

# Influência das condições climáticas e de acidentes na caracterização do comportamento do tráfego em rodovias

Felipe Caleffi<sup>1</sup>, Shanna Trichês Lucchesi<sup>2</sup>, Michel José Anzanello<sup>3</sup> e Helena Beatriz Bettella Cybis<sup>4</sup>

**Resumo:** Dados meteorológicos e relacionados a questões de segurança têm sido integrados a análises que buscam caracterizar as relações fundamentais do fluxo de tráfego em rodovias, com o objetivo de compreender melhor o comportamento do tráfego em condições adversas (como chuvas), e em períodos com ocorrências de incidentes. Neste contexto, este artigo propõe a aplicação de métodos de agrupamento de observações (clusterização) para caracterizar o comportamento do tráfego da rodovia BR-290/RS, na região metropolitana da cidade de Porto Alegre – RS. As técnicas de clusterização aplicadas na análise são *K-means* e *fuzzy clustering* (FCM). Através da clusterização, foi possível determinar padrões comportamentais influenciados pelas variáveis (fluxo, velocidade, acidentes e chuvas). Clusters relacionados a períodos com chuva indicam redução de velocidade em cerca de 10 km/h. Cluster de períodos relativos a condições de acidentes indicam que as velocidades reduzem para intervalos entre 10 e 35 km/h, com volumes sempre inferiores a 1000 veic/h.

*Palavras-chave:* clusterização, acidentes, chuvas, relação fluxo-velocidade.

**Abstract:** Meteorological and safety-related data have been integrated to several techniques aimed at characterizing the fundamental relationships of traffic flow on highways, in order to better understand traffic behavior in adverse conditions such as rain, and when incidents occurs. In this context, this paper proposes the use of clustering methods to characterize the traffic behavior in highway BR-290/RS, located in the metropolitan area of Porto Alegre - RS. The clustering techniques tested, K-means and fuzzy clustering (FCM), enabled the determination of behavioral patterns influenced by variables (flow, speed, accidents and rain). Clusters related to periods with rain indicate speed drop by about 10 km/h. Clusters periods related to accident conditions indicate speeds between 10 and 35 km/h, and volumes below 1000 veic/h.

*Keywords:* clustering, accidents, rain, speed-flow relationship.

## 1. INTRODUÇÃO

Classificar as condições do tráfego é essencial para caracterizar o desempenho operacional de uma rodovia. Esta classificação baseia-se tipicamente em variáveis, tais como velocidade, volume e densidade. Através destas caracterizações, é possível entender melhor o comportamento do tráfego na rodovia, e assim, determinar períodos com diferentes condições de tráfego – períodos em que a rodovia opera em condições de fluxo livre, na capacidade, ou em condições de fluxo congestionado (anda/para) (Xia e Chen, 2007a).

Tais classificações, como as apresentadas por Xia e Chen (2007a), Xia e Chen (2007b), Sun e Zhou (2005), Park (2002), Kianfar e Edara (2013), Weijermars e van Berkum (2005) e Chunchun *et al.* (2011), contudo, ignoram eventos importantes, como períodos com acidentes e condições adversas que incluem chuvas ou baixa visibilidade. Períodos em que ocorrem estes eventos normalmente são classificados apenas levando em consideração suas características de fluxo e, portanto, ignoram se a rodovia estava operando com um número reduzido de faixas (em caso de acidentes), ou se a rodovia estava operando com velocidades baixas,

devido à baixa visibilidade e/ou grandes volumes de chuvas. Agregando a ocorrência de acidentes e obstrução de faixas, além de períodos com chuva e baixa visibilidade, é possível aprimorar a classificação das condições de fluxo da rodovia e, principalmente, compreender de forma mais detalhada o comportamento do tráfego em períodos de colapso do fluxo onde ocorrem os congestionamentos.

Progressos em mineração de dados têm disponibilizado ferramentas eficazes para a caracterização do comportamento do tráfego. A mineração de dados é o processo de descobrir e extrair padrões e relacionamentos anteriormente desconhecidos, em conjuntos de dados grandes e complexos. Dentre as técnicas de mineração de dados, técnicas de clusterização apresentam grande potencial para identificar semelhanças entre conjuntos de dados complexos e fornecer caracterizações de tráfego confiáveis. Analisando dados históricos por estes métodos, dados de tráfego podem ser agrupados através de distintas características, com cada grupo representando um padrão de comportamento típico (Caceres *et al.*, 2012).

A possibilidade de identificar informações sobre padrões dos grupos (*clusters*), e as características de cada um destes grupos faz das técnicas de clusterização uma ferramenta adequada para estimar as condições de operação e comportamento do tráfego em rodovias. Para classificar os dados em diferentes *clusters*, procedimentos de agrupamento normalmente fazem uso das características do tráfego como variáveis de entrada, as quais tipicamente são coletadas por detectores fixos (laços indutivos, câmeras de vídeo, sistemas de radar, etc.) (Azimi e Zhang, 2010).

