Poster Application of Graph Neural Networks for Representing and Analyzing the Internet Topology via the BGP Protocol

Valentina Esteban
DCC, Universidad de Chile & NICLabs
Santiago, Chile
valentina@niclabs.cl

Ivana Bachmann
DCC, Universidad de Chile & NICLabs
Santiago, Chile
ibachman@dcc.uchile.cl

Sebastián Ferrada
IDIA, Universidad de Chile & IMFD
Chile
Santiago, Chile
sebastian.ferrada@uchile.cl

Abstract

The relationships between Autonomous Systems (ASes) are a crucial aspect of the Internet, as they reveal how it operates and influence in the routing decisions, as well as identifying BGP anomalies. However, most of the time this information is confidential, given that each AS is independently managed by different entities. This work aims to infer the types of relationships between ASes using Graph Neural Networks (GNNs).

The Type of Relationship (ToR) problem has been a topic studied for the past two decades, with most solutions being heuristic. One of the biggest challenges this problem presents is the lack of ground truth information to validate the results.

Our preliminary results show an accuracy of 0.943 for binary classification and 0.936 for multiclass classification.

Keywords

AS relationships, routing policies, GNN

ACM Reference Format:
Valentina Esteban, Ivana Bachmann, and Sebastián Ferrada. 2024. Poster Application of Graph Neural Networks for Representing and Analyzing the Internet Topology via the BGP Protocol. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3646547.3689680

1 Introduction

The Internet consists of thousands of interconnected Autonomous Systems (AS). Each AS consists of a set of IPs addresses that share a common routing policy and are managed by a single entity. BGP is the protocol through which ASes advertise their routing tables and changes in their AS Paths to reach specific IP addresses. Consequently, each AS receives these announcements from all its BGP neighbors and makes decisions on the best way to route its packets.

Within the Internet's graph of Autonomous Systems, packet paths between nodes are often not the shortest due to commercial agreements between ASes. These agreements typically fall into three relationship types: 1) Provider-to-Customer (P2C), where the customer pays the provider AS for access to the Internet; 2) Peer-to-Peer (P2P), where ASes exchange traffic directly among each other and with their customers, but not with providers or other peers; and 3) Sibling-to-Sibling (S2S), where two ASes belong to the same administrative domain. Understanding the relationships between ASes is essential when evaluating the structure of the Internet and its performance, resilience, and evolution globally.

In this research, we propose the use of Graph Neural Networks (GNNs) to classify relationships between Autonomous Systems on the Internet. Previous studies have relied primarily on heuristic methods, while more recent research has explored Deep Learning approaches using BGP2Vec for node representation. However, the application of GNNs for this purpose remains unexplored. Through this study, we aim to improve the accuracy and efficiency of classifying relationships among Autonomous Systems.

2 Related Work

Early studies focused on solving this problem through heuristic methods, with the first being Gao [6]. Gao gathered information from public BGP tables and used the concept of "valid paths" for heuristic inference. His approach was based on the idea that a provider Autonomous System would have a higher degree in the graph than a customer, and that peers would have approximately the same degree. The algorithm managed to locally identify the "top" providers and then classify the relationships between Autonomous Systems. Later, Subramanian et al. [10] formulated the problem as an optimization task, defining it as the "the type of relationship (ToR) problem." To simplify the problem, they removed S2S relationships and focused on C2P and P2P relationships.

Di Battista et al.[5] and Erlebach et al.[8] later proved that the ToR problem is NP-complete. Shavitt [9] applied Deep Learning techniques for the classification task. They used BGP2Vec to create representations of Autonomous Systems and then fed these learned embeddings into an Artificial Neural Network. Just as a word within a sentence provides context in Word2Vec, an Autonomous System is defined by the AS Paths to which it belongs,where the context is provided by its neighbors in BGP2Vec. This study achieved an accuracy of 95.2%.

We believe GNNs could achieve better accuracy, as they not only capture the topology of the graph but also incorporate the representations of neighboring nodes. Additionally, GNNs offer greater flexibility, precision, and generalization capabilities, providing opportunities to explore this problem from new perspectives, such as introducing packet flow dynamics.

We propose a novel technique to resolve the inference of relationships between ASes using Graph Neural Networks (GNNs).We believe GNNs could achieve better accuracy, as they not only capture the topology of a graph but also incorporate the representations

IMC '24, November 4–6, 2024, Madrid, Spain

ACM ISBN 979-8-4007-0592-2/24/11
https://doi.org/10.1145/3646547.3689680

of their neighbors. In addition, GNNs offer great flexibility, precision, and generalization capabilities, providing opportunities to explore this problem from new perspectives, such as introducing packet flow dynamics.

3 The Dataset

To create the dataset, we combined the CAIDA AS-relationships dataset [2] with the dataset from the study by Giakatos et al.[4]. The latter consists of heterogeneous features, including normalized numerical attributes such as ASN, customer cone prefixes, customer cone addresses, total number of neighbors, etc., as well as categorical attributes encoded as one-hot vectors, such as RIR region, location continent, traffic ratio, scope, network type, and others. All this data was collected from various publicly available sources[1–3, 7] to represent a complete Internet topology as of July 2022, along with its features at that time.

