UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Gustavo Henrique Martins da Costa

Caracterização da relação entre interações sociais, mobilidade e aspectos socioeconômicos de indivíduos em contextos urbanos

São João del-Rei

2022

UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Gustavo Henrique Martins da Costa

Caracterização da relação entre interações sociais, mobilidade e aspectos socioeconômicos de indivíduos em contextos urbanos

Dissertação apresentada como requisito para obtenção do título de mestre em Ciências no Curso de Mestrado do Programa de Pós Graduação em Ciência da Computação da UFSJ.

Orientador: Vinícius da Fonseca Vieira

Universidade Federal de São João del-Rei – UFSJ

Mestrado em Ciência da Computação

São João del-Rei

2022

Ficha catalográfica elaborada pela Divisão de Biblioteca (DIBIB)
e Núcleo de Tecnologia da Informação (NTINF) da UFSJ,
com os dados fornecidos pelo(a) autor(a)

C837c

Costa, Gustavo.
Caracterização da relação entre interações sociais, mobilidade e aspectos socioeconômicos de indivíduos em contextos urbanos / Gustavo Costa ; orientador Vinícius Vieira. -- São João del-Rei, 2022.
55 p.

Dissertação (Mestrado - Ciência da Computação) -- Universidade Federal de São João del-Rei, 2022.

1. Redes Complexas. 2. Mobilidade urbana. I. Vieira, Vinícius , orient. II. Título.

Gustavo Henrique Martins da Costa

Caracterização da relação entre interações sociais, mobilidade e aspectos socioeconômicos de indivíduos em contextos urbanos

Dissertação apresentada como requisito para obtenção do título de mestre em Ciências no Curso de Mestrado do Programa de Pós Graduação em Ciência da Computação da UFSJ.

Trabalho aprovado. São João del-Rei, 13 de maio de 2022:

Vinícius da Fonseca Vieira
Orientador

Márcio Roberto Toledo
Convidado 1

Leonardo Goliatt da Fonseca
Convidado 2

Carolina Ribeiro Xavier
Convidada 3

São João del-Rei
2022

Resumo

Palavras-chaves: CDR, redes complexas, mobilidade urbana.

A compreensão da organização urbana é fundamental para um melhor planejamento e definição de políticas públicas que visem maior bem estar da população e mitigação de problemas socioeconômicos. Modelos computacionais capazes de integrar dados de diferentes fontes podem auxiliar a caracterização da complexa organização geográfica e socioeconômica da população em áreas urbanas com grande resolução e aplicabilidade em contextos práticos. Este trabalho apresenta a análise da complexidade urbana representada pelas relações sociais e de mobilidade através da investigação de registros de ligações telefônicas armazenadas como Call Detail Records (CDR). Além disso, visando compreender como aspectos econômicos afetam tanto a mobilidade quanto a interação entre os indivíduos, utilizou-se dados do Censo Demográfico Brasileiro de 2010, onde os setores censitários das regiões estudadas foram classififcados conforme sua renda média. Considerando duas cidades com características distintas, experimentos realizados a partir da caracterização da mobilidade individual e econômica das pessoas permitem observar uma forte interdependência entre seu comportamento de mobilidade no espaço urbano e o comportamento exibido pelas pessoas que definem sua rede social.

Abstract

Key-words:CDR, complex networks, urban mobility.

The understanding of urban organization is essential for better planning and definition of public policies that aim at greater well-being of the population and mitigation of socioeconomic problems. Computational models capable of integrating data from different sources can help to characterize the complex geographic and socioeconomic organization of the population in urban areas with high resolution and applicability in practical contexts. This work presents the analysis of urban complexity represented by social and mobility relations through the investigation of telephone call records stored as Call Detail Records (CDR). Furthermore, in order to understand how economic aspects affect both mobility and as for the interaction between individuals, data from the 2010 Brazilian Demographic Census were used, where the census sectors of the regions studied were classified according to your average income. Considering two cities with different characteristics, experiments carried out based on the characterization of people's individual and economic mobility allow us to observe a strong interdependence between their mobility behavior in the urban space and the behavior exhibited by the people who define their social network.

Lista de ilustrações

Figura 1 – Representação do problema de *Koninsberg* através de um grafo 16

Figura 2 – Amostragem da base CDR de São João del-Rei 22
Figura 3 – Algumas características sociais da rede. (a) Exemplo de uma amostra da rede usando *snowball sampling* com seis saltos (o vértice em preto é o ego). (b) Distribuição da reciprocidade na rede.). 24
Figura 4 – Distribuição cumulativa complementar de probabilidades do número médio de ligações: (a) todos os registros; (b) registros divididos por classes (retirado de (RODRIGUES, 2018)). 26

Figura 5 – Densidade demográfica Juiz de Fora 31
Figura 6 – Densidade demográfica São João del-Rei 32
Figura 7 – Região imediata de Juiz de Fora 33
Figura 8 – Região imediata de São João del-Rei 34
Figura 9 – Distribuição de graus por classe na rede não direcionada de Juiz de Fora 37
Figura 10 – Distribuição do número de ligações por classe em Juiz de Fora 37
Figura 11 – Distribuição de grau de entrada na rede direcionada de Juiz de Fora . . 37
Figura 12 – Distribuição de grau de saída na rede direcionada de Juiz de Fora . . . 37
Figura 13 – Distribuição de graus por classe na rede não direcionada de São João del-Rei 38
Figura 14 – Distribuição do número de ligações por classe em São João del-Rei . . . 38
Figura 15 – Distribuição de grau de entrada na rede direcionada de São João del-Rei 38
Figura 16 – Distribuição de grau de saída na rede direcionada de São João del-Rei . 38
Figura 17 – Distribuição da porcentagem de presença da classe predominante em cada comunidade por classe de renda - SJDR 40
Figura 18 – Distribuição da porcentagem de presença da classe predominante em cada comunidade por classe de renda - JF 40
Figura 19 – Raio de giração por classe de renda - Juiz de Fora 42
Figura 20 – Raio de giração por classe de renda - São João del-Rei 42
Figura 21 – Distância máxima percorrida por classe de renda - Juiz de Fora 43
Figura 22 – Distância máxima percorrida por classe de renda - São João del-Rei . . 43
Figura 23 – Distância máxima percorrida de casa a partir da residância por classe de renda - Juiz de Fora 43
Figura 24 – Distância máxima percorrida de casa por classe de renda - São João del-Rei 43
Figura 25 – Número de locais visitados por classe de renda - Juiz de Fora 44
Figura 26 – Número de locais visitados por classe de renda - São João del-Rei . . . 44

Figura 27 – Entropia por classe de renda - Juiz de Fora 45
Figura 28 – Entropia por classe de renda - São João del-Rei 45
Figura 29 – Similaridade de trajetórias entre as classes de renda na região imediata de Juiz de Fora 46
Figura 30 – Similaridade de trajetórias entre as classes de renda na região imediata de São João del-Rei 46
Figura 31 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 1 46
Figura 32 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 2 46
Figura 33 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 3 47
Figura 34 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 4 47
Figura 35 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 5 47
Figura 36 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 1 48
Figura 37 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 2 48
Figura 38 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 4 48
Figura 39 – Correlação de trajetórias entre as classes de renda - Juiz de Fora 49
Figura 40 – Correlação de trajetórias entre as classes de renda - São João del-Rei . 49

Lista de tabelas

Tabela 1 – Divisão de classes de renda. Os intervalos abertos à esquerda significam que os valores devem ser maiores do que o limite inferior à esquerda . . 25

Tabela 2 – População por cidade da região imediata de São João del-Rei 29
Tabela 3 – População por cidade da região imediata de Juiz de Fora 30
Tabela 4 – Características das bases cdr de cada região imediata 32
Tabela 5 – Divisão de indivíduos por classe de renda - Juiz de Fora 35
Tabela 6 – Divisão de indivíduos por classe de renda - São João del-Rei 35
Tabela 7 – Características básicas das redes não direcionas 36
Tabela 8 – Características básicas das redes não direcionas 36
Tabela 9 – Características básicas das redes não direcionas 39

Sumário

1 Introdução 10

2 Conceitos e trabalhos relacionados 14

2.1 *Call Detail Records* 14

2.2 Redes sociais e mobilidade 15

2.2.1 Grafos: definição matemática 15

2.2.2 Redes de comunicação 17

2.2.3 Redes urbanas 18

2.3 Modelos de mobilidade humana em espaços urbanos 19

2.4 Mobilidade e aspectos socioeconômicos 20

3 Metodologia 22

3.1 Pré-Processamento e Caracterização das relações sociais com base nos dados CDR 22

3.2 Presunção de residência e caracterização do estrato social 24

3.3 Caracterização básica da mobilidade dos indivíduos 25

3.4 Caracterização da relação entre aspectos socioeconômicos, rede social e padrões de mobilidade 27

4 Resultados e Discussão 29

4.1 Caracterização das regiões geográficas e do conjunto de dados 29

4.2 Presunção das residências dos indivíduos 35

4.3 Características básicas das redes sociais 36

4.4 Relação entre aspectos socioeconômicos e utilização do espaço urbano . . . 41

4.5 Relação dos aspectos socioeconômicos com as interações entre indivíduos . 45

5 Conclusões e trabalhos futuros 50

Referências 51

1 Introdução

Os dados gerados pelas novas tecnologias têm sido objeto de estudo de diversos trabalhos que têm por objetivo entender a organização dos espaços urbanos. A compreensão das dinâmicas das cidades é de fundamental importância para um melhor planejamento urbano e o estabelecimento de políticas públicas que possibilitem a antecipação e mitigação de problemas sociais na escala em que ocorrem nos centros urbanos dos mais variados portes. Essa grande massa de dados tem ampliado a utilização do coneceito de *Smart Cities*, que busca desenvolver cidades criativas e sustentáveis, que fazem uso da tecnologia em seu processo de planejamento com a participação dos cidadãos, mesmo que em alguns casos de maneira involuntária.

O Brasil, assim como vários outros países em desenvolvimento, experimentou um expressivo aumento na taxa de urbanização durante a segunda metade do século XX. Para se ter uma noção, a população urbana do país no ano de 1940 representava 31,24% do total. Em 2010, 84,36% dos brasileiros residiam em cidades, segundo dados do Instituto Brasileiro de Geografia e Estatística, o IBGE. (FARBER S., 2013) e (SILM S. & AHAS, 2014) apontam que as condições de infraestrutura dos grandes centros urbanos crescem de maneira sublinear à expansão da população e da área das cidades. A combinação dessas informações explica um pouco da dinâmica dos centros urbanos brasileiros, onde as melhorias em infraestrutura não acompanharam o ritmo de crescimento urbano-populacional, deixando as populações menos favorecidas economicamente distantes de uma série de bens e serviços ofertados e garantidos por lei. Dado esse cenário de segregação espacial, se faz necessária uma maior compreensão das dinâmicas de mobilidade da população das cidades. Soluções que envolvam Big Data oferecem a possibilidade de incrementar o leque de informações sobre os indivíduos e também grupos de indivíduos.

Modelos computacionais capazes de integrar e processar dados de múltiplas origens podem auxiliar a caracterização da organização geográfica e socioeconômica da população com grande resolução e aplicabilidade em contextos práticos (BLONDEL, 2015; GONZALEZ, 2008). Inovações na área de armazenamento e processamento de grandes volumes de dados propiciaram a coleta de dados em tempo real, o que vem servindo de insumo para soluções inovadoras para cidades baseadas em dados, o que envolve algoritmos e metodologias que são, ainda, foco de pesquisa em muitas áreas (ALESSANDRETTI, 2017; LENORMAND; RAMASCO, 2016).

A modelagem do sistema urbano considerada neste trabalho toma como base de dados os registros de ligações telefônicas (*Call Detail Records* - CDR), armazenados por companhias de telefonia móvel com o objetivo de identificar as antenas nas quais as

chamadas telefônicas são realizadas e, com isso, efetuar cobranças adequadas.

Diversos estudos caracterizam a mobilidade dos indivíduos nas cidades buscando entender como esses utilizam o espaço urbano. (GONZALEZ, 2008) foi um dos pioneiros em estudos com o objetivo de compreender a cidade a partir de dados obtidos pela observação das ações dos indivíduos. Gonzalez *et. al* demonstram em seu trabalho que as trajetórias das pessoas nas cidades tem um alto grau de regularidade espaço-temporal. Outros trabalhos de grande contribuição na área são (SCHNEIDER, 2013), que explora a ideia de utilização de *motifs* para a caracterização das trajetórias dos indivíduos e (BOEING G., 2021) que traz um debate histórico sobre a mobilidade urbana nos grandes centros.

Como dito anteriormente, o rápido processo de urbanização vivido no Brasil durante a segunda metade do século passado, trouxe enormes desigualdades econômicas e problemas sociais que foram acentuados ao longo dos anos devido a falta de um planejamento adequado de agentes tomadores de decisões públicas. A compreensão dos processos que produzem segregação nas cidades e o estudo de seus efeitos são fundamentais para um melhor planejamento e a proposta de políticas públicas para mitigar problemas sociais severos no país. A distribuição espaço-temporal dos indivíduos em um território não é uniforme e à medida em que as cidades crescem, elas se tornam mais diversas e muitas vezes menos integradas, o que é resultado da combinação de muitos fatores, tais como os locais de residência e trabalho, a infraestrutura do sistema de transporte, a mobilidade diária e o planejamento urbano (REARDON S. F. & O'SULLIVAN, 2018). Na última década, essa conjunção de fatores tornou as cidades desiguais e segregadas. Embora a intensidade das desigualdades possa variar de um lugar para o outro, a existência de vizinhanças ricas e pobres na mesma cidade pode ser observada independentemente da idade da cidade e do período histórico considerado. A segregação e a distinção na forma como as pessoas utilizam o espaço urbano, embora fácil de observar, é difícil de quantificar, já que o sistema urbano é muito complexo e aspectos globais não podem ser negligenciados.

Espera-se nesse trabalho, representar a complexidade urbana através de uma perspectiva econômica. Para isso, será analisado qual o impacto da renda dos indivíduos em suas trajetórias no espaço urbano e também como fatores econômicos influenciam nas interações entre os indivíduos. Portanto, pretende-se compreender como os indivíduos de diferentes classes de renda interagem com o espaço urbano, o que pode auxiliar na caracterização da organização das cidades e dos indivíduos. E além disso, as informações compiladas ao longo da construção do trabalho podem servir de subsídio para ações de agentes públicos tomadores de decisões, assim como fomentar trabalhos de pesquisadores das mais diversas que a natureza multidisciplinar do trabalho se insere.

Visando entender como a condição econômica de um indivíduo influencia sua mobilidade, analisaremos dados econômicos das cidades que compõem o grupo das regiões

imediatas estudadas, tendo como insumo os resultados do Censo Demográfico Brasileiro de 2010, o último realizado no país. De acordo com o IBGE, regiões geográficas imediatas agrupam municípios que possuem relações de dependência e deslocamento da população em busca de bens, prestação de serviços e trabalho e estão ligados a um município base. Dada a grande quantidade de estudos direcionados para municípios de grande porte e a ausência de trabalhos desse tipo tomando base cidades do interior mineiro, escolheu-se as regiões imediatas de Juiz de Fora e São João Del Rei como objeto de estudo.

Muitos trabalhos na literatura são dedicados a estudar a segregação de grupos sociais e as diferentes formas que utilizam o espaço urbano (MASSEY D. S. & DENTON, 1988; REARDON S. F., 2004; WONG D. W. S. & SHAW, 2011). Alguns autores, como Wong e Shaw (WONG D. W. S. & SHAW, 2011) e Farber *et al.* (FARBER, 2015) investigam a segregação de grupos étnicos através da análise de locais visitados por indivíduos em um período de tempo. Mais especificamente, a perspectiva de redes complexas para dados sociais e de mobilidade a partir de dados CDR pode nos ajudar a trazer conclusões valiosas sobre a forma como diferentes grupos utilizam o espaço urbano, combinando conceitos e técnicas frequentemente utilizados na geografia, ciência da computação e ciência de redes (BOEING G., 2021; BARBOSA, 2018). Nesse sentido, é importante apontar o trabalho de Lotero *et al.* (LOTERO-VéLEZ, 2016), que combinou dados censitários com informações de origem e destino de viagens de indivíduos em duas cidades na Colômbia para caracterizar a forma como diferentes estratos sociais utilizam os diferentes modais de transporte, considerando um modelo baseado em redes complexas.

