

Análise de Mobilidade Urbana em Diferentes Estratos Sociais a partir de uma Rede de Telefonia Móvel

Francielly Rodrigues¹, Carolina R. Xavier², Alexandre G. Evsukoff³,
Artur Ziviani¹, Vinícius da F. Vieira²

¹Laboratório Nacional de Computação Científica (LNCC)
Petrópolis – RJ – Brazil

²Dep. de Ciência da Computação – Universidade Federal de São João del Rei (UFSJ)
São João del Rei – MG – Brazil

³COPPE – Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro – RJ – Brazil.

{fmunique, ziviani}@lncc.br; {vinicius, carolinaxavier}@ufs.j.edu.br;

alexandre.evsukoff@coc.ufrj.br

Resumo. *A mobilidade urbana de um cidadão pode ser compreendida como as sequências de trajetos percorridos por ele. Os estratos sociais nos quais os indivíduos estão contextualizados, contudo, configuram um fator determinante para a organização espacial de uma cidade. O objetivo deste estudo é investigar padrões de mobilidade e analisar características dos estratos sociais no estado do Rio de Janeiro utilizando dados de telefonia móvel, de forma a analisar não só a movimentação da população como um todo, mas também realizar comparações entre diferentes classes sociais. Foi identificado que as classes sociais obedecem o mesmo padrão de distâncias percorridas diariamente, porém com as classes mais altas percorrendo menores distâncias. Outro ponto analisado foi o número médio de ligações, cuja distribuição de probabilidades apresenta uma cauda longa, comportamento frequente em análises de comunicação. A principal contribuição do presente estudo é a metodologia desenvolvida, que pode funcionar como um padrão para várias bases de dados georreferenciadas, bastando somente ajustá-la à granularidade desejada.*

Abstract. *Urban mobility of a citizen can be understood as the sequences of trips followed by him. The social classes in which individuals are contextualized, however, denote a significant factor for the spatial organization of a city. The goal of this study is to investigate mobility patterns and analyze characteristics of social classes in the Rio de Janeiro State using cell phone data together with census data, in order to analyze not only the movement of the population as a whole but also to make comparisons between different social classes. It is clear that the classes follow the same pattern of distances traveled daily, but with higher classes traveling shorter distances. Another analyzed topic was the average number of calls, whose probability distribution follows a heavy tail, a frequent behavior in communication analysis. The main contribution of the present study is the developed methodology, which can work for several georeferenced databases, by simply adjusting it to the desired granularity.*

1. Introdução

Para viver em sociedade, as pessoas precisam se comunicar e se locomover. A comunicação é a base dos relacionamentos humanos e, com a popularidade dos dispositivos móveis, comunicar-se com outras pessoas tornou-se uma tarefa simples, que influencia diretamente na formação de grupos sociais. Por outro lado, pertencer ou não a um determinado grupo social tem forte relação à formação dos padrões de locomoção de um indivíduo. Assim, os possíveis deslocamentos necessários para realizar atividades cotidianas tais como frequentar a escola ou o trabalho, fazer compras ou mesmo se divertir podem ser inferidos a partir da similaridade dos trajetos entre indivíduos que possuem algum vínculo social entre si [Toole et al., 2015b]. O conjunto dos trajetos percorridos em uma área urbana por um indivíduo para deslocar-se entre locais associados às suas atividades caracteriza a mobilidade urbana.

A mobilidade urbana pode ser entendida como as trajetórias adotadas pelos seres humanos diariamente, ou seja, a natureza dos deslocamentos dentro de uma região metropolitana [Duarte et al., 2012, Csáji et al., 2013]. A sequência de locais visitados por um indivíduo constitui o seu padrão de mobilidade. Padrões de mobilidade normalmente podem ser generalizados para grupos de pessoas e são obtidos através da análise da movimentação humana diária, que pode ser capturada na análise de dados georreferenciados, ou seja, dados com informações complementares de localização [Hill, 2009].

A crescente utilização de serviços de comunicação como Internet, ligações e mensagens em dispositivos móveis gerou, e ainda gera, enormes bancos de dados georreferenciados [Barlacchi et al., 2015]. Nesse contexto, dados georreferenciados muito importantes nos dias de hoje são os dados de *Call Detail Record* (CDR), armazenados por diferentes operadoras de telefonia que devem trocar informações com a finalidade de efetuar a cobrança devida pela utilização das antenas pelos seus clientes [Barlacchi et al., 2015].

Como na sociedade atual as pessoas estão conectadas quase todo o tempo, é possível traçar trajetórias utilizando as sequências de localizações das torres com as quais o dispositivo se comunicou. Esses registros podem auxiliar tanto na tomada de decisões para melhorar diversos fatores urbanos quanto para compreender como indivíduos se locomovem e qual a estrutura da comunicação entre eles. Dentre outras finalidades, isto possibilita o estudo da mobilidade urbana aplicada a várias áreas como planejamento urbano [Gonzalez et al., 2008], espalhamento de doenças [Vazquez-Prokopec et al., 2013] e análise de interação entre grupos ou estratos sociais [Loterio et al., 2015]. Sabendo que a distribuição e organização dos estratos sociais definem a estrutura das cidades, caracterizar a mobilidade urbana levando em consideração esta distribuição espacial é também fundamental para estudar o comportamento da sociedade. Tal estudo é possível através do cruzamento de dados censitários e dados georreferenciados [Loterio et al., 2015].

