

Introducción

En la sociedad actual el Internet juega un papel esencial en la vida cotidiana, facilitando la comunicación, la colaboración y el intercambio de información. En los últimos años, su uso y desarrollo ha crecido exponencialmente, convirtiéndose en una herramienta indispensable. Esta creciente interconexión global ha hecho que las redes sean fundamentales para el funcionamiento de la sociedad moderna. En este contexto, es crucial entender cómo está formado el Internet, ya que conocer su funcionamiento y estructura permite preservar su integridad y eficiencia. Además, garantizar un funcionamiento continuo y óptimo.

El Internet está conformado por miles de Sistemas Autónomos (SA) interconectados entre sí. Cada SA consiste en un conjunto de IPs que comparten un mismo protocolo de enrutamiento y están administradas por una misma entidad, como proveedores de servicios de Internet (ISP), empresas comerciales, universidades, entre otros. A cada uno de estos Sistemas Autónomos se le asigna un número y prefijos de direcciones IP, los cuales anuncia a sus vecinos a través del Border Gateway Protocol (BGP). BGP es un protocolo dinámico de enrutamiento externo en el que los SA anuncian sus tablas de ruteo y cambios en sus AS-Paths para alcanzar direcciones IP específicas. De esta manera, cada SA recibe estos anuncios de todos sus vecinos BGP y toma decisiones sobre la mejor forma de direccionar sus paquetes.

Dentro del grafo de Sistemas Autónomos que conforma Internet, el camino que un paquete recorre de un nodo a otro no suele ser el más corto debido a los acuerdos comerciales que cada SA, como entidad independiente, establece con sus vecinos. Estos acuerdos se clasifican en tres tipos de relaciones: 1) Provider-to-customer (P2C), en el cual el cliente paga al SA proveedor para que este enlace permita el tráfico de sus paquetes hacia el resto de Internet. 2) Peer-to-peer (P2P), donde los SA intercambian tráfico entre sí y con sus clientes, pero no con sus proveedores u otros pares. 3) Sibling-to-sibling (S2S), cuando dos SA pertenecen al mismo dominio. Gao [1] propuso reglas para modelar estas relaciones entre SA, que reflejan cómo suelen configurarse en BGP, lo que permite inferir las posibles rutas seleccionadas por este protocolo. Sin embargo, estas soluciones se basan principalmente en el cálculo de heurísticas. Por otro lado, estudios más recientes como el de Shapira y Shavitt [Unveiling the Type Relationship Between Autonomous Systems Using Deep Learning] proponen técnicas de Deep Learning, creando representaciones de los SA que luego son utilizadas en una Red Neuronal.

Aquí entra en juego un nuevo enfoque que podría cambiar nuestra forma de analizar datos: el uso de Redes Neuronales de Grafos (GNNs). Las GNNs están diseñadas específicamente para trabajar con datos organizados en forma de grafos. Al ser Redes Neuronales, tienen la capacidad de encontrar representaciones efectivas de la información y descubrir patrones en datos estructurados de esta manera. A diferencia de las Redes Neuronales convencionales, las GNNs pueden aprovechar la información de los nodos vecinos, lo que les permite entender mejor la estructura del grafo y realizar análisis más detallados.

Así nace esta tesis, con el objetivo de explorar el comportamiento de las Redes Neuronales de Grafos (GNNs) utilizando datos de BGP para representar Internet y las relaciones entre los Sistemas Autónomos que lo componen. Esta área es aún poco explorada y presenta varias dificultades, principalmente debido a la necesidad de aplicar conocimiento tanto en redes como en Deep Learning. En particular, obtener una representación precisa de la topología de Internet requiere comprender a fondo el funcionamiento del protocolo BGP y saber cómo y dónde obtener datos, que no siempre están disponibles públicamente. Así como también tener un conocimiento de teoría de grafos y Deep Learning necesario al momento de la implementación de un modelo de GNNs y diferentes técnicas para la tarea requerida.

Motivación

(está copiado el problema, es similar pero revisar de cambiar la forma en cómo se plantea)

El problema a abordar mediante esta investigación radica en la necesidad crítica de comprender, medir y evaluar la degradación en la calidad de servicio de Internet en situaciones específicas donde los nodos de la red experimentan desconexiones. En la actualidad, Internet se ha convertido en un componente esencial para individuos y organizaciones en todo el mundo, desempeñando un papel fundamental en la vida cotidiana y en el funcionamiento de diversas aplicaciones y servicios críticos.

La resiliencia de Internet, entendida como su capacidad para mantener la calidad de servicio incluso en condiciones desafiantes, se ve comprometida cuando los nodos experimentan desconexiones. Esta situación se traduce en fenómenos perjudiciales como la pérdida de paquetes, retrasos en la transmisión de datos y fluctuaciones en el rendimiento, todos los cuales impactan directamente en la calidad de la experiencia del usuario. La investigación busca explorar a fondo estas situaciones, clasificando distintos escenarios de desconexión y definiendo métricas específicas para evaluar cómo se manifiesta la degradación en la QoS en cada caso.

La investigación se propone analizar y clasificar diversas situaciones de desconexión, elegir métricas específicas para evaluar la QoS, realizar pruebas realistas en un simulador, y comparar enfoques existentes.

Hipótesis

Las Redes Neuronales de Grafos (GNNs) pueden ofrecer un rendimiento superior en comparación con las metodologías del estado del arte para la inferencia del tipo de relación entre Sistemas Autónomos.

Objetivos

Objetivo general

El objetivo principal de este estudio es evaluar diversas arquitecturas de Redes Neuronales de Grafos (GNNs) para determinar su viabilidad en la inferencia del tipo de relación de tráfico entre dos Sistemas Autónomos. Esto se logrará mediante el análisis de características específicas de cada Sistema Autónomo, la información de actualizaciones BGP, la topología y los cambios en esta.

Objetivos específicos

1. Obtención de datos: Recopilar datos de fuentes confiables como

que correspondan a Sistemas Autónomos representativos de la Red de Internet. Esto implica obtener datos sobre nodos, características y relaciones entre ellos. Asimismo, obtener información relevante sobre flujos de paquetes BGP.

item Preparación de datos: Mejorar la calidad de los datos mediante el uso de técnicas de normalización, conversión de atributos categóricos a numéricos, manejo de desequilibrio de clases, entre otros. Además, construir el grafo y definir cómo se proporcionarán los datos de entrada a nuestros modelos GNNs.

1. Diseño e implementación de modelos: Diseñar e implementar modelos GNN y framework específicos que permita la inferencia del tipo de relación que dos Sistemas Autónomos comparten.
2. Evaluación de performance: Comparar el desempeño de diferentes arquitecturas de GNNs en las inferencias, identificando los parámetros de mayor relevancia.
3. Análisis de resultados: Comprender los resultados obtenidos mediante el estudio y la comparación con los valores esperados y estado del arte

Metodología

Contribuciones

Estructura del trabajo

la tesis está organizada de la siguiente manera:

- capítulo 2 blabla
- capítulo 3 blablabla
- capítulo 4 no existe todavía