Neste artigo, discute-se o uso de métodos de clusterização para caracterização do comportamento do tráfego da rodovia BR-290/RS, na região metropolitana da cidade de Porto Alegre – RS. Este estudo busca agregar dados refe-

<sup>1</sup> Felipe Caleffi, Laboratório de Sistema de Transportes, UFRGS. (f.caleffi@yahoo.com)

<sup>2</sup> Shanna Trichês Lucchesi, Laboratório de Sistema de Transportes, UFRGS. (slucchesi@gmail.com)

<sup>3</sup> Michel José Anzanello, Departamento de Engenharia de Produção, UFRGS. (michel.anzanello@gmail.com)

<sup>4</sup> Helena Beatriz Bettella Cybis, Laboratório de Sistema de Transportes, UFRGS. (helenabc@producao.ufrgs.br)

Manuscrito recebido em 29/03/2016 e aprovado para publicação em 11/11/2016.

Este artigo é parte de TRANSPORTES v. 24, n. 4, 2016. ISSN: 2237-1346 (online). DOI:10.14295/transportes.v24i4.1104

rentes a acidentes e condições adversas como chuva e visibilidade às características típicas de tráfego (fluxo e velocidade), para assim determinar padrões de comportamento do tráfego, e caracterizar de forma detalhada as características da rodovia. As técnicas de clusterização *K-Means* e *Fuzzy C-Means* foram empregadas para classificar as características do tráfego da rodovia em estudo. Estes métodos de clusterização são aplicados em um banco de dados contendo os registros das características do tráfego (fluxo, velocidade, visibilidade horizontal, incidência e quantidade de chuva) da rodovia BR-290/RS.

Este artigo está dividido em cinco seções. A seção 2 apresenta os métodos de clusterização aplicados. A seção 3 apresenta uma revisão da literatura, e a seção 4 uma análise da aplicação destes métodos. A seção 5 apresenta as considerações finais.

## 2. MÉTODOS DE CLUSTERIZAÇÃO

Clusterização, ou agrupamento de dados, é a classificação não-supervisionada de padrões (observações, itens de dados, ou vetores de características) em grupos (*clusters*) com características similares. O objetivo da clusterização é atribuir observações multidimensionais para diferentes agrupamentos tais que os pontos dentro de um aglomerado sejam semelhantes e difiram de pontos em outros *clusters*. Os agrupamentos são divididos em duas categorias principais: particionais e hierárquicas. Agrupamento hierárquico utiliza *clusters* previamente estabelecidos para construir os novos. Por outro lado, agrupamentos particionais determinam todos os *clusters* de uma só vez. Duas das mais conhecidas e aplicadas técnicas de agrupamento particionais, são *K-Means* e *Fuzzy C-Means* (Jain *et al.*, 1999). Estas técnicas são apresentadas brevemente nas seções seguintes, e foram utilizadas para análise por apresentarem características de clusterização distintas entre si, permitindo assim realizar diferentes avaliações no banco de dados. Estas técnicas também apresentaram resultados pertinentes para análises na área de transportes, como descrito na seção 3.

### 2.1. O algoritmo *K-Means*

O algoritmo *K-Means* foi desenvolvido por Lloyd (1982). Esta técnica de clusterização é capaz de manipular grandes conjuntos de dados de maneira simples e rápida. *K-means* é uma técnica que usa o algoritmo de agrupamento de dados por *K-médias*. O objetivo deste algoritmo é encontrar a melhor divisão de  $P$  dados em  $K$  grupos  $C_i$ ,  $i = 1, \dots, K$ , de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada.

Tal como muitos outros métodos de agrupamento, *K-Means* requer a definição do número de *clusters* ( $K$ ) a ser formado no momento de inicialização da análise. Determinar o número correto de *clusters* pode ser uma tarefa desafiadora e necessitar de algoritmos específicos a este fim. Após determinar o número de *clusters* ( $K$ ) a serem analisados, o algoritmo define  $K$  pontos aleatórios como os centros de *cluster* (centroides). Em seguida, calcula a distância euclidiana de cada observação para o centroide mais próximo e recalcula os novos centroides. Este processo é repetido

com vistas à minimização de uma função objetivo. O algoritmo é interrompido quando não há mais melhorias na função objetivo (Jain *et al.*, 1999).

A função objetivo da clusterização *K-Means* apoiada na distância euclidiana é expressa na equação (1):

$$\arg_{\min} = \sum_{i=1}^K \sum_{j=1}^{n_i} \|v_j - c_i\|^2 \quad (1)$$

onde: “ $K$ ” é o número de *clusters*; “ $n_i$ ” é o número de pontos (observações) no *cluster*  $i$ ; “ $v_j$ ” é o  $j$ ésimo vetor de observação; e “ $c_i$ ” é o centroide do *cluster*  $i$ .