O objetivo deste estudo é analisar os padrões de mobilidade e interação social de indivíduos de diferentes estratos sociais do Rio de Janeiro a partir de informações obtidas através do uso de uma rede de telefonia móvel, que gera uma enorme base de dados CDR. A escolha do Rio de Janeiro é justificada por sua formação heterogênea no que diz respeito a existência de comunidades carentes convivendo lado a lado com áreas onde predomina um padrão de vida economicamente elevado. A principal contribuição é a metodologia desenvolvida, que pode funcionar como um padrão para várias bases de dados georreferenciadas, bastando somente ajustá-la à granularidade desejada. Com a análise combinada

desses dados de CDR junto a informações de classes sociais pretende-se identificar não somente tendências de comportamento e rotina, mas investigar se os indivíduos formam padrões de acordo com o seu nível sócio-econômico. O propósito geral é agregar conhecimento sobre questões que podem ser resolvidas, ou pelo menos analisadas, utilizando dados georreferenciados, dados censitários e informações sobre mobilidade urbana.

Foi identificado que as classes sociais obedecem o mesmo padrão de distâncias percorridas diariamente, porém com as classes mais altas percorrendo menores distâncias. Outro ponto analisado foi o número médio de ligações, cuja distribuição de probabilidades segue uma lei de potência, comportamento frequente em análises de comunicação. Adicionalmente, analisando os padrões de locomoção diária mais frequentes, também conhecidos como motifs [Schneider et al., 2013], foi possível perceber que a tendência é que a população siga padrões pequenos, de uma ou duas localizações distintas.

O restante do artigo está organizado da seguinte maneira. A Seção 2 aborda os principais conceitos discutidos ao longo do estudo, juntamente com uma revisão da literatura apresentando alguns trabalhos relacionados. A Seção 3 apresenta a metodologia utilizada para o desenvolvimento deste estudo, bem como as ferramentas e métricas adotadas. A Seção 4 apresenta os resultados e uma discussão sobre as análises realizadas. Por fim, na Seção 5, são apresentadas as conclusões, contribuições e limitações deste trabalho.

2. Fundamentos Teóricos e Trabalhos Relacionados

A localização geográfica complementa outros tipos de informação [Longley, 2005, Hill, 2009], sendo útil no estudo de padrões de mobilidade a fim de melhor compreender e apoiar decisões nas áreas urbanas. Grandes eventos, por exemplo, são capazes de temporariamente modificar o comportamento esperado nas rotinas diárias dos indivíduos e conseqüentemente o fluxo de pessoas em um determinado local, o que afeta a organização urbana de forma significativa [Xavier et al., 2012, Marques-Neto et al., 2018].

Para obter esse tipo de informação é preciso realizar o georreferenciamento, que consiste em relacionar dados com as respectivas localizações geográficas. As coordenadas de latitude e longitude são as informações mais utilizadas para construir mapas e calcular distâncias e escalas de representação [Longley, 2005].

Vários tipos de dados podem ser representados utilizando informações georreferenciadas de latitude e longitude aliadas à escolha do sistema de referência correto, como é o caso de dados do tipo CDR (*Call Detail Records*). Dados CDR são registros de atividades de comunicação entre um dispositivo e uma torre telefônica, que representam a atividade telefônica da população em determinada região, isto é, ligações, mensagens de texto e internet. Eles são gerados por empresas telefônicas que armazenam as atividades dos clientes de acordo com a antena na qual os mesmos se conectam nesses períodos ativos com o propósito de permitir a correta cobrança pelos serviços prestados pelas operadoras aos clientes.

Embora sejam fontes de informações bastante ricas sobre a utilização da rede de telefonia, é importante ressaltar que registros CDR individualmente tem muitas limitações quando utilizados para análise do comportamento de pessoas. Primeiramente, eles são gerados apenas quando indivíduos realizam alguma atividade telefônica, o que faz com que os mesmos se tornem invisíveis para os registros o restante do tempo. Outra limitação

é a granularidade das localizações registradas, pois é atribuída ao usuário a localização da antena na qual ele se conectou, causando uma incerteza da real localidade do indivíduo, visto que o alcance das antenas pode chegar a 3km.

Ainda assim esse tipo de registro é frequentemente utilizado na literatura para investigação de padrões de mobilidade. Seu uso se justifica não só pela disponibilidade de grandes bases de dados, mas também por possuírem informações de toda a cidade, o que torna viável a análise para uma grande área geográfica com baixo custo. Adicionalmente, esse tipo de dado reflete observações provenientes de comunicações reais de milhares de pessoas, ao invés de informações reportadas pelos próprios usuários. Tal combinação de dados pessoais e dados de localização fornecida pelos CDR fazem dos mesmos uma rica fonte de informação [Blondel et al., 2015].

Em diversos trabalhos relacionados à modelagem do comportamento de pessoas em contextos urbanos, dados CDR tem sido utilizados para presumir localizações como a casa e o trabalho dos indivíduos. O trabalho de Csáji *et al.* [Csáji et al., 2013] utiliza uma grande base de dados de telefonia móvel para correlacionar características do comportamento humano. Um ponto importante desenvolvido nesse estudo é o pré-processamento dos dados. Uma vez que a posição da antena não é necessariamente a posição do usuário, foi preciso aplicar um filtro de média ponderada que funciona atenuando as posições de cada usuário ao longo do tempo. Nesse trabalho, os autores também desenvolvem um procedimento para extrair locais frequentes dos usuários e através de clusterização semanal confirmam a hipótese geralmente assumida de que há apenas dois lugares identificados com clareza: casa e trabalho.