O presente trabalho, assim como o trabalho de Lotero *et al.* (LOTERO-VéLEZ, 2016), também visa caracterizar as formas como diferentes estratos sociais utilizam o espaço urbano, porém, algumas diferenças essenciais devem ser destacadas. Uma delas está no objeto a ser investigado. Enquanto o trabalho de Lotero *et al.* busca caractrerizar a forma como os diferentes estratos sociais utilizam os diferentes modais de transporte, este trabalho busca caracterizar o impacto da classificação social baseada em renda nas relações sociais estabelecidas pelos indivíduos nas cidades, além dos diferentes padrões de mobilidade observado para os indivíduos. Outra diferença está nos materiais tomados como base para os trabalhos. Enquanto o trabalho de Lotero *et al.* utiliza informações de origem e destino a partir de questionários ativamente aplicados para a identificação das viagens dos indivíduos, este trabalho baseia-se em uma metodologia para a obtenção das informações sobre mobilidade a partir de dados de telefonia celular, passivamente fornecidos pelos indivíduos.

O uso de questionários para a obtenção de informações sobre as viagens dos indivíduos traz algumas vantagens, como o fato de que a comunicação deliberada das pessoas tende a introduzir menos erros e ruídos, podendo gerar resultados mais sólidos. Por outro lado, há uma limitação no número de indivíduos para os quais é viável se aplicar um

questionário, o que traz uma limitação no tamanho da base de dados a ser considerada e dificulta a generalização de conclusões potencialmente obtidas. Em contrapartida, o uso de dados massivos, mesmo que fornecidos de maneira não deliberada pelos usuários, pode trazer informações mais generalizáveis quando aplicados métodos para eliminação de ruídos, preservando princípios éticos de anonimicidade dos indivíduos, quando dados sobre suas identidades são omitidos.

Com os dados econômicos do censo de 2010, os setores censitários - menor divisão residencial dentro de uma cidade - dos municípios das regiões imediatas serão classificados de acordo com a renda média de cada um deles. Dessa forma, combinando com a metodologia para detecção de residência de (VANHOOF, 2018), poderemos analisar os diferentes padrões de mobilidade e interação dos individuos por classe econômica.

Considerando a análise de duas regiões imediatas, uma de pequeno e outra de médio porte, sob a perspectiva das relações socioeconômicas e de mobilidade, duas hipóteses, tratadas como Questões de Pesquisa (QP), são investigadas neste trabalho. Primeiramente, serão investigadas as relações sociais entre os indivíduos e como a condição econômica desses influencia em seus contatos dentro de uma rede social (QPI). Logo após, as trajetórias e distâncias percorridas pelos indivíduos, medidas individualmente e coletivamente, serão analisadas com o objetivo de identificar padrões de mobilidade nas diferentes classes de renda encontradas (QP2). Também será discutido se a classificação econômica dos indivíduos impacta no uso do espaço urbano. Pretende-se com isso compreender como aspectos econômicos podem influenciar na forma como os indivíduos utilizam a cidade.

O presente trabalho está organizado da seguinte forma: O capítulo 2 traz as referências que subsidiaram a construção deste; o capítulo 3 apresenta a metodologia aplicada e o quarto capítulo demonstra os resultados dos experimentos realizados.

2 Conceitos e trabalhos relacionados

Nas seções a seguir serão apresentados uma série de trabalhos e conceitos que fundamentaram a criação deste.

2.1 *Call Detail Records*

Conforme aponta (CHAVES, 2018), os dados de CDR são gerados em grande quantidade como subproduto de bilhetagem das operadoras e contêm, dentre outras informações, a localização aproximada da realização da chamada telefônica, assim como a data e a hora dessa atividade.

O primeiro estudo que utiliza como insumo para análise dados de telefonia é o de (HOLLINGSHEAD A. B.,). Realizado em 1949 e fazendo uso de registros de telegráfos, Hollingshead *et. al*, descobrem duas leis de potência entre:

- *a)* quantidade de mensagens trocadas versus distância entre um telégrafo e outro;
- *b)* soma da quantidade de dinheiro gasto em passagens versus distância entre a origem e o destino.

Desde então, com o avanço dos modos de comunicação e o surgimento dos celulares, o número de trabalhos que exploram os registros telefônicos cresceu consideravelmente. A proporção de pessoas ao redor do mundo que utilizam esse aparelho saltou de 12% em 2000 para 96% em 2014, é o que aponta (ITU, 2015). Esse significativo incremento permitiu que a comunidade científica modelasse das mais diversas formas a mobilidade humana, sobretudo nos centros urbanos.

Dentre os modelos construídos para representar a mobilidade urbana, destaca-se o apresentado por (YANG Y., 2007), que além de extrair a mobilidade dos indivíduos, também foi capaz de inferir os motivos mais frequentes das movimentações, estabelecendo parâmetros sobre as cirscunstâncias das ligações, como horário, duração e local. Isso é de fundamental importância pois dado que se sabe a motivação do deslocamento, passa a se conhecer um aspecto comportamental das pessoas e torna-se possível uma generalização do padrão de deslocamento que posssa subsidiar políticas públicas voltadas para a mobilidade urbana.

Entretanto, apesar da grande disponibilidade de dados, existem algumas restrições a esses *datasets* que devem ser ressaltadas. (BLONDEL, 2015) apresenta primeiramente que o conteúdo dos registros telefônicos é inacessível, devido a utilização de SMS ou pelo

fato que as conversas telefônicas não são gravadas. Em seguida, Blondel *et. al* esclarece que o fato das operadoras telefônicas serem as proprietárias desses conjuntos de dados, acarreta na limitação de acesso aos mesmos, variando conforme as políticas de privacidade das empresas e dos países em que operam.

Um dos principais problemas na utilização de dados *CDR* é a precisão dos locais de chamada. (BARBOSA, 2018) destaca que há uma variabilidade significativa nas áreas cobertas pelas torres de celular – com cobertura variando de dezenas de metros nos bairros mais densos das áreas urbanas, até alguns quilômetros nas áreas rurais – um usuário registrado em uma área rural pode transmitir toda a comunicação por meio de uma única torre durante sua rotina diária, enquanto move exatamente a mesma área de um usuário urbano que pode se movimentar por áreas cobertas por várias torres. Assim, no primeiro caso, apesar do movimento possivelmente significativo por parte do usuário, para fins de análise eles são considerados estáticos, o que é obviamente enganoso.

O trabalho de (GONZALEZ, 2008) foi um dos pioneiros no que tange o estudo da mobilidade através de dados *CDR*. Em seu estudo, Gonzalez *et. al*, tomam como referência a trajetória de 100.000 indivíduos e detecta que tais trajetórias tem um alto grau de regularidade temporal e espacial, isto é, em geral os indivíduos percorrem distâncias e visitam locais padrões em seus percursos diários.

2.2 Redes sociais e mobilidade

2.2.1 Grafos: definição matemática

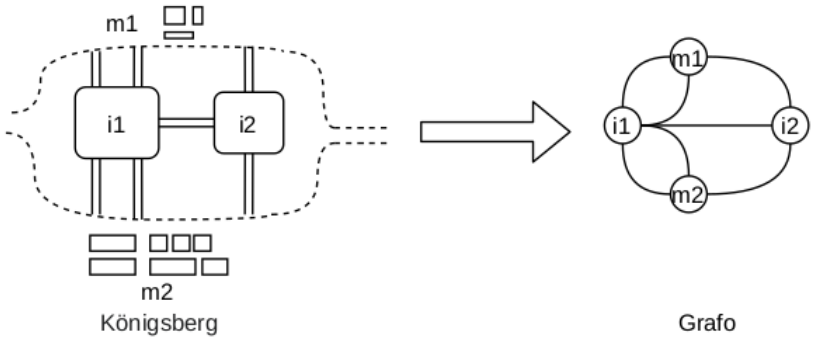
Para um melhor entendimento dos experimentos aqui realizados, é necessário que alguns conceitos chaves sejam explicados. O primeiro deles, é o conceito de grafo. Um grafo *G(V,A)* é definido pelo par de conjuntos V e A:

- V - conjunto não vazio: os vértices ou nodos do grafo;
- A - conjunto de pares ordenados α =(v,w), v e w ϵ V: as arestas do grafo;

A teoria dos grafos surgiu do problema das sete pontes de *königsberg*, em que os moradores dessa cidade, localizada na época na Prússia Oriental, hoje *Kalingrad* na Rússia, buscavam uma rota turística que permitisse passar pelas sete pontes da cidade que ligavam as suas duas ilhas e retornasse ao local de origem passando uma única vez em cada ponte. Apesar de aparentar ser simples, este problema desafiava os intelectuais da cidade, sendo resolvido somente pelo grande matemático Leonard Euler. Para solucionar esse desafio, Euler modelou o problema com uma estrutura elegante e gráfica de linhas e pontos, que resultaria no que hoje é conhecido como grafo. Euler demonstrou que o

problema proposto não tinha uma solução dado que as pontes que ligavam as porções de terra sempre provinham um número ímpar de passagens, portanto em algum momento não existiria um caminho de retorno (BARBOSA, 2022).

Figura 1 – Representação do problema de *Koninsberg* através de um grafo



Os grafos podem ser divididos em dois subgrupos: Os direcionados e não direcionados. Os direcionados são aqueles em que as arestas tem um ponto de partida e um ponto de saída, diferentemente dos não direcionados. É importante destacar que existe uma equivalência entre esses dois tipos de grafos.

Um grafo também pode ser ponderado, ou seja, possuir um peso ou custo nas suas arestas indicando um custo de deslocamento em determinado caminho.

Um conceito intrínseco a teoria dos grafos e de fundamental importância nesse trabalho é o grau. O grau é uma propriedade inerente a cada vértice do grafo, se referindo ao número de conexões de um vértice. Em outras palavras, esta propriedade representa o número de arestas que incidem ou saem de determinado vértice. Em um grafo direcionado, ocorre distinção entre o grau de saída, o número de arestas que saem deste vértice, e o grau de entrada, o número de arestas que incidem sobre este vértice. O grau pode ser utilizado diretamente para a definição da importância de um vértice no grafo. Por exemplo, de maneira simples, pode-se dizer que um vértice com grau 5 pode ser mais importante que um vértice de grau 2.

As redes complexas são grafos que apresentam propriedades topográficas bastante particulares, não encontradas em grafos mais simples. A Figura 3 é uma representação de uma rede complexa não direcionada.

Dentro do estudo de redes complexas, um tema sempre muito abordado é sobre comunidade. Comunidade pode ser definida como agrupamento de nós aos quais podemos chamar de subgrafos, em que os nós de uma mesma comunidade irão possuir mais ligações com os nós que estão presentes no mesmo grupo, do que com os nós que estão presentes em outras comunidades.

Nas análises feitas sobre as comunidades, uma métrica conhecida como Modularidade é utilizada para partição do grafo em comunidades, em que se permite identificar grupos com alta coesão quanto as ligações internas e baixa externa. Essa métrica permite identificar a qualidade das comunidades, onde uma comunidade com alta modularidade, represente agrupamento de nós em que se possui um grande número de ligações internas (com nós que estão no mesmo grupo) e um baixo número de ligações externas (nós que estão em outro grupo)(BARBOSA, 2022).

Outro conceito importante é o de matriz de adjcência, que é uma matriz $n \times n$, denotada por

$$X = [X_{ij}] \tag{2.1}$$

As células dessas matrizes são preenchidas com 1 caso haja arestas entre dois vértices e 0 caso contrário.

A representação de grafos através de matrizes está associada à necessidade de realização de cálculos envolvendo dados estruturais. A matriz de adjacência, quando construída de forma ingênua, não é muito econômica do ponto de vista computacional e ocupa n^2 posições de memória. Entretanto pode-se utilizar representações que tirem vantagem da esparsidade do grafo, ou seja, que não represente as posições correspondentes a vértices não adjacentes (VIEIRA, 2013).

2.2.2 Redes de comunicação

Redes de comunicação buscam modelar através de grafos o fluxo de troca de normas, valores e ideias de vários recursos sociais e culturais das sociedades, conforme explica Onnela *et al.* (ONNELA, 2007). Nas redes de comunicação desenvolvidas a partir de dados de telefonia móvel, cada vértice representa um indivíduo, e há aresta entre dois indivíduos se houve ligações telefônicas entre os mesmos.

Na construção de redes de comunicação, Onnela *et. al* utilizaram da abordagem de grafos ponderados, onde o peso das arestas era quantificado através de cálculos que envolviam o tempo total e a quantidade de ligações entre dois indivíduos.

Em sua análise sobre os vértices das redes construídas, Onnela *et. al* constatou que a maioria dos indivíduos mantêm contato com poucos indivíduos, enquanto que uma pequena minoria se comunica com mais de uma dezena, reforçando a ideia de que a maioria das pessoas consegue manter contato regularmente com poucos.

Blondel *et al.* (BLONDEL, 2008) têm grande contribuição na construção de redes complexas baseados em dados *CDR*, com um discussão extremamente profunda e interessante sobre a relevância de estudos que tem como insumo redes de telefonia móvel.

A proposta central de Blondel *et. al* é extrair comunidades a partir de redes consideravelmente grandes, utilizando o algoritmo de *Louvain*. Esta abordagem tem como

objetivo a maximização da modularidade de cada comunidade. O algoritmo pode ser explicado em duas etapas: na primeira parte cada nó é atribuído em uma comunidade, em que o número total de comunidades, será igual ao número de nós. Após este ponto os nós são deslocados entre as comunidades, buscando maximizar a modularização. Este processo ocorre de maneira repetitiva, parando ao se encontrar o máximo local de modularidade. Somente são considerados válidos, os deslocamentos que resultam em aumento de modularidade. A segunda etapa consiste na construção de uma nova rede considerando as comunidades encontradas.

2.2.3 Redes urbanas

Há uma série de trabalhos que buscam modelar as relações humanas e as interações destes com o espaço urbano. (GONZALEZ, 2008) é referência de destaque na maior parte dos trabalhos que tentam enxergar padrões nas trajetórias humanas nos centros urbanos. Esse trabalho apresenta a ideia de que, apesar das incontáveis diferenças entre os indivíduos e suas histórias, o ser humano segue padrões de mobilidade altamente reproduzíveis e simples, sendo possível caracterizar estatisticamente as trajetórias humanas aplicando-se técnicas apropriadas na manipulação dos dados *CDR*.

Essa semelhança entre os padrões de viagens podem impactar todos os fenômenos impulsionados pela mobilidade humana, desde a prevenção de epidemias até a resposta a emergências, planejamento urbano e modelagem baseada em agentes.

(BARBOSA, 2018) faz uma revisão dos princípios fundamentais de modelagem da mobilidade humana e traz uma coleção de métodos técnicos aplicáveis a problemas específicos relacionados à mobilidade.

Nesse trabalho, Barbosa *et. al* exploram diversos estudos e metodologias aplicados na análise da mobilidade urbana utilizando diversas fontes de dados, tanto de mobilidade (dados de GPS, ligações telefônicas, interações online e pequisa sobre viagens locais) quanto de aspectos socioeconômicos (dados de censo demográfico e de órgãos fiscais, como a Receita Federal dos Estados Unidos).

Quando a fonte de dados é obtida sem interação direta com os indivíduos, como no trabalho de Gonzalez *et. al*, denomina-se essa forma de coleta como passiva. Utilizando dessa abordagem e tomando como insumo dados oriundos de trajetórias traçadas através de *GPS*, (SHIN S. HONG, 2008) registra com uma acurácia inferior a 3 metros a movimentação de 44 indivíduos em dois campi universitários, uma feira estadual, um parque temático e uma região metropolitana nos EUA.

Rhee *et. al* demonstram que muitas características estatísticas das caminhadas humanas seguem uma lei de potência, independente do cenário onde essas caminhadas venham a ocorrer.

A forma de coleta de dados passiva privilegia a escala em detrimento da resolução, facilitando a generalização. Porém pouco se sabe sobre os indivíduos estudados.

