

# Caracterização da relação entre redes sociais e mobilidade de indivíduos em contextos urbanos

José Mauro Ribeiro<sup>1</sup>, Ricardo Alencar<sup>2</sup>, Gustavo Martins<sup>1</sup>,  
Carolina Xavier<sup>1</sup>, Alexandre Evsukoff<sup>2</sup>, Vinícius Vieira<sup>1</sup>

<sup>1</sup>Dep. de Ciência da Computação - Universidade Federal de São João del-Rei (UFSJ)  
São João del-Rei – MG – Brasil

<sup>2</sup>COPPE – Universidade Federal do Rio de Janeiro  
Rio de Janeiro – RJ – Brasil

vinicius@ufsj.edu.br,

**Abstract.** *The understanding of the urban organization is fundamental for better planning and definition of public policies that aim the well-being of a population and the mitigation of social problems. Computational models able to integrate data from multiple sources can help the characterization of the complex geographic and socioeconomic organization of the population in urban areas with great resolution and applicability to practical contexts. This work presents the analysis of the urban complexity represented by the social and mobility relations by the investigation of Call Detail Records (CDR). Considering two cities with distinct characteristics, experiments performed by the characterization of the individual mobility of people allow us to observe a strong interdependency between their mobility behavior in the urban space and the behavior exhibited by their social networks.*

**Resumo.** *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 sociais. 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). Considerando duas cidades com características distintas, experimentos realizados a partir da caracterização da mobilidade individual 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.*

## 1. Introdução

O crescimento da população em áreas urbanas tem imposto um grande desafio para planejadores urbanos e responsáveis por tomadas de decisão de políticas públicas, o que pode trazer enormes desigualdades, problemas sociais e de saúde pública, caso aconteça sem um planejamento adequado. A compreensão da organização das cidades é, então, fundamental para um melhor planejamento urbano e o estabelecimento de políticas públicas

para que seja possível antecipar e mitigar problemas sociais na escala em que ocorrem em grandes cidades [Lenormand and Ramasco 2016, Sapiezynski et al. 2015].

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 et al. 2015, Gonzalez et al. 2008]. Nos últimos anos, novas tecnologias para armazenamento e processamento de grandes volumes de dados permitiram que a informação coletada em tempo real seja utilizada para o desenvolvimento de soluções inovadoras para cidades baseadas em dados, o que envolve algoritmos e metodologias que são, ainda, foco de pesquisa em muitas áreas [Alessandretti et al. 2017, Lenormand and Ramasco 2016]. Essas bases de dados, embora possam colocar uma série de questões relacionadas a ética e privacidade, constituem uma inegável fonte de informação para a compreensão de fenômenos espaciais de maneira sem precedentes.

A modelagem do sistema urbano considerado neste trabalho toma como base dados de 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. A complexidade urbana é representada sob duas perspectivas: social e de mobilidade. A perspectiva social permite que se compreenda como se dá a interação entre pares de indivíduos, revelando aspectos importantes específicos sobre as relações humanas em contextos urbanos. Sob a perspectiva de mobilidade é possível melhor compreender como os indivíduos interagem com o espaço urbano e o que pode auxiliar na caracterização da organização das cidades e dos indivíduos.

A organização humana em áreas urbanas pode ser investigada através da exploração de plataformas para geração e armazenamento de dados massivos com diferentes graus de resolução e originados de diferentes fontes. Por exemplo, *emails*, mensagens em redes sociais *online*, como Twitter e Facebook, e registros de ligações telefônicas podem evidenciar relações sociais [Blondel et al. 2015, Barbosa et al. 2018]. *Check-ins* por aplicativos de avaliação de serviços e sistemas de rastreamento revelam a forma como indivíduos transitam nas cidades [Sapiezynski et al. 2015]. Apesar de haver um grande número de trabalhos na literatura com o objetivo de capturar e modelar aspectos espaciais e temporais dos indivíduos no espaço urbano [Gonzalez et al. 2008, Barbosa et al. 2018], muitos deles ignoram aspectos ligados às relações sociais, que podem melhorar significativamente nossa compreensão sobre a forma como indivíduos utilizam o espaço urbano, como tem sido apontado por diversos autores [Grabowicz et al. 2014, Cornacchia et al. 2020, Toole et al. 2015]. De maneira ainda mais específica, o presente trabalho é o primeiro a realizar uma análise dessa natureza em cidades no Brasil e fornece uma importante fonte para a validação da natureza social e de mobilidade de indivíduos em espaços urbanos brasileiros, especialmente considerando modelos baseados em dados provenientes da comunicação entre pessoas.

