

CLEVeR: Aprimorando a Eleição de Líderes de *Cluster* em Redes AD Hoc

Vinicius B. Jesus¹, Lucas M. Almeida¹, Wesley F. Santos¹,
Joahannes B. D. da Costa², Geraldo P. Rocha Filho³, Maycon L. M. Peixoto¹

¹Departamento de Ciência da Computação – Universidade Federal da Bahia (UFBA)
Salvador – BA – Brasil

{viniciusbj, lucasmascarenhas, wesleisantos, maycon.leone}@ufba.br

²Universidade Federal de São Paulo (UNIFESP)

joahannes.costa@unifesp.br

³Universidade Estadual do Sudoeste da Bahia (UESB)

geraldrocha@uesb.edu.br

Abstract. *Vehicular Ad Hoc Networks (VANETs) require efficient leader election mechanisms to mitigate communication overhead in highly dynamic environments. Existing strategies typically rely on topological centrality metrics while overlooking mobility stability, which can result in leaders that quickly lose representativeness. This paper presents CLEVeR, a cluster-based predictive leader election model that integrates connectivity and vehicular mobility using a Gradient Boosting Regressor to estimate a leadership score. By replacing purely metric-based selection with supervised inference, the approach produces more stable and representative leaders without introducing significant computational overhead. Simulations conducted in realistic urban scenarios demonstrate that CLEVeR achieves higher leadership scores and reduces mobility instability by more than 40% compared to Betweenness- and Closeness-based methods, while also exhibiting lower variability across executions.*

Resumo. *Redes Veiculares Ad Hoc (VANETs) requerem mecanismos eficientes de eleição de líderes para mitigar a sobrecarga de comunicação em ambientes altamente dinâmicos. Estratégias existentes geralmente se baseiam em métricas topológicas de centralidade, mas negligenciam a estabilidade da mobilidade, o que pode resultar em líderes que rapidamente perdem representatividade. Este artigo apresenta o CLEVeR, um modelo preditivo de eleição de líderes baseado em clusters que integra conectividade e mobilidade veicular por meio de um Gradient Boosting Regressor para estimar um score de liderança. Ao substituir a seleção puramente baseada em métricas por inferência supervisionada, a abordagem produz líderes mais estáveis e representativos sem introduzir sobrecarga computacional significativa. Simulações conduzidas em cenários urbanos realistas demonstram que o CLEVeR alcança scores de liderança superiores e reduz a instabilidade da mobilidade em mais de 40% quando comparado a métodos baseados em Betweenness e Closeness, além de apresentar menor variabilidade entre as execuções.*

1. Introdução

A popularização de veículos conectados tem impulsionado soluções para a gestão de dados em tempo real. Em redes veiculares, ou *Vehicular Ad Hoc Networks* (VANETs), a

comunicação eficiente é essencial para a segurança e a otimização do tráfego urbano. Entretanto, a alta mobilidade dos veículos e a densidade variável da rede impõem desafios como sobrecarga de comunicação, latência e perda de pacotes, especialmente em cenários urbanos dinâmicos. Nesse contexto, estudos recentes destacam que a eficácia de sistemas V2X depende de estratégias de coordenação distribuída e redução de redundância na comunicação [Feng et al. 2024, Peixoto et al. 2023].

Em ambientes urbanos densos, a mobilidade elevada e a conectividade intermitente degradam a qualidade do enlace, resultando em redundância de transmissão, picos de latência e perda de pacotes que comprometem decisões em tempo quase real. A mitigação desses efeitos requer mecanismos de organização local e processamento na borda para reduzir tráfego desnecessário e aliviar gargalos de comunicação [Liu et al. 2021, Yousefpour et al. 2019]. Além disso, a sensibilidade dos sistemas veiculares à dinâmica de mobilidade torna insuficiente considerar apenas a topologia estática, reforçando a necessidade de estratégias que conciliem estabilidade de velocidade e conectividade local para manter a qualidade de serviço [Aburukba et al. 2020, Liu et al. 2018].

Esse cenário tem impulsionado o desenvolvimento de soluções mais adaptativas para a organização da comunicação em VANETs, capazes de responder às variações do ambiente, da densidade de veículos e das condições de mobilidade. Nessa direção, abordagens baseadas em processamento na borda e agrupamento têm mostrado potencial para reduzir a sobrecarga de transmissão [Peixoto et al. 2020], ao mesmo tempo em que a complexidade desse ambiente torna essencial avaliar tais mecanismos em cenários experimentais mais realistas [Araújo et al. 2023]. Esse contexto reforça que a eficiência da comunicação veicular depende não apenas da formação de grupos, mas de outros fatores.

Entre as abordagens propostas, o FogJam [Peixoto et al. 2023] utiliza técnicas de *clustering* para reduzir o volume de dados e detectar congestionamentos em fluxos contínuos, mas depende da transmissão por múltiplos veículos, o que pode gerar redundância e perda de representatividade. O algoritmo LEADER [Ferreira et al. 2024], por sua vez, reduz o tráfego ao eleger um único líder por *cluster* com base em métricas de centralidade, porém desconsidera variáveis da dinâmica veicular, como a estabilidade de velocidade. Assim, apesar dos avanços obtidos, ambas as abordagens negligenciam aspectos ligados à mobilidade, abrindo espaço para soluções que integrem centralidade e variabilidade de velocidade de forma mais equilibrada.

Embora o LEADER [Ferreira et al. 2024] represente um avanço ao reduzir o tráfego por meio da seleção de um único líder por *cluster*, sua estratégia ainda apresenta limitações em redes veiculares dinâmicas. A eleição baseada apenas em métricas topológicas, como *Betweenness* e *Closeness*, identifica nós estruturalmente relevantes, mas ignora a estabilidade de velocidade, essencial para manter a representatividade ao longo do tempo. Em cenários de alta mobilidade, pequenas variações de movimento comprometem a consistência dos dados agregados, tornando a centralidade isolada insuficiente. Diante disso, este artigo propõe uma abordagem baseada em aprendizado de máquina que utiliza *Gradient Boosting* [Friedman 2001] para predizer um score de liderança que integra centralidade e variabilidade de velocidade, explorando informações locais de conectividade, posição relativa e dinâmica veicular.