A cada iteração do algoritmo, o novo centroide do *cluster*  $i$  é calculado por:

$$c_i = \frac{1}{n_j} \sum_{j=1}^{n_j} v_j \quad (2)$$

Baseado na função objetivo, o método minimiza as distâncias para a média do grupo de observações dentro do grupo e, finalmente, retorna o número pré-determinado de *clusters* com seus correspondentes pontos atribuídos.

### 2.2. O algoritmo *Fuzzy C-Means*

O algoritmo *Fuzzy C-Means* (FCM) foi desenvolvido por Dunn (1973), o qual apoia-se nos conceitos da lógica *Fuzzy*. No algoritmo *K-Means*, se atribui grau de pertinência de “1” a um ponto se este pertencer a um *cluster*, e “0” se não pertencer ao *cluster*. No FCM, contudo, cada ponto não pertence totalmente a um único *cluster*, mas apresenta um grau de pertinência a cada um dos grupos (valor que pode variar entre 0 e 1). Portanto, os pontos que estão mais perto do centroide têm um maior grau de pertinência ao *cluster* em comparação aos pontos mais próximos dos limites do *cluster*.

A pertinência total a diferentes *clusters* para cada ponto deve somar 1. O algoritmo inicia a clusterização com um pré-determinado número de *clusters*, que podem ser obtidos de forma semelhante ao método *K-Means*. A função objetivo do FCM é apresentada na equação (3):

$$\text{Arg min} = \sum_{i=1}^K \sum_{j=1}^{n_i} u_{ji}^m \|v_j - c_i\|^2, \quad 1 \leq m \leq \infty, \quad (3)$$

onde: “ $K$ ” é o número de *clusters*; “ $n_i$ ” é o número de pontos (observações) no *cluster*  $i$ ; “ $u_{ji}$ ” é o grau de pertinência de  $v_j$  no *cluster*  $i$ ; “ $m$ ” é qualquer número real maior que 1; “ $v_j$ ” é o  $j$ ésimo vetor de observação; e “ $c_i$ ” é o centroide do *cluster*  $i$ .

## 3. REVISÃO DA LITERATURA

A aplicação de métodos de clusterização para avaliar as características das condições de tráfego em rodovias tem apresentado resultados positivos (Azimi e Zhang, 2010). Contudo, estes estudos não levam em conta eventos como acidentes, chuvas ou baixa visibilidade. O foco destes estudos é a caracterização da rodovia com base apenas em dados de fluxo, velocidade e densidade.

Tabela 1. Fração do banco de dados

Hora do Dia	Visibilidade (metros)	Chuva (milímetros/hora)	Fluxo (veículos/hora)	Velocidade (km/h)	Dia/Noite	Acidente
09:00	10000	0	552	33,21	1	1
08:00	7000	0	480	81,07	1	0
22:20	50	0	4464	67,77	0	0
09:30	3000	9,9	2820	30,60	1	0

Xia e Chen (2007b) aplicaram o método de clusterização *K-Means* em um banco de dados de (velocidade – densidade), de um segmento de freeway, os resultados mostraram que a velocidade tem um impacto maior no resultado dos *clusters*. Xia e Chen (2007a) também aplicaram análises de *cluster* em um banco de dados de fluxo, velocidade e ocupação, para classificar diferentes condições operacionais de fluxo de veículos. Por sua vez, Sun e Zhou (2005) propuseram a aplicação do algoritmo de *K-Means* para classificar um conjunto de dados de tráfego (velocidade – densidade) para dois e três *clusters*, a fim de fornecer uma ferramenta para a determinação de pontos de colapso no fluxo de tráfego. Park (2002) aplicou *Fuzzy C-Means* para desenvolver um método de previsão de volume de tráfego, usando dados de fluxo, velocidade e densidade. O emprego do método FCM visou reduzir o tempo dispendido com análises do banco de dados.

Kianfar e Edara (2013) aplicaram as técnicas de *clusterização K-means* e *General Mixture Model (GMM)* para particionar dados de fluxo, velocidade e ocupação de faixas em regimes de congestionamento e fluxo livre, para melhor entender as características do tráfego. Weijermars e van Berkum (2005) também propuseram uma análise de *cluster*, através do procedimento hierárquico de *Ward*, para determinar padrões de fluxo do tráfego para uma rodovia situada em uma região metropolitana, concluindo que uma pré-classificação do banco de dados em “dias de trabalho (segunda a sexta-feira)” e “finais de semana” melhora substancialmente a análise.

Com propósitos semelhantes, Chunchun *et al.* (2011) propôs um novo algoritmo de clusterização *Fuzzy* para avaliar as características das condições de tráfego de rodovias e encontrar um padrão de distribuição espacial do fluxo de tráfego, enquanto que Montazeri-Gh e Fotouhi (2011), aplicaram *K-Means* buscando reconhecer padrões no fluxo de tráfego. Ambos os estudos demonstraram boa capacidade de segregar em *clusters* os diferentes estados do tráfego (congestionamento e fluxo livre).

