COLING 2022

Knowledge Graph Embeddings for NLP: From Theory to Practice

Luca Costabello
Accenture Labs
@lukostaz

Sumit Pai
Accenture Labs
@sumitppai

Eda Bayram
Accenture Labs
@bayrameda_ee

Adrianna Janik
Accenture Labs
@adri_janik

COLING 2022

kge4nlp-coling22.github.io

accenture

Knowledge Graph Embeddings for NLP: From Theory to Practice

Luca Costabello
Accenture Labs
@lukostaz

Sumit Pai
Accenture Labs
@sumitppai

Eda Bayram
Accenture Labs
@bayrameda_ee

Adrianna Janik
Accenture Labs
@adri_janik

Outline

Foundations **40m**

- Introduction
- Anatomy of a Knowledge Graph Embedding Model
- Evaluation Protocol and Metrics

Advanced KGE Topics **20m**

- Advanced KGE Topics
- Open Research Questions

KGs for NLP **30m**

Break **30m**

Live Q&A **10m**

Applications & Software Ecosystem **20m**

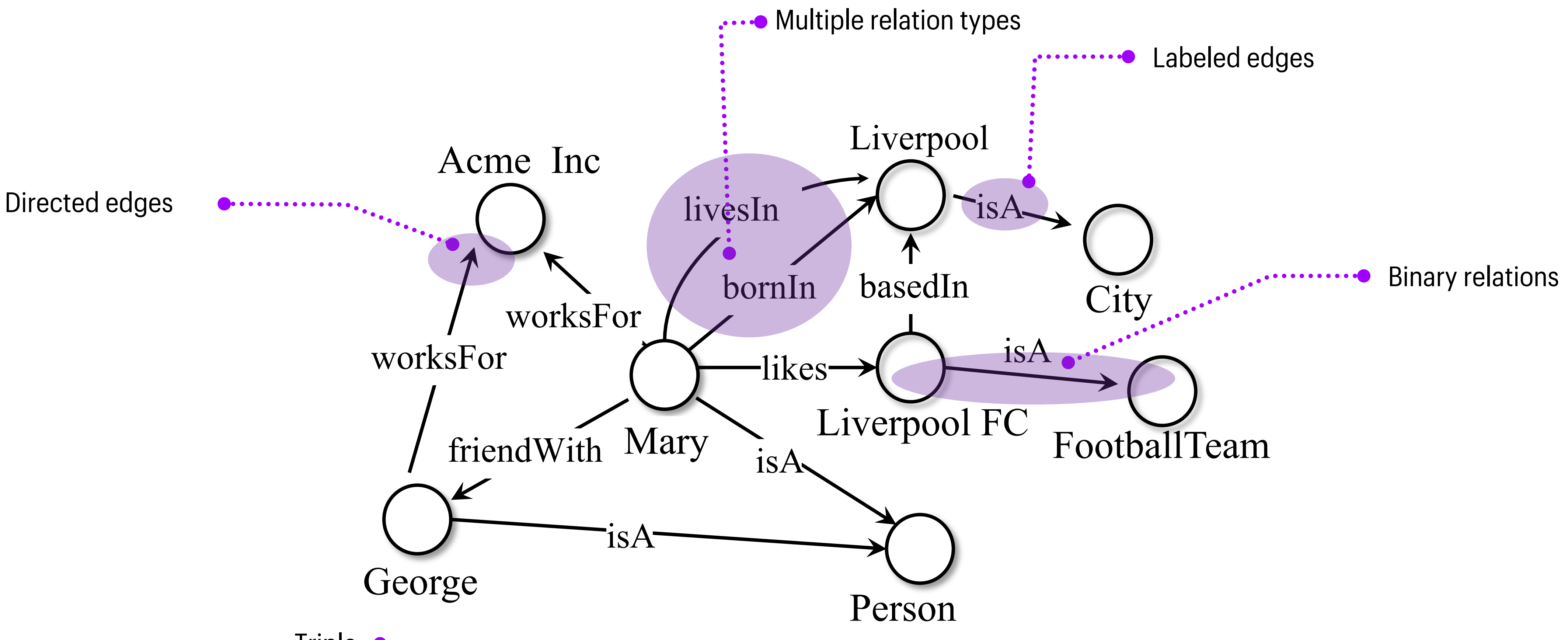
Hands-on Session **45m**

Live Q&A **10m**

Outline

Foundations	**40m**
• Introduction • Anatomy of a Knowledge Graph Embedding Model • Evaluation Protocol and Metrics	
Advanced KGE Topics	**20m**
KGs for NLP	**30m**
Break	**30m**
Live Q&A	**10m**
Applications & Software Ecosystem	**20m**
Hands-on Session	**45m**
Live Q&A	**10m**

Knowledge Graph



$$\mathcal{G} = \{(s, p, o)\} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$$

$\mathcal{E}$: set of entities of $\mathcal{G}$

$\mathcal{R}$: set of relations of $\mathcal{G}$

In-depth overview of Knowledge Graphs in [Hogan et al. 2020]

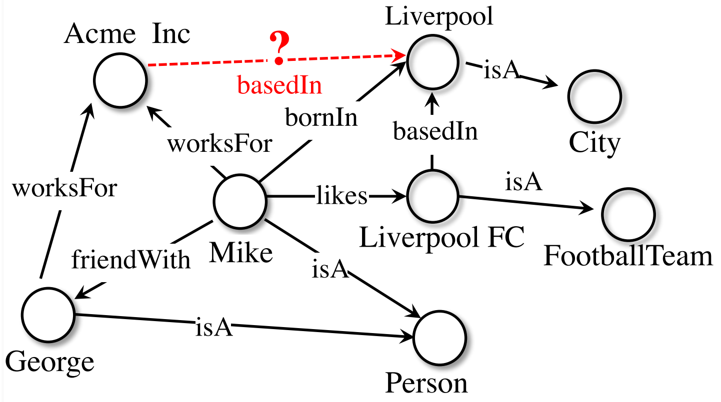
	Statements	Entities
yago select knowledge	343 M	67 M
WIKIDATA	14 B	99 M
DBpedia	3 B	38 M

https://yago-knowledge.org/downloads/yago-4
https://www.wikidata.org/wiki/Wikidata:Statistics
http://wikidata.dbpedia.org/about

Machine Learning on Knowledge Graphs: Tasks

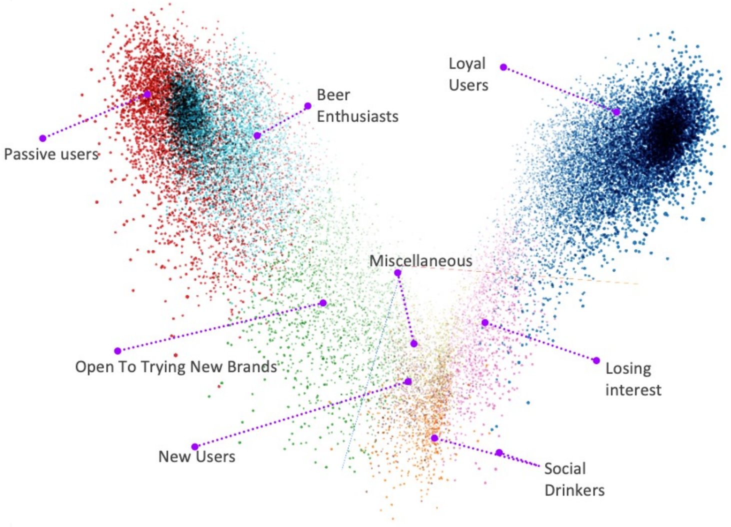
LINK PREDICTION / TRIPLE CLASSIFICATION

- Knowledge graph completion
- Content recommendation
- Knowledge discovery



COLLECTIVE NODE CLASSIFICATION / LINK-BASED CLUSTERING

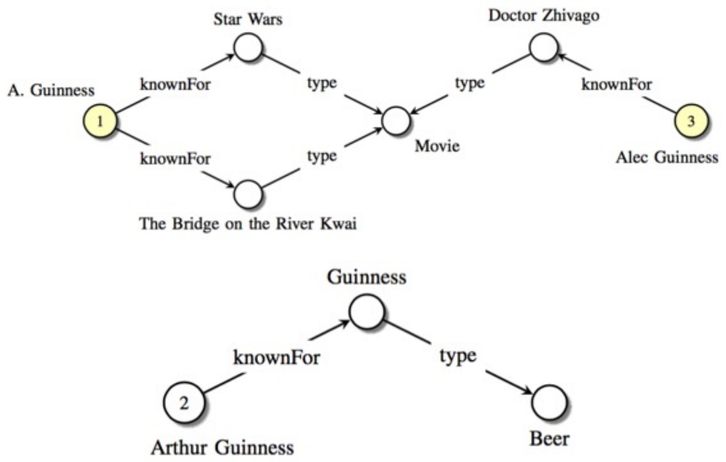
- Customer segmentation



[Pai et al. 2022]

ENTITY MATCHING

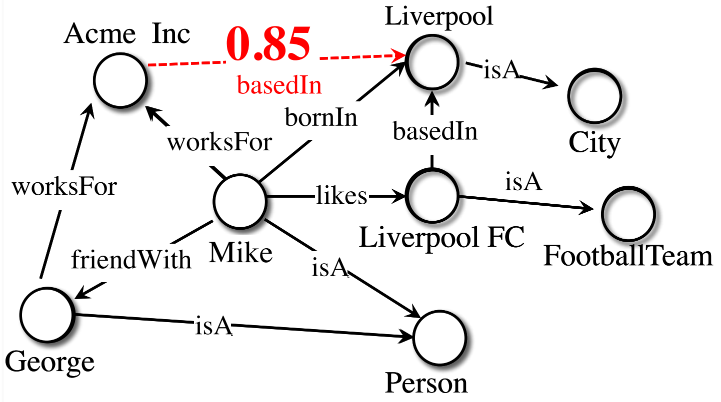
- Duplicate detection
- Inventory items deduplication



[Nickel et al. 2016a]

LINK PREDICTION / TRIPLE CLASSIFICATION

- Knowledge graph completion
- Content recommendation
- Question answering



Assigning a score proportional to the likelihood that an unseen triple is true.