Recentemente, Barlacchi *et al.* [Barlacchi et al., 2015] disponibilizaram uma série de dados relacionados aos contextos urbanos de Milão e Trentino, como dados CDR, dados de Twitter, dados de consumo de eletricidade, entre outros. Nesse trabalho, os autores enfatizam os dados CDR como uma nova área para a análise de comportamento de pessoas em cenários urbanos. Já em [Silveira et al., 2016] é elaborada uma nova família de modelos orientados a dados, chamada MobHet, a partir de fontes de dados heterogêneas realizando uma combinação de recursos. Esse tipo de abordagem, porém, requer alta disponibilidade de várias fontes de dados diferentes de uma mesma localização.

Levando em consideração que dados CDR são criados sempre que uma nova atividade é detectada pelas torres de celular, várias técnicas foram desenvolvidas para obter latitude e longitude dos usuários utilizando esse tipo de dado, com destaque para trabalhos baseados no cell-ID, que é o identificador de uma torre disponível nos dados CDR. Presumir os locais mais frequentes dos indivíduos viabiliza o surgimento de padrões de mobilidade.

Já [Schneider et al., 2013] consideram a mobilidade urbana como uma rede e identificam 17 padrões de mobilidade, ou seja, sub-redes que ocorrem mais frequentemente entre os usuários. Essas sub-redes são representadas em forma de grafos direcionados, onde os vértices são as localizações das antenas nas quais houve conexão naquele dia e as arestas indicam a sequência do caminho percorrido. Esses grafos são chamados de MOTIFS, e conseguem representar até 95% da mobilidade urbana dos dados analisados. É portanto proposto um modelo de perturbação que reproduz esses padrões. Esse modelo, porém, reproduz a atividade durante um único dia, e tem a limitação de ser baseado em

um período curto de observação, não podendo predizer o comportamento ao longo do tempo.

Nenhum dos trabalhos citados, porém, relaciona o comportamento observado no estudo da mobilidade urbana com dados censitários. Neste trabalho, a mobilidade de indivíduos no espaço urbano é analisada considerando os estratos sociais nos quais as pessoas estão contextualizadas. Informações de estratificação social são um bom exemplo de dados pertinentes ao estudo da mobilidade, pois tornam possível realizar uma análise do ponto de vista socioeconômico. Segundo [Saunders, 2006], a estratificação social é uma área de estudo que refere-se às diferentes camadas, ou estratos, dos grupos sociais. O papel dessa área é estudar como esses grupos se relacionam uns com os outros. No presente estudo, os estratos sociais estão relacionados à classe econômica dos indivíduos.

Para obter informações de estratos sociais, dados de rendimento são capturados como parte do Censo Demográfico, uma Pesquisa realizada pelo IBGE [IBGE, 2010] a cada dez anos onde são reunidas informações sobre toda a população brasileira. A coleta é feita por meio da aplicação de um questionário em visitas aos domicílios do país. Depois, essas informações são analisadas e divulgadas em publicações no próprio site do IBGE, de acesso livre. Dentre os dados coletados estão os rendimentos, de onde são extraídas as informações dos estratos sociais.

As tabelas do IBGE informam dados de renda por setores censitários. "O setor censitário é a menor unidade territorial, formada por área contínua, integralmente contida em área urbana ou rural, com dimensão adequada à operação de pesquisas e cujo conjunto esgota a totalidade do Território Nacional"[IBGE, 2011].

3. Metodologia

A metodologia proposta neste trabalho pode ser dividida em três etapas: (i) obtenção e tratamento dos dados; (ii) extração de informação útil a partir dos dados; e (iii) análise dos padrões encontrados nos dados. Dessa forma, de posse dos arquivos CDR em formato cru, é necessário modelá-los de forma que se possa tratar a base e limpar os registros incorretos ou inúteis. A partir da base de dados tratada, pode-se capturar trajetórias, presumir residências e realizar outras análises. A esses resultados preliminares, associa-se os dados censitários, que permitem a estratificação social dos identificadores, possibilitando assim as análises de comunicação e mobilidade.

Para a execução da metodologia proposta, duas bases de dados georreferenciadas são essenciais. Uma delas corresponde aos registros de atividades em uma rede de telefonia móvel realizadas pelos indivíduos de uma determinada região, que nesse trabalho são provenientes de uma operadora de telefonia móvel em formato CDR. A outra corresponde a informações sociais dos indivíduos, que nesse trabalho são definidas por dados censitários, disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) [IBGE, 2010] e englobam informações sócio-econômicas e geográficas da população, permitindo identificar as faixas de renda de cada setor.

Neste trabalho, a filtragem dos dados CDR corresponde à eliminação de registros duplicados e registros com campos incompletos. Além disso, foram eliminados identificadores com atividade telefônica em apenas um dia, uma vez que não é possível caracterizar padrões de mobilidade e residência para indivíduos com um número insuficiente de registros. Com esses dados, é possível realizar algumas análises básicas sobre a utilização

da rede de telefonia móvel, como número de ligações realizadas e recebidas por um ID, número de ligações diárias bem como trajetórias e distâncias percorridas.

A partir dos registros CDR, foi possível presumir as residências dos indivíduos a partir dos registros de utilização diária da rede de telefonia. A identificação das residências foi feita da seguinte maneira: para cada identificador de origem (ID1) da base, foram selecionados os registros entre 20:00 e 07:00 horas [Toole et al., 2015a], gerando uma nova tabela contendo apenas os registros noturnos. O par de coordenadas (*latitude, longitude*) escolhido como a residência foi aquele que, dentre os registros noturnos, apareceu durante mais dias distintos. Vale ressaltar que uma nova limpeza fez-se necessária após este passo, visto que um par de coordenadas que aparece em apenas um dia não caracteriza informação suficiente para presumir a residência.

