



# Tecnológico de Monterrey

## Maestría en Inteligencia Artificial Aplicada

Análisis de Grafos y Extrapolación de Conocimiento  
para identificar Relaciones No Explícitas de  
información

Investigación y documentación extendida del estado del  
arte

Profesor Dr. Gerardo Jesús Camacho González

Profesor Dr. Eusebio Vargas Estrada

Equipo 32

José Adan Vega Pérez	A01796093
Silvia Xochitl Ibañez Vara	A01795200
Diego Andrés Bernal Díaz	A01795975



<i>Introducción</i> .....	5
<b>1. Knowledge Graph Embedding clásico: el paradigma transductivo</b> .....	5
Translating Embeddings for Modeling Multi-Relational Data (TransE) .....	6
<b>2. Aprendizaje relacional y los límites del supuesto de mundo cerrado</b> .....	7
A Review of Relational Machine Learning for Knowledge Graphs .....	7
<b>3. Aprendizaje inductivo en grafos generales</b> .....	8
Inductive Representation Learning on Large Graphs (GraphSAGE) .....	8
<b>4. Redes neuronales sobre grafos multirrelacionales</b> .....	9
Modeling Relational Data with Graph Convolutional Networks (R-GCN) .....	10
<b>5. Redes neuronales sobre grafos multirrelacionales</b> .....	11
Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach nowledge Transfer for Out-of-Knowledge-Base Entities .....	11
<b>6. Inducción centrada en relaciones</b> .....	12
INGRAM: Inductive Knowledge Graph Embedding via Relation Graphs. (Ago. 2023) .....	12
<b>7. Razonamiento estructural basado en subgrafos</b> .....	13
Link Prediction Based on Graph Neural Networks .....	13
Inductive Relation Prediction by Subgraph Reasoning (Grail). .....	14
<b>8. Razonamiento secuencial y búsqueda de caminos en grafos de conocimiento</b> .....	15
DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning .....	16
<b>9. Completado de grafos de conocimiento en escenarios open-world</b> .....	17
Open-World Knowledge Graph Completion for Unseen Entities and Relations via Attentive Feature Aggregation .....	17
<b>10. Extrapolación temporal y meta-aprendizaje en grafos de conocimiento</b> .....	18
Meta-learning based knowledge extrapolation for temporal knowledge graphs. ....	18
<b>11. Evaluación rigurosa y benchmarks para generalización</b> .....	20
CoDEX: A comprehensive knowledge graph completion benchmark.....	20
<b>12. Marco unificador de la extrapolación de conocimiento en grafos</b> .....	21
Generalizing to Unseen Elements: A Survey on Knowledge Extrapolation for Knowledge Graphs.....	21
<b>Glosario Analítico: Terminología de Extrapolación en Grafos de Conocimiento</b> .....	22

A. Paradigmas y Conceptos Fundamentales .....	22
B. Definición de Problemas y Escenarios .....	23
C. Arquitecturas y Modelos Específicos.....	24
D. Evaluación .....	26
<b>Resumen Integral de Técnicas y Evolución hacia la Extrapolación de Conocimiento .</b>	<b>26</b>
<b>Referencias .....</b>	<b>29</b>

## Introducción

Se revisó de manera progresiva los enfoques existentes para el modelado y la inferencia en grafos de conocimiento, con énfasis en las **limitaciones de generalización** que surgen cuando los supuestos clásicos dejan de cumplirse. Se parte de los modelos de *knowledge graph embedding* tradicionales, desarrollados bajo un paradigma transductivo y de mundo cerrado, que sentaron las bases del campo pero presentan restricciones fundamentales frente a escenarios dinámicos y abiertos. A partir de estas limitaciones, el estado del arte evoluciona hacia enfoques que buscan desacoplar el aprendizaje de representaciones de la identidad específica de las entidades.

Posteriormente, se analizan los avances en **aprendizaje inductivo**, primero en grafos generales y luego en grafos multirrelacionales, que permiten generar representaciones para componentes no observados mediante funciones aprendidas sobre la estructura del grafo. Sobre esta base, se revisan formulaciones explícitas del problema de **entidades fuera de la base de conocimiento (OOKB)** y enfoques que desplazan el foco de la inducción hacia relaciones o patrones estructurales compartidos. Estas líneas introducen mecanismos de razonamiento más abstractos, incluyendo el razonamiento estructural basado en subgrafos y el razonamiento secuencial mediante recorridos explícitos, que buscan capturar regularidades transferibles más allá de entidades individuales.

Finalmente, el capítulo aborda escenarios de **mayor complejidad**, como el completado de grafos en contextos *open-world*, la extrapolación temporal y el uso de meta-aprendizaje para grafos dinámicos, así como la importancia de *benchmarks* diseñados para evaluar generalización real. El recorrido concluye con marcos de síntesis que unifican estas contribuciones bajo el concepto de **knowledge extrapolation**, proporcionando una taxonomía que permite comparar enfoques, identificar supuestos implícitos y delimitar con precisión los vacíos que motivan la presente investigación.

### 1. Knowledge Graph Embedding clásico: el paradigma transductivo

Los primeros enfoques para el modelado de grafos de conocimiento se desarrollaron bajo el paradigma de **knowledge graph embedding**, cuyo objetivo principal fue facilitar tareas como la predicción de enlaces mediante representaciones vectoriales continuas. Estos métodos comparten un supuesto fundamental: el grafo es **cerrado y estático**, y el conjunto de entidades y relaciones se conoce completamente durante el entrenamiento. Este escenario, comúnmente denominado **transductivo**, define el punto de partida del estado

del arte y permite entender tanto los logros iniciales del campo como las limitaciones estructurales que motivan el surgimiento de enfoques inductivos y de extrapolación de conocimiento, desarrollados en las secciones posteriores.

## Translating Embeddings for Modeling Multi-Relational Data (TransE)

Bordes et al. (2013)

El trabajo de Bordes et al. (2013) introduce **TransE**, uno de los primeros y más influyentes modelos de *knowledge graph embedding* para el modelado de datos multi-relacionales. La idea central del modelo es representar entidades y relaciones en un espacio vectorial continuo, de tal forma que una relación se interprete como una traslación entre la entidad origen y la entidad destino. Formalmente, para un triplete  $(h, r, t)$ , el modelo busca que el embedding de la entidad cabeza más el embedding de la relación sea cercano al embedding de la entidad cola.

Este planteamiento simple permite escalar a grafos de gran tamaño y ofrece una interpretación geométrica intuitiva de las relaciones. El modelo se entrena maximizando la separación entre tripletes verdaderos y tripletes corruptos generados artificialmente. TransE mostró resultados competitivos frente a modelos previos más complejos y se convirtió en un referente fundamental para el desarrollo posterior de métodos de embeddings en grafos de conocimiento. No obstante, el modelo asume un conjunto fijo de entidades y relaciones durante el entrenamiento y presenta limitaciones al modelar relaciones uno a muchos, muchos a uno y muchos a muchos.

