Stability and Generalization Capability of Subgraph Reasoning Models for Inductive Knowledge Graph Completion

Minsung Hwang, Jaejun Lee, and **Joyce Jiyoung Whang***
School of Computing, KAIST
The 42nd International Conference on Machine Learning (ICML 2025)

▼ BDILab

Main Contributions

- Propose a general framework for **subgraph reasoning models**
  - Derive their **stability** w.r.t. the perturbations of the **subgraph structure**
- **Introduce RTMD** designed for subgraph reasoning models
  - Use RTMD to **compute the stability** of subgraph reasoning models
- Analyze the **generalization bound** of the subgraph reasoning model
  - Discuss the impact of **stability** on their **generalization capability**
- Empirically **validate our theoretical findings** on real-world KGs
  - Examine how the **stability impacts generalization capability**

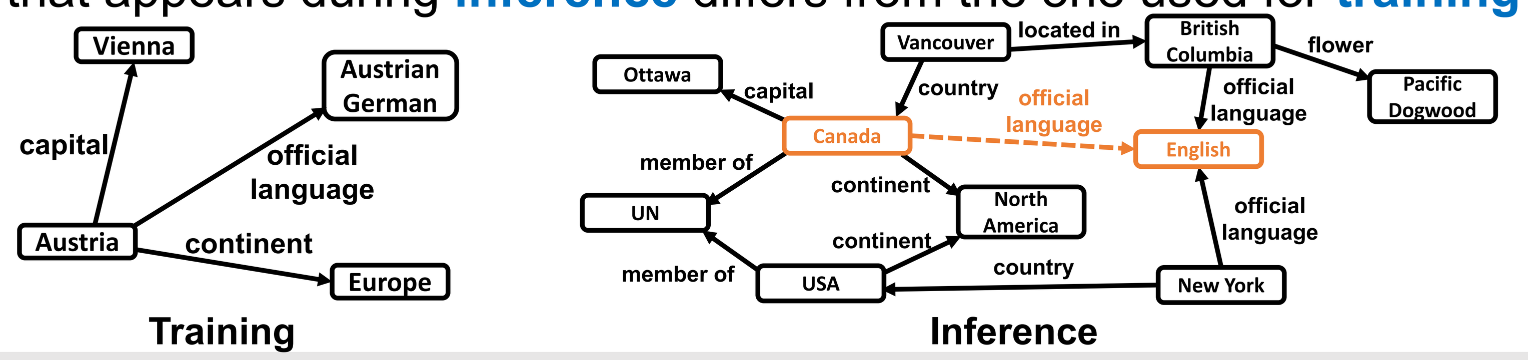
Inductive Knowledge Graph Completion

Knowledge Graph (KG)

- Represents **real-world knowledge** by **relationships between entities**

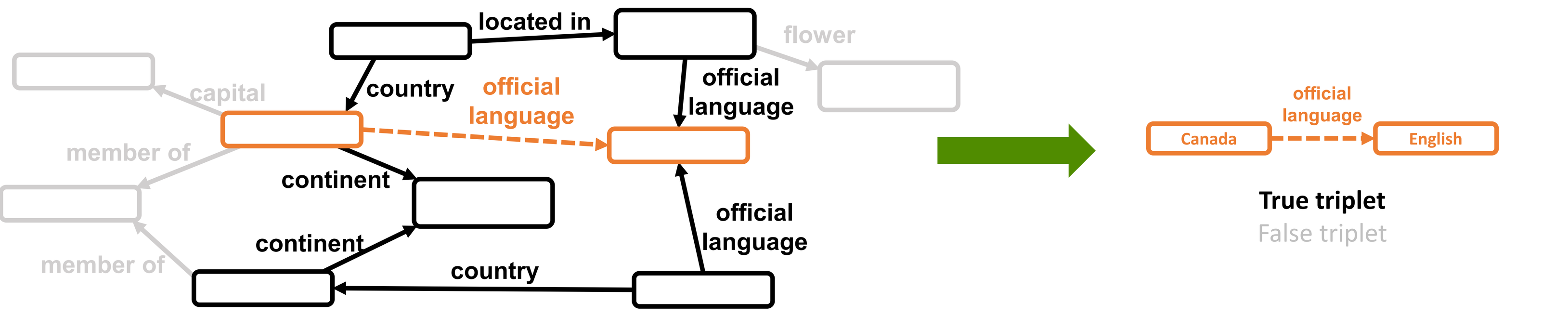
Inductive Knowledge Graph Completion (Inductive KGC)