As informações sociais dos indivíduos foram obtidas a partir da base de dados do CENSO 2010,¹ que disponibiliza um grande conjunto de atributos sociais e econômicos em diferentes níveis de granularidade (como unidade de federação, município, subdistrito, bairro e setor censitário). Para este trabalho, foi considerada a granularidade de setores censitários, a menor granularidade disponível, e o rendimento nominal médio do setor foi utilizado como uma estimativa do rendimento das residências associadas àquele setor.

A associação entre cada indivíduo e sua informação de rendimento, para sua classificação em um estrato social, foi feita com o auxílio do *software* Quantum GIS (QGIS) [Quantum, 2013], o que permitiu relacionar atributos de renda aos identificadores e às respectivas localizações geográficas de suas residências. A saída obtida a partir do uso do QGIS é uma nova tabela contendo, para cada indivíduo, o setor ao qual ele pertence e o par de coordenadas da residência presumida. Assim, é feita uma junção dessa tabela resultante com a tabela de renda do IBGE atribuindo para cada indivíduo sua informação de renda de acordo com o local da residência de forma a identificar o estrato social de cada um.

A divisão de classes foi a mesma adotada pelo IBGE, tendo como base o salário mínimo de 2010 (data do censo mais recente), de R\$510,00. O salário limite para cada uma das sete classes pode ser visto na Tabela 1.

Para calcular as distâncias percorridas pelos indivíduos por dia, utilizou-se a equação de Haversine [Shumaker e Sinnott, 1984], uma das mais utilizadas na literatura com esse propósito. De posse das distâncias das antenas, uma trajetória de pares (*latitude, longitude*) das conexões do usuário às torres foi criada, onde a distância percorrida é a distância entre todos os pares de antenas, de forma sequencial, conforme proposto por [Isaacman et al., 2011]. Os pares na trajetória foram organizados por hora da comunicação e agrupados por dia. Assim, para saber a distância diária total percorrida, basta analisar os pares de antenas dois a dois na trajetória e consultar a tabela de distâncias entre antenas para descobrir o valor a ser somado à distância total do dia analisado.

Conhecendo os caminhos diários dos usuários, é possível identificar padrões frequentes na mobilidade dos mesmos. A combinação desses padrões, também chamados de *motifs* [Schneider et al., 2013], é útil para analisar e compreender a movimentação durante a rotina da cidade. Para caracterizar o *motif* de cada trajetória diária, as coordenadas distintas do caminho receberam cada uma seu identificador único. Uma sequência

¹Disponível em ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010

Tabela 1. Divisão de classes (ou estratos) sociais. Os intervalos abertos à esquerda significam que os valores devem ser maiores do que o limite inferior à esquerda.

Classe social	Descrição	Intervalo de renda
1	Até 1 salário mínimo	[R\$0,00, R\$510,00]
2	De 1 a 2 salários mínimos	(R\$510,00, R\$1020,00]
3	De 2 a 3 salários mínimos	(R\$1020,00, R\$1530,00]
4	De 3 a 5 salários mínimos	(R\$1530,00, R\$2550,00]
5	De 5 a 10 salários mínimos	(R\$2550,00, R\$5100,00]
6	De 10 a 20 salários mínimos	(R\$5100,00, R\$10200,00]
7	Mais de 20 salários mínimos	(R\$10200,00, $+\infty$)

de identificadores foi construída de acordo com a ordem das coordenadas na trajetória. A construção do motif para uma dada trajetória pode ser vista na Figura 1. Um ponto importante a ser destacado é o fato de que trajetórias com ciclo entre um mesmo par de vértices são atribuídas ao mesmo motif, não importando quantas vezes aquele ciclo tenha se repetido. Logo, um motif cuja sequência fosse, por exemplo, 0,1,0,1,0,1,2 teria a mesma representação visual do passo 3 da Figura 1.

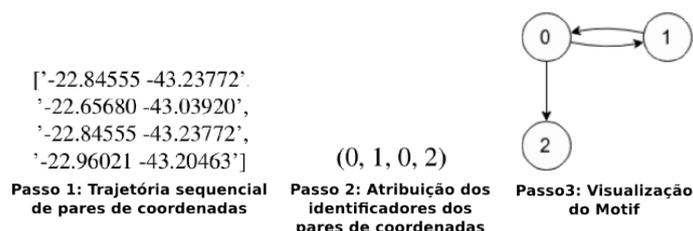


Figura 1. Construção de um motif a partir de pares (latitude, longitude).

4. Resultados e Discussão

A metodologia proposta neste trabalho, descrita na Seção 3, foi executada no contexto urbano do estado do Rio de Janeiro e, para isso, foi utilizada uma base de dados de registros de atividade em uma rede de telefonia móvel de uma grande operadora naquele estado. Cada registro representa uma atividade de conexão entre o dispositivo e a torre. Foram utilizados os seguintes atributos da conexão: data e hora do início da atividade na rede, hora do início da comunicação, IDs dos dispositivos de origem e destino das conexões, latitude e longitude da torre em coordenadas decimais. Essa base conta com registros do período de 01 a 19 de abril de 2012, totalizando 19 dias e abrangendo todo o estado. Algumas características da base de dados são descritas na Tabela 2.

O número total de registros da base ultrapassa os 200 milhões, com aproximadamente um milhão e trezentos mil identificadores distintos, cada identificador representando um indivíduo único na base. O número médio de ligações por ID apresenta um desvio padrão alto, indicando alta variabilidade dos dados. A moda auxilia a confirmar

Tabela 2. Algumas estatísticas da base de dados de chamadas de celular.