Em continuidade aos avanços do FogJam [Peixoto et al. 2023] e do LEADER

[Ferreira et al. 2024], este trabalho apresenta o CLEVeR *Cluster-based Leader Election model for VEhicular netwoRks*, que mantém os benefícios da organização em *clusters* e da representatividade na comunicação, ao mesmo tempo em que integra explicitamente a dinâmica de mobilidade ao processo de eleição de líderes. A abordagem utiliza um modelo treinado para equilibrar métricas topológicas e de mobilidade, permitindo escolhas mais consistentes em cenários altamente dinâmicos. Com isso, o CLEVeR constitui um avanço prático sobre soluções existentes, ao reduzir a redundância comunicacional e aumentar a robustez da comunicação em VANETs sem elevar o custo operacional.

2. Trabalhos Relacionados

Geralmente diversas abordagens como [Rui et al. 2018, Mohanty et al. 2019, Peixoto et al. 2021, Feng et al. 2024, Sreenivasamurthy and Obraczka 2024, Kaur et al. 2025, Devi et al. 2025] concentram-se na clusterização de veículos e na eleição de nós representativos para mediar a troca de informações, buscando reduzir sobrecarga de comunicação, latência e perda de pacotes. Estudos recentes exploram a comparação de métricas para seleção de *cluster head* em cenários urbanos dinâmicos, considerando velocidade, estabilidade e posição do centróide [Kaur et al. 2025], bem como o uso de agrupamento com suporte de *fog computing* para aumentar a segurança e a eficiência do roteamento em V2X [Devi et al. 2025].

Nesse contexto, o FogJam [Peixoto et al. 2023] avança na redução de tráfego ao operar na borda com suporte de fog computing e processamento contínuo. A solução combina amostragem ao CDS-DBSCAN, uma adaptação do DBSCAN para fluxos contínuos e agrupamentos sensíveis à densidade espacial e temporal, reduzindo o volume de dados em até 70% sem comprometer a acurácia na detecção de congestionamentos. Ainda assim, a transmissão simultânea de múltiplos veículos por *cluster* mantém sobrecarga residual e pode comprometer a representatividade em cenários com maior dispersão espacial.

O algoritmo LEADER [Ferreira et al. 2024] propõe a eleição de um único representante por *cluster* para comunicação com a infraestrutura, preservando o CDS-DBSCAN na formação dos grupos. A abordagem baseia-se em métricas de centralidade, como *Betweenness* e *Closeness*, alcança redução de tráfego de até 45% e opera em tempo quase real sobre arquiteturas de borda. Contudo, o cálculo contínuo dessas centralidades impõe custo computacional elevado, e a ausência de uma medida explícita de estabilidade de velocidade limita a representatividade sob alta dinâmica de mobilidade. A Tabela 1 apresenta os trabalhos do estado da arte:

Diferentemente das abordagens anteriores, o CLEVeR integra formação de *clusters* com CDS-DBSCAN, eleição de líder e redução de dados em execução na borda e em tempo real. Sua principal inovação é o uso de aprendizado supervisionado para inferir o *score* de liderança a partir de medidas leves de conectividade e vizinhança, incorporando explicitamente a estabilidade de velocidade e evitando o custo de cálculos contínuos de centralidade. Assim, a proposta concilia representatividade estrutural e dinâmica de mobilidade, oferecendo uma solução mais eficiente e abrangente para ambientes veiculares.

3. CLEVeR

O crescimento exponencial do número de veículos conectados tem intensificado os desafios de escalabilidade, confiabilidade e eficiência da comunicação, especialmente em

Tabela 1. Comparação entre o CLEVeR e os trabalhos relacionados. (1) VANETs (2) *Clustering* (3) Centralidade (4) Eleição de Líder (5) Redução de dados (6) *Fog Computing* (7) Tempo real (8) Medidas de Velocidade (9) Aprendizado de Máquina

Referência	1	2	3	4	5	6	7	8	9
[Rui et al. 2018]	✓	✓						✓	
[Mohanty et al. 2019]	✓	✓							
[Peixoto et al. 2021]	✓	✓			✓	✓			
[Feng et al. 2024]	✓	✓				✓			
[Sreenivasamurthy and Obraczka 2024]	✓	✓		✓		✓			
[Devi et al. 2025]	✓	✓				✓			
[Kaur et al. 2025]	✓	✓	✓	✓					✓
[Peixoto et al. 2023]	✓	✓			✓	✓	✓		
[Ferreira et al. 2024]	✓	✓	✓	✓	✓	✓	✓		
CLEVeR	✓	✓	✓	✓	✓	✓	✓	✓	✓

cenários urbanos densos e altamente dinâmicos. Abordagens tradicionais baseadas em comunicação ponto-a-ponto ou em estruturas de controle estáticas tornam-se rapidamente inviáveis, devido ao aumento do tráfego redundante, à sobrecarga de sinalização e à instabilidade causada pela mobilidade dos nós. Nesse contexto, o **CLEVeR *Cluster-based Leader Election model for Vehicular netwoRks*** é proposto como um mecanismo de gerenciamento adaptativo de comunicação, que combina agrupamento dinâmico de veículos com técnicas de aprendizado de máquina para realizar a eleição eficiente e estável de líderes.

3.1. Modelo Preditor: *Gradient Boosting Regressor*

O CLEVeR baseia-se em aprendizado de máquina supervisionado para estimar um score de liderança associado a cada veículo dentro de um cluster. Diferentemente das abordagens tradicionais baseadas em métricas globais de centralidade, como *Betweenness* e *Closeness*, o CLEVeR utiliza um modelo de regressão treinado para capturar padrões locais relacionados à estabilidade e representatividade do líder. O objetivo do modelo é prever, a partir de características locais do veículo e de sua vizinhança, um valor escalar que represente sua adequação para atuar como líder do cluster. O veículo com maior score estimado é então selecionado como líder.