Outro trabalho de grande importância na modelageem de redes urbanas é o de (LOTERO-VéLEZ, 2016) que através de questionários busca caracterizar a mobilidade urbana tomando como referência os modais de transporte nas cidades de Medellin e Bogotá, na Colômbia.

Em seu trabalho, Lotero *et. al* exploram a caracterização do espaço urbano através das redes de transporte, criando uma rede para cada modal de transporte nas cidades referidas acima. Com as informações econômicas obtidas nos questionários, eles caracterizam padrões de mobilidade para cada subconjunto da população em cada rede de transporte.

Assim como Lotero *et. al*, (LIANG J. ZHAO, 2013) faz uso de métodos ativos para a coleta de dados. Através de questionários detalhados onde era averiguado o motivo da viagem, o meio de tranporte utilizado e o horário de saída e chegada, Liang *et. al* caracterizaram 46.000 viagens entre 2.017 zonas dentro da cidade de Los Angeles, nos Estados Unidos.

A coleta de dados ativa, conforme feito nos trabalhos de Lotero e Liang *et. al* fornece um conjunto de dados mais detalhado, onde é possível averiguar até o motivo da viagem. Entretanto, esse aumento de resolução e precisão sacrifica a escala da pesquisa, dificultando a generalização dos padrões (BARBOSA, 2018).

2.3 Modelos de mobilidade humana em espaços urbanos

Em muitos trabalhos que visam a compreensão da mobilidade humana nos espaços urbanos são utilizadas técnicas para relacionar os indivíduos estudados a um local de residência, para a partir daí caracterizar a mobilidade destes. Em seu trabalho (VANHOOF, 2018) procura atribuir locais de residência dos indivíduos tomando como critério o dia e horário em que estes realizam atividades telefônicas. Para isso, ele considera que ligações realizadas em períodos dados como de descanso - como, por exemplo, a noite e durante fins de semana - são feitas do local de moradia dos indivíduos.

A presunção de residência facilita a aferição de uma série de métricas - raio de giração, distância máxima percorrida, distância máxima percorrida de casa - que podem subsidiar a caracterização da mobilidade humana.

Matrizes de origem e destino são amplamente utilizadas para a caracterização da mobilidade humana nas cidades. (LOTERO-VéLEZ, 2016) e (CHAVES, 2018) fazem uso desse conceito para modelar as trajetórias dos indivíduos nos contextos estudados. Na modelagem dessas matrizes, são utilizados grafos onde cada nó representa uma localidade urbana, e existe uma aresta entre esses nós se houver, para uma determinada trajetória

de indivíduo (ou via de transporte, no caso de Lotero *et. al*) o deslocamento de uma localidade de origem para uma outra de destino.

Os trabalhos de (NOULAS S. SCELLATO, 2012) e (HAWELKA I. SITKO, 2014) buscam caracterizar a mobilidade urbana através de dados de redes sociais online. O desenvolvimento de funcionalidades de *GPS* e placas de wifi em smartphones, propiciou uma riquissíma fonte de dados para a análise de deslocamento dos indivíduos.

Além dos dados de deslocamento dos indivíduos, os dados extraídos através de redes sociais online podem revelar aspectos sociais e econômicos. Noulas e Hawelka *et. al* salientam que através desse tipo de dado é possível caracterizar os contatos dos indivíduos, dado a possibilidade de marcar pessoas em publicações em redes sociais como instagram e facebook e também pode-se caracterizar os indivíduos economicamente conforme os locais que visitam.

(BARBOSA, 2018) menciona uma outra forma de medir a mobilidade dos indivíduos, porém em um escala espaço-temporal maior do que as até aqui citadas. Utilizando dados da Receita Federal Norte-Americana, são comparados as residências dos cidadões entre um ano fiscal e outro, e havendo diferença entre ambos, traça-se matrizem de origem e destino.

Em relação as métricas estatísticas que foram aferidas, como raio de giração e distância percorrida (serão detalhadas no capítulo 3), que visam quantificar distâncias percorridas, número de locais visitados entre outras características dos indivíduos, é importante mencionar o trabalho de (PAPPALARDO, 2019), que aponta o repositório python *scikit-mobility* como um centralizador de funções que implementam essas métricas, facilitando o desenvolvimento de trabalhos na área.

2.4 Mobilidade e aspectos socioeconômicos

O trabalho de (LOTERO-VéLEZ, 2016) é uma das principais referências na caracterização da mobilidade humana nos grandes centros urbanos sob uma perspectiva socioeconômica. Como mencionado anteriormente, ele busca caracterizar a utilização dos modais de transportes das cidades estudadas para cada conjunto de classe econômica.

Assim como Lotero *et. al*, porém tomando a cidade do Rio de Janeiro como objeto de estudo, (RODRIGUES, 2018) busca estratificar a população em classes de renda para analisar a mobilidade dos indivíduos. Entretanto, diferentemente de Lotero *et .al*, a obtenção dos dados das trajetórias dos indivíduos se dá de uma maneira passiva. Neste mesmo trabalho, Rodrigues *et. al* fazem uso de dados *CDR* para a obtenção das trajetórias dos indivíduos. A caracterização socioeconômica se dá através de dados do Censo Demográfico Brasileiro de 2010.

Ainda no Rio de Janeiro, (MOTTE B., 2016) faz uma análise interessante sobre as trajetórias estratificando os indivíduos pelo o nível de capacitação profissional e pelo o regime de contratação (formal, quando há registro previdenciário e informal, quando não há).

Utilizando de métodos ativos para a obtenção dos dados, Motte *et. al* revelam que os moradores mais pobres e menos qualificados precisam realizar os deslocamentos pendulares mais longos e penosos e, ainda, que os trabalhadores informais realizam deslocamentos pendulares mais curtos e dispersos quando comparados aos trabalhadores formais.

Outro trabalho que visa caracterizar a mobilidade humana através de um viés socioeconômico é o de (RODRIGUES, 2021). Também fazendo uso de dados de telefonia móvel, Rodrigues *et. al* divide os usuários entre residentes de favelas e não residentes de favelas - tomando como insumo dados do IPTU (Imposto Predial e Territorial Urbano) da cidade de São Paulo - e compara as trajetórias, distâncias percorridas e locais visitados desses indivíduos em diferentes períodos do dia e em diferentes dias (de semana ou fim de semana). Nesse trabalho, Rodrigues *et. al* constataram que na maioria dos cenários experimentados as trajetórias dos indivíduos residentes em favelas são maiores do que dos outros, confirmando um antigo e conhecido problema brasileiro, que é a exclusão/privação geográfica que os moradores de áreas carentes vivenciam em nossas grandes cidades.

(STEELE J. E., 2017) em seu mapeamento sobre a distribuição da pobreza na cidade indiana de Bangladesh utiliza diversas fontes para caracterização econômica dos indivíduos. Steele *et. al* fazem uso de diferentes questionários afim de classificar as famílias em um ranking que toma como referência certas métricas para o estabelecimento da pobreza. Dados *CDR* e de satélites também foram usados com o objetivo de identificar características ambientais e físicas sobre as residências estudadas.

Nesse trabalho, Steele *et. al* procuram demonstrar como a combinação entre dados públicos e privados podem ser usados com a finalidade de oferecer informações relevantes sobre a distribuição espacial da pobreza em contextos urbanos.

3 Metodologia

Este trabalho toma como base de dados os registros de chamadas telefônicas (*Call Detail Records - CDR*) realizadas em regiões geográficas imediatas no Brasil e dados socioeconômicos obtidos por informações demográficas dessas regiões, através do Censo Demográfico Brasileiro de 2010, realizado pelo IBGE.

A partir dos dados *CDR* e de informações censitárias, propõe-se uma série de passos metodológicos, descritos nas seções seguintes.

3.1 Pré-Processamento e Caracterização das relações sociais com base nos dados CDR

Dados *CDR* veem sendo usados como insumo para caracterização das relações sociais entre os indivíduos através de redes complexas em muitos estudos. Nesses trabalhos, grafos são utilizados para representar a interação entre indivíduos, onde cada vértice é uma pessoa e há uma aresta entre dois vértices se ocorreu uma ligação telefônica entre eles.

A fim de averiguar as relações sociais e as trajetórias dos indivíduos, foram utilizados os seguintes atributos da base CDR: data e hora da realização da chamada telefônica, duração da ligação, IDs dos dispositivos de origem e destino das conexões, e ID da antena para qual a chamada é direcionada. Com o ID da antena, é possível encontrar suas coordenadas geográficas. A figura 2 apresenta uma amostra da base *CDR* de São João del-Rei:

Figura 2 – Amostragem da base CDR de São João del-Rei

DATE	HOUR	DURATION	DDD_USER_FROM	USER_FROM	DDD_USER_TO	USER_TO	ANTENNA
2013-03-21	11:51:30	0.370000	32	5B5F2C071D12AF13219DF5EBE05132AF	32	9FB3B96B6D5E16C9DD564AA3E84F1954	50141
2013-03-21	21:03:30	0.130000	61	B3299B0E587D7275E3E4D530E9EECF50	98	6432F1DF21BA38368D9A165C739EEBB3	51601
2013-03-21	17:33:20	0.330000	61	85D5C50A6D882CA8E4BB00BCA3574417	61	0D8583F810B9720A8032BB939F12B3FF	51601
2013-03-21	11:58:26	1.270000	61	6C10A9E9F325CAA3CCB7F9A0D6983D2A	61	B0C50ED1DEC9E06E4C64E7419DDC4B09	51601
2013-03-21	16:06:59	0.170000	61	29122DCF4189448865B03F24C2DFEAE6	61	E0B097D86BF9C02A23011610EBAF47FD	51601
2013-03-21	11:14:19	2.120000	61	07B468660874AD7C8FE958741A768181	61	728614A7ED24715CC8502F6B4DD53491	51601
2013-03-21	20:30:29	7.020000	61	BA4DEA1F710C1216F74F1255150C49D6	61	4F6DCBA36292B8487743E13319E67DF6	51601

A grande quantidade de dados permite uma generalização de comportamento dos usuários. Porém, tal volume de registros contêm dados irrevelantes ou desnecessários para qualquer análise. Um passo muito importante ao se trabalhar com dados CDR é filtrar

apenas usuários dos quais pode-se, de fato, extrair informação útil no contexto explorado. Assim, é fundamental realizar um pré-processamento dos dados que elimine da análise registros de ligações telefônicas que possam conter erros de anotações ou que representem ligações telefônicas realizadas por engano. Portanto, a seguinte filtragem é feita no dataset:

- Exclusão de ligações as quais o tempo de duração ficou fora do intervalo de 3 segundos e 120 munutos;
- Exclusão de indivíduos que realizaram menos que 3 e mais que 500 ligações no período de tempo analisado;
- Exclusão de indivíduos que realizaram ligações em menos que 3 dias distintos.

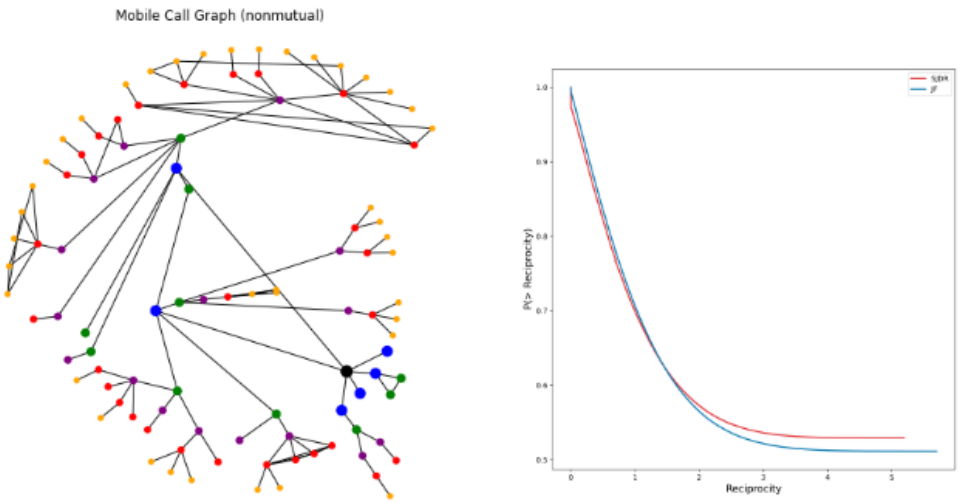
Além disso, considerando que os registros de uma comunicação entre dois indivíduos devem refletir uma interação social, pode-se tomar como base os resultados dos trabalhos de Robin Dunbar (CARRON, 2016), mais especificamente o conhecido número de Dunbar, eliminando da rede vértices que mantêm mais do que 150 relações sociais, que poderiam indicar centrais de atendimento e números com ramais.

Após esses filtros foram desenvolvidas redes sociais em suas versões direcionada e não direcionada. Sobre essas redes foram aplicados novos filtros, visando eliminar indivíduos que não realizaram ligações. Para isso, excluiu-se vértices com grau de saída nulo.

Foi utilizado o método de Louvain para detecção de comunidades (BLONDEL, 2008), que recebe como entrada uma rede - nesse caso específico, utilizamos as redes não direcionadas - e tem como resultado um agrupamento otimizado dos indivíduos em comunidades.

Conforme apontado por (ONNELA, 2007), redes sociais baseadas em ligações telefônicas apresentam um aspecto parecido com uma árvore - uma estrutura topológica baseada em hierarquia de redes e sub-redes onde existam nós centrais e os demais nós se ramificam a partir destes -, o que pode fazer com que algoritmos clássicos para detecção de comunidades produzam resultados insatisfatórios neste tipo de redes. A Figura 3 a) apresenta uma amostra em *snowball* da rede social SJDR. O vértice marcado na cor preta representa o ego, a partir do qual são obtidos vértices em até seis saltos. Buscando aprofundar ainda mais a discussão sobre a interação entre os indivíduos, a figura 3 b) demonstra a distribuição da reciprocidade da mesma rede mencionada anteriormente, onde podemos notar que os valores de reciprocidade descrescem a medida que a distância social entre os indivíduos aumenta.

Figura 3 – Algumas características sociais da rede. (a) Exemplo de uma amostra da rede usando *snowball sampling* com seis saltos (o vértice em preto é o ego). (b) Distribuição da reciprocidade na rede.).



(a) (b)

3.2 Presunção de residência e caracterização do estrato social

Dados CDR são armazenados por diferentes operadoras de telefonia móvel, que devem trocar informações com o propósito de efetuar a cobrança devida pela utilização das antenas por seus clientes e geralmente armazenam informação de um ID do usuário que realizou a chamada, um ID do usuário que recebeu a chamada, assim como sua duração. Dados CDR também armazenam o ID da antena que o dispositivo móvel utilizou para realizar a chamada o que, juntamente com sua localização, pode ser utilizado para estimar a localização do usuário.

Neste trabalho pretende-se investigar aspectos socioeconômicos apenas dos usuários residentes em uma determinada região imediata, que é definida segundo o IBGE como estruturas a partir de centros urbanos próximos para a satisfação das necessidades imediatas das populações, tais como: compras de bens de consumo duráveis e não duráveis; busca de trabalho; procura por serviços de saúde e educação; e prestação de serviços públicos, como postos de atendimento do Instituto Nacional do Seguro Social - INSS, do Ministério do Trabalho e de serviços judiciários, entre outros.Conforme mencionado na 1, as regiões imediatas de São João del-Rei e Juiz de Fora foram selecionadas devido a necessidade de tentar compreender as dinâmicas populacionais de centros urbanos menores, privilegiando a região de atuação da UFSJ. Assim, é importante descartar usuários dos quais não é possível presumir residência, pressupondo assim que estes não são residentes das regiões estudadas. Neste trabalho, a presunção de residência será baseada nos trabalhos de Vanhoof *et al.* (VANHOOF, 2018). Para isso, identifica-se cada indivíduo que efetue no

mínimo 5 e no máximo 50 chamadas em 7 dias distintos. Será considerada sua residência o local onde ele realizou mais que 50% das chamadas nos domingos ou das 19h às 6h no restante da semana.

As informações de estratos sociais dos indivíduos serão obtidas a partir da base de dados do CENSO 2010, 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 setor censitário, e o rendimento nominal médio do setor censitário foi utilizado como uma estimativa do rendimento das residências associadas àquele setor. A divisão de classes será 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.