Our approach uses this data to create a directed graph representing the Internet through its ASes and incorporates data from CAIDA AS Relationships to label the edges between Autonomous Systems (ASes). In contrast to previous studies, we have the opportunity to not only incorporate node degree and transit degree into the information of an AS, as done in prior research to infer the ToR problem, but also include more information that could help us in the inference process. Additionally, Graph Neural Networks allow us to integrate topology information to enhance the representation of a node.

4 Preliminary Results

We present preliminary results for inferring the type of relationship in both binary and multiclass classification problems.

For edge classification, our pipeline begins with an input consisting of a graph of ASes from the CAIDA AS Relationships dataset, along with the attributes of each AS and its corresponding edge labels. Then, this data passes through a two-layer GNN model, which acts as an encoder. The model generates node embeddings for each AS. These embeddings are subsequently passed to a predictor that acts as a decoder, combining the nodes that form an edge to produce the final classification output. Different GNN models and predictors were tested.

During our review of the dataset, we found that many of the non-categorical data for the ASes were incomplete. Therefore, we decided to include only those attributes whose occurrence percentage in the dataset per AS was greater than 80%.

For the binary classification task, edges between nodes were considered a single class if they were P2C or C2P, with the other type being P2P. We obtained an AUC (Area Under the Curve) of 0.9833 and an accuracy of 0.943 (Figure 1). For the multiclass classification, we obtained an accuracy of 0.9369. The confusion matrix for these results is shown in Figure 2.

The method we utilized shows results similar to those achieved with the state-of-the-art technique presented in [9]. However, we are still in the early stages of this research. As such, there are several areas where we need to improve. For instance, we need to evaluate the performance of edge degree and transit degree attributes alone in the model and compare these results with those obtained using heuristic methods based on these metrics. Additionally, we plan to explore sampling methods, techniques for handling class imbalance, and strategies to avoid overfitting, which are crucial for refining our results. Furthermore, we intend to incorporate traffic node data into the graph as well as investigate alternative frameworks for this problem.

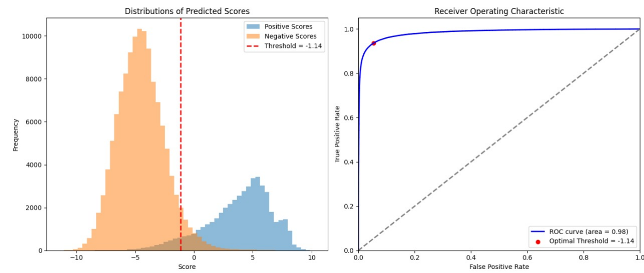

Figure 1: ROC AUC curve for binary classification.

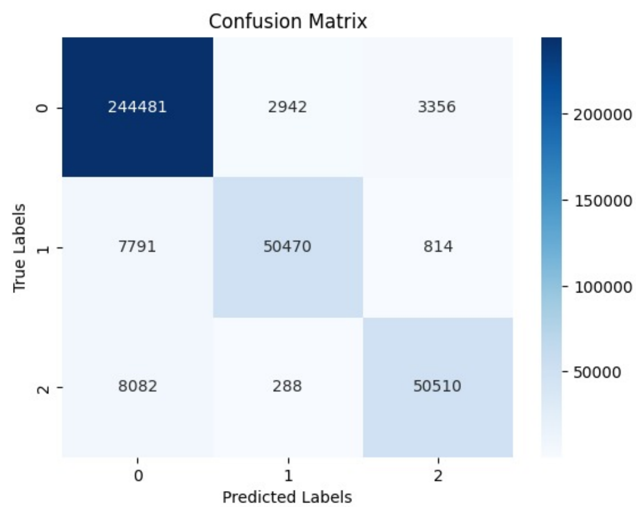

Figure 2: Confusion matrix for multiclass classification.

References

[1] CAIDA. 2022. *AS-rank dataset.* https://asrank.caida.org/
[2] CAIDA. 2022. *AS-relationships dataset.* https://publicdata.caida.org/
[3] CAIDA. 2022. *PeeringDB dataset.* https://publicdata.caida.org/datasets/peeringdb/
[4] Pavlos Sermpezis Athena Vakali Dimitrios P. Giakatos, Sofia Kostoglou. December 2022. Benchmarking Graph Neural Networks for Internet Routing Data. *GNNet '22: Proceedings of the 1st International Workshop on Graph Neural Networking.* (December 2022).
[5] M. Patrignani G. Di Battista and M. Pizzonia. 2003. Computing the Types of the Relationships between Autonomous Systems. *IEEE INFOCOM* (2003).
[6] Lixin Gao. 2001. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking* (2001).
[7] PeeringDB. 2022. *The Interconnection database.* https://www.peeringdb.com/
[8] A. Hall T. Erlebach and T. Schank. 2002. Classifying Customer-Provider Relationships in the Internet. *Proceedings of the IASTED International Conference on Communications and Computer Networks (CCN)* (2002).
[9] Yuval Shavitt Tal Shapira. 2020. Unveiling the Type of Relationship Between Autonomous Systems Using Deep Learning. *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium* (2020).
[10] Characterizing the Internet hierarchy from multiple vantage points. 2002. Characterizing the Internet hierarchy from multiple vantage points. *IEEE INFOCOM* (2002).