O score de liderança combina métricas topológicas e dinâmicas do veículo, sendo definido como

$$y_i = \alpha \cdot \text{Betweenness}(v_i) + \beta \cdot \text{Closeness}(v_i) + \gamma \cdot \text{RMSE}_v(v_i),$$

onde $\alpha, \beta, \gamma \in \mathbb{R}^+$ são coeficientes de ponderação que controlam a influência relativa de cada métrica. As métricas de *Betweenness* e *Closeness* capturam a representatividade topológica do veículo no grafo de comunicação *intra-cluster*, enquanto $\text{RMSE}_v(v_i)$ representa o erro quadrático médio da velocidade do veículo em relação à média de seus vizinhos, funcionando como um indicador de estabilidade cinemática. Para este trabalho foram definidos para a equação os valores: $\alpha = 0.3$, $\beta = 0.3$ e $\gamma = 0.4$. Essa escolha mantém a importância da centralidade e conexão do veículo, ao passo que busca uma maior estabilidade para o líder do *cluster*.

Cada amostra de treinamento corresponde a um veículo v_i pertencente a um *cluster* C_k durante uma janela temporal Δt . O veículo é descrito por um vetor de características $X_i = [x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}]$, composto por atributos derivados da topologia intra-*cluster*: posição espacial, velocidade, grau de conectividade local e variação de velocidade em relação aos vizinhos. O rótulo associado a cada amostra é o score de liderança y_i , previamente calculado a partir da função de score definida neste trabalho.

O treinamento do GBR ocorre de forma iterativa e segue os seguintes passos. Inicialmente, o modelo é inicializado com uma predição constante $f_0(X)$, geralmente definida como a média dos valores de y_i no conjunto de treinamento. Em seguida, a cada iteração t , um novo regressor fraco $h_t(X)$, tipicamente uma árvore de decisão de baixa profundidade, é ajustado para aproximar o erro residual do modelo acumulado até então. Esse erro é definido como o gradiente negativo da função de perda em relação às predições atuais.

No contexto deste trabalho, a função de perda adotada é o erro quadrático médio (MSE), de modo que, a cada iteração, o modelo busca minimizar a discrepância entre o score real y_i e o score previsto \hat{y}_i . A atualização do modelo ocorre de acordo com $f_t(X) = f_{t-1}(X) + \eta \cdot h_t(X)$, onde $\eta \in (0, 1]$ é a taxa de aprendizado que controla a contribuição de cada regressor fraco para o modelo final.

Esse processo é repetido por um número pré-definido de iterações, resultando em um modelo composto pela soma ponderada de múltiplas árvores de decisão. Como consequência, o GBR é capaz de capturar relações não lineares e interações complexas entre as características dos veículos, mantendo baixo custo computacional durante a fase de inferência.

Após o treinamento, o modelo passa a operar em modo preditivo durante a execução do CLEVeR. Para cada veículo v_i pertencente a um *cluster* ativo, o vetor de características X_i é extraído localmente e fornecido ao modelo treinado, que estima o score de liderança $\hat{y}_i = f(X_i)$. A eleição do líder é então realizada selecionando-se o veículo que maximiza o score previsto dentro de cada *cluster*.

3.2. Agrupamento de Veículos

Cada *cluster* C_i possui um único veículo líder, denotado por R_i , responsável por centralizar a comunicação intra-*cluster* e atuar como intermediário entre os veículos do grupo e a infraestrutura de rede. Essa estratégia reduz transmissões redundantes ao substituir múltiplas comunicações diretas por um fluxo agregado conduzido pelo líder, diminuindo a sobrecarga de comunicação e melhorando a eficiência global da VANET.

A VANET é modelada como um grafo dinâmico, dirigido e ponderado, representado por $G(t) = (V(t), E(t), W(t))$, onde $V(t)$ corresponde ao conjunto de veículos ativos no instante de tempo t , $E(t) \subseteq V(t) \times V(t)$ representa os enlaces de comunicação dirigidos, e $W(t) = \{w_{ij}(t) \in \mathbb{R}^+\}$ é o conjunto de pesos associados às arestas.

Cada peso $w_{ij}(t)$ reflete o custo de comunicação entre os veículos v_i e v_j , sendo influenciado por fatores como a distância euclidiana $d_{ij}(t)$, intensidade do sinal, interferência e condições ambientais. A topologia da rede incorpora ainda um conjunto de interseções viárias $\mathcal{I} = \{i_1, i_2, \dots, i_{|\mathcal{I}|}\}$, que representam pontos estratégicos para roteamento e gerenciamento do tráfego.

A comunicação entre dois veículos ocorre por meio de um caminho dirigido $\mathcal{P}_{ij}(t) = (e_{ij}^{(1)}, e_{ij}^{(2)}, \dots, e_{ij}^{(k)})$, cujo custo total é dado por $W(\mathcal{P}_{ij}) = \sum_{l=1}^k w_{ij}^{(l)}(t)$. Devido à mobilidade veicular e às variações ambientais, o grafo $G(t)$ é intrinsecamente dinâmico, exigindo mecanismos adaptativos capazes de atualizar a topologia em tempo real.

A operação é discretizada em janelas temporais de duração fixa Δt . Ao final de cada intervalo, o conjunto de veículos ativos $V(t)$ é particionado em um conjunto de *clusters* $\mathcal{C}(t) = \{C_1(t), C_2(t), \dots, C_{K(t)}(t)\}$, de modo que $C_i(t) \subseteq V(t)$ e $C_i(t) \cap C_j(t) = \emptyset$ para $i \neq j$. Essa discretização temporal permite que o processo de agrupamento acompanhe a dinâmica da mobilidade veicular, mantendo a consistência dos *clusters* ao longo do tempo.

O agrupamento considera critérios de proximidade geográfica e conectividade local, permitindo que os *clusters* se adaptem dinamicamente às variações de densidade veicular. Para essa tarefa, utiliza-se o algoritmo CDS-DBSCAN (*Connectivity- and Density-Sensitive DBSCAN*) [Peixoto et al. 2023], uma adaptação do DBSCAN clássico projetada especificamente para ambientes veiculares móveis. Diferentemente de abordagens baseadas em partições fixas, o CDS-DBSCAN não requer a definição prévia do número de *clusters* e é capaz de identificar agrupamentos arbitrários, característica essencial em cenários urbanos heterogêneos.

