

Classificação de indivíduos segundo comportamento individual relativo a viagens a partir de dados em painel obtidos por *smartphones*

Lucas Assirati¹, Cira Souza Pitombo²

¹Universidade de São Paulo, Escola de Engenharia de São Carlos, assirati@usp.br

²Universidade de São Paulo, Escola de Engenharia de São Carlos, cirapitombo@usp.br

Recebido:

9 de maio de 2018

Aceito para publicação:

29 de março de 2019

Publicado:

31 de agosto de 2019

Editor de área:

Bruno Vieira Bertoncini

Palavras-chaves:

Comportamento individual relativo a viagens,
Coleta de dados passiva,
Smartphones,
Análise de *Cluster*,
K-médias.

Keywords:

Individual travel behavior,
Passive data collection,
Smartphones,
Cluster analysis,
K-means.

DOI:10.14295/transportes.v27i2.1679



RESUMO

A caracterização comportamental relativa a viagens é uma questão importante nas análises baseadas em atividades e, comumente, é a variável dependente nos modelos de estimativa de demanda por transportes. A classificação individual, segundo comportamentos relacionados aos deslocamentos, pode ser realizada com dados seccionais, considerando diferentes fatores como distâncias, modos utilizados e atividades realizadas, ou com dados em painel, através de valores médios nos múltiplos dias ou atividades frequentes, por exemplo. Dados em painel constituem importante ferramenta em análises comportamentais relativas às viagens urbanas, propiciando dimensão analítica extra relativo à heterogeneidade temporal individual. Todavia, a obtenção desses dados não é trivial, demandando recursos monetários e de tempo. Assim, o objetivo principal deste trabalho é classificar indivíduos segundo comportamento relativo a viagens a partir de dados em painel. O objetivo secundário associa-se à obtenção do painel através de smartphones. A potencialidade da proposta é validada por um estudo de caso contemplando estudantes universitários em São Carlos – SP, Brasil. Mediante dados fornecidos pelos estudantes, utilizou-se o algoritmo k-médias considerando quatro variáveis associadas às viagens realizadas em três dias úteis consecutivos. Obtiveram-se três grupos comportamentais distintos com diferenças quanto ao grau de motorização, recorrência de localidades, número de viagens realizadas e distâncias médias percorridas.

ABSTRACT

The characterization of travel behavior is a major issue in the activity-based travel analysis and, generally, is the response variable on travel demand modeling. The individual travel behavior classification can be realized using sectional data such as trip distances, travel mode choice or performed activities. Also, can be done using panel data, such as average values for multiple days or recurrent activities. Panel data are an important tool in behavioral analysis related to urban trips, providing extra dimension of analysis related to the individual temporal heterogeneity. However, obtaining these data is not trivial, requiring monetary and time resources. Thus, the main goal of this study is to classify individuals according to travel behavior from panel data. The secondary goal is related with panel data collection through smartphones. The potential of the study is validated by a case study with undergraduate and PhD students from São Carlos - SP, Brazil. With data voluntarily provided by the students, a k-means algorithm was employed considering, as input, four variables associated with trips carried out in three consecutive working days. Three different behavioral groups were obtained with differences concerning degree of motorization, recurrence of localities, number of trips performed, and average distances traveled.

1. INTRODUÇÃO E BACKGROUND

Diariamente, pessoas realizam múltiplos padrões de deslocamentos visando cumprir suas atividades dentro de um ambiente urbano. Na área de modelagem de demanda por transportes,

muitas pesquisas já se dedicaram ao estudo de padrões comportamentais de viagens de indivíduos, sendo caracterizados de diversas maneiras. A caracterização de tais deslocamentos determina o comportamento individual relativo a viagens. O comportamento relativo a viagens vem do termo “*travel behavior*” e caracteriza aspectos das escolhas dos indivíduos, associadas às viagens. A importância da sua caracterização adequada está estritamente relacionada à busca do seu entendimento e das variáveis que influenciam tais escolhas. Em geral, o comportamento individual, associado às viagens urbanas, é a variável dependente, incorporada aos modelos de previsão de demanda por viagens, baseadas em atividades. Diversos estudos buscam estruturar e entender esse comportamento a partir de diferentes modelos e ferramentas (Hägerstrand, 1970; Hensher, 2001; Ortuzar e Willumsen, 2001).

Para caracterização do comportamento relativo a viagens, destacam-se os estudos utilizando um único dia de coletas de dados, múltiplos dias de coleta de dados e composição de sequência de deslocamentos.

Kockelman (1997), considerando entrevistas relativas a um único dia útil, caracterizou o comportamento relativo a viagens pela distância percorrida, propósito da viagem (trabalho ou outros) e modo utilizado (veículo motorizado ou não motorizado). Posteriormente, tal comportamento foi definido através de modelos lineares e *logit* e variáveis explicativas, tais como densidade populacional, acessibilidade e índices entrópicos para uso do solo.

Daisy *et al.* (2018) conduziram uma pesquisa em uma grande universidade canadense considerando funcionários e alunos. Aplicando um questionário *on-line*, elaboraram diários de atividades para um dia de deslocamentos e analisaram frequências de viagens realizadas por automóvel ou por transporte ativo (caminhadas e bicicleta), através de um modelo binomial. Neste caso, o comportamento de viagens foi caracterizado através de uma variável numérica que mensurava frequência de viagens por modo de transporte.

Ao longo dos anos, o comportamento associado a viagens de um único dia também foi mensurado em termos de cadeias ou sequências de viagens. Arentze e Timmermans (2005), propuseram um modelo onde cadeias de viagens são formadas, baseadas em valores de atratividade em um mapa simulado, sendo a descrição das viagens fundamentada na quantificação dessas cadeias visando determinar o propósito para o deslocamento (atividades de trabalho, sociais, compras, lazer e recreação). Ramadurai e Ukkusuri (2010), também atuando em um modelo aplicado a um regime simulado, representavam cadeias de atividades considerando deslocamentos a partir do domicílio para locais de trabalho ou outro propósito.

Pitombo *et al.* (2011), utilizando uma pesquisa Origem/Destino relativa a um dia de deslocamentos, definem padrões comportamentais através do estabelecimento de cadeias de deslocamentos usando codificações para o modo de deslocamento (privado, coletivo ou não-motorizado), motivação para o deslocamento (residência, trabalho, estudo, outras atividades) e a distância dos deslocamentos. As sequências são agrupadas em padrões distintos por meio de uma árvore de decisão.

No entanto, a utilização de apenas um único dia de dados não permite aprofundar os estudos quanto a eventuais variabilidades existentes no mesmo indivíduo. Assim, graças ao advento de tecnologias para coleta de dados mais modernas, novos trabalhos passam a utilizar múltiplos dias nas suas análises.

Huff e Hanson (1986), atuando em uma cidade sueca, constituíram uma base de dados que consistia de um conjunto de informações de dias consecutivos, coletados por um período maior

que um mês. Através desses dados, valendo-se da avaliação acerca das repetições das ações realizadas pelas pessoas, os autores conseguiram dar interpretações com respeito à variabilidade intrapessoal dos indivíduos, no período considerado no estudo, corroborando o estudo prévio (Hanson e Huff, 1981) que propunha que os padrões comportamentais poderiam ser caracterizados por repetições ou por variabilidade.

Visando à compreensão dos deslocamentos com motivação para trabalho, Pas (1987) compôs uma base de dados com indivíduos que realizavam atividades remuneradas, considerando um diário de atividades de cinco dias úteis. Computando o número de viagens diárias por indivíduos, para cada um dos cinco dias, e aplicando um “índice de variância da taxa de geração de viagens”, os autores mensuraram quais elementos apresentavam maiores ou menores variações de geração de viagens de um dia para o outro.

Susilo e Kitamura (2005), valendo-se de uma base de dados de 6 semanas, constituída em duas cidades alemãs (Axhausen *et al.*, 2002), caracterizaram o espaço dos locais de realização de atividades dos indivíduos por meio de um “índice de variabilidade do espaço de atividades” e concluíram que a dimensão do espaço de atividades é influenciada pela motivação do deslocamento. Atividades obrigatórias como alimentação, estudo e trabalho confinam o espaço de realização de atividades, enquanto atividades não obrigatórias como lazer, aos finais de semana, estendem o espaço de realização de atividades.

