

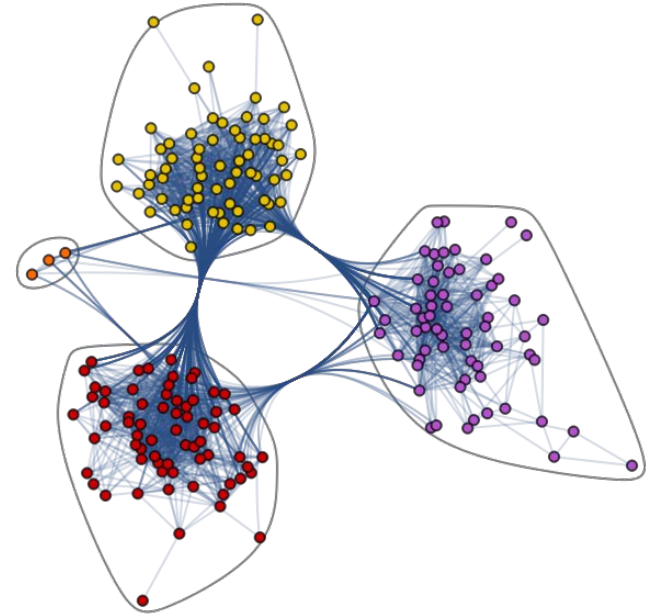
Predicción de Enlaces en Grafos: Aplicación de Redes Neuronales con SEAL, GCN y DGCNN

Edwin Andrés Mora
Iván Andrés Espinosa

Definición de Grafo

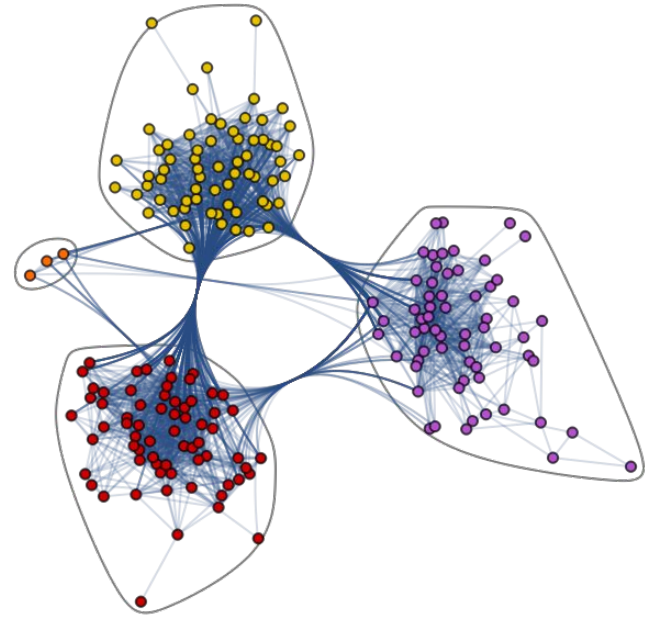
¿Qué es un Grafo?

- Un grafo es una estructura matemática que modela entidades (nodos) y las conexiones (aristas) entre ellas.
- De manera formal, un grafo G se define como un par $G = (V, E)$, donde V es un conjunto de nodos y E es un conjunto de aristas que conectan pares de nodos.



Definición de Grafo

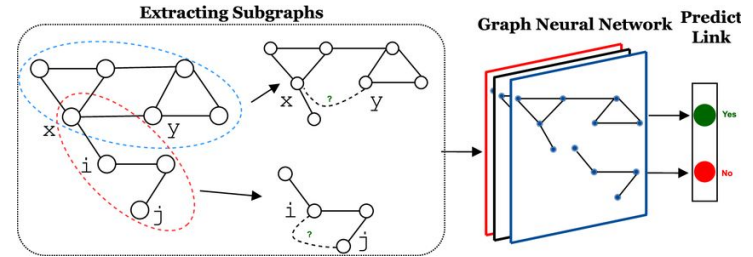
- **Nodos (V):** Representan entidades individuales en la red, como personas en una red social o artículos en una red de citación.
- **Aristas (E):** Representan las conexiones entre nodos, como relaciones de amistad en una red social o citas entre artículos en una red académica.
- **Matriz de Adyacencia (A):** Una representación matricial de un grafo, donde $A[i][j] = 1$ si existe una arista entre el nodo i y el nodo j , y 0 en caso contrario.