Durante cada janela Δt , os veículos transmitem mensagens *beacon* contendo posição (x_i, y_i) , velocidade v_i e identificador do nó. Essas mensagens são coletadas pela RSU mais próxima e constituem o conjunto de entrada \mathcal{B} do algoritmo de agrupamento. A partir dessas informações, o CDS-DBSCAN avalia a densidade local de veículos considerando um raio de vizinhança ε e um número mínimo de pontos *minPts*, formando *clusters* apenas quando há conectividade suficiente para sustentar a comunicação intra-*cluster*.

O funcionamento do CDS-DBSCAN é detalhado no Algoritmo 1. Para cada *beacon* recebido, o algoritmo identifica o conjunto de vizinhos localizados a uma distância inferior ou igual a ε . Caso o número de vizinhos satisfaça o critério $|\mathcal{N}_\varepsilon(b_i)| \geq \text{minPts}$, um novo *cluster* é iniciado e expandido de forma recursiva, incorporando veículos que atendam ao mesmo critério de densidade. Esse processo é repetido até que todos os *beacons* tenham sido analisados.

Algoritmo 1: CDS-DBSCAN

Entrada: Conjunto de *beacons* \mathcal{B} coletados em Δt
 Raio de vizinhança ε
 Número mínimo de pontos *minPts*
Saída : Conjunto de *clusters* $\mathcal{C}(t)$

- 1 Inicializar $\mathcal{C}(t) \leftarrow \emptyset$
- 2 **foreach** *beacon* $b_i \in \mathcal{B}$ **do**
- 3 Determinar a vizinhança $\mathcal{N}_\varepsilon(b_i) = \{b_j \in \mathcal{B} \mid \text{dist}(b_i, b_j) \leq \varepsilon\}$
- 4 **if** $|\mathcal{N}_\varepsilon(b_i)| \geq \text{minPts}$ **then**
- 5 Criar novo *cluster* C_k
- 6 Expandir C_k adicionando recursivamente os elementos de $\mathcal{N}_\varepsilon(b_i)$
- 7 Atualizar $\mathcal{C}(t) \leftarrow \mathcal{C}(t) \cup \{C_k\}$
- 8 **end**
- 9 **end**

Veículos que não satisfazem o critério de densidade são classificados como ruído

e não pertencem a nenhum *cluster*. Essa característica torna o CDS-DBSCAN particularmente adequado para ambientes urbanos, nos quais a densidade veicular pode variar significativamente ao longo do espaço e do tempo. Ao produzir *clusters* coesos e adaptativos, essa etapa estabelece a base para a eleição eficiente de líderes, abordada na subseção seguinte. A Figura 1 sintetiza o fluxo completo do CLEVeR, desde a coleta de mensagens *beacon*, passando pelo agrupamento baseado em densidade, até a extração das métricas utilizadas na predição do score de liderança e na eleição do líder de cada *cluster*.

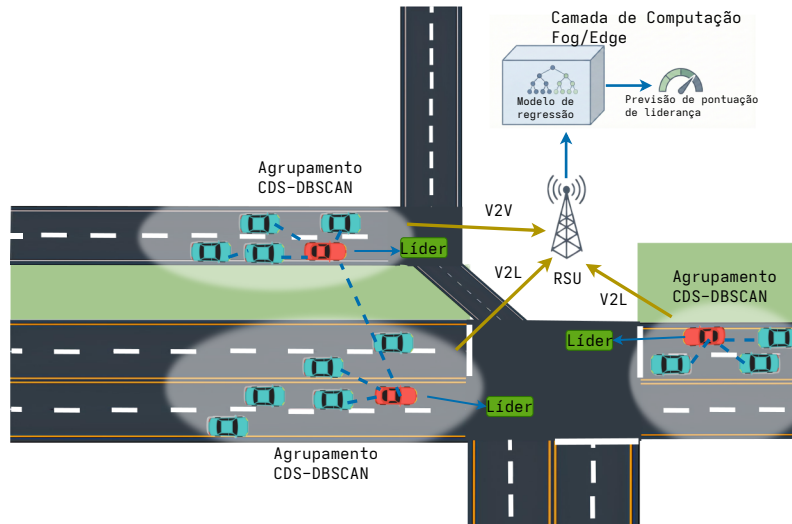


Figura 1. Representação esquemática do processo de agrupamento e da eleição de líderes em redes veiculares.

3.3. Formulação do Problema e Eleição de Líder

Após a formação dos *clusters*, a comunicação na VANET passa a ser organizada de forma hierárquica em dois níveis complementares. A comunicação veículo-para-veículo (V2V) ocorre exclusivamente no interior de cada *cluster*, enquanto a comunicação veículo-para-infraestrutura (V2I) é mediada por um único veículo líder. Essa organização tem como objetivo reduzir o volume de transmissões redundantes e centralizar a interação com a infraestrutura, aumentando a eficiência e a escalabilidade da rede.

Nesse contexto, o problema central abordado pelo CLEVeR consiste em selecionar, para cada *cluster* C_i , um único veículo $R_i \in C_i$ que seja capaz de representar o grupo de forma eficiente e estável. Essa seleção deve levar em consideração tanto a posição topológica do veículo no grafo de comunicação quanto sua estabilidade cinemática ao longo do tempo.

Formalmente, a eleição do líder é formulada como um problema de maximização de um score de liderança. Para cada veículo $v_j \in C_i$, define-se um score y_j , e o líder do *cluster* é determinado como $R_i = \arg \max_{v_j \in C_i} y_j$.

A partir dessa formulação, a eleição do líder é realizada conforme descrito no Algoritmo 2. Para cada *cluster* ativo, o score de liderança é estimado para todos os veículos pertencentes ao grupo pela predição realizada pelo modelo de aprendizado supervisionado. O veículo que apresenta o maior score é selecionado como líder, e apenas ele permanece responsável pela comunicação V2I durante a janela temporal corrente.