Classe social	Descrição	Intervalo de renda
1	Até 1 salário mínimo	R$0,00 a R$510,00
2	De 1 a 2 salários mínimos	R$510,00 a R$1020,00
3	De 2 a 3 salários mínimos	R$1020,00 a R$1530,00
4	De 3 a 5 salários mínimos	R$1530,00 a R$2550,00
5	De 5 a 10 salários mínimos	R$2550,00 a R$5100,00
6	De 10 a 20 salários mínimos	R$5100,00 a R$10200,00
7	Acima de 20 salários mínimos	R$10200,00

Tabela 1 – Divisão de classes de renda. Os intervalos abertos à esquerda significam que os valores devem ser maiores do que o limite inferior à esquerda

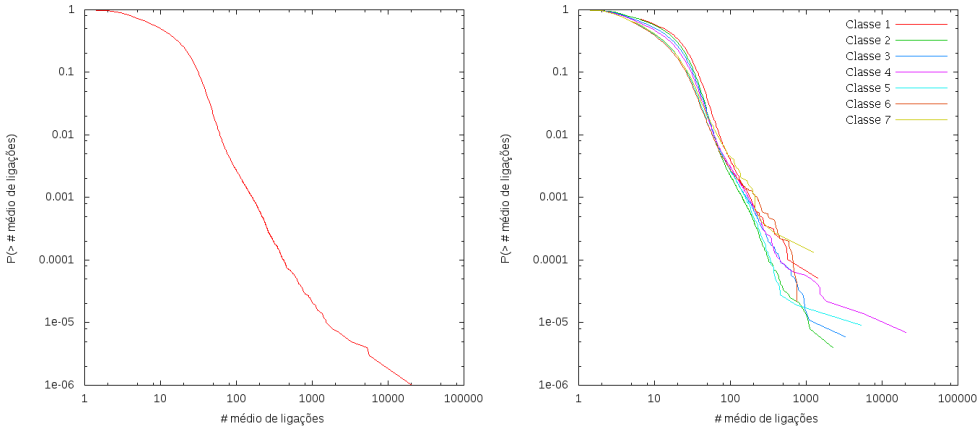
Com a classificação de renda de cada setor sensitário e a presunção de residência dos indivíduos definida, adicionamos aos vértices da rede a classe de renda atribuída ao setor sensitário de sua residência, possibilitando a análise sob um viés econômico.

3.3 Caracterização básica da mobilidade dos indivíduos

Após o pré-processamento da base de dados e a estratificação social dos indivíduos, definida pela presunção de residência, uma série de caracterizações pode ser realizada, com o objetivo de obter uma compreensão mais profunda sobre a forma como os indivíduos de diferentes estratos utilizam o espaço urbano. Para isso, uma série de ferramentas de caracterização de mobilidade descritas na literatura será utilizada. A seguir, são apresentadas algumas das medidas de caracterização que serão exploradas neste trabalho.

Incialmente, é importante caracterizar como se dá a distriubuição cumulativa complementar do número médio de ligações para a base de dados *CDR* completa e estratificada por classe de renda. Assim, espera-se obter um resultado que seja semalhante ao apresentado por Rodrigues *et al.* (RODRIGUES, 2018), como pode ser visto na Figura 4,

onde é possível observar um comportamento semelhante a uma *power law* em uma parte significativa do intervalo.



(a) Todos os registros (b) Registros divididos por classes

Figura 4 – Distribuição cumulativa complementar de probabilidades do número médio de ligações: (a) todos os registros; (b) registros divididos por classes (retirado de (RODRIGUES, 2018)).

Considerando especificamente a forma como os indivíduos se movimentam nas cidades, o que é um aspecto fundamental neste trabalho, diversas medidas podem ser encontradas na literatura com o objetivo de caracterizar a mobilidade sob uma perspectiva individual ou coletiva e serão tomadas como base para construir a metodologia proposta.

Uma das medidas mais básicas para a caracterização da mobilidade individual é o raio de giração. Considerando um indivíduo u, o raio de giração r_g mede a distância característica por ele percorrida e pode ser definido como

$$r_g = \sqrt{\frac{1}{n_u}\sum_{i=1}^{n^u} dist(r_i(u) - r_{cm}(u))^2}, \tag{3.1}$$

onde $r_i(u)$ representam as n_u posições de u e r_{cm} é o centro de massa da trajetória de u.

Dentro da área de teoria da informação, a entropia busca avaliar o grau de casualidade e indeterminação que pode ser observado em uma fonte (SHANNON, 2001). Quanto maior a entropia em um conjunto, maior é a imprevisibilidade de seus elementos. Por exemplo, jogando uma moeda honesta para cima, o grau de incerteza dos resultados de cara ou coroa, é menor do que o grau de incerteza do resultado do lançamento de um dado honesto de seis faces. Desse modo o dado oferecerá uma maior incerteza e entropia. Considerando o contexto da mobilidade de indivíduos no espaço urbano, podemos definir a entropia de um indivíduo u como

$$E(u) = -\sum_{T_u'} P(T_u')log_2 P_{T_u'}, \tag{3.2}$$

onde $P(T_u)$ é a probabilidade de encontrar uma sequência de localizações $T_u^{'}$ na trajetória T_u.

Será calculada também a distância máxima d_{max} percorrida por cada indivíduo u, definida como

$$d_{max} = \max_{1 \leq i < j < n_u} dist(r_i, r_j). \tag{3.3}$$

, e a distância máxima de casa, com a seguinte fórmula:

$$d_{max} = \max_{1 \leq i < j < n_u} dist(r_i, h(u)). \tag{3.4}$$

Considerando medidas de caracterização individuais, também serão calculados a distribuição do número de localizações visitadas por cada indivíduo. A partir da agregação de medidas individuais, os indivíduos serão caracterizados coletivamente e, para isso, serão consideradas medidas como o deslocamento médio dos indivíduos e a entropia de visitação das localizações.

3.4 Caracterização da relação entre aspectos socioeconômicos, rede social e padrões de mobilidade

Após o pré-processamento dos dados CDR e a construção da rede social em suas versões direcionada e não-direcionada, os passos metodológicos adotados neste trabalho são baseados na análise da combinação entre informações relacionadas às interações sociais com os padrões de mobilidade dos indivíduos. A cada vértice da rede é atribuída uma matriz de localização $\mathbf{L}$, definida por Toole *et al.* (TOOLE, 2015) como uma matriz $48 \times L$, onde L é o número de antenas de localização única no espaço observado e cada linha de $\mathbf{L}$ corresponde a uma hora t de seu dia de semana típico e seu fim de semana típico. Cada elemento $\mathbf{L}_{t,j}$ da matriz de localização de um usuário armazena a quantidade de vezes que um indivíduo realizou uma ligação utilizando a antena da localização j, o que é utilizado para estimar uma visita do usuário à localização j. Como discutido em (TOOLE, 2015), a matriz de localização pode ser de grande utilidade para o cálculo de diversas propriedades de mobilidade dos usuários, como número de ligações telefônicas realizadas por um indivíduo e a frequência de visitação dos usuários a determinadas localizações. Uma dessas propriedades, de particular importância neste trabalho, é o vetor de trajetórias $\mathbf{v}$, obtido através da soma das linhas da matriz de localização. O vetor de trajetórias, calculado para cada um dos usuários da rede, armazena a quantidade de vezes que um indivíduo visitou cada uma das localizações durante o período de observação e pode ser utilizado como uma forma de representar seu padrão de interação com o espaço urbano.

Considerando usuários representados por dois vértices v_i e v_j, pode-se medir sua similaridade de mobilidade através do cálculo da similaridade de seus vetores de trajetória.

Assim como proposto em (TOOLE, 2015), neste trabalho a similaridade entre os vetores de localização $\mathbf{v_i}$ e $\mathbf{v_j}$ é calculada através do cosseno do ângulo formado entre eles, como

$$cos\theta_{i,j} = \frac{\mathbf{v_i} \cdot \mathbf{v_j}}{|\mathbf{v_i}||\mathbf{v_j}|}. \tag{3.5}$$

A similaridade de cosseno mostra-se como uma medida bastante adequada para o cálculo da similaridade entre dois vetores de similaridade, quando comparada a outras medidas de similaridade, já que não leva em conta a magnitude dos vetores de trajetória, desconsiderando, assim, o número de ligações realizadas pelos usuários. Além disso, a similaridade de cosseno entre dois vetores não é afetada pelas posições vazias nos vetores, concentrando a análise sobre os lugares visitados.

Nesse trabalho em específico estamos interessados em investigar a relação entre as relações sociais e os padrões de mobilidade sob uma perspectiva socioeconômica, com o objetivo de identificar como fatores socioecnômicos influenciam na relação entre indivíduos e desses com o espaço urbano. Assim, os experimentos realizados visam explorar o impacto do fator econômico nos padrões de mobilidade e interações sociais.

4 Resultados e Discussão

A metodologia apresentada no Capítulo 3 foi aplicada no contexto das regiões imediatas de duas cidades - uma de porte pequeno e uma de porte médio - no estado de Minas Gerais, no Brasil: São João del-Rei e Juiz de Fora. Os experimentos apresentados neste capítulo podem ser divididos em dois conjuntos, visando responder de maneira mais clara às Questões de Pesquisa apresentadas no Capítulo 1. O primeiro conjunto de experimentos tem como objetivo descrever as relações entre os aspectos sociais e os aspectos econômicos nos espaços urbanos observados. O segundo conjunto de experimentos busca caracterizar a relação entre os padrões de utilização dos espaços urbanos observados e a estratificação social de seus indivíduos. Para uma melhor contextualização e compreensão dos resultados, a próxima seção apresenta algumas caracterizações básicas das regiões geográficas investigadas neste trabalho e do conjunto de dados considerado.

4.1 Caracterização das regiões geográficas e do conjunto de dados

Algumas características básicas das regiões imediatas de Juiz de Fora e São João del-Rei, sobre as quais se apoiam os experimentos realizados neste estudo, foram identificadas com o objetivo de contextualizar a investigação. As Tabelas 3 e 2 apresentam a população de cada cidade que compõem as regiões imediatas.

Cidade	População estimada (2021)	População do Censo de 2010
Conceição da Barra de Minas	3.939	3.954
Coronel Xavier Chaves	3.448	3.301
Lagoa Dourada	13.115	12.256
Madre de Deus de Minas	5.119	4.904
Nazareno	8.710	7.954
Piedade do Rio Grande	4.436	4.709
Prados	9.128	8.391
Resende Costa	11.578	10.913
Ritápolis	4.521	4.925
Santa Cruz de Minas	8.723	7.865
São João del-Rei	90.897	84.469
São Vicente de Minas	7.876	7.008
Tiradentes	8.160	6.961
Total	179.650	148.832

Tabela 2 – População por cidade da região imediata de São João del-Rei

Cidade	População estimada (2021)	População do Censo de 2010
Andrelândia	12.189	12.173
Aracitaba	2.056	2.058
Arantina	2.779	2.823
Belmiro Braga	3.442	3.403
Bias Fortes	3.043	3.793
Bocaina de Minas	5.088	5.007
Bom Jardim de Minas	6.444	6.501
Chácara	3.216	2.792
Chiador	2.657	2.785
Coronel Pacheco	3.095	2.983
Ewbank da Câmara	3.932	3.753
Goianá	4.014	3.659
Juiz de Fora	577.532	516.247
Liberdade	4.994	5.346
Lima Duarte	1.6794	16.149
Matias Barbosa	14.626	13.435
Olaria	1.694	1.976
Oliveira Fortes	2.128	2.123
Paiva	1.517	1.558
Passa Vinte	2.024	2.079
Pedro Teixeira	1.804	1.785
Piau	2.179	2.841
Rio Novo	8.964	8.712
Rio Preto	5.493	5.292
Santa Bárbara do Monte Verde	3.212	2.788
Santa Rita de Jacutinga	4.843	4.993
Santana do Deserto	3.985	3.860
Santos Dumont	46.357	46.284
Simão Pereira	2.621	2.537
Total	751.872	673.586

Tabela 3 – População por cidade da região imediata de Juiz de Fora

Além de pouco populosas, as regiões estudadas também são pouco povoadas, como pode se observar nos mapas de densidade demográfica a seguir:

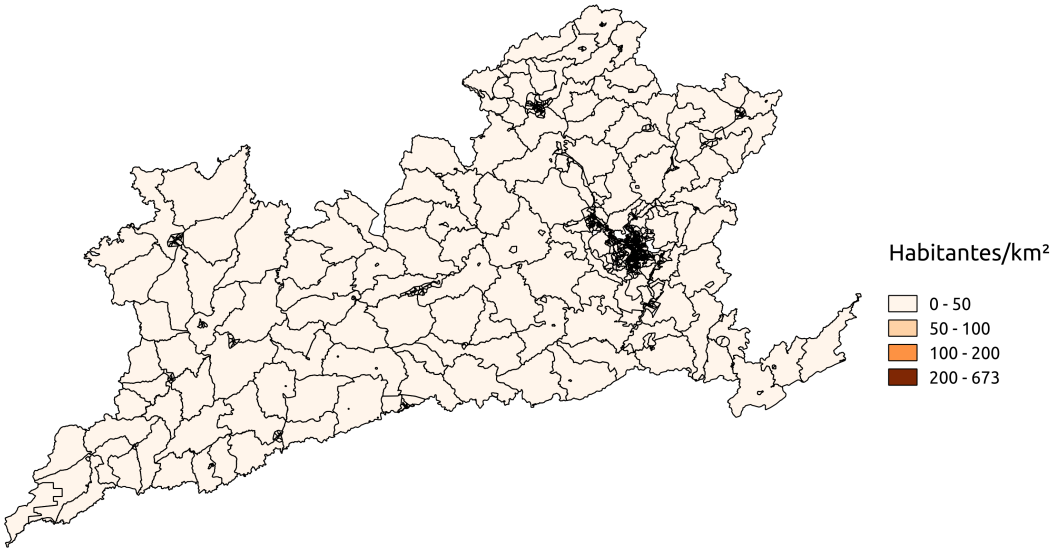


Figura 5 – Densidade demográfica Juiz de Fora

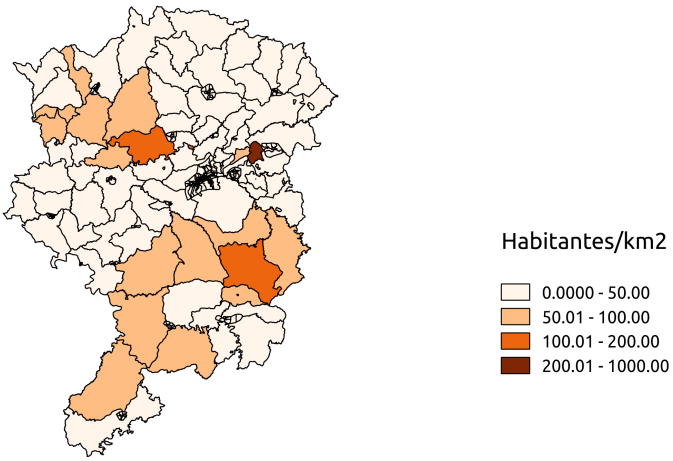


Figura 6 – Densidade demográfica São João del-Rei

Os dados *CDR* utilizados como insumo para os experimentos realizados neste trabalho remetem ao período entre 21 de março de 2013 a 19 de abril do mesmo ano. A base de ligações de Juiz de Fora tem um total de 63.588.996 registros, enquanto a de São João del-Rei contém 9.293.317 ligações. A Tabela 4 apresenta algumas características estatísticas básicas das bases de dados.

Região imediata	Quantidade de indivíduos	Média de ligações	Desvio padrão do número de ligações
São João del-Rei	1.303.639	20,876	51,454
Juiz de Fora	3.060.046	57,850	100,770

Tabela 4 – Características das bases cdr de cada região imediata

Apesar das regiões imediatas serem compostas por diversas cidades, as antenas - foram encontradas 12 na região imediata de São João Del Rei e 349 em Juiz de Fora - utilizadas para direcionar as ligações se concentram nos principais centros urbanos de cada região. Isso indica uma alta densidade demográfica em locais onde são ofertados os principais bens de consumo, que implica também em maior necessidade de mão-de-obra e consequentemente uma infraestrutura (nesse caso, de telecomunicação) mais robusta. Os

mapas apresentados nas Figuras 7 e 8 mostram essa concentração, além de classificar os setores censitários por classe de renda.

