GNN em Segurança de Redes*

Lucas A. Jaques da Costa[1], **Helen C. de Mattos Senefonte**[1]

[1]Departamento de Computação – Universidade Estadual de Londrina (UEL)
Caixa Postal 10.011 – CEP 86057-970 – Londrina – PR – Brasil

lucas.a.jaques@uel.br, helen@uel.br

Abstract. *Graph Neural Network (GNN) is an Artificial Inteligence model that combines the theory of graphs with the concepts of Neural Network. This model is already utilized in molecular chemistry, traffic prediction, recommendation systems, social classification, etc. But it's not common to find GNNs being applied in cybersecurity, so this project's objective is to explore, analyze and develop the usage of GNNs in Network Security.*

Resumo. *Graph Neural Networks (GNN) é um modelo de Inteligência Artificial que usa a Teoria de Grafos junto dos conceitos de Rede Neural. Esse modelo já é usado em química molecular, previsão de tráfegos, sistemas de recomendação, relações sociais, etc. Porém não é comum encontrar o uso de GNNs na área de cybersegurança, então esse trabalho visa explorar, analisar e desenvolver o uso de GNN na área de Segurança de Redes.*

1. Introdução

Com o mundo cada vez mais presente na internet, onde negócios, transações, conversas e acordos acontecem parcialmente ou completamente no meio digital. Logo, a segurança nesses meios fica cada vez mais difícil, tanto pelo volume de dados quanto pelo avanço de técnicas de cyber-ataques.

Uma das áreas em Segurança de Redes abordadas é a Detecção de Anomalias, onde o objetivo [8] é encontrar padrões do acesso dos usuários fora do normal. Várias técnicas utilizando Aprendizado de Máquina [4] foram usadas para resolver esse problema: como o uso de **K-Nearest Neighbor** para classificar se uma atividade é um cyber-ataque ou não; e **Bayesian Network** também pode ser usado para distinguir se uma atividade é um ataque ou não por meio de comparações de métricas amostradas durante a fase de treinamento do modelo.

Um outro modelo utilizado também para a detecção de anomalias é a **Surpervised Neural Network** [4], que tem uma boa tolerância a dados imprecisos e incertos, e uma habilidade de encontrar boas soluções sem ter um conhecimento prévio das regularidades dos dados.

Com esses exemplos, é notável que Aprendizado de Máquinas e Redes Neurais são ferramentas muito poderosas para a análise de anomalias, porém um modelo ainda

*Este meta-artigo descreve o estilo a ser usado no projeto de TCC do curso de Bacharelado em Ciência da Computação da UEL. É o estilo usado para publicação nos anais das conferências organizadas pela SBC (Disponível em http://www.sbc.org.br/index.php?option=com_jdownloads&Itemid=0&task=view.download&catid=32&cid=38), com alguns pacotes LaTeX úteis e recomendações quanto ao conteúdo do projeto, baseadas no texto de [**?**].

não muito utilizado para isso é **Graph Neural Network** (GNN), ele basicamente é um modelo de Rede Neural que utiliza grafos na estruturação da rede, essa estruturação particular pode ser útil quando se é necessário descobrir a correlação de elementos, como [7] vulnerabilidade em códigos e em Smart Contracts.

Nesse trabalho temos as seguintes seções: A Seção 2 apresenta as Fundamentações Teóricas importantes para a conclusão desse trabalho; A Seção 3 define de forma direta quais os focos do trabalho; A Seção 4 explica em mais detalhes quais os passos de como obter meus objetivos; A Seção 5 mostra em formato de tabela quanto tempo cada tarefa levará até a conclusão do TCC.

2. Fundamentação Teórico-Metodológica e Estado da Arte

Para a produção desse trabalho é necessário compreender três áreas principais: Grafos, Segurança de Redes e Rede Neural.

2.1. Segurança de Redes

Uma das áreas mais importantes da Computação, Segurança de Redes abrange uma série de formas de prevenir ataques, como apontado por Abhinav V. Deshpande [2], por exemplo em detecção de acesso não autorizado e manipulação de dados.

Ela está presente em tudo: em computadores domésticos na forma de firewalls e antivírus; finanças com detecções de fraude de cartão de crédito; redes sociais na detecção de bots e spam; em bancos para detecção de transações inválidas.

Para esse trabalho focarei na área de Detecção de Anomalias [8], que tem como objetivo descobrir ações ou padrões ”diferentes do normal”com o objetivo de identificar e prevenir danos maiores. Foi escolhido esse tema em específico pois nele é possível trabalhar com GNNs para a detecção dessas anomalias, podendo ter resultado até melhores do que com outras técnicas já utilizadas.

2.2. Grafo

Para iniciar o estudo de GNN, é necessário primeiro entender o que é um Grafo [7][3], que nada mais é que uma Estrutura de Dados que usa o grupo de informações (V, E, U), onde **V** é um grupo de **vértices**, **E** são **arestas** que representam a conexão de 2 elementos do grupo V, e **U** são os as informações **globais** do grafo (que pode ser o número de nós, de arestas, o caminho entre dois nós, etc).