- **Predicts missing triplets** within knowledge graphs
- KG that appears during **inference** differs from the one used for **training**



Subgraph Reasoning Model

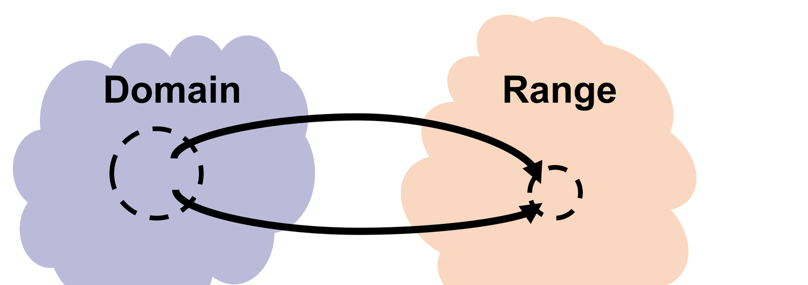
- Determines the validity of a triplet **using the subgraph**
  - **Extracts a subgraph** associated with a target triplet
  - **Relabels the entities** within the subgraph
  - **Computes a score** of the subgraph through **message-passing**



Theoretical Properties

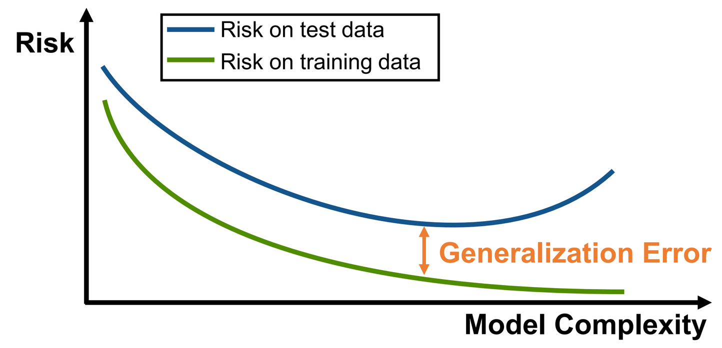
Stability

- **Consistency** of the model's output
- Measured by **Lipschitz constant**



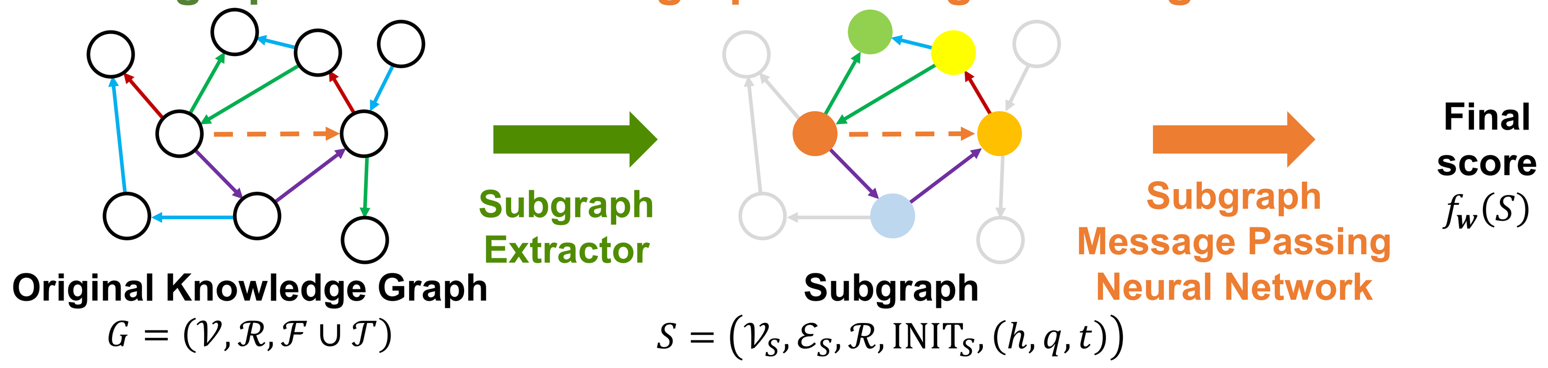
Generalization Capability

- **Performance discrepancy** between training and test data
- Measured by **generalization bounds**



General Framework for Subgraph Reasoning Model

- **Decomposing the subgraph reasoning model** into two parts
  - **Subgraph Extractor** / **Subgraph Message Passing Neural Network**



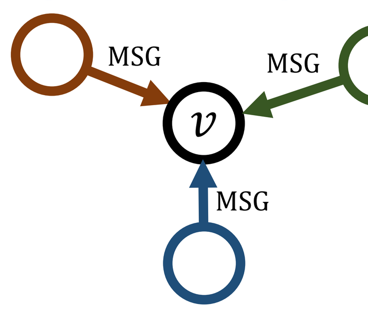
Subgraph Extractor

- A non-parametric function that **maps a triplet to a subgraph**