Total de Registros	203.497.693
Total de IDs	1.359.532
Período	01 a 19 de abril de 2012
Média do número médio de ligações por ID	149,68
Mediana do número médio de ligações por ID	44
Moda do número médio de ligações por ID	1 (80.856 vezes)
Variância do número médio de ligações por ID	198.711,44
Desvio padrão do número médio de ligações por ID	445,77

esta afirmação, uma vez que o número de ligações por ID mais frequente é de apenas uma ligação, comportamento seguido por mais de 80.000 IDs, o que equivale a aproximadamente 6% do total. Assim, a base possui muitos IDs com poucas ligações diárias e poucos IDs com um número elevado de ligações. Os registros armazenados são referentes a um período de 19 dias, de 01 de abril a 19 de abril de 2012 e é importante observar que há um feriado no dia 06 de abril de 2012 (sexta-feira da paixão, um feriado cristão).

Algumas características da utilização da rede de telefonia móvel pelas diferentes classes sociais puderam ser observadas após a associação dos dados CDR com a base de dados censitários, o que está ilustrado na Tabela 3. Nela, é possível perceber que a classe social com maior número de IDs na base é a classe 2, representando sozinha quase 33% do total de indivíduos. Essa nível de representatividade é a mesma apresentada pelo IBGE no censo 2010 [IBGE, 2011] para a classe 2, validando o modelo de classes sociais adotado e o método de atribuição das residências aos indivíduos. O método para presumir as residências é validado, pois consiste em atribuir primeiro o código do setor censitário ao ID de acordo com sua localização na malha e, em seguida, adicionar a informação de rendimento referente ao setor censitário determinado. Logo, após a execução do método para presumir as residências, caso a mesma não representasse a localidade real ou aproximada do indivíduo, o ID estaria localizado incorretamente sobre a malha. A atribuição do código de setor censitário seria também incorreta e a amostra não refletiria os resultados o IBGE para a fração de indivíduos em cada classe social.

Tabela 3. Distribuição e comparação de algumas métricas por classes sociais.

	Classes sociais						
	1	2	3	4	5	6	7
Número de IDs	19.261	238.589	175.129	139.237	105.643	45.542	7.532
Porcentagem de IDs	2,60%	32,50%	23,91%	19,00%	14,42%	6,25%	1,02%
Média de ligações por dia	17	15	14	14	12	11	12
Distância média percorrida (km)	249	224	203	188	141	139	144

A partir dos resultados apresentados na Tabela 3, pode-se notar que a média de ligações por dia decresce suavemente com o aumento da renda, assim como a distância média percorrida pelos indivíduos durante os 19 dias da base de dados. Nas quatro primeiras classes, a distância é reduzida cerca de 20 km menor em relação a classe anterior. Nas três classes mais altas (5, 6 e 7), porém, os valores para a distância média ficam bem próximos uns dos outros e distantes dos demais. Uma possível explicação pode ser que para maiores faixas salariais, a condição de vida permite aos indivíduos se locomover menos, por possuir trabalho e lazer próximos ao seu local residencial, o que não acontece com indivíduos de rendas menores, que precisam se deslocar por mais longas distâncias para realizar suas atividades diárias.

A distância média percorrida por cada classe em cada um dos 19 dias estudados é apresentada na Figura 2 e uma investigação superficial desse resultado revela um mesmo padrão para todas as classes sociais. É possível observar que durante os dias 1, 6, 7, 8, 14 e 15 a distância média percorrida é significativamente menor do que os outros dias. Isto pode ser justificado pela análise do calendário de abril, onde verifica-se que tais dias correspondem aos finais de semana e feriados do mês. Logo, como esperado, a distância percorrida é maior nos dias úteis e diminui aos fins de semana e feriados.

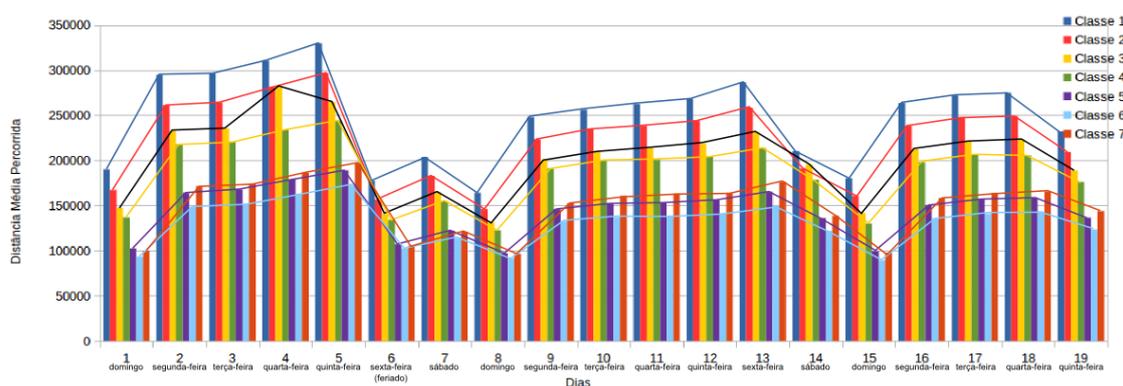


Figura 2. Distâncias médias por dia das classes 1 a 7.