Considerando a análise de duas regiões urbanas, uma de pequeno porte e uma de médio porte, sob perspectiva das relações sociais e de mobilidade, duas hipóteses, tratadas como Questões de Pesquisa (QP), são investigadas neste trabalho. Primeiramente, será investigado se há (e como se dá) correlação entre as relações sociais apresentadas pelos indivíduos e a maneira como utilizam o espaço urbano (QPI). Além disso, pretende-se

identificar se os padrões sociais e de mobilidade exibidos pelos indivíduos se assemelham aos padrões observados por outros autores em bases de dados similares em outras regiões urbanas (QPII). Caso isso seja verificado, pretende-se obter uma forma de validação dos dados CDR aqui utilizados para que possam ser aplicados em outros estudos, servindo como base para a realização de simulação de fenômenos de diversas naturezas nos sistemas urbanos.

O presente trabalho está organizado da seguinte forma. A Seção 2 apresenta algumas características gerais, um conjunto de passos de pré-processamento da base de dados utilizada, assim como a metodologia proposta. A Seção 3 apresenta os experimentos conduzidos para a investigação dos aspectos sociais dos indivíduos. Também são apresentados experimentos para relacionar aspectos sociais a aspectos ligados à mobilidade dos indivíduos no espaço urbano. Por fim, algumas conclusões e direções futuras do presente trabalho são apresentadas na Seção 4.

## **2. Dados CDR para análise social e de mobilidade**

### **2.1. Descrição dos dados CDR, pré-processamento e características gerais**

Os dados considerados neste trabalho compreendem registros de detalhes de chamadas telefônicas (*Call Detail Records* - CDR) realizados em duas regiões imediatas localizadas no estado de Minas Gerais, no Brasil: a região imediata de São João del-Rei (de porte pequeno) e a região imediata de Juiz de Fora (de porte médio). De acordo com o IBGE (Instituto Brasileiro de Geografia e Estatística), 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. A região geográfica imediata de São João del-Rei (aquí chamada de SJDR) é composta de 14 municípios (Conceição da Barra de Minas, Coronel Xavier Chaves, Lagoa Dourada, Madre de Deus de Minas, Nazareno, Piedade do Rio Grande, Prados, Resende Costa, Ritópolis, Santa Cruz de Minas, São João del-Rei, São Tiago, São Vicente de Minas, Tiradentes) e estima-se que possua uma população de aproximadamente 184.210 habitantes, sendo 90.497 no município base de São João del-Rei. A região geográfica de Juiz de Fora (aquí chamada de JF) é composta de 29 municípios (Andrelândia, Aracitaba, Arantina, Belmiro Braga, Bias Fortes, Bocaina de Minas, Bom Jardim de Minas, Chácara, Chiador, Coronel Pacheco, Ewbank da Câmara, Goianá, Juiz de Fora, Liberdade, Lima Duarte, Matias Barbosa, Olaria, Oliveira Fortes, Paiva, Passa Vinte, Pedro Teixeira, Piau, Rio Novo, Rio Preto, Santa Bárbara do Monte Verde, Santa Rita de Jacutinga, Santana do Deserto, Santos Dumont, Simão Pereira), dos quais estima-se uma população de 749.569 habitantes, sendo 573.285 no município base de Juiz de Fora.

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.

Os dados considerados consistem em 9.293.317 registros de ligações telefônicas na região de São João del-Rei e 63.588.996 registros na região de Juiz de Fora, observados

durante um período de 30 dias. Com base nesses registros, de maneira imediata, é possível construir redes de comunicação direcionadas representando os usuários como nós e as ligações telefônicas como arestas. Cada aresta  $(v_i, v_j)$  indica que o indivíduo representado pelo vértice  $v_i$  realizou uma chamada para o indivíduo representado pelo vértice  $v_j$  e armazena duas informações: a quantidade e o tempo total das chamadas de  $v_i$  para  $v_j$ .

