

DEEP Q-NETWORK BASED ROUTING IN A SIMPLIFIED VEHICULAR NETWORK: A PROOF-OF-CONCEPT STUDY

ALUNAS: ANA CECÍLIA SILVEIRA FERNANDES
BIANCA SABRINA DE CÁSSIA DA SILVA



AGENDA

Motivação

Trabalhos Relacionados

Solução Proposta

Resultados

Conclusão



A network diagram on the left side of the slide. It features several colored nodes (red, green, blue) connected by black lines, representing a network topology. The nodes are arranged in a way that suggests a dynamic and interconnected system.

MOTIVAÇÃO

- **Alta mobilidade e variação topológica** tornam o roteamento em VANETs um desafio.
- **Protocolos tradicionais (AODV, DSR)** não se adaptam bem a mudanças rápidas na rede.
- **Deep Reinforcement Learning (DRL)** oferece adaptação dinâmica e tomada de decisão inteligente.

A network diagram on the left side of the slide. It features several colored nodes (red, green, blue) connected by black lines, representing a complex network structure. The nodes are arranged in a way that suggests a central hub or a complex web of connections.

MOTIVAÇÃO

- Estudos existentes usam **simulações complexas e de alto custo**, dificultando a análise do aprendizado.
- Este trabalho propõe um **ambiente simplificado e interpretável** para estudar o comportamento de um agente **DQN** em roteamento veicular dinâmico.
- Objetivo: **entender, visualizar e aprimorar** o processo de aprendizado em redes veiculares.

Ref.	Method	Objective	Environment
[10]	DQN	Delay-aware adaptive routing	SUMO + NS-3
[3]	Deep RL	Delay-tolerant routing under mobility	IoV topology
[12]	Double DQN	Stable relay selection and link adaptation	Urban simulation
[13]	Dueling DQN	Reduce overestimation in heterogeneous VANETs	Hybrid model
This work	DQN	Learn optimal path in simplified VANET	Custom environment

*TRABALHOS
RELACIONADOS*

SOLUÇÃO PROPOSTA

- Função Q-ótima:

$$Q^*(s, a) = \mathbb{E} \left[r_t + \gamma \max_{a'} Q^*(s', a') \right]$$

- Estima o retorno esperado ao executar a ação a no estado s ;
- r_t : recompensa imediata;
- γ : fator de desconto (importância das recompensas futuras).

- Função de perda:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')} [(y_t - Q(s, a; \theta))^2]$$

- Mede o erro entre o valor alvo y_t e o valor previsto pela rede;
- Usada para **ajustar os pesos** da rede via **gradiente descendente**;
- Permite ao agente **aprender políticas de roteamento mais eficientes**.

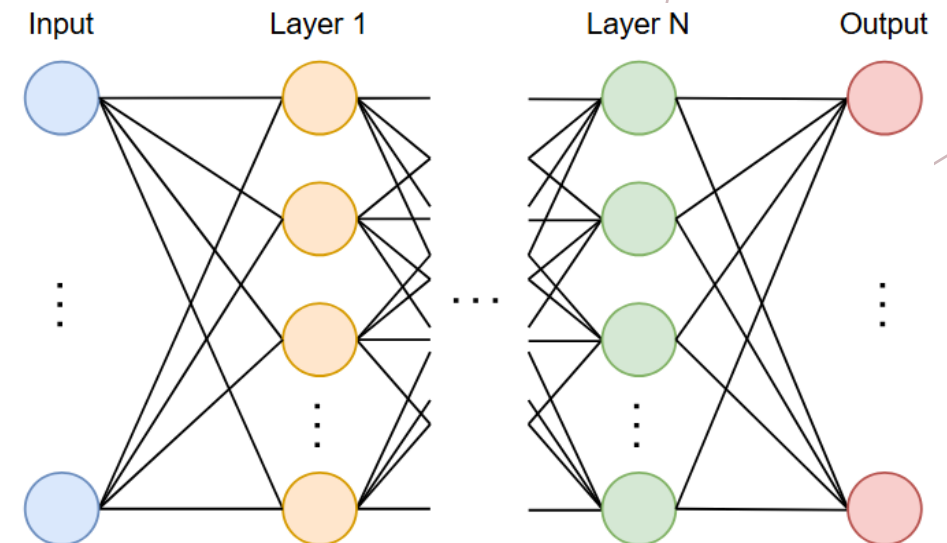
SOLUÇÃO PROPOSTA

Rede totalmente conectada (Fully Connected NN):