Predicción de Enlaces

¿Qué es la Predicción de Enlaces?

- La predicción de enlaces es el proceso de identificar conexiones potenciales entre nodos en un grafo, basándose en la estructura existente y en los patrones observados.

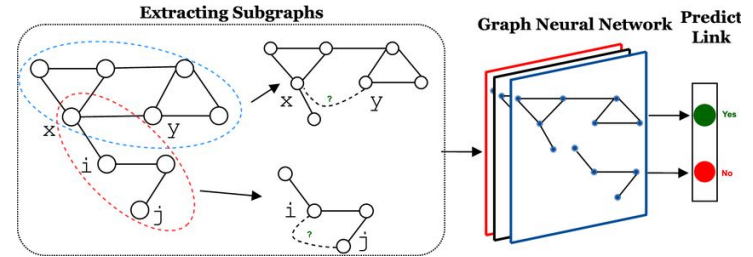


Predicción de Enlaces

Para qué nos sirve?

- **Identificar Nuevas Conexiones:**
Predice qué aristas (enlaces) entre dos nodos podrían formarse en el futuro.
- **Uso de Información Existente:** Se basa en las aristas ya observadas y en la información de los nodos para hacer estas predicciones.

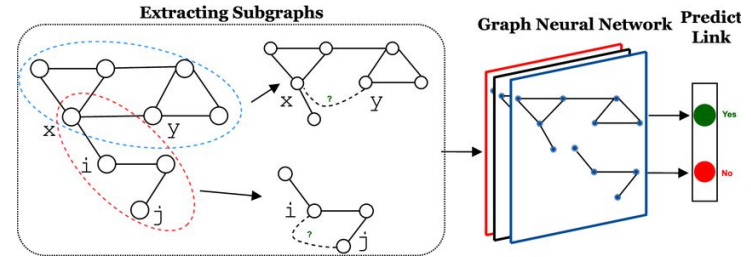
○



Predicción de Enlaces

Métodos Básicos:

- **Heurísticos:** Métodos simples, como el conteo de vecinos comunes o la distancia en el grafo, que dan una medida de similitud entre los nodos para predecir si se conectarán.
- **Modelos Avanzados:** Redes neuronales de grafos (GNNs) que aprenden automáticamente las características relevantes para la predicción.



Aplicaciones de la Predicción de Enlaces

- **Redes Sociales:** La predicción de enlaces se usa para sugerir nuevas conexiones entre usuarios. Por ejemplo, si dos personas tienen muchos amigos en común, es probable que se sugiera que se conecten.
- **Redes de Citación Académica:** Predice qué autores podrían citarse mutuamente en el futuro, basándose en las citas previas y en las temáticas de investigación.



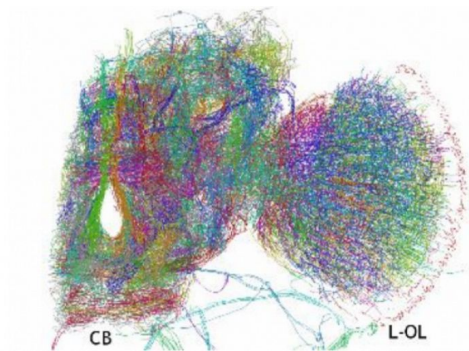
Redes Neuronales de Grafos (GNNs)

¿Qué son las Redes Neuronales de Grafos?

- Las Redes Neuronales de Grafos (GNNs) son una clase de modelos de aprendizaje profundo diseñados específicamente para trabajar con datos estructurados en forma de grafos.