É importante também observar que a movimentação do sábado após o feriado é maior que a movimentação do feriado, como era de se esperar, pois supõe-se que a atividade comercial e empresarial é em sua maioria restabelecida após o dia de folga. O sábado seguinte, por outro lado, tem movimentação bem maior que o sábado pós feriado, o que pode ser explicado pelo fato de que, após o feriado, nem toda a atividade comercial e empresarial foi restabelecida, ou seja, alguns indivíduos que rotineiramente se deslocam aos sábados permaneceram em casa ou reduziram a movimentação característica desse dia. Apesar de haver apenas um sábado padrão (sem a presença de feriados) para análise, é notável que, mesmo que faça parte do fim de semana, a movimentação nesse dia é mais próxima dos dias úteis do que do domingo, o que se explica pelo fato de que sábado é um dia de trabalho e atividades para muitos brasileiros, mas não todos, apresentando uma pequena diminuição com relação aos dias úteis porém não suficiente para se aproximar mais do domingo.

O padrão de movimentação durante as duas primeiras semanas iniciais se manteve em todas as classes. A primeira semana foi a que apresentou as maiores distâncias médias,

sendo o último dia útil o dia de maior distância percorrida tanto na primeira semana quanto na segunda, exceto para a classe 3, onde o penúltimo dia teve maior distância. Distâncias maiores percorridas no último dia útil da semana podem ser justificadas pela hipótese de que há uma necessidade da população de resolver seus problemas antes no início do final de semana. Curiosamente, o último dia da base, uma quinta-feira, tem uma distância reduzida se comparado as outras duas quintas-feiras. Como é o último dia disponível da base, é possível que os dados não estejam completos para este dia, o que faz com que nem toda a informação de distância seja adicionada à trajetória dos indivíduos, reduzindo a média do dia.

O número médio diário de ligações de cada indivíduo também foi investigado. A Figura 3 apresenta o gráfico da função de distribuição cumulativa complementar do número médio de ligações para a base de dados CDR completa, isto é, sem a separação por classes (Figura 3(a)) e o mesmo gráfico para a base de dados CDR dividida por classes (Figura 3(b)). Em todos os resultados apresentados na Figura 3, é possível observar um comportamento de cauda longa em uma parte significativa do intervalo. Esse resultado converge com o comportamento apresentado no trabalho de Onnela *et al.* [Onnela et al., 2007], onde uma rede social é construída a partir de dados CDR e observa-se que a probabilidade de indivíduos se comunicarem com um número x de contatos decresce com o aumento de x . Aqui, a probabilidade de um indivíduo realizar n ligações decresce com o aumento de n , indicando que a maioria dos indivíduos tende a realizar poucas ligações para poucas pessoas, e o número de indivíduos que realizam muitas ligações para mais pessoas é reduzido.

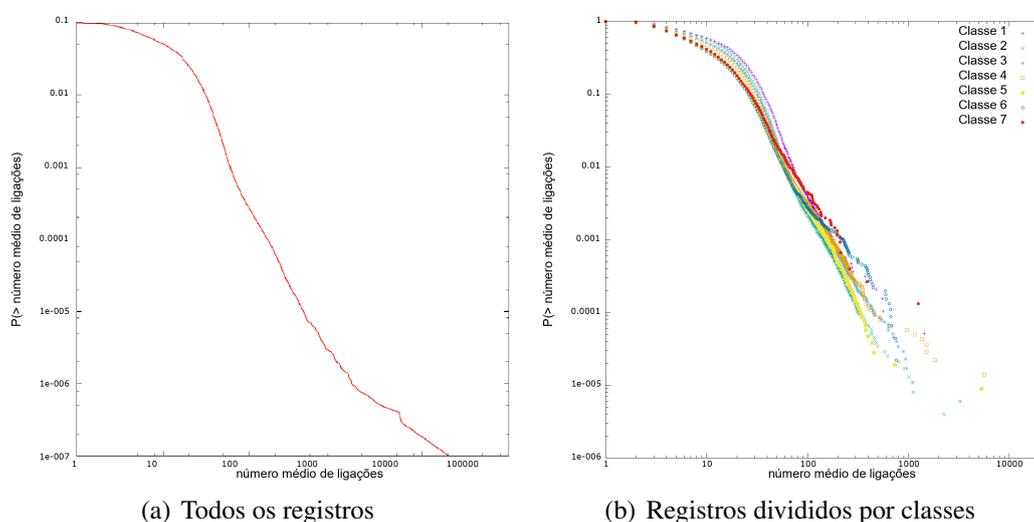


Figura 3. Distribuição cumulativa complementar de probabilidades do número médio de ligações: (a) todos os registros; (b) registros divididos por classes.

A análise do número médio diário de ligações efetuadas para os indivíduos das diferentes classes sociais (Figura 3(b)) revelou também uma tendência semelhante entre elas.

A partir da análise da Figura 3(b), pode-se observar que as classes apresentam curvas semelhantes e com o mesmo comportamento para um grande intervalo de valores. Um comportamento distinto para as classes pode ser observado nas caudas dos gráficos

ilustrados na Figura 3(b). Da classe 1 até a classe 4, temos um crescimento no número máximo de ligações. A partir da classe 5, esses valores começam a decrescer. Uma hipótese para tentar justificar tal fenômeno pode ser dada a partir da própria metodologia utilizada para definir informações de estratos sociais através dos setores censitários a partir de sua residência presumida, o que faz com que as ligações por classe também sejam divididas por região censitária. A Figura 4 mostra os indivíduos distribuídos pelo mapa de setores censitários de acordo com as residências presumidas. Nesse mapa é possível observar que os pontos identificados como parte das classes sociais intermediárias (1, 2, 3 e 4) estão concentrados nos grandes centros urbanos, como as cidades do Rio de Janeiro, Niterói e São Gonçalo.