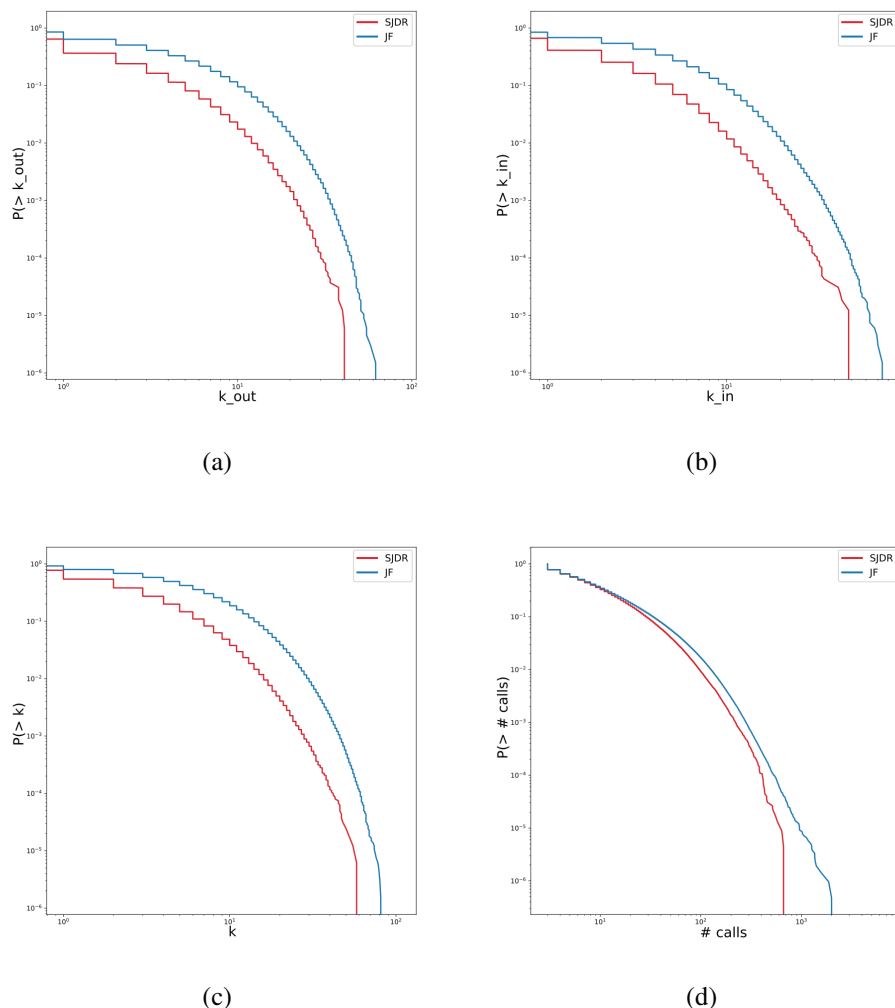
Diversos trabalhos podem ser encontrados na literatura com o objetivo de obter uma rede social a partir de uma rede de comunicação construída através de CDRs [Onnela et al. 2007, Blondel et al. 2015] e uma das maiores preocupações nesse sentido é evitar que a rede social represente relações que claramente não reflitam interações entre pessoas. Assim, tomando como base os resultados dos trabalhos de Robin Dunbar [Mac Carron et al. 2016], mais especificamente o conhecido número de Dunbar, eliminamos da rede vértices que mantinham mais do que 150 relações sociais, que poderiam indicar centrais de atendimento e números com ramais. Além disso, eliminamos da rede arestas entre usuários com menos do que três ligações telefônicas e arestas entre usuários com duração total de chamada inferior a 30 segundos, com o objetivo de reduzir o número de registros de ligações telefônicas realizadas por engano ou interações esporádicas. Considerando  $n$  como o número de nós e  $m$  como o número de arestas, os tamanhos das redes direcionadas foram  $n = 1.303.294/m = 689.952$  para a rede SJDR e  $n = 3.080.014/m = 4.395.760$  para a rede JF, após essas filtrações.

Um aspecto particular observado na base de dados CDR considerada neste trabalho é o fato de há o registro da localização apenas da antena que realizou a chamada, mas não da antena que recebeu a chamada. Embora isso não seja um grande problema para a caracterização das relações sociais entre os usuários, é impossível analisar qualquer característica dos indivíduos em relação à utilização do espaço, caso não tenham realizado chamadas, o que impacta diretamente na investigação de padrões de mobilidade proposta neste trabalho. Sendo assim, foram também eliminados da rede direcionada vértices que apresentavam grau de saída nulo, ou seja, não realizaram nenhuma ligação. Após essa filtração, as redes direcionadas resultantes possuem  $n = 161.948/m = 297.427$ , para a região de SJDR e  $n = 662.212/m = 2.770.314$ , para a região de JF.

Apesar de, em um primeiro momento, a comunicação entre indivíduos em uma rede de telefonia se dar de maneira direcionada, já que um indivíduo realiza a ligação e outro recebe, neste trabalho consideramos que, em diversas situações, as relações sociais, mesmo provenientes de chamadas telefônicas, podem ser consideradas não-direcionadas. Por isso, obtivemos também uma versão não-direcionada da rede. Assim, as arestas da rede não direcionada armazenam o número de ligações entre os indivíduos e o tempo total das chamadas realizadas por eles, não importando quem realizou e que recebeu as chamadas. A Figura 1 apresenta a distribuição de graus para as versões direcionada e não-direcionada da rede e a distribuição do número de chamadas entre indivíduos.

