



Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Escola de Informática & Computação

Curso Superior de Tecnologia em Sistemas para Internet (CST-SI)

Aprendizado de Representações em Grafos por Redes Neurais: uma Análise Comparativa

Aluno: Augusto José Moreira da Fonseca

Professor orientador: Eduardo Bezerra, D.Sc.



Sumário

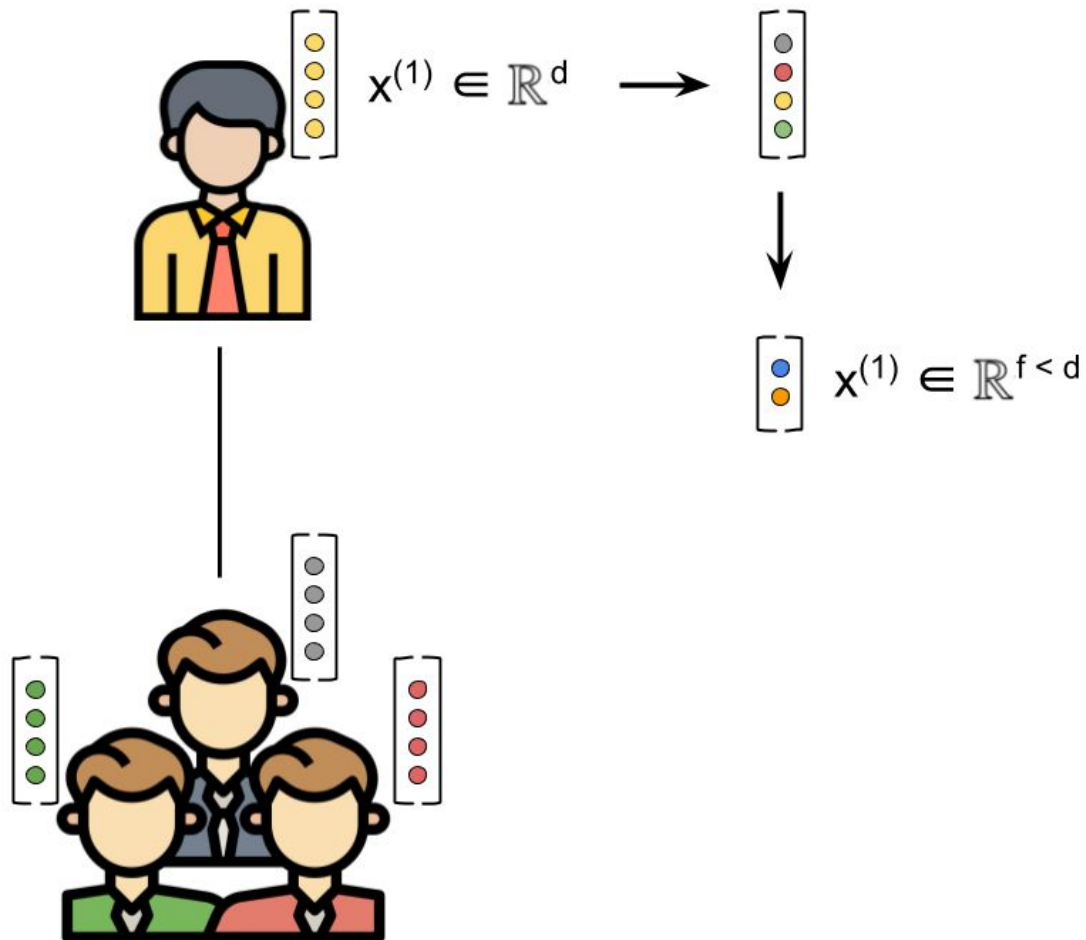
1. Motivação
2. Objetivos
3. Fundamentação teórica
4. Trabalhos relacionados
5. Metodologia
6. Avaliação Experimental
7. Análise retrospectiva e trabalhos futuros



Motivação

- Grafo como representação de objetos e relações
- Aplicação de redes neurais a dados estruturados em grafos
- *Graph Neural Network* (GNN)
 - Característica principal: aprendizado de representações.
- O que são essas representações?

Representações



Motivação

- Representações servem como entrada para métodos de AM
- GNNs em geral demandam altos custos computacionais em termos de tempo de treinamento e consumo de memória



Processo recursivo



Armazena várias instâncias de representações para um mesmo objeto

- Grafos de grandes proporções em termos de número de vértices e arestas podem tornar-se inviáveis para treinamento em GNNs



Motivação

Se eu usar um método de aprendizado de representações que não leva em consideração as relações, a perda no resultado final é relevante?

Motivação

- *Autoencoder Neural Network* (AENN) como arquitetura para o aprendizado de representações



Processo não recursivo



Armazena em memória uma única instância de representação por objeto



Não incorpora as informações dos objetos relacionados



Motivação

Usar uma GNN deve resultar em maior poder preditivo do que usar uma AENN. Mas a que custo?

É possível determinar quando uma arquitetura torna-se mais adequada que a outra?



Objetivos

- Realizar uma análise experimental comparativa utilizando as arquiteturas GNN e AENN no aprendizado de representações, em tarefas de classificação.
- Explorar o trade-off entre custo computacional e poder preditivo entre as arquiteturas citadas.
- Análise comparativa em termos de tempo de treinamento e inferência, consumo de memória e poder preditivo.