Ainda utilizando a mesma base de dados, Susilo e Axhausen (2014) avaliaram o grau de repetição das escolhas individuais para realização de atividades por meio do “índice Herfindahl-Hirschman”. Esse índice é herdado da Economia e sua aplicação em transportes revela os indivíduos que realizam escolhas tipicamente recorrentes no intervalo de 6 semanas de análise, frente indivíduos que realizam escolhas atípicas no mesmo intervalo de tempo.

O leitor que se interessar por maior formalismo matemático acerca das métricas “índice de variância da taxa de geração de viagens”, “índice de variabilidade do espaço de atividades” e “índice Herfindahl-Hirschman” deve consultar o trabalho de Pizzol (2018) em que há um maior detalhamento sobre os mesmos.

Fica claro que a utilização de múltiplos dias é benéfica para as pesquisas, especialmente aquelas que se destinam à compreensão de padrões comportamentais. Dados em corte propiciam a análise da heterogeneidade entre indivíduos, dados em painel, contudo, propiciam ampliar o conhecimento dessa heterogeneidade incorporando a mudança de comportamento no tempo.

Assim, este trabalho destina-se a classificar indivíduos, segundo comportamento individual relativo a viagens, a partir de dados em painel obtidos por *smartphones*. O objetivo secundário, associado a esta pesquisa, é promover uma maneira prática de obter dados em painel. A partir de tais dados, torna-se possível posterior classificação dos indivíduos em grupos, baseados na análise comportamental de seus deslocamentos considerando três dias úteis.

Sua obtenção será viabilizada por *smartphones*, pois são um meio viável de se estabelecer um conjunto de dados em painel para a pesquisa. Isso é possível dado que os aparelhos atuais são dotados de sensores *GPS* e conectividade com a *internet* móvel.

Assim, por meio do uso de aplicativos adequados, é possível que o telefone colete informações de deslocamentos e as armazene, tornando-se, assim, uma ferramenta para a descrição e estudos de viagens diárias. Seu uso justifica-se por duas razões: a primeira é a de que há considerável presença desse dispositivo na população. Pesquisas dão conta que no Brasil, 94% das

peças declaram ter um aparelho móvel, sendo que 62% é proprietário de, ao menos, um aparelho do tipo *smartphone* (Nogueira, 2018); a segunda é a de que as pessoas realizam seus deslocamentos em posse de seus *smartphones*, tornando a coleta de dados passiva, constante e precisa.

O potencial do uso de *smartphones* em pesquisas de viagens é explorado, comumente, na literatura atual, mostrando-se eficiente para aquisição de dados relacionados aos deslocamentos, com algumas restrições metodológicas (Assemi *et al.*, 2018; Safi *et al.*, 2014).

2. MATERIAIS E MÉTODO

Este trabalho foi conduzido na cidade de São Carlos – SP. Segundo o último levantamento do Instituto Brasileiro de Geografia e Estatística (IBGE), o município tem uma população de 221.950 pessoas, possui área territorial de 1136,907 km² e PIB per capita de R\$40.435,49. A cidade abriga duas grandes instituições de ensino superior: A Universidade de São Paulo (USP) que conta com dois campi e a Universidade Federal de São Carlos (UFSCar). Juntas, as instituições possuem cerca de 40 mil estudantes o que representa aproximadamente 18% da população do município (Assirati, 2018).

Pretendendo estabelecer um espaço amostral que fornecesse subsídios para o estabelecimento da nova proposta de coleta de dados e posterior agrupamentos, baseados na análise comportamental individual relativa a viagens, contou-se com a participação de 104 alunos, pertencentes ao curso de Graduação ou Pós-graduação em Engenharia civil na Escola de Engenharia de São Carlos, localizada no campus 1 da Universidade de São Paulo. Os alunos, voluntariamente, colaboraram com a coleta de informações durante um período de três dias úteis (terça, quarta e quinta-feira) do mês de agosto de 2017.

A escolha do mês de agosto se deu por ser um período letivo com ausência de feriados, época conveniente para a realização de análises acerca de estudantes. A escolha dos dias úteis é baseada no trabalho de Ribeiro e Cunha (2016). Os autores afirmam que, frente uma análise de dados históricos para tráfego de veículos, os dias de terça, quarta e quinta-feira têm comportamentos similares, sendo que para meses sem feriados, refletem padrões bastante distintos dos dias de final de semana e das segundas e terças-feiras.

As ações propostas para esse trabalho abrangem:

- Coleta de dados através de *Smartphones* e composição do painel;
- Construção e quantificação das cadeias de viagens;
- Algoritmo de agrupamento k-médias.

2.1 Coleta de dados via *Smartphones* e composição do painel

A coleta de dados foi realizada com aparelhos *Smartphone*. Não há a necessidade de nenhuma marca ou modelo em particular. Para que o telefone seja utilizado como uma ferramenta capaz de descrever os deslocamentos diários, basta que ela seja dotada de um sensor de GPS e um aplicativo de mapas. Para realizar a coleta, o usuário deve ativar a ferramenta de “*histórico de localização*” no aplicativo de mapas de seu dispositivo. Com tal ferramenta ativa, o telefone passa a armazenar informações de horário, latitude e longitude, referentes aos deslocamentos realizados no período de três dias. O processo é feito de maneira automática, não havendo a

necessidade de invocar a aplicação a todo momento, nem da tela permanecer ligada. O indivíduo, realizando suas viagens em posse do aparelho, é suficiente para que o trajeto seja arquivado, compondo um rico e preciso banco de dados.

Difere-se substancialmente da abordagem tradicional de diário de atividades por questionário por três motivos: (i) Questionários geralmente contemplam apenas um dia de atividades, já a nova abordagem abrange três dias úteis; (ii) Por muitas vezes, informações úteis ao desenvolvimento do projeto são perdidas devido aos esquecimentos por parte do indivíduo no ato da entrevista. Esquecimento esse que é eliminado na aquisição proposta, pois, desde que os deslocamentos sejam feitos portando o *Smartphone*, a coleta contemplará todas as movimentações sem perda de dados. Reforça-se que não há a necessidade de um plano de dados. Sendo o aparelho dotado do sensor de GPS, a coleta é realizada *off-line*, com armazenamento no próprio *Smartphone*; (iii) A questão do tempo dos deslocamentos: na modalidade de entrevista, é comum o entrevistado prover informações imprecisas, pois tem uma percepção equivocada dos tempos de deslocamento. Nessa proposta essa imprecisão é eliminada, pois a posse do *Smartphone* durante as movimentações computa o tempo real despendido.

Ao término dos registros de três dias, o usuário é instruído a submeter seus dados, em caráter de anonimato, em uma plataforma on-line. Durante essa submissão, um questionário socioeconômico (compreendendo informações sobre idade, sexo, grau de escolaridade, renda, tipo de moradia, execução de atividade remunerada, quantidade de automóveis disponíveis na residência e quantidade de motocicletas disponíveis na residência) é preenchido pelos indivíduos.

Ao receber os dados brutos, é necessário pré-processar esses dados a fim de torná-los compatíveis com os rumos que se deseja para a pesquisa. Seguindo técnicas apresentadas no trabalho de Anda, Erath e Fourie (2017), desenvolveu-se um algoritmo para a “clusterização” de tempos e distâncias para identificar os locais de atividades dos indivíduos a partir de seus dados brutos.

Quando um indivíduo se encontra em uma determinada localidade por muito tempo (por exemplo, um aluno no campus da universidade), este gera uma quantidade grande de registros (pontos georreferenciados) espacialmente próximos. É preciso então aglutinar esses registros todos, pois referem-se à mesma localidade em um grande intervalo de tempo.

Para promover essa aglutinação, define-se uma cerca geográfica (no caso desse estudo, o valor foi de 150 metros). Os pontos aglutinados são aqueles que estão compreendidos dentro desta cerca, que tem seu centro nas coordenadas do primeiro ponto de uma suposta localidade de atividades.