É possível perceber que as distribuições de graus são semelhantes a distribuições de graus tipicamente observadas em outras redes sociais relatadas na literatura [Newman and Park 2003]. Também é importante notar que há uma queda acentuada nas probabilidades observadas para os graus e para o número de ligações maiores que 10, indicando que apenas uma pequena quantidade de indivíduos apresenta atividade mais intensa, o que é particularmente relevante na discussão sobre a relação entre características sociais e de mobilidade conduzidas neste trabalho, mais especificamente na Seção 3.

**Figura 1. Distribuição de quantidades. (a) Grau de saída (rede direcionada). (b) Grau de entrada (rede direcionada). (c) Grau (rede não direcionada). (d) Número de ligações (rede não direcionada).**



Nas próximas seções, serão discutidos aspectos metodológicos que permitem uma investigação para a caracterização das relações de maneira mais específica no contexto deste trabalho, tanto do ponto de vista social, como de mobilidade.

## 2.2. Metodologia

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  $L$ , definida por Toole *et al.* [Toole et al. 2015] como uma matriz  $48 \times L$ , onde  $L$  é o número de antenas de localização única no espaço observado e cada linha de  $L$  corresponde a uma hora  $t$  de seu dia de semana típico e seu fim de semana típico. Cada elemento  $L_{t,j}$  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 et al. 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 et al. 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

$$\cos\theta_{i,j} = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{|\mathbf{v}_i||\mathbf{v}_j|}. \quad (1)$$

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.

Particularmente, neste trabalho, estamos interessados em investigar a relação entre as relações sociais e os padrões de mobilidade. Assim, os experimentos realizados visam explorar o impacto de indivíduos que possuem relação social relevante nos padrões de mobilidade de outros indivíduos. Por isso, a rede social será fundamental para a sistematização da escolha dos indivíduos a serem comparados.

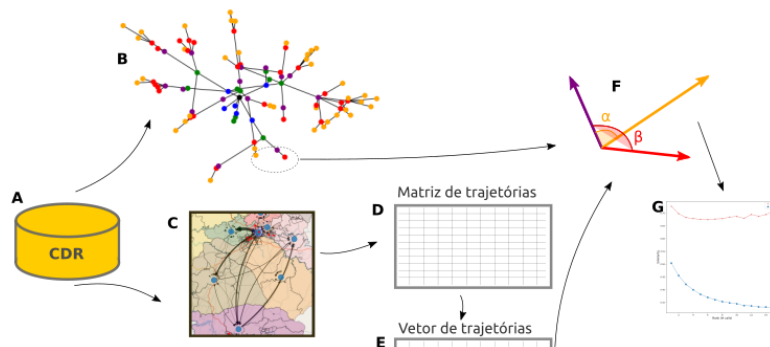
A Seção 3 apresenta os resultados dos experimentos realizados neste trabalho, considerando a metodologia aqui descrita e apresentada de maneira geral na Figura 2. Primeiramente, é realizado um pré-processamento dos dados CDR (Figura 2A), definindo as relações sociais, modeladas pela rede social (Figura 2B). São também identificadas as localizações visitadas pelos usuários no espaço urbano (Figura 2C), que são representadas pelas matrizes de localização (Figura 2D), para que, após, os vetores de trajetórias sejam calculados (Figura 2E). Tomando como base os vetores de trajetórias e as interações sociais a serem investigadas, são calculadas as similaridades entre os indivíduos (Figura 2F), para que os experimentos possam ser realizados (Figura 2G).

### 3. Experimentos e discussão

#### 3.1. Análise da rede social

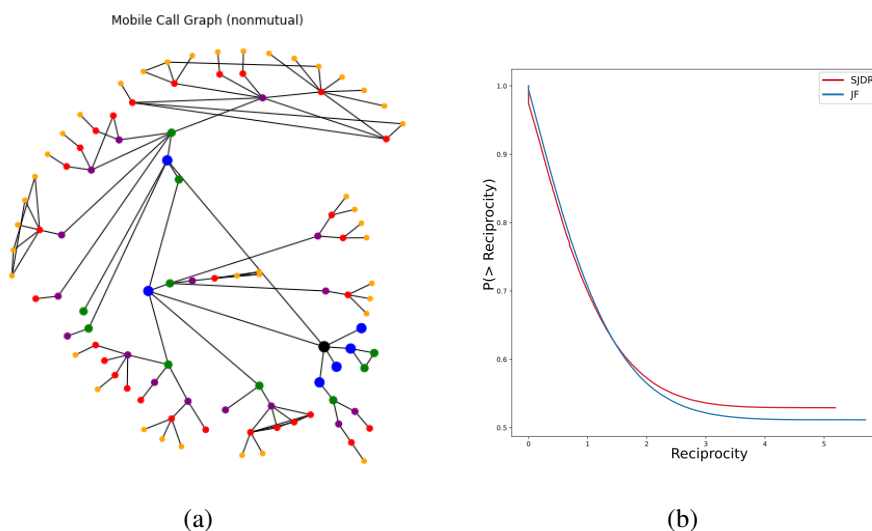
Uma das características mais frequentemente investigadas em redes é a organização de seus vértices em comunidades, isto é, grupos com uma alta densidade interna e uma baixa densidade externa. Considerando o contexto explorado neste trabalho, a estrutura de comunidades pode revelar como os contatos sociais estão distribuídos, o que poderia revelar aspectos relacionados à organização espacial e administrativa da cidade e, integrada a outras fontes de dados, poderia auxiliar a compreensão de aspectos sociológicos e culturais