- Generates an **initial embedding** for each entity

Subgraph Message Passing Neural Network (SMPNN)

- Generalizes the scoring functions that utilize **message-passing**

$$x_S^{(0)}(v) = \mathrm{INIT}_S$$
$$\mathcal{M}_S^{(l)}(v) = \{\{\mathrm{MSG}^{(l)}\left(x_S^{(l-1)}(u), x_S^{(l-1)}(v), r, q\right) \mid (r,u) \in \mathcal{N}_S(v)\}\}$$
$$x_S^{(l)}(v) = \mathrm{UPD}^{(l)}\left(x_S^{(\theta(l))}(v), \mathrm{AGG}^{(l)}\left(\mathcal{M}_S^{(l)}(v)\right)\right)$$
$$f_w(S) = \mathrm{RD}\left(x_S^{(L)}(h), x_S^{(L)}(t), \mathrm{GRD}\left(\{\{x_S^{(L)}(u) \mid u \in \mathcal{V}_S\}\}\right), q\right)$$



Instantiation of Subgraph Reasoning Models

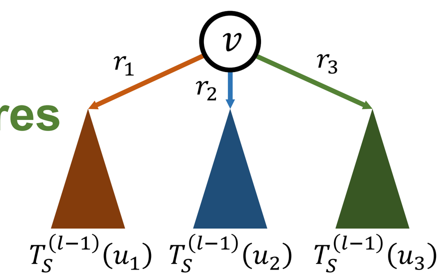
- By **appropriately configuring** the **subgraph extractor** and **SMPNN**, well-known subgraph reasoning models can be represented.

Models	Subgraph	MSG	AGG	UPD	GRD	RD
GraIL (ICML 2020)	Enclosing subgraph	Attention	Sum	Linear	Mean	Linear
NBFNet (NeurIPS 2021)	Union	TransE / DistMult / RotatE	Sum / Mean / PNA	Linear	-	MLP
RED-GNN (WWW 2022)	Union	Attention	Sum	Linear	-	Linear

Relational Tree Mover's Distance (RTMD)