Como apontado por Agarwal *et al.* (2005), Gettman *et al.* (2008), Jia *et al.* (2014), e Thakuriah *et al.* (2008), acidentes e condições climáticas adversas afetam as condições operacionais da rodovia, reduzindo as médias de velocidade e a capacidade. Assim, um profundo conhecimento dos impactos destas condições sobre os padrões de tráfego é necessário para estimar as reduções de velocidade e capacidade, e quais são seus efeitos na caracterização das condições de tráfego das rodovias.

#### 4. ANÁLISES DE CLUSTERIZAÇÃO

O banco de dados utilizado para análise é proveniente de sensores instalados no km 96 da rodovia BR-290/RS, que se situa na região metropolitana da cidade de Porto Alegre – RS. Estes sensores fornecem contagens volumétricas e medidas de velocidade. Este trecho de rodovia tem caracte-

rísticas planas e possui três faixas de tráfego com 3,75 metros de largura cada, e um acostamento com 3 metros de largura. Os maiores volumes de tráfego diários deste segmento ocorrem durante os dias úteis. Este banco de dados conta com quatro meses de contagem de dados (de abril a julho). Este período é caracterizado por grandes incidências de neblina, que reduz a visibilidade na rodovia. Este banco de dados possui sete variáveis, como apresentado na Tabela 1.

Dados de fluxo-velocidade foram agrupados em intervalos de cinco minutos. Precipitações de chuva foram segregadas em duas intensidades: chuva leve < 5 mm/h; e chuva forte ≥ 5 mm/h. Dados referentes à visibilidade horizontal e precipitação são fornecidos pela Rede de Meteorologia do Comando da Aeronáutica (Redemet). Estes dados foram coletados no Aeroporto Salgado Filho, situado aproximadamente 2,5 km do ponto de coleta dos contadores de tráfego na rodovia. Dados de acidentes são binários, com “1” para período com ocorrência de acidentes. São computados os acidentes registrados no raio de 1 km em torno do ponto de coleta dos dados de fluxo-velocidade, pois não há registros de acidentes a montante deste raio e, a jusante, acima deste raio as ocorrências não causam influência nos volumes e velocidades no local de aquisição dos dados. O banco de dados também possui uma separação binária entre dia e noite (“1” para o período diurno). O banco de dados é composto por 23101 observações.

Como os dados possuem variáveis binárias (Dia/Noite e Acidentes), e variáveis com diferentes escalas e magnitudes (metros, milímetros, km/h, veículos/hora), os dados foram normalizados para posterior aplicação dos métodos de clusterização.

##### 4.1. Clusterização do banco de dados

A avaliação da qualidade dos agrupamentos gerados utilizou a métrica *Silhouette Index (SI)*; tal indicador também pode ser utilizado para determinar o número recomendado de *clusters* a ser gerado. O SI fornece uma representação gráfica sucinta de quão bem cada objeto se encontra dentro de seu *cluster*, além de ser representado em um intervalo entre -1 e 1 – quanto mais próximo de 1, mas bem inserida a observação ao seu *cluster* de destino (Rousseeuw, 1987). A Figura 1 apresenta o perfil SI para o banco de dados em análise.

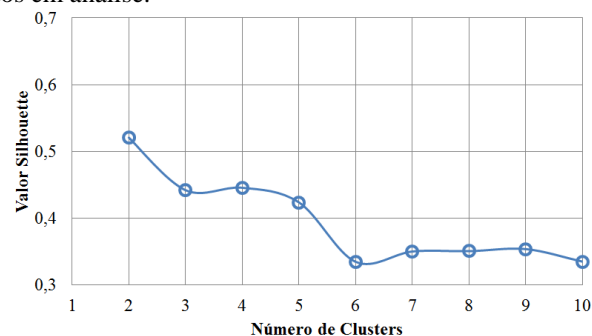


Figura 1. *Silhouette* index para o banco de dados

**Tabela 2.** Descrição dos clusters para *K-Means*

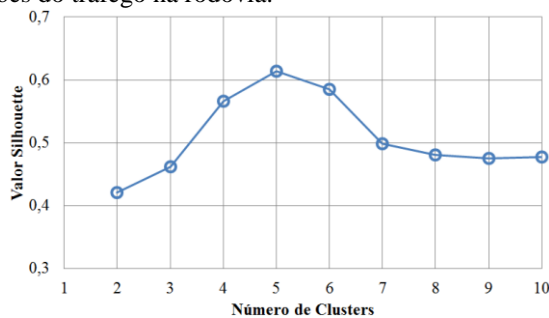
Cluster	Descrição	Observações	Proporção
Cluster 1	Todas as observações em que havia acidentes na rodovia	37	1,05%
Cluster 2	Observações entre (06:30 e 08:55) onde não havia incidência de chuva	2266	64,08%
Cluster 3	Observações entre (09:00 e 10:00) onde não havia incidência de chuva	988	27,94%
Cluster 4	Todas as observações onde havia incidência de chuva leve	188	5,32%
Cluster 5	Todas as observações onde havia incidência de chuva forte	57	1,61%
<b>Total</b>		<b>3536</b>	<b>100,00%</b>