Aporte al estado del arte:

La principal aportación de TransE al estado del arte es haber establecido el **paradigma de embeddings traslacionales**, que sirvió como base para una amplia familia de modelos posteriores de *knowledge graph embedding*. Sin embargo, desde la perspectiva de la extrapolación de conocimiento, TransE representa claramente el **enfoque transductivo clásico**, ya que requiere que todas las entidades y relaciones estén presentes durante el entrenamiento. Esta limitación es precisamente la que motiva trabajos posteriores como el de Hamaguchi et al. (2017), que buscan extender los embeddings a entidades fuera de la base de conocimiento, y es señalada de forma explícita en el survey de Chen et al. (2024) como uno de los supuestos estructurales que deben romperse para lograr una generalización real. En este sentido, TransE funciona como el **punto de contraste**

**fundamental** frente a los enfoques inductivos y de extrapolación que constituyen el núcleo de esta línea de investigación.

## 2. Aprendizaje relacional y los límites del supuesto de mundo cerrado

Antes de la adopción de enfoques inductivos basados en redes neuronales sobre grafos, el modelado de grafos de conocimiento se desarrolló dentro del marco del **aprendizaje automático relacional**, cuyo objetivo era capturar dependencias entre entidades y relaciones en estructuras complejas. Este cuerpo de trabajo permitió sistematizar distintas familias de modelos y sentó las bases conceptuales del campo. Sin embargo, la mayoría de estos enfoques comparten un **supuesto de mundo cerrado**, en el que el conjunto de entidades y relaciones es fijo y conocido. Esta característica resulta adecuada para grafos estáticos, pero limita la capacidad de generalización en escenarios dinámicos y abiertos, motivando la transición hacia enfoques inductivos y de extrapolación de conocimiento que se revisan en las secciones siguientes.

### A Review of Relational Machine Learning for Knowledge Graphs

Nickel et al. (2016)

Nickel et al. (2016) presentan una revisión integral de los enfoques de **aprendizaje automático relacional** aplicados a grafos de conocimiento, abarcando métodos simbólicos, probabilísticos y basados en *embeddings*. El survey sistematiza modelos de factorización matricial, modelos translacionales y enfoques basados en tensores, proporcionando un marco unificado para entender cómo se representa y razona sobre conocimiento estructurado. Asimismo, discute aplicaciones clave como *link prediction*, *entity resolution* y *knowledge base completion*.

El trabajo identifica fortalezas y limitaciones de cada familia de métodos, destacando que la mayoría de los enfoques asumen un conjunto fijo de entidades y relaciones, operando bajo un escenario transductivo. Aunque el survey precede a la adopción masiva de GNNs, establece claramente los retos fundamentales asociados con la escalabilidad, la generalización y la capacidad de razonamiento, anticipando la necesidad de modelos más flexibles y expresivos para grafos dinámicos.

Aporte al estado del arte:

La principal aportación de Nickel et al. (2016) al estado del arte es ofrecer un **marco conceptual fundacional** que sistematiza el aprendizaje relacional previo a los enfoques inductivos y extrapolativos. Desde la perspectiva del problema OOKB abordado por Hamaguchi et al. (2017), este survey permite identificar con claridad el supuesto de **mundo cerrado** que domina los modelos clásicos de embeddings. Asimismo, el survey de Chen et al. (2024) se apoya implícitamente en esta categorización para argumentar por qué la extrapolación de conocimiento requiere romper con los supuestos establecidos en esta literatura temprana. En este sentido, Nickel et al. funcionan como el **punto de referencia histórico** frente al cual se mide el avance de los enfoques modernos de *knowledge extrapolation*.

### 3. Aprendizaje inductivo en grafos generales

La transición desde modelos transductivos hacia enfoques capaces de generalizar a componentes no observados se materializa con la introducción explícita del **aprendizaje inductivo en grafos**. En este paradigma, el objetivo deja de ser aprender representaciones asociadas a nodos específicos y pasa a centrarse en el aprendizaje de **funciones generadoras** que explotan la estructura local del grafo. Este cambio conceptual demuestra que es posible producir representaciones para nodos no vistos sin reentrenar el modelo completo, sentando las bases para una generalización más flexible. Aunque estos enfoques fueron desarrollados inicialmente para grafos homogéneos y no contemplan de forma explícita la semántica relacional de los grafos de conocimiento, constituyen un habilitador metodológico clave que permite, posteriormente, adaptar la inducción al contexto multirrelacional y abordar escenarios de extrapolación más complejos.

#### Inductive Representation Learning on Large Graphs (GraphSAGE)

Hamilton et al. (2017)

Introduce **GraphSAGE**, uno de los primeros marcos generales para el aprendizaje **inductivo** de representaciones en grafos. A diferencia de modelos transductivos previos, GraphSAGE aprende funciones de agregación que permiten generar embeddings para nodos no vistos durante el entrenamiento, utilizando únicamente la información de sus vecindarios locales. El modelo define un conjunto de operadores de agregación—como media, LSTM y *pooling*—que combinan las representaciones de los vecinos de un nodo de forma aprendible y escalable.

Durante el entrenamiento, GraphSAGE muestra subgrafos locales y optimiza las funciones de agregación para preservar proximidad estructural y semántica. En la fase de inferencia, estas funciones pueden aplicarse directamente a nodos nuevos, sin necesidad de reentrenar el modelo completo. Los autores demuestran que GraphSAGE escala a grafos con millones de nodos y que supera a métodos transductivos en tareas de clasificación de nodos en escenarios inductivos. El trabajo establece así una separación clara entre el aprendizaje de parámetros globales y la generación de representaciones locales, sentando las bases para modelos inductivos posteriores en grafos complejos.

Aporte al estado del arte:

La principal aportación de GraphSAGE al estado del arte es introducir un **paradigma inductivo explícito** para el aprendizaje de representaciones en grafos, demostrando que es posible generalizar a nodos no vistos mediante funciones de agregación aprendidas. Este enfoque influye directamente en trabajos como el de Hamaguchi et al. (2017), que adaptan la idea de agregación inductiva al contexto de grafos de conocimiento para manejar entidades *Out-of-Knowledge-Base*. No obstante, GraphSAGE no aborda de manera explícita la extrapolación semántica ni la aparición de nuevas relaciones, manteniéndose en un escenario estructuralmente cerrado. Desde la perspectiva del survey de Chen et al. (2024), GraphSAGE representa un habilitador metodológico clave, pero insuficiente por sí mismo para la **knowledge extrapolation**, ya que no distingue entre generalización a nuevos nodos y extrapolación a nuevos elementos conceptuales. Así, el trabajo funciona como un **puente técnico** entre los modelos transductivos clásicos y los enfoques inductivos y extrapolativos más avanzados.

#### 4. Redes neuronales sobre grafos multirrelacionales

Si bien los enfoques inductivos revisados en la sección anterior demuestran que es posible generalizar a nodos no observados en grafos homogéneos, los grafos de conocimiento presentan una complejidad adicional derivada de la presencia de **múltiples tipos de relaciones** con semántica diferenciada. Modelar adecuadamente esta heterogeneidad requiere extender los mecanismos inductivos para incorporar información relacional explícita en el proceso de propagación de mensajes. En este contexto, las redes neuronales sobre grafos multirrelacionales surgen como un paso necesario para trasladar la inducción al dominio de los grafos de conocimiento, proporcionando la base arquitectónica sobre la cual se desarrollan posteriormente los enfoques orientados a entidades fuera de la base de conocimiento y a la extrapolación estructural.