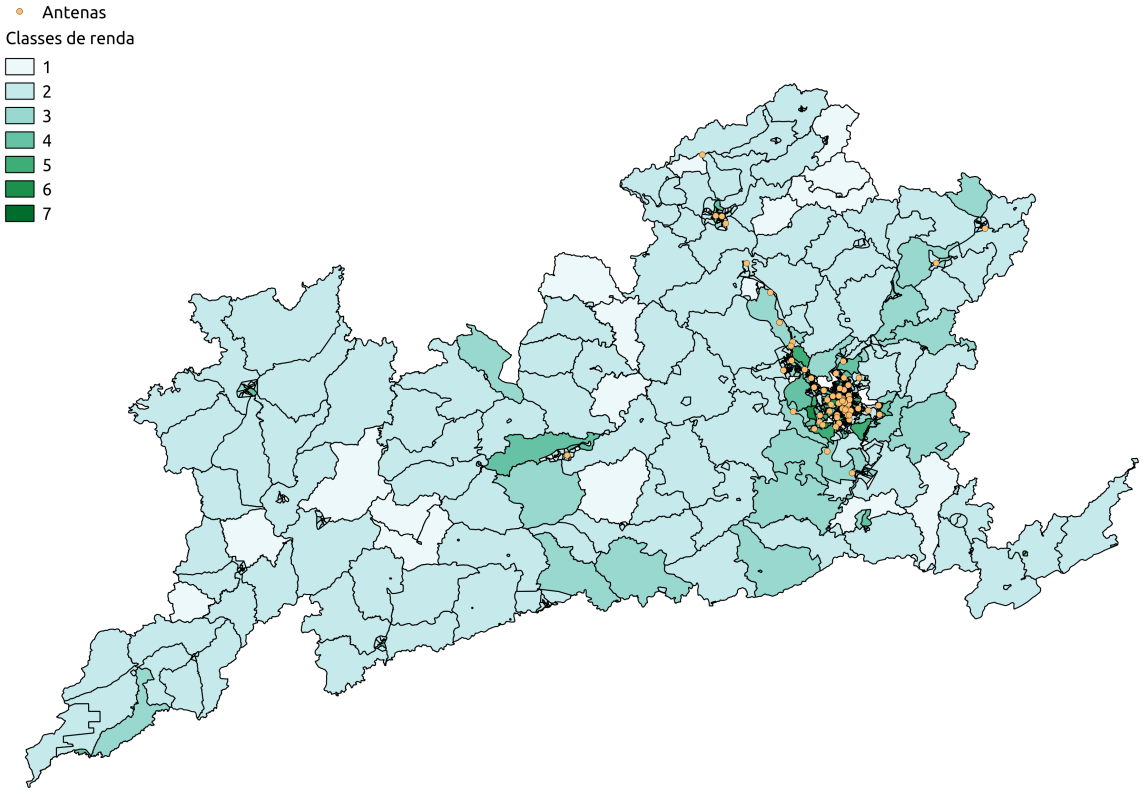


Figura 7 – Região imediata de Juiz de Fora

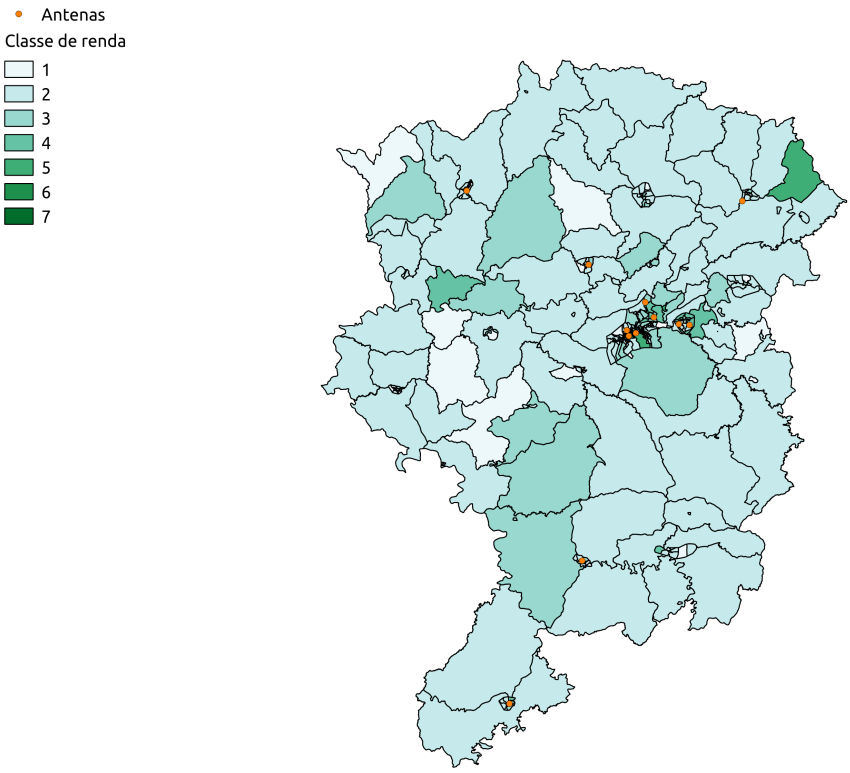


Figura 8 – Região imediata de São João del-Rei

Como visto no capítulo 2, um dos principais problemas ao se tentar compreender a mobilidade dos indivíduos a partir de dados *CDR* é justamente a concentração de antenas em locais mais densamente povoados. Isso pode provocar um falha na contabilização dos deslocamentos de indivíduos residente em áreas menos povoadas ou rurais, dado que uma mesma antena cobre um área mais extensa.

Dito isso, é importante salientar os vazios demográficos e áreas rurais características de ambas regiões imediatas (sobretudo de São João del-Rei). Das 29 cidades que compõem a região imediata de Juiz de Fora, 23 tem menos que 10.000 habitantes, enquanto que em São João del-Rei temos 10 em 13.

Além disso, uma característica das cidades mineiras de pequeno porte é que ainda é comum a presença de agricultores familiares, onde a família ou membros dela trabalham na prdodução e comercialização de produtos agrícolas, sendo que todo o processo de produção é feito em pequenas propriedades rurais. Essa característica intrínseca aos pequenos municípios das regiões estudadas dificulta mais ainda a acurácia das movimentações diárias dos indivíduos residentes em tais áreas.

4.2 Presunção das residências dos indivíduos

Com a aplicação do algoritmo para a presunção de residência descrito na Seção 3.2, foi possível identificar a residência de 166.563 e 546.356 indivíduos nas redes de São João del-Rei e Juiz de Fora, respectivamente. Em cada uma das regiões geográficas, cada região censitária foi associada a uma antena mais próxima, gerando uma associação de cada antena a um ou mais setores censitários. A partir disso, foi possível definir uma renda média estimada para a região associada a cada antena, obtida através da média das rendas de seus setores censitários. A cada um dos indivíduos foi, então, atribuída uma classe de renda, correspondente à renda associada à antena mais próxima de sua residência presumida. As Tabelas 5 e 6 apresentam a distribuição de indivíduos por classe de renda e a porcentagem de cada no total por região imediata.

Classe social	Quantidade absoluta	Porcentagem
1	8764	1,60
2	107819	19,73
3	168962	30,92
4	94967	17,38
5	165844	30,35
6	0	0
7	0	0

Tabela 5 – Divisão de indivíduos por classe de renda - Juiz de Fora

Classe social	Quantidade absoluta	Porcentagem
1	136427	81,90
2	26769	16,07
3	0	0
4	3367	2,02
5	0	0
6	0	0
7	0	0

Tabela 6 – Divisão de indivíduos por classe de renda - São João del-Rei

É importante ressaltar que o fato de que classes de renda sem indivíduos não significa de fato que não existam indivíduos dessa classe. Conforme dito na seção 3.2, a cada indivíduo foi atribuído a renda média do setor de possível residência.

Na análise da tabela da quantidade de indivíduos por classe de renda de São João del-Rei, podemos notar a ausência de indivíduos da classe 3, que pode ser explicado pelos fatos descritos na seção anterior e no capítulo 2, que a concentração de antenas prejudica a caracterização da mobilidade de indivíduos residentes em áreas de baixa densidade demográfica.

Já na análise da tabela 5, percebe-se um alto número de indivíduos da classe 5. Um fato que explica isso é a concentração de população da região presumida na cidade de Juiz de Fora, obviamente mais abastada economicamente que suas cidades "satélites". Além disso, há também a possibilidade de que o caráter noturno da cidade e a concentração de opções de lazer em áreas de maior capacidade econômica façam com que a presunção de residência de indivíduos que trabalham nesses locais seja errônea, visto que podem ser dados como residentes nos setores censitários de seu local de trabalho.

4.3 Características básicas das redes sociais

Uma das questões mais importantes a serem exploradas neste trabalho é a relação entre a forma como indivíduos se organizam socialmente no espaço urbano e os estratos sociais nos quais estão inseridos, o que pode ser observado na QP1, apresentada no Capítulo 1. Para isso, a metodologia para a obtenção das relações sociais, modeladas como redes complexas, foi aplicada aos dados CDR, como descrito na Seção 3.4.

Algumas características básicas das redes sociais obtidas para as regiões imediatas de Juiz de Fora e São João del-Rei são apresentadas nas tabelas 7 e 8, que descrevem o número de vértices, o número de arestas, o grau médio (de entrada e saída para as redes direcionadas), e o coeficiente de *clustering* dessas redes.

Rede	Número de nós	Número de arestas	Grau médio	Média do coeficiente *clustering*
SJ	161.948	227.412	2,8085	0,116
JF	662.212	2.093.618	6,3231	0.122

Tabela 7 – Características básicas das redes não direcionas

Rede	Número de nós	Número de arestas	Grau médio de entrada	Grau médio de saída	Média do coeficiente *clustering*
SJ	161.948	297.427	1,8366	1,8366	0,090
JF	662.212	2.770.314	4,1834	4,1834	0,101

Tabela 8 – Características básicas das redes não direcionas

Uma das propriedades topológicas mais fundamentais e importantes em redes sociais é o número de vértices aos quais os vértices se ligam na rede, que podem revelar características do fenômeno estudado. As Figuras 9-16, apresentam as distribuições de graus para as redes construídas de Juiz de Fora e São João del-Rei, em suas versões direcionadas e não direcionadas. As distribuições de graus são também apresentadas para cada um dos estratos isoladamente, com o objetivo de contribuir com a discussão conduzida para os experimentos a serem apresentados na sequência deste trabalho. São também

apresentadas as distribuições de números de chamadas dos indivíduos em cada uma das regiões imediatas estudadas e cada um dos estratos sociais.

A seguir são apresentadas as curvas de Funções de Distribuição Cumulativas Complementares (CCDF - *Complementary Cumulative Distribute Function*) para as redes na região de Juiz de Fora em sua versão não-direcionada (Figura 9) e direcionada (Figuras 11 e 12). A Figura 10 apresenta a distribuição no número de ligações dos indivíduos da mesma região.

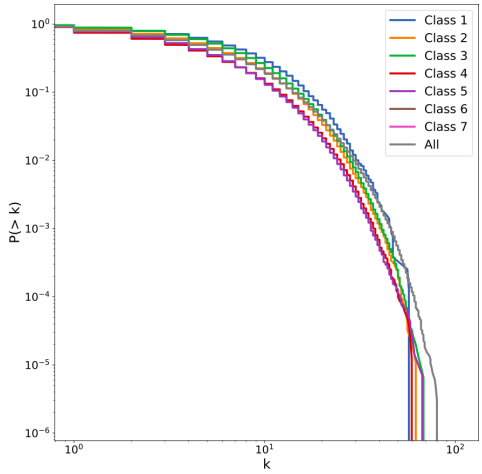


Figura 9 – Distribuição de graus por classe na rede não direcionada de Juiz de Fora

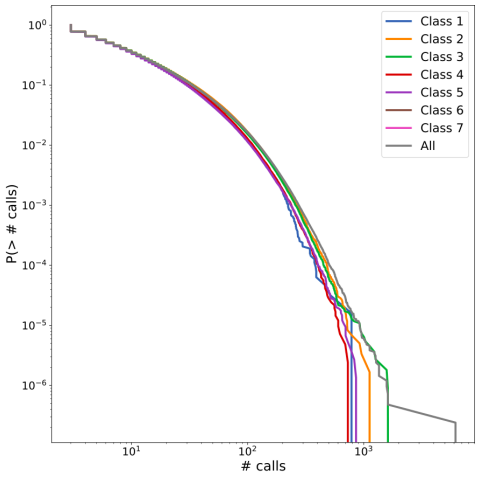


Figura 10 – Distribuição do número de ligações por classe em Juiz de Fora

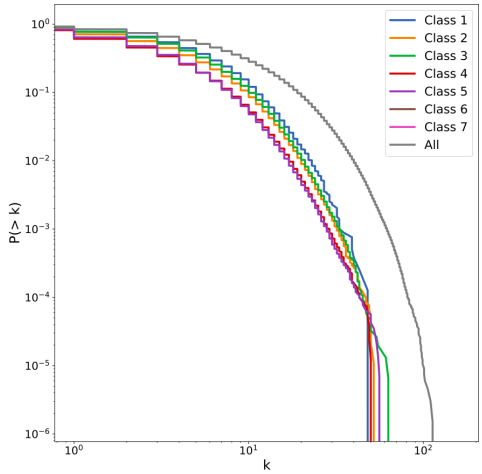


Figura 11 – Distribuição de grau de entrada na rede direcionada de Juiz de Fora

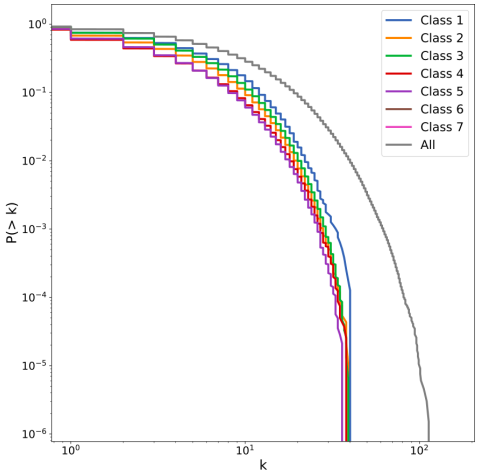


Figura 12 – Distribuição de grau de saída na rede direcionada de Juiz de Fora

Podemos notar que os indivíduos analisados em conjunto, sem segregação por classe, apresentam os maiores valores na distribuição de graus. Analisando os resultados

de cada estrato, temos que as classes 1 (com o menor número de indivíduos) e 3 (com o maior número de indivíduos) são aquelas com os maiores valores de grau na rede não-direcionada. Além disso, a classe 3 é a com os maiores números na distribuição de ligações, de grau de entrada e grau de saída.

Também foram obtidas as curvas CCDF para os graus e número de ligações nas redes para a região de São João del-Rei. As distribuições de graus são ilustradas pela Figuras, 13, 15 e 16. A distribuição de número de ligações é ilustrada pela Figura 14.

