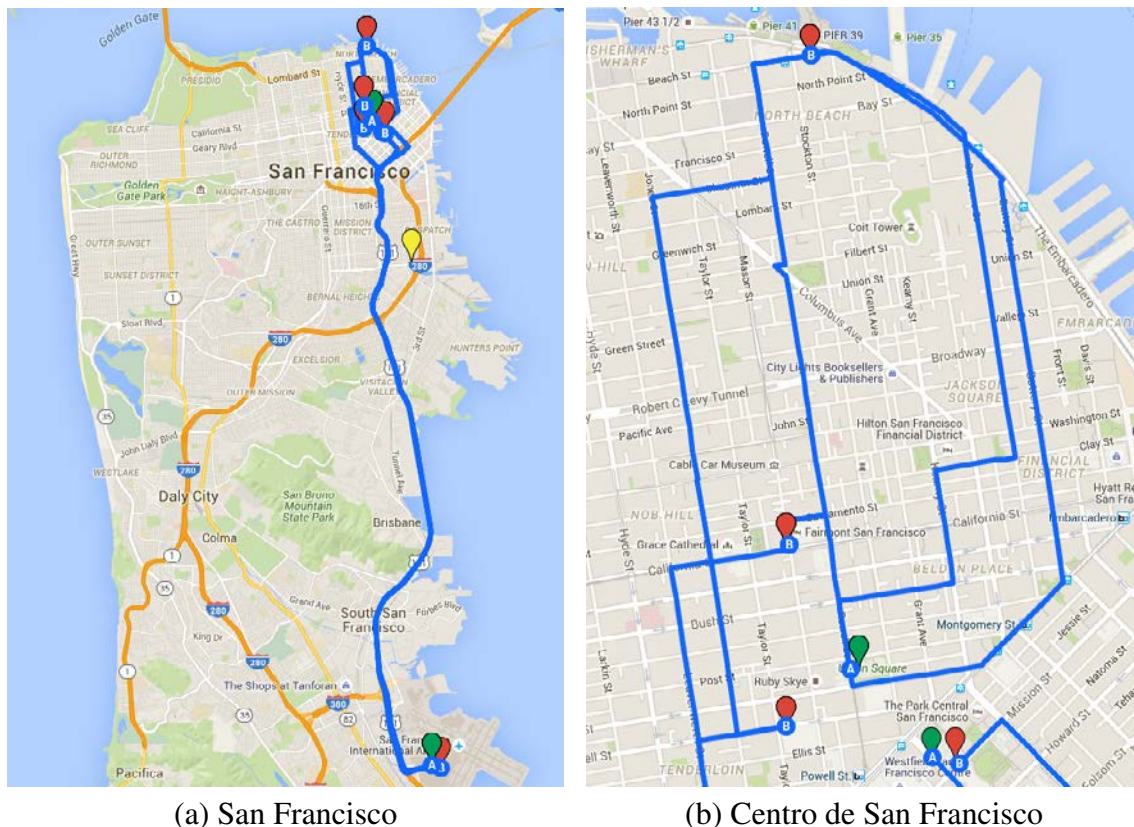


Figura 2. Rotas mais frequentes de San Francisco

rotas: “Pacific-Union Club” e o hotel “Hilton San Francisco Union Square”.

Todas as rotas identificadas ficam localizadas na região leste da cidade. Para todas as rotas que ligam o centro de San Francisco ao aeroporto, o Google Maps sugeriu utilizar a rodovia 101. Pela diferença de cores no mapa de calor da Figura 1(a), imagina-se que os táxis também escolhem a rodovia 101 para realizar esse trajeto.

O fato da maioria das rotas mais frequentes compartilharem os mesmos locais de início e fim indica que certos pontos de interesse de San Francisco como o Aeroporto Internacional, hotéis, o Union Square e o *shopping* “Pier 39” representam fontes importantes de clientes para os serviços de táxi. Além disso, as análises revelam uma relação de rotas entre o aeroporto e os hotéis da cidade. Acreditamos que essa relação se deve principalmente por turistas durante a chegada e saída de San Francisco.

4.2. Análise Temporal em Ambas as Bases de Dados

Compreender as dinâmicas dos táxi no decorrer do dia e da semana é importante para o gerenciamento das próprias agências táxi, para o planejamento de transportes públicos concorrentes, e para residentes e visitantes que utilizam ou dependem do serviço de táxi.