Algoritmo 2: Eleição de Líder

Entrada: Conjunto de clusters $\mathcal{C}(t)$
Saída : Conjunto de líderes $\mathcal{R}(t)$

- 1 Inicializar $\mathcal{R}(t) \leftarrow \emptyset$
- 2 **foreach** cluster $C_i \in \mathcal{C}(t)$ **do**
- 3 **foreach** veículo $v_j \in C_i$ **do**
- 4 Estimar score $y_j = f(X_j)$
- 5 **end**
- 6 Selecionar líder $R_i = \arg \max_{v_j \in C_i} y_j$
- 7 Atualizar $\mathcal{R}(t) \leftarrow \mathcal{R}(t) \cup \{R_i\}$
- 8 **end**

3.4. Exemplo Ilustrativo

A Figura 2 apresenta um exemplo ilustrativo composto por sete veículos que formam um *cluster* veicular. Cada veículo possui uma velocidade distinta, variando entre valores baixos, médios e altos, o que cria um ambiente heterogêneo semelhante ao observado em situações reais de tráfego. Nesse contexto, a tarefa consiste em escolher um líder capaz de representar o *cluster* e para realizar a comunicação com a RSU mais próxima. Esse exemplo destaca o impacto que diferenças de velocidade exercem na dinâmica interna do *cluster*, uma vez que veículos muito rápidos ou muito lentos tendem a gerar maior instabilidade na coesão do grupo. Assim, o cenário sintetiza o desafio central do processo de eleição de líder ao evidenciar como a variabilidade de mobilidade influencia a representatividade e a estabilidade necessárias para selecionar o veículo que servirá como ponto de agregação e encaminhamento de dados.

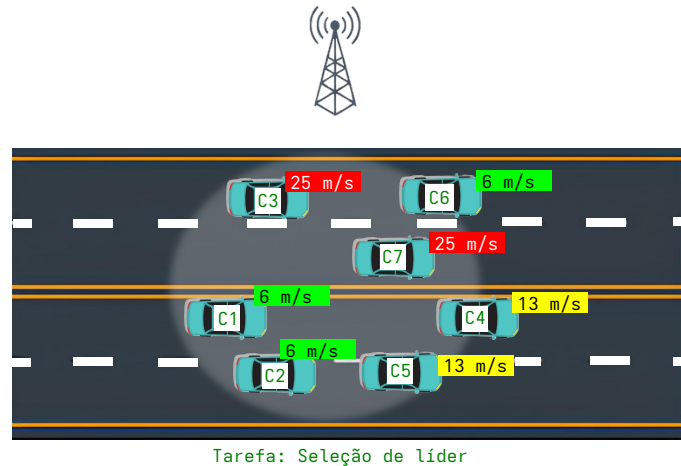


Figura 2. Exemplo ilustrativo de um *cluster* veicular composto por sete veículos com velocidades distintas, ilustrando o desafio da seleção de um líder para comunicação com a RSU.

O cenário inicial mostrado na Figura 2 serve como ponto de partida para observar como diferentes algoritmos selecionam o líder, enquanto as Figuras 3 e 4 mostram o passo seguinte após essa escolha. Na Figura 3a, o algoritmo *Closeness*, seleciona o veículo C7 por estar mais ao centro do *cluster*, embora ele apresente velocidade acima da média e tenda a se afastar rapidamente do grupo, tornando-se uma escolha instável. Já o

algoritmo *Betweenness*, Figura 3b, elege o veículo C1 por ser o nó com maior conectividade interna, mas sua velocidade muito inferior à dos demais faz com que ele fique para trás com o avanço do tráfego, reduzindo sua capacidade de representar o *cluster* de forma consistente. Esses comportamentos evidenciam como métodos puramente topológicos ignoram a dinâmica de mobilidade e podem selecionar líderes que se descolam do grupo logo após a eleição.

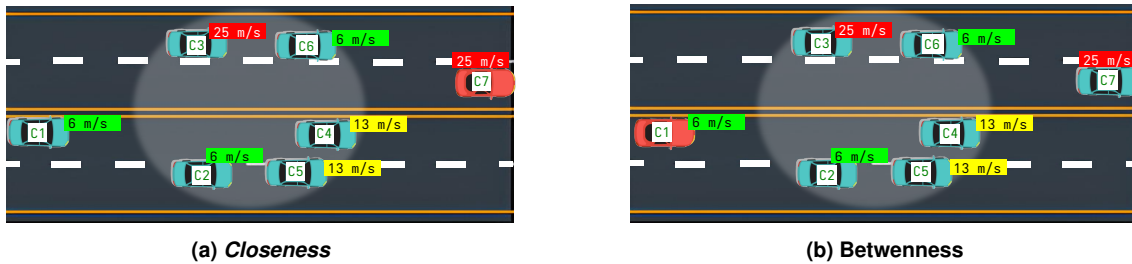


Figura 3. Seleção de líderes baseada em métricas topológicas. (a) *Closeness* escolhe o veículo central C7 e (b) *Betweenness* seleciona o veículo C1 com maior conectividade, ambos resultando em líderes que se afastam do *cluster* devido à discrepância de velocidade.

A Figura 4 destaca como o CLEVeR supera as limitações observadas nas abordagens puramente topológicas ao selecionar como líder o veículo C4, que apresenta o melhor score conforme a função definida na Seção 3.3. Essa pontuação já integra as métricas de *Betweenness* e *Closeness* com a estabilidade de velocidade representada pelo RMSE, o que permite identificar um veículo que está bem posicionado no *cluster* e ao mesmo tempo mantém um comportamento de mobilidade mais previsível. Ao escolher um líder que combina representatividade e estabilidade, o CLEVeR evita tanto o afastamento rápido observado no *Closeness* quanto o atraso causado pelas baixas velocidades do *Betweenness*. Esse resultado mostra que considerar também a dinâmica de mobilidade torna a escolha mais estável e permite selecionar líderes que permanecem representativos. Com isso, é notável que a escolha baseada no score preditivo oferece uma solução mais robusta e alinhada ao comportamento real do tráfego veicular.