¿Para qué nos sirven?

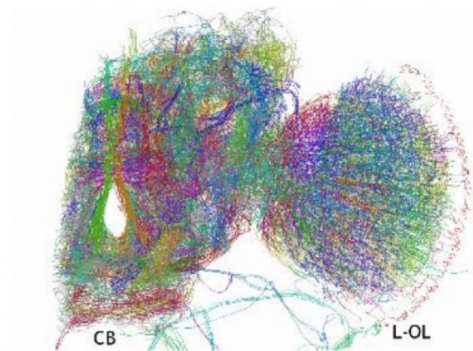
- **Procesamiento de Grafos:** Las GNNs pueden capturar y modelar las relaciones complejas entre los nodos y aristas de un grafo, actualizando las representaciones de los nodos (node embeddings) en función de su entorno.



¿Cómo Funcionan las GNNs?

Actualización de Embeddings:

- **Paso 1:** Inicialmente, cada nodo tiene un embedding basado en sus características propias.
- **Paso 2:** En cada capa de la GNN, los embeddings de los nodos se actualizan al combinar su información con la de sus vecinos. Este proceso se conoce como "**paso de mensajes**".



¿Cómo Funcionan las GNNs?

Propagación de Información:

- **Paso de Mensajes:** En cada paso, un nodo recibe información de sus nodos vecinos y actualiza su embedding para reflejar tanto su información como la de su contexto.
- **m-Capas:** Después de m-capas, cada nodo tiene una representación que captura información de sí mismo y de su entorno K-saltos (nodos a K pasos de distancia).

Aplicación en Predicción de Enlaces:

- **Objetivo:** Las GNNs son ideales para la predicción de enlaces porque pueden aprender automáticamente características estructurales complejas que predicen si una nueva conexión es probable.

Modelo GCN (Graph Convolutional Network)

¿Qué es un GCN?

- **Definición:** El Graph Convolutional Network (GCN) es un tipo de red neuronal diseñada para operar directamente sobre la estructura de un grafo, aplicando operaciones de convolución a los nodos del grafo para aprender representaciones (embeddings) de los nodos que capturan tanto sus características como la estructura del grafo.
- **Propósito:** El objetivo de un GCN es actualizar la representación de cada nodo mediante la agregación de información de sus nodos vecinos, permitiendo capturar patrones locales de la estructura del grafo.

Modelo GCN (Graph Convolutional Network)

¿Cómo Funciona un GCN?

- **Propagación de Información:** En un GCN, cada nodo actualiza su embedding en función de una operación de convolución que agrega las características de los nodos vecinos ponderadas por un conjunto de pesos entrenables.

Modelo GCN (Graph Convolutional Network)

Fórmula de propagación:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$$

$\tilde{A} = A + I$: Matriz de adyacencia con bucles.

\tilde{D} : Matriz diagonal de grados normalizada.

$H^{(l)}$: Embeddings en la capa l .

$W^{(l)}$: Matriz de pesos entrenables.

$\sigma(\cdot)$: Función de activación, como ReLU.

Modelo DGCNN (Deep Graph Convolutional Neural Network)

¿Qué es el DGCNN?

- **Definición:** DGCNN es un modelo de red neuronal convolucional profunda diseñado para trabajar con grafos, que se destaca por su capacidad de capturar y ordenar las características estructurales del grafo de manera que puedan ser procesadas por redes neuronales tradicionales.
- **Propósito:** El DGCNN se construye sobre la base del GCN, pero introduce técnicas adicionales para manejar de manera más efectiva la estructura de los grafos, particularmente cuando los grafos no tienen un orden fijo de nodos.

Modelo DGCNN (Deep Graph Convolutional Neural Network)

Estructura del Modelo DGCNN

Convolución de Grafos: Similar a las capas del GCN, el DGCNN utiliza convoluciones para agregar información de los nodos vecinos y actualizar las representaciones (embeddings) de los nodos.