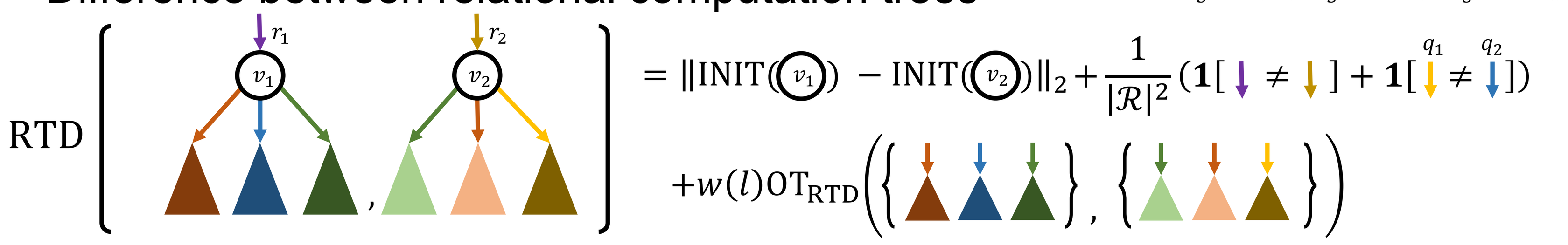
Relational Computation Tree

- Modeling how **SMPNNs process the subgraph structures**



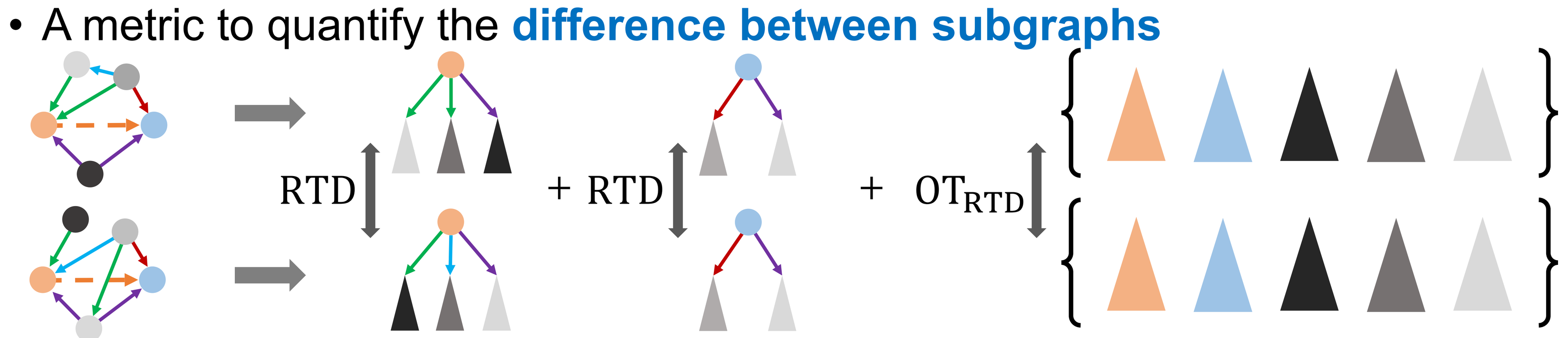
Relational Tree Distance (RTD)

- Difference between relational computation trees



Relational Tree Mover's Distance (RTMD)

- A metric to quantify the **difference between subgraphs**



Stability of Subgraph Reasoning Model

Lipschitz continuity of subgraph reasoning models

- **A score difference** is bounded by **an RTMD between subgraphs**

The upper bound of the Lipschitz constant

- Bounded by the **Lipschitz constant of each function** of the SMPNNs.

Theorem 4.5 Given $G_{\mathrm{tr}} = (\mathcal{V}_{\mathrm{tr}}, \mathcal{R}, \mathcal{F}_{\mathrm{tr}} \cup \mathcal{T}_{\mathrm{tr}})$, $G_{\mathrm{inf}} = (\mathcal{V}_{\mathrm{inf}}, \mathcal{R}, \mathcal{F}_{\mathrm{inf}} \cup \mathcal{T}_{\mathrm{inf}})$, and an SMPNN f_w with L layers, if the message, aggregation, update, global-readout, and readout function of f_w are Lipschitz continuous, then f_w is Lipschitz continuous with the **Lipschitz constant** η_f and the following holds:

$$\eta_f \le \prod_{l=1}^{L+1} \eta^{(l)} \ \ if\ \theta(k) = k-1, \qquad \eta_f \le (L+1)\prod_{l=1}^{L+1} \eta^{(l)} \ \ if\ \theta(k) = 0$$
$$\eta^{(l)} = \max\left(A_{\mathrm{upd}}^{(l)} + d_{\max} B_{\mathrm{upd}}^{(l)} A_{\mathrm{agg}}^{(l)} B_{\mathrm{msg}}^{(l)}, B_{\mathrm{upd}}^{(l)} A_{\mathrm{agg}}^{(l)} A_{\mathrm{msg}}^{(l)}, |\mathcal{R}|^2 B_{\mathrm{upd}}^{(l)} A_{\mathrm{agg}}^{(l)} C_{\mathrm{msg}}^{(l)}, |\mathcal{R}|^2 B_{\mathrm{upd}}^{(l)} A_{\mathrm{agg}}^{(l)} D_{\mathrm{msg}}^{(l)}, 1\right)$$
$$\eta^{(L+1)} = \max\left(A_{\mathrm{rd}}, B_{\mathrm{rd}}, C_{\mathrm{rd}} A_{\mathrm{grd}}, \frac{|\mathcal{R}|^2 D_{\mathrm{rd}}}{2 + \max(|\mathcal{V}_{\mathrm{tr}}|, |\mathcal{V}_{\mathrm{inf}}|)}\right)$$

where $1 \le l \le L$, and A, B, C, D are **the Lipschitz constants of the corresponding function**.

Generalization Bound of Subgraph Reasoning Models

Expected Risk Discrepancy

- **Discrepancy** between the **expected risks** measured on each KG

Expected Risk Discrepancy

$$D(\mathcal{P}, \lambda, \gamma) = \ln\left(\mathbb{E}_{w\sim\mathcal{P}}\left[\exp\left(\lambda\left(\mathcal{L}_{G_{\mathrm{tr}}}\left(f_w, \frac{\gamma}{2}\right) - \mathcal{L}_{G_{\mathrm{inf}}}(f_w, \gamma)\right)\right)\right]\right)$$

Generalization Bound of Subgraph Reasoning Model

- Depends on the **KL divergence** and **expected risk discrepancy**

Theorem 5.3 Given G_{tr}, G_{inf}, and a subgraph reasoning model with a subgraph extractor g and an SMPNN f_w, for any prior distribution $\mathcal{P}$ and posterior distribution $\mathcal{Q}$ on the parameter space of f_w constructed by adding random noise $\ddot{w}$ to w such that $\mathbb{P}\left(\max\left(\max_{e\in\mathcal{T}_{\mathrm{tr}}}|f_{\tilde{w}}(g(G_{\mathrm{tr}},e)) - f_w(g(G_{\mathrm{tr}},e))|, \max_{e\in\mathcal{T}_{\mathrm{inf}}}|f_{\tilde{w}}(g(G_{\mathrm{inf}},e)) - f_w(g(G_{\mathrm{inf}},e))|\right)\right)$, and $\gamma, \lambda > 0$, the following holds with probability at least $1-\delta$

$$\mathcal{L}_{G_{\mathrm{inf}}}(f_w, 0) \le \hat{\mathcal{L}}_{G_{\mathrm{tr}}}(f_w, \gamma) + \frac{1}{\lambda}\left(2\mathrm{KL}(\mathcal{Q}|\mathcal{P}) + \ln\frac{4}{\delta} + \frac{\lambda^2}{4|\mathcal{T}_{\mathrm{tr}}|} + D\left(\mathcal{P}, \lambda, \frac{\gamma}{2}\right)\right)$$

where $D\left(\mathcal{P}, \lambda, \frac{\gamma}{2}\right)$ is the expected risk discrepancy between G_{tr} and G_{inf}, and $\mathrm{KL}(\mathcal{Q}|\mathcal{P})$ is a KL divergence of $\mathcal{Q}$ from $\mathcal{P}$.