## Modeling Relational Data with Graph Convolutional Networks (R-GCN)

Schlichtkrull et al. (2018)

Propone **Relational Graph Convolutional Networks (R-GCN)**, una extensión de las redes convolucionales sobre grafos diseñada específicamente para manejar **datos relacionales y grafos de conocimiento**. A diferencia de los GCN estándar, R-GCN introduce transformaciones dependientes del tipo de relación, permitiendo que cada tipo relacional tenga un impacto distinto en el proceso de *message passing*. Para controlar el crecimiento del número de parámetros en grafos con muchas relaciones, el modelo emplea técnicas de regularización como *basis decomposition* y *block-diagonal decomposition*.

El modelo se evalúa en tareas de clasificación de entidades y *link prediction*, mostrando que puede capturar información estructural y relacional de manera más efectiva que enfoques que ignoran la heterogeneidad de las relaciones. R-GCN mantiene un esquema de aprendizaje mayormente transductivo, aunque su arquitectura permite generar representaciones a partir de vecindarios locales. Este trabajo se convierte en una referencia fundamental para el uso de GNNs en grafos de conocimiento y sienta las bases técnicas para modelos posteriores que incorporan atención, inducción y razonamiento estructural.

Aporte al estado del arte:

La principal aportación de R-GCN al estado del arte es proporcionar una **arquitectura general para el aprendizaje neuronal en grafos multirrelacionales**, resolviendo una limitación clave de modelos inductivos previos como GraphSAGE, que no consideran explícitamente tipos de relación. Este avance técnico resulta esencial para trabajos posteriores orientados a la extrapolación, como el de Hamaguchi et al. (2017), que emplean GNNs para generar representaciones de entidades *Out-of-Knowledge-Base* en grafos de conocimiento. Sin embargo, desde la perspectiva de la *knowledge extrapolation* sistematizada por Chen et al. (2024), R-GCN sigue operando bajo un supuesto de mundo cerrado, ya que no aborda la aparición de nuevas relaciones ni la generalización semántica fuera del conjunto de entrenamiento. En este sentido, R-GCN actúa como un **habilitador arquitectónico clave**, pero no como una solución directa al problema de extrapolación de conocimiento.

## 5. Redes neuronales sobre grafos multirrelacionales

La incorporación de mecanismos inductivos en grafos de conocimiento abre la posibilidad de generalizar más allá de los elementos observados durante el entrenamiento; sin embargo, en escenarios reales, la novedad se manifiesta con frecuencia en la aparición de **entidades completamente nuevas** que deben integrarse al grafo existente. Este escenario, conocido como *Out-of-Knowledge-Base* (OOKB), introduce un desafío específico que no se resuelve únicamente con arquitecturas multirrelacionales, ya que requiere generar representaciones para entidades sin historial previo en la base de conocimiento. La formulación explícita de este problema marca un punto de inflexión en el campo, al trasladar la inducción desde un objetivo implícito hacia un desafío central en el modelado de grafos de conocimiento abiertos.

### Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach nowledge Transfer for Out-of-Knowledge-Base Entities

Hamaguchi et al. (2017)

El artículo aborda el problema de entidades fuera de la base de conocimiento (*Out-of-Knowledge-Base*, OOKB) en la tarea de Knowledge Base Completion (KBC), que consiste en predecir hechos sobre entidades que no estuvieron visibles durante el entrenamiento del modelo de embeddings y sin necesidad de volver a entrenar por completo el sistema. En modelos clásicos de knowledge graph embedding, las entidades desconocidas en el conjunto de entrenamiento no tienen vectores representativos, lo que impide inferir relaciones para ellas. Para superar esta limitación, los autores proponen un **método basado en Graph Neural Networks (GNNs) que calcula embeddings de las entidades nuevas en función de su vecindario estructural** y un conjunto limitado de hechos auxiliares disponibles en tiempo de prueba, aprovechando la información del grafo ya aprendido. En sus experimentos, este enfoque demuestra mejor desempeño en el escenario OOKB frente a métodos tradicionales y también consigue resultados competitivos en tareas estándar de KBC en conjuntos como WordNet11.

Aporte al estado del arte:  
Este trabajo fue uno de los primeros en formalizar y resolver el problema OOKB dentro de embeddings de grafos de conocimiento, creando un puente entre técnicas de embeddings y modelos basados en propagación de grafos. A diferencia de métodos clásicos de KBC que

requerían que todas las entidades aparecieran durante el entrenamiento, el uso de GNNs permite transferir conocimiento estructural a entidades no vistas, lo que es crucial para aplicaciones dinámicas de grafos donde la aparición de nuevas entidades es constante. Esto supone un avance significativo respecto a enfoques previos que dependían de recursos externos o retraining costoso, y sienta una base metodológica para posteriores estudios inductivos y extrapolativos en grafos —como modelos OOKB extendidos y métodos de extrapolación de entidades y relaciones— que buscan abordar precisamente la generalización más allá del conjunto entrenado.

## 6. Inducción centrada en relaciones

Aunque los enfoques OOKB permiten generar representaciones para entidades nuevas a partir de su contexto local, su efectividad depende en gran medida de la disponibilidad y calidad de dicho contexto. Para reducir esta dependencia, surge una línea de trabajo que desplaza el foco de la inducción desde las entidades hacia las **relaciones**, bajo la premisa de que la semántica relacional es más estable y transferible que las identidades individuales de los nodos. Esta perspectiva introduce un cambio conceptual relevante, al buscar patrones compartidos a nivel relacional que puedan reutilizarse para generalizar a entidades no observadas, sentando las bases para enfoques de extrapolación menos dependientes de información local específica.

### INGRAM: Inductive Knowledge Graph Embedding via Relation Graphs. (Ago. 2023)

Lee et al. (2020)

El artículo propone InGram, un método para embeddings inductivos en grafos de conocimiento que genera representaciones tanto para **nuevas entidades como para nuevas relaciones en tiempo de inferencia** —es decir, sin requerir que hayan aparecido durante el entrenamiento del modelo. Este enfoque supera una limitación importante de métodos inductivos previos, que permiten inferir hechos sobre entidades no vistas pero no generan vectores para relaciones nuevas. Para lograr esto, InGram construye **un grafo de relaciones —donde cada nodo corresponde a una relación y los pesos capturan su afinidad— y aprende a agregar información estructural de vecinos usando un mecanismo de atención que combina el grafo original de conocimiento y el grafo de relaciones**. Con ello, InGram puede proyectar entidades y relaciones no vistas a un espacio

de embedding útil para completar tripletas faltantes, demostrando rendimiento superior a 14 métodos de referencia en múltiples escenarios inductivos.