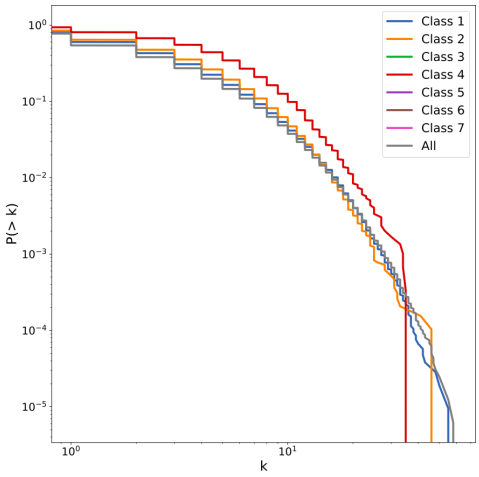


Figura 13 – Distribuição de graus por classe na rede não direcionada de São João del-Rei

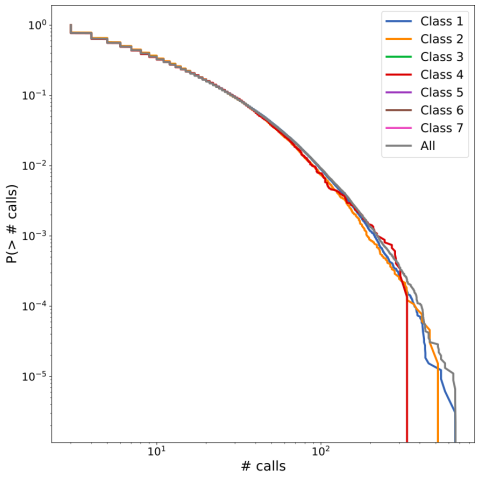


Figura 14 – Distribuição do número de ligações por classe em São João del-Rei

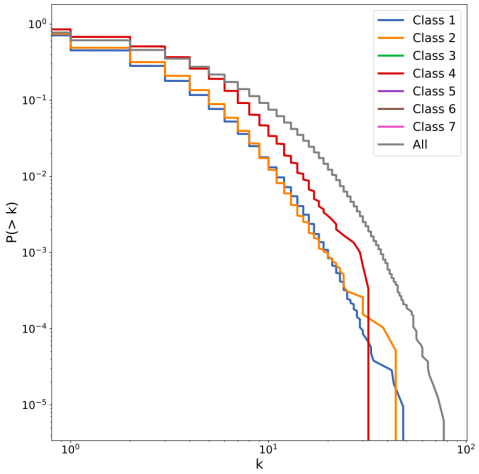


Figura 15 – Distribuição de grau de entrada na rede direcionada de São João del-Rei

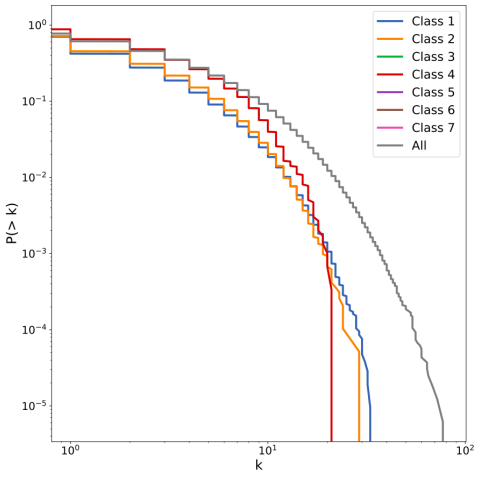


Figura 16 – Distribuição de grau de saída na rede direcionada de São João del-Rei

Nota-se que, assim como nas redes para a região de Juiz de Fora, nas redes de São João del-Rei os maiores valores estão primeiramente com a classe geral e em seguida com

a classe mais populosa, que no caso de São João del-Rei é a classe 2. Além disso, todos os estratos de renda respeitam a *power law*, como previsto na seção 3.3.

Conforme já discutido por Ribeiro *et al.* (RIBEIRO, 2021), talvez pelo menor volume de dados, os vértices da rede SJDR apresentam graus sensivelmente menores que os vértices da rede JF. A quantidade de dados também parece influenciar nos resultados por classe de renda, onde aquelas classes que têm a maior quantidade de individuos têm o grau mais alto.

Frequentemente, redes sociais apresentam grupos de vértices com densidades locais maiores do que o esperado caso as ligações acontecessem ao acaso, indicando a formação de conjuntos de vértices que podem representar interesses e propriedades em comum. Com o objetivo de investigar a formação de comunidades nas redes sociais modeladas a partir das ligações telefônicas estudadas, foi aplicada a metodologia apresentada na Seção 3.4. Buscando verificar se a divisão da rede em comunidades de fato representa a realidade, usou-se a modularidade como métrica de análise. A equação abaixo demonstra a definição dessa métrica:

$$Q = \sum_{c=1}^{n} [\frac{L_c}{m} - \gamma(\frac{\kappa_c}{2m})^2] \tag{4.1}$$

, onde a soma itera sobre todas as comunidades c, m é o número de arestas, L_c é o número de links intracomunitários para a comunidade c, κ_c é a soma dos graus dos nós na comunidade e γ é o parâmetro de resolução. A Tabela 9 apresenta a modularidade e o número de comunidades identificadas para as redes sociais construídas. Como a modularidade é uma caracteristica única das redes não direcionadas, as métricas tiradas englobam somente esse tipo de rede.

Rede	Número de Comunidades	Modularidade
SJ	161.948	0,563
JF	662.212	0,387

Tabela 9 – Características básicas das redes não direcionas

Tendo em vista a limitação da modularidade como uma medida para capturar a qualidade das comunidades identificadas em redes sociais formadas a partir de CDR, com o objetivo de investigar a divisão dos indivíduos em grupos, neste trabalho as comunidades foram avaliadas com uma estratégia que visa avaliar a prevalência de indivíduos de classes de rendas específicas em cada comunidade isoladamente.

As Figuras 17 e 18 apresentam *boxplots* para a porcentagem de indivíduos da classe predominante, i.e., a classe com mais indivíduos, em cada comunidade, para as redes não direcionadas de São João del-Rei e Juiz de Fora, especificamente.

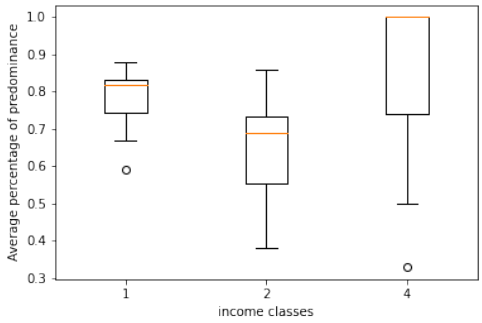


Figura 17 – Distribuição da porcentagem de presença da classe predominante em cada comunidade por classe de renda - SJDR

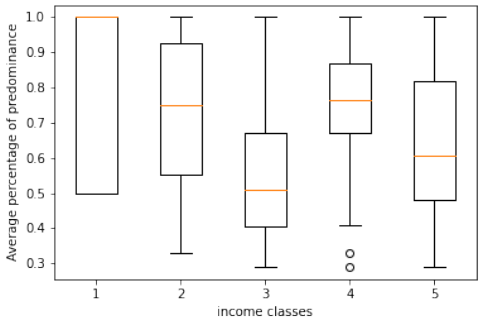


Figura 18 – Distribuição da porcentagem de presença da classe predominante em cada comunidade por classe de renda - JF

De uma maneira geral, em análise de dados, uma das tarefas mais importantes é encontrar uma propriedade que discrimine o comportamento dos dados. Considerando, mais especificamente, os dados CDR investigados neste trabalho, a principal hipótese explorada é a de que informações referentes às classes de renda dos indivíduos são bastante relevantes para caracterizar a forma como os indivíduos se relacionam entre si e com o espaço urbano. É claro que, devido à alta complexidade de fenômenos urbanos, é impossível destacar apenas uma propriedade que seja capaz de discriminar com eficiência os indivíduos a partir dos dados estudados. Porém, os resultados observados nas Figuras 17 e 18 permitem afirmar que a classe de renda na qual um indivíduo se insere é, sim, um fator muito importante para definir a forma como as pessoas se organizam na cidade. Pode-se observar que a mediana, para todas as classes de renda, está acima de 0,5, sinalizando a alta taxa de concentração de indivíduos da mesma classe dentro das comunidades e indicando que, independente da medida que se use para avaliar a qualidade das comunidades encontradas, é possível afirmar que as comunidades são capazes de discriminar indivíduos por suas classes de renda. Ou seja, mesmo que não seja possível apontar propriedades de fato conclusivas para a divisão de indivíduos nas cidades, o fato das comunidades apresentarem altas porcentagens de indivíduos de uma mesma classe, é mais um indicativo de que a classe de renda é uma variável capaz de refletir a maneira como as pessoas se organizam.

A combinação dos resultados obtidos na distribuição de graus e na análise das comunidades pode indicar que a interação é maior entre invíduos de uma mesma classe de renda do que de classes de renda distintas. Dado que as classes de renda com mais indivíduos têm os maiores valores na distribuição de grau e as comunidades são bem definidas em relação a classifcação econômica, podemos presumir que a interação entre indivíduos da mesma classe é muito mais frequente do que de classes diferentes.

4.4 Relação entre aspectos socioeconômicos e utilização do espaço urbano

Um dos objetivos principais deste trabalho é investigar o impacto de aspectos socioeconômicos, aqui modelados como as classes de renda das regiões de residências onde presume-se que os indivíduos vivem, sobre a forma como os indivíduos utilizam o espaço urbano, modelada como o conjunto de localizações por eles visitadas (QP2).

Para isso, diversas medidas clássicas frequentemente utilizadas para análise de mobilidade individual foram consideradas, conforme apresentadas na Seção 3.3. As medidas foram obtidas com o auxílio da biblioteca Python scikit-mobility[1].

As medidas de mobilidade investigadas foram calculadas para cada uma das regiões imediatas aqui consideradas (Juiz de Fora e São João del-Rei) e foram calculadas para cada uma das classes de renda isoladamente, com o objetivo de facilitar uma análise comparativa entre os diferentes estratos. Nesse trabalho, objetiva-se entender a mobilidade urbana do dia-a-dia, então faz sentido eliminar alguns usuários que representem fluxos de deslocamentos incompatíveis com o padrão que espera-se dos indivíduos das regiões estudadas. Para isso aplicamos alguns filtros sobre os deslocamentos dos indivíduos. Em Juiz de Fora, para cada usuário foi contabilizado somente trajetórias em dias que ele tenha realizado no minímo 7 atividades telefônicas e a trajetória máxima percorrida nesse dia seja inferior a 75 km. Já em São João del-Rei, temos uma quantidade considerável de indivíduos em que a distância diária percorrida é igual a 0. Visando eliminar esses indivíduos e outros que tem uma quilometragem de deslocamento muita acima da média, considerou-se apenas trajetórias diárias que fossem superiores a 0 e inferiores a 20km e também que os indivíduos donos dessas trajetórias tenham realizado no mínimo 3 ligações no dia. Tais filtros acabaram por eliminar os indivíduos da classe de renda 4 da análise. Obviamente, a metodologia aqui aplicada tem alguns pontos limitantes. Porém, o objetivo do trabalho é aumentar a gama de possibilidades para se compreender os padrões de deslocamento dos indivíduos nas cidades. Portanto, os resultados que são apresentados a seguir contribuem para a melhora da discussão.

As Figuras 19 e 20 apresentam os boxplots para os valores de raio de giração observados para cada uma das classes de indivíduos, respectivamente, na região de Juiz de Fora e São João del-Rei.

[1] https://scikit-mobility.github.io/scikit-mobility/

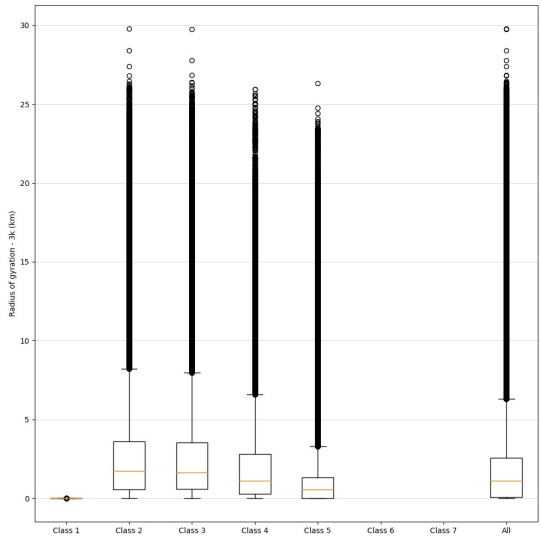


Figura 19 – Raio de giração por classe de renda - Juiz de Fora

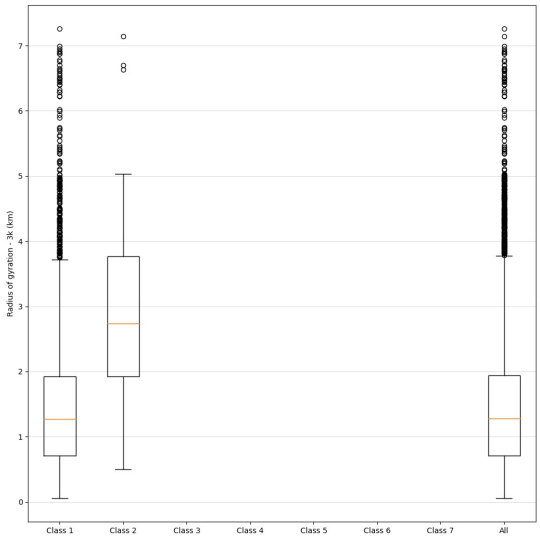


Figura 20 – Raio de giração por classe de renda - São João del-Rei

Analisando o raio de giração das duas regiões imediatas podemos observar que em Juiz de Fora, as classes de renda 4 e 5 apresentam valores médios menos elevados que a classe 2 e 3, indicando possivelmente uma menor necessidade de deslocamento. Em São João del-Rei a classe 1 apresenta valores médios menores que as demais classes, o que pode ser explicado por uma condição peculiar: Bairros com alta densidade demográfica e reconhecidamente mais carentes (como Tejuco e Senhor dos Montes) se localizam próximo ao centro econômico da cidade, enquanto outros bairros considerados de classes econômicas mais altas(colônia, por exemplo) se localizam distantes desse centro.

Os boxplots dos valores de distância máxima percorrida para as regiões de Juiz de Fora e São João del-Rei são apresentados, respectivamente, pelas Figuras 21 e 22.

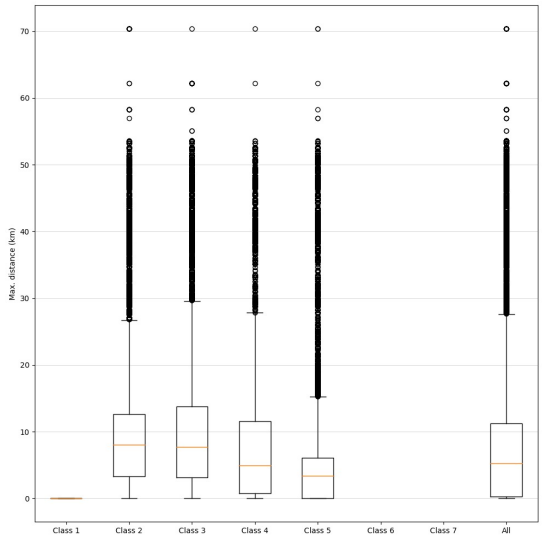


Figura 21 – Distância máxima percorrida por classe de renda - Juiz de Fora

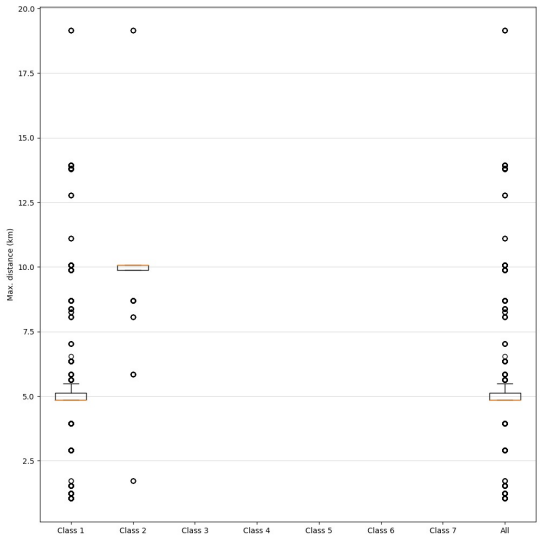


Figura 22 – Distância máxima percorrida por classe de renda - São João del-Rei

As Figuras 23 e 24 também apresentam os boxplots para os valores de distância máxima percorrida observados para cada uma das classes de indivíduos, respectivamente, na região de Juiz de Fora e São João del-Rei, porém considerando a residência presumida para os indivíduos como ponto de partida para as movimentações.