- Camada de entrada: representa o estado atual do nó;
- 2 camadas ocultas com **64 neurônios** e ativação **ReLU**;
- Camada de saída: representa todas as ações de roteamento possíveis.

Treinamento:

- Otimiza os pesos usando o **erro quadrático médio (MSE)**;
- Utiliza **target network** para maior estabilidade no aprendizado;
- Amostras são retiradas da **memória de replay** para evitar correlação.



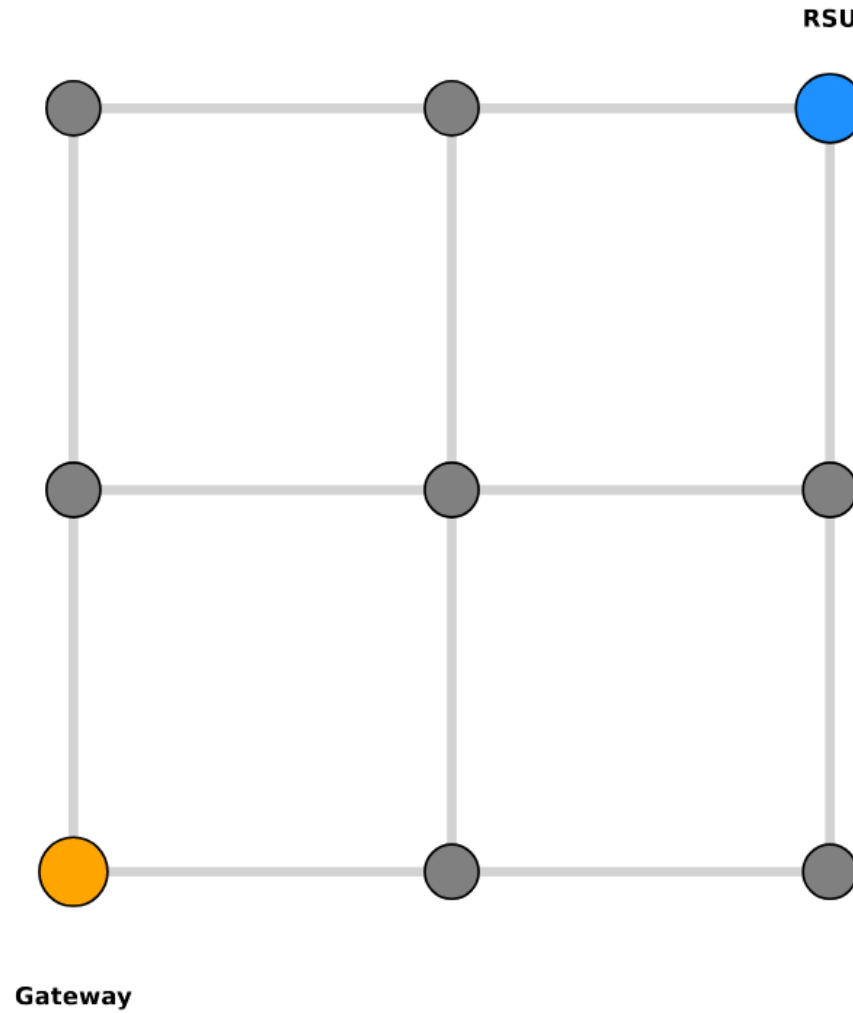
SOLUÇÃO PROPOSTA

$$a_t = \begin{cases} \text{ação aleatória,} & \text{com probabilidade } \epsilon \\ \arg \max_a Q(s_t, a; \theta), & \text{caso contrário} \end{cases}$$