Analisando a hipótese, a presença de estabelecimentos comerciais nesses locais pode justificar tanto o crescente valor do número de ligações nos extremos retratado na Figura 3 como os maiores valores de distância média percorrida para as classes 1, 2, 3 e 4 na Figura 2. Maiores valores para distância diária nessas classes podem indicar uma necessidade de os indivíduos de se locomoverem mais para a realização de atividades diárias. Adicionalmente, a presença de indivíduos de classes mais altas (5, 6 e 7) nas zonas litorâneas endossa a suposição de que tais indivíduos não necessitam se locomover tanto quanto os das classes de menor poder aquisitivo, possuindo opções de trabalho, lazer e outras atividades mais próximas de suas residências.

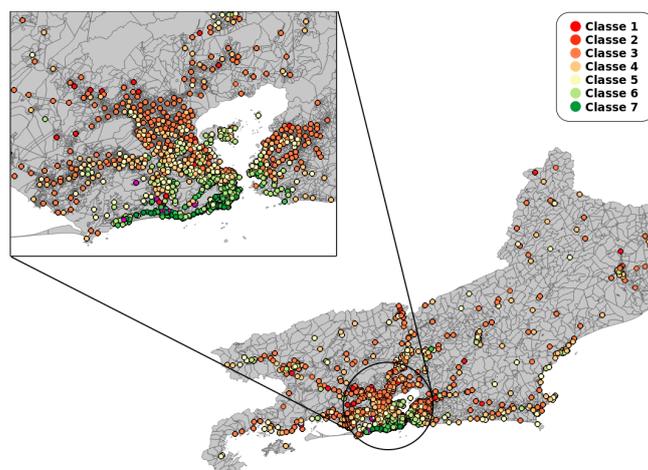


Figura 4. Mapa de setores censitários do Rio de Janeiro com indivíduos em intervalos de renda separados por cores.

Analisou-se também os padrões de movimentação dos indivíduos de toda população, representados pelos motifs. A Figura 5 apresenta a probabilidade $p(ID)$ de cada motif representado ser encontrado. Todo motif possui um nó central, que é a origem do deslocamento.

Foram encontrados 3392584 padrões distintos e, apesar da grande quantidade de padrões de trajetórias, com poucos deles, cerca de 30, é possível representar 50% de toda a movimentação durante os 19 dias. Este fato indica que há uma tendência da maioria dos indivíduos em seguir uma rotina que apresenta padrões parecidos.

Apesar de poucos padrões, esse número não é tão significativo quanto o encontrado em [Schneider et al., 2013], onde 17 padrões representaram 95% da movimentação.

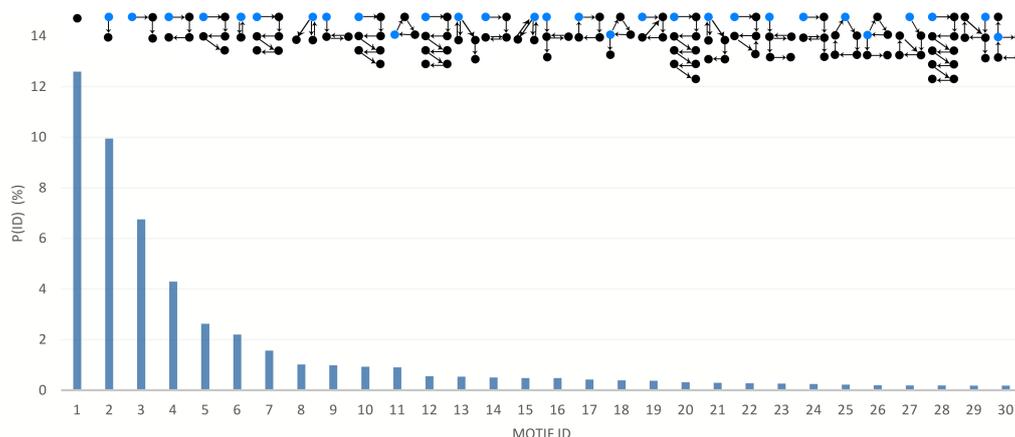


Figura 5. Porcentagem de padrões diários de deslocamento mais frequentes.

Isto pode ser justificado pelo curto período de tempo da base, de apenas 19 dias, que não permitiu caracterizar com precisão comportamentos relacionados a rotina da movimentação da população. Mesmo assim, com apenas 19 dias, é perceptível que a tendência é que a população siga padrões pequenos, sendo os mais frequentes apenas uma (0) ou duas (0,1) localizações.

Observa-se também que em alguns motifs encontrados os indivíduos não retornam ao local de origem, como era de se esperar. Um possível fator que explica esse fenômeno é a limitação dos dados, que por serem telefônicos, podem não refletir toda a movimentação dos usuários. Ainda sim, muitos dos padrões mais frequentes são coerentes com os encontrados em [Schneider et al., 2013].

5. Conclusões e Trabalhos Futuros

Neste estudo, os padrões de mobilidade e comunicação de diferentes classes sociais do estado do Rio de Janeiro foram analisados e comparados. Um ponto percebido no estudo foi a alta representatividade das classes sociais 2 e 3, em especial a classe social 2, que sozinha somou quase 33% do total de indivíduos. Descobriu-se que a distância média percorrida decresce conforme a classe social é mais elevada, o que pode ser um indício de uma menor necessidade de deslocamento por parte dos indivíduos de classes sociais mais altas. Um padrão de movimentação semanal também surgiu, caracterizando uma movimentação significativamente maior durante os dias úteis, possivelmente por conta de atividades profissionais.

