Contrastive Language-Entity Pre-training for Richer Knowledge Graph Embedding

Andrea Papaluca, Daniel Krefl, Artem Lensky, Hanna Suominen

Australian National University

Multimodal Learning

Multimodal Learning

- Integration of multiple data modalities

Multimodal Learning

- Integration of multiple data modalities
- Text and Images, or Text and Graph data for instance

Multimodal Learning

- Integration of multiple data modalities
- Text and Images, or Text and Graph data for instance
- Xie et al, Yao et al, Shen et al, Wang et al

  → Textual descriptions improve KG representation learning

Multimodal Learning

- Integration of multiple data modalities
- Text and Images, or Text and Graph data for instance
- Xie et al, Yao et al, Shen et al, Wang et al

  → Textual descriptions improve KG representation learning
- CLIP (Radford et al)

  → Joint multimodal text-image space

  → Stable Diffusion

Paradisiac beach of a tropical South-Korean island

Multimodal Learning

- Integration of multiple data modalities
- Text and Images, or Text and Graph data for instance
- Xie et al, Yao et al, Shen et al, Wang et al
  → Textual descriptions improve KG representation learning
- CLIP (Radford et al)
  → Joint multimodal text-image space
  → Stable Diffusion
- Can we learn a similar multimodal text-graph space
  for Knowledge Graphs?

Paradisiac beach of a tropical South-Korean island

Multimodal Learning

- Integration of multiple data modalities
- Text and Images, or Text and Graph data for instance
- Xie et al, Yao et al, Shen et al, Wang et al

  → Textual descriptions improve KG representation learning
- CLIP (Radford et al)

  → Joint multimodal text-image space

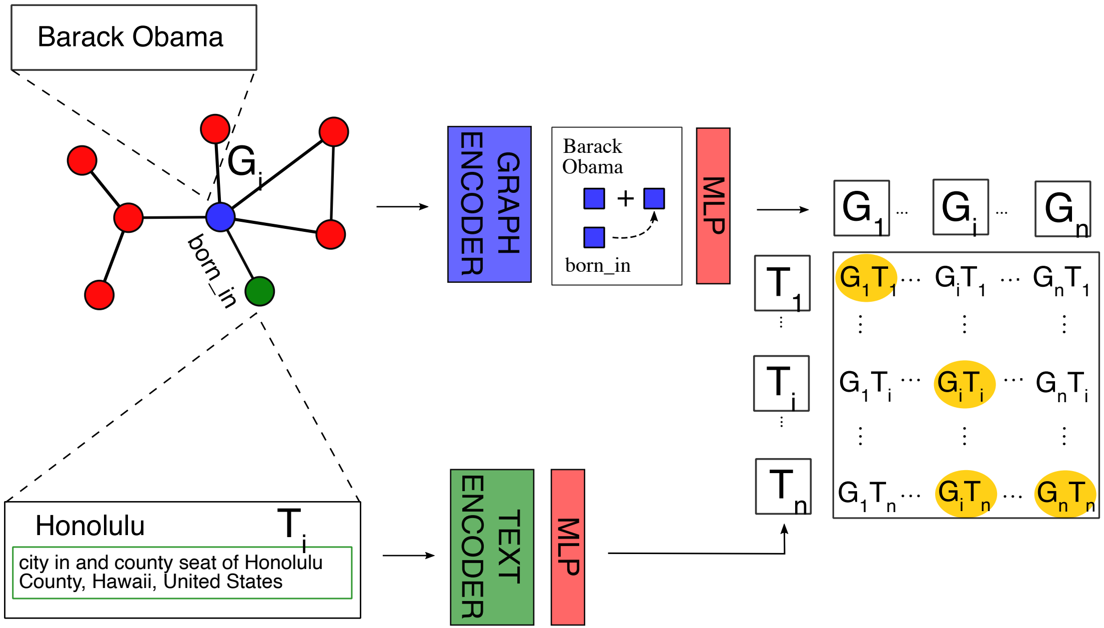
  → Stable Diffusion
- Can we learn a similar multimodal text-graph space

  for Knowledge Graphs?
- Does it implicitely yield better KG representations?

Paradisiac beach of a tropical South-Korean island

The CLEP Architecture

$(Barack\ Obama,\ born\_in,\ Honolulu)$



A Forward Pass

e^{head} : head node of the relational triplet

d^{tail} : description of the tail node

Batch of KG triplets

$$\{(e_1^{head}, r_1, d_1^{tail}),\ \ldots\ ,(e_n^{head}, r_n, d_n^{tail})\}$$

A Forward Pass

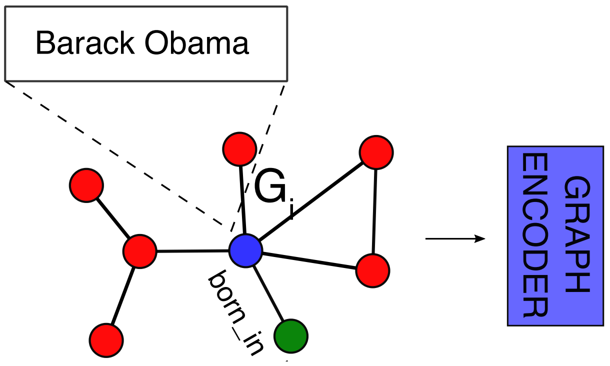
e^{head} : head node of the relational triplet