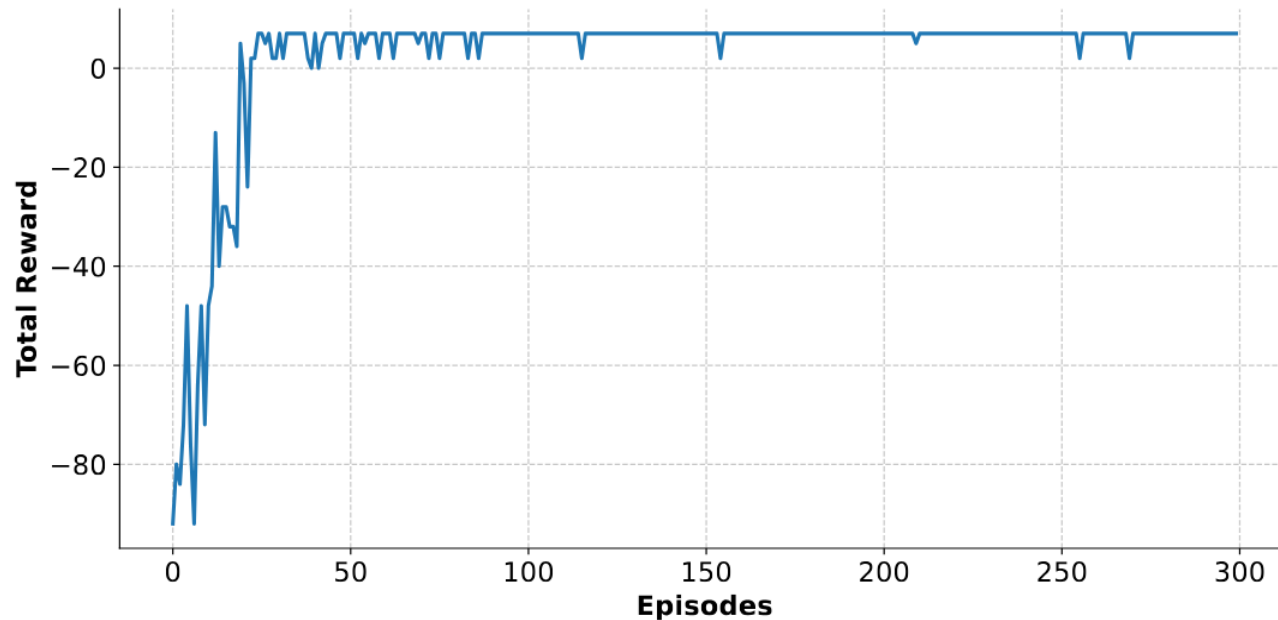
- Garante o **equilíbrio entre exploração e exploração**;
- O agente escolhe **ações aleatórias** no início (alta ϵ);
- ϵ **decai ao longo dos episódios**, favorecendo decisões aprendidas;
- Essencial para **evitar mínimos locais** e melhorar a **eficiência da política**.

SOLUÇÃO PROPOSTA

$$r_t = \begin{cases} +10, & \text{if } v_t = v_d \\ -5, & \text{if } (v_t, a_t) \notin E \\ -1, & \text{otherwise.} \end{cases}$$