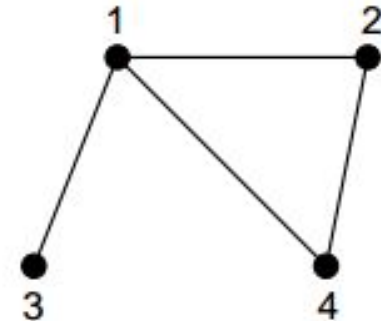
Grafos

Representação: $G = (V, E)$

$V: \{1, 2, 3, 4\}$

$E: \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 4\}\}$

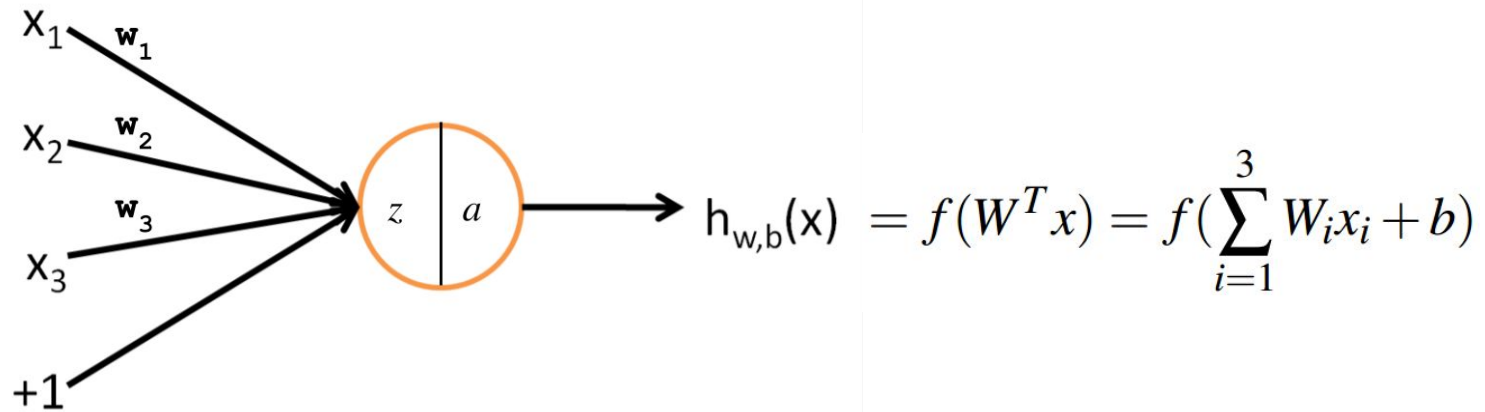
- Adjacência
- Matriz de Adjacência A
 - Dimensão $n \times n$, $n = |V|$
 - $a_{i,j} = 1$, se existe aresta entre os vértices v_i e v_j , 0 caso contrário
- Variações (Multigrafos, Grafos dirigidos)



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

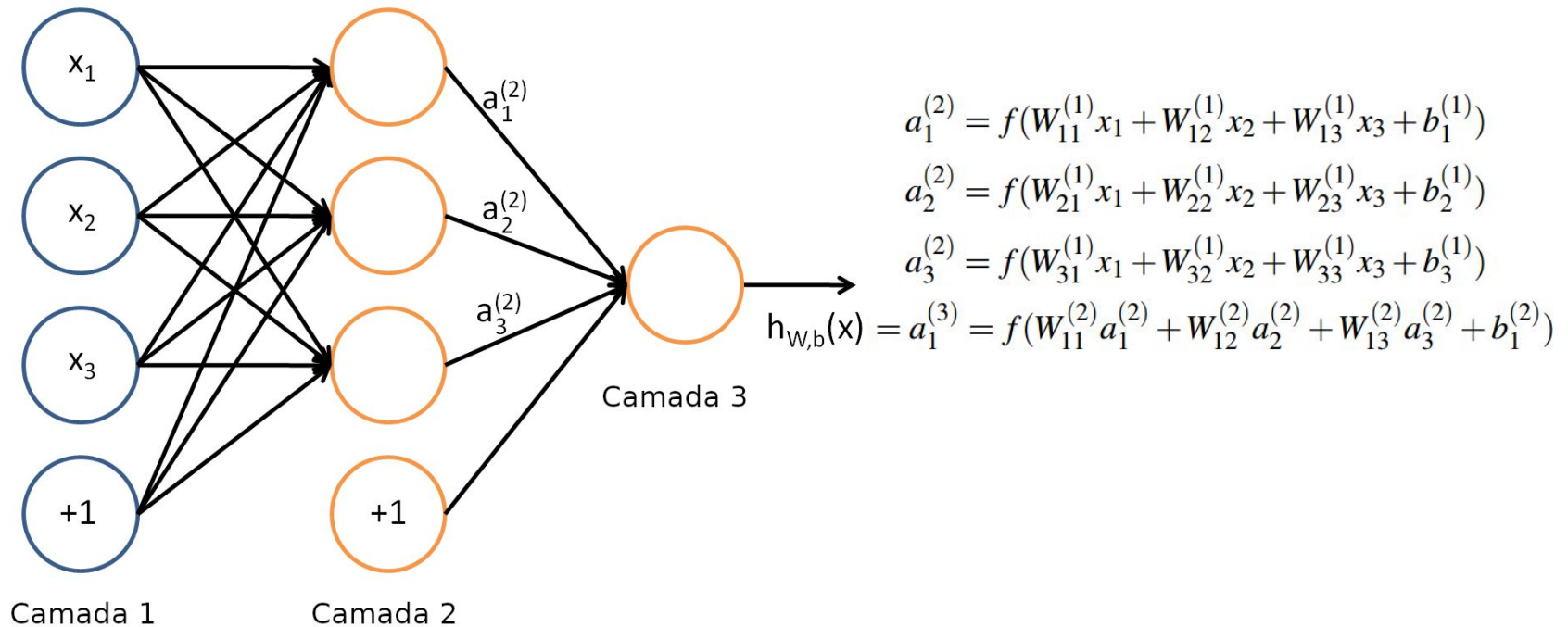
Perceptron

- Modelo matemático inspirado no neurônio biológico
- Proposto por Rosenblatt em 1958
- Capaz de aprender funções lineares



Multilayer Perceptron

- Capaz de aprender funções não lineares



Aprendizado com retropropagação

- O treinamento consiste em aprender os valores (W, b) de forma que $h(x^{(i)})$ seja aproximada para $y^{(i)}$ (treinamento supervisionado)

FUNÇÃO DE CUSTO

$$J(W, b; x, y) = \frac{1}{2} \|h_{W,b}(x) - y\|^2 \quad (1)$$

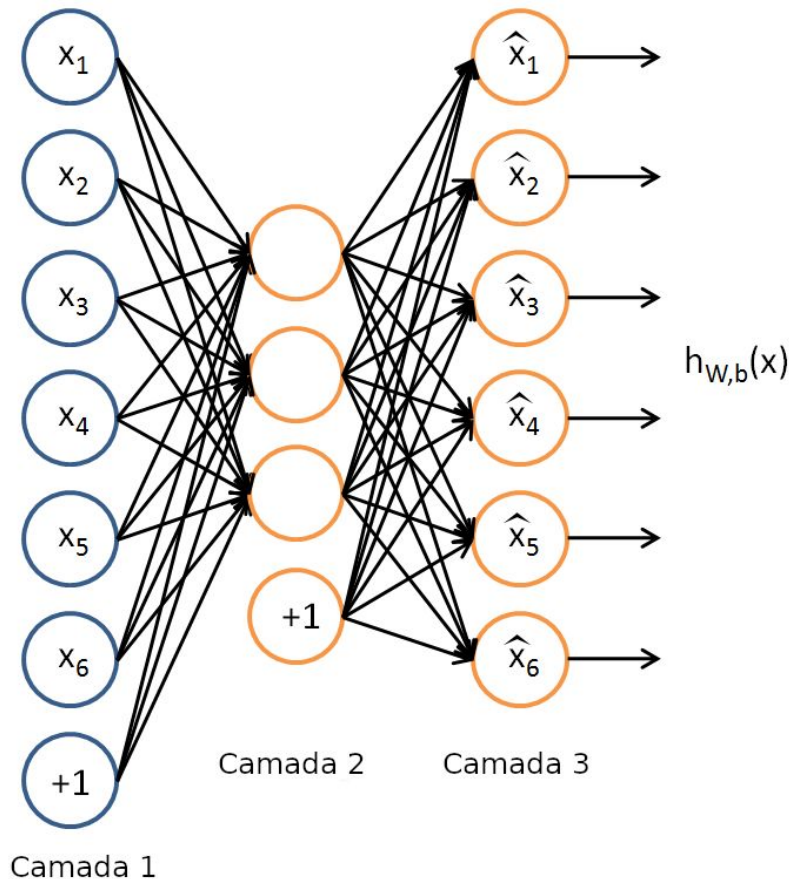
$$J(W, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right) \quad (2)$$