A madrugada é o momento de menor quantidade de registros diários em ambas as bases de dados de San Francisco e Roma. A menor presença de táxis na rua em San Francisco foi entre 03:00 e 04:00 horas (2,1% da quantidade diária). Em relação à Roma, a menor quantidade diária foi das 04:00 às 05:00 (1,05% do total diário). As menores taxas de táxis ocupados da base de dados de San Francisco ocorrem das 03:00 até antes

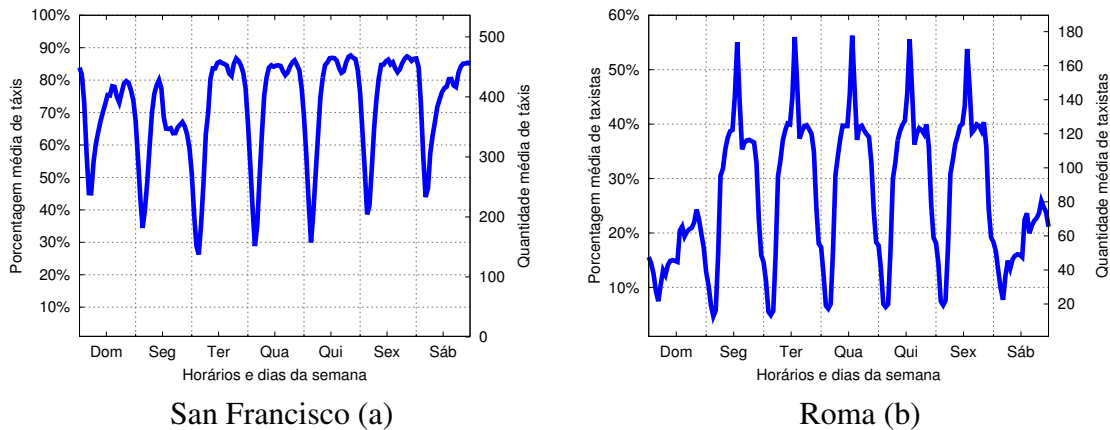


Figura 3. Quantidade de táxis e taxistas no decorrer da semana

das 07:00; nesse período, as taxas de ocupação por hora oscilam de 26,1% a 27,94%. Após as 04:00, o número de registros aumenta regularmente e se estabiliza por volta das 09:00 em ambas as bases de dados. Acreditamos que esse comportamento se deve à menor demanda de passageiros durante a madrugada. Portanto, além de haver menos rotas com os táxis ocupados, os taxistas parecem preferir estacionar o táxi e esperar até que um passageiro surja em vez de trafegar pelas ruas procurando o próximo passageiro.

Em San Francisco, a quantidade de registros oscila pouco das 09:00 até as 16:00. Uma hipótese para a demanda de táxis não ter aumentado durante o horário do almoço é de que a população local evita almoçar em locais distantes do serviço, não precisando utilizar táxis. A Figura 3(a) apresenta uma queda na quantidade de táxis em todos os dias da semana (principalmente na segunda-feira) próximo ao horário de almoço. Imaginamos que essa queda seja causada por troca de taxistas que dirigem o mesmo táxi, ou uma pausa dos próprios taxistas para almoçar. A quantidade de registros cresce regularmente das 16:00 até as 20:00, horário da maior porcentagem diária (5,02%). Das 17:00 até as 23:00 a quantidade de táxis ocupados ultrapassa a quantidade de táxis desocupados. Após as 18:00, a quantidade de registros cai até as 04:00 do dia seguinte. Essa demanda de táxis, principalmente ocupados, pode ser o reflexo das pessoas deixando os locais de trabalho após o horário comercial e dirigindo-se para suas casas ou outros destinos.

Em relação à Roma, o maior percentual de registros ocorre por volta das 13:00 (5,95%), tendo uma queda às 15:00 (4,65%) e voltando a subir até às 17:00 (5,93%), de onde cai até o menor valor no dia seguinte. A Figura 3(b) ilustra esse comportamento com a quantidade média de taxistas no decorrer da semana. Nota-se que, diferentemente de San Francisco, há um pico de demanda de taxistas no horário do almoço e os dados seguem um padrão muito semelhante de segunda a sexta, variando somente nos sábados e domingos, quando a porcentagem média de taxistas não chega a 30%. Acreditamos que essa diferença pode ser explicada devido à população de Roma ter mais costume de almoçar em locais distantes de onde trabalha, e pelos dados não serem referentes a táxis, e sim a taxistas que seguem jornadas de trabalho controladas durante o dia e semana.

A segunda-feira é o dia da semana com menor quantidade média de táxis em San Francisco, com a média de táxis variando entre 65% e 80% do total. A média dos táxis não alcança 90% dos 536 táxis em nenhum dia da semana. Para todos os dias da

semana, a quantidade média de táxis desocupados é maior que a quantidade média de táxis ocupados. As menores porcentagens diárias diminuem do começo de domingo até terça-feira. Após a terça-feira, essas menores porcentagens diárias começam a aumentar até o final da semana. Um comportamento diferente se percebe com as maiores porcentagens diárias, as quais oscilam pouco de terça à sábado tendo uma queda no domingo e segunda. Esse padrão pode ser o reflexo de eventos sociais da população de San Francisco que aumentam à medida em que o final de semana se aproxima.

Outro ponto diferente entre as duas cidades está nas porcentagens de uso do sistema de táxi. Enquanto em Roma as porcentagens médias não chegam à 60%, em San Francisco a porcentagem ultrapassa os 80% ao menos uma vez por dia. Em nenhum momento o percentual de taxistas operando em Roma supera a quantidade de táxis de San Francisco. Embora essas duas bases de dados tenham características semelhantes, os padrões temporais analisados das rotas de táxi indicam costumes e rotinas diferentes dentre as duas metrópoles.