Após, compara-se os registros de horários dos pontos aglutinados. Se a diferença entre o primeiro e o último for maior que um determinado limite temporal (neste caso, o valor foi de 10 minutos), considera-se aquela aglutinação como um local de realização de atividades.

Por fim, substitui-se a aglutinação por um único registro, “geolocalizado” nas coordenadas do primeiro ponto, e com duração temporal começando no horário do primeiro ponto e se encerrando no horário do último.

Dispondo do banco de dados, deve-se compor o painel derivando informações a partir dos dados georreferenciados submetidos. Em sua forma original, os dados contêm as informações de data/hora e latitude/longitude (Figura 1A). A partir dos dados originais, são derivadas as seguintes informações: tempo e distância de deslocamentos, velocidade média dos deslocamentos, recorrência dos locais de atividades e escolha do modo de transporte (motorizado/não-

motorizado). A seguir, essas informações embasam a composição das cadeias de viagens, cuja quantificação é parâmetro de entrada para o algoritmo de agrupamento. A obtenção das informações derivadas e o método de composição das cadeias de viagens, bem como sua quantificação são descritos na sequência da seção e ilustrados nas Figuras 1 e 2.

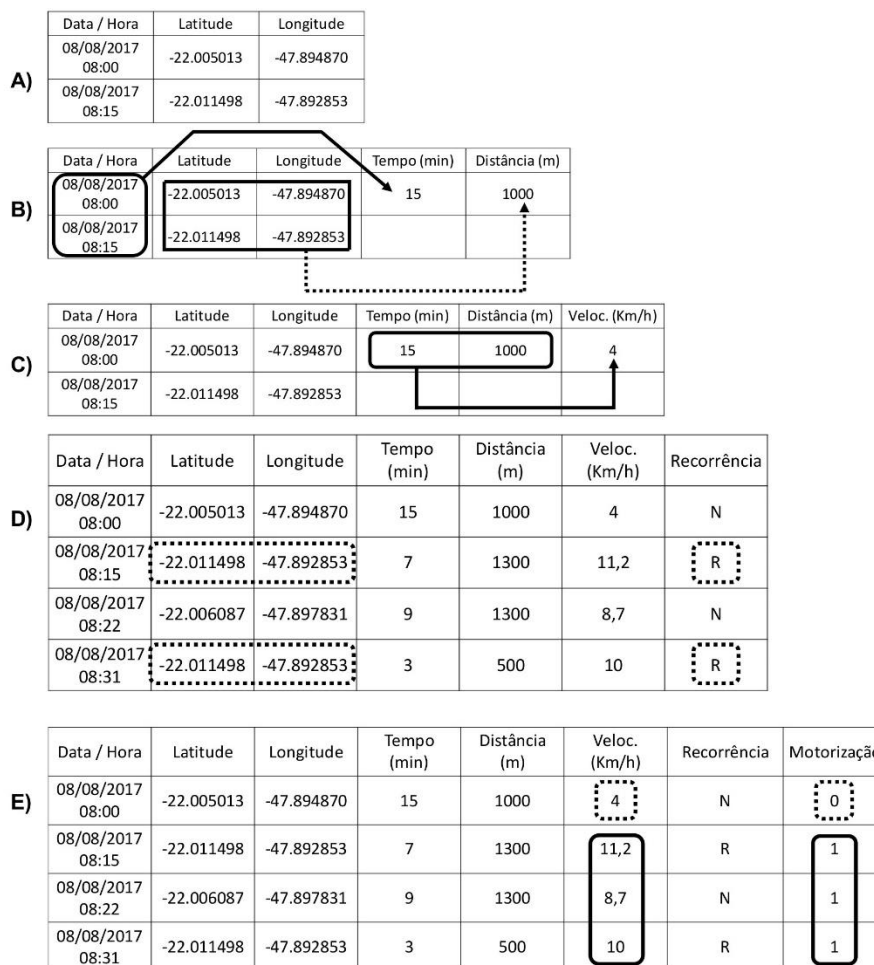


Figura 1: Exemplo de dados submetidos pelos participantes e posterior tratamento até a composição do painel.

O tempo de deslocamento (em minutos) entre duas localidades distintas é obtido pela subtração dos horários de partida da localidade inicial e chegada da localidade final. Já a distância (em metros) é dada pela dimensão da ortodromia entre os pontos, segundo as coordenadas de origem e destino. A Figura 1B exemplifica a derivação do tempo e da distância a partir de exemplos de registros.

A seguir, estima-se a velocidade média do deslocamento (km/h), que é calculada a partir da razão entre as variáveis distância e tempo, com conversões de unidade apropriadas. A Figura 1C ilustra a obtenção dessa informação.

Ao se estudar padrões comportamentais de deslocamento é relevante ter ciência se a movimentação foi feita para um local recorrente ou não-recorrente. Para a realização dessa tarefa foram examinados todos os pontos georreferenciados, por participantes, buscando por repetições. Caso a localidade se repetisse no intervalo de três dias, era rotulada como "R" por ser um local recorrente. Caso contrário, era rotulada como "N" de não-recorrente. A Figura 1D exemplifica a detecção de um local recorrente.

Finalmente, o painel de dados se completa com o caráter de motorização acerca dos deslocamentos de cada indivíduo. O caráter de motorização distingue os deslocamentos entre motorizados e não-motorizados e é baseada na velocidade média do deslocamento.

Assumiu-se que uma pessoa caminha com velocidade média de 6 km/h, assim, examinando todos os registros coletados, considerou-se que os deslocamentos que têm velocidade média igual ou inferior aos 6 km/h foram feitos por modos de transporte não-motorizados, sendo rotulados com o binário “0”, caso contrário foram rotulados com o binário “1”, indicando modos motorizados. Essa inferência sobre a escolha modal não leva em conta deslocamentos não-motorizados realizados por bicicletas, uma vez que essas, podem atingir velocidades equivalentes aos veículos motorizados em fluxo regular e que, portanto, seriam classificadas equivocadamente como um deslocamento motorizado. A Figura 1E exibe exemplos de deslocamentos motorizados/não-motorizados.

2.2. CONSTRUÇÃO E QUANTIFICAÇÃO DAS CADEIAS DE VIAGENS

Após o estabelecimento do painel completo, iniciou-se o processo de análise acerca dos comportamentos individuais. Para tal, deve-se estabelecer e quantificar as cadeias de viagens, que são um encadeamento dos deslocamentos, realizados por cada indivíduo no período de três dias.