ATUALIZAÇÃO

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W, b) \quad (3)$$

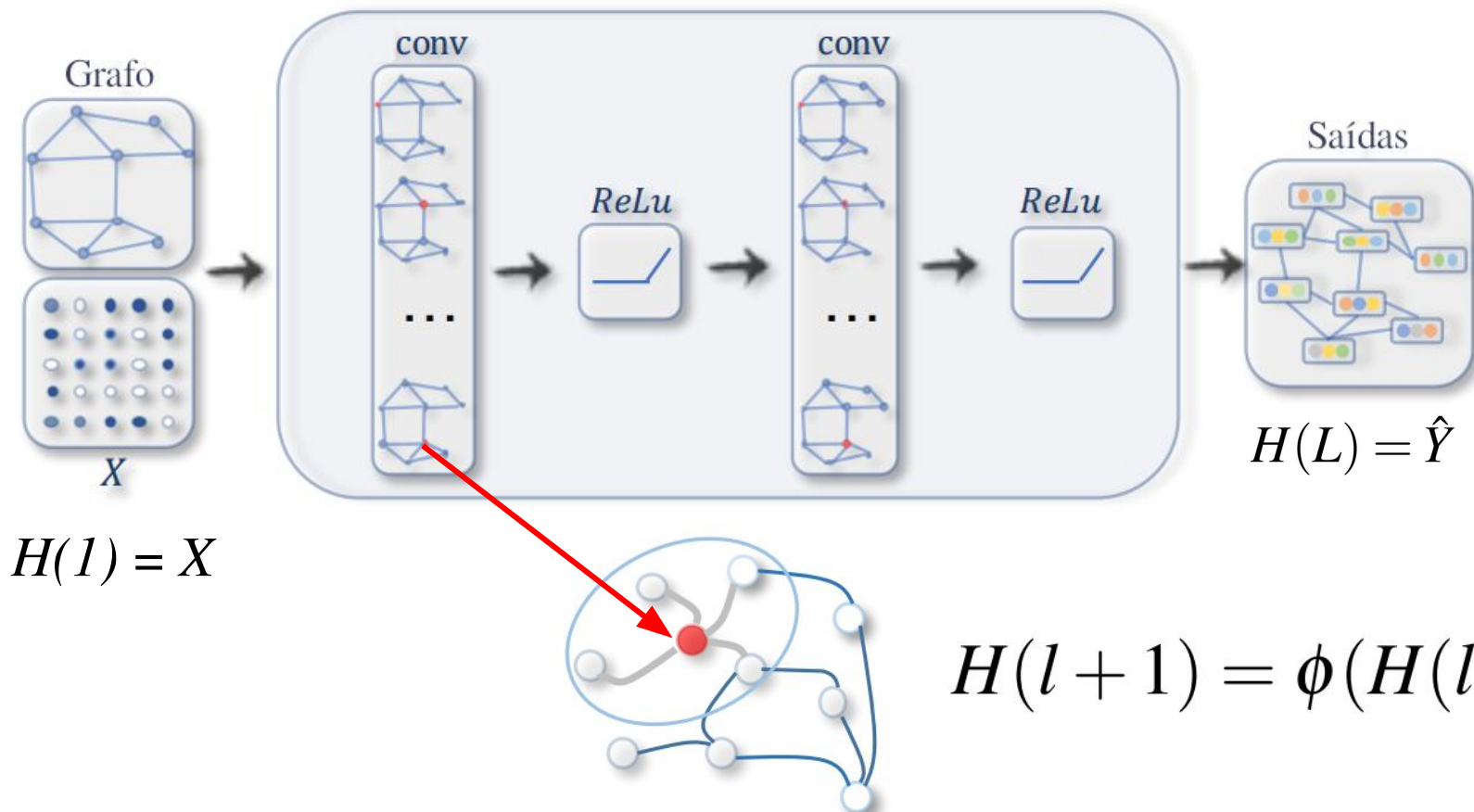
$$b_i^{(l)} = b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W, b),$$

Autoencoder Neural Network



- Arquitetura de rede neural que pode ser treinada para aprender representações
- Treinamento não supervisionado. Aplica retropropagação utilizando $\hat{y}^{(i)} = x^{(i)}$
- $a^{(2)}$ como representação de baixa dimensionalidade