4. Configuração dos Experimentos

Para validar a abordagem proposta, foi desenvolvido um modelo de simulação de tráfego online utilizando uma arquitetura distribuída integrada às ferramentas VEINS¹ (v5.1), SUMO² (v0.32) e OMNet++³ (v5.3). O ambiente de simulação adotado segue a mesma configuração utilizada no algoritmo LEADER, permitindo uma comparação direta entre os métodos. O cenário foi modelado a partir de dados do OpenStreetMap, convertidos para o formato do SUMO, garantindo uma simulação realista da mobilidade veicular em um ambiente urbano.

A simulação foi conduzida em um modelo baseado na Grade de Manhattan, representando um tráfego congestionado com múltiplas faixas e comunicação entre veículos e infraestrutura (V2I) a uma taxa de 1 Hz. Para testar a escalabilidade da abordagem, foram

¹<https://veins.car2x.org/>

²<https://sumo.dlr.de/docs/>

³<https://omnetpp.org/>

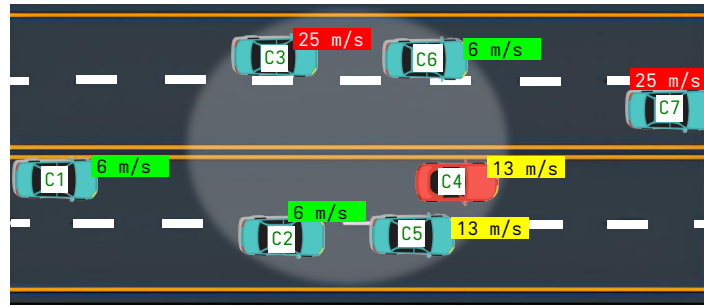


Figura 4. CLEVeR: escolhe o veículo C4 como líder ao combinar métricas topológicas e estabilidade de mobilidade no cálculo do score de liderança.

utilizadas duas configurações de densidade veicular: 400 e 800 veículos. Além disso, foram testados dois intervalos de comunicação entre os veículos: 5 segundos e 10 segundos, buscando avaliar como diferentes frequências de troca de mensagens impactam a escolha do líder. Os experimentos foram replicados dez vezes com diferentes sementes aleatórias, garantindo uma análise estatisticamente válida dos resultados. Para validar a robustez das métricas, foram calculados intervalos de confiança sobre as réplicas experimentais, permitindo avaliar a estabilidade da abordagem proposta.

Na fase de experimentação, os algoritmos foram avaliados quanto à redução de dados, eficiência da comunicação e precisão na eleição do líder. O modelo de regressão foi implementado utilizando GradientBoostingRegressor do Scikit-Learn⁴ (v1.5.2). O treinamento do modelo seguiu um fluxo estruturado para garantir a extração eficiente das características dos veículos e a correta definição do score de liderança. Primeiramente, os dados coletados da simulação foram pré-processados, onde cada veículo dentro de um *cluster* teve suas features extraídas com base em sua posição e velocidade. Então foi calculado o score de liderança para cada veículo, sendo utilizado como a variável alvo no treinamento do modelo.

Para garantir uma avaliação justa do modelo, o conjunto de dados foi dividido em duas partes: 80% dos dados foram utilizados para o treinamento, enquanto os 20% restantes foram reservados para teste. Essa divisão permitiu que o modelo aprendesse os padrões subjacentes na fase de treinamento, enquanto sua capacidade de generalização foi validada em um conjunto de dados não utilizado anteriormente.

5. Resultados

Nesta seção, são apresentados os resultados da avaliação de desempenho da abordagem proposta, comparando-a com estratégias tradicionais de eleição de líderes baseadas em métricas de centralidade (*Betweenness* e *Closeness*). As simulações foram conduzidas em cenários urbanos realistas, variando a densidade veicular (400 e 800 veículos) e os intervalos de disseminação de mensagens de controle (5 e 10 segundos). A análise foca inicialmente na qualidade da liderança eleita, avaliada através do *Score* médio e da estabilidade da velocidade (RMSE), seguida pela validação da precisão do modelo de regressão através do coeficiente de determinação (R^2). Os dados obtidos buscam demonstrar a capacidade do CLEVeR em selecionar líderes mais robustos e representativos para a rede.

⁴<https://scikit-learn.org/1.5/>

Embora os experimentos tenham sido conduzidos considerando dois intervalos distintos de comunicação entre veículos (5 s e 10 s), não foram observadas diferenças estatisticamente relevantes para nenhuma das abordagens avaliadas. Em ambos os intervalos, os resultados permaneceram praticamente inalterados, indicando que a frequência de envio de mensagens de controle, não influencia de forma significativa a qualidade da liderança eleita nem a estabilidade de mobilidade dos líderes. Esse comportamento sugere que o processo de eleição de líderes é dominado pela estrutura dos *clusters* e pela dinâmica espacial dos veículos, sendo pouco sensível a variações na periodicidade de comunicação. Dessa forma, a Figura 5 apresenta a comparação das médias do *Score* (Figura 5a) e do RMSE (Figura 5b), variando apenas a densidade veicular (D.400 e D.800) e considerando três abordagens: (i) líder escolhido pelo CLEVeR, (ii) líder com maior *Betweenness* e (iii) líder com maior *Closeness*.

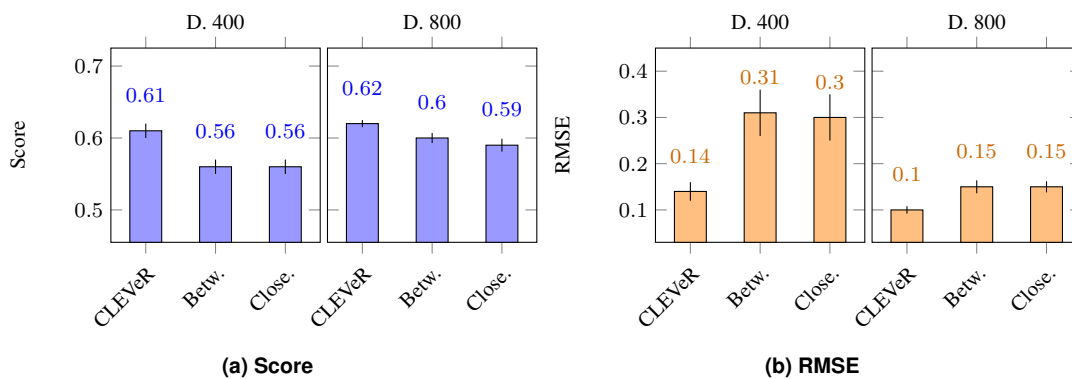


Figura 5. Comparação de desempenho para (a) Score e (b) RMSE para algoritmos de eleição de líder com densidades de 400 e 800 veículos.