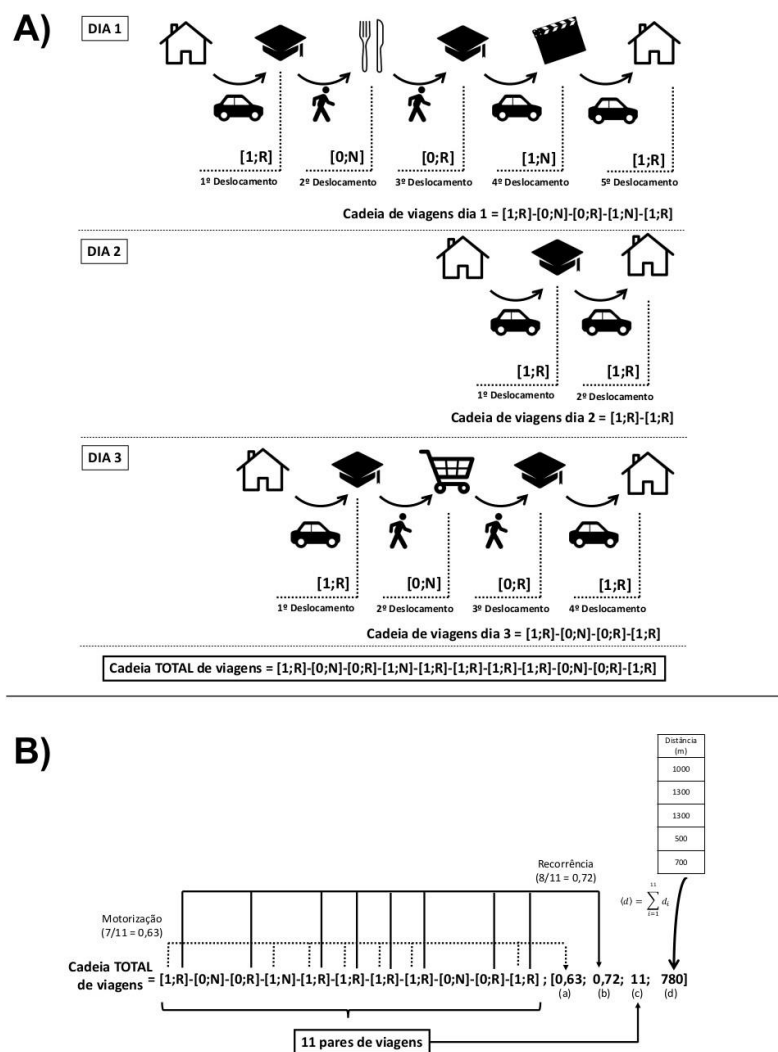


Figura 2: Exemplo de formação (A) e quantificação (B) da cadeia de viagens

Para cada indivíduo, a cadeia é formada concatenando os pares de rótulos de motorização (0 – Não motorizado; 1 – Motorizado) e recorrência (R – local recorrente; N – local não-recorrente), considerando-se todas as movimentações registradas no painel. A Figura 2A exemplifica a codificação de uma cadeia completa para um estudante, durante o período de três dias.

No dia 1, em seu primeiro deslocamento, o indivíduo desloca-se de automóvel (1) de sua residência para a universidade (R). Em seu segundo deslocamento, o indivíduo desloca-se caminhando (0) da universidade para um restaurante fora do campus (N). Em seu terceiro deslocamento, o indivíduo desloca-se caminhando (0) do restaurante de volta para a universidade (R). Em seu quarto deslocamento, o indivíduo desloca-se de automóvel (1) da universidade para um cinema (N). Em seu quinto deslocamento, o indivíduo desloca-se de automóvel (1) do cinema de volta para sua residência (R). O conjunto de cinco pares compõem a cadeia de viagens do primeiro dia $\rightarrow [1;R]-[0;N]-[0;R]-[1;N]-[1;R]$.

No dia 2, o primeiro deslocamento é feito de automóvel (1) da residência para a universidade (R). Lá o aluno passa o dia, retornando (segundo deslocamento) de automóvel (1) para sua residência (R). Os dois pares compõem a cadeia de viagens do segundo dia $\rightarrow [1;R]-[1;R]$.

No dia 3, há um primeiro deslocamento feito de automóvel (1) da residência para a universidade (R). Após, há um segundo deslocamento feito a pé (0) para um mercado (N). A seguir, há um terceiro deslocamento feito a pé (0) do mercado para a universidade (R) sucedido do quarto deslocamento, realizado de automóvel (1) da universidade para residência (R). O conjunto de quatro pares compõem a cadeia de viagens do terceiro dia $\rightarrow [1;R]-[0;N]-[0;R]-[1;R]$.

A concatenação das cadeias dos três dias resulta na cadeia total de viagens do indivíduo avaliado $\rightarrow [1;R]-[0;N]-[0;R]-[1;N]-[1;R]-[1;R]-[1;R]-[1;R]-[0;N]-[0;R]-[1;R]$.

Acrescenta-se, ainda, ao final da estrutura total, quatro informações que quantificam a cadeia de viagens:

- A porcentagem de motorização (razão entre a contagem do número de deslocamentos motorizados e a contagem total dos pares de rótulos);
- A porcentagem de recorrência (razão entre a contagem do número de deslocamentos para locais recorrentes e a contagem total dos pares de rótulos);
- O número total de viagens realizadas (contagem total dos pares de rótulos);
- A distância média de deslocamentos (média das distâncias de todos os deslocamentos realizados - extraída do painel de dados).

A Figura 2B ilustra o acréscimo das quatro informações quantitativas aplicadas ao exemplo da cadeia total da Figura 2A:

- Porcentagem de motorização (0,63) advinda de sete deslocamentos motorizados em um total de onze;
- Porcentagem de recorrência (0,72) obtida de três deslocamentos para locais recorrentes em um total de onze;
- Número de deslocamentos (11) proveniente da simples contagem do número de pares de viagem;
- Distância média dos deslocamentos, calculada a partir da média dos valores registrados no painel para o indivíduo.

A Tabela 1 sumariza as características dos dados, obtidos no período de três dias. Nota-se que há quem se locomova integralmente apenas por modos não-motorizados, entretanto, não há quem se locomova somente por modos motorizados. Para a variável “Grau de recorrência”,

não há quem somente vá a locais corriqueiros, nem quem faça deslocamentos apenas para locais infrequentes. “Número de viagens” e “distância média de deslocamentos” têm grande variação entre valor máximo e mínimo. A “faixa etária” dos 18 aos 22 anos é predominante e o “sexo” tem distribuição quase uniforme. Com relação ao “grau de escolaridade”, a vasta maioria dos alunos tem o ensino médio completo, o que significa que estão matriculados no curso de graduação. A pequena taxa de indivíduos com ensino superior completo são os alunos do curso de Pós-graduação. A grande maioria dos indivíduos não possui “renda” nem “exerce atividade remunerada”. Relativo à variável “moradia”, há um número razoável de alunos que residem em residência estudantil, entretanto, a maioria reside em casas/apartamentos. Quanto à “presença de automóveis na residência”, pouco mais da metade dos alunos não tem acesso a este veículo. Finalmente, relativo à “presença de motocicletas na residência”, é pequena a parcela de alunos que tem acesso a esse veículo.

Tabela 1: Sumarização das características dos dados coletados.

| Município | São Carlos – SP | | | |
|---|---|----------------------|--------------------------|-----|
| Estrato populacional | Estudantes universitários - Campus Universidade de São Paulo | | | |
| Registros | 104 | | | |
| Variáveis | Mínimo | Máximo | Média | |
| Grau de motorização | 0 | 0,83 | 0,42 | |
| Grau de recorrência | 0,2 | 0,92 | 0,67 | |
| Número de viagens | 4 | 20 | 11,68 | |
| Distância média dos deslocamentos (m) | 307 | 1099 | 585,19 | |
| Faixas etárias | 18-22 anos | 23-27 anos | 28+ anos | |
| | 78% | 19% | 3% | |
| Sexo | Masculino | | Feminino | |
| | 56% | | 44% | |
| Escolaridade | Médio completo | | Superior completo | |
| | 95% | | 5% | |
| Renda | Sem renda | 0-2 Salários Mínimos | 2,1 – 4 Salários Mínimos | |
| | 91% | 6% | 3% | |
| Moradia | Casa/Apartamento | | Residência Estudantil | |
| | 82% | | 18% | |
| Atividade remunerada | Sim | | Não | |
| | 9% | | 91% | |
| Número de automóveis disponíveis no domicílio | 0 | 1 | 2 | 3 |
| | 54% | 27% | 6% | 13% |
| Número de motocicletas disponíveis no domicílio | 0 | | 1 | |
| | 90% | | 10% | |

2.3. Algoritmo de agrupamento k-médias

O estabelecimento das cadeias de viagens é de suma relevância para a etapa de agrupamento, pois cada uma das quatro informações quantitativas, acrescidas às cadeias servirá como parâmetro de entrada para o algoritmo k-médias (MacQueen, 1967). Tal algoritmo faz parte das mais úteis ferramentas de técnicas de Aprendizagem de Máquinas não supervisionadas e foi usado muitas vezes na literatura tanto para encontrar relações entre variáveis explicativas e comportamento relativo a viagens, obtendo-se padrões de indivíduos e uso do solo e determinados comportamentos relativos aos deslocamentos (Park *et al.*, 2018; Sesham *et al.*, 2014; Ding e Zhang, 2016) como na caracterização do comportamento individual relativo a viagens, especialmente para grandes bases de dados (Goulet-Langlois, 2015; Agard *et al.*, 2013).