Aprendizado de representações em grafos

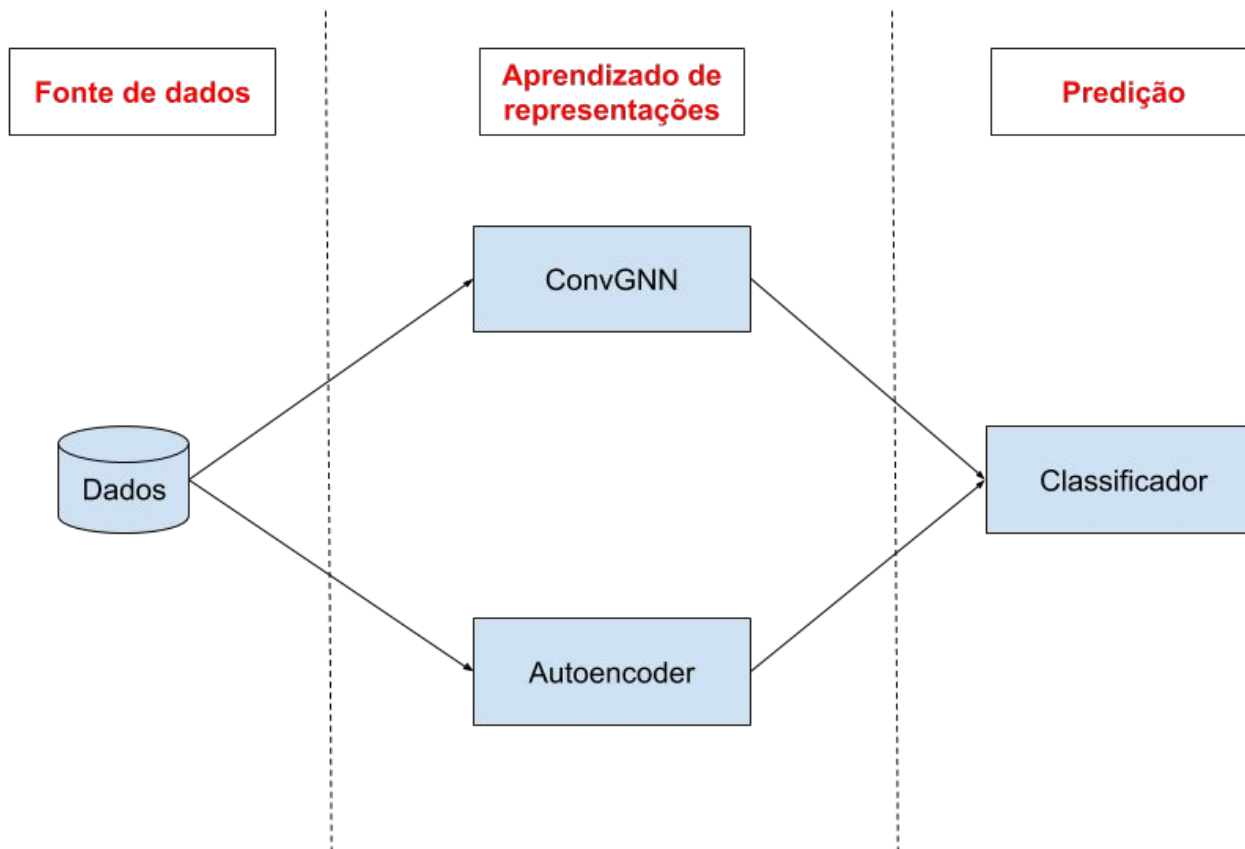




Trabalhos relacionados

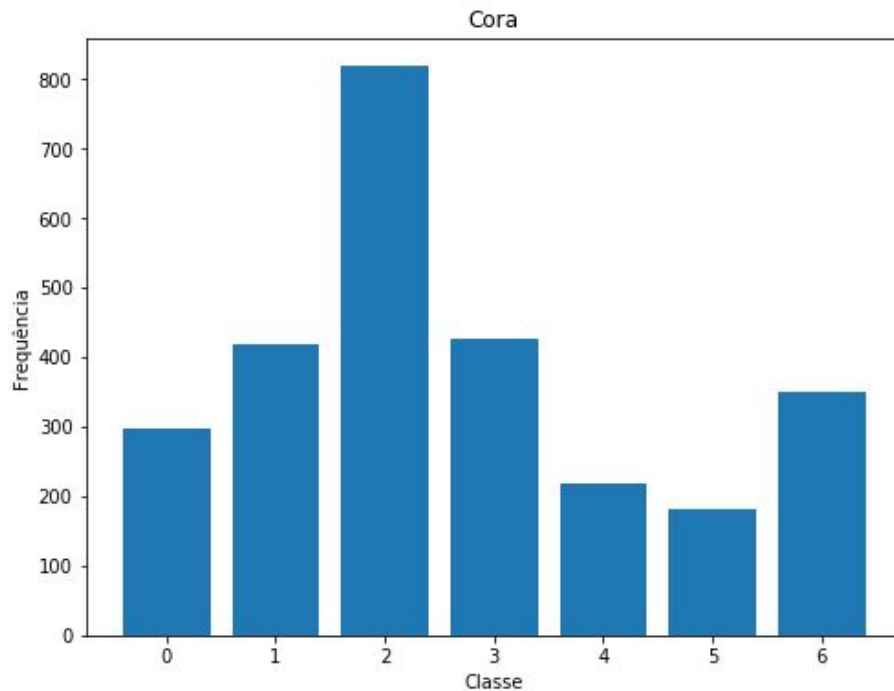
- Micheli [2009] realizou o primeiro trabalho
 - Convolução por somas sucessivas dos vetores de características dos vértices vizinhos.
- Hamilton et al. [2017] propôs o GraphSAGE
 - Amostragem dos vértices vizinhos
- Chiang et al. [2019] propôs o ClusterGCN
 - Utilizou algoritmos de agrupamento para particionar os vértices do grafo.
 - Cada iteração treina com partições escolhidas aleatoriamente
 - Significativa diminuição do consumo de memória e tempo de treinamento, além de viabilizar a manipulação de grafos com mais de 2 milhões de vértices.