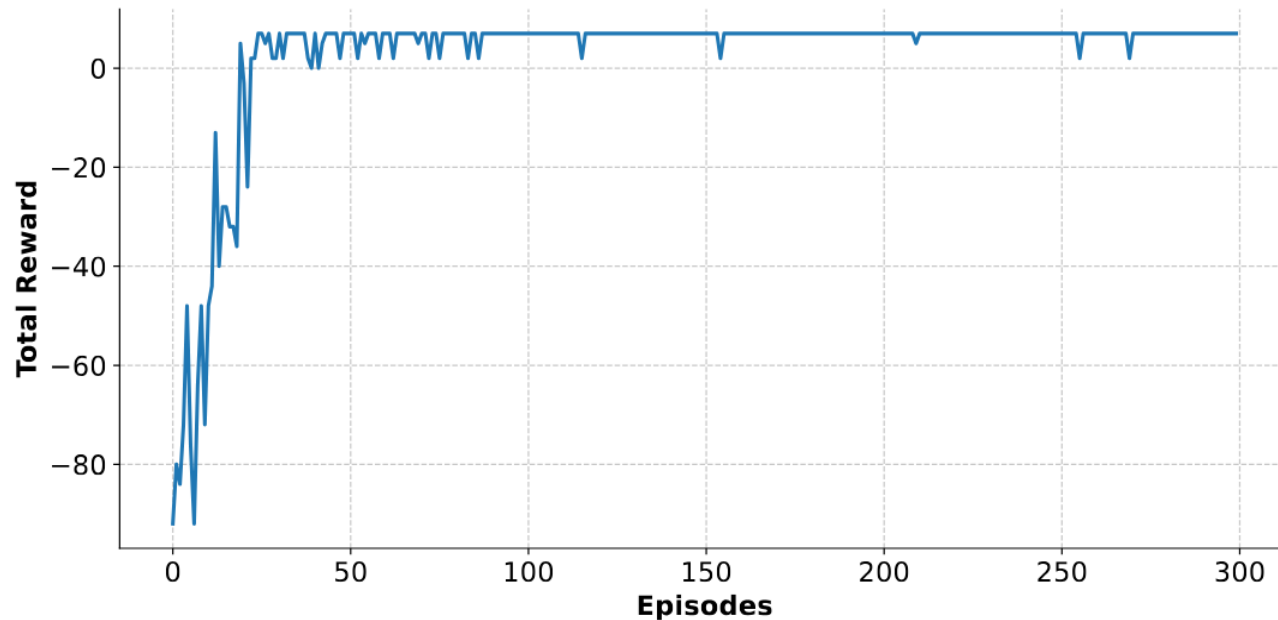


RESULTADOS

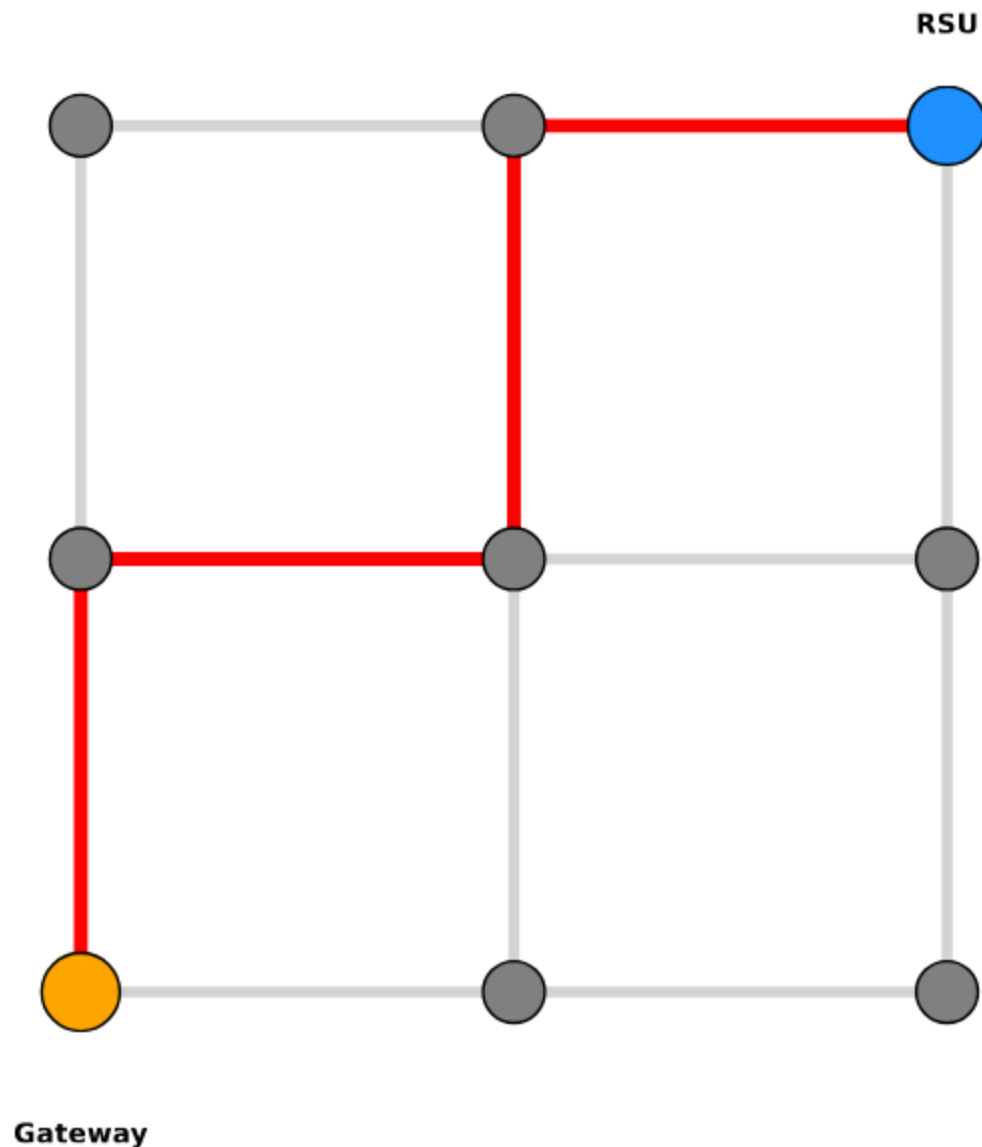


- Inicialmente, o agente apresenta **recompensas baixas ou negativas** devido à alta taxa de exploração ($\epsilon \approx 1,0$);
- À medida que o treinamento avança, o valor de **ϵ diminui**, reduzindo a aleatoriedade das ações;
- O agente passa a **explorar o conhecimento adquirido**, resultando em aumento gradual das recompensas;