d^{tail} : description of the tail node

Batch of KG triplets

$$\{(e_1^{head}, r_1, d_1^{tail}), \ldots, (e_n^{head}, r_n, d_n^{tail})\}$$

$$(h_i^{(g)}, \rho_i^{(g)}) = \text{GraphEncoder}(e_i^{head}, r_i)$$



A Forward Pass

e^{head} : head node of the relational triplet

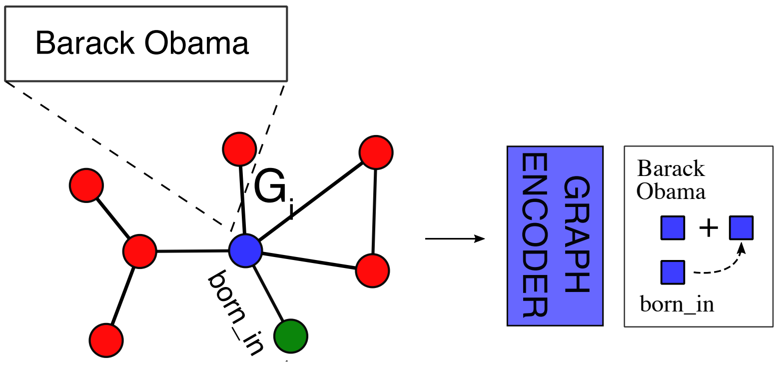
d^{tail} : description of the tail node

Batch of KG triplets

$$\{(e_1^{head}, r_1, d_1^{tail}), \ldots, (e_n^{head}, r_n, d_n^{tail})\}$$

$$(h_i^{(g)}, \rho_i^{(g)}) = \text{GraphEncoder}(e_i^{head}, r_i)$$

$$x_i^{(g)} = h_i^{(g)} + \rho_i^{(g)}$$



A Forward Pass

e^{head} : head node of the relational triplet

d^{tail} : description of the tail node

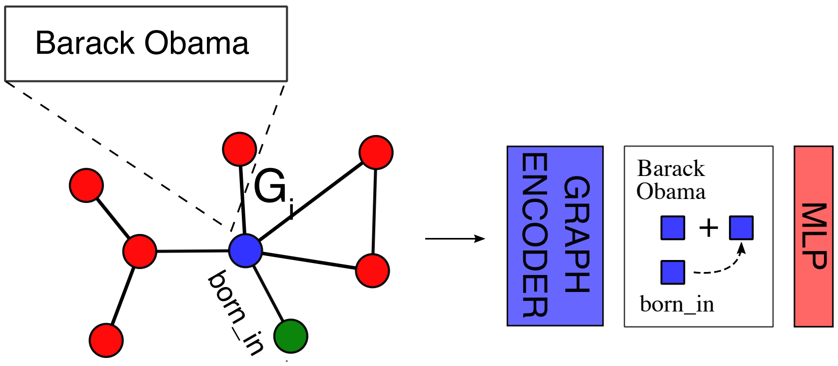
Batch of KG triplets

$$\{(e_1^{head}, r_1, d_1^{tail}), \ldots, (e_n^{head}, r_n, d_n^{tail})\}$$

$$(h_i^{(g)}, \rho_i^{(g)}) = \text{GraphEncoder}(e_i^{head}, r_i)$$

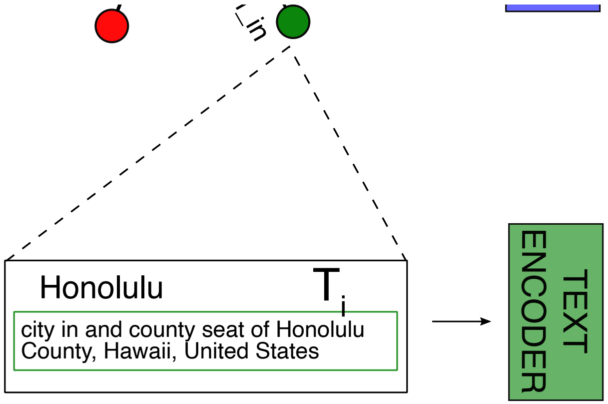
$$x_i^{(g)} = h_i^{(g)} + \rho_i^{(g)}$$

$$\tilde{x}_i^{(g)} = \text{MLP}_g(x_i^{(g)})$$



A Forward Pass

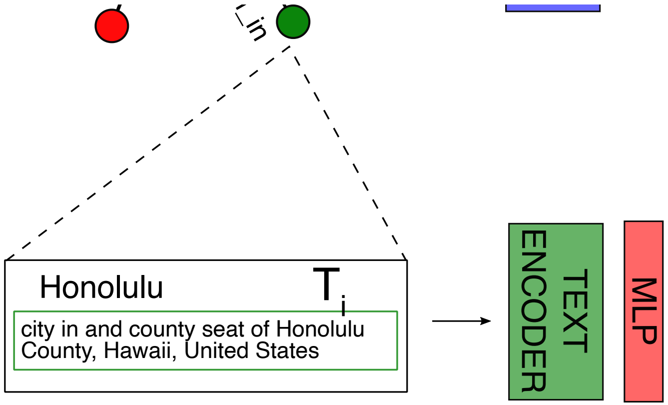
$$x_i^{(t)} = \text{TextEncoder}\left(d_i^{tail}\right)$$



A Forward Pass

$$x_i^{(t)} = \text{TextEncoder}\left(d_i^{tail}\right)$$

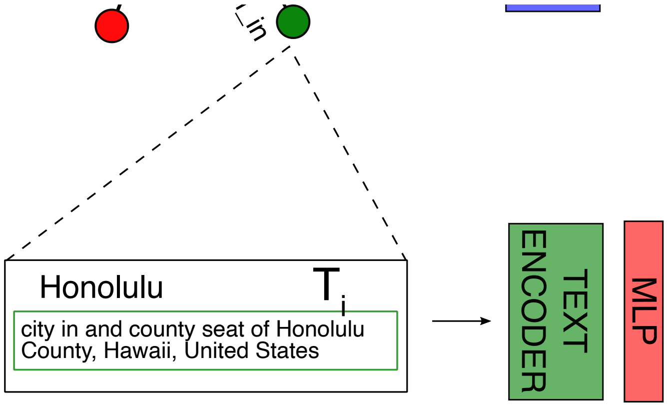
$$\tilde{x}_i^{(t)} = \text{MLP}_t\left(x_i^{(t)}\right)$$