Visão geral da proposta



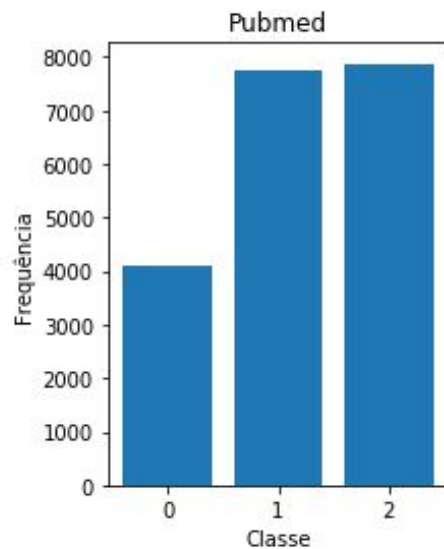
Conjuntos de dados

- Cora [Sen et al., 2008]
 - 2.708 vértices, 5.429 arestas



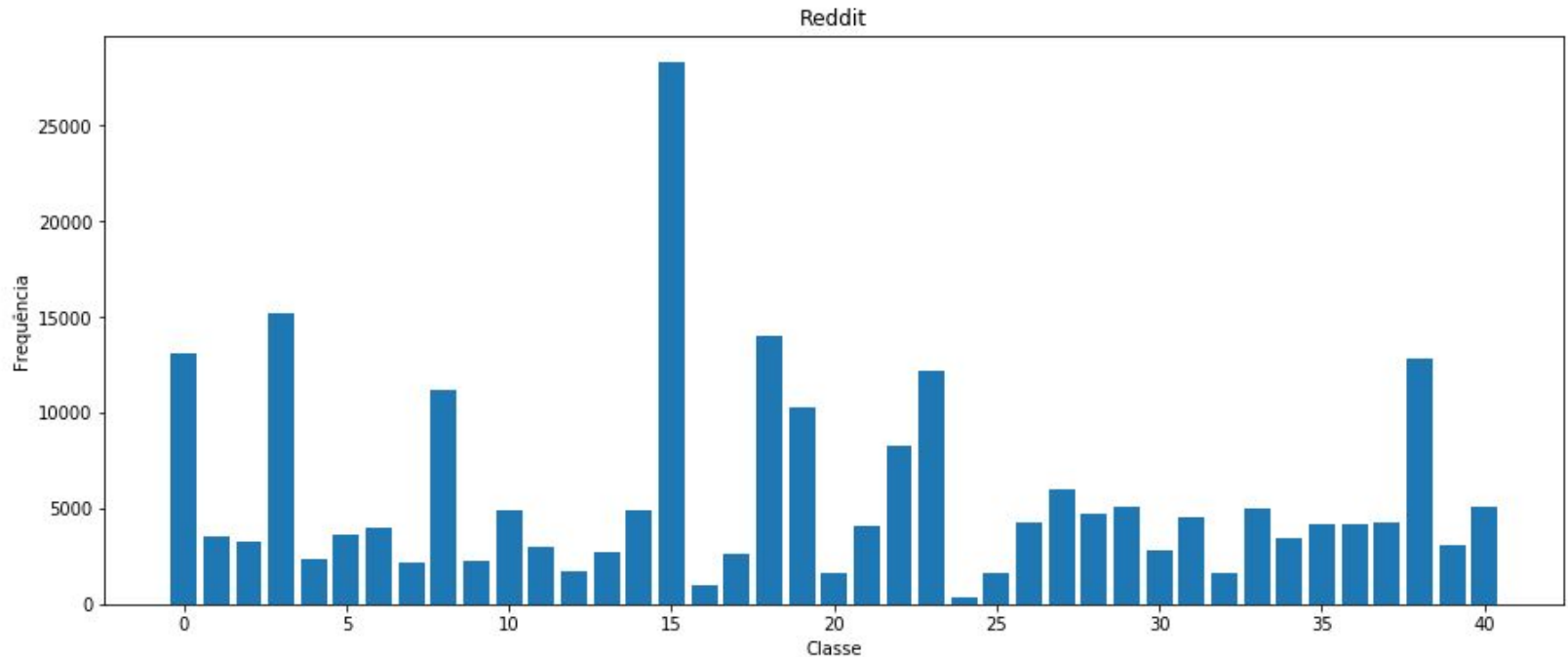
Conjuntos de dados

- Pubmed [Sen et al., 2008]
 - 19.717 vértices, 44.338 arestas



Conjuntos de dados

- Reddit [Hamilton et al., 2017]
 - 232.965 vértices, 11.606.919 arestas





Arquitetura GNN - ClusterGCN

- Obteve melhor acurácia em conjuntos de dados utilizados em trabalhos anteriores
- Baixa complexidade de tempo e memória comparado aos trabalhos anteriores
- Única arquitetura que conseguiu trabalhar como GNN profunda e conjunto de dados com mais de 2 milhões de vértices
- Código fonte disponível no Github



Arquitetura AENN

- Implementada de forma a:
 - ajustar-se às fontes de dados empregadas pelo ClusterGCN
 - empregar o mesmo método de aprendizado para a tarefa de classificação



Hiperparâmetros, treinamento e métricas

- Foram definidas combinações de hiperparâmetros para as duas arquiteturas, sendo:
 - 108 combinações para o ClusterGCN
 - 12 combinações para a AENN
- Após a busca pelos melhores hiperparâmetros, 6 redes foram selecionadas, uma para cada par arquitetura/conjunto de dados.



Hiperparâmetros, treinamento e métricas

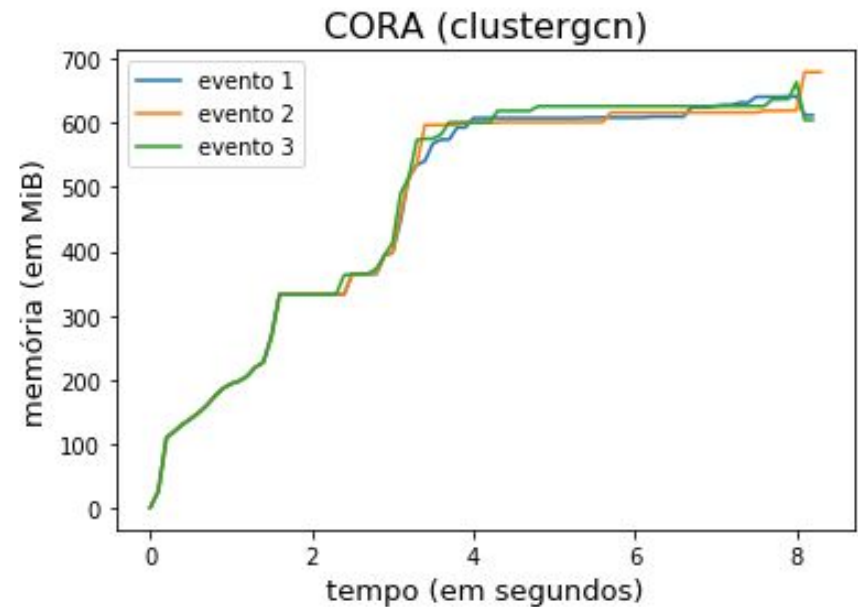
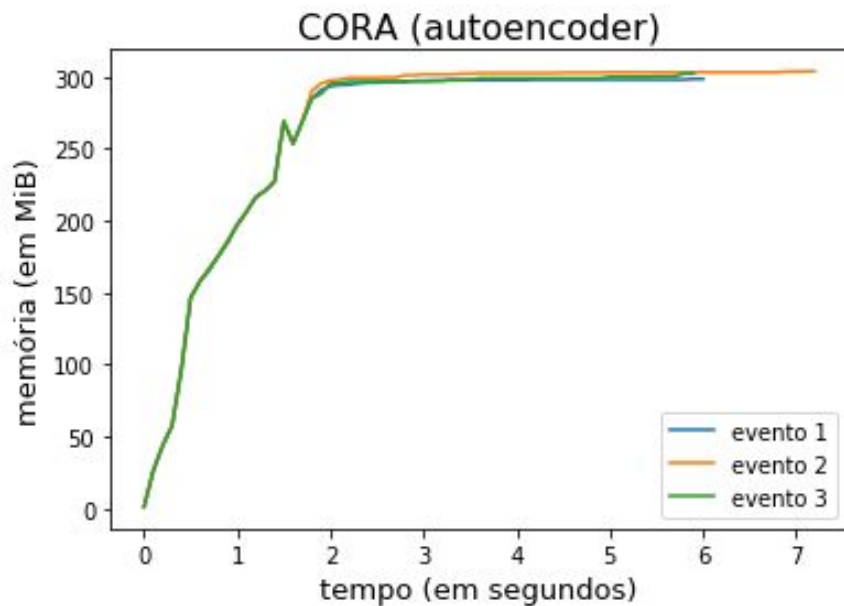
- Métricas
 - Consumo de memória
 - Memória RAM alocada ao longo do tempo de execução do processo
 - Tempo de treinamento e inferência
 - Tempo cronometrado
 - Poder preditivo
 - *F1 Score*