RESULTADOS

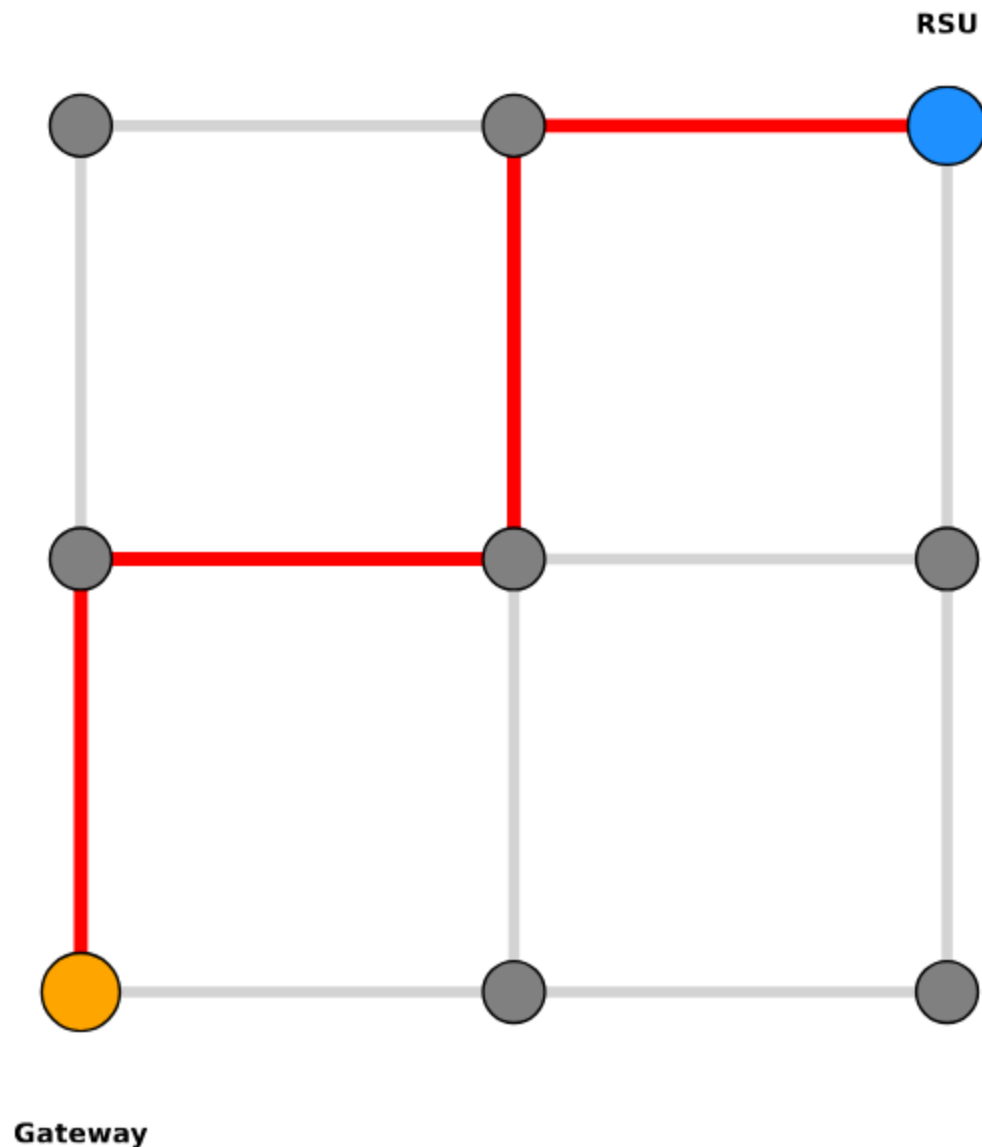


- Após cerca de **50 episódios**, a curva se estabiliza, indicando uma **política próxima do ótimo**;
- Pequenas oscilações refletem a **natureza estocástica** da política ϵ -greedy e do **replay buffer**;
- O padrão de convergência confirma o **equilíbrio entre exploração e exploração** e a **estabilidade do aprendizado**.



RESULTADOS

- Avaliação realizada no ambiente **MiniCity 3×3** após o treinamento;
- O agente DQN convergiu para o **caminho mais curto e confiável** entre o Gateway e a RSU;
- Caminho ótimo aprendido:
0–3–4–7–8;



RESULTADOS

- Preferência explicada pela **maior conectividade dos nós intermediários**, reduzindo transições inválidas;
- A estratégia foi aprendida **sem conhecimento prévio da topologia**, apenas pela interação com o ambiente;
- Demonstra a **capacidade do DQN em aprender políticas ótimas e robustas de roteamento**.

CONCLUSÃO E TRABALHOS FUTUROS

Conclusão:

- Framework proposto: **roteamento adaptativo com agente DQN** em ambiente leve baseado em grafos;
- **MiniVANET** permite aprendizado autônomo de rotas sem conhecimento prévio da topologia;
- **DQN alcançou convergência estável** e encontrou consistentemente o caminho ótimo;
- Ambiente simplificado garante **baixo custo computacional, reprodutibilidade e interpretabilidade**.

Trabalhos Futuros:

- Expandir para **redes maiores e mais realistas**, com mobilidade e variação de enlaces;
- Testar **arquiteturas avançadas**: Double/Dueling DQN e **Graph Neural Networks**;
- Explorar **coordenação multiagente** e **otimização entre camadas** (cross-layer);
- Aplicar o método em cenários **V2X** e **comunicações veiculares 6G**.



OBRIGADA!

ana.fernandes@mtel.inatel.br

bianca.sabrina@dtel.inatel.br