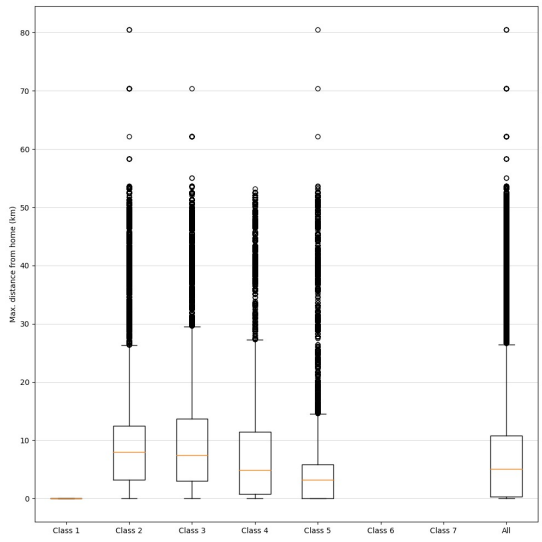


Figura 23 – Distância máxima percorrida de casa a partir da residência por classe de renda - Juiz de Fora

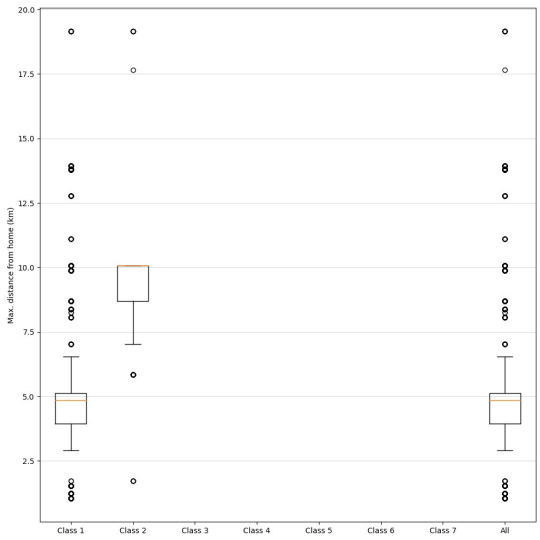


Figura 24 – Distância máxima percorrida de casa por classe de renda - São João del-Rei

Analisando a distância máxima e a distância máxima de casa das duas regiões imediatas, nota-se o mesmo padrão de deslocamento que o percebido no raio de giração:

Em Juiz de Fora as classes mais elevadas percorrem menores distâncias que as classes menos elevadas, e o oposto é observado para São João del-Rei.

Rodrigues *et al.* (RODRIGUES, 2021) apresentam um panorama parecido com o encontrado em Juiz de Fora, só que tomando como base de estudo a cidade de São Paulo: nesse estudo observou-se que durante o horário comercial os moradores de bairros de baixa renda estão, em média, mais distantes de seu local de residência do que os residentes de bairros de melhor classificação econômica. Esse padrão não foi observado em São João del-Rei, provavelmente devido as características geográficas da cidade, e também pela quantidade de dados significativamente inferiores em comparação com as outras cidades mencionadas.

As Figuras 25 e 26 apresentam os boxplots para o número de locais visitados pelos indivíduos de cada uma das classes de renda, respectivamente, na região de Juiz de Fora e São João del-Rei.

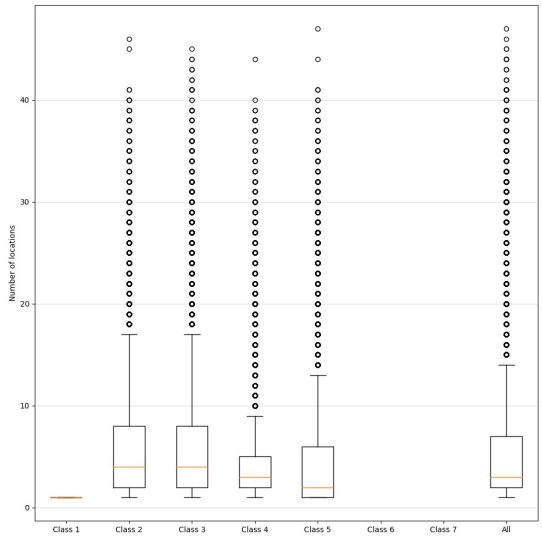


Figura 25 – Número de locais visitados por classe de renda - Juiz de Fora

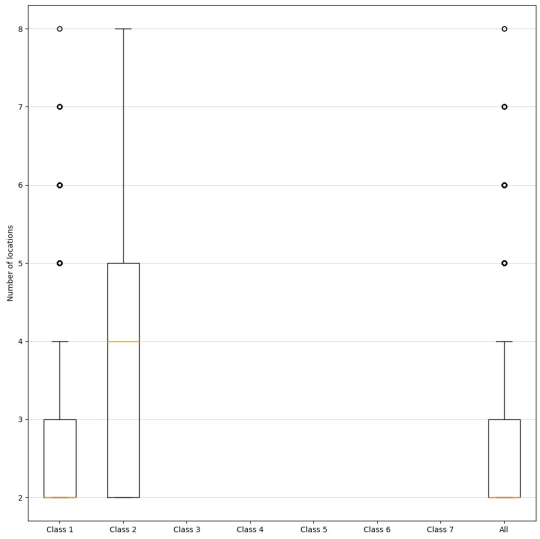


Figura 26 – Número de locais visitados por classe de renda - São João del-Rei

Analisando o número de locais visitados, pode-se afirmar que o padrão dos gráficos segue o dos experimentos anteriores. Um ponto interessante a se destacar é em relação a comparação dos valores médios encontrados para Juiz de Fora e São João del-Rei nos experimentos realizados até aqui. Juiz de Fora apresenta distâncias e locais visitados bem superiores aos encontrados em São João del-Rei, reforçando a ideia de que quanto a maior a cidade, maior a necessidade de deslocamento dos indivíduos.

As Figuras 27 e 28 apresentam os boxplots para as entropias observadas nas movimentações dos indivíduos de cada uma das classes de renda, respectivamente, na região de Juiz de Fora e São João del-Rei.

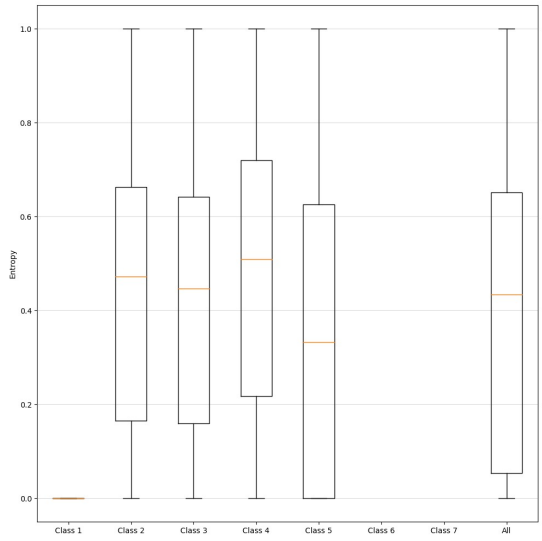


Figura 27 – Entropia por classe de renda - Juiz de Fora

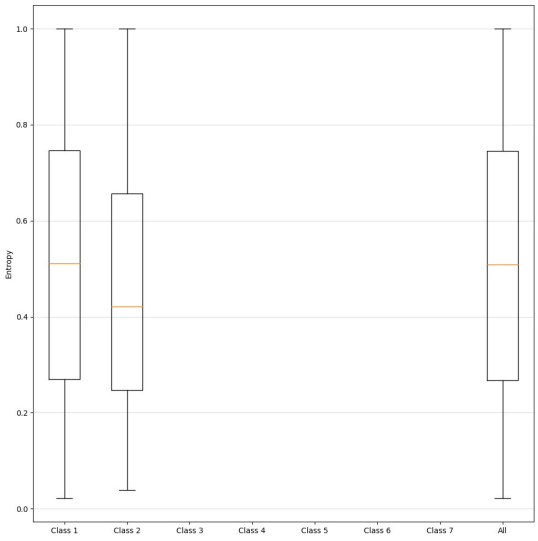


Figura 28 – Entropia por classe de renda - São João del-Rei

O valor médio de entropia para todas as classes das duas regiões imediatas está entre o intervalo de 0,45 e 0,5. Esses valores indicam que as trajetórias dos indivíduos tem um grau considerável de casualidade e imprevisibilidade.

4.5 Relação dos aspectos socioeconômicos com as interações entre indivíduos

Aqui fizemos uma série de análises buscando verificar se a condição econômica de um indívidiuo influencia na sua trajetória. Para isso, primeiramente comparamos a similaridade da trajetória de indivíduos de uma classe de renda, com todas as outras. Os gráficos das Figuras 29 e 30 evidenciam que as trajetórias são mais similiares entre indivíduos de uma mesma classe de renda do que entre indivíduos de classes de renda diferentes.

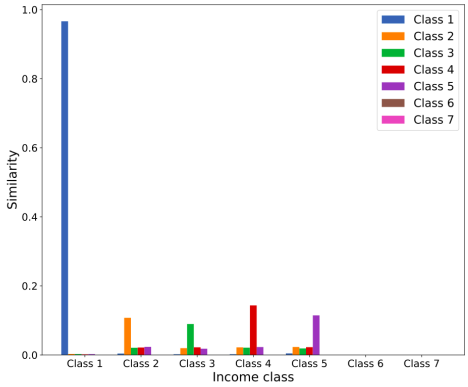


Figura 29 – Similaridade de trajetórias entre as classes de renda na região imediata de Juiz de Fora

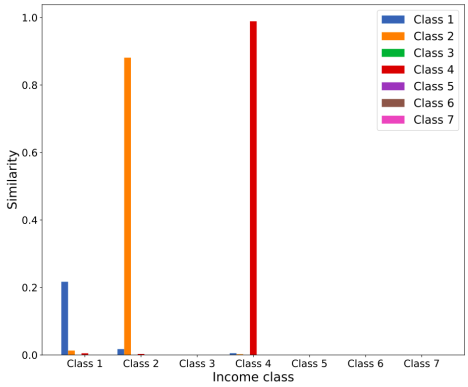


Figura 30 – Similaridade de trajetórias entre as classes de renda na região imediata de São João del-Rei

Além disso, buscamos relacionar o ranking de amizade dos indivíduos e a similaridade de trajetória entre eles, agrupando os resultados por classe de renda. O objetivo é verificar se o grau de interação entre dois indivíduos pode influenciar na trajetória de ambos. As Figuras 31 a 35 mostram os resultados de Juiz de Fora, enquanto as Figuras 36 a 38 apresentam os de São João del-Rei.

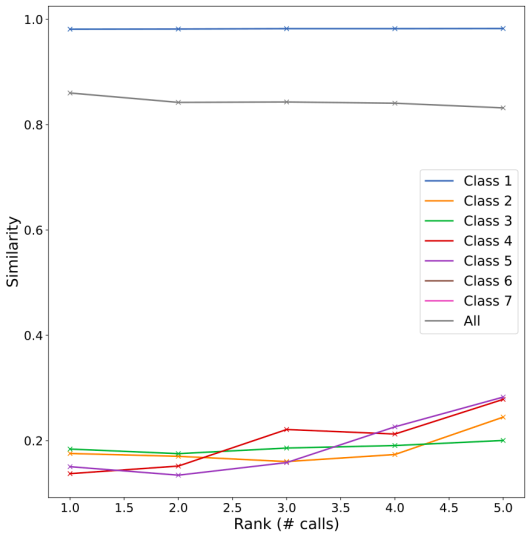


Figura 31 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 1

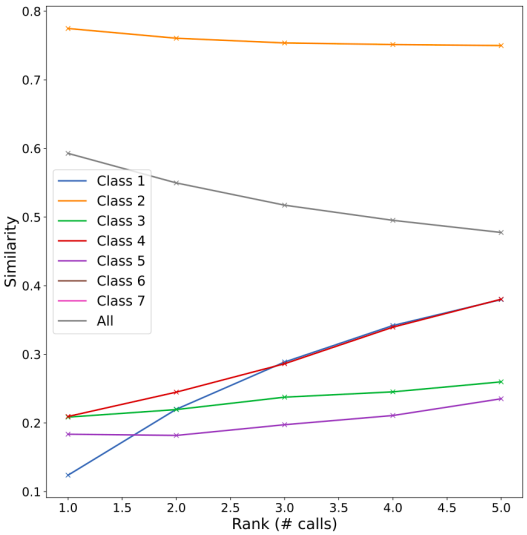


Figura 32 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 2

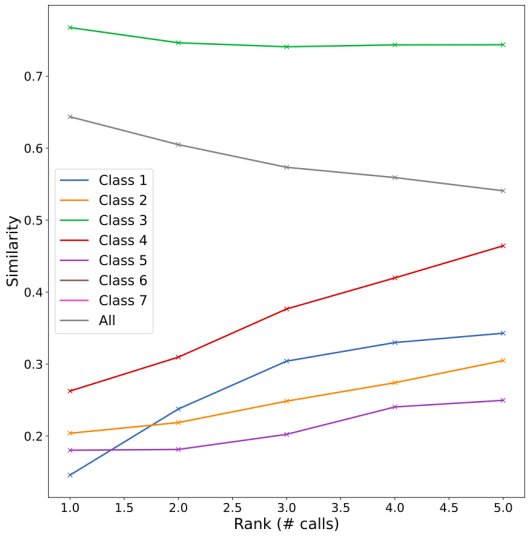


Figura 33 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 3

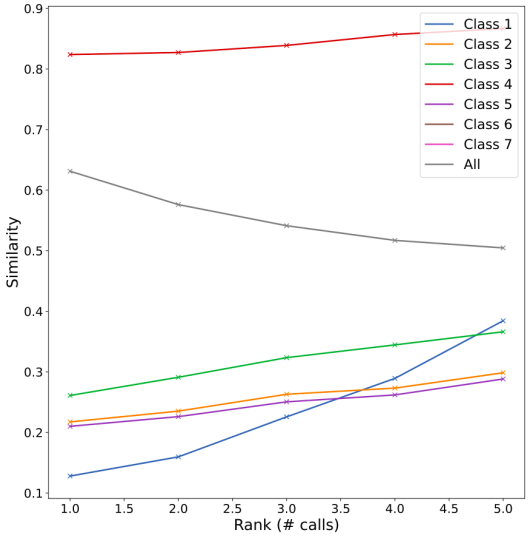


Figura 34 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 4

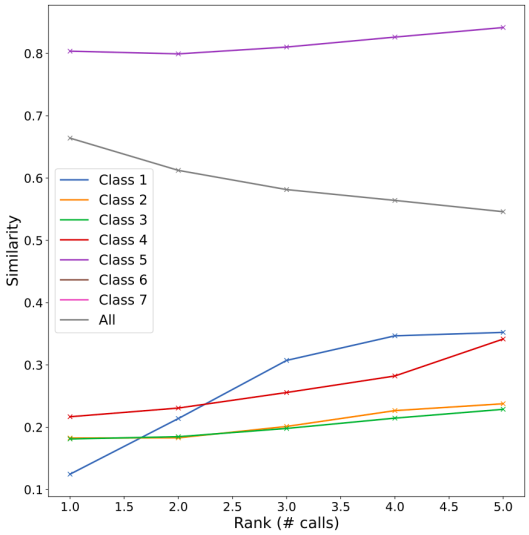


Figura 35 – Similaridade de trajetórias x Ranking de amigos - JF - Classe de renda 5

Analisando os dados de Juiz de Fora podemos constatar que a similaridade de trajetória é sempre mais alta entre indivíduos da mesma classe do que entre classes distintas. Para a classe 1 a similaridade não se modifica a medida que os valores no ranking de amizade crescem. Para as classes 2 e 3, a similaridade decresce à medida que o ranking de amizade aumenta, enquanto que as classes 4 e 5 apontam para valores mais altos a medida que vai crescendo o ranking de amizade. Entretanto, tanto o decréscimo quanto o acréscimo do coeficiente de similaridade são suaves, indicando que a trajetória de um indivíduo é bastante semelhante a dos seus cinco melhores amigos.

Tais resultados convergem para aquilo que foi apresentado por (RIBEIRO, 2021) que observou nas mesmas redes que quanto melhor rankeado é um indivíduo no ranking de melhores amigos de outro indivíduo - tomando como parâmetro o número de ligações - mais similar é a trajetória de ambos.

Em relação as análises de trajetórias feitas na rede de São João, constatou-se também que a similaridade é mais alta para indivíduos da mesma classe, sendo que essa métrica se mantêm constante ao longo do ranking de amizade para as classes 2 e 4 e sofre uma leve queda para a classe de renda 1.