O objetivo dessa Estrutura é codificar relações. Por exemplo, poderíamos usar grafos para representar a distancia de cidades entre si, como mostrado na Figura 1, onde os vértices são as cidades, e as arestas são as distâncias entre cada cidade. Grafos também podem ser utilizados [5] em analise de imagens, descrição de cenas, engenharia de software, e processamento de linguagem natural.

2.3. Rede Neural

Rede Neural (ou melhor, Rede Neural Artificial) [1] é um algoritmo de Aprendizado de Máquina inspirado pelas Redes Neurais encontradas na biologia. Como apresentado na Figura 2, uma Rede Neural possui nós que se comunicam entre si usando ”conexões”, esses nós são separados em múltiplas camadas onde: a primeira camada é a de entrada

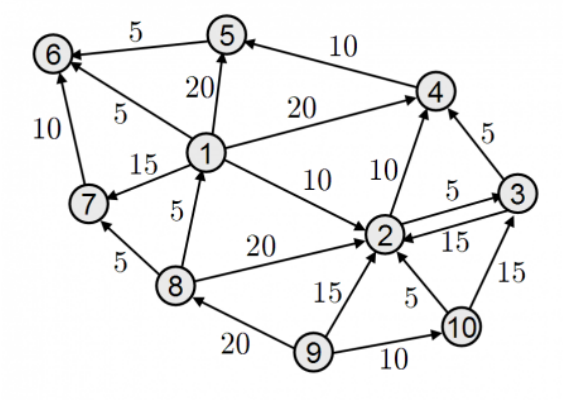


Figura 1. Representação visual de um grafo

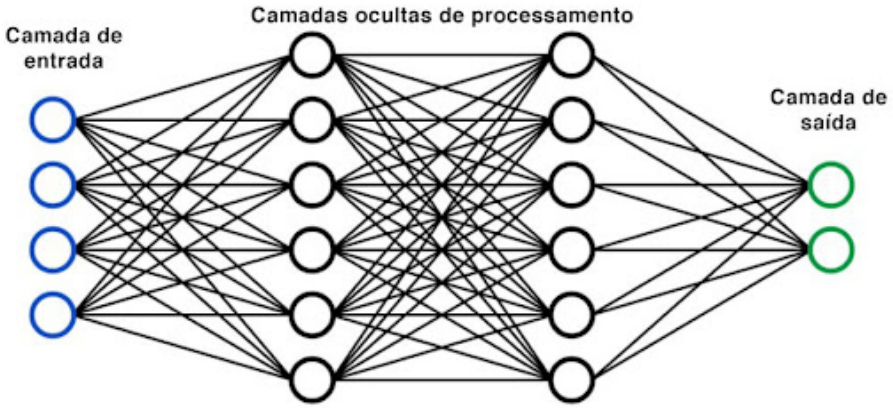


Figura 2. Representação visual de uma Rede Neural

(nós azuis), uma de saída (nós verdes), e entre as duas várias intermediárias para o processamento do dado (nós pretos).

Na maioria das aplicações envolvendo esse algoritmo os dados sempre movem para uma direção. Eles são inseridos na camada de entrada, processados nas camadas intermediárias, e apresentados na camada de saída.

2.4. Graph Neural Network

GNNs usam uma estrutura semelhante de uma Rede Neural, só que ao invés de usar nós que se conectam apenas com a camadas antecessora e posterior, cada camada é um uma rede de nós (ou seja, um grafo), que além de se conectarem com nós de camadas posterior e antecessor, também se conectam com nós da própria camada, como mostra a Figura 3.

Logo, [7] a ideia é que temos um grafo como entrada, e outro grafo como saída, onde as informações são inseridas em nós, arestas e/ou no contexto global. Com esse modelo [7][6], é possível criar aplicações que tem como objetivo de calcular um resultado baseado no tipo de tarefa a ser aprendido:

- **Node-level**: Tarefa focada nos nós, onde é analisado a identidade ou papel de cada nó no grafo. Nessa forma de aprendizado pode-se abordar diferentes objetivos:

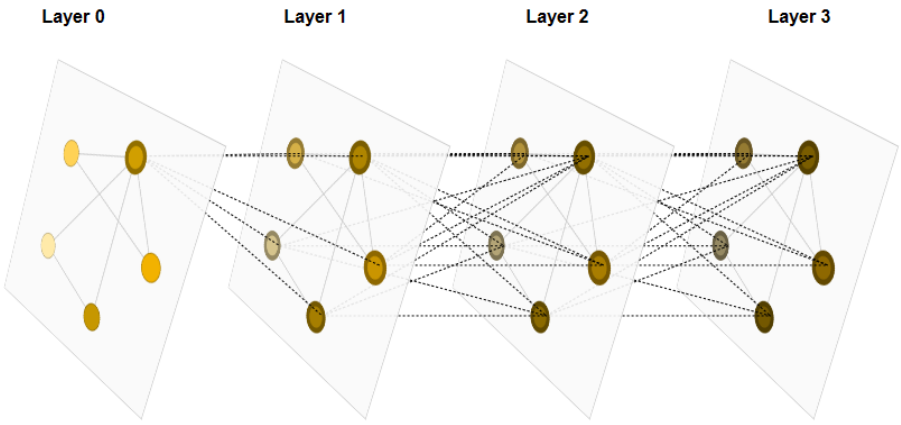


Figura 3. Representação visual de uma Graph Neural Network

categorização dos nós, agrupamento de nós e, mais especificamente, análise de redes sociais;