A Figura 1 mostra que o maior valor de *silhouette* ocorre para 2 *clusters*, recomendando a geração deste número de agrupamentos. Analisando o banco de dados após a aplicação do método de *K-Means* para 2 *clusters*, nota-se que houve uma grande separação dos dados apenas em observações diurnas e noturnas, devido ao fato de que a rodovia atinge sua capacidade sempre durante o dia, e que existe pouca incidência de congestionamentos durante o período da noite. Durante a noite, predominam as condições de fluxo livre de veículos, enquanto que no período diurno, é possível notar todas as condições de tráfego (fluxo livre, operação na capacidade, condições de anda/para).

Para o segmento da rodovia estudada no sentido (região metropolitana – Porto Alegre), há apenas um pico de volume, no período da manhã, entre 06:30 e 10:00 horas. Este pico é responsável por aproximadamente 28% do volume diário da rodovia, e é apenas neste período que a rodovia opera na sua capacidade e, conseqüentemente, onde acontecem congestionamentos. Para melhor entender este período crítico, e avaliar se acidentes e condições climáticas adversas têm influência sobre as condições de tráfego da rodovia, o banco de dados foi fracionado, e uma análise foi realizada apenas para referido intervalo (pico da manhã).

**4.2. Análise do pico da manhã**

O fracionamento do banco de dados para o período de pico da manhã (06:30 a 10:00) contemplou 3536 observações. Novamente a métrica SI foi aplicada para determinar o número recomendado de *clusters*; posteriormente foram aplicados os algoritmos *K-Means* e *Fuzzy C-Means*. A Figura 2 apresenta os valores de SI para o novo banco de dados, indicando que o maior valor ocorre para 5 *clusters* (sugerindo a formação de cinco grupos). Na sequência, os agrupamentos foram gerados utilizando os métodos de *K-Means* e *Fuzzy C-Means*, buscando padrões de comportamento do tráfego que possam caracterizar as diferentes condições do tráfego na rodovia.



**Figura 2.** *Silhouette* index para o pico da manhã

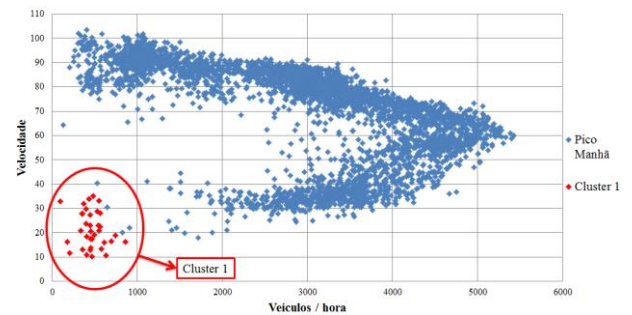
Devido à proximidade dos valores de SI para 4, 5 e 6 *clusters*, análises foram realizadas para as três opções de agrupamentos. Porém, as análises com 4 e 6 grupos apresentaram resultados pouco satisfatórios. Para 4 *clusters*

houve um mau agrupamento das observações, não levando em conta alguma variável ou a magnitudes das variáveis. Já para 6 *clusters*, houve uma subdivisão de *clusters* que visualmente pertenceriam ao mesmo grupo. Portanto, são apresentados os resultados decorrentes da formação de 5 *clusters*.

**4.2.1. Clusterização utilizando K-Means**

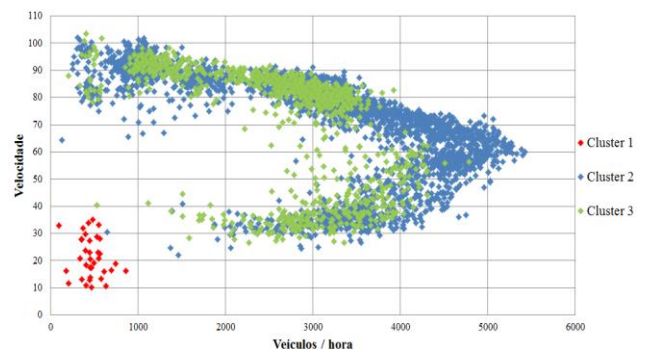
A Tabela 2 apresenta uma descrição de cada um dos cinco *clusters* formados, bem como a proporção de observações em cada *cluster*. A visibilidade horizontal demonstrou não possuir influência na clusterização. Uma posterior análise no banco de dados confirmou que a visibilidade baixa não exerce influência direta nas velocidades e nos fluxos, tão pouco na ocorrência de acidentes. Portanto, no contexto analisado, a visibilidade horizontal é um fator que não altera as condições prévias de tráfego.