Aporte al estado del arte:

Este trabajo amplía el marco de aprendizaje inductivo en Knowledge Graphs al permitir la generalización conjunta a nuevos nodos y tipos de relaciones, algo que muchos métodos inductivos previos no contemplaban y que limita su aplicabilidad en grafos dinámicos. Al modelar explícitamente las relaciones entre relaciones en un grafo de afinidad y usar atención para generar embeddings estructurales, InGram ofrece una solución más flexible y expresiva para el problema de extrapolación de conocimiento en grafos, donde tanto entidades como relaciones emergentes deben ser incorporadas sin retraining completo. Esto lo hace especialmente relevante para tu proyecto, pues hace explícita la necesidad de capturar relaciones no explícitas y patrones estructurales complejos —vinculándose con conceptos como OOKB e inducido inductiveinductive y ayudando a delimitar el estado del arte en estrategias inductivas y extrapolativas para grafos de conocimiento.

## 7. Razonamiento estructural basado en subgrafos

Más allá de la inducción centrada en entidades o relaciones individuales, una línea de investigación propone que la **estructura local del grafo**, por sí misma, puede capturar regularidades semánticas suficientes para la inferencia. Bajo esta perspectiva, la predicción de enlaces se aborda como un problema de razonamiento sobre **patrones estructurales** compartidos, en lugar de depender de representaciones explícitas de entidades. Este enfoque desplaza el énfasis desde la generación de embeddings hacia el aprendizaje de **funciones de razonamiento transferibles**, marcando un paso clave hacia formas más generales de extrapolación de conocimiento en grafos.

### Link Prediction Based on Graph Neural Networks

Zhang y Chen (2018)

Proponen un enfoque para *link prediction* que desplaza el foco desde los embeddings globales de nodos hacia el **razonamiento estructural sobre subgrafos locales**. El método extrae, para cada par de nodos objetivo, un subgrafo inducido que captura los caminos y patrones estructurales relevantes entre ellos. Sobre este subgrafo, los autores aplican una Graph Neural Network que aprende representaciones discriminativas capaces de predecir la existencia de un enlace.

El trabajo demuestra que este enfoque permite a las GNNs aprender de forma implícita heurísticas clásicas de predicción de enlaces, como *common neighbors*, sin codificarlas manualmente. A diferencia de los modelos de embeddings tradicionales, el método no depende directamente de representaciones globales de nodos, sino de la estructura local que conecta a los nodos objetivo. Los resultados experimentales muestran mejoras consistentes frente a baselines basados en embeddings y heurísticas manuales, destacando el potencial del razonamiento por subgrafos como alternativa al aprendizaje puramente vectorial.

Aporte al estado del arte:

La principal aportación de Zhang y Chen (2018) al estado del arte es demostrar que la **estructura local del grafo contiene suficiente información semántica** para realizar predicción de enlaces, sin depender de embeddings globales de entidades. Este resultado responde indirectamente a las limitaciones del enfoque de Hamaguchi et al. (2017), al sugerir que la generalización no tiene por qué basarse en la transferencia de representaciones de entidades OOKB, sino en patrones estructurales reutilizables. Desde la perspectiva del survey de Chen et al. (2024), este trabajo constituye un **antecedente fundamental del razonamiento estructural**, que posteriormente se extiende al dominio de grafos de conocimiento multirrelacionales y escenarios inductivos estrictos en modelos como GrAIL. Así, Zhang y Chen establecen el fundamento conceptual para entender la extrapolación como un problema de **aprendizaje de reglas implícitas sobre subgrafos**, más que de generación de embeddings.

## Inductive Relation Prediction by Subgraph Reasoning (GrAIL).

K. K. Teru et al. (2020)

El trabajo Inductive Relation Prediction by Subgraph Reasoning aborda una de las principales limitaciones de los métodos tradicionales de Knowledge Graph Completion: su dependencia de un conjunto fijo de entidades durante el entrenamiento, lo que restringe su capacidad de generalización en grafos dinámicos. Frente a este problema, los autores proponen GrAIL, un modelo basado en redes neuronales sobre grafos que reformula la predicción de relaciones como un problema de razonamiento inductivo sobre subgrafos locales, en lugar de aprendizaje de embeddings de entidades.

La hipótesis central del enfoque es que la evidencia semántica necesaria para inferir una relación entre dos entidades se encuentra en los patrones estructurales de los caminos que las conectan, y que dichos patrones son independientes de la identidad específica de las

entidades. Para operacionalizar esta idea, GrailL extrae el subgrafo envolvente que conecta un par de nodos objetivo, etiqueta los nodos según su posición relativa respecto a estos y emplea una GNN relacional con mecanismos de atención para aprender composiciones relevantes de relaciones. A diferencia de los métodos de embeddings, el modelo no utiliza atributos ni representaciones preentrenadas de entidades, lo que lo obliga a aprender semántica relacional puramente a partir de la estructura del grafo.

El paper demuestra teóricamente que GrailL es capaz de representar un subconjunto significativo de reglas lógicas basadas en caminos, lo que le permite razonar de forma inductiva sobre relaciones no observadas. Experimentalmente, el modelo es evaluado en escenarios estrictamente inductivos, donde los conjuntos de entidades de entrenamiento y prueba son disjuntos, mostrando un desempeño superior frente a métodos de inducción de reglas y enfoques diferenciables existentes. Adicionalmente, los autores muestran que GrailL aporta un sesgo inductivo complementario a los métodos clásicos de embeddings, mejorando su desempeño cuando ambos enfoques se combinan.

Aporte al estado del arte:

El principal aporte de Inductive Relation Prediction by Subgraph Reasoning Teru et al.(2020) al estado del arte consiste en desplazar el foco de la extrapolación de conocimiento desde la transferencia de representaciones hacia el aprendizaje de funciones de razonamiento estructural. En contraste con enfoques como el de Hamaguchi et al. (2017), que buscan generar embeddings para entidades OOKB a partir de vecinos conocidos, GrailL elimina la necesidad de representar explícitamente a las entidades y demuestra que la generalización inductiva puede lograrse mediante patrones estructurales independientes de la identidad. Asimismo, el trabajo extiende y consolida las ideas de razonamiento por subgrafos introducidas por Zhang y Chen, llevándolas al contexto de grafos de conocimiento multirrelacionales y proporcionando tanto evidencia empírica como garantías teóricas de su capacidad expresiva. De este modo, el paper establece un marco más general y robusto para la predicción inductiva de relaciones, alineándose con escenarios realistas de grafos abiertos y en evolución.

## 8. Razonamiento secuencial y búsqueda de caminos en grafos de conocimiento

Una aproximación complementaria al razonamiento estructural considera la inferencia en grafos de conocimiento como un proceso de **composición secuencial de relaciones**,

donde la semántica emerge a partir de los caminos que conectan entidades. En este enfoque, el razonamiento se modela explícitamente como una secuencia de decisiones que recorre el grafo, priorizando interpretabilidad y trazabilidad de las inferencias. Aunque estos métodos no están diseñados principalmente para la extrapolación inductiva, introducen la noción de razonamiento como exploración estructurada, que influye en enfoques posteriores orientados a capturar patrones relacionales transferibles.

## DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning

Xiong et al. (2018)

Xiong et al. (2018) introducen **DeepPath**, un método para el razonamiento en grafos de conocimiento que modela la predicción de relaciones como un problema de **búsqueda de caminos guiada por aprendizaje por refuerzo**. En este enfoque, un agente aprende a navegar el grafo desde una entidad origen hasta una entidad destino mediante una secuencia de relaciones, maximizando una señal de recompensa asociada a la predicción correcta del enlace objetivo. El modelo utiliza representaciones neuronales para estados, acciones y políticas, permitiendo aprender estrategias de exploración que priorizan caminos semánticamente relevantes.

DeepPath reduce el espacio de búsqueda mediante técnicas de poda y *reward shaping*, haciendo viable el razonamiento en grafos de gran escala. Los experimentos muestran que el modelo puede descubrir rutas interpretables que explican la existencia de relaciones, ofreciendo una alternativa más explicable frente a los métodos de embeddings. Sin embargo, el enfoque asume que las entidades y relaciones están presentes durante el entrenamiento y depende de la exploración explícita del grafo, lo que limita su escalabilidad y su capacidad de generalización inductiva estricta.

Aporte al estado del arte:

La principal aportación de DeepPath al estado del arte es introducir el **razonamiento secuencial basado en caminos** como mecanismo explícito para inferir relaciones en grafos de conocimiento, reforzando la idea de que la semántica relacional puede emergir de composiciones de relaciones. En contraste con Hamaguchi et al. (2017), DeepPath no aborda directamente el escenario OOKB ni la generación inductiva de representaciones para entidades nuevas, ya que opera bajo un conjunto cerrado de entidades y relaciones. Desde la perspectiva del survey de Chen et al. (2024), DeepPath se sitúa como un antecedente de los enfoques de **razonamiento estructural**, aunque carece de mecanismos explícitos de extrapolación. Su relevancia radica en haber influido

conceptualmente en modelos posteriores, como GrAIL, que sustituyen la búsqueda explícita de caminos por razonamiento estructural aprendido y más generalizable.

## 9. Completado de grafos de conocimiento en escenarios open-world

Los enfoques inductivos revisados hasta ahora asumen, en distinta medida, la disponibilidad de estructura suficiente dentro del grafo y un conjunto de relaciones predefinido. Sin embargo, en escenarios reales esta suposición rara vez se cumple, ya que los grafos de conocimiento pueden incorporar **entidades y relaciones completamente nuevas**, incluso en ausencia de conectividad estructural significativa. El escenario *open-world* amplía el problema de generalización al requerir mecanismos capaces de integrar señales externas y manejar niveles más altos de incertidumbre, situando la extrapolación de conocimiento en un contexto más amplio que el abordado por los enfoques OOKB y estructurales.

### Open-World Knowledge Graph Completion for Unseen Entities and Relations via Attentive Feature Aggregation

Hwang et al.,( 2021)

Estudian el problema de **Open-World Knowledge Graph Completion (OW-KGC)**, donde tanto entidades como relaciones pueden no haber sido observadas durante el entrenamiento. A diferencia de los enfoques transductivos tradicionales, el modelo propuesto busca generalizar a componentes completamente nuevos incorporando **características externas y semánticas**, además de la estructura del grafo. El método introduce un mecanismo de **agregación atencional de características**, que combina información estructural del grafo con atributos textuales y descriptivos asociados a entidades y relaciones.

El modelo aprende a ponderar dinámicamente distintas fuentes de información según su relevancia para la predicción de enlaces, permitiendo manejar escenarios donde la conectividad estructural es limitada o inexistente. Los experimentos muestran que el enfoque supera a baselines inductivos y de embeddings en escenarios open-world, especialmente cuando se presentan relaciones no vistas. No obstante, el método depende

de la disponibilidad y calidad de características externas, lo que introduce un supuesto adicional sobre el acceso a información semántica auxiliar.

Aporte al estado del arte:

La principal aportación de Hwang et al. (2021) al estado del arte es **formalizar y atacar explícitamente el escenario open-world**, ampliando el alcance del problema OOKB introducido por Hamaguchi et al. (2017). Mientras que Hamaguchi se centra en generar embeddings para entidades nuevas conectadas al grafo, este trabajo permite generalizar incluso cuando aparecen **relaciones no vistas**, apoyándose en información semántica externa. Desde la taxonomía propuesta en el survey de Chen et al. (2024), este enfoque se sitúa en una categoría de extrapolación más amplia, aunque dependiente de atributos auxiliares, en contraste con métodos puramente estructurales como GrailL. Así, el paper representa una línea complementaria dentro de la *knowledge extrapolation*, donde la generalización se logra mediante la integración de señales externas más que exclusivamente a partir de la estructura del grafo.

## 10. Extrapolación temporal y meta-aprendizaje en grafos de conocimiento

Las secciones anteriores han abordado la extrapolación de conocimiento principalmente desde una perspectiva estructural y estática, donde el grafo se asume fijo y la generalización se plantea en términos de entidades o relaciones no observadas. No obstante, en escenarios reales los grafos de conocimiento evolucionan en el tiempo, incorporando nuevos hechos y modificando patrones relacionales existentes. Esta dinámica introduce un desafío adicional, ya que la extrapolación no solo debe capturar estructura y semántica, sino también **regularidades temporales transferibles**. En este contexto, la extrapolación temporal amplía el problema OOKB planteado por Hamaguchi et al. al incorporar explícitamente la dimensión temporal, y se posiciona, según la taxonomía de *knowledge extrapolation* de Chen et al., como una forma más general y exigente de generalización en grafos de conocimiento.

Meta-learning based knowledge extrapolation for temporal knowledge graphs.

Chen et al. (2023)

Este artículo aborda el problema de la extrapolación de conocimiento en grafos de conocimiento temporales (TKGs), donde no solo aparecen nuevas entidades y relaciones, sino también nuevos eventos en el tiempo. Los autores identifican que los métodos existentes de Temporal Knowledge Graph Embedding operan bajo un supuesto de mundo cerrado y, por tanto, no pueden generalizar cuando emergen componentes no vistos durante el entrenamiento.

Para resolver este problema, el trabajo propone MTKGE, un modelo de extrapolación temporal basado en meta-aprendizaje y redes neuronales sobre grafos, cuyo objetivo es aprender meta-conocimiento transferible entre relaciones. La hipótesis central es que existen patrones estructurales y temporales universales, independientes de las entidades específicas, que pueden reutilizarse para representar componentes no observados. En particular, el modelo captura dos tipos de patrones: (1) patrones de posición relativa entre relaciones, que reflejan regularidades estructurales en el grafo, y (2) patrones de secuencia temporal, que modelan el orden y la co-ocurrencia temporal de eventos.

Durante la fase de meta-entrenamiento, MTKGE aprende estos patrones a partir de tareas de predicción de enlaces muestreadas de TKGs existentes, simulando escenarios de extrapolación. Posteriormente, el meta-conocimiento aprendido se transfiere para generar embeddings de entidades y relaciones no vistas en grafos emergentes. Un módulo GCN integra estos embeddings con información contextual para realizar la predicción final de enlaces temporales.