Tal algoritmo objetiva classificar em k grupos um conjunto de n observações, definidas em termos de seus respectivos vetores de características. O intuito do algoritmo é definir k centroides, um para cada um dos k grupos. Esses centroides inicialmente são distribuídos randomicamente pelo espaço delimitado pelo conjunto das n observações. O próximo passo é pegar cada um dos n pontos e associá-los ao centroide mais próximo. Quando nenhum ponto está pendente, o primeiro passo é concluído e um agrupamento inicial é feito. Neste ponto, precisa-se recalcular k novos centroides como baricentros dos grupos resultantes da etapa anterior. Depois de se ter esses k novos centroides, uma nova ligação deve ser feita entre os mesmos pontos de conjunto de dados e o novo centroide mais próximo. Esses passos devem ser repetidos enquanto os k centroides mudarem sua localização quando comparados aos centroides da iteração anterior. O algoritmo se encerra quando os centroides não mais se moverem.

Neste trabalho, os 104 registros do painel geraram as $n=104$ cadeias totais de viagens, que são definidas pelo vetor de quatro características: porcentagem de motorização, porcentagem de recorrência, número de viagens e distância média das viagens.

Por se tratar de uma análise multidimensional, considerou-se que as características deveriam ter o mesmo peso analítico, portanto, decidiu-se por normalizar os dados de forma que todas os valores do vetor de características estivessem dentro de uma extensão que varia de 0 a 1.

3. RESULTADOS E DISCUSSÕES

Foi obtido um particionamento para $k=3$ grupos, os quais são exibidos (Figura 3) em diagramas de Voronoi, plotados em plano bidimensional para cada uma das características consideradas. Tal número de agrupamentos se deu considerando o valor de k que convergiu em um menor número de iterações para o cenário de centroides fixos. Esse valor foi corroborado posteriormente pelo *Elbow Method* que apontou não haver mais mudanças significativas de variância dos valores dos elementos de cada grupo após o valor $k=3$.

O algoritmo de agrupamento, para o presente trabalho, deve ser considerado efetivo pois segmentou três grupos com homogeneidade interna e heterogeneidade entre grupos considerando as quatro variáveis que caracterizam as viagens. Tal afirmação sustenta-se uma vez que cada uma das células de Voronoi (região delimitada por elementos pertencentes a um mesmo grupo), está bem demarcada, havendo pouca ou nenhuma superposição de elementos (fator que induz classificações equivocadas) em todos os diagramas.

Convencionou-se as nomenclaturas:

- Grupo1 (+verde) – Estudantes universitários com baixo grau de motorização, pouca recorrência de localidades, poucas viagens e distâncias curtas: *Estudantes não motorizados com motivos diversos de viagens, predominantemente curtas.*
- Grupo2 (◇vermelho) – Estudantes universitários com grau de motorização moderada, recorrência moderada, número alto de viagens e distâncias médias: *Estudantes com opções motorizadas, com viagens recorrentes e também motivos variados e distâncias médias.*
- Grupo3 (●azul) – Estudantes universitários com alto grau de motorização, alta recorrência das viagens, poucas viagens e distâncias longas: *Estudantes motorizados com viagens longas, predominância das atividades obrigatórias (recorrentes).*

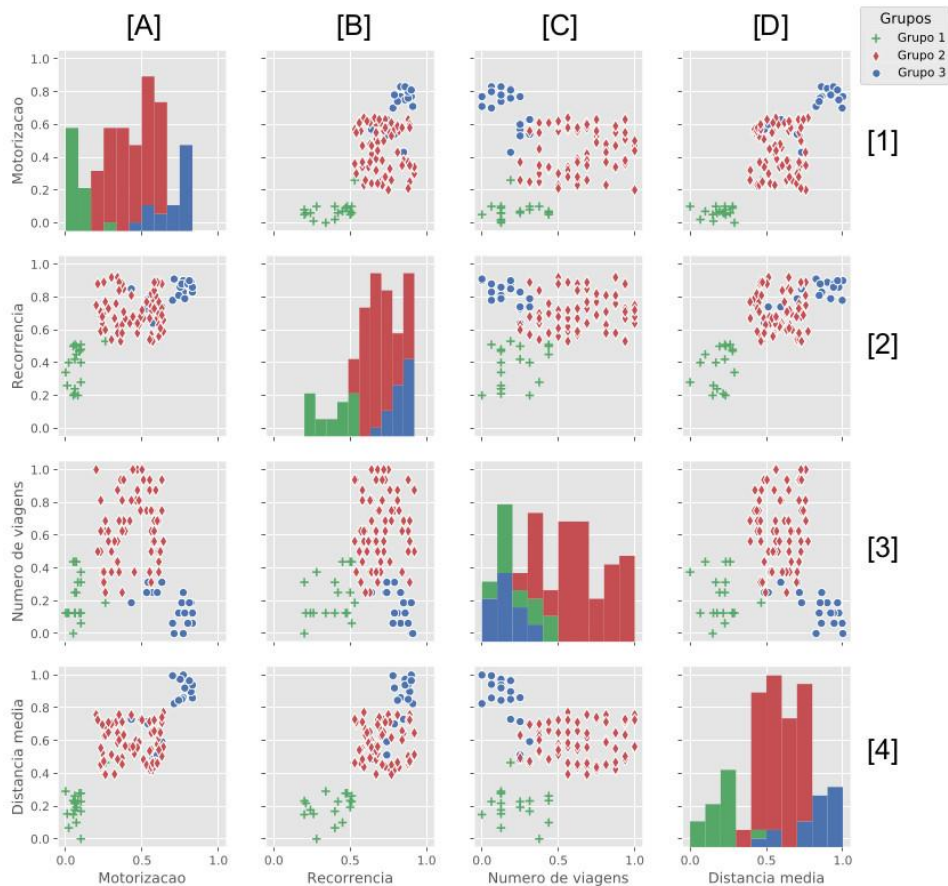


Figura 3: Particionamento estabelecido pelo algoritmo k-médias

A Tabela 2 apresenta a mesma sumarização de dados da Tabela 1, porém agora, aplicada aos grupos estipulados:

Tabela 2: Sumarização das características dos dados coletados discriminadas nos três grupos estabelecidos.

| Município | São Carlos - SP | | | |
|---------------------------------------|--|----------------------|-----------------------|---------|
| Estrato populacional | Estudantes universitários - Campus Universidade de São Paulo | | | |
| Registros | 104 | | | |
| Grupo 1 | Estudantes não motorizados com motivos diversos de viagens, predominantemente curtas | | | |
| Elementos | 18 | | | |
| Variáveis | Mínimo | Máximo | Média | |
| Grau de motorização | 0 | 0,26 | 0,07 | |
| Grau de recorrência | 0,2 | 0,53 | 0,38 | |
| Número de viagens | 4 | 11 | 7,5 | |
| Distância média dos deslocamentos (m) | 307 | 536 | 408,66 | |
| Faixas etárias | 18-22 anos | 23-27 anos | 28+ anos | |
| | 72% | 28% | 0% | |
| Sexo | Masculino | | Feminino | |
| | 61% | | 39% | |
| Escolaridade | Médio completo | | Superior completo | |
| | 94% | | 6% | |
| Renda | Sem renda | 0-2 Salários Mínimos | 2,1 - 4 Salários | Mínimos |
| | 89% | 11% | 0% | |
| Município | São Carlos - SP | | | |
| Moradia | Casa/Apartamento | | Residência Estudantil | |
| | 78% | | 22% | |
| Atividade remunerada | Sim | | Não | |
| | 11% | | 89% | |