Nota-se que no primeiro *cluster* foram agrupadas todas as observações em que havia acidentes na rodovia. Mesmo sendo apenas 1,05% das observações, estes eventos são responsáveis por grandes perturbações no tráfego, uma vez que podem interromper faixas e causar lentidão dos veículos. A Figura 3 apresenta a relação fluxo-velocidade para o pico da manhã, com o Cluster 1 evidenciado.



**Figura 3.** Relação fluxo-velocidade para hora pico e “Cluster 1”

Uma interessante distinção feita pelo método *K-Means* foi a separação em dois *clusters* (“Cluster 2” e “Cluster 3”) da faixa horária quando não havia incidência de chuvas. A Figura 4 apresenta esta separação.



**Figura 4.** Clusters 1, 2 e 3

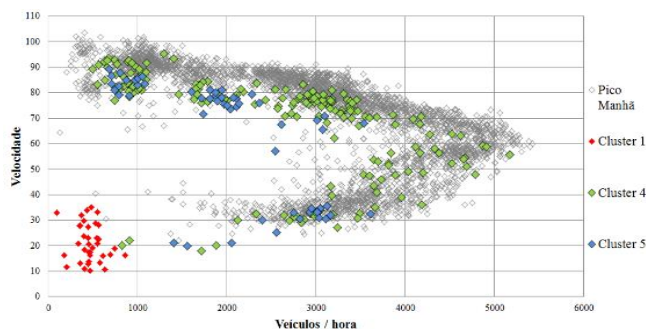


Figura 5: Pico da manhã, “Cluster 1 – acidentes”, “Cluster 4 – chuva leve” e “Cluster 5 – chuva forte”

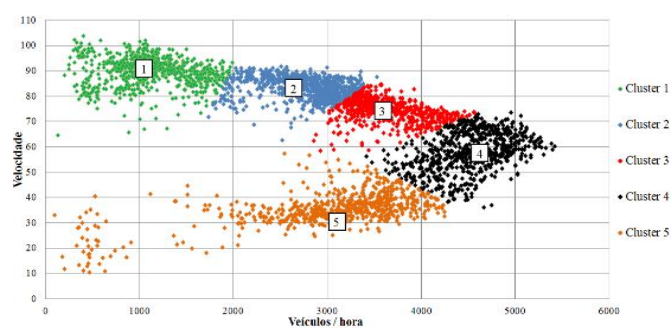


Figura 6: Relação fluxo-velocidade, resultantes da clusterização Fuzzy C-Means

Tabela 3. Descrição dos clusters para Fuzzy C-Means

Cluster	Descrição	Observações	Proporção
Cluster 1	Representa a condição de fluxo livre dos veículos	696	19,68%
Cluster 2	Representa a condição onde os veículos entram em <i>car following</i> , e passam a sofrer influência dos veículos a sua volta	585	16,54%
Cluster 3	Representa a condição que precede a ocorrência de congestionamentos	769	21,75%
Cluster 4	Representa a condição de operação na capacidade	791	22,37%
Cluster 5	Representa a condição de congestionamento	695	19,65%
<b>Total</b>		<b>3536</b>	<b>100,00%</b>

É possível notar que para o “cluster 3” (referente ao período entre 09:00 e 10:00 horas), há uma redução na capacidade da rodovia em cerca de 1000 veíc/h. Contudo, as características do tráfego em fluxo livre e congestionamento são semelhantes para os Clusters 2 e 3. A redução da capacidade neste período é originada por congestionamentos diários que acontecem à jusante do segmento de coleta dos dados, onde a rodovia torna-se um dos principais acessos à cidade de Porto Alegre – RS. Este acesso acaba caracterizando-se como um gargalo, causando essa redução da capacidade à montante que foi revelada pelo Cluster 3.

Outra separação feita pelo algoritmo foi a dos Clusters 4 e 5, referentes às observações onde havia chuva leve e forte, respectivamente. A Figura 5 apresenta a relação total de fluxo-velocidade para o pico, e os Clusters 5 e 6.

A incidência de chuvas leves não altera a capacidade da rodovia, porém é possível notar uma redução nas velocidades médias, principalmente na porção superior da curva. Essa velocidade tem uma redução de cerca de 9 km/h em média. Para a incidência de chuvas fortes, nota-se que não há registros em torno da capacidade, o que pode ser um indicativo de que a capacidade se reduz quando há chuvas com alta intensidade. Os maiores volumes registrados durante chuvas fortes são de 3650 veic/hora. Para estudar a capacidade durante chuvas fortes, faz-se necessário um conjunto maior de registros com estas precipitações. As velocidades médias da porção superior da curva também reduzem, em cerca de 10 km/h. Portanto, é possível concluir que as chuvas (leves ou fortes) reduzem a velocidade média da rodovia nas condições de fluxo livre e nas condições que precedem a ocorrência de congestionamento para o mesmo patamar.