Los experimentos en dos conjuntos de datos temporales diseñados específicamente para extrapolación muestran que MTKGE supera de manera consistente a métodos de extrapolación estáticos, a modelos temporales adaptados y a baselines basados en embeddings, demostrando una capacidad robusta para manejar entidades, relaciones y eventos temporales no observados.

Aporte al estado del arte:

La principal contribución de este trabajo al estado del arte es extender el problema de la knowledge extrapolation al dominio temporal, integrando de forma explícita la dimensión del tiempo en escenarios inductivos. A diferencia de enfoques previos como el de Hamaguchi et al. (2017), que se centran en transferir representaciones de entidades OOKB en grafos estáticos, MTKGE propone aprender meta-conocimiento relacional y temporal que puede transferirse a componentes completamente nuevos. En relación con enfoques inductivos basados en razonamiento estructural como Grall, este trabajo amplía el alcance del razonamiento inductivo incorporando patrones de secuencia temporal, demostrando que la extrapolación efectiva en grafos dinámicos requiere capturar no solo estructura, sino también regularidades temporales entre relaciones. Con ello, el paper establece un puente

entre meta-aprendizaje, extrapolación inductiva y razonamiento temporal, ofreciendo un marco más general para grafos de conocimiento abiertos y en evolución.

## 11. Evaluación rigurosa y benchmarks para generalización

Las mejoras en inducción y extrapolación de conocimiento en grafos requieren ser evaluadas bajo **criterios y conjuntos de datos adecuados**, ya que los *benchmarks* tradicionales suelen contener sesgos estructurales que inflan el desempeño de los modelos sin reflejar capacidades reales de razonamiento o generalización. En particular, cuando el objetivo es extraer información a entidades, relaciones o estructuras no vistas, la evaluación bajo supuestos transductivos puede resultar engañosa. En este contexto, la evaluación se convierte en un componente central del estado del arte, no como un aspecto metodológico secundario, sino como un mecanismo para distinguir entre ajuste a distribuciones sesgadas y **generalización genuina**. Esta sección revisa trabajos que abordan explícitamente la necesidad de *benchmarks* más rigurosos, en línea con el marco de *knowledge extrapolation* sistematizado por Chen et al., y que permiten evaluar de forma más realista los escenarios abiertos y dinámicos discutidos a lo largo de este capítulo.

### CoDEx: A comprehensive knowledge graph completion benchmark

Safavi et al. (2020)

El artículo CoDEx: A Comprehensive Knowledge Graph Completion Benchmark introduce CoDEx, un conjunto de benchmarks diseñados para evaluar de forma más rigurosa los modelos de Knowledge Graph Completion. Los autores argumentan que los conjuntos de datos tradicionales presentan sesgos estructurales que permiten a los modelos obtener buenos resultados sin realizar razonamiento semántico real. Para abordar este problema, CoDEx se construye a partir de Wikidata con un proceso de curaduría que garantiza relaciones semánticamente diversas, entidades no triviales y una menor dependencia de atajos estadísticos. El benchmark se presenta en tres tamaños (Small, Medium y Large) y se evalúa con una amplia gama de modelos de embeddings clásicos y neuronales. Los resultados muestran que muchos métodos con alto desempeño en benchmarks tradicionales experimentan caídas significativas en CoDEx, evidenciando que la tarea de completado de grafos sigue siendo un desafío abierto cuando se eliminan sesgos artificiales del conjunto de datos.

Aporte al estado del arte:

La principal aportación de CoDEx al estado del arte es establecer un estándar de evaluación más exigente y realista para modelos de completado de grafos de conocimiento. Aunque el paper no propone un nuevo modelo ni aborda directamente la extrapolación inductiva, su contribución es fundamental para trabajos como Hamaguchi, Grail o enfoques de knowledge extrapolation, al proporcionar un benchmark que reduce atajos estructurales y obliga a los modelos a capturar semántica relacional genuina. En este sentido, CoDEx funciona como una infraestructura crítica de evaluación, alineada con la necesidad de medir capacidades reales de generalización y razonamiento, más allá del ajuste a distribuciones sesgadas.

## 12. Marco unificador de la extrapolación de conocimiento en grafos

Las líneas de investigación revisadas hasta este punto abordan la generalización en grafos de conocimiento desde perspectivas diversas, incluyendo inducción estructural, razonamiento por subgrafos, escenarios *open-world* y extrapolación temporal. No obstante, la coexistencia de múltiples enfoques, supuestos y terminologías dificulta una comparación sistemática y una comprensión unificada del problema. En este contexto, surge la necesidad de un **marco conceptual integrador** que permita organizar estas contribuciones bajo un mismo lenguaje y delimitar claramente los distintos tipos de extrapolación. Esta sección introduce trabajos de síntesis que sistematizan el campo de la *knowledge extrapolation*, proporcionando una taxonomía que articula los escenarios, capacidades y limitaciones de los modelos existentes, y que sirve como base para identificar de manera precisa los vacíos que motivan la presente investigación.

### Generalizing to Unseen Elements: A Survey on Knowledge Extrapolation for Knowledge Graphs

Chen et al. (2024)

El artículo presenta una revisión exhaustiva del área de Knowledge Extrapolation aplicado a grafos de conocimiento (KGs), lista un conjunto de técnicas que buscan generalizar modelos de representación o razonamiento para manejar elementos no vistos durante el entrenamiento —como entidades o relaciones emergentes— sin volver a entrenar el modelo. Tradicionalmente, los métodos de knowledge graph embedding (KGE) asumen una base de conocimientos estática y conocida de antemano, lo que limita su uso en contextos

dinámicos o con elementos nuevos. Los autores proponen un marco taxonómico unificado, agrupando trabajos según escenarios como extrapolación de entidades, extrapolación de relaciones, inductivo, few-shot y zero-shot, y describen benchmarks relevantes para evaluar estos enfoques, así como relaciones entre métodos.

Aporte al estado del arte:

Este survey sistematiza y **unifica terminología** y enfoques en un campo fragmentado — incluyendo lineamientos para el problema out-of-knowledge-base (OOKB), inductivo, few-shot y zero-shot— proporcionando una visión clara de los desafíos y soluciones para extender modelos de KG más allá de los elementos vistos en entrenamiento. Tal sistematización es especialmente valiosa para investigaciones que buscan comprender cómo los KG pueden evolucionar con nuevas entidades y relaciones sin perder capacidad de predicción y extrapolación, conectando directamente con trabajos como el enfoque OOKB de Hamaguchi et al. (2017) y modelos inductivos recientes. Al ofrecer una taxonomía, comparación de benchmarks y sugerencias de líneas futuras, este paper funciona como referencia clave para delimitar el estado del arte en extrapolación de conocimiento en grafos, y contextualiza las limitaciones de los métodos clásicos frente a la necesidad de generalización real en aplicaciones dinámicas de KG.

## Glosario Analítico: Terminología de Extrapolación en Grafos de Conocimiento

Este análisis conceptual clasifica los términos clave del estado del arte en tres niveles: Paradigmas de Aprendizaje, Problemas Específicos y Arquitecturas/Modelos.