Avaliação Experimental

- Cada uma das 6 redes foi novamente treinada por 3 vezes, de forma a monitorar:
 - Valor **médio** do tempo de treinamento e inferência
 - Valor **máximo** do consumo de memória
- A seguir são apresentados os resultados obtidos para essas 6 redes

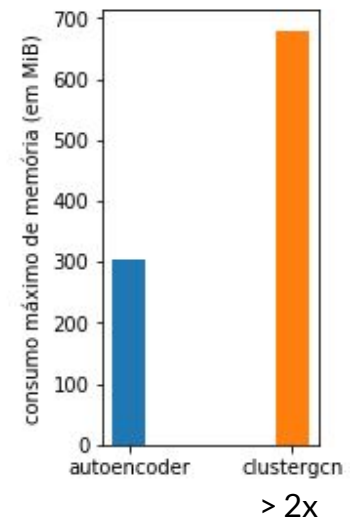
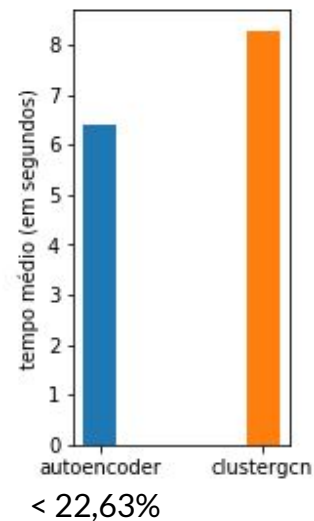
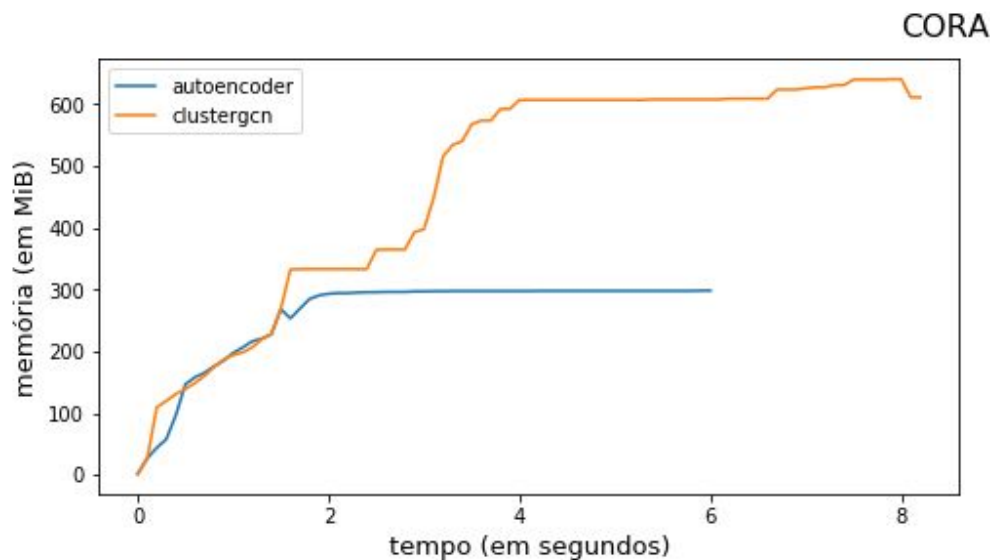
Avaliação Experimental



Consumo de memória em função do tempo de treinamento e inferência.

Cada gráfico apresenta 3 execuções da mesma rede para o conjunto de dados Cora.

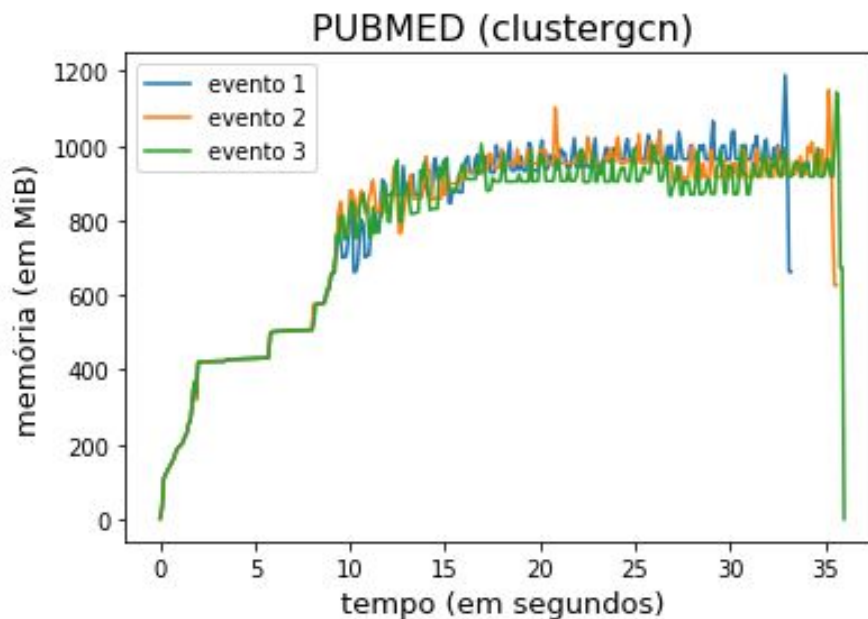
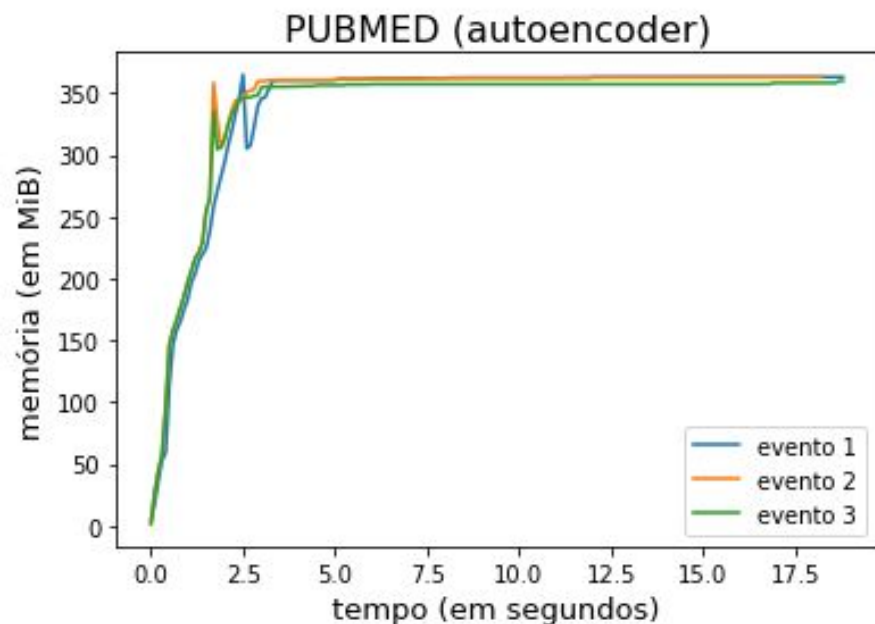
Avaliação Experimental