5. Conclusão

Neste trabalho foram comparados os padrões espaciais e temporais da circulação de táxis de San Francisco e Roma; foi discutido o efeito da inferência de táxis parados ao calcular a distância das rotas; e foram apresentadas as rotas mais frequentes da base de dados de San Francisco. As análises espaciais mostram a maior atividade de táxis nos centros das cidades e que se distribuem pelas maiores avenidas de trânsito rápido das regiões. As análises temporais evidenciam diferenças no padrão de utilização dos serviços de táxis nas duas cidades, que podem refletir diferenças culturais ou mesmo legais desse serviço.

Em relação às rotas de táxi, foram apresentados resultados da eficiência de inferir táxis parados para evitar erros no cálculo de distâncias das rotas. Assim como o esperado, foram encontradas evidências estatísticas com 5% de significância de que as distâncias das rotas geradas após inferir e desconsiderar os trechos com o táxi parado são diferentes que as distâncias convencionais.

Como trabalhos futuros, sugere-se ajustar os parâmetros τ e δ da inferência de táxis parados para diferentes situações do tráfego de veículos nas estradas tais como congestionamentos. Também propõe-se que, embora haja o *status* de ocupação nos dados de San Francisco, seja inferido quando cada rota iniciou e terminou a fim de reduzir a quantidade de rotas anômalas da base de dados. Além disso, sugerimos que seja apresentada a variação dos locais de início e fim das rotas mais frequentes de San Francisco no decorrer dos horários dos dias e no decorrer dos dias da semana.

6. Agradecimentos

Os autores agradecem ao CEFET-MG e aos financiadores dos projetos INCT InWeB (MCT/CNPq 573871/2008- 6) e MASWeb (FAPEMIG/PRONEX APQ-01400-14).

Referências

- 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.
- Bracciale, L., Bonola, M., Loreti, P., Bianchi, G., Amici, R., and Rabuffi, A. (2014). CRAWDAD dataset roma/taxi (v. 2014-07-17). Downloaded from <http://crawdad.org/roma/taxi/20140717>.

- Castro, P. S., Zhang, D., Chen, C., Li, S., and Pan, G. (2013). From Taxi GPS Traces to Social and Community Dynamics: A Survey. *ACM Comput. Surv.*, 46(2):17:1–17:34.
- Chen, C. (2014). *Understanding social and community dynamics from taxi GPS data*. PhD thesis, Evry, Institut national des télécommunications.
- Cunha, F. D., Alvarenga, D. A., Viana, A. C., Mini, R. A., and Loureiro, A. A. (2015). Understanding Interactions in Vehicular Networks Through Taxi Mobility. In *Proceedings of the 12th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, pages 17–24. ACM.
- Ding, L., Fan, H., and Meng, L. (2015). Understanding taxi driving behaviors from movement data. In *AGILE 2015*, pages 219–234. Springer.
- Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric statistical inference*. Marcel Dekker, New York.
- Kellner, L. and Egger, R. (2016). Tracking Tourist Spatial-Temporal Behavior in Urban Places, A Methodological Overview and GPS Case Study. In *Information and Communication Technologies in Tourism 2016*, pages 481–494. Springer.
- Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., and Yang, Q. (2011). Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 63–68. IEEE.
- Liu, S., Wang, S., Liu, C., and Krishnan, R. (2015). Understanding taxi drivers’ routing choices from spatial and social traces. *Frontiers of Computer Science*, 9(2):200–209.
- Murashkin, M. and Geroliminis, N. (2015). Analysis of taxi drivers’ behavior with GPS data from a megacity. In *hEART*, Lyngby. European Association for Research in Transportation (hEART), Department of Transport at Technical University of Denmark (DTU Transport).
- Piorkowski, M., Sarafjanovic-Djukic, N., and Grossglauser, M. (2009). CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from <http://crawdad.org/epfl/mobility/20090224>.
- Robusto, C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):38–40.
- Rossi, L., Walker, J., and Musolesi, M. (2015). Spatio-temporal techniques for user identification by means of gps mobility data. *EPJ Data Science*, 4(1):1–16.
- Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., and Fotheringham, A. S. (2015). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, pages 1–26.
- Thang, N. V., Thang, P. M., and Tan, T. D. (2012). The performance improvement of a lowcost ins/gps integration system using the street return algorithm. *Vietnam Journal of Mechanics*, 34(4):271–280.
- Valero, B., Luis, J., Julián, A., Belén, A., Villén, G., and Natalia (2014). *GNSS. GPS: Fundamentos y aplicaciones en geomática*. Editorial de la Universidad Politécnica de Valencia, Valencia.
- Wikipedia (2015). Decimal Degrees. https://en.wikipedia.org/wiki/Decimal_degrees.
- Zhang, Y. (2014). How do taxis work in beijing? Master of urban and regional planning, University of California, Los Angeles, Los Angeles.
- Zheng, Z., Rasouli, S., and Timmermans, H. (2015). Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data. *Geographical Analysis*.
- Zhu, Y., Zheng, Y., Zhang, L., Santani, D., Xie, X., and Yang, Q. (2012). Inferring taxi status using gps trajectories. *arXiv preprint arXiv:1205.4378*.