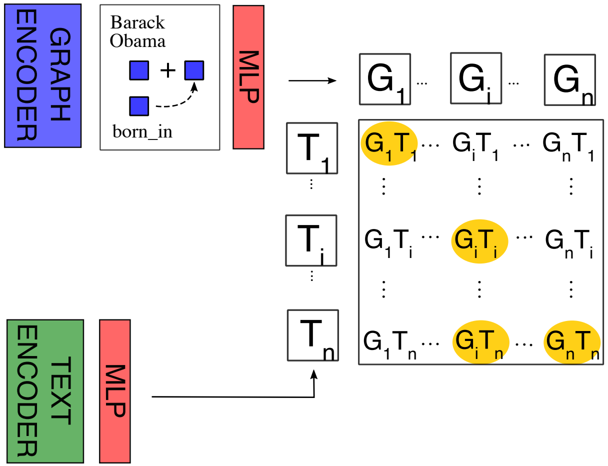


A Forward Pass

$$x_i^{(t)} = \mathrm{TextEncoder}\left(d_i^{tail}\right)$$

$$\tilde{x}_i^{(t)} = \mathrm{MLP}_t\left(x_i^{(t)}\right)$$





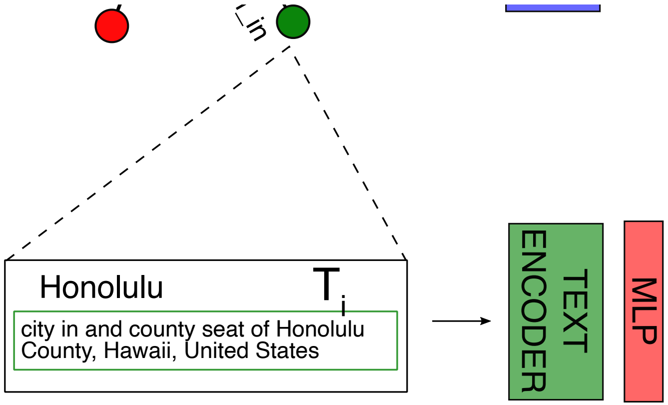
Cosine similarity matrix

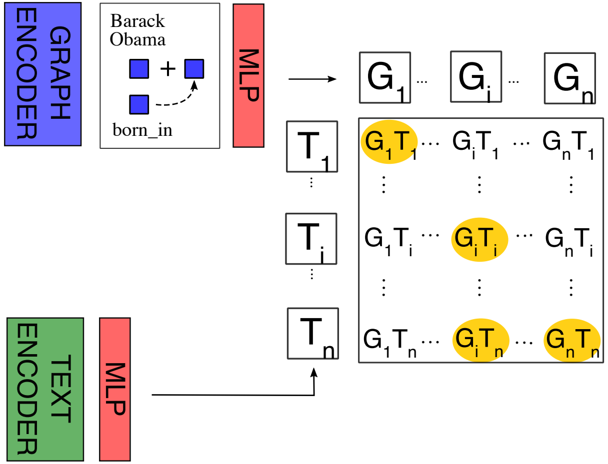
$$m_{i,j} = \frac{\tilde{x}_i^{(g)} \cdot \tilde{x}_j^{(t)}}{\|\tilde{x}_i^{(g)}\| \|\tilde{x}_j^{(t)}\|} \cdot e^{\tau}$$

A Forward Pass

$$x_i^{(t)} = \text{TextEncoder}\left(d_i^{tail}\right)$$

$$\tilde{x}_i^{(t)} = \text{MLP}_t\left(x_i^{(t)}\right)$$





Cosine similarity matrix

$$m_{i,j} = \frac{\tilde{x}_i^{(g)} \cdot \tilde{x}_j^{(t)}}{\|\tilde{x}_i^{(g)}\| \|\tilde{x}_j^{(t)}\|} \cdot e^{\tau}$$

τ : temperature scaling the logits

A Forward Pass

Row-wise Cross Entropy (CE)

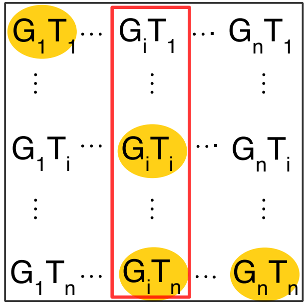
$$\mathrm{CE}(M) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{e^{m_{i,i}}}{\sum_{j=1}^{n} e^{m_{i,j}}}$$

$$\begin{bmatrix} G_1T_1 & \cdots & G_iT_1 & \cdots & G_nT_1 \\ \vdots & & \vdots & & \vdots \\ G_1T_i & \cdots & G_iT_i & \cdots & G_nT_i \\ \vdots & & \vdots & & \vdots \\ G_1T_n & \cdots & G_iT_n & \cdots & G_nT_n \end{bmatrix}$$

A Forward Pass

Row-wise Cross Entropy (CE)

$$\mathrm{CE}(M) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{e^{m_{i,i}}}{\sum_{j=1}^{n} e^{m_{i,j}}}$$

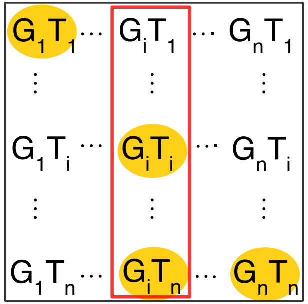


The column-wise CE is obtained by simply taking $M \to M^T$

A Forward Pass

Row-wise Cross Entropy (CE)

$$\mathrm{CE}(M) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{e^{m_{i,i}}}{\sum_{j=1}^{n} e^{m_{i,j}}}$$



The column-wise CE is obtained by simply taking $M \rightarrow M^T$

$$\mathcal{L} = \frac{1}{2}\Big(\mathrm{CE}(M) + \mathrm{CE}(M^\top)\Big)$$

$\longrightarrow$ Enforces minimization of incorrect entity-description associations simultaneously in rows and columns!

The aligned Text-Graph space

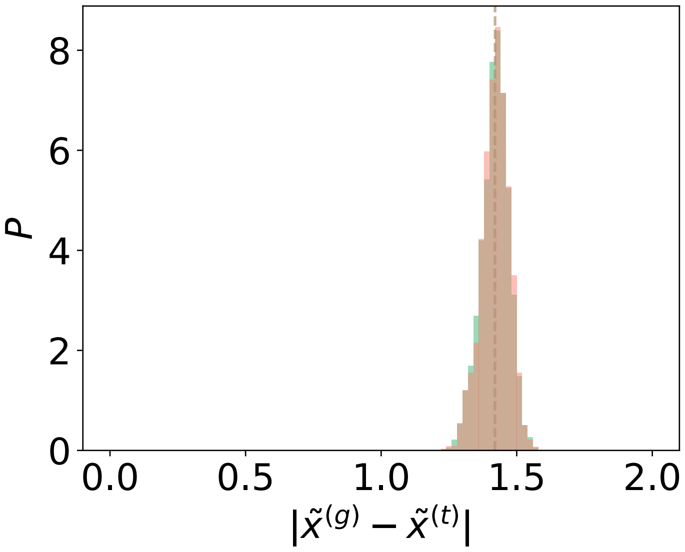
Euclidean distance of the correct/incorrect entity-description associations

$$P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\|\right) \qquad P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}\right)$$

The aligned Text-Graph space

Euclidean distance of the correct/incorrect entity-description associations

$$P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\|\right) \qquad P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}\right)$$