Os resultados da Figura 5a indicam que os líderes eleitos pelo modelo apresentam scores médios superiores aos líderes baseados exclusivamente em centralidade. Em D.400, o CLEVeR alcança score médio de aproximadamente 0.61, enquanto as abordagens baseadas em *Betweenness* e *Closeness* atingem cerca de 0.56. Em D.800, observa-se comportamento semelhante, com o modelo obtendo aproximadamente 0.62, frente a valores entre 0.59 e 0.60 para as abordagens de centralidade. Além disso, as barras de erro associadas ao CLEVeR são sistematicamente menores, especialmente no cenário de maior densidade, indicando menor dispersão entre as execuções e maior consistência na escolha dos líderes. Isso sugere que o modelo consegue balancear melhor a representatividade da liderança, selecionando veículos que não apenas têm boa posição na rede, mas também possuem mobilidade mais estável.

A Figura 5b compara a média do RMSE por *cluster* entre o modelo proposto, o líder *Betweenness* e o líder *Closeness*. O objetivo dessa métrica é avaliar a estabilidade das escolhas de líderes, verificando se o modelo reduz a variabilidade dos scores dentro dos *clusters*. Verifica-se que o modelo obtém menores RMSEs em comparação com as abordagens baseadas em centralidade. Em D.400, o CLEVeR apresenta RMSE médio de aproximadamente 0.14, enquanto *Betweenness* e *Closeness* atingem cerca de 0.31 e 0.30, respectivamente. Em D.800, o modelo mantém vantagem com RMSE em torno de 0.10, frente a valores próximos de 0.15 para ambas as abordagens de centralidade. Além disso, as barras de erro do CLEVeR tendem a ser menores, especialmente em D.800, sugerindo menor dispersão entre execuções e maior consistência das escolhas. Isso indica

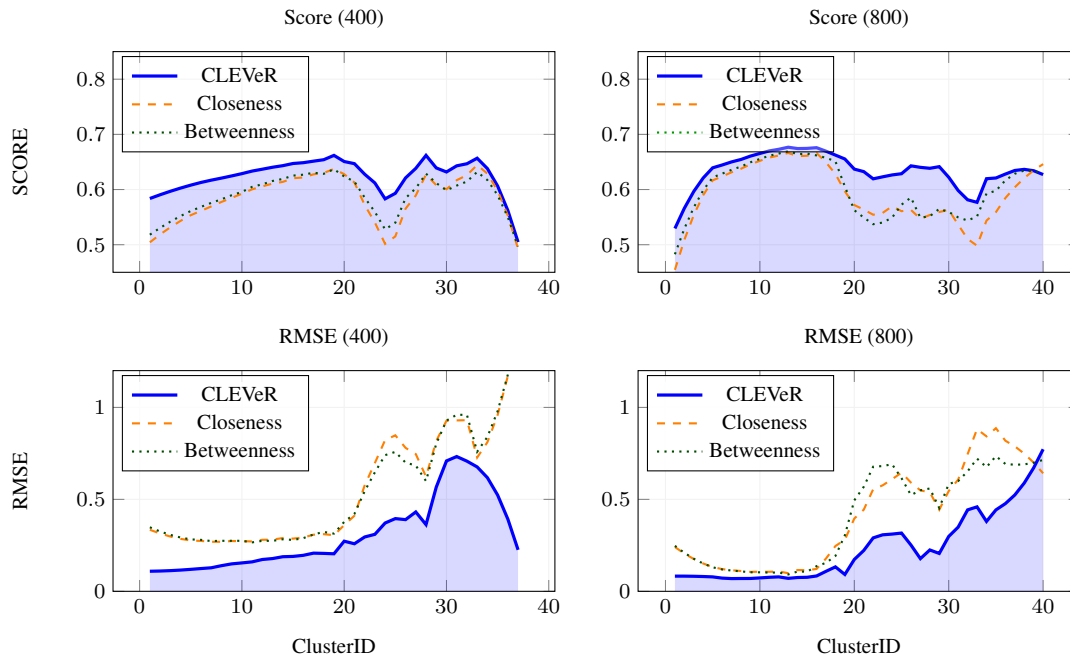


Figura 6. Comparação entre CLEVeR, *Closeness* e *Betweenness* para Score e RMSE nos cenários 400 e 800.

que suas previsões são mais consistentes e menos sensíveis a variações extremas nos dados, reforçando a confiabilidade do modelo na eleição do líder.

A Figura 6 apresenta uma comparação detalhada entre os métodos CLEVeR, *Closeness* e *Betweenness* nos cenários 400 e 800, considerando as métricas de *Score* e *RMSE* ao longo dos diferentes *ClusterIDs*. Para reduzir a influência de flutuações locais e ruído experimental inerentes ao processo de seleção de líderes, todas as séries foram suavizadas utilizando o filtro de Savitzky–Golay.

No que se refere à métrica de *Score*, observa-se que o CLEVeR apresenta desempenho superior ou equivalente aos métodos baseados em centralidade, tanto no cenário 400 quanto no cenário 800 (Figura 6, painéis *Score*–400 e *Score*–800). A suavização evidencia regiões de estabilidade prolongada, nas quais o CLEVeR mantém valores elevados de score com menor variabilidade intercluster. Esse comportamento indica que o método é menos sensível a perturbações topológicas locais, refletindo uma estratégia de seleção de líderes mais robusta e coerente ao longo da progressão dos *clusters*. O efeito do aumento de escala do cenário (de 400 para 800) mostra que os métodos de *Closeness* e *Betweenness* apresentam oscilações mais pronunciadas e quedas localizadas de desempenho, o CLEVeR mantém uma tendência global mais suave e monotônica, sugerindo melhor capacidade de adaptação a cenários com maior complexidade estrutural e heterogeneidade.