À esquerda, o gráfico comparativo entre as arquiteturas para o conjunto de dados de Cora.

À direita, o tempo médio de treinamento e o consumo máximo de memória.

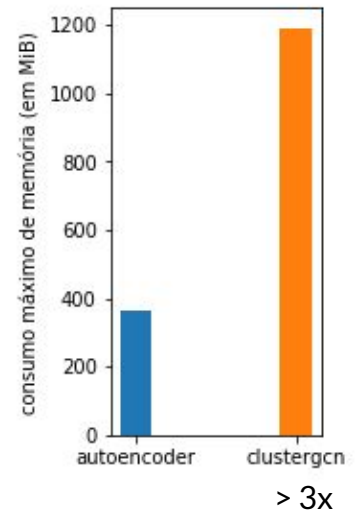
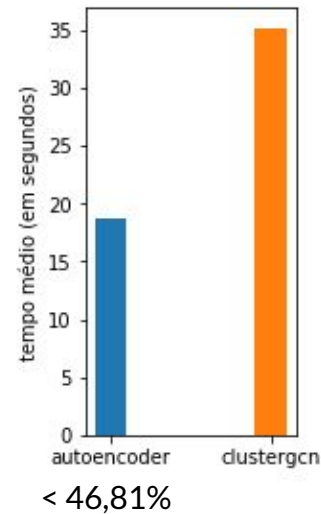
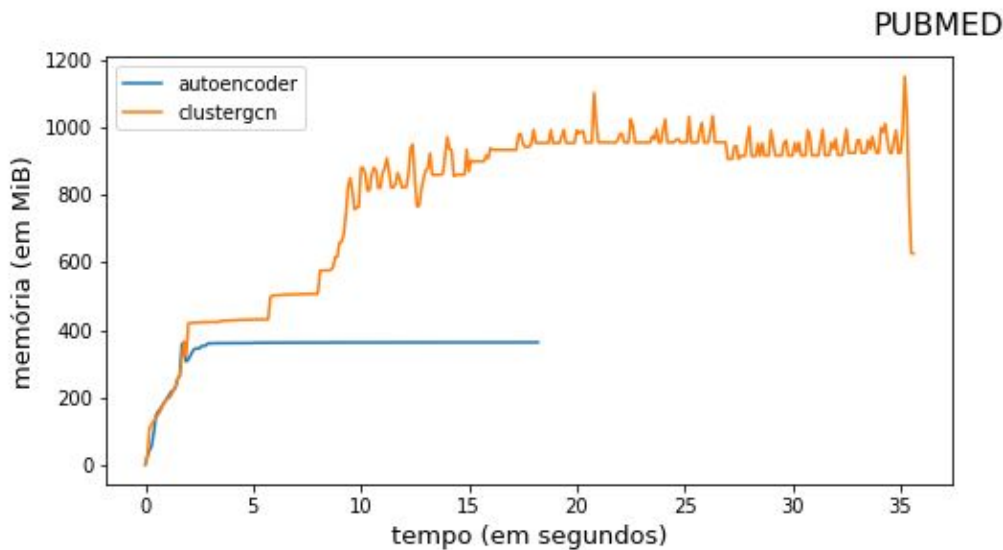
Avaliação Experimental



Consumo de memória em função do tempo de treinamento e inferência.

Cada gráfico apresenta 3 execuções da mesma rede para o conjunto de dados Pubmed.

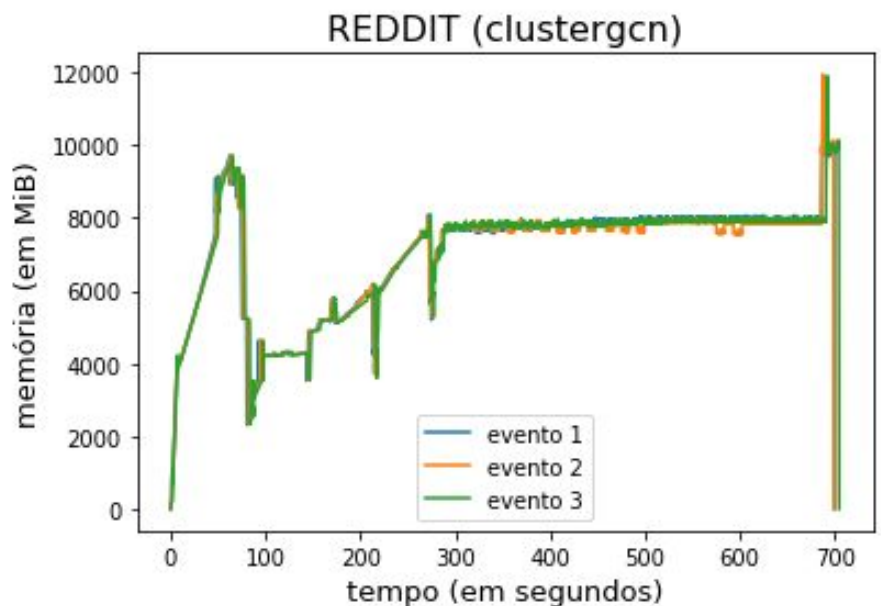
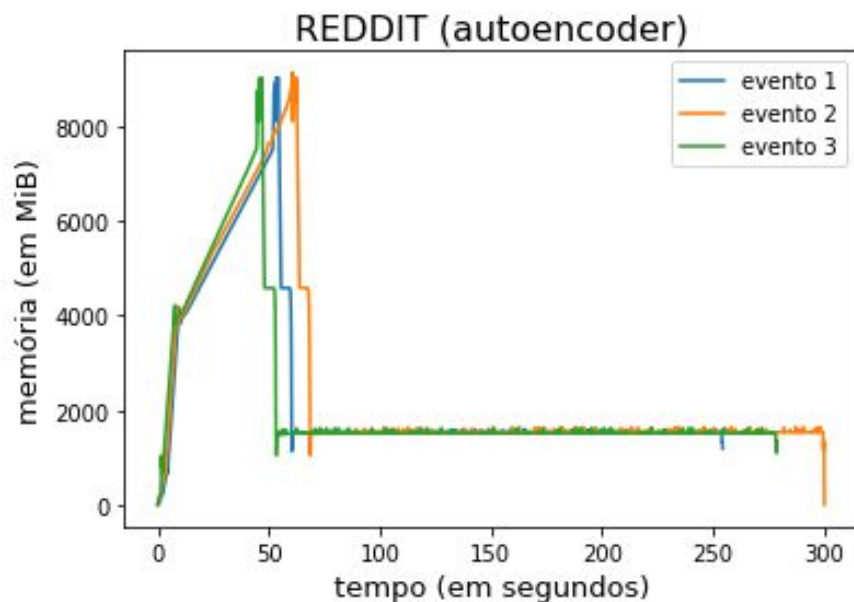
Avaliação Experimental