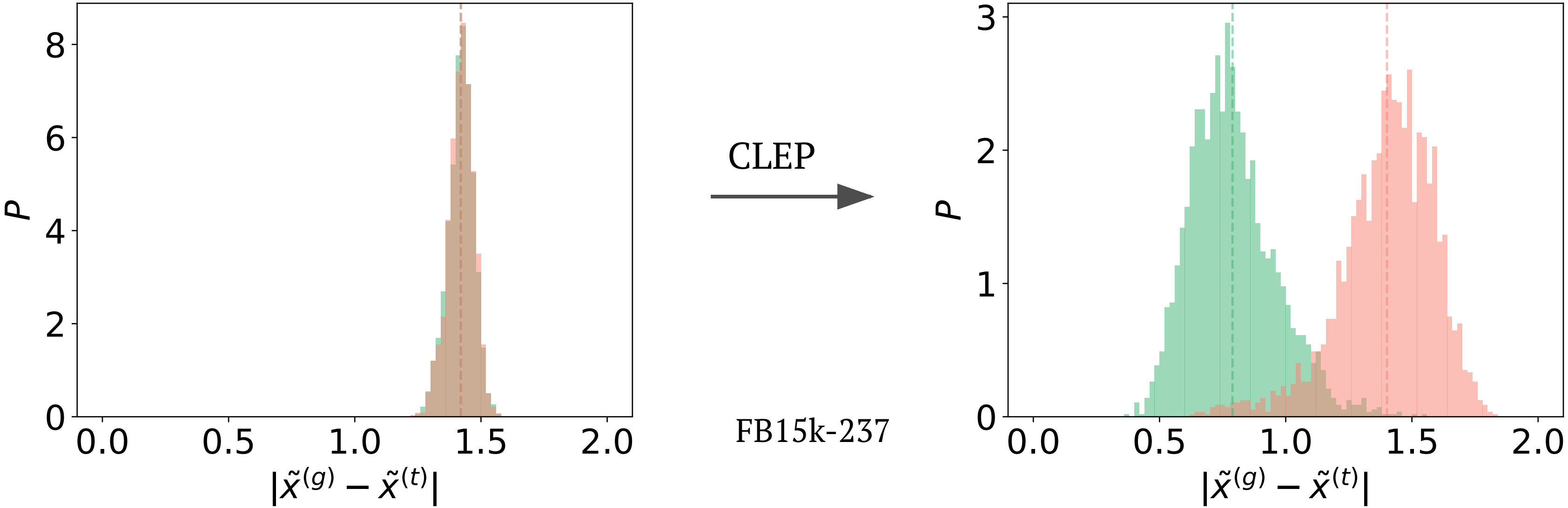


FB15k-237

The aligned Text-Graph space

Euclidean distance of the correct/incorrect entity-description associations

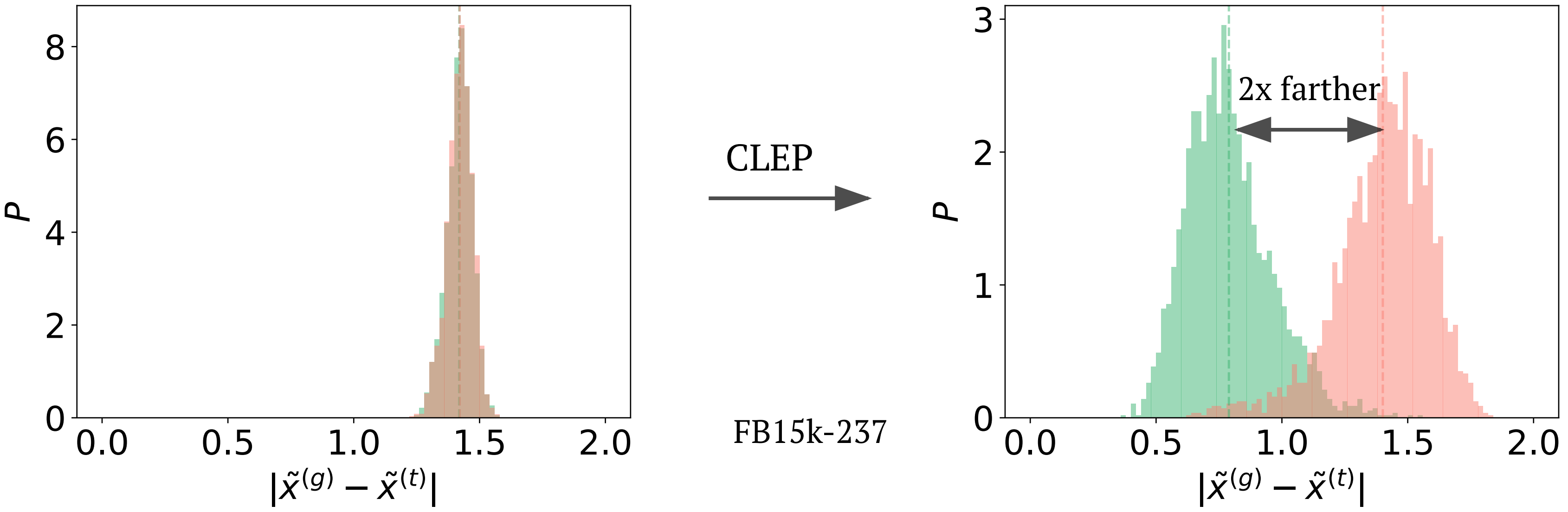
$$P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\|\right) \qquad P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}\right)$$



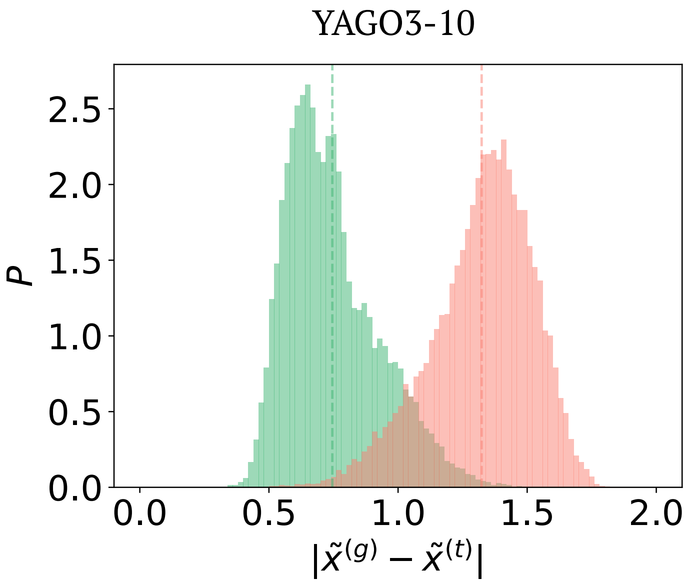
The aligned Text-Graph space

Euclidean distance of the correct/incorrect entity-description associations

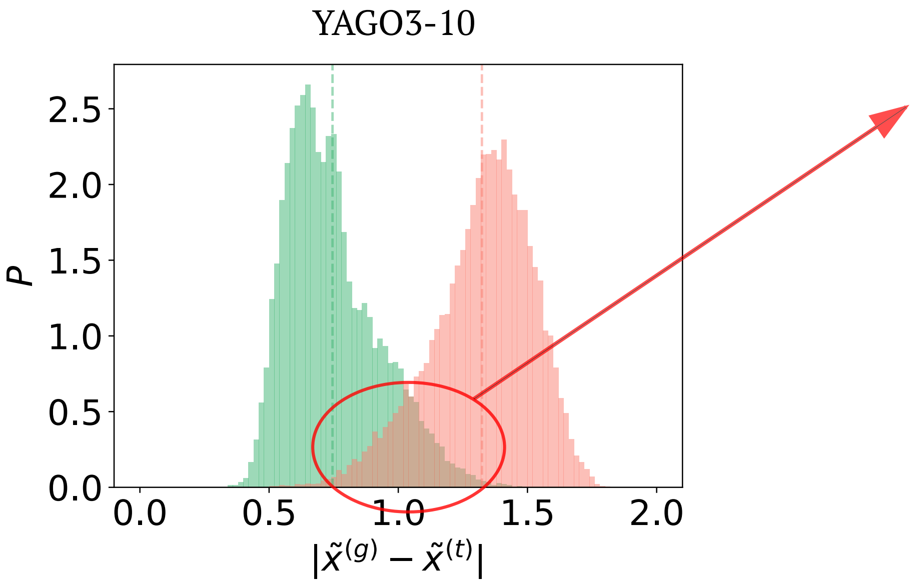
$$P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\|\right) \qquad P\left(\|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}\right)$$