Upper Bound of the Expected Risk Discrepancy

- As **the stability increases**, **the expected risk discrepancy decreases**

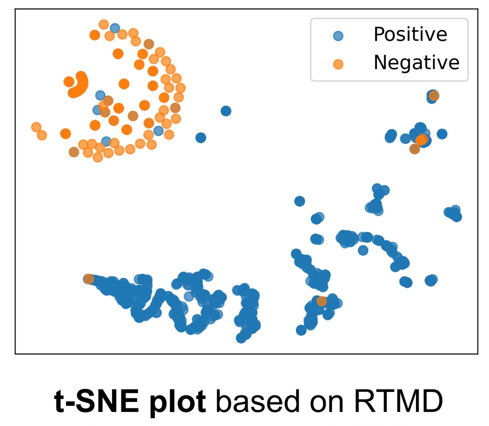
Theorem 5.5 Given G_{tr}, G_{inf}, and a subgraph reasoning model with a subgraph extractor g and an SMPNN f_w with stability C_f, for any prior distribution $\mathcal{P}$ and posterior distribution $\mathcal{Q}$ on the parameter space of f_w, and $\lambda > 0$, the following holds:

$$D(\mathcal{P}, \lambda, \gamma) \le \lambda\left(\max\left(0, \frac{|\mathcal{T}_{\mathrm{tr}}|}{|\mathcal{T}_{\mathrm{inf}}|} - 1\right) + \frac{2\mathrm{OT}_{\mathrm{RTMD}}(\psi(\mathcal{T}_{\mathrm{inf}}, \mathcal{T}_{\mathrm{tr}}))}{\gamma C_f \max(|\mathcal{T}_{\mathrm{inf}}|, |\mathcal{T}_{\mathrm{tr}}|)}\right)$$

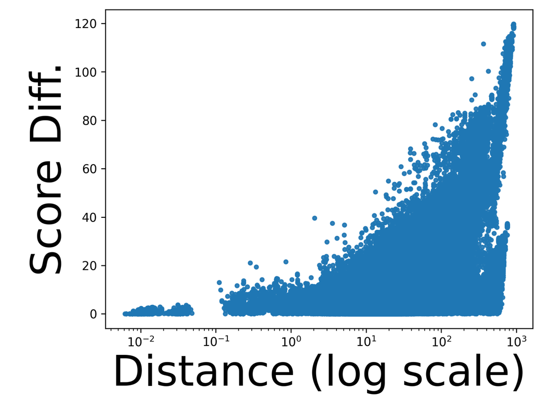
Experiments

Empirically validate our theoretical findings

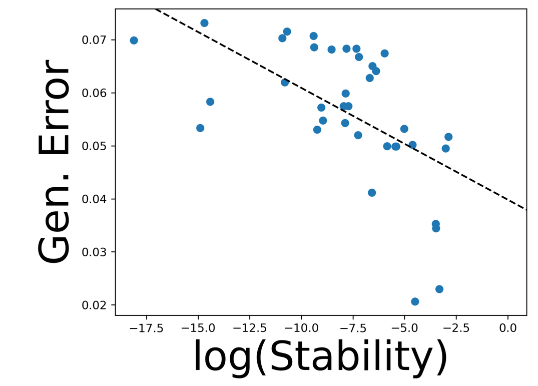
- Conduct experiments on **real-world KGs** (FB15K237, WN18RR, NELL-995)



t-SNE plot based on RTMD
Classification acc: 0.8205



Compare **score differences** and **RTMD**



Compare **stability** and **gen. error**
Corr: -0.5759, p-value: 0.00019

RTMD is a valid metric for quantifying the distance between subgraphs

SMPNN is Lipschitz continuous with respect to the RTMD

A **more stable subgraph reasoning model** tends to exhibit **better generalization capability**