À esquerda, o gráfico comparativo entre as arquiteturas para o conjunto de dados de Pubmed.

À direita, o tempo médio de treinamento e o consumo máximo de memória.

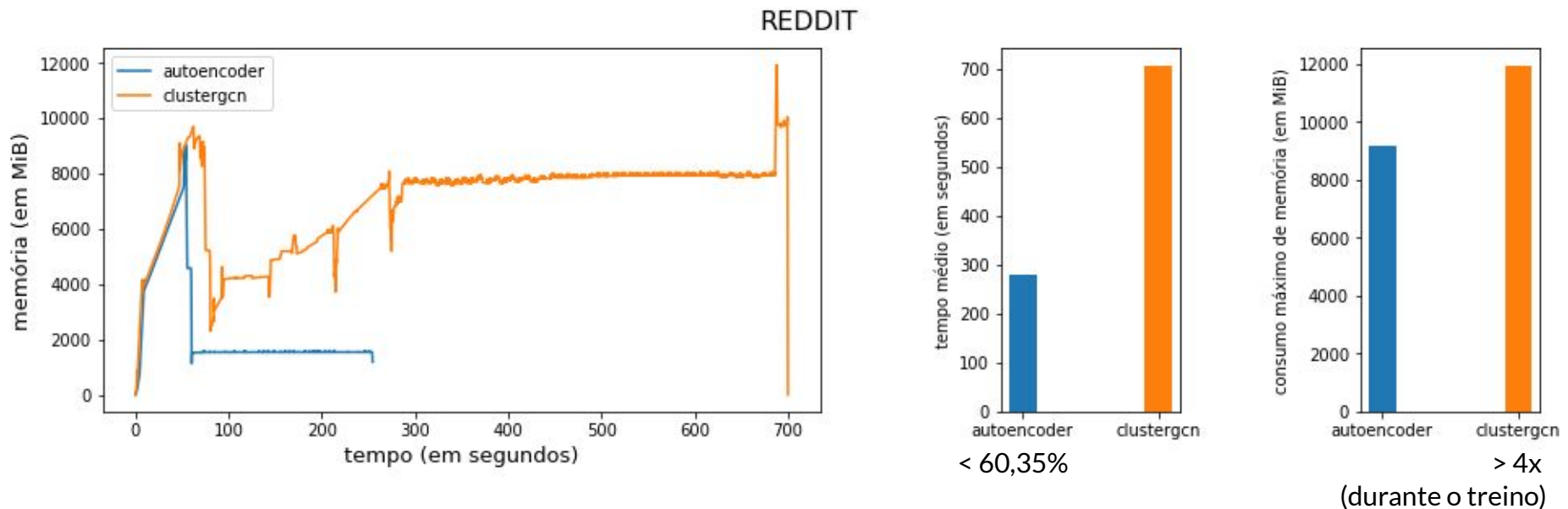
Avaliação Experimental



Consumo de memória em função do tempo de treinamento e inferência.

Cada gráfico apresenta 3 execuções da mesma rede para o conjunto de dados Reddit.

Avaliação Experimental



À esquerda, o gráfico comparativo entre as arquiteturas para o conjunto de dados de Reddit.

À direita, o tempo médio de treinamento e o consumo máximo de memória.



Avaliação Experimental

Acurácia dos modelos (F1 Score)

	ClusterGCN	AENN
Cora	86,90	14,02
Pubmed	87,47	85,19
Reddit	95,73	61,95



Análise Retrospectiva

- O ClusterGCN foi capaz de gerar modelos satisfatórios em termos de poder preditivo.
- A técnica de emprego de partições do grafo como amostras no processo de treinamento reduziu o uso de recursos computacionais.
- Para os padrões de hardware atuais, não haveria justificativa em não empregar o ClusterGCN.
- Devemos descartar a possibilidade de emprego das AENN?



Análise Retrospectiva

- A AENN obteve um valor *F1 Score* muito próximo do ClusterGCN no conjunto de dados Pubmed.
- Hipótese 1
 - O resultado alcançado pela AENN para o Pubmed pode estar relacionado às características do conjunto de dados
 - Apenas três classes distintas
 - Duas classes balanceadas



Análise Retrospectiva

- A quantidade de exemplos no conjunto de dados Pubmed é cerca de 7,3 vezes maior do que no Cora.
- Hipótese 2
 - O resultado insatisfatório alcançado pela AENN para o Cora pode estar relacionado à quantidade insuficiente de exemplos para o devido aprendizado de representações.



Análise Retrospectiva

- A quantidade de exemplos no conjunto de dados Reddit é cerca de 12 vezes maior do que no Pubmed.
- Hipótese 3
 - O resultado insatisfatório alcançado pela AENN para o Reddit pode estar relacionado à necessidade de uma rede maior, em termos de número de camadas, para o devido aprendizado de representações.



Trabalhos futuros

- Avaliar conjuntos de dados similares ao Pubmed (reduzido número de classes, balanceado).
- Aplicar técnicas de balanceamento aos conjuntos de dados.
- Avaliar AENNs com maior número de camadas.
- Avaliar diferentes técnicas de pré-processamento dos dados.



Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Escola de Informática & Computação

Curso Superior de Tecnologia em Sistemas para Internet (CST-SI)

Aprendizado de Representações em Grafos por Redes Neurais: uma Análise Comparativa

Aluno: Augusto José Moreira da Fonseca

Professor orientador: Eduardo Bezerra, D.Sc.