**Figura 2. Visão geral da metodologia. (A) Base de dados CDR. (B) Rede social. (C) Mobilidade. (D) Matriz de localização. (E) Vetor de trajetórias. (F) Similaridade de cosseno. (G) Experimento.**



da população. Entretanto, a estrutura das redes sociais obtidas a partir de ligações telefônicas apresenta características que dificultam a identificação das comunidades, como discutido por Tibely *et al.* [Tibély et al. 2011] e Blondel *et al.* [Blondel et al. 2015]. Conforme apontado por Onnela *et al.* [Onnela et al. 2007], redes sociais baseadas em ligações telefônicas apresentam um aspecto parecido com uma árvore, o que pode fazer com que algoritmos clássicos para detecção de comunidades produzam resultado insatisfatório 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.

**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.**



O método de Louvain [Blondel et al. 2008] foi aplicado às redes SJDR e JF e, após o cálculo da modularidade  $Q$  de Newman [Newman and Park 2003] considerando as estruturas de comunidades obtidas, foi possível obter partições aparentemente modu-

lares, com  $Q = 0,8437$  para a rede SJDR e  $Q = 0,6014$  para a rede JF. Entretanto, as partições apresentaram um número de comunidades extremamente alto (42704 e 58293 para SJDR e JF, respectivamente) e as comunidades resultantes não são muito significativas. Conforme discutido por Blondel *et al.* [Blondel et al. 2015], tais resultados levantam dúvidas sobre a utilidade de métodos para detecção de comunidades em redes de telefonia e vão ao encontro de uma profunda discussão sobre o que é, de fato, uma comunidade e como elas podem ser melhor estudadas, em alternativa aos métodos tradicionalmente utilizados [Vieira et al. 2020, Cherifi et al. 2019].

Ainda com o objetivo de investigar o comportamento de usuários com base no comportamento de suas ligações telefônicas, é possível quantificar a reciprocidade das relações. Tradicionalmente, a reciprocidade é utilizada como uma forma de quantificar as arestas do tipo  $(v_j, v_i)$  que ocorrem em razão de arestas  $(v_i, v_j)$  em um grafo direcionado. Considerando a medida de reciprocidade tradicional, foram observados valores de 0,4708 e 0,4885 para as rede SJDR e JF, respectivamente, para a versão da rede direcionada sem vértices com grau de saída nulo. Por outro lado, alguns autores discutem a caracterização da reciprocidade de relações de uma maneira mais refinada [Wang et al. 2011, Kovanen et al. 2010]. Wang *et al.* [Wang et al. 2011] propõem uma medida de reciprocidade para avaliar o desbalanceamento no relacionamento entre dois usuários:

$$R_{ij} = |\ln(p_{ij}) - \ln(p_{ji})|, \quad (2)$$

onde  $p_{ij}$  é a probabilidade de ligação de um  $v_i$  para um  $v_j$ . A Figura 3(b) apresenta a distribuição da reciprocidade dos usuários nas redes SJDR e JF, onde valores mais baixos indicam relações mais balanceadas. É possível perceber que a maior parte das relações apresenta reciprocidade moderada, enquanto poucas relações são extremamente balanceadas ou desbalanceadas. Esses resultados, quando comparados com os apresentados em [Wang et al. 2011], em que os autores também realizam uma análise de características sociais em uma base de telefonia celular, revelam uma forte semelhança entre as redes sociais investigadas neste trabalho e as redes investigadas por outros autores. Apesar de uma discussão mais aprofundada sobre as causas e o impacto da reciprocidade na rede social estar além do escopo do estudo, os resultados obtidos em relação à reciprocidade são bastante importantes, representam mais uma forma de validar a metodologia aqui conduzida para a construção de uma rede social a partir de dados de telefonia celular.