### A. Paradigmas y Conceptos Fundamentales

#### **Knowledge Extrapolation (Extrapolación de Conocimiento)**

Definición: Es el término paraguas (propuesto por Chen et al., 2024) que unifica todas las técnicas diseñadas para que un modelo funcione correctamente sobre elementos

(entidades, relaciones o eventos temporales) que no existían o no fueron vistos durante la fase de entrenamiento.

Relación: Es la evolución del término Inductive Learning. Mientras que "inductivo" suele referirse a la capacidad técnica del modelo, "extrapolación" define el problema macro de generalizar a dominios o tiempos desconocidos.

### **Transductivo vs. Inductivo**

Transductivo: Escenario clásico (ej. TransE). Asume que el grafo es estático y completo (Supuesto de Mundo Cerrado). Si entrenas con un conjunto de entidades {A,B,C}, el modelo solo puede predecir sobre {A,B,C}. No sabe qué hacer con una entidad nueva D.

Inductivo: Escenario moderno (ej. GraphSAGE, GrailL). El modelo aprende "reglas" o funciones de agregación. Si aparece una entidad nueva D, el modelo observa sus vecinos y calcula su representación en el momento, sin necesidad de reentrenar.

### **Open-World Assumption (Supuesto de Mundo Abierto)**

Definición: La premisa de que la base de conocimiento está incompleta por naturaleza. Lo que no está en el grafo no es necesariamente falso, solo "desconocido".

En este documento se asocia a modelos como el de Hwang et al., que utilizan información externa (texto, descripciones) para llenar los vacíos estructurales que los modelos puramente matemáticos no pueden resolver.

## **B. Definición de Problemas y Escenarios**

### **OOKB (Out-of-Knowledge-Base)**

Definición: Un problema específico dentro de la extrapolación que ocurre cuando aparecen entidades nuevas en tiempo de prueba.

Análisis: Fue formalizado por Hamaguchi et al.. La solución estándar para OOKB es usar la estructura de los vecinos (vecindario) para inferir la identidad del nuevo nodo.

Limitación: Originalmente, OOKB solo contemplaba entidades nuevas, pero asumía que las relaciones (los tipos de aristas) eran conocidas.

### **Few-Shot / Zero-Shot Learning**

Definición: Aprendizaje con pocos (few) o ningunos (zero) ejemplos previos.

Contexto: Se menciona en relación con el meta-aprendizaje (MTKGE) y INGRAM. Es la capacidad de entender una nueva relación (ej. "es\_padre\_de") habiendo visto solo uno o dos ejemplos de ella, transfiriendo conocimiento de relaciones similares (ej. "es\_madre\_de").

## **C. Arquitecturas y Modelos Específicos**

### **GNN / R-GCN (Relational Graph Convolutional Networks)**

GNN: Redes Neuronales en Grafos. La tecnología base que permite pasar mensajes entre nodos.

R-GCN: Una variante crucial para Knowledge Graphs. A diferencia de una GNN normal (que trata todas las aristas igual), la R-GCN tiene pesos diferentes para cada tipo de relación.

Importancia: Es el "motor" de la mayoría de los modelos modernos. Hamaguchi y GrailL usan variantes de GNNs/R-GCNs para procesar la información local.

### **INGRAM (Inductive Knowledge Graph Embedding via Relation Graphs)**

Concepto Clave: A diferencia de otros que solo miran nodos, INGRAM construye un Grafo de Relaciones.

Explicación: Imagina un grafo donde cada nodo es un concepto como "Padre", "Madre", "Hijo". INGRAM aprende qué tan cerca están estos conceptos.

Aporte: Permite la extrapolación de relaciones. Si aparece un tipo de relación nueva, INGRAM puede inferir cómo funciona basándose en su similitud con relaciones conocidas en su "grafo de relaciones".

### **GrAIL (Graph Inductive Learning)**

Concepto Clave: Razonamiento sobre subgrafos.

Explicación: GrAIL no aprende "quién es quién" (identidad), sino "qué forma tiene la conexión". Aprende patrones topológicos (ej. triángulos, caminos cerrados).

Aporte: Es el estándar de oro para la inducción pura. Al ser agnóstico a la identidad de las entidades, puede entrenarse en un grafo (ej. Biología) y funcionar razonablemente bien en otro distinto (ej. Química) si comparten estructuras lógicas subyacentes.

### **DeepPath**

Concepto Clave: Aprendizaje por Refuerzo (RL) para búsqueda de caminos.

Explicación: Un agente "camina" por el grafo saltando de nodo en nodo buscando el destino.

Diferencia: Aportar explicabilidad (traza el camino exacto de la decisión), pero es computacionalmente costoso y menos apto para extrapolación masiva que las GNNs.

### **MTKGE (Meta-learning based Knowledge Extrapolation)**

Concepto Clave: Meta-aprendizaje (aprender a aprender).

Explicación: En lugar de aprender un grafo fijo, aprende a adaptarse rápidamente a nuevas "instantáneas" (snapshots) de un grafo temporal.

Aporte: Resuelve el problema de la Extrapolación Temporal. Entiende que las reglas del grafo pueden cambiar con el tiempo.

## D. Evaluación

CoDEx (Comprehensive Knowledge Graph Completion Benchmark)

Definición: Un conjunto de datos (benchmark) curado para ser más difícil y realista.

Propósito: Eliminar los "sesgos fáciles" de datasets antiguos (como FB15k). En datasets viejos, los modelos podían "hacer trampa" memorizando patrones simples (ej. la relación inversa). CoDEx obliga al modelo a razonar de verdad, siendo fundamental para validar si un modelo de extrapolación funciona realmente o solo tiene suerte.

## Resumen Integral de Técnicas y Evolución hacia la Extrapolación de Conocimiento

El estado del arte revisado presenta una narrativa evolutiva clara en el campo de Knowledge Graph Reasoning, transitando desde modelos estáticos y cerrados hacia arquitecturas dinámicas capaces de generalizar ante lo desconocido. Este análisis desglosa los términos y técnicas clave, enfocándose en su interrelación y en los avances más recientes.

### 1. Del Paradigma Transductivo a la Necesidad Inductiva

El punto de partida técnico se define por el Knowledge Graph Embedding (KGE) clásico bajo un escenario transductivo. Modelos fundacionales como TransE y revisiones como la de Nickel et al. establecen representaciones vectoriales basadas en la identidad de las entidades.

Limitación Crítica: Operan bajo el Supuesto de Mundo Cerrado (Closed World Assumption). Si una entidad o relación no está presente durante el entrenamiento, el modelo no puede inferir nada sobre ella.

**Habilitadores Arquitectónicos:** La introducción de GraphSAGE (en grafos generales) y R-GCN (en grafos multirrelacionales) marca el primer quiebre técnico. Estos modelos cambian el objetivo de aprendizaje: en lugar de optimizar embeddings fijos para cada nodo, aprenden funciones de agregación (mensajes) sobre vecindarios locales. Aunque R-GCN se usa mayormente de forma transductiva, su arquitectura de paso de mensajes es el precursor necesario para la inducción.

## 2. El Salto a la Inducción: OOKB y Relaciones

La verdadera modernización comienza al abordar el problema de Out-of-Knowledge-Base (OOKB).