SortPooling: La capa de SortPooling es una característica distintiva del DGCNN. Esta capa ordena los nodos del grafo en un orden consistente basado en sus características, lo que permite convertir un grafo de tamaño variable en una representación de tamaño fijo.

Modelo DGCNN (Deep Graph Convolutional Neural Network)

Función: Ordena los nodos de acuerdo con su importancia estructural, lo que facilita el uso de capas tradicionales de redes neuronales para la clasificación o predicción.

Capas Convolucionales y Densas Tradicionales: Después de SortPooling, los nodos ordenados son pasados a través de capas convolucionales y densas tradicionales, como en una red neuronal clásica, para realizar la tarea de predicción.

Modelo DGCNN (Deep Graph Convolutional Neural Network)

Fórmula de propagación:

$$Z = f(\tilde{D}^{-1}AXW)$$

Donde $\tilde{A} = A + I$ es la matriz de adyacencia del grafo con auto-bucles, \tilde{D} es la matriz diagonal con $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$, $W \in \mathbb{R}^{c \times c'}$ es una matriz de parametros convolucionales entrenables para el grafo, f es una función de activación no lineal, y $Z \in \mathbb{R}^{n \times c'}$ es la matriz de activación de salida.

Framework SEAL

¿Qué es SEAL?

- **Definición:** SEAL (Subgraphs, Embeddings and Attributes for Link prediction) es un framework diseñado para mejorar la predicción de enlaces en grafos utilizando redes neuronales de grafos (GNNs). Este framework se enfoca en capturar las características estructurales de subgrafos locales alrededor de enlaces objetivo.
- **Objetivo Principal:** SEAL busca aprender automáticamente una función que pueda predecir la existencia de enlaces basándose en la estructura de subgrafos, en lugar de depender de heurísticas predefinidas.

Nota: Al ser este un Framework, se permite el uso de heurísticas y diferentes estrategias de etiquetado de nodos, dando paso a diversas combinaciones y comparar resultados

Implementación de SEAL y Dataset Utilizado

Implementación en Google Colab

Proceso de Entrenamiento:

- **Configuración del Entrenamiento:** El modelo fue entrenado durante 50 épocas. Se utilizaron un 50% del dataset para entrenamiento, un 20% para validación y un 30% para pruebas finales.
- **Métrica de Evaluación:** El rendimiento del modelo fue evaluado usando la métrica AUC (Área bajo la Curva ROC), que es estándar para evaluar la calidad de la predicción de enlaces.

Implementación de SEAL y Dataset Utilizado

Resultados

- La implementación del framework SEAL en este dataset tiene como objetivo evaluar la capacidad del modelo para predecir con precisión las relaciones (enlaces) entre actores que no están explícitamente representadas en el dataset original.
- Se espera comparar los resultados obtenidos por los modelos GCN y DGCNN en términos de la métrica AUC, para determinar cuál ofrece un mejor rendimiento en la tarea de predicción de enlaces en este contexto específico.

Comparación de modelos GCN y DGCNN

Observando tabla que contiene los resultados al probar la predicción de enlaces en el dataset Actor con los modelos GCN y DGCNN, notamos que hay una pequeña ventaja por parte del modelo DGCNN

Modelo	Validación	Test
GCN	83.22 %	82.42 %
DGCNN	83.41 %	82.61 %

Conclusiones

Las redes neuronales de grafos, utilizando el framework SEAL y los modelos GCN y DGCNN, son herramientas efectivas para la predicción de enlaces en grafos, los dos modelos tienen un rendimiento parecido, aunque DGCNN presenta una ligera superioridad debido a su capacidad para ordenar y unificar representaciones de grafos. La implementación práctica en Google Colab demuestra que estas técnicas pueden aplicarse con éxito en contextos reales, ofreciendo soluciones viables para problemas complejos de predicción en redes

GRACIAS!!