The aligned Text-Graph space



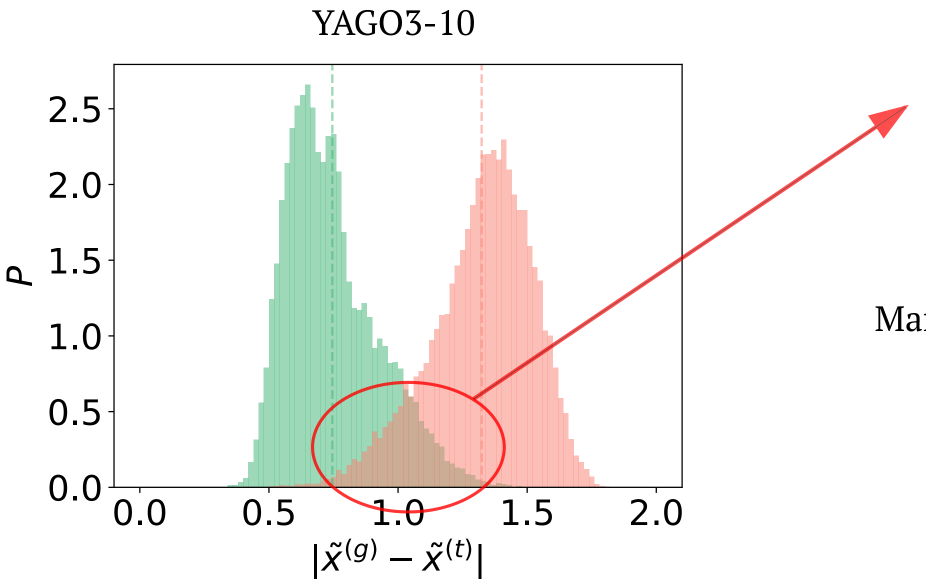
The aligned Text-Graph space



Incorrect pairs closer than correct ones

$$\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\| \;\geq\; \|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}$$

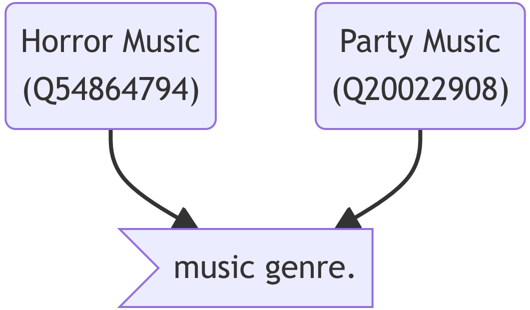
The aligned Text-Graph space



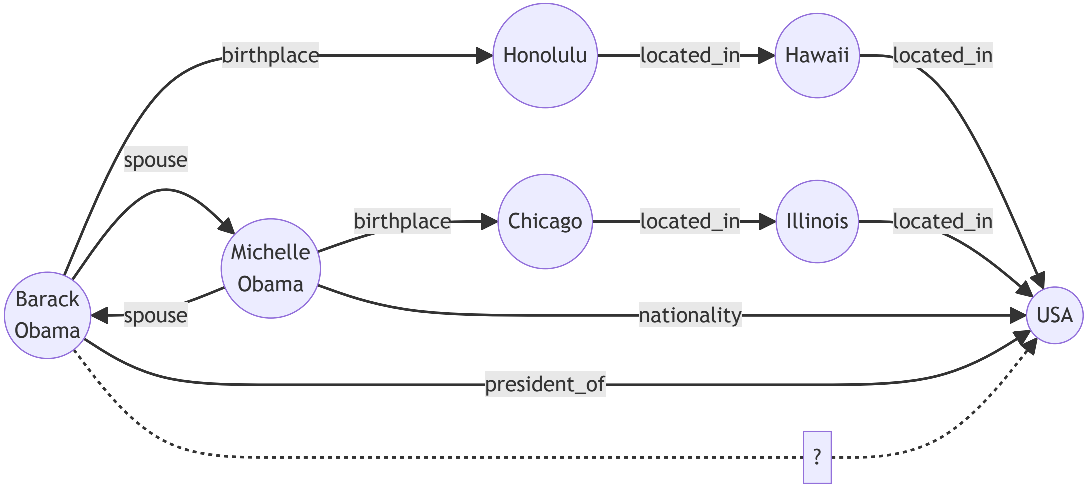
Incorrect pairs closer than correct ones

$$\|\tilde{x}_i^{(g)} - \tilde{x}_i^{(t)}\| \geq \|\tilde{x}_i^{(g)} - \tilde{x}_j^{(t)}\|_{i \neq j}$$

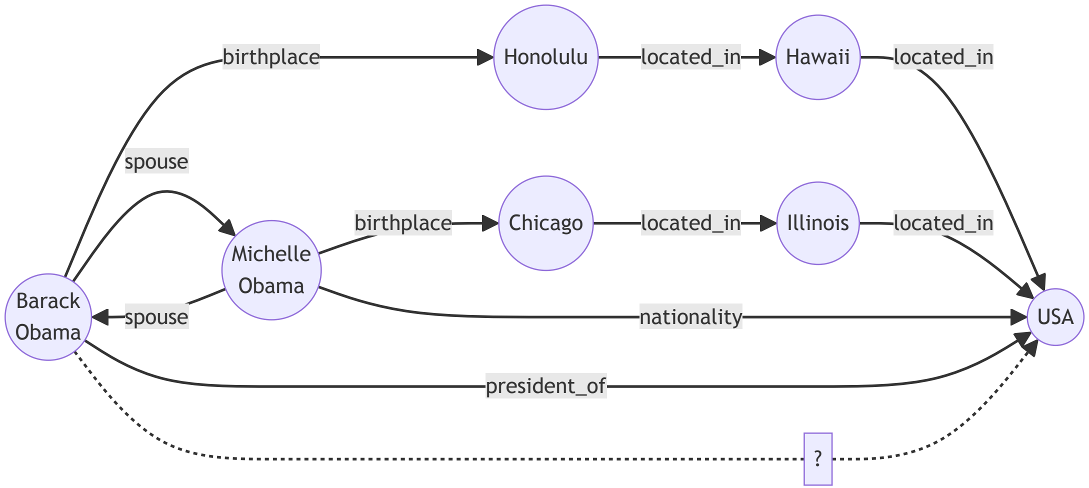
Many descriptions are shared over different entities