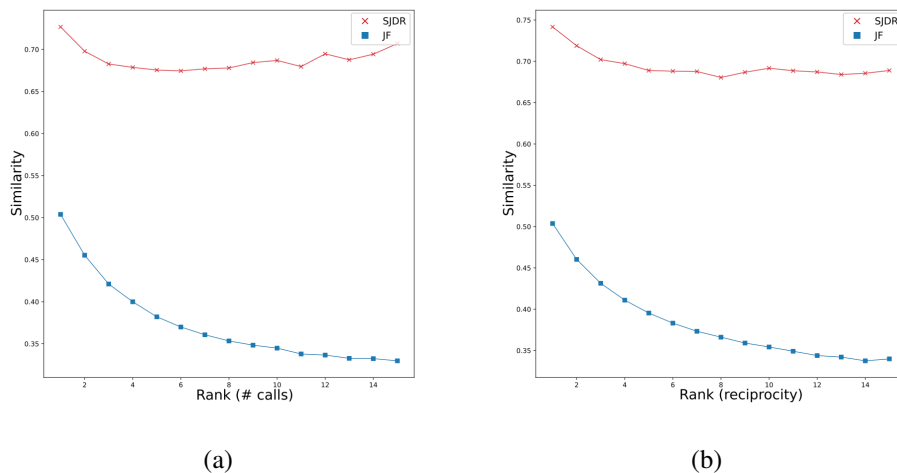
### 3.2. Análise de aspectos sociais e mobilidade

A partir da definição de vetores de trajetórias entre quaisquer pares de indivíduos, é possível calcular a similaridade entre seus padrões de mobilidade, tomando como base a metodologia descrita na Seção 2.2. Nesta seção, são apresentados os resultados de uma série de experimentos que têm como objetivo mostrar que aspectos relacionados à mobilidade dos indivíduos estão fortemente correlacionados às suas interações sociais. Para isso, as relações sociais são exploradas sob diferentes pontos de vista.

Primeiramente, investigamos como indivíduos que possuem relação direta com um outro afetam sua mobilidade, o que é apresentado na Figura 4, sob duas perspectivas: números de ligações (Figura 4(a)) e reciprocidade (Figura 4(b)). Tomando como base cada indivíduo  $v_i$  de uma rede, identificamos todos os seus vizinhos  $v_j$  e elaboramos um *rank* de amizade (com base no número de ligações e na reciprocidade) e calculamos as similaridades entre os vetores de trajetória  $t_i$  e  $t_j$ .



**Figura 4. Similaridade da mobilidade X rank dos contatos sociais. (a) Número de ligações. (b) Reciprocidade.**



Como é possível observar pela Figura 4, os indivíduos tendem a apresentar uma maior similaridade de trajetória com seus contatos mais altos no *rank* e essa similaridade é reduzida a medida em que a força das relações cai, tanto considerando ao número de ligações entre indivíduos, como em relação às reciprocidades das relações, o que permite que possamos caminhar em direção a uma resposta para a QPI, apresentada na Seção 1. Na rede JF, esse comportamento é preservado até a posição 15 do *rank*, enquanto na rede SJDR isso é observado apenas nas primeiras posições do *rank*. Uma possível explicação pode ser obtida pela observação da Figura 1, que mostra que, talvez pelo menor volume de dados, os vértices da rede SJDR apresentam graus sensivelmente menores que os vértices da rede JF. Assim, é interessante observar, combinando os resultados das Figura 1 e 4, a enorme queda no número de contatos sociais, e o seu impacto na definição de padrões de mobilidade, em torno de 15. Esse número corrobora as descobertas de Dunbar [Mac Carron et al. 2016], mais especificamente no que diz respeito ao que o autor afirma que, tipicamente, as pessoas são capazes de manter contato com alguma regularidade com apenas outras 15. Assim, como há apenas poucos vértices com grau alto, qualquer comportamento desses indivíduos acaba por enviesar a análise de maneira desproporcional, o que é conhecido como “viés do sobrevivente” (“*survivor bias*”) [Mehrabi et al. 2019].