Link Prediction

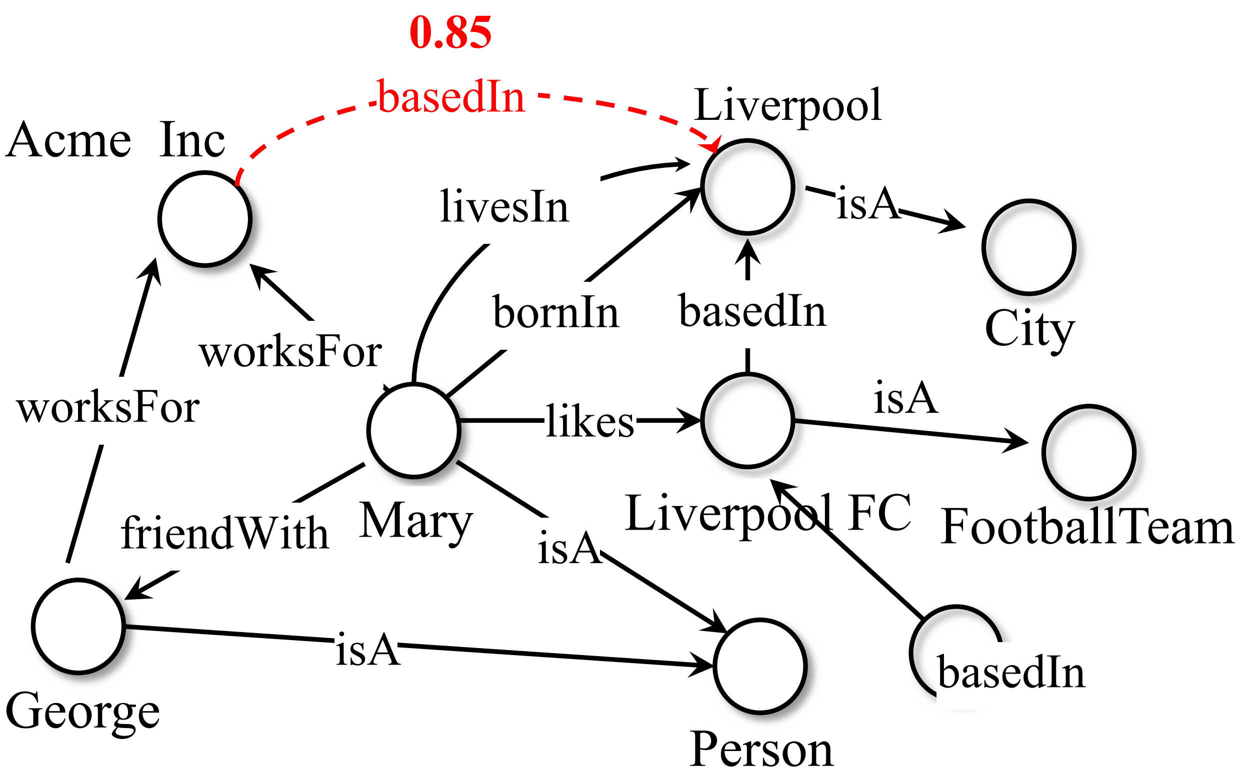
- Learning to rank problem
- Information retrieval metrics
- No ground truth negatives in test set required

Triple Classification

- Binary classification task
- Binary classification metrics
- Test set requires positives and ground truth negatives

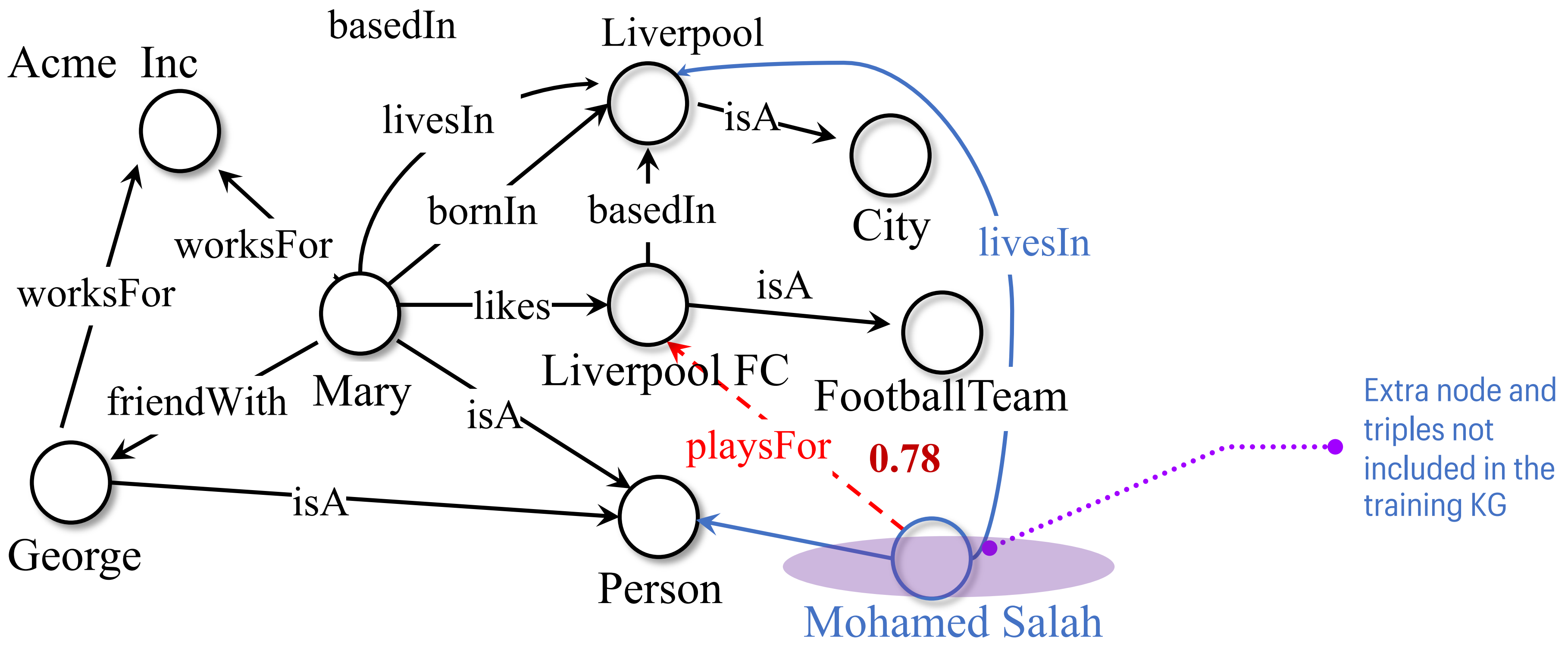
Link Prediction: Transductive vs Inductive

***Transductive* Link Prediction:** when both subject and object of the predicted link occur in the training knowledge graph

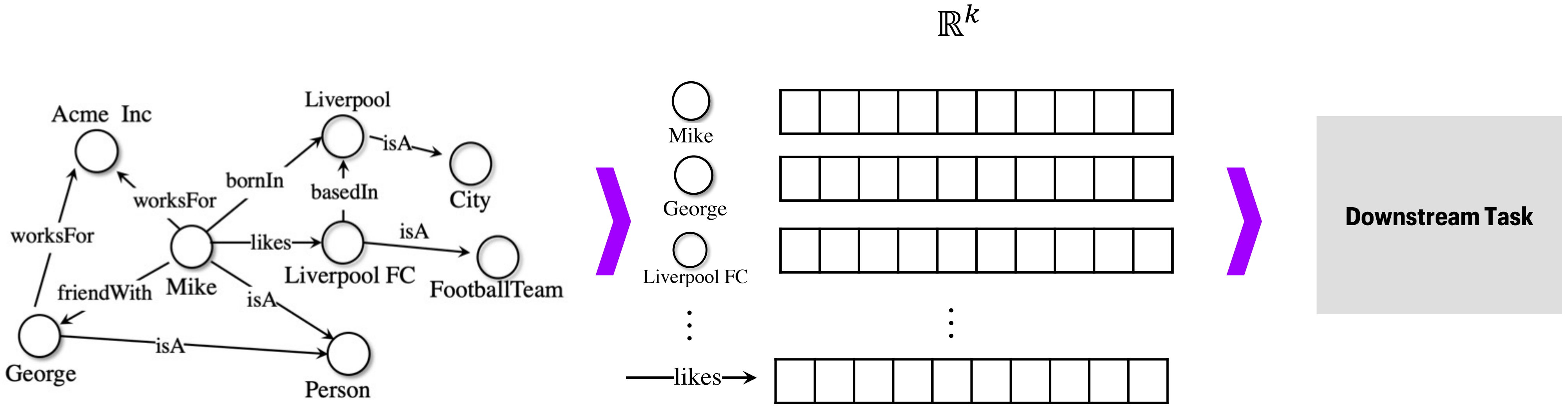


Link Prediction: Transductive vs Inductive

***Inductive* Link Prediction:** when either subject or object of the predicted link do not occur in the training knowledge graph