#### 4.2.2. Clusterização utilizando Fuzzy C-Means (FCM)

O FCM foi utilizado para gerar 5 clusters. Embora os dois algoritmos (*K-Means* e FCM) sejam semelhantes quanto ao processo de clusterização, o resultado dos clusters para FCM resultaram em uma caracterização distinta. Para a análise Fuzzy, o agrupamento foi gerado baseado

apenas nas variáveis “Fluxo” e “Velocidade”, ignorando as demais variáveis do banco de dados. Após uma avaliação no banco de dados quanto à correlação entre variáveis, conclui-se que as variáveis “Fluxo” e “Velocidade” possuem um alto índice de correlação – maior do que para as demais variáveis – portanto, este agrupamento baseado em apenas duas variáveis pode ser explicado por esta alta correlação. Esta correlação tem influência significativa na posição dos centroides dos clusters no método Fuzzy, fazendo com que a clusterização, ou grau de pertencimento de cada objeto a um cluster, seja também influenciada por estas variáveis correlacionadas (Dunn, 1973). A Figura 6 apresenta a relação fluxo-velocidade, oriunda da clusterização.

Apesar de o método Fuzzy C-Means não levar em consideração os dados referentes a “Acidentes”, “Chuvas” e “Visibilidade”, devido à grande correlação entre “Fluxo” e “Velocidade”, os resultados oriundos desta análise são pertinentes, e com isso foram mantidos na análise. Através da Tabela 3, é possível notar que a clusterização gerou 5 clusters que representam diferentes condições do tráfego. Interessante notar que a clusterização gerou agrupamentos que seguem uma lógica semelhante aos níveis de serviço do *Highway Capacity Manual* (HCM, 2010), apesar da diferença no número de níveis.

O “Cluster 1” representa a condição de fluxo livre dos veículos, na qual os condutores possuem liberdade espacial na rodovia para imprimir as velocidades por eles desejada. O “Cluster 2”, representa a condição onde os veículos passam a sofrer uma influência dos veículos a sua volta e, por isso, muitas vezes não são mais capazes de imprimir as velocidades desejadas. Desta forma, acontece uma pequena queda nas médias de velocidade, em relação às condições de fluxo livre.

O “Cluster 3” representa as condições que precedem a ocorrência de congestionamentos na rodovia. Nestas condições as velocidades médias diminuem ainda mais (em relação aos Clusters 1 e 2). São nestas condições que técnicas como as de gerenciamento ativo de tráfego entram em operação (Mirshahi *et al.*, 2007), com o objetivo de retardar ou mesmo eliminar a ocorrência de congestionamentos. O “Cluster 4” representa as condições onde a rodovia opera na

sua capacidade. Estas condições representam os momentos que antecedem a ocorrência de fluxos congestionados (Cluster 5), e as velocidades médias passam a reduzir significativamente.

## 5. CONSIDERAÇÕES FINAIS

Dois métodos de clusterização foram aplicados neste estudo, com o objetivo de caracterizar o comportamento do tráfego numa região metropolitana da rodovia BR-290/RS. A contribuição deste estudo se dá no fato de agregar, aos dados tradicionais de fluxo e velocidade, dados relacionados a acidentes e condições adversas como chuva e visibilidade. Foi possível determinar padrões comportamentais influenciados pelas variáveis (fluxo, velocidade, acidentes e chuvas), permitindo caracterizar de forma detalhada as características da rodovia.

Ao aplicar o método de clusterização *K-Means* foi possível classificar satisfatoriamente a rodovia em distintas condições de tráfego. A análise dos resultados das clusterizações permitiu concluir que quando ocorrem acidentes – no raio de 1 km em torno do ponto de coleta dos dados – a rodovia passa a operar em condições diferentes do seu padrão normal. Nestas condições deve-se esperar que as velocidades variem entre 10 e 35 km/h. Em função do bloqueio de faixas e da redução de velocidade natural dos motoristas, a análise indicou que a demanda da rodovia nestas condições não ultrapassa volumes maiores que 1000 veic/h.

Através do *K-Means*, também foi possível determinar diferentes comportamentos do tráfego durante os horários de pico, provando que a técnica de clusterização é capaz de identificar dissimilaridades temporárias numa mesma variável dentro do banco de dados. Além disso, foi possível identificar padrões distintos de comportamento do tráfego quando há ocorrência de precipitações, em especial para altas intensidades de chuva, onde existe um indicativo de redução na capacidade. É possível concluir também que chuvas (leves ou fortes) reduzem as médias de velocidade dos veículos para cerca de 10 km/h. A variável visibilidade horizontal não demonstrou ter nenhum efeito no comportamento do tráfego na rodovia.

Com o *Fuzzy C-Means*, foi possível identificar diferentes características da rodovia, uma vez que a análise resultou em uma caracterização apenas das variáveis de “Fluxo” e “Velocidade” – forçadas pela grande correlação entre estas variáveis. Embora esta técnica não tenha considerado acidentes e chuva, ela forneceu uma caracterização pertinente da rodovia ao representar diferentes condições do tráfego visíveis na curva que relaciona o fluxo com a velocidade. Com esta análise, por exemplo, é possível traçar um paralelo com dados de densidade da rodovia, para buscar determinar diferentes níveis de serviço da rodovia.