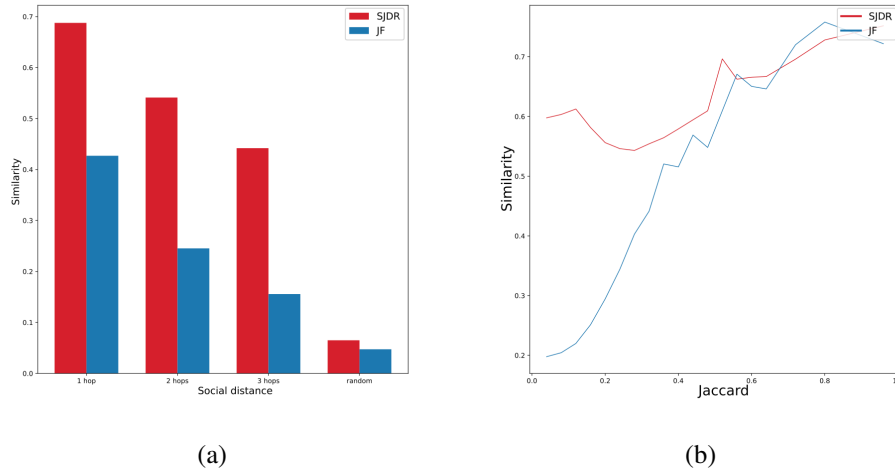
Foram também realizados experimentos para avaliar se, além do comportamento dos usuários com os quais um determinado indivíduo mantém contato direto, isto é, seus vizinhos na rede, a topologia da rede social tem um impacto significativo sobre a mobilidade dos indivíduos. Ou seja, investigamos se padrões de mobilidade de indivíduos  $v_j$  com os quais um vértice  $v_i$  sequer mantém relações sociais afetam o comportamento de  $v_i$ , o que é apresentado na Figura 5. Para isso, duas ideias foram exploradas. A Figura 5(a) apresenta uma investigação da variação da similaridade entre indivíduos em razão de sua distância na rede social, considerando 1 salto (seus vizinhos), 2 saltos e 3 saltos, além de sua similaridade com indivíduos aleatórios. Além disso, investigamos como a similaridade entre dois indivíduos  $v_i$  e  $v_j$  é afetada pelo número de vizinhos que  $v_i$  e  $v_j$  têm em

comum (Figura 5(b)). Para isso, utilizamos o coeficiente de Jaccard:

$$J_{i,j} = \frac{|F_i \cap F_j|}{|F_i \cup F_j|}, \quad (3)$$

onde  $F_i$  é o conjunto dos vizinhos do vértice  $v_i$ .

**Figura 5. Similaridade da mobilidade X similaridade na rede social. (a) Número de saltos. (b) Coeficiente de Jaccard.**



Primeiramente, é possível verificar pela Figura 5 que a mobilidade de um indivíduo  $v_i$  pode ser afetada por outro indivíduo  $v_j$ , mesmo que eles não mantenham relações na rede social e, assim, é possível ter a noção da importância da topologia da rede social na definição de comportamentos em espaços urbanos. Outro ponto relevante, observado na Figura 5(a), é que mesmo com três saltos de separação, a similaridade entre indivíduos ainda é significativamente mais alta do que entre dois indivíduos aleatoriamente selecionados. Além disso, as curvas apresentadas para os coeficientes de Jaccard na Figura 5(b) mesmo não apresentando comportamento monotônico, indicam tendência crescente, corroborando as observações anteriores de que a similaridade social está positivamente correlacionada à similaridade de mobilidade. Assim, podemos avançar ainda mais em uma resposta para a QPI. É importante observar que os resultados dos experimentos conduzidos neste trabalho são coerentes com resultados obtidos por outros autores que também investigam a relação entre a interação social e padrões de mobilidade [Toole et al. 2015, Grabowicz et al. 2014]. Mais especificamente, é interessante comparar os resultados obtidos neste trabalho com os resultados descritos em Toole *et al.* [Toole et al. 2015]. Lá, os autores investigam três regiões, das quais duas se assemelham muito com o comportamento exibido para a região de JF deste trabalho e uma se assemelha ao comportamento exibido para a região de SJDR. Em [Toole et al. 2015], os autores argumentam que a região que apresentou comportamento diferente é a única para a qual os dados CDR não possuem geolocalização dos receptores de chamadas telefônicas e que isso justificaria a distinção de comportamento. Se isso fosse verdade, era esperado que ambas regiões exploradas neste trabalho (SJDR e JF) apresentassem comportamento

similar, já que nenhuma delas possui tal informação para os recebedores. Assim, outras hipóteses precisam ser investigadas para avaliar tal diferença nos experimentos.

Na tentativa de compreender as diferenças no comportamento dos usuários de diferentes regiões, pode-se levantar como hipótese o baixo volume de dados para a região de SJDR, que faz com que o “viés do sobrevivente” tenha um impacto muito mais significativo. Ainda nesse sentido, é impossível ignorar o fato que, em todos os experimentos realizados, a similaridade observada para os usuários de telefonia na região de SJDR é muito maior do que para os usuários de JF. Uma das justificativas para isso pode vir do fato de que o centro urbano da região de SJDR possui uma extensão muito menor do que o centro urbano da região de JF, fazendo com que todos os indivíduos, de fato, tendam a apresentar padrões mais similares de mobilidade, ou seja, uma inadequação na granularidade dos dados. Outra justificativa pode vir do fato de o centro urbano de SJDR possuir uma menor complexidade socioeconômica em relação ao centro urbano de JF, reduzindo o impacto das relações sociais nos padrões de mobilidade. De qualquer forma, o presente trabalho levanta uma série de questões que precisam ser exploradas mais profundamente em trabalhos futuros. De qualquer forma, é possível afirmar que os resultados obtidos neste trabalho são coerentes com outros resultados encontrados na literatura e que as relações sociais e de mobilidade observadas nos experimentos realizados vão ao encontro de resultados observados por outros autores. Assim, além de permitir uma mais clara compreensão dos dados de telefonia armazenados no CDR e da forma como as pessoas se relacionam e utilizam o espaço urbano, os resultados obtidos neste trabalho permitem que seja respondida a QP11, sendo base para uma validação dos dados em outras aplicações.