Graph Representation Learning
Learning representations of nodes and edges



Node Representation/Graph Feature based Methods
PRA, LINE, DeepWalk, node2vec

Graph Neural Networks (GNNs)
GCNs, Graph Attention Networks

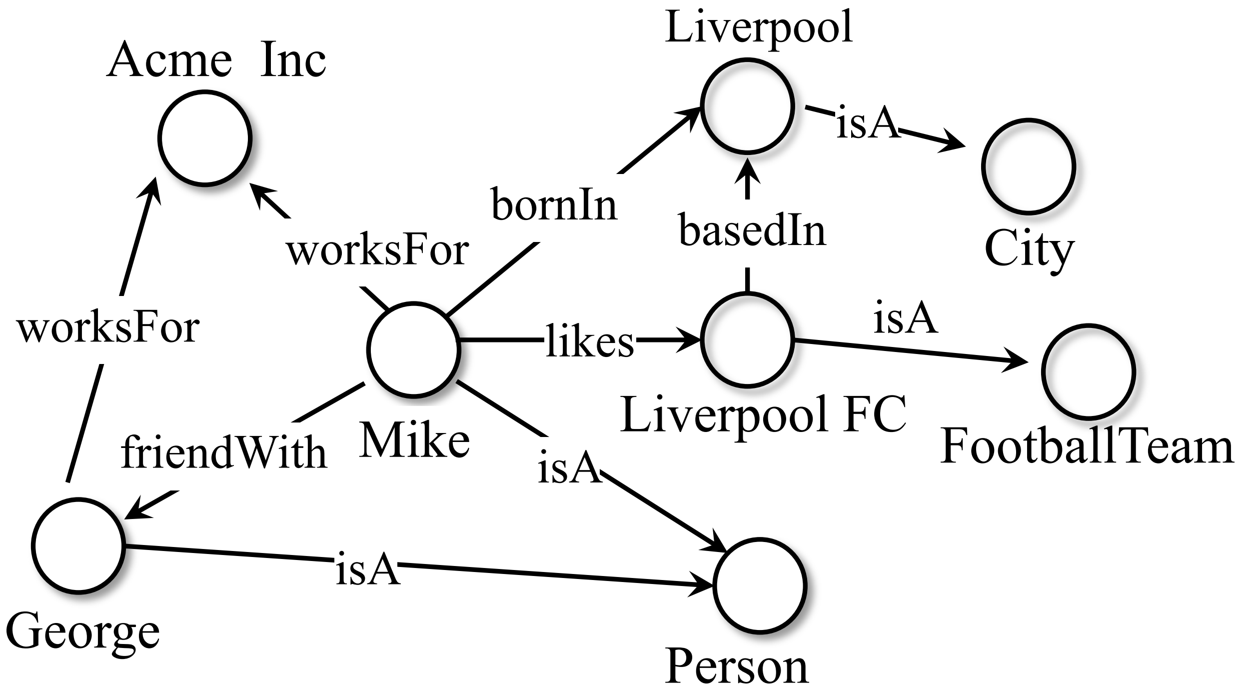
Knowledge Graph Embeddings (KGE)
TransE, DistMult, ComplEx, ConvE

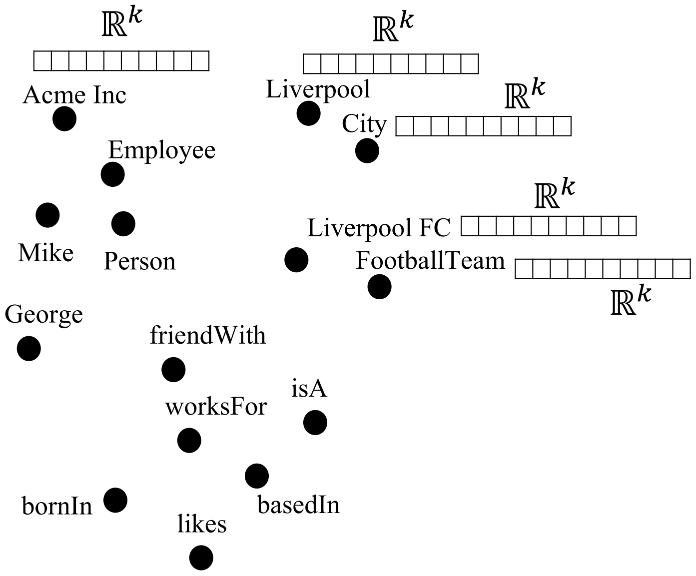
Scope of this tutorial

For a complete overview of graph feature-based models and GNNs:
[Hamilton & Sun 2019]
[Hamilton 2020]

Knowledge Graph Embeddings (KGE)

Automatic, supervised learning of **embeddings**, i.e. projections of entities and relations into a continuous low-dimensional space $\mathbb{R}^k$.





KGE Design Rationale: Capture KG Patterns

Symmetry `<Alice marriedTo Bob>`

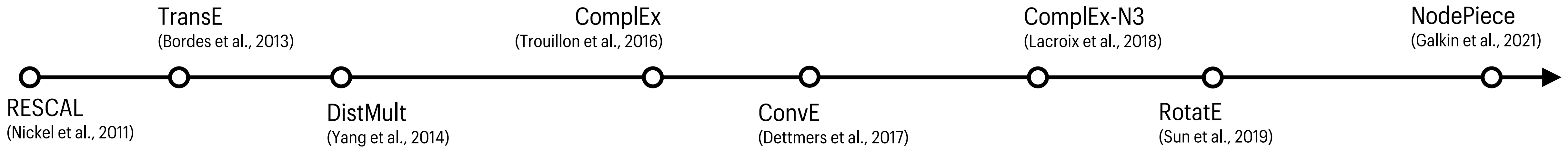
Asymmetry `<Alice childOf Jack>`

Inversion
`<Alice childOf Jack>`
`<Jack fatherOf Alice>`

Composition
`<Alice childOf Jack>`
`<Jack siblingOf Mary>`
`<Alice nieceOf Mary>`

But also:
- Hierarchies
- Type constraints
- Transitivity
- Homophily
- Long-range dependencies

Popular KGE models in recent published literature



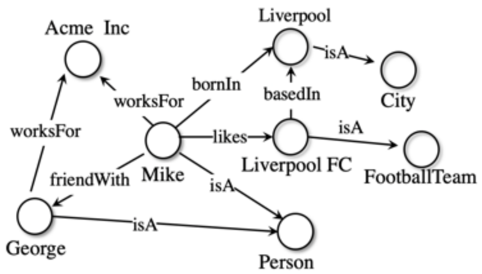
Model	Symmetry	Antisymmetry	Inversion	Composition
SE	✗	✗	✗	✗
TransE	✗	✓	✓	✓
TransX	✓	✓	✗	✗
DistMult	✓	✗	✗	✗
ComplEx	✓	✓	✓	✗
RotatE	✓	✓	✓	✓

[Sun et al. 2019]

Outline

Foundations	**40m**
• Introduction	
• **Anatomy of a Knowledge Graph Embedding Model**	
• Evaluation Protocol and Metrics	
Advanced KGE Topics	**20m**
KGE & NLP Tasks	**30m**
Break	**30m**
Live Q&A	**10m**
Applications & Software Ecosystem	**20m**
Hands-on Session	**45m**
Live Q&A	**10m**

At a Glance



Optimizer

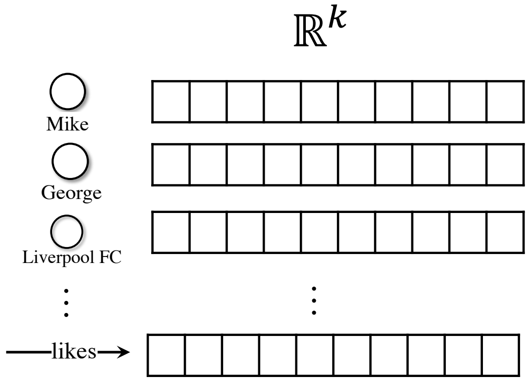
Lookup Layer

Scoring Layer
$f(s,p,o) \in \mathbb{R}$

Loss Functions
$\mathcal{L}$

Negatives Generation

Training



Downstream Tasks
(Link Prediction)

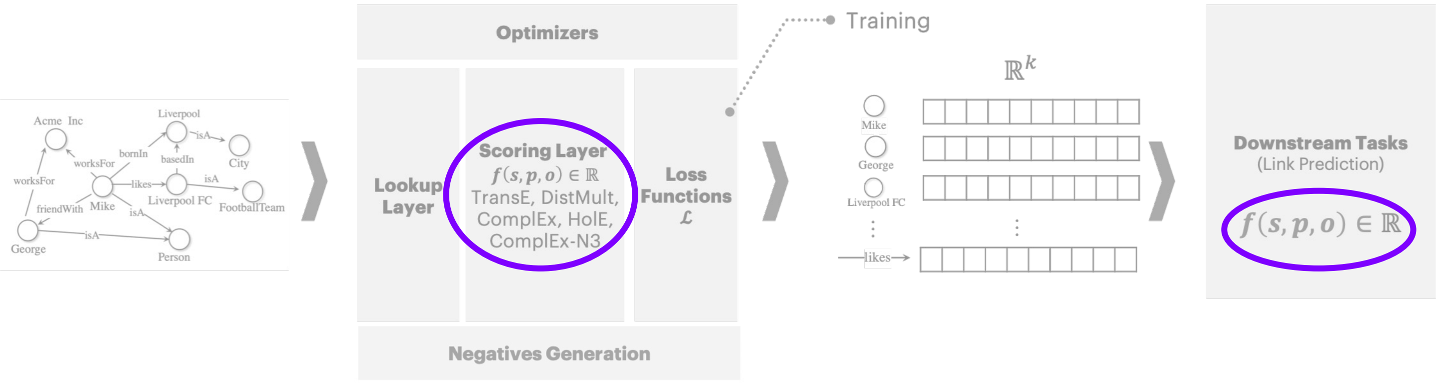
$f(s,p,o) \in \mathbb{R}$

Inference

Anatomy of a Knowledge Graph Embedding Model

- Knowledge Graph (KG) $\mathcal{G}$
- Scoring function for a triple $f(t)$
- Loss function $\mathcal{L}$ (Translation-based, Factorization-based, Deep)
- Optimization algorithm
- Negatives generation strategy