Os métodos de clusterização aplicados apresentaram resultados distintos, porém pertinentes. Com isso, destaca-se a importância de utilizar múltiplos métodos em análises de bancos de dados, uma vez que estes métodos podem ser complementares entre si.

## AGRADECIMENTOS

Os autores agradecem o apoio da Concessionária CONCEPA e do REDEMET, que disponibilizaram os dados de tráfego e climáticos.

## REFERÊNCIAS

- Agarwal, M., Maze, T. H., e R. R. Souleyrette (2005) Impacts of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity. Final Technical Report, Center for Transportation Research and Education Iowa State University, USA.
- Azimi, M., e Y. Zhang (2010) Categorizing Freeway Flow Conditions by Using Clustering Methods. Transportation Research Record: Journal of the Transportation Research Board, 2173, 105–114. doi:10.3141/2173-13
- Caceres, N., Romero, L. M., e F. G. Benitez (2012) Estimating Traffic Flow Profiles According to a Relative Attractiveness Factor. Procedia - Social and Behavioral Sciences, 54, 1115–1124. doi:10.1016/j.sbspro.2012.09.826
- Chunchun, H., Nianxue, L., Xiaohong, Y., e S. Wenzhong (2011) Traffic Flow Data Mining and Evaluation Based on Fuzzy Clustering Techniques. International Journal of Fuzzy Systems, 13(4), 344–349.
- Dunn, J. C. (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics, 3(3), 32–57. doi:10.1080/01969727308546046
- Gettman, D., Pu, L., Sayed, T., e S. Shelby (2008) Surrogate Safety Assessment Model and Validation: Final Report. Federal Highway Administration, Report No. FHWA-HRT-08-051.
- HCM, Highway Capacity Manual (2010). 3rd Ed., National Research Council, Transportation Research Board, Washington, D.C.
- Jain, A. K., Murty, M. N., e P. J. Flynn (1999) Data Clustering: A Review. ACM Computing Surveys (CSUR), 31(3), 264–323.
- Jia, Y., Wu, J., Du, Y., e G. Qi (2014) Impacts of rainfall weather on urban traffic in beijing: analysis and modeling. Physics and Society, 1–13.
- Kianfar, J., e P. Edara (2013) A Data Mining Approach to Creating Fundamental Traffic Flow Diagram. Procedia - Social and Behavioral Sciences, 104, 430–439. doi:10.1016/j.sbspro.2013.11.136
- Lloyd, S. (1982) Least squares quantization in PCM. IEEE Transactions on Information Theory, 28(2), 129–137. doi:10.1109/TIT.1982.1056489
- Mirshahi, M., Obenberger, J., Fuhs, C. A., Howard, C. E., Krammes, R. A., Kuhn, B. T., Mayhew, R. M., Moore, M. A., Sahebjam, K., Stone, C. J., e J. L. Yung (2007) Active Traffic Management: The Next Step in Congestion Management. Federal Highway Administration, FHWA-PL-07-012, Washington D.C.
- Montazeri-Gh, M., e A. Fotouhi (2011) Traffic condition recognition using the -means clustering method. Scientia Iranica, 18(4), 930–937. doi:10.1016/j.scient.2011.07.004

Park, B. (2002) Hybrid Neuro-Fuzzy Application in Short-Term Freeway Traffic Volume Forecasting. *Transportation Research Record*, 1802, 190–196. doi:10.3141/1802-21

Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Sun, L., e J. Zhou (2005) Development of Multiregime Speed-Density Relationships by Cluster Analysis. *Transportation Research Record*, 1934, 64–71. doi:10.3141/1934-07

Thakuriah, P., Yanos, G., Lin, J., Metaxatos, P., Sulo, R., Pu, W., e L. Mbekeani (2008) Including Weather Effects In a Real-Time Traffic Information Environment - Exploratory Analysis and Alternative Models. In *Proceedings of the 15th Intelligent Transportation Systems World Congress, New York City - USA*.

Weijermars, W., e E. van Berkum (2005) Analyzing highway flow patterns using cluster analysis. *Proceedings. 2005 IEEE Intelligent Transportation Systems*, 831–836. doi:10.1109/ITSC.2005.1520157

Xia, J., e M. Chen (2007a) A Nested Clustering Technique for Freeway Operating Condition Classification. *Computer-Aided Civil and Infrastructure Engineering*, 22(6), 430–437. doi:DOI: 10.1111/j.1467-8667.2007.00498.x

Xia, J., e M. Chen (2007b) Defining Traffic Flow Phases Using Intelligent Transportation Systems-Generated Data. *Journal of Intelligent Transportation Systems*, 11, 15–24. doi:10.1080/15472450601122322