- **Edge-level**: Tarefa focada nas arestas, onde se busca a classificação das arestas e/ou predição de conexões. Na classificação pode-se buscar como cada par de nós se conecta, e na predição é verificado se um par de nós pode ou não possuir uma conexão.
- **Graph-level**: Tarefa que busca analisar o grafo como um todo. Por exemplo em uma representação de grafo de uma molécula, Graph-Level pode ser usado para dizer que cheiro uma molecula pode ter.

3. Objetivos

O objetivo desse trabalho é estudar, analisar e desenvolver a aplicação de GNNs no contexto de identificar e barrar anomalias em um sistema.

4. Procedimentos metodológicos/Métodos e técnicas

Inicialmente, é necessário se aprofundar e buscar novas Referências Bibliográficas sobre o tema do trabalho, com o objetivo de ter um melhor entendimento de GNNs e como aplicá-la da melhor forma na detecção de anomalias.

Depois, pesquisar uma base de dados para realizar o treinamento e testes com o modelo. No momento já foi encontrada uma base no **Kaggle**, na qual ela possui registros de acesso de micro-serviços via API de uma aplicação. Ainda é necessário explorará-la para ver se ela se encaixa com o problema que é proposto a ser resolvido.

Com essa base em mãos, precisaremos realizar o pré-processamento dessa base para extrair quais informações são ou não são relevantes para serem usados na detecção de anomalia usando a GNN.

Após o pré-processamento, será iniciado o treinamento e teste do modelo. Aqui deve-se estudar a melhor forma para modelar os dados e como a GNN dev e ser configurada para melhor identificar as anomalias.

E por fim, será feita a analise dos resultados obtidos e as considerações finais. Identificando se os resultados são satisfatórios para o problema, e abordando os lados

positivos e negativos de se usar GNNs para a detecção de anomalias, a comparando com outros modelos de Rede Neural também utilizados para isso.

5. Cronograma de Execução

Atividades:

1. **Atividade 1:** Revisão Bibliográfica;
2. **Atividade 2:** Pesquisar bases de dados;
3. **Atividade 3:** Pré-processamento dos dados;
4. **Atividade 4:** Treinamento e teste do modelo;
5. **Atividade 5:** Analise dos dados e considerações finais;

Tabela 1. Cronograma de Execução

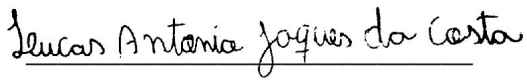
	set	out	nov	dez	jan	fev	mar	abr	mai
Atividade 1	X	X	X	X					
Atividade 2	X								
Atividade 3	X	X	X						
Atividade 4				X	X	X	X	X	
Atividade 5							X	X	X

6. Contribuições e/ou Resultados esperados

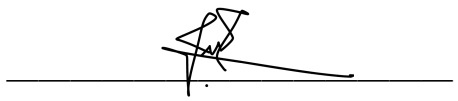
Espera-se, com esse trabalho, analisar o quão útil GNNs podem ser na área de Segurança de Redes, e criar uma aplicação que exemplifique com sucesso esse modelo de Rede Neural nessa área da Computação.

7. Espaço para assinaturas

Londrina, *data_por_extenso*.



Aluno



Orientador

Referências

[1] Rene Y. Choi1; Aaron S. Coyner; Jayashree Kalpathy-Cramer; Michael F. Chiang; and J. Peter Campbell. *Introduction to Machine Learning, Neural Networks, and Deep Learning*. Translational Vision Science Technology, 2015.

[2] Abhinav V. Deshpande. *Introduction to Network Security*. International Journal on Computer Science and Engineering, 2015.

[3] Darij Grinberg. *An introduction to graph theory*. Drexel University, 2023.

[4] Salima Omar; Asri Ngadi; Hamid H. Jebur. *Machine Learning Techniques for Anomaly Detection: An Overview*. International Journal of Computer Applications, 2013.

[5] Franco Scarselli; Marco Gori; Ah Chung Tsoi; Markus Hagenbuchner; Gabriele Monfardini. *The graph neural network model*. University of Wollongong, 2009.

[6] Jie Zhou; Ganqu Cui; Shengding Hu; Zhengyan Zhang; Cheng Yang; Zhiyuan Liu; Lifeng Wang; Changcheng Li; Maosong Sun. *Graph neural networks: A review of methods and applications*. ScienceDirect, 2020.

[7] Benjamin Sanchez; Lengeling; Emily Reif; Adam Pearce; Alexander B. Wiltschko. *A Gentle Introduction to Graph Neural Networks*. Distill, 2021.

[8] Shen Wang; Philip S. Yu. *Graph Neural Networks in Anomaly Detection*. Springer Nature, 2022.