Scoring function f



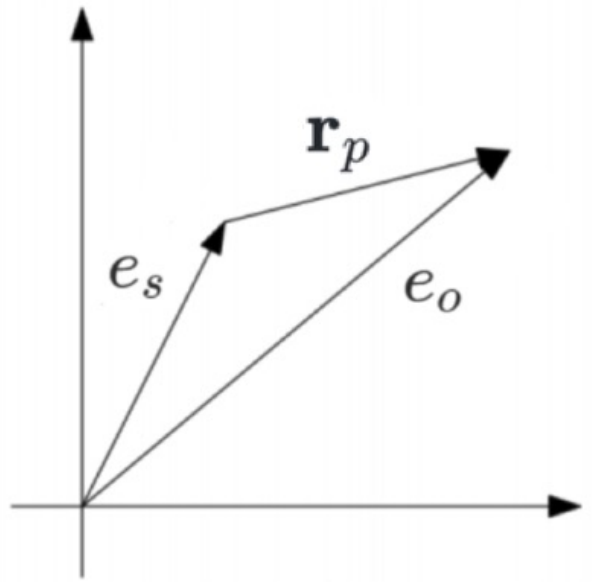
f assigns a score to a triple (s, p, o)

High score = triples is very likely to be factually correct

Translation-based Scoring Functions

- **TransE**: Translating Embeddings [Bordes et al. 2013]

$$f_{TransE} = -||(\mathbf{e}_s + \mathbf{r}_p) - \mathbf{e}_o||_n$$

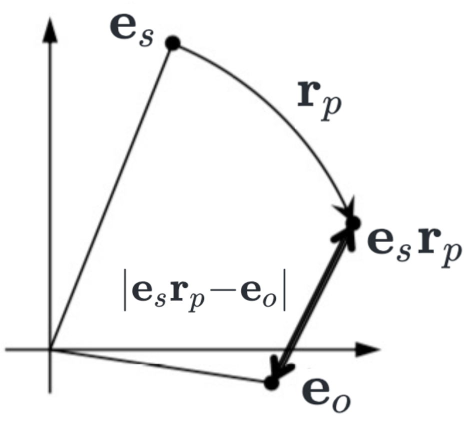


Translation-based Scoring Functions

- **RotatE**: relations modelled as *rotations* in complex space $\mathbb{C}$: element-wise product between complex embeddings.

[Sun et al. 2019]

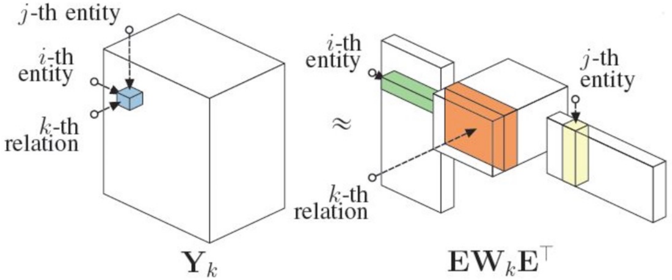
$$f_{RotatE} = -||\mathbf{e}_s \circ \mathbf{r}_p - \mathbf{e}_o||_n$$



Factorization-based Scoring Functions

- **RESCAL**: low-rank factorization with tensor product

$$f_{RESCAL} = \mathbf{e}_s^T \mathbf{W}_r \mathbf{e}_o$$



[Nickel et al. 2011]

- **DistMult**: bilinear diagonal model. Dot product.

[Yang et al. 2015]

$$f_{DistMult} = \langle \mathbf{r}_p, \mathbf{e}_s, \mathbf{e}_o \rangle$$

- **ComplEx**: Complex Embeddings (Hermitian dot product): (i.e. extends DistMult with dot product in ℂ)

$$f_{ComplEx} = Re(\langle \mathbf{r}_p, \mathbf{e}_s, \overline{\mathbf{e}_o} \rangle)$$

[Trouillon et al. 2016]

"Deeper" Scoring Functions

- **ConvE**: reshaping + convolution

[Dettmers et al. 2017]

Non-linearity

$$f_{ConvE} = \langle \sigma \left(vec \left(g \left([\overline{\mathbf{e}_s}; \overline{\mathbf{r}_p}] * \Omega \right) \right) \mathbf{W} \right) \mathbf{e}_o \rangle$$

2D reshaping

Linear convolution

- **ConvKB**: convolutions and dot product

[Nguyen et al. 2018]

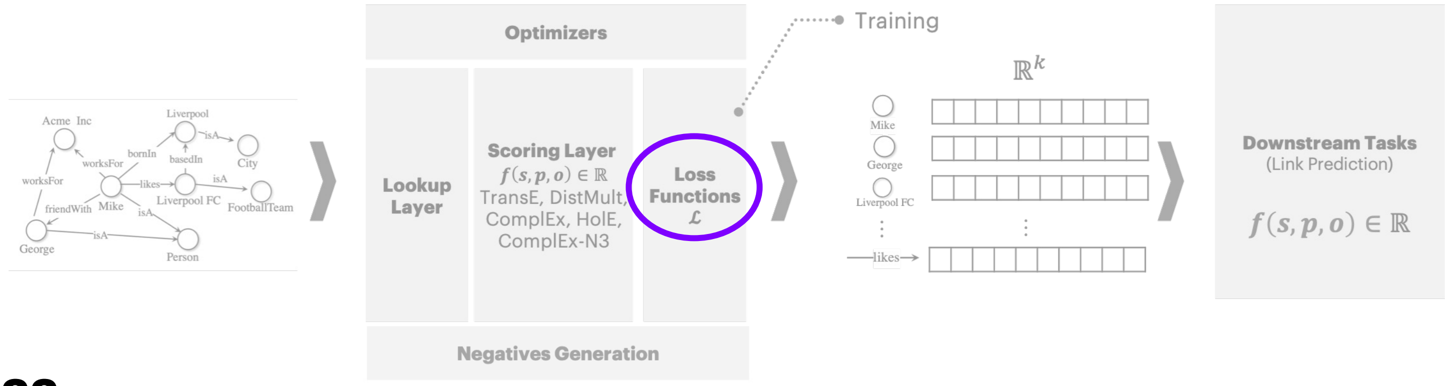
$$f_{ConvKB} = concat \left(g \left([\mathbf{e}_s, \mathbf{r}_p, \mathbf{e}_o] \right) * \Omega \right) \cdot W$$

Computationally expensive!

Other Recent Models

- HolE [Nickel et al. 2016]
- SimplE [Kazemi et al. 2018]
- QuatE [Zhang et al. 2019]
- MurP [Balaževic´ et al. 2019]
- NodePiece [Galkin´ et al. 2021]
- …

Loss function $\mathcal{L}$



Pairwise Margin-Based Hinge Loss

Pays a penalty if score of positive triple < score of synthetic negative by a margin γ

$$\mathcal{L}(\Theta) = \sum_{t^+ \in \mathcal{G}} \sum_{t^- \in \mathcal{C}} max(0, [\gamma + \underbrace{f(t^-;\Theta)}_{\text{Score assigned to a synthetic negative}} - \underbrace{f(t^+;\Theta)}_{\text{Score assigned to true triple}}])$$

Score assigned to a **synthetic negative**

Score assigned to **true** triple

[Bordes et al. 2013]

Negative Log-Likelihood / Cross Entropy

$$\mathcal{L}(\Theta) = \sum_{t \in \mathcal{G} \cup \mathcal{C}} log(1 + exp(-\underbrace{y}_{\text{Label of the triple t}}\, f(t;\Theta)))$$

Label of the triple t $\quad y \in \{-1, 1\}$

[Trouillon et al. 2016]

Loss function $\mathcal{L}$

Binary Cross-Entropy

$$\mathcal{L} = -\frac{1}{N} \sum_{t \in \mathcal{G} \cup \mathcal{C}}^{N} y \cdot log(\sigma(f(t;\Theta))) + (1-y) \cdot log(1 - f(t;\Theta))$$

[Dettmers et al. 2017]

Self-Adversarial

$$\mathcal{L} = -log\, \sigma(\gamma + f(t^{+};\Theta)) - \sum_{t \in \mathcal{G}}^{N} \underbrace{p(t^{-};\Theta)}_{\text{Weight for the negative sample } t^{-}} log\, \sigma(-f(t^{-};\Theta) - \gamma)$$

[Sun et al. 2019]

Many more: Multiclass Negative Log-likelihood, Absolute Margin, etc.

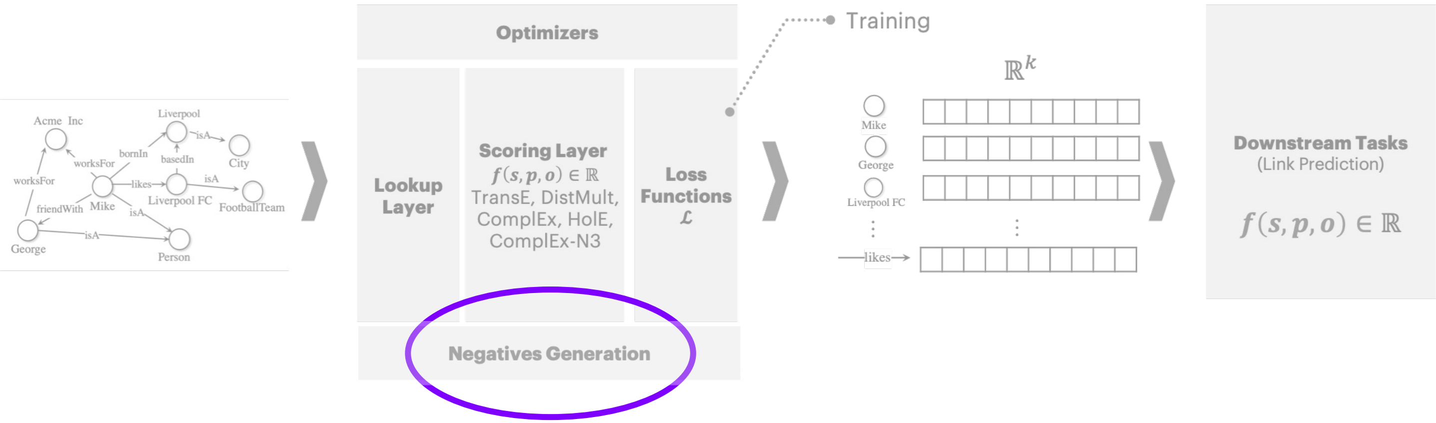
Regularizers

- L1, L2
- L3 [Lacroix et al. 2018]
- Dropout (ConvE) [Dettmers et al. 2017]