Link Prediction across spaces



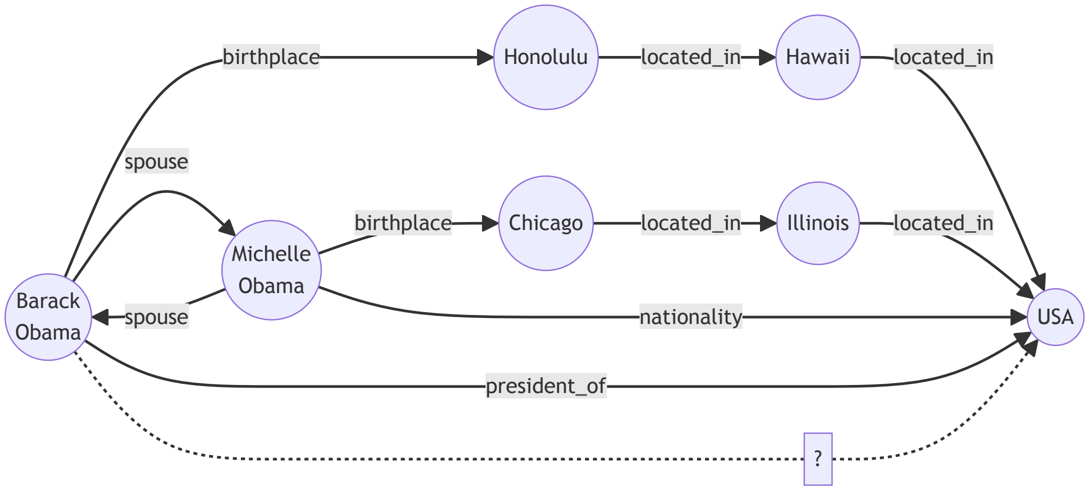
Link Prediction across spaces



What's Barack Obama's Nationality?

$$f_s(\text{Barack Obama},\ \text{nationality},\ v) \quad \forall v \in \mathcal{G}$$

Link Prediction across spaces



What's Barack Obama's Nationality?

$$f_s(\text{Barack Obama},\ \text{nationality},\ v) \quad \forall v \in \mathcal{G}$$

Rank	f_s	Link
1	0.91	(Barack Obama, nationality, USA)
2	0.53	(Barack Obama, nationality, Hawaii)
3	0.44	(Barack Obama, nationality, Illinois)
.	.	.
.	.	.
.	.	.
n	0.11	(Barack Obama, nationality, Michelle Obama)

Link Prediction across spaces

- CLEP is trained to align head entities with tails descriptions $e^{head} + r \sim d^{tail}$

$$f_s(\text{Barack Obama},\ \text{nationality},\ d(v)) \quad \forall v \in \mathcal{G}$$

Link Prediction across spaces

- CLEP is trained to align head entities with tails descriptions $e^{head} + r \sim d^{tail}$

$$f_s(\text{Barack Obama},\ \text{nationality},\ d(v)) \quad \forall v \in \mathcal{G}$$

node $\in$ graph space

description $\in$ text space

Link Prediction across spaces

- CLEP is trained to align head entities with tails descriptions $e^{head} + r \sim d^{tail}$

$$f_s(\text{Barack Obama},\ \text{nationality},\ d(v)) \quad \forall v \in \mathcal{G}$$

$\text{node} \in \text{graph space}$

$\text{description} \in \text{text space}$

Cosine Similarity score

$$f_s(h, r, t) = \frac{\mathrm{CLEP}_g(h,\ r)\ \cdot\ \mathrm{CLEP}_t(d(t))}{\|\mathrm{CLEP}_g(h,\ r)\|\ \|\mathrm{CLEP}_t(d(t))\|}$$

Link Prediction across spaces

- CLEP is trained to align head entities with tails descriptions $e^{head} + r \sim d^{tail}$

$$f_s(\text{Barack Obama},\ \text{nationality},\ d(v)) \quad \forall v \in \mathcal{G}$$

$\text{description} \in \text{text space}$

$\text{node} \in \text{graph space}$

Cosine Similarity score

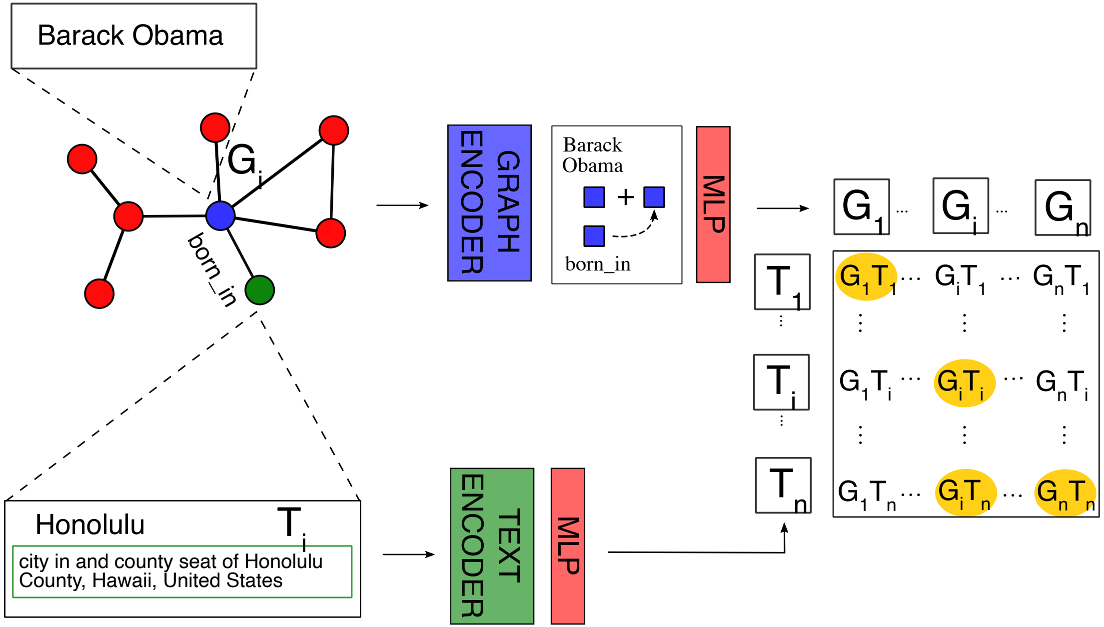
$$f_s(h, r, t) = \frac{\mathrm{CLEP}_g(h,\ r)\ \cdot\ \mathrm{CLEP}_t(d(t))}{\|\mathrm{CLEP}_g(h,\ r)\|\ \|\mathrm{CLEP}_t(d(t))\|}$$

	MR	MRR	hits@1	hits@10
CLEP	**198**	0.222	0.137	0.396
RGCN + Distmult	315	**0.237**	**0.156**	**0.407**

FB15k-237

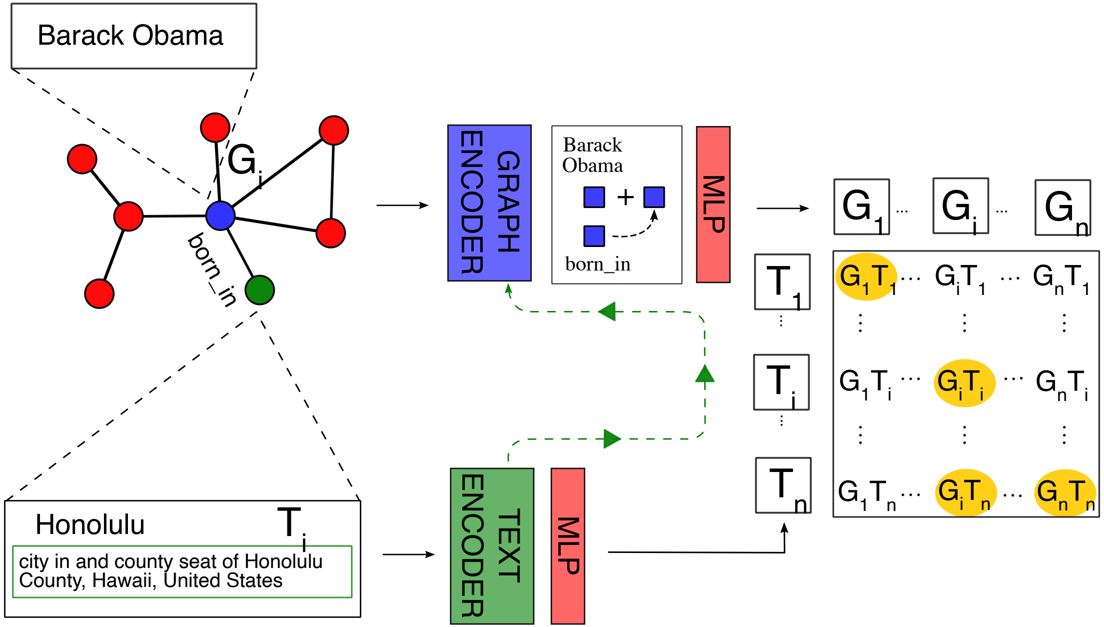
Link Prediction Finetuning

- Pretrain with CLEP



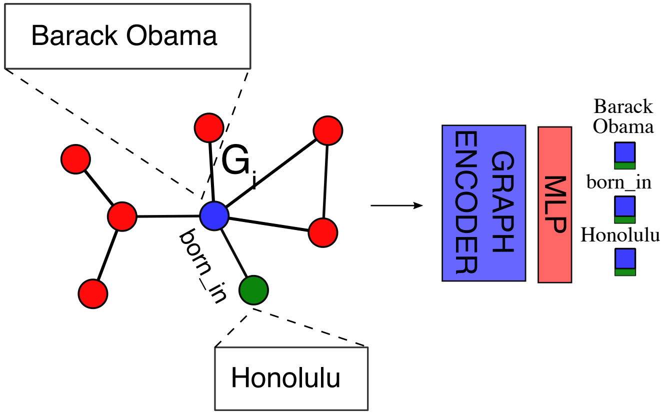
Link Prediction Finetuning

- Pretrain with CLEP



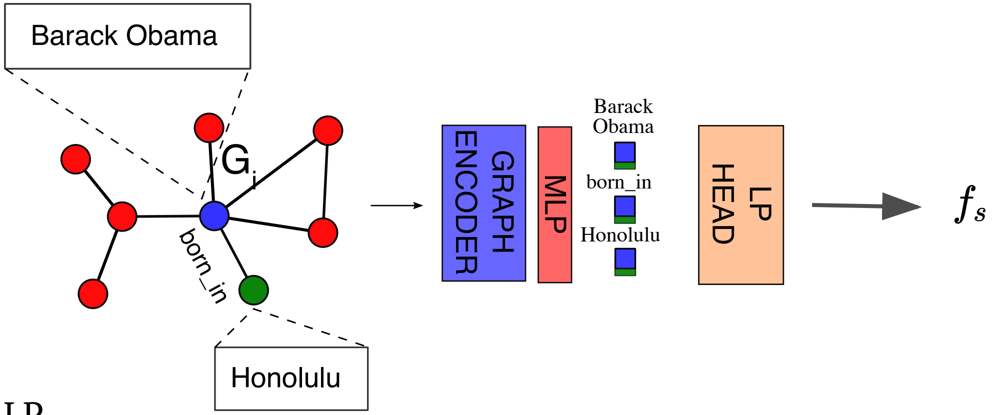
Link Prediction Finetuning

- Pretrain with CLEP



Link Prediction Finetuning

- Pretrain with CLEP



- Finetune on pure LP

RESCAL / DistMult

$$f_s(h, r, t) = h^T M_r t$$

TransE

$$f_s(h, r, t) = \|h + r - t\|$$

Link Prediction Finetuning

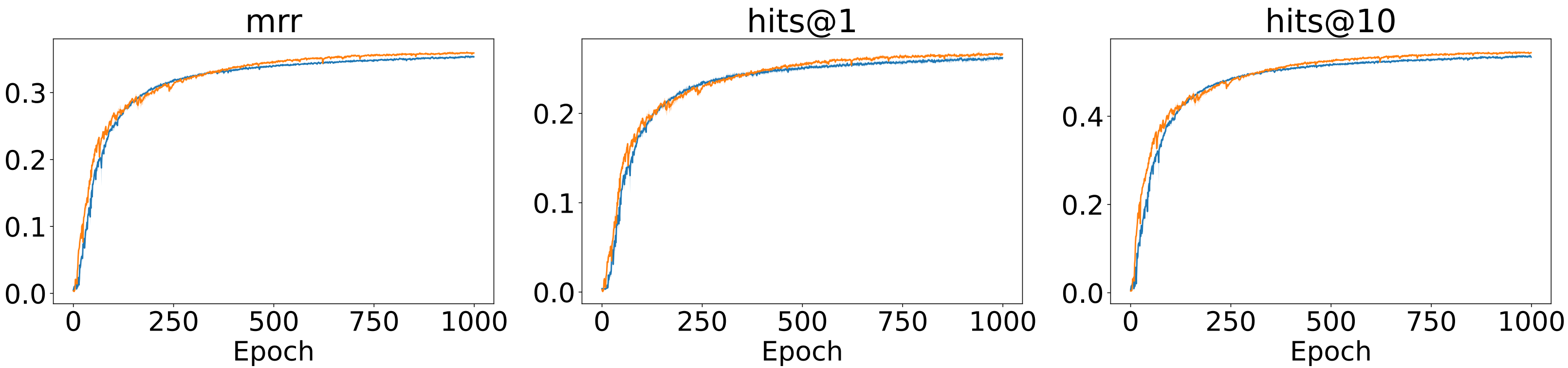
- Randomly initialized CompGCN
- CLEP pretrained CompGCN

Link Prediction Finetuning

Randomly initialized CompGCN

CLEP pretrained CompGCN

FB15k-237



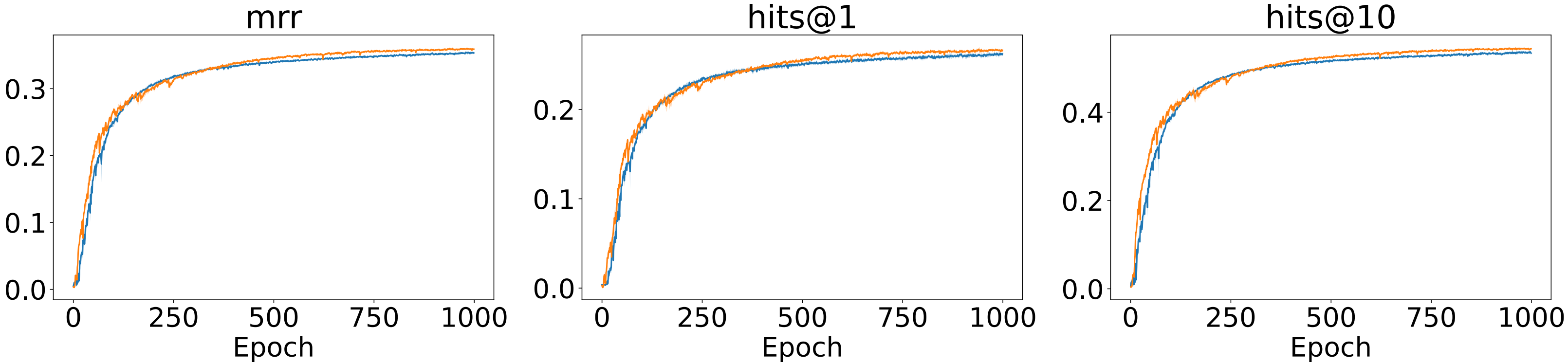
$\sim +1 \text{ - } 2\%$

Link Prediction Finetuning

Randomly initialized CompGCN

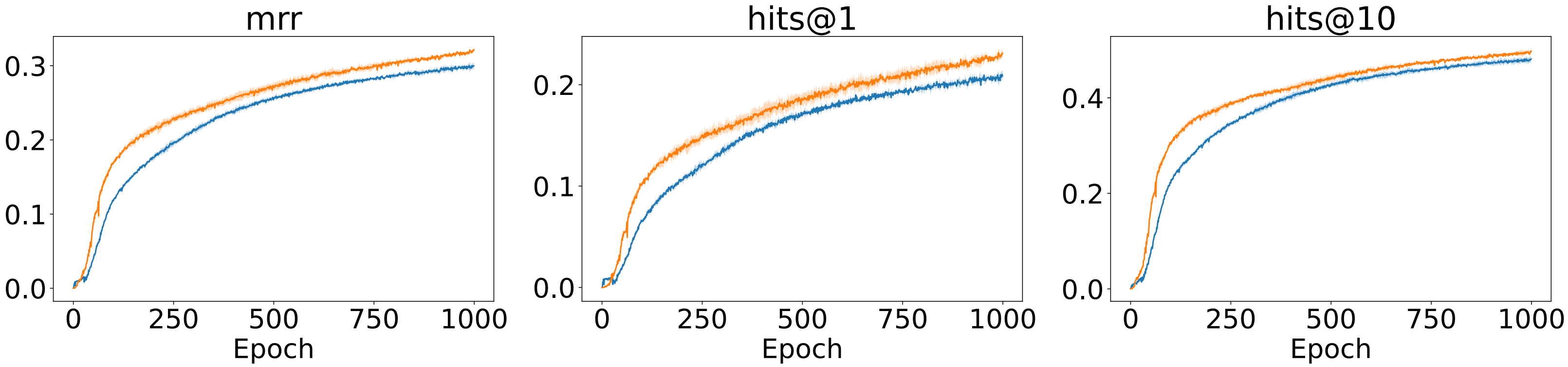
CLEP pretrained CompGCN

FB15k-237



$\sim +1\text{-}2\%$

YAGO3-10



$\sim +4\text{-}10\%$

Conclusion and Future Work

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space
- Properties of the original spaces are preserved: *e.g.* composition of entities and relations

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space
- Properties of the original spaces are preserved: *e.g.* composition of entities and relations
- Some of the textual information is transferred to the graph encoder during the pre-training

  → improved performance on downstream tasks without additional textual inputs

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space
- Properties of the original spaces are preserved: *e.g.* composition of entities and relations
- Some of the textual information is transferred to the graph encoder during the pre-training

  → improved performance on downstream tasks without additional textual inputs
- Is the text encoder expected to manifest a similar transfer?

  → finetuning for instance on Question Answering?

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space
- Properties of the original spaces are preserved: *e.g.* composition of entities and relations
- Some of the textual information is transferred to the graph encoder during the pre-training

  → improved performance on downstream tasks without additional textual inputs
- Is the text encoder expected to manifest a similar transfer?

  → finetuning for instance on Question Answering?
- Any zero-shot capability enabled?

  → zero-shot Entity Linking: comparison of entity mentions and node embeddings

Conclusion and Future Work

- CLEP allows for learning an aligned multi-modal Text-Graph space
- Nodes and corresponding textual descriptions are embedded close in this space
- Properties of the original spaces are preserved: *e.g.* composition of entities and relations
- Some of the textual information is transferred to the graph encoder during the pre-training

  → improved performance on downstream tasks without additional textual inputs
- Is the text encoder expected to manifest a similar transfer?

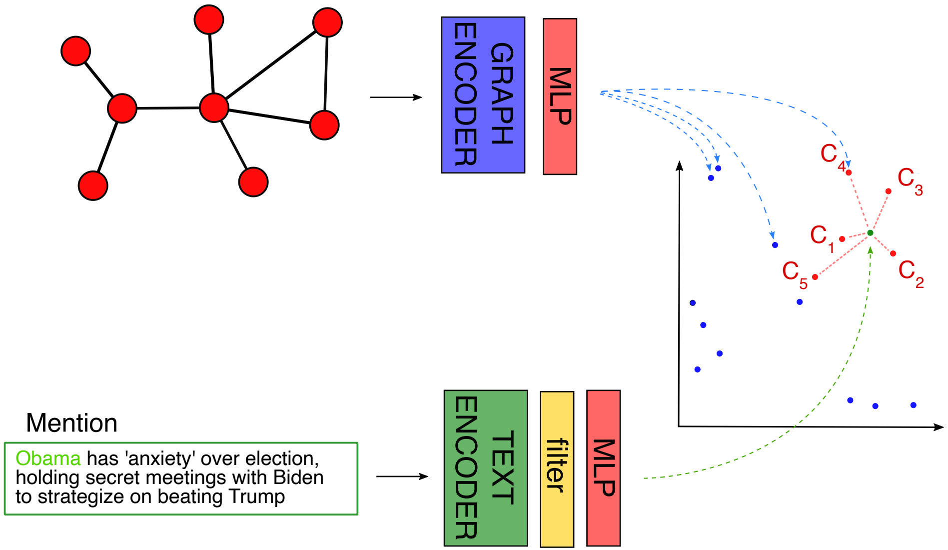
  → finetuning for instance on Question Answering?
- Any zero-shot capability enabled?

  → zero-shot Entity Linking: comparison of entity mentions and node embeddings
- Stable diffusion based Graph Generative Model for Information Extraction

Thank you for the Attention!

Zero-shot Entity Linking

- Candidates generation $(C_1, C_2, \ldots, C_n)$ through calculating the distance from the mention m



Stable Diffusion for Graph Generation