A análise da métrica de *RMSE* (Figura 6, painéis *RMSE*–400 e *RMSE*–800) reforça essas observações. As séries mostram que o erro associado ao CLEVeR cresce de forma mais gradual ao longo dos *ClusterIDs*, enquanto os métodos baseados em centralidade exibem aumentos mais abruptos e maior variância. De forma geral, a utilização de séries suavizadas não altera as conclusões quantitativas obtidas a partir dos dados ori-

ginais, mas desempenha um papel crucial na interpretação qualitativa dos resultados. A suavização torna explícita a estabilidade do CLEVeR em diferentes escalas e métricas, reforçando sua superioridade em termos de robustez, previsibilidade e consistência ao longo dos cenários avaliados.

6. Conclusão

Este trabalho apresentou o *CLEVeR*, um modelo de eleição de líderes baseado em *clusters* para redes veiculares que introduz inteligência preditiva no processo de decisão, avançando além de abordagens tradicionais fundamentadas exclusivamente em métricas topológicas. A principal contribuição reside em demonstrar que a qualidade da liderança em VANETs não é apenas um problema estrutural, mas também um fenômeno fortemente dependente da dinâmica de mobilidade, e que ambos podem ser tratados de forma integrada por meio de aprendizado supervisionado. Ao incorporar explicitamente a estabilidade de velocidade na função de liderança, o CLEVeR produz escolhas mais consistentes ao longo do tempo, reduzindo oscilações na representatividade dos *clusters* e tornando a comunicação mais previsível sob alta mobilidade. Os resultados experimentais confirmam essa mudança de paradigma, evidenciando ganhos sistemáticos em qualidade de liderança e estabilidade, com reduções de instabilidade superiores a 40% em cenários densos, sem introduzir sobrecarga computacional ou penalizar a eficiência de comunicação. Em síntese, o CLEVeR estabelece um novo ponto de equilíbrio entre representatividade estrutural e estabilidade cinemática, oferecendo uma base sólida para mecanismos de coordenação veicular mais robustos, escaláveis e alinhados às exigências de sistemas V2X em tempo real.

Disponibilidade de Artefatos

Em aderência aos princípios da Ciência Aberta, o código-fonte e o dataset utilizados neste trabalho podem ser acessados em: <https://github.com/viniciusbj/CLEVeR>.

Agradecimentos

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Foi também apoiado em parte pela bolsa FAPESB e pelo grant FAPESB INCITE PIE0002/2022, e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil, grant #403231/2023-0.

Referências

- Aburukba, H., Al-Jaroodi, J. Z., and Zomaya, A. Y. (2020). Scheduling internet of things requests to minimize latency in hybrid fog–cloud computing. *Future Generation Computer Systems*, 111:539–551.
- Araújo, G. B., Peixoto, M. L. M., and Sampaio, L. (2023). A comprehensive and configurable simulation environment for supporting vehicular named-data networking applications. *Computer Networks*, 235:109949.
- Devi, A., Kait, R., and Ranga, V. (2025). Secure and efficient routing in fog-enabled VANETs: A clustering-based approach. *International Journal of Current Science Research and Review*. Accessed: 2026-01-15. Open access PDF.

- Feng, Z., Li, K., and Li, B. (2024). A spectral clustering-based deployment strategy for roadside units in vehicular edge computing environments. *Ad Hoc Networks*, 158:103483.
- Ferreira, J. V., Freire, M., Cruz, E., Prazeres, C., Figueiredo, G. B., and Peixoto, M. (2024). Leading the way: Reducing network traffic in vehicular ad hoc networks through cluster leader algorithms. available at ssrn: <https://ssrn.com/abstract=4937250> or <http://dx.doi.org/10.2139/ssrn.4937250>. *Social Science Research Network*.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.
- Kaur, G., Khurana, M., and Kaur, A. (2025). Cluster head selection metric in VANETs under dynamic environmental conditions. *Journal of Transformative Technologies and Sustainable Development*, 9:Article 5.
- Liu, L., Chen, C., Pei, Q., Maharjan, S., and Zhang, Y. (2021). Vehicular edge computing and networking: A survey. *Mob. Netw. Appl.*, 26(3):1145–1168.
- Liu, L., Chen, C., Qiu, T., Zhang, M., Li, S., and Zhou, B. (2018). A data dissemination scheme based on clustering and probabilistic broadcasting in vanets. *Vehicular Communications*, 13:78–88.
- Mohanty, A., Mahapatra, S., and Bhanja, U. (2019). Traffic congestion detection in a city using clustering techniques in vanets. *Indonesian Journal of Electrical Engineering and Computer Science*, 13(2):884–891.
- Peixoto, M., Mota, E., Maia, A., Lobato, W., Salahuddin, M., Boutaba, R., and Villas, L. (2023). Fogjam: A fog service for detecting traffic congestion in a continuous data stream vanet. *Ad Hoc Networks*, 140:103046.
- Peixoto, M. L. M., Cruz, E. M., Maia, A. H. O., Santos, M. C. A., Lobato, W., and Villas, L. A. (2020). Exploiting fog computing with an adapted dbSCAN for traffic congestion detection system. In *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pages 1–5.
- Peixoto, M. L. M., Maia, A. H., Mota, E., Rangel, E., Costa, D. G., Turgut, D., and Villas, L. A. (2021). A traffic data clustering framework based on fog computing for vanets. *Vehicular Communications*, 31:100370.
- Rui, L., Zhang, Y., Huang, H., and Qiu, X. (2018). A new traffic congestion detection and quantification method based on comprehensive fuzzy assessment in vanet. *KSII Transactions on Internet & Information Systems*, 12(1).
- Sreenivasamurthy, S. and Obraczka, K. (2024). Clustering at the edge: Load balancing and energy efficiency for the IoT. *Ad Hoc Networks*, 156:103433.
- Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., Kong, J., and Jue, J. P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. *Journal of Systems Architecture*, 98(December 2018):289–330.