**Inducción basada en Entidades:** El trabajo de Hamaguchi et al. formaliza el problema OOKB. La técnica clave aquí es transferir conocimiento estructural: un nodo nuevo se representa mediante la agregación de sus vecinos conocidos.

**Inducción centrada en Relaciones (INGRAM):** Una innovación reciente (Lee et al.) aborda un vacío que los modelos OOKB anteriores ignoraban: la aparición de nuevas relaciones. La técnica distintiva es el uso de un Grafo de Relaciones (un grafo donde los nodos son las relaciones mismas) y mecanismos de atención para inferir afinidad entre tipos de relaciones no vistas. Esto permite extrapolar no solo a quién (entidades) sino a cómo se conectan (relaciones).

## 3. Razonamiento Estructural y Desacoplamiento de Identidad

Las técnicas más robustas y recientes desplazan el foco de los "embeddings de nodos" al razonamiento sobre subgrafos. Esta es la tendencia dominante actual para lograr una generalización genuina.

**Patrones sobre Identidades:** Zhang & Chen y posteriormente Grail (Teru et al.) demuestran que la capacidad de predecir un enlace reside en la estructura local (el subgrafo envolvente) y no en la identidad latente del nodo.

**Técnica de Grail:** Extrae subgrafos alrededor del par objetivo y aprende a identificar patrones topológicos. Al ser agnóstico a la identidad, el modelo puede operar en grafos completamente disjuntos (entrenar en un grafo A, probar en un grafo B), lo cual representa el estado más puro de aprendizaje inductivo.

Razonamiento Secuencial (DeepPath): Complementariamente, se explora el Reinforcement Learning (RL) para navegar el grafo. Aquí, el "razonamiento" se modela como la búsqueda de caminos explícitos, aportando interpretabilidad a la predicción, aunque con menor escalabilidad que los enfoques de subgrafos.

#### 4. Escenarios de Alta Complejidad: Open-World y Temporalidad

Las fronteras más actuales del estado del arte (2021-2023) integran dimensiones adicionales al problema.

Open-World KGC (Hwang et al.): Supera la inducción puramente estructural al integrar características semánticas externas (texto, descripciones). La técnica clave es la Attentive Feature Aggregation, que pondera dinámicamente la importancia de la estructura del grafo frente a la información externa, vital cuando la conectividad estructural es pobre o nula.

Extrapolación Temporal (MTKGE): Chen et al. (2023) introducen la dimensión tiempo en grafos dinámicos (TKGs). La innovación técnica es el uso de Meta-Learning (meta-aprendizaje). El modelo aprende "meta-conocimiento" sobre patrones secuenciales y estructurales que se transfieren rápidamente a nuevas instantáneas del grafo. Esto redefine la extrapolación como un problema de adaptación rápida a cambios evolutivos.

#### 5. Unificación y Evaluación Rigurosa

Taxonomía Unificadora (Chen et al., 2024): Introduce formalmente el término Knowledge Extrapolation. Este marco conceptual organiza todas las técnicas anteriores (OOKB, Inductivo, Few-shot) bajo un mismo paraguas, distinguiendo entre generalización a elementos no vistos (nodos/relaciones) y generalización en el tiempo.

Benchmarking (CoDEx): Se establece que las mejoras técnicas deben medirse sin sesgos. CoDEx elimina los "atajos estadísticos" de los datasets antiguos, obligando a los modelos (como Grail o R-GCN) a realizar un razonamiento semántico real y no solo memorización estructural.

#### Conclusión del Análisis:

El estado del arte muestra una transición desde la memorización de entidades (TransE) hacia el aprendizaje de reglas de razonamiento (Grail, MTKGE). Las técnicas actuales ya no buscan simplemente completar huecos en una base de datos estática, sino aprender la

física de las relaciones (estructurales, semánticas y temporales) para aplicarlas en escenarios abiertos y en constante cambio. La relación entre R-GCN (arquitectura), GraIL (lógica estructural) y Meta-learning (adaptabilidad) constituye el núcleo técnico de la investigación moderna.

## Referencias

- Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., & Yakhnenko, O. (2013). *Translating embeddings for modeling multi-relational data*. In **Advances in Neural Information Processing Systems** (Vol. 26). <https://papers.nips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html>
- Chen, M., Zhang, W., Geng, Y., Xu, Z., Pan, J. Z., & Chen, H. (2023). *Generalizing to unseen elements: A survey on knowledge extrapolation for knowledge graphs*. arXiv. <https://arxiv.org/abs/2302.01859>
- Chen, Z., Xu, C., Su, F., Huang, Z., & Dou, Y. (2023). *Meta-learning based knowledge extrapolation for temporal knowledge graphs*. In **Proceedings of the ACM Web Conference 2023 (WWW '23)** (pp. 1–11). Association for Computing Machinery. <https://doi.org/10.1145/3543507.3583279>
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). *Inductive representation learning on large graphs*. In **Advances in Neural Information Processing Systems** (Vol. 30). <https://arxiv.org/abs/1706.02216>
- Hamaguchi, T., Oiwa, H., Shimbo, M., & Matsumoto, Y. (2017). *Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach*. arXiv. <https://arxiv.org/abs/1706.05674>
- Hwang, J., Lee, J., Kim, S., & Kim, J. (2021). *Open-world knowledge graph completion for unseen entities and relations via attentive feature aggregation*. In **Proceedings of the AAAI Conference on Artificial Intelligence**, 35(5), 3797–3804. <https://ojs.aaai.org/index.php/AAAI/article/view/16507>

- Lee, J., Chung, C., & Whang, J. J. (2020). *INGRAM: Inductive knowledge graph embedding via relation graphs*. In **Proceedings of the AAAI Conference on Artificial Intelligence**, **34**(03), 4054–4061. <https://doi.org/10.1609/aaai.v34i03.5865>
- Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). *A review of relational machine learning for knowledge graphs*. **Proceedings of the IEEE**, **104**(1), 11–33. <https://doi.org/10.1109/JPROC.2015.2483592>
- Safavi, T., Koutra, D., & Koutra, D. (2020). *CoDEx: A comprehensive knowledge graph completion benchmark*. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)** (pp. 8328–8340). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.669>
- Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). *Modeling relational data with graph convolutional networks*. In **The Semantic Web – ESWC 2018** (Lecture Notes in Computer Science, Vol. 10843, pp. 593–607). Springer. [https://doi.org/10.1007/978-3-319-93417-4\\_38](https://doi.org/10.1007/978-3-319-93417-4_38)
- Teru, K. K., Denis, E., & Hamilton, W. L. (2020). Inductive relation prediction by subgraph reasoning. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 9448–9457). PMLR. <https://arxiv.org/abs/1911.06962>
- Xiong, W., Hoang, T., & Wang, W. Y. (2018). *DeepPath: A reinforcement learning method for knowledge graph reasoning*. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)** (pp. 1793–1802). Association for Computational Linguistics. <https://aclanthology.org/D18-1212/>
- Zhang, M., & Chen, Y. (2018). *Link prediction based on graph neural networks*. In **Advances in Neural Information Processing Systems** (Vol. 31). <https://arxiv.org/abs/1802.09691>