Tabela 2: Sumarização das características dos dados coletados discriminadas nos três grupos estabelecidos. (continuação)

| Município | São Carlos - SP | | | |
|---|--|----------------------|--------------------------|-----|
| | 0 | 1 | 2 | 3 |
| Número de automóveis disponíveis no domicílio | 83% | 11% | 0% | 6% |
| Número de motocicletas disponíveis no domicílio | 0 | | 1 | |
| | 94% | | 6% | |
| Grupo 2 | Estudantes com opções motorizadas, com viagens recorrentes e também motivos variados e distâncias médias | | | |
| Elementos | 70 | | | |
| Variáveis | Mínimo | Máximo | Média | |
| Grau de motorização | 0,2 | 0,64 | 0,44 | |
| Grau de recorrência | 0,53 | 0,92 | 0,7 | |
| Número de viagens | 7 | 20 | 13,97 | |
| Distância média dos deslocamentos (m) | 510 | 687 | 596,54 | |
| Faixas etárias | 18-22 anos | 23-27 anos | 28+ anos | |
| | 84% | 11% | 5% | |
| Sexo | Masculino | | Feminino | |
| | 53% | | 47% | |
| Escolaridade | Médio completo | | Superior completo | |
| | 97% | | 3% | |
| Renda | Sem renda | 0-2 Salários Mínimos | 2,1 - 4 Salários Mínimos | |
| | 99% | 1% | 0% | |
| Moradia | Casa/Apartamento | | Residência Estudantil | |
| | 81% | | 19% | |
| Atividade remunerada | Sim | | Não | |
| | 1% | | 99% | |
| Número de automóveis disponíveis no domicílio | 0 | 1 | 2 | 3 |
| | 56% | 24% | 4% | 16% |
| Número de motocicletas disponíveis no domicílio | 0 | | 1 | |
| | 91% | | 9% | |
| Grupo 3 | Estudantes motorizados com viagens longas, predominância das atividades obrigatórias (recorrentes) | | | |
| Elementos | 16 | | | |
| Variáveis | Mínimo | Máximo | Média | |
| Grau de motorização | 0,53 | 0,83 | 0,73 | |
| Grau de recorrência | 0,73 | 0,91 | 0,83 | |
| Número de viagens | 4 | 9 | 6,37 | |
| Distância média dos deslocamentos (m) | 599 | 1099 | 734,12 | |
| Faixas etárias | 18-22 anos | 23-27 anos | 28+ anos | |
| | 56% | 44% | 0% | |
| Sexo | Masculino | | Feminino | |
| | 63% | | 38% | |
| Escolaridade | Médio completo | | Superior completo | |
| | 88% | | 13% | |
| Renda | Sem renda | 0-2 Salários Mínimos | 2,1 - 4 Salários Mínimos | |
| | 62% | 19% | 19% | |
| Moradia | Casa/Apartamento | | Residência Estudantil | |
| | 88% | | 12% | |
| Atividade remunerada | Sim | | Não | |
| | 38% | | 62% | |
| Número de automóveis disponíveis no domicílio | 0 | 1 | 2 | 3 |
| | 12% | 56% | 19% | 13% |
| Número de motocicletas disponíveis no domicílio | 0 | | 1 | |
| | 81% | | 19% | |

Verifica-se, na Tabela 2, que valores mínimos globais são todos advindos do Grupo 1. O Grupo 2 colabora com os valores máximos globais das variáveis “Grau de recorrência” e “Número de viagens”. O Grupo 3 colabora com os valores máximos globais das variáveis “Grau de motorização” e “Distância média dos deslocamentos”.

A convenção de nomenclatura estipulada é reforçada pela estatística referente ao número de automóveis disponíveis no domicílio: apenas 17% dos indivíduos do Grupo 1 têm acesso a pelo

menos um automóvel. Esse número aumenta para 44% no Grupo 2 até culminar em 88% no Grupo 3. Estabelecidas as nomenclaturas, procede-se às seguintes análises referentes aos diagramas da Figura 3. Em seguida, encontra-se o Quadro 1 que sintetiza as análises realizadas.

(i) Figura 3[A][4] - Grau de motorização versus distância média de viagens: Há uma correlação positiva entre o grau de motorização e a distância média das viagens, uma vez que a medida que o grau de motorização aumenta, a distância média de viagens também aumenta, pois, a posse de um veículo motorizado privilegia deslocamentos mais extensos. Ressalta-se também a distinção dos grupos neste gráfico, onde o Grupo 1 está concentrado no canto inferior esquerdo e gradativamente vai aumentando ambas as características, passando pelo Grupo 2, até culminar no Grupo 3 (canto superior direito).

(ii) Figura 3[A][3] - Grau de motorização versus número de viagens: do diagrama percebe-se que o maior número de viagens está associado ao Grupo 2. Indivíduos deste conjunto combinam dois tipos de modos de transporte ao longo do dia, motorizados e não-motorizados. Segundo a Tabela 2, 19% dos elementos desse grupo residem em moradia estudantil, aumentando as probabilidades destes indivíduos terem acesso a pelo menos um automóvel, experimentando viagens feitas por meio de carona. Rotineiramente, o modo principal de locomoção deste grupo é não motorizado, entretanto, sempre que possível, valem-se de viagens motorizadas apesar de muitos não possuírem veículos, visto que 99% não exercem atividade remunerada.

Os indivíduos do Grupo 1, por não terem acesso à motocicletas ou automóveis, geralmente realizam um número reduzido de viagens. Os indivíduos do Grupo 3, também realizam um número reduzido de viagens, visto que seus deslocamentos são em média os mais longos (análise i). Como há uma estreita relação entre maiores distâncias percorridas e um aumento do tempo de viagem, estes acabam por realizar menos viagens a fim de não despenderem muito tempo nas mesmas. O elevado grau de motorização desse grupo explica-se pelo fato deste concentrar a maior taxa de alunos de pós-graduação (13% possui ensino superior completo) e também a maior taxa de indivíduos que exercem atividade remunerada (38%).

(iii) Figura 3[A][2] - Grau de motorização versus Grau de recorrência: Estas são duas grandezas que também têm correlação positiva entre si, sendo que os elementos pertencentes ao Grupo 1 são os mais propensos a um comportamento menos recorrente, uma vez que estes estão mais aptos a realizarem viagens de oportunidade, pois a não motorização privilegia grande variação de caminhos realizados e locais visitados. Ambas as variáveis aumentam gradativamente até chegarem no comportamento dos elementos do Grupo 3 que normalmente restringem-se a realizar deslocamentos para desempenhar suas atividades regulares, já que residem mais distantes das localidades recorrentes (Figura 4).

(iv) Figura 3[B][4] - Grau de recorrência versus Distância média: Outro gráfico que apresenta correlação positiva entre as variáveis, evidencia o fato de que viagens para locais não recorrentes normalmente têm uma menor distância média, enquanto viagens para locais recorrentes estão associados aos maiores deslocamentos.

Pela análise i, infere-se que deslocamentos para localidades não recorrentes são realizados por meios não-motorizados, meios estes que, segundo a análise iii, contemplam menores distâncias, expressos apenas por indivíduos pertencentes ao Grupo 1.

Já o comportamento relativo aos locais recorrentes com distância média elevada, associados ao modo motorizado, são expressos pelos elementos dos Grupos 2 e 3.

(v) Figura 3[B][3] – Grau de recorrência versus Número de viagens: Este diagrama esclarece que ambos os extremos, viajantes que majoritariamente deslocam-se para locais não recorrentes e viajantes que majoritariamente deslocam-se para locais recorrentes, estão associados a um menor número de viagens realizadas por dia, sendo contemplados pelos Grupos 1 e 3, respectivamente. O comportamento intermediário que condiz a um misto entre viagens com propósitos recorrentes e não-recorrentes, concentra o maior número de deslocamentos no período analisado, e manifesta-se pelos elementos do Grupo 2.

(vi) Figura 3[C][4] – Número de viagens versus Distância média: Este diagrama demonstra que um baixo número de viagens está associado a curtas distâncias ou longas distâncias, enquanto que valores mais altos de número de viagens correlacionam-se a distâncias médias percorridas.

Segundo a análise ii, a ausência da motorização (Grupo 1) ou a presença com uso predominante (Grupo 3) está ligado a um menor número de viagens, que de acordo com a análise v, tem íntima relação com os extremos de baixo e alto grau de recorrência. Já o Grupo 2, que combina dois modos, realiza um número maior de viagens, sendo que tais deslocamentos não são nem tão curtos como o dos indivíduos do Grupo 1 e, nem tão longos como o dos indivíduos do Grupo 3.