#### **4. Conclusões e direções futuras**

Este trabalho apresenta uma investigação da relação entre redes sociais e aspectos de mobilidade de indivíduos em espaços urbanos, tomando como base registros de ligações telefônicas (*Call Detail Records* - CDR). Experimentos realizados considerando duas cidades permitem caracterizar aspectos sociais relevantes dos indivíduos nas cidades e revelam uma forte interdependência entre seus padrões de utilização do espaço urbano e os padrões exibidos pelos seus contatos sociais. Mais ainda, é possível verificar uma forte similaridade entre os padrões de mobilidade de um indivíduo e os padrões de mobilidade dos indivíduos de sua rede social, mesmo que eles não estabeleçam um contato direto. É importante enfatizar que esse é o primeiro trabalho dessa natureza realizado em cidades no Brasil e que os resultados encontrados são bastante coerentes com os resultados encontrados nos trabalhos de outros autores realizados em espaços urbanos em outras regiões do mundo e permitem um melhor entendimento sobre as cidades aqui investigadas, mas também fornecem uma forma de validação dos dados CDR em outras aplicações relacionadas a redes sociais e mobilidade urbana. Uma série de questões, principalmente relacionadas ao impacto de características socioeconômicas dos indivíduos sobre sua mobilidade ainda precisam ser investigadas em trabalhos futuros. Além disso, pretende-se aplicar a metodologia apresentada a outras cidades, com o objetivo de generalizar os resultados encontrados.

#### **Referências**

Alessandretti, L., Sapiezynski, P., Lehmann, S., and Baronchelli, A. (2017). Multi-scale spatio-temporal analysis of human mobility. *PLOS ONE*, 12(2):1–17.

- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74.
- Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4:1–55.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Stat. Mech.*, 2008(10):P10008.
- Cherifi, H., Palla, G., Szymanski, B. K., and Lu, X. (2019). On community structure in complex networks: challenges and opportunities. *Applied Network Science*, 4(117).
- Cornacchia, G., Rossetti, G., and Pappalardo, L. (2020). Modelling human mobility considering spatial, temporal and social dimensions. *arXiv:2007.02371*.
- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.
- Grabowicz, P. A., Ramasco, J. J., Gonçalves, B., and Eguíluz, V. M. (2014). Entangling mobility and interactions in social media. *PLOS ONE*, 9(3):1–12.
- Kovanen, L., Saramaki, J., and Kaski, K. (2010). Reciprocity of mobile phone calls. *arXiv: Physics and Society*.
- Lenormand, M. and Ramasco, J. J. (2016). Towards a better understanding of cities using mobility data. *Built Environment*, 42:356–364(9).
- Mac Carron, P., Kaski, K., and Dunbar, R. (2016). Calling dunbar’s numbers. *Social Networks*, 47:151–155.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2019). A survey on bias and fairness in machine learning. *CoRR*, abs/1908.09635.
- Newman, M. and Park, J. (2003). Why social networks are different than other types of networks. *Physical review. E*, 68:036122.
- Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *PNAS*, 104:7332–6.
- Sapiezynski, P., Stopczynski, A., Gatej, R., and Lehmann, S. (2015). Tracking human mobility using wifi signals. *PLOS ONE*, 10(7):1–11.
- Tibély, G., Kovanen, L., Karsai, M., Kaski, K., Kertész, J., and Saramäki, J. (2011). Communities and beyond: Mesoscopic analysis of a large social network with complementary methods. *Physical Review E*, 83(5).
- Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., and González, M. C. (2015). Coupling human mobility and social ties. 12(105):20141128.
- Vieira, V. F., Xavier, C. R., and Evsukoff, A. G. (2020). A comparative study of overlapping community detection methods from the perspective of the structural properties. *Appl. Netw. Sci.*, 5(1):51.
- Wang, C., Strathman, A., Lizardo, O., Hachen, D., Toroczkai, Z., and Chawla, N. (2011). Weighted reciprocity in human communication networks.