A probabilidade de um indivíduo realizar um número médio diário x de ligações decresce com o aumento de x , revelando um comportamento de cauda longa para a curva de probabilidade. Isso acontece para todas as classes, e para as classes de 1 a 4 há um aumento no número máximo de ligações diárias, enquanto que nas classes mais altas esse número diminui. Isso pode ser justificado pela hipótese de que as 4 primeiras classes estão localizadas nos locais dos grandes centros comerciais, e os identificadores responsáveis por esse número máximo podem ser na verdade estabelecimentos comerciais.

A análise dos motifs revelou, apesar de grande variabilidade de padrões, que metade da movimentação da população pode ser representada por apenas 30 padrões de mobilidade, o que equivale a aproximadamente 0,0008% do total de padrões. A possível

justificativa para tal resultado está no fato de que os indivíduos geralmente seguem alguns padrões de movimentação de acordo com a sua rotina, frequentando na maioria das vezes poucos locais em uma ordem geralmente pré estabelecida.

O cálculo da localização residencial dos indivíduos teve sua precisão limitada devido a granularidade escolhida, de setores censitários, pois as localizações disponibilizadas pelos dados telefônicos se referem às antenas das conexões, podendo ser diferentes da localização real do usuário devido ao alcance das antenas. Esta porém teve que ser utilizada, uma vez que a informação de rendimento estava associada aos setores. O curto período de tempo dos dados coletados, de apenas 19 dias, também contribuiu para que algumas variáveis não apresentassem conclusões tão claras em seus resultados. É importante mencionar que as hipóteses aqui levantadas como possíveis justificativas para o comportamento da população não puderam ser testadas e validadas, não sendo este o escopo do estudo.

Portanto, a principal contribuição do presente estudo é a metodologia desenvolvida, que pode funcionar como um padrão para várias bases de dados CDR, bastando somente ajustá-la à granularidade desejada. Com uma base de dados mais volumosa e completa, poderão ser aplicadas, por exemplo, técnicas de triangulação para aproximação das residências. A utilização de uma base de dados contendo registros de conexões de internet e a aplicação de técnicas de aproximação das antenas e de cálculo não linear das distâncias entre elas são pretendidos como trabalhos futuros na área. Um estudo mais aprofundado dos motivos por classe social também se faz necessário a fim de comparar os padrões encontrados para a população em geral com os padrões de cada classe social.

6. Agradecimentos

Os autores agradecem ao CNPq, à FAPEMIG, à FAPERJ e à CAPES pelo apoio financeiro.

Referências

- Barlacchi, G., Nadai, M. D., Larcher, R., Casella, A., Chitic, C., Torrissi, G., Antonelli, F., Vespignani, A., Pentland, A., e Lepri, B. (2015). A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific Data* 2.
- Blondel, V. D., Decuyper, A., e Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10.
- Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J.-C., Huens, E., Van Dooren, P., Smoreda, Z., e Blondel, V. D. (2013). Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, 392(6):1459–1473.
- Duarte, F., Liberdi, R., e Sánchez, K. (2012). *Introdução à Mobilidade Urbana*. Juruá Editora. pg. 13.
- Gonzalez, M. C., Hidalgo, C. A., e Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- Hill, L. L. (2009). *Georeferencing: The geographic associations of information*. MIT Press. Capítulo 1.
- IBGE (2010). Censo demográfico 2010.

- IBGE (2011). Base de informações do censo demográfico 2010: resultados do universo por setor censitário.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., e Varshavsky, A. (2011). Ranges of human mobility in los angeles and new york. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 88–93. IEEE.
- Longley, P. (2005). *Geographic information systems and science*. John Wiley & Sons.
- Lotero, L., Cardillo, A., Hurtado, R., e Gomez-Gardenes, J. (2015). Several multiplexes in the same city: the role of wealth differences in urban mobility. *Interconnected Networks*.
- Marques-Neto, H. T., Xavier, F. H. Z., Xavier, W. Z., Malab, C. H. S., Ziviani, A., Silveira, L. M., e Almeida, J. M. (2018). Understanding human mobility and workload dynamics due to different large-scale events using mobile phone data. *Journal of Network and System Management*. Aceito para publicação.
- Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., De Menezes, M. A., Kaski, K., Barabási, A.-L., e Kertész, J. (2007). Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9(6):179.
- Quantum, G. (2013). Development team, 2012. quantum gis geographic information system. open source geospatial foundation project. *Free Software Foundation, India*.
- Saunders, P. (2006). *Social class and stratification*. Routledge. Capítulo 1.
- Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., e González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246.
- Shumaker, B. e Sinnott, R. (1984). Astronomical computing: 1. computing under the open sky. 2. virtues of the haversine. *Sky and telescope*, 68:158–159.
- Silveira, L. M., Almeida, J. M., Marques-Neto, H. T., Sarraute, C., e Ziviani, A. (2016). Mobhet: Predicting human mobility using heterogeneous data sources. *Computer Communications*, 95:54–68.
- Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., e González, M. C. (2015a). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58:162–177.
- Toole, J. L., Herrera-Yaque, C., Schneider, C. M., e González, M. C. (2015b). Coupling human mobility and social ties. *Journal of the Royal Society Interface*, 12.
- Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Ramirez-Paredes, J., Halsey, E. S., Kochel, T. J., Scott, T. W., e Kitron, U. (2013). Using gps technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PLOS ONE*, 8(4):1–10.
- Xavier, F. H. Z., Silveira, L. M., Almeida, J. M., Ziviani, A., Malab, C. H. S., e Marques-Neto, H. T. (2012). Analyzing the workload dynamics of a mobile phone network in large scale events. In *Proc. of the Workshop on Urban Networking (UrbaNe), ACM CoNEXT*.