(vii) Figura 3[A][1] – Histograma da variável grau de motorização: A distribuição do histograma evidencia que há uma transição entre o grau de motorização e os elementos de cada grupo. As caixas mais a esquerda são completamente compostas por elementos do Grupo 1. Taxas centrais são compostas predominantemente por indivíduos do Grupo 2, havendo alguma superposição da baixa motorização nas caixas à esquerda e da alta motorização à direita. Caixas mais a direita são completamente compostas por elementos do Grupo 3.

(viii) Figura 3[B][2] – Histograma da variável grau de recorrência: As taxas de baixa recorrência são dominadas pelos indivíduos do Grupo 1. À medida que a taxa chega a um valor médio começa a haver predominância dos indivíduos do Grupo 2. Tal predominância estende-se até os mais elevados graus de recorrência, quando a composição é mista entre elementos dos Grupos 2 e 3.

(ix) Figura 3[C][3] – Histograma da variável número de viagens: Os menores valores de viagens realizadas são compostos por elementos dos Grupos 1 e 3. Para os valores intermediários, há uma superposição entre os três grupos. Os valores elevados são predominantemente realizados pelo Grupo 2.

(x) Figura 3[D][4] – Histograma da variável distância média de viagens: Outro histograma que apresenta transição entre o grau de motorização e os elementos dos grupos. Caixas mais a esquerda, predominantemente compostas por indivíduos do Grupo 1, apontam os menores deslocamentos. Caixas mais a direita, completamente compostas por indivíduos do Grupo 3, apontam os maiores deslocamentos. As caixas mais altas, que indicam maior predominância no conjunto de dados, apontam que a maior parte dos deslocamentos são de caráter médio, vastamente realizados por elementos do Grupo 2, apresentando alguma intersecção com os outros grupos.

Quadro 1. Resumo das análises dos três grupos

| Grupos | Grau de motorização | Distâncias | Número de Viagens | Recorrência | Posse de auto/moto |
|---------|---------------------|------------|-------------------|-------------|--------------------|
| Grupo 1 | Baixo | Baixas | Baixo | Baixa | Baixa |
| Grupo 2 | Moderado | Médias | Alto | Moderada | Moderada |
| Grupo 3 | Alto | Altas | Baixo | Alta | Alta |

Uma vez que se conta com a localidade da residência de cada registro, é possível construir um mapa (Figura 4) onde observa-se a distribuição geográfica dos grupos. Por essa perspectiva é possível corroborar muitas das análises relativas à Figura 3.

O conjunto de dados analisado compete estudantes universitários, portanto é natural que o seu local de estudo (circunferência alaranjada: campus 1 da Universidade de São Paulo) seja uma localidade recorrente e um ponto agregador de viagens.

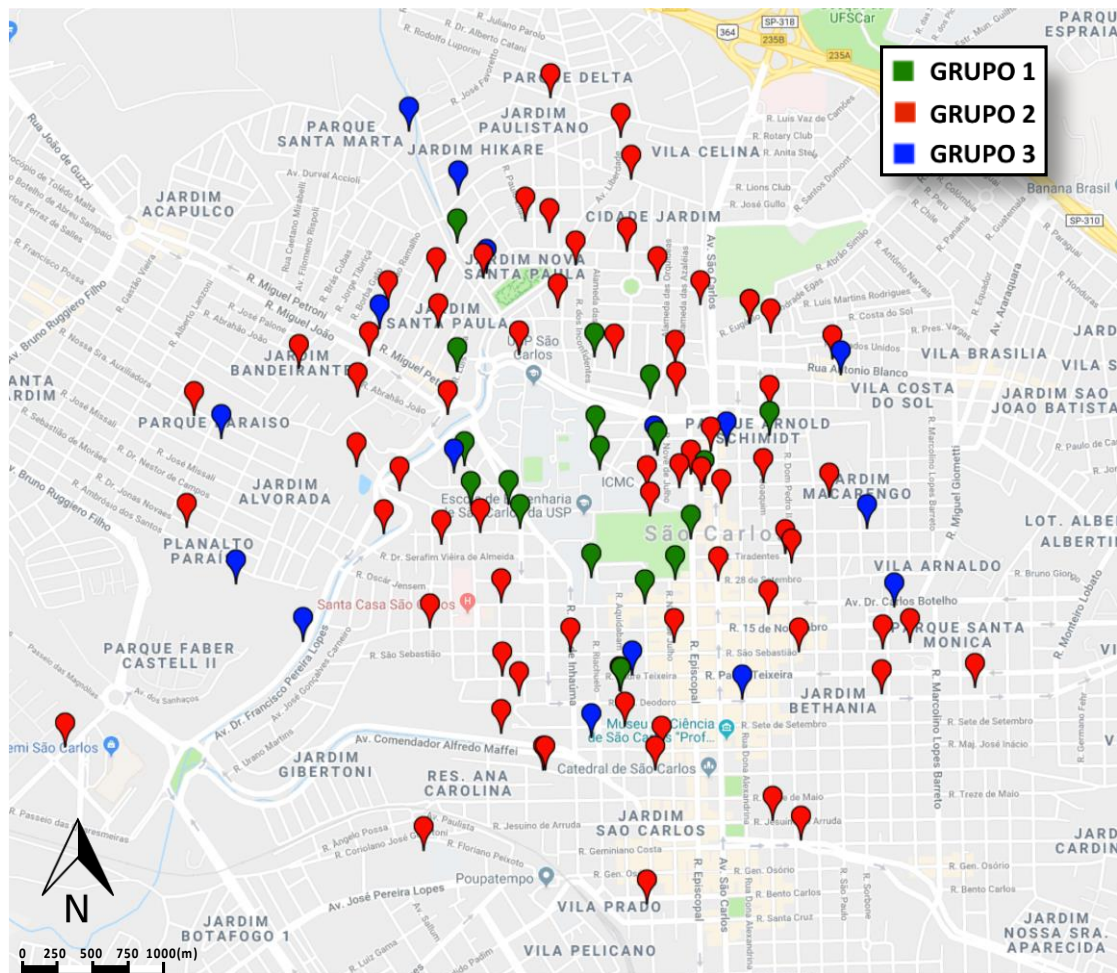


Figura 4: Mapa construído a partir das coordenadas da residência dos indivíduos estudados.

Os elementos do Grupo 1 (marcadores verdes) em sua maioria residem nas proximidades do campus pois não estão aptos à realização de longos deslocamentos.

Já os elementos representados pelos marcadores vermelhos, têm comportamento misto, visto que muitos elementos moram nas proximidades do campus e tantos outros residem em localidades mais afastadas, o que está de acordo com a categoria de motorização moderada deste agrupamento (Grupo 2), onde os indivíduos alternam entre deslocamentos motorizados e não-motorizados.

Finalmente, os marcadores em azul, em grande maioria, distam bastante do campus, estando em conformidade com o Grupo 3, que uma vez dotados do acesso aos modos motorizados, tornam-se aptos às maiores movimentações.

4. CONCLUSÕES

Este artigo foi delineado com base em duas lacunas de pesquisa, associadas ao objetivo principal e secundário, respectivamente: (1) Classificar indivíduos segundo comportamento relacionado a viagens a partir de múltiplos dias e; (2) Obter dados em painel através de *smartphones*.

Em relação à primeira lacuna e objetivo principal, a principal contribuição obtida foi a caracterização numérica de cadeias de viagens, considerando diversas características dos deslocamentos de múltiplos dias. Tal caracterização pode ser replicada, futuramente, em outros trabalhos e utilizada em outras ferramentas, podendo ser, inclusive, variável dependente em modelos paramétricos associados ao comportamento individual relacionado a viagens. A aplicação sequencial de um algoritmo de agrupamentos para classificar os indivíduos em grupos, concernentes aos seus comportamentos, constitui uma contribuição metodológica, pois trabalhos anteriores basearam-se em caracterizar o comportamento de múltiplos dias com base em indicadores e valores médios referentes às escolhas dos deslocamentos (Pizzol, 2018; Susilo e Axhausen, 2014).