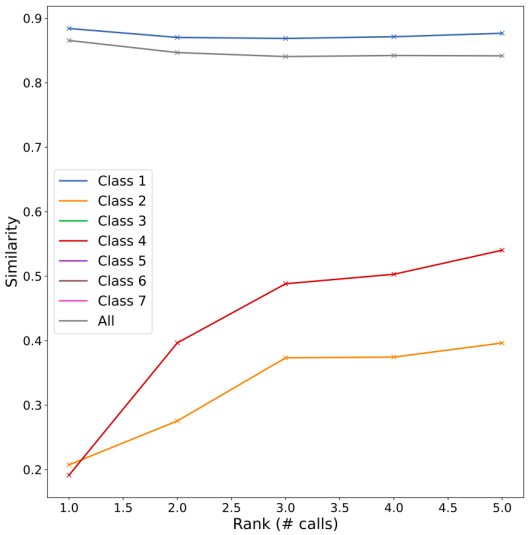


Figura 36 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 1

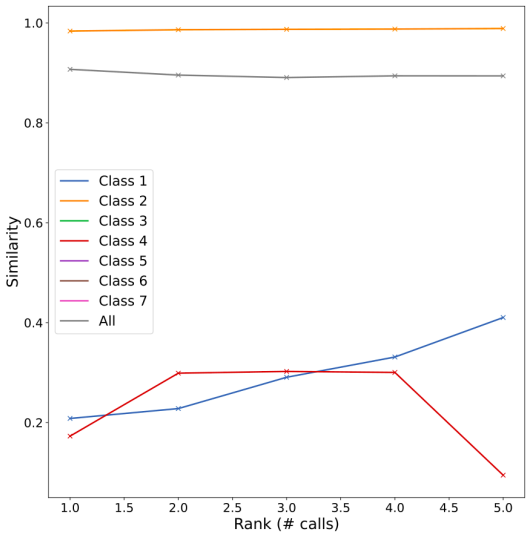


Figura 37 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 2

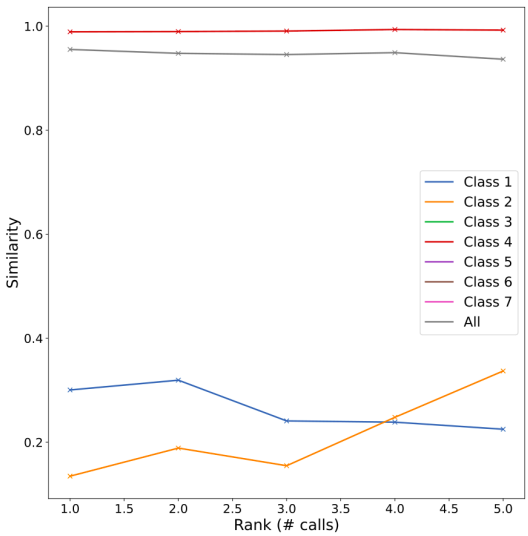


Figura 38 – Similaridade de trajetórias x Ranking de amigos - SJDR - Classe de renda 4

Esses resultados, somados às análises feitas sobre a correlação de trajetórias entre as classes de renda - apresentadas nos mapas de calor das figuras 39 e 40 - reforçam a ideia de que a condição econômica de fato influencia nas trajetórias diárias dos indivíduos, visto que os maiores valores sempre estão na diagonal principal.

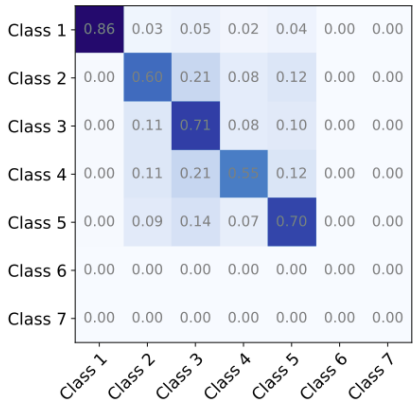


Figura 39 – Correlação de trajetórias entre as classes de renda - Juiz de Fora

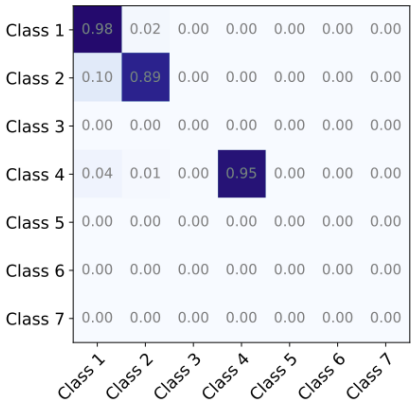


Figura 40 – Correlação de trajetórias entre as classes de renda - São João del-Rei

5 Conclusões e trabalhos futuros

Pode-se afirmar, a partir dos resultados obtidos, que a classe de renda é sim de fato um fator preponderante nas trajetórias diárias dos indivíduos nos centros urbanos. Além de impactar nas trajetórias, a classe de renda influencia nas relações entre os indivíduos, visto que a interação se da principalmente entre indivíduos com a mesma classificação. Além disso, foi confirmado que a maioria dos indivíduos mantêm relação constante com outros poucos. Também foi possível verificar que quanto mais próximos são dois indivíduos socialmente, mais semelhante são suas trajetórias.

Um ponto importante que deve ser destacado é que a quantidade de dados impacta diretamente na qualidade dos resultados, o que dificulta esse tipo de trabalho em regiões de pequeno e médio porte. Entretanto, a metodologia aqui aplicada pode formular um *framework* para análises de mobilidade urbana sob um viês econômico, sendo possível a aplicação em outras regiões. Dessa forma, além dos resultados obtidos, outro importante resultado desse trabalho é o incremento de formas de modelagem de aspectos urbanos a partir de dados.

Por fim, vale destacar que os resultados obtidos permitem identificar diferentes padrões de deslocamento dos indivíduos das classes econômicas, possibilitando a compreensão das desigualdades sociais e como elas afetam a trajetória diária das pessoas. Qualquer agente público que esteja determinado a implementar os objetivos de desenvolvimento da ONU, que, entre outros pontos, buscam a redução da desigualdade e a criação de cidades e comunidades sustentáveis, podem usar esses resultados para a promoção de politícas que visem mitigar/resolver problemas relacionados a mobilidade e distribuição de renda nas cidades.

Referências

RODRIGUES, F.; XAVIER, C. R.; EVSUKOFF, A. G.; ZIVIANI, A.; VIEIRA, V. da F. Análise de mobilidade urbana em diferentes estratos sociais a partir de uma rede de telefonia móvel. In: *Anais do II Workshop de Computação Urbana*. Porto Alegre, RS, Brasil: SBC, 2018. ISSN 2595-2706. Disponível em: <https://sol.sbc-.org.br/index.php/courb/article/view/233>. Citado 4 vezes nas páginas 6, 20, 25 e 26.

FARBER S., N. T. M. H. J. . L. X. The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. *Ann. Assoc. Am. Geogr*, Ann. Assoc. Am.Geogr, v. 1, n. 103, p. 483–504, 07 2013. Disponível em: <https://doi.org/10.1080/00045608.2012.68923>. Citado na página 10.

SILM S. & AHAS, R. The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset. *Social Science Research*, Social Science Research, v. 9, n. 0, p. 30–43, 03 2014. Disponível em: <https://www.sciencedirect.com/science/article-/pii/S0049089X1400071>. Citado na página 10.

BLONDEL, V. D.; DECUYPER, A.; KRINGS, G. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, v. 4, p. 1–55, 2015. Citado 2 vezes nas páginas 10 e 14.

GONZALEZ, M. C.; HIDALGO, C. A.; BARABASI, A.-L. Understanding individual human mobility patterns. *Nature*, Nature Publishing Group, v. 453, n. 7196, p. 779–782, June 2008. Citado 4 vezes nas páginas 10, 11, 15 e 18.

ALESSANDRETTI, L.; SAPIEZYNSKI, P.; LEHMANN, S.; BARONCHELLI, A. Multi-scale spatio-temporal analysis of human mobility. *PLOS ONE*, Public Library of Science, v. 12, n. 2, p. 1–17, 02 2017. Citado na página 10.

LENORMAND, M.; RAMASCO, J. J. Towards a better understanding of cities using mobility data. *Built Environment*, v. 42, p. 356–364(9), 10 2016. Citado na página 10.

SCHNEIDER, C. M.; BELIK, V.; COURONNÉ, T.; SMOREDA, Z.; GONZÁLEZ, M. C. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, v. 10, n. 84, jul. 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles-/PMC3673164>. Citado na página 11.

BOEING G., B. M. J. S. S. L. Urban analytics: History, trajectory, and critique. *Ann. Assoc. Am. Geogr*, Ann. Assoc. Am.Geogr, v. 1, n. 103, p. 483–504, 07 2021. Disponível em: <https://doi.org/10.1080/00045608.2012.68923>. Citado 2 vezes nas páginas 11 e 12.

REARDON S. F. & O'SULLIVAN, D. The scaling of income distribution in australia: Possible relationships between urban allometry, city size, and economic inequality. *Environ. Plan. B Urban Anal. City Sci.*, Environ. Plan. B Urban Anal. City Sci., v. 45, n. 11, p. 603–622, 01 2018. Disponível em: <https://doi.org/10-.1177%2F026581351667648>. Citado na página 11.

MASSEY D. S. & DENTON, N. A. The dimensions of residential segregation on jstor. *Soc. Forces*, Soc. Forces, v. 67, n. 0, p. 281–315, 01 1988. Disponível em: <https://doi.org/10.2307/257918>. Citado na página 12.

REARDON S. F., O. D. Measures of spatial segregation. *Sociol. Methodol*, Sociol. Methodol, v. 34, n. 0, p. 121–162, 01 2004. Disponível em: <https://doi.org/10.1111/j-.0081-1750.2004.00150.>. Citado na página 12.

WONG D. W. S. & SHAW, S. L. Measuring segregation: An activity space approach. *J. Geogr. Syst*, J. Geogr. Syst, v. 13, n. 0, p. 127–145, 06 2011. Disponível em: <https://doi.org/10.1007/s10109-010-0112->. Citado na página 12.

FARBER, S.; O'KELLY, M.; MILLER, H.; NEUTENS, T. Measuring segregation using patterns of daily travel behavior: A social interaction based model of exposure. *Journal of Transport Geography*, v. 49, p. 26–38, 12 2015. Citado na página 12.

BARBOSA, H.; BARTHELEMY, M.; GHOSHAL, G.; JAMES, C. R.; LENORMAND, M.; LOUAIL, T.; MENEZES, R.; RAMASCO, J. J.; SIMINI, F.; TOMASINI, M. Human mobility: Models and applications. *Physics Reports*, v. 734, p. 1–74, 2018. ISSN 0370-1573. Human mobility: Models and applications. Disponível em: <https://www-.sciencedirect.com/science/article/pii/S037015731830022>. Citado 5 vezes nas páginas 12, 15, 18, 19 e 20.

LOTERO-VéLEZ, L.; CARDILLO, A.; HURTADO, R.; GóMEZ-GARDEñES, J. t. In: _____. [S.l.: s.n.], 2016. p. 149–164. ISBN 978-3-319-23945-3. Citado 3 vezes nas páginas 12, 19 e 20.

VANHOOF, M.; REIS, F.; PLOETZ, T.; SMOREDA, Z. Assessing the quality of home detection from mobile phone data for official statistics. *Journal of Official Statistics*, Walter de Gruyter GmbH, v. 34, n. 4, p. 935–960, Dec 2018. Citado 3 vezes nas páginas 13, 19 e 24.

CHAVES, J. C. *Metodologia para Estimativa de Matrizes Origem-Destino a partir de Dados de Telefonia Celular*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, 2018. Citado 2 vezes nas páginas 14 e 19.

HOLLINGSHEAD A. B., Z. G. K. Timegeo:modeling urban mobility without travel surveys. Citado na página 14.

ITU. *ITU: Committed to connecting the world*. 2015. Disponível em: <http://www.itu.int/en/Pages/default.asp>. Citado na página 14.

YANG Y., J. S. V. D. e. a. Human behavior and the principle of least effort: An introduction to human ecology. *Proceedings of the National Academy of Sciences of the United States of AmericaISSN*, 2007. Citado na página 14.

BARBOSA, C. M. G. Sara: Um framework para análise de redes sociais online. *UFSJ*, UFSJ, v. 1, n. 1, p. 1–122, 2022. Citado 2 vezes nas páginas 16 e 17.

VIEIRA, V. d. F. DetecÇÃo de comunidades em redes complexas de larga escala. *UFRJ*, UFRJ, v. 1, n. 1, p. 1–126, 2013. Citado na página 17.

ONNELA, J.-P.; SARAMäKI, J.; HYVöNEN, J.; SZABó, G.; LAZER, D.; KASKI, K.; KERTéSZ, J.; BARABáSI, A.-L. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, p. 7332–6, 06 2007. Citado 2 vezes nas páginas 17 e 23.

BLONDEL, V. D.; GUILLAUME, J.-L.; LAMBIOTTE, R.; LEFEBVRE, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, IOP Publishing, v. 2008, n. 10, p. P10008, Oct 2008. ISSN 1742-5468. Citado 2 vezes nas páginas 17 e 23.

SHIN S. HONG, K. L. S. C. I. R. R. On the levy-walk nature of human mobility: Do humans walk like monkeys? *IEEE INFOCOM*, IEEE INFOCOM, v. 1, n. 1, p. 924–932, 2008. Citado na página 18.

LIANG J. ZHAO, L. D. K. X. X. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 2013. Citado na página 19.

NOULAS S. SCELLATO, R. L. M. P. C. M. A. A tale of many cities: universal patterns in human urban mobility. *PloS one*, IEEE INFOCOM, v. 7, n. 5, 2012. Citado na página 20.

HAWELKA I. SITKO, E. B. S. S. P. K. C. R. B. Geolocated twitter as proxy for global mobility patterns. *Cartography and Geo- graphic Information Science*, Cartography and Geo- graphic Information Science, v. 41, 2014. Citado na página 20.

PAPPALARDO, L.; SIMINI, F.; BARLACCHI, G.; PELLUNGRINI, R. scikit-mobility: a python library for the analysis, generation and risk assessment of mobility data. *Cornel University*, p. 5–12, 2019. Citado na página 20.

MOTTE B., A. A. B. O. . N. C. D. Commuting patterns in the metropolitan region of rio de janeiro. what differences between formal and informal jobs? *Journal of Transport Geography*, Journal of Transport Geography, 2016. Citado na página 21.

RODRIGUES, A. L.; GIANNOTTI, M.; BARBOZA, M. H. C. C.; ALVES, B. B. Measuring mobility inequalities of favela residents based on mobile phone data. *BMC infectious diseases*, p. 37–37, 2021. Citado 2 vezes nas páginas 21 e 44.

STEELE J. E., S. P. R. P. C. A. V. A. B. T. J. B. J. B. L. Mapping poverty using mobile phone and satellite data. *Cartography and Geo- graphic Information Science*, Journal of The Royal Society Interface, v. 14, 2017. Citado na página 21.

CARRON, P. M.; KASKI, K.; DUNBAR, R. Calling dunbar's numbers. *Social Networks*, Elsevier BV, v. 47, p. 151–155, Oct 2016. ISSN 0378-8733. Disponível em: <http://dx.doi.org/10.1016/j.socnet.2016.06.00>. Citado na página 23.

SHANNON, C. E. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, ACM New York, NY, USA, v. 5, n. 1, p. 3–55, 2001. Citado na página 26.

TOOLE, J. L.; HERRERA-YAQÜE, C.; SCHNEIDER, C. M.; GONZÁLEZ, M. C. Coupling human mobility and social ties. The Royal Society, v. 12, n. 105, p. 20141128, 2015. Citado 2 vezes nas páginas 27 e 28.

RIBEIRO, J. M.; ALENCAR, R.; MARTINS, G.; XAVIER, C.; EVSUKOFF, A.; VIEIRA, V. Caracterização da relação entre redes sociais e mobilidade de indivíduos em contextos urbanos. *BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING*, BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, v. 9, p. 32–43, 07 2021. Disponível em: <https://doi.org/10.5753/brasnam-.2021.1612>. Citado 2 vezes nas páginas 39 e 48.