Initialization

- Random (Uniform)
- Random (Normal)
- Glorot

Negatives Generation



Where do negative examples come from? (i.e. false facts)

"Local Closed-World" Assumption: the KG is only *locally* complete

"Corrupted" versions of a triple as synthetic negatives:

$$\mathcal{C} = \{(\hat{s}, p, o) | \hat{s} \in \mathcal{E}\} \cup \{(s, p, \hat{o}) | \hat{o} \in \mathcal{E}\}$$

"corrupted subject"

"corrupted" object

The predicate is unaltered

Synthetic Negatives: Example

$$\mathcal{E} = \{Mike, Liverpool, AcmeInc, George, LiverpoolFC\}$$
$$\mathcal{R} = \{bornIn, friendWith\}$$

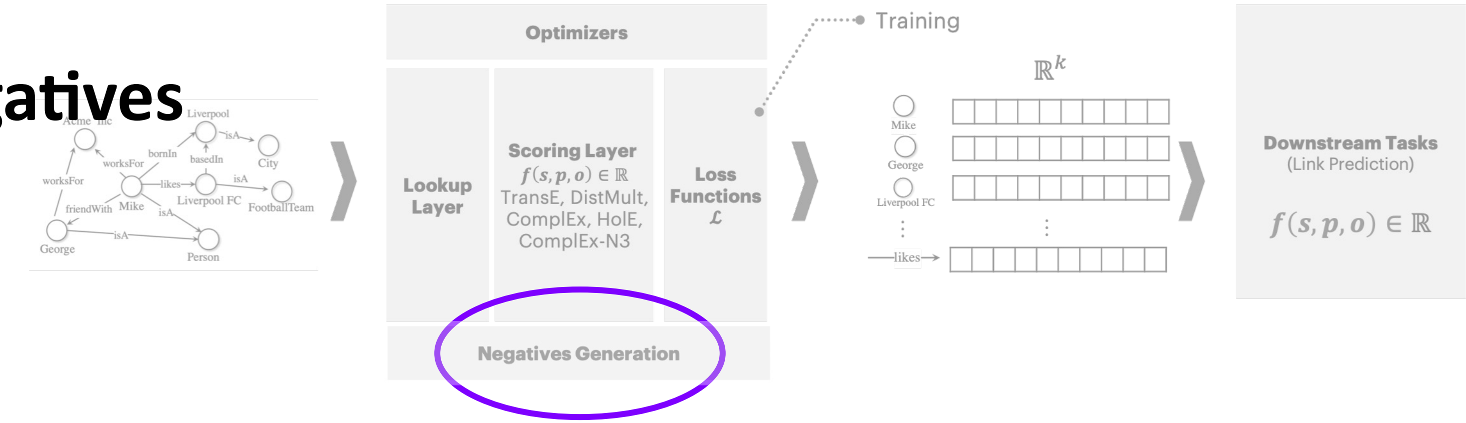
$t \in \mathcal{G} =$ `(Mike bornIn Liverpool)`

$\mathcal{C}_t =$

```
Mike      bornIn   AcmeInc
Mike      bornIn   LiverpoolFC
George    bornIn   Liverpool
AcmeInc   bornIn   Liverpool
```

Set of corruptions for `t` (in this example we generate four corruptions)

Training with Synthetic Negatives

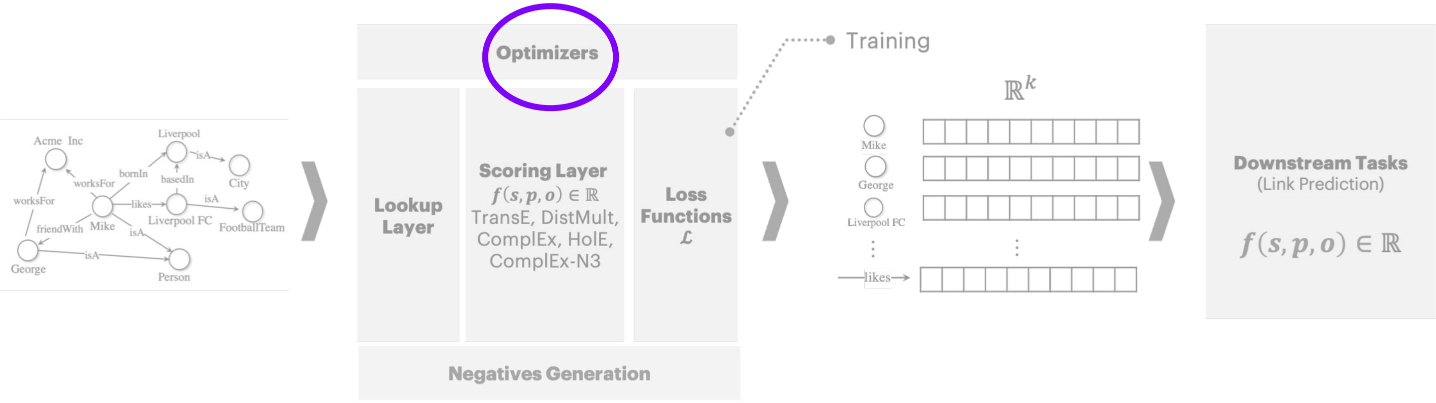


Uniform sampling: generate all possible synthetic negatives and sample n negatives for each positive t.

Complete set: no sampling. Use all possible synthetic negatives for each positive t. (mind scalability)

1-n scoring: batches of (s, p, *) or (*, p, o) labeled as positives (if included in training KG) or negatives (if not in training KG).
[Dettmers et al. 2017]

Training Procedure and Optimizer



Optimizer: learn optimal parameters (e.g. embeddings). Off-the-shelf SGD variants: (AdaGrad, Adam)

$$\min_{\Theta} \mathcal{L}(\Theta)$$

Reciprocal Triples

Injection of reciprocal triples in training set.

`<Alice childOf Jack>`
`<Jack childOf`$^{-1}$` Alice>`

[Dettmers et al. 2017]
[Lacroix et al. 2018]

Model Selection

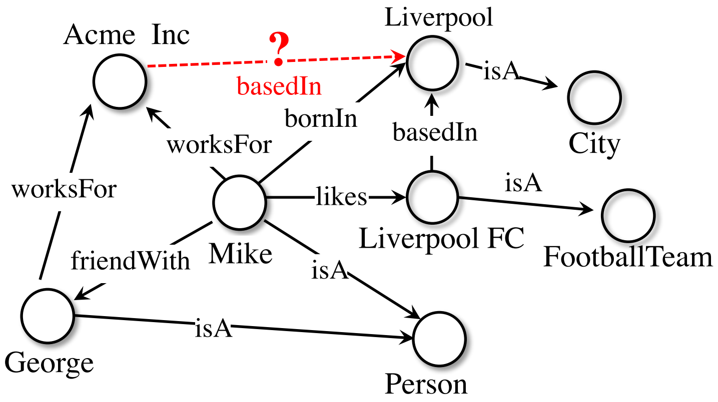
- Grid search
  - Mind the size of the grid!
  - Early stopping
- Random search
- Quasi-random + Bayesian [Ruffinelli et a. 2020]

Outline

Foundations	**40m**
• Introduction	
• Anatomy of a Knowledge Graph Embedding Model	
• **Evaluation Protocol and Metrics**	
Advanced KGE Topics	**20m**
• Advanced KGE Topics	
• Open Research Questions	
KGs for NLP	**30m**
Break	**30m**
Live Q&A	**10m**
Applications & Software Ecosystem	**20m**
Hands-on Session	**45m**
Live Q&A	**10m**

The Task

LINK PREDICTION / TRIPLE CLASSIFICATION



Assigning a score proportional to the likelihood that an unseen triple is true.

Link Prediction

- Learning to rank problem
- Information retrieval metrics
- No ground truth negatives in test set required

Triple Classification

- Binary classification task
- Binary classification metrics
- Test set requires positives and ground truth negatives

Learning–To-Rank problem:

How well are positive triples ranked against **synthetic negatives** built under the **Local Closed World Assumption.**

Same procedure used in training

Evaluation Metrics

Mean Rank (MR)

$$MR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} rank_{(s,p,o)_i}$$

Mean Reciprocal Rank (MRR)

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_{(s,p,o)_i}}$$

Hits@N

$$Hits@N = \sum_{i=1}^{|Q|} 1 \text{ if } rank_{(s,p,o)_i} \leq N$$

Example: How unseen, test positive triples rank against **synthetic negatives**? (four negatives/positive)

s	p	o	score	rank	
Mike	friend_with	George	0.901	1	*
Mike	friend_with	Jim	0.345	2	
Acme	friend_with	George	0.293	3	
Mike	friend_with	Liverpool	0.201	4	
France	friend_with	George	0.156	5	

s	p	o	score	rank	
Mike	born_in	Leeds	0.789	1	
Mike	born_in	Liverpool	0.753	2	*
Mike	born_in	Germany	0.695	3	
George	born_in	Liverpool	0.456	4	
Mike	born_in	George	0.234	5	

Positive triples from test set

Test set = {
`<Mike friend_with George>`
`<Mike born_in Liverpool>`
}

$$MR = 1.5$$
$$MRR = 0.75$$
$$Hits@1 = 0.5$$
$$Hits@3 = 1.0$$

Benchmark Datasets

Freebase

WordNet

YAGO

	FB15K-237	WN18RR	YAGO3-10
Training	272,115	86,835	1,079,040
Validation	17,535	3,034	5,000
Test	20,466	3,134	5,000
Entities	14,541	40,943	123,182
Relations	237	11	37

Transductive Link Prediction: SOTA Results

	FB15K-237	WN18RR	YAGO3-10
Literature Best	**0.35***	0.48*	0.49*
TransE (AmpliGraph)	0.31	0.22	**0.51**
DistMult (AmpliGraph)	0.31	0.47	0.50
ComplEx (AmpliGraph)	0.32	**0.51**	0.49
HolE (AmpliGraph)	0.31	0.47	0.50
ConvE (AmpliGraph)	0.26	0.45	0.30
ConvE (1-N, AmpliGraph)	0.32	0.48	0.40
ConvKB (AmpliGraph)	0.23	0.39	0.30

https://docs.ampligraph.org/en/latest/experiments.html

		FB15K-237		*WNRR*	
		MRR	Hits@10	MRR	Hits@10
First	RESCAL (Wang et al., 2019)	27.0	42.7	42.0	44.7
	TransE (Nguyen et al., 2018)	29.4	46.5	22.6	50.1
	DistMult (Dettmers et al., 2018)	24.1	41.9	43.0	49.0
	ComplEx (Dettmers et al., 2018)	24.7	42.8	44.0	51.0
	ConvE (Dettmers et al., 2018)	32.5	50.1	43.0	52.0
Ours	RESCAL	**35.7**	**54.1**	46.7	51.7
	TransE	31.3	49.7	22.8	52.0
	DistMult	34.3	53.1	45.2	53.1
	ComplEx	34.8	53.6	**47.5**	**54.7**
	ConvE	33.9	52.1	44.2	50.4
Recent	TuckER (Balazevic et al., 2019)	**35.8**	**54.4**	47.0	52.6
	RotatE (Sun et al., 2019a)	33.8	53.3	**47.6**	**57.1**
	SACN (Shang et al., 2019)	35.0	54.0	47.0	54.4
Large	DistMult (Salehi et al., 2018)	35.7	54.8	45.5	54.4
	ComplEx-N3 (Lacroix et al., 2018)	**37.0**	**56.0**	**49.0**	**58.0**

[Ruffinelli et a. 2020]

Comparing SOTA Results is Tricky

- Different training strategies (e.g. synthetic negatives)
- Reciprocal relations in training set?
- Unfair or suboptimal hyperparameters selection
- Evaluation protocol: how to behave with tie ranks?
- Ablation studies!

Read discussion in [Ruffinelli et al 2020]

Outline

Foundations 40m

Advanced Knowledge Graph Embeddings Topics 20m

- Multimodality
- Uncertainty
- Time
- Explainability
- Robustness
- Memory
- Reasoning
- Open Research Questions

KGs for NLP 30m

Break 30m

Live Q&A 10m

Applications & Software Ecosystem 20m

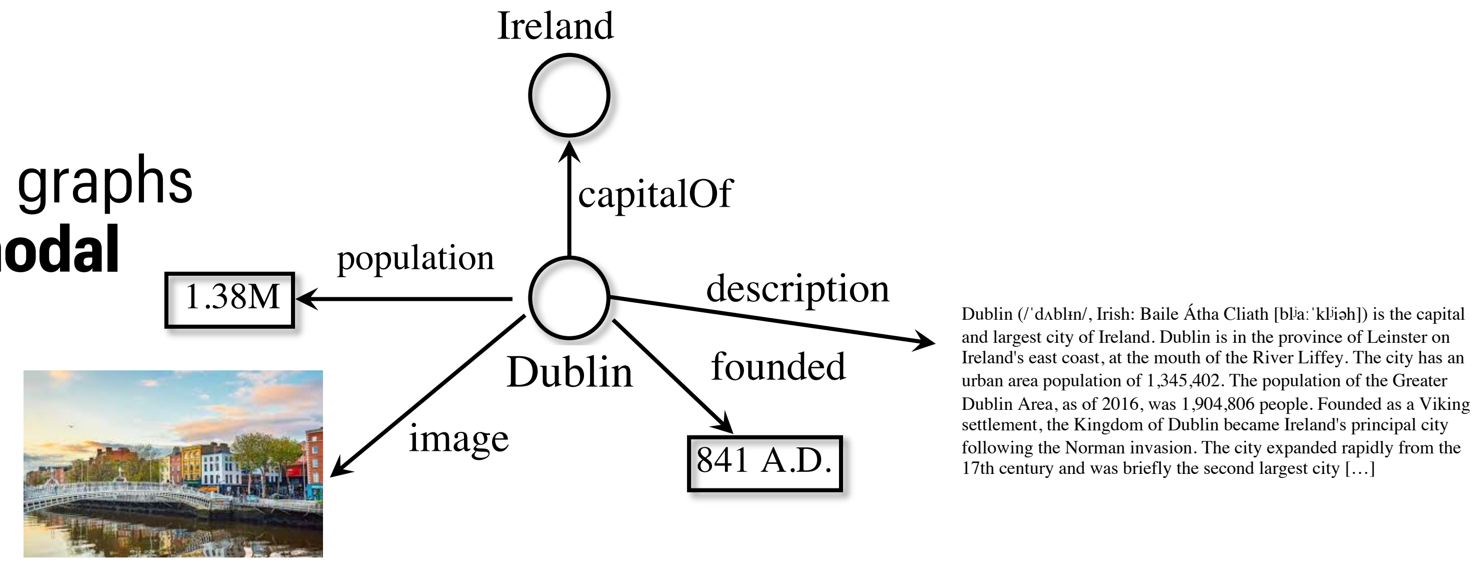
Hands-on Session 45m

Live Q&A 10m

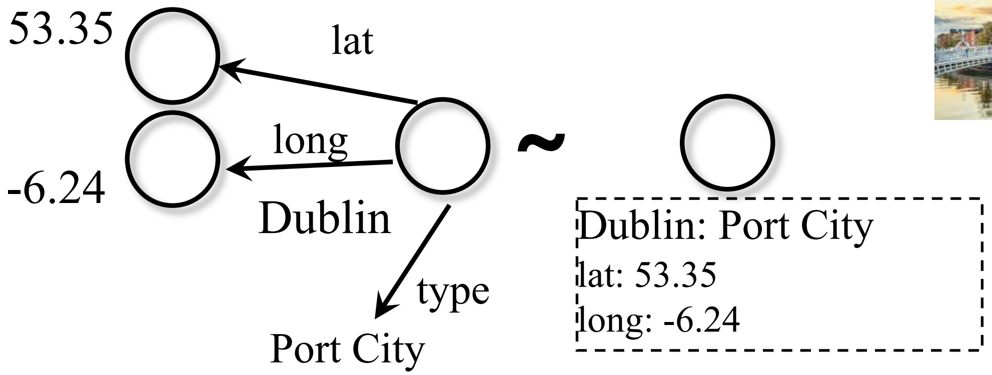
Multimodality

Support for Heterogenous Data

Many real-world graphs include **multi-modal attributes**.



[Gesese et al. 2019] surveys recent literature



Example approach is to use modality-specific encoders to embed objects.

[Pezeshkpour et al. 2018]

Nodes and edges can also have properties...

[Hogan et al. 2022] surveys recent literature

...which also can be missing.

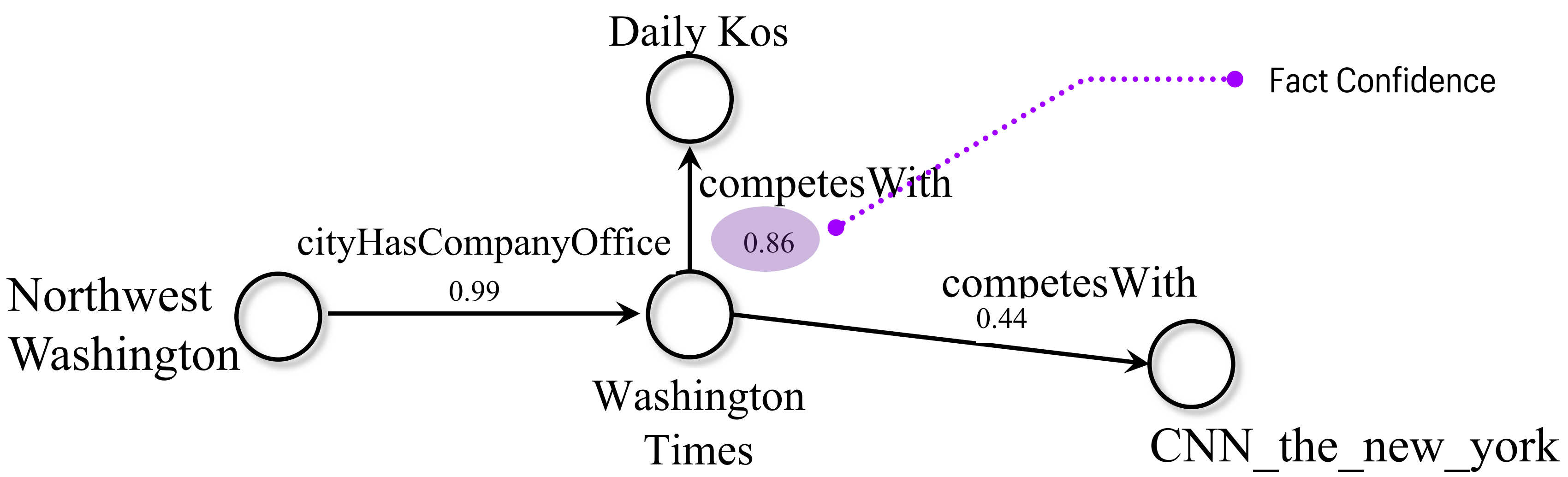
[Bayram et al. 2020]

Alternative approach uses deep learning.

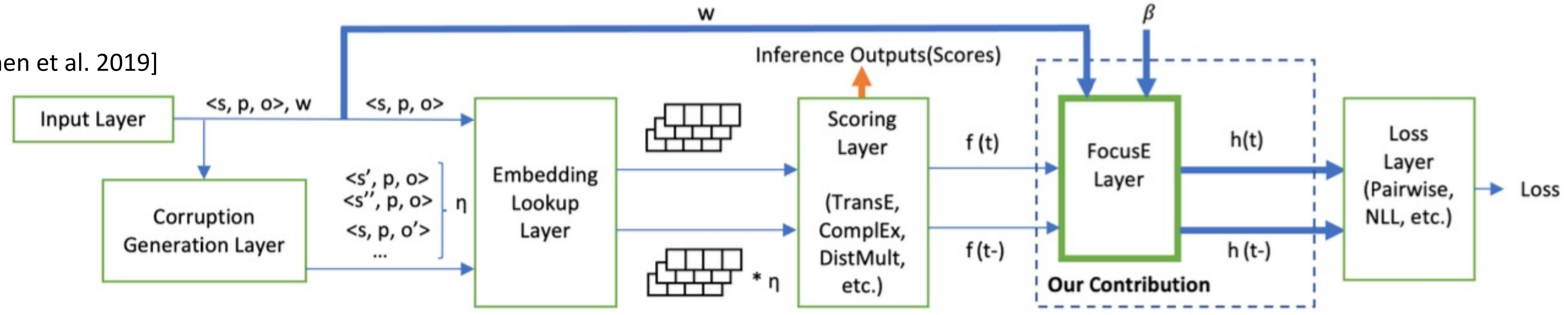
[Wilcke et al. 2021]

Automatic KG generation may lead to *uncertain* facts.

Many real-world graphs include numeric information on edges (e.g. “strength”)



[Chen et al. 2019]

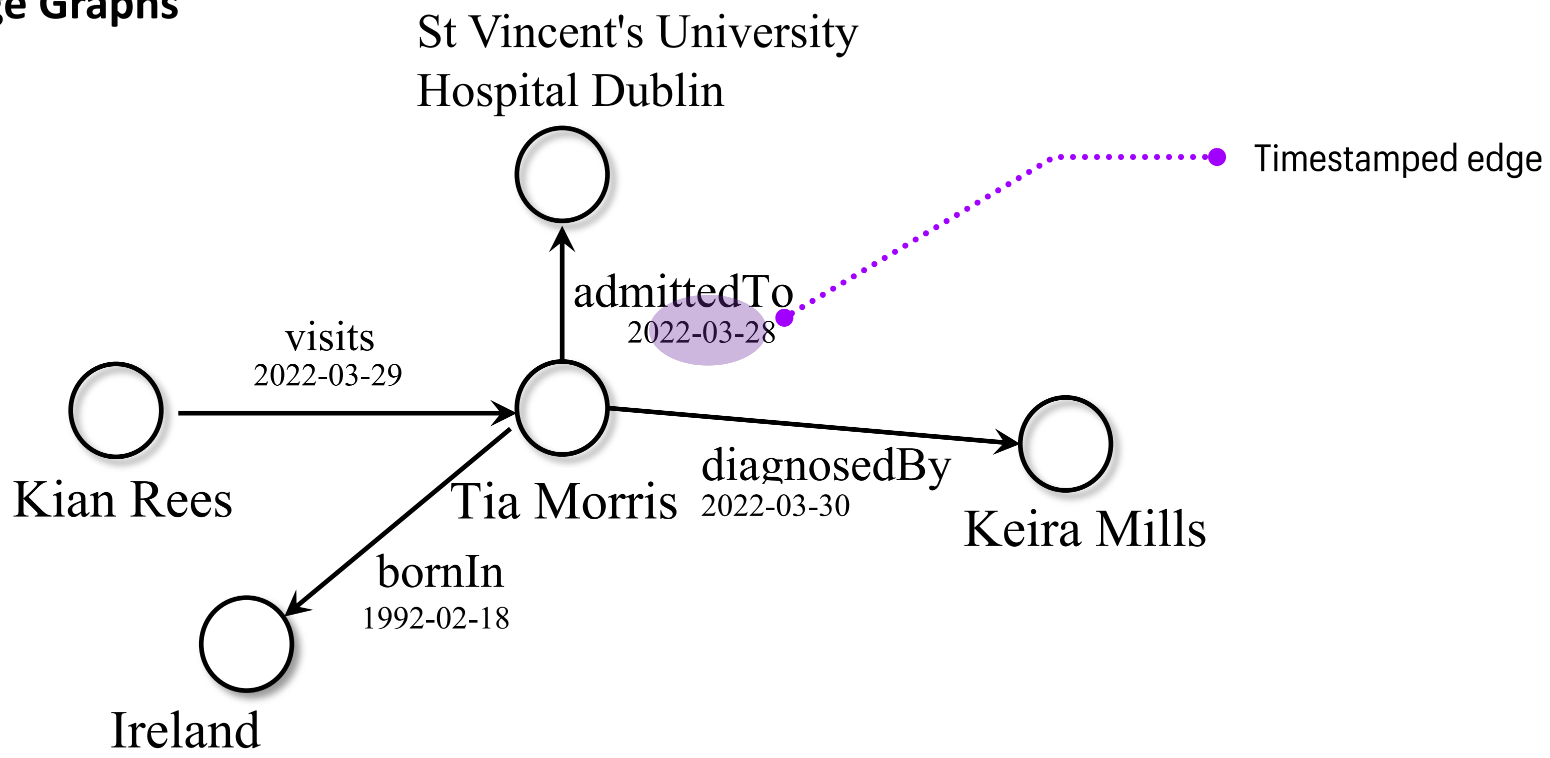


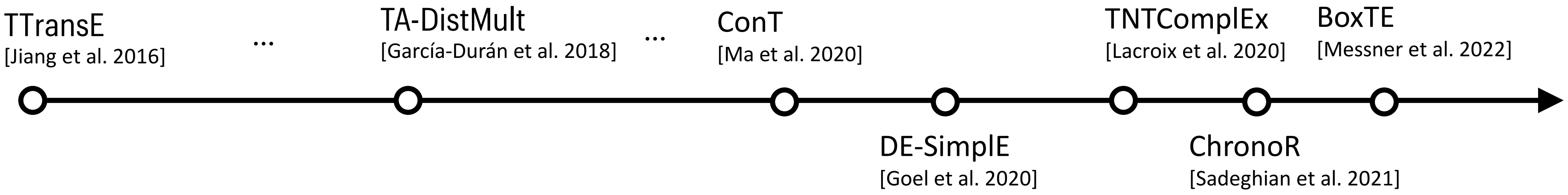
[Pai & Costabello IJCAI-21]

Time

Temporal Aspects of Knowledge Graphs

Many real-world graphs represent timestamped concepts.



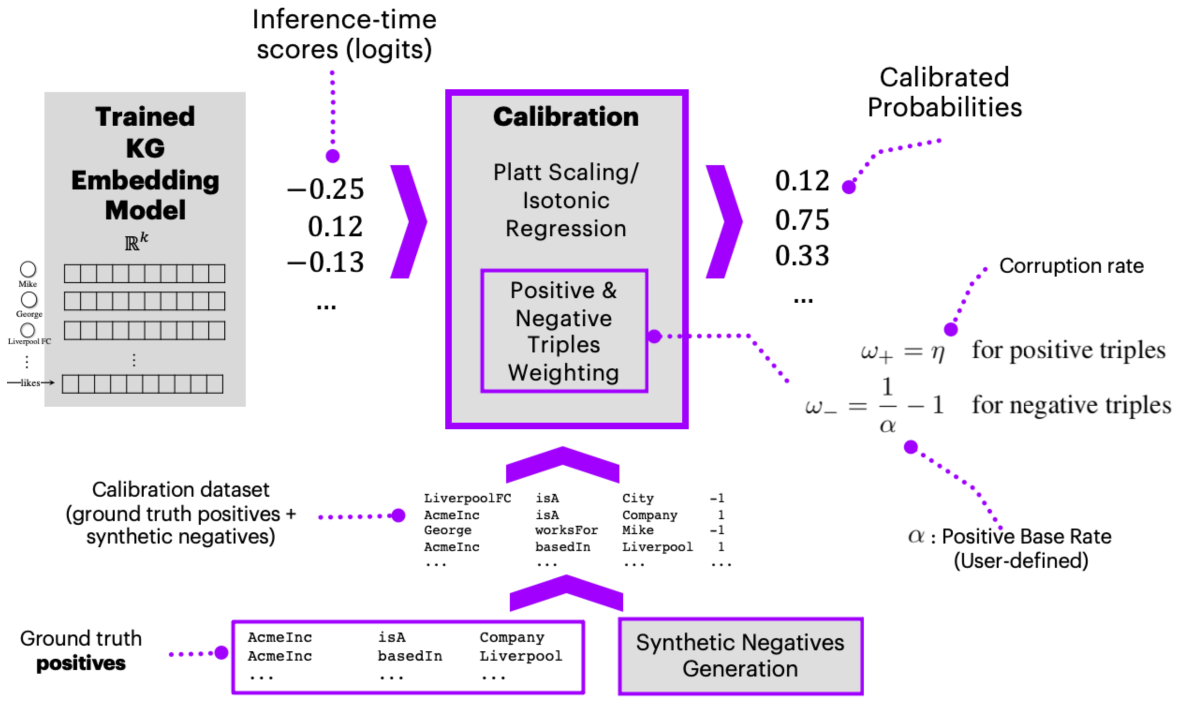


Not an exhaustive list

Explainability
from Interpreting and Understanding Predictions...

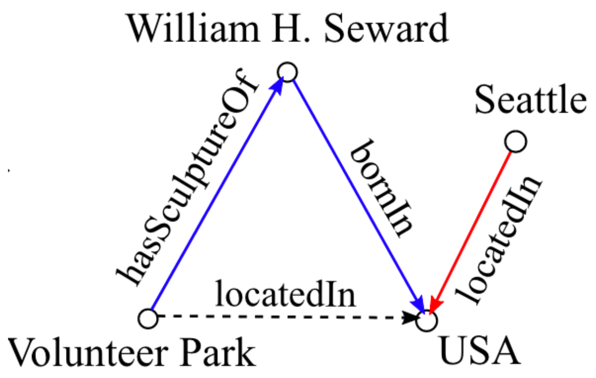
Predictions from trained KGE are hard to interpret, calibration allows to estimate output probabilities.

[Tabacof & Costabello ICLR 2020]



Estimate influence of triples.

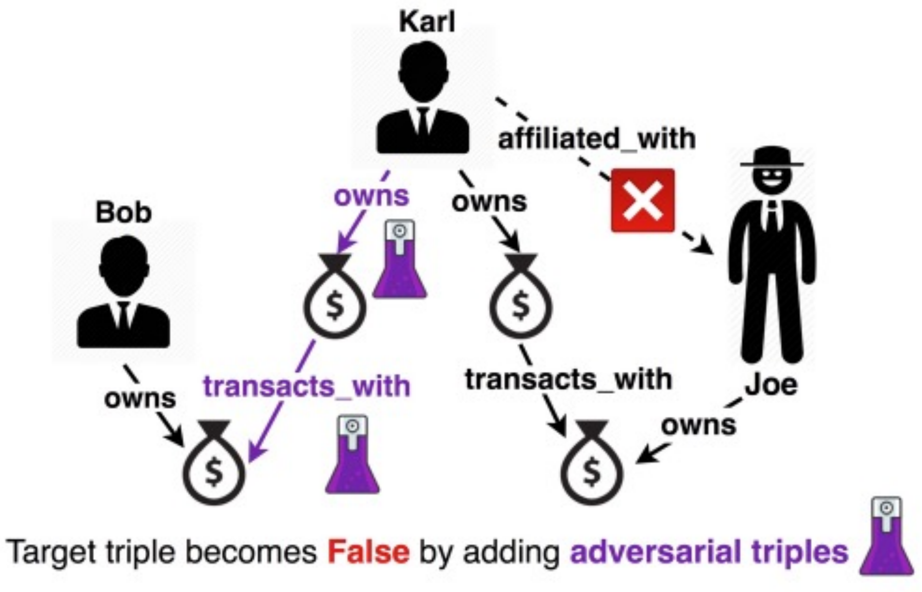
Gradient Rollback
[Lawrence et al. 2021]



… through adversarial attacks and robustness…

Identify most influential triple…

…and delete it.



[Bhardwaj EMNLP-21]
[Bhardwaj ACL-21]

… to adversarial explanations.

Identify regularities in the knowledge graph and attack.

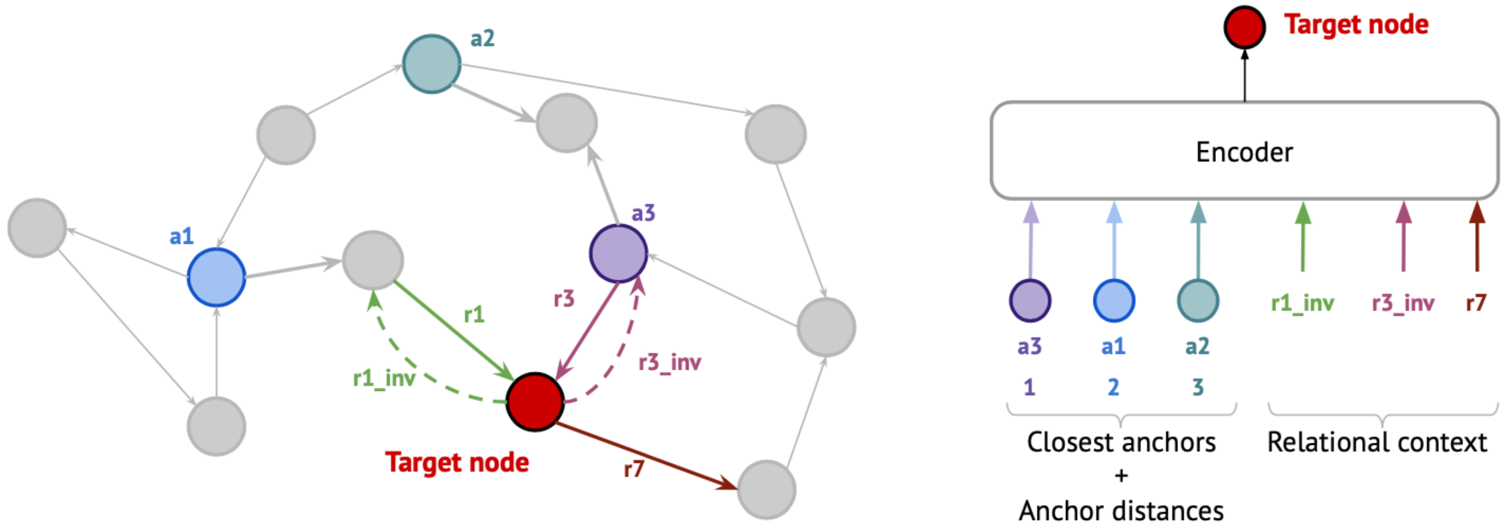
(E1) t_1: *nationality(Sawashiro, Japan)*
$\mathcal{E}_1$: *born(Sawashiro, Tokyo), located(Tokyo, Japan)*
ϕ_1^*: *nationality(X, Y)* $\leftarrow$ *born(X,Z)* $\wedge$ *located(Z,Y)* [0.76]

[Betz et. al IJCAI-22]

Memory
Learning More Compact Representations

NodePiece: up to 10x less paramters by learning a target node embedding by passing a hashed sequence of top-k closest nodes, relation types, and distances to target to an encoder.

[Galkin et al. 2021]



KGE & Neuro-Symbolic Reasoning

Background knowledge injection with Soft Constraints

[Minervini et al. 2017]

Manually provide rules (or mine with AMIE+) and inject into loss function:

(a) Axioms extracted from YAGO3		
HASNEIGHBOR	$\equiv$	HASNEIGHBOR$^-$
ISMARRIEDTO	$\equiv$	ISMARRIEDTO$^-$
PLAYSFOR	$\equiv$	ISAFFILIATEDTO
ISCONNECTEDTO	$\equiv$	ISCONNECTEDTO$^-$

(b) Axioms extracted from DBPEDIA		
ASSOC. BAND	$\equiv$	ASSOC. MUSICAL ARTIST
MUSICAL BAND	$\equiv$	MUSICAL ARTIST

We extend $\mathcal{L}$ with the **regularization term** $\mathcal{R}_{\mathcal{S}}$:

$$\mathcal{L}_{\mathcal{S}}(\Theta) = \mathcal{L}(\Theta) + \lambda \mathcal{R}_{\mathcal{S}}(\Theta)$$

$\lambda = \infty$ hard constraints

$\lambda = 0$ original model

$$\mathcal{R}_{\mathcal{S}}(\Theta) \triangleq \sum_{p \equiv q \in \mathcal{A}_1} D\left[\mathbf{r}_p \| \mathbf{r}_q\right] + \sum_{p \equiv q^- \in \mathcal{A}_2} D\left[\mathbf{r}_p \| \Phi(\mathbf{r}_q)\right]$$

$D[x||y] = ||x - y||_2^2$: *Divergence measure*

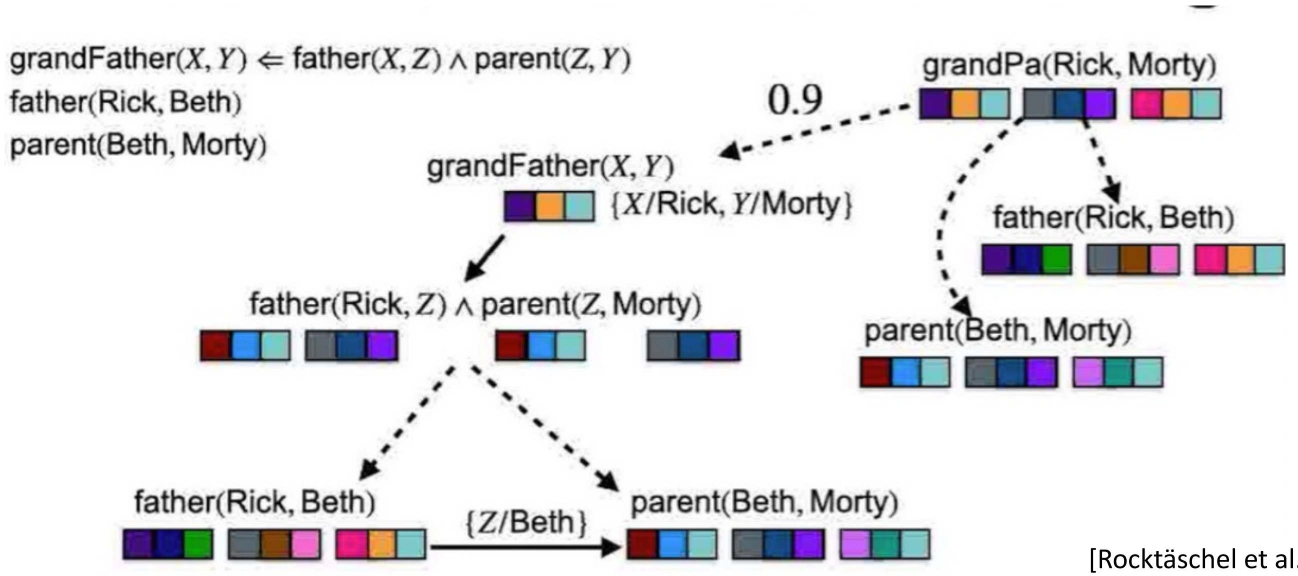
$\Phi(\cdot)$: *Model-dependent transformation*

$\mathcal{A}_1$: equivalent axioms set

$\mathcal{A}_2$: inverse axioms set

KGE & Neuro-Symbolic Reasoning: Neural Theorem Provers (NTP)

- Rule-based models + KGE
- Interplay of KGE strengths (good generalization power, scalability) with rule-based interpretability ("small data" capabilities).
- NTP implement reasoning (e.g. backward chaining) in fully differentiable architectures
  - Symbols replaced by embeddings
  - Compare embeddings in Prolog backward chaining instead of matching symbols