A aquisição de dados em painel via *smartphones* mostra-se uma alternativa viável para a construção de um painel de dados que permita o estudo de padrões comportamentais de deslocamentos. A economia de recursos monetários se dá pelo elevado número de aparelhos que a população detém. Utilizados em seus deslocamentos, mostram-se uma ferramenta em potencial para aquisição de precisos dados sem a necessidade de dispender recursos com entrevistadores, podendo coletar informações referentes a vários dias, enriquecendo a análise comportamental. Embora sejam encontrados, na literatura recente, trabalhos provenientes de dados em painel, coletados por *smartphones*, o tratamento, inferências e caracterização destes ainda é uma questão importante (Andrade *et al.*, 2017; He *et al.*, 2017; Vlassenroot *et al.*, 2014).

Ressalva-se que as discussões apresentadas se aplicam à amostra limitada, de oportunidade, não sendo possível generalizar tais comportamentos para toda a população do município. Entretanto, os procedimentos de aquisição e processamento de dados, algoritmo de agrupamento e mecanismos de análises são de cunho geral, sendo aplicáveis a outro estrato populacional ou região que se deseje compreender padrões de deslocamentos.

AGRADECIMENTOS

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Os autores também agradecem à turma de Engenharia Civil da EESC - USP, bem como os alunos do Programa de Pós-Graduação em Engenharia de Transportes da EESC USP, pela participação voluntária na coleta de dados.

REFERÊNCIAS

- Anda, C., Erath, A., & Fourie, P. J. (2017). Transport modelling in the age of big data. *International Journal of Urban Sciences*, v. 21, p. 19-42. DOI:10.1080/12265934.2017.1281150
- Andrade, P. F. L., Gogoy, L. A., Giannotti, M. A., Cunha, C. B., & Yoshizaki, H. T. Y. (2017). Análise e visualização de dados de rastreamento para caracterização da logística urbana. *Transportes*, v. 25, n. 3, p. 24-35. DOI: 10.14295/transportes.v25i3.1353
- Agard, B., Partovi Nia, V., & Trépanier, M. (2013). *Assessing public transport travel behaviour from smart card data with advanced data mining techniques*. In World Conference on Transport Research, v13, p. 15-18.
- Arentze, T. A. e H. J. P. Timmermans (2005) Information gain, novelty seeking and travel: a model of dynamic activity-travel behavior under conditions of uncertainty. *Transportation Research Part A*, v.39, n.2-3, p.125-145. DOI:10.1016/j.tra.2004.08.002
- Assemi, B., Jafarzadeh, H., Mesbah, M., & Hickman, M. (2018). *Participants' perceptions of smartphone travel surveys*. *Transportation research part F: traffic psychology and behavior*. v. 54, p. 338-348. DOI: 10.1016/j.trf.2018.02.005
- Assirati, L. (2018). *Análise da influência da vizinhança no comportamento individual relativo a viagens através de dados em painel*. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, Springer, v. 29, n. 2, p. 95–124, 2002. DOI: 10.1023/A:1014247822322
- Daisy, N. S.; M. H. Hafezi; L. Liu e H. Millward (2018) Understanding and Modeling the Activity-Travel Behavior of University Commuters at a Large Canadian University. *Journal of Urban Planning and Development*, v.144, n.2, p.04018006. DOI:10.1061/(asce)up.1943-5444.0000442
- Ding, L., & Zhang, N. (2016). *A travel mode choice model using individual grouping based on cluster analysis*. *Procedia engineering*, v. 137, p. 786-795, 2016. DOI: 10.1016/j.proeng.2016.01.317
- Goulet-Langlois, G. (2015). Exploring regularity and structure in travel behavior using smart card data. Tese de Doutorado - Massachusetts Institute of Technology.
- Hagerstrand, T. (1970) What about people in regional science? *Papers of the Regional Science Association*, v.24, p. 7-21. DOI: 10.1007/BF01936872
- Hanson, S., & Huff, J. O. (1981). *Assessing day-to-day variability in complex travel patterns*. *Transportation Research Record*, v. 891, p. 18–24, 1981.
- He, S., Miller, J. E., Scott, D. (2017). Big data and travel behaviour. *Travel Behaviour and Society*, v. 11. DOI: 10.1016/j.tbs.2017.12.003.
- Hensher, D., & King, J. (2001) *The Leading Edge of Travel Behavior Research*. *Travel behaviour research: The leading edge*. Amsterdam: Elsevier, p. 1-13. DOI: 10.1016/b978-008043924-2/50000-6
- Huff, J. O., & Hanson, S. (1986). *Repetition and Variability in Urban Travel*. Wiley Online Library, v. 18, n. 2, p. 97–114. DOI:10.1111/j.1538-4632.1986.tb00085.x
- Kockelman, K. (1997) Travel behavior as function of accessibility, land use mixing, and land use balance: evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board*, n.1607, p.116-125. DOI:10.3141/1607-16
- MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, v.1, n.14, p.281-297.
- Nogueira, E. L. P. (2018). *O uso da calculadora gráfica GeoGebra no smartphone como ferramenta para o ensino das funções exponencial e logarítmica* (Master's thesis, Brasil).
- Ortúzar, J. de D., e Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons. DOI:10.1002/9781119993308
- Park, K., Ewing, R., Scheer, B. C., & Ara Khan, S. S. (2017). *Travel Behavior in TODs vs. Non-TODs: Using Cluster Analysis and Propensity Score Matching*. *Transportation Research Record*, p. 0361198118774159. DOI:10.1177/0361198118774159
- Pas, E. I. (1987). *Intrapersonal variability and model goodness-of-fit*. *Transportation Research Part A: General*, Elsevier, v. 21, n. 6, p. 431–438, 1987. DOI:10.1016/0191-2607(87)90032-x
- Pitombo, C. S.; E. Kawamoto e A. J. Sousa (2011) An exploratory analysis of relationships between socioeconomic, land use, activity participation variables and travel patterns. *Transport Policy*, v.18, n.2, p.347-357. DOI:10.1016/j.transpol.2010.10.010
- Pizzol, B. (2018) *Padrões de atividades de residentes de Paraisópolis: Análise de dados de múltiplos dias coletados com smartphones*. Dissertação (Mestrado) — Universidade de São Paulo.
- Ramadurai, G. e S. Ukkusuri (2010) Dynamic user equilibrium model for combined activity-travel choices using activity-travel supernetwork representation. *Networks and Spatial Economics*, v.10, n.2, p.273-292. DOI:10.1007/s11067-008-9078-3
- Ribeiro E. e A. L. Cunha (2016) Análise exploratória de método para definição de dia típico utilizando transformada wavelet e análise de agrupamento. *Anais do XXX Congresso de Pesquisa e Ensino em Transportes*, ANPET, Rio de Janeiro, v.1, p.1502–1513.
- Safi, H., Assemi, B., Mesbah, M., & Ferreira, L. G. (2014). *A framework for smartphone-based travel surveys: an empirical comparison with alternative methods in New Zealand*. In 10th International Conference on Transport Survey Methods.
- Sesham, A., Padmanabham, P., Goverdhan, A., & Sai Hanuman, A. (2014). *Performance of clustering algorithms on home interview survey data employed for travel demand estimation*. *International Journal of Computer Science & Information Technology*, v. 5, n. 3, p. 2767-2771.
- Susilo, Y. O., & Axhausen, K. W. (2014). *Repetitions in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman Index*. *Transportation*, Springer, v. 41, n. 5, p. 995–1011. DOI:10.1007/s11116-014-9519-4
- Susilo, Y. O., & Kitamura, R. (2005). *Analysis of day-to-day variability in an individual's action space: exploration of 6-week Mobilize travel diary data*. *Journal of the Transportation Research Board*, TRB, n. 1902, p. 124–133. DOI: 10.3141/1902-15.
- Vlassenroot, S., Gillis, D., Bellens, R., Gautama, S. (2014). The Use of Smartphone Applications in the Collection of Travel Behaviour Data. *International Journal of Intelligent Transportation Systems Research*. v. 13, p. 1-11. DOI: 10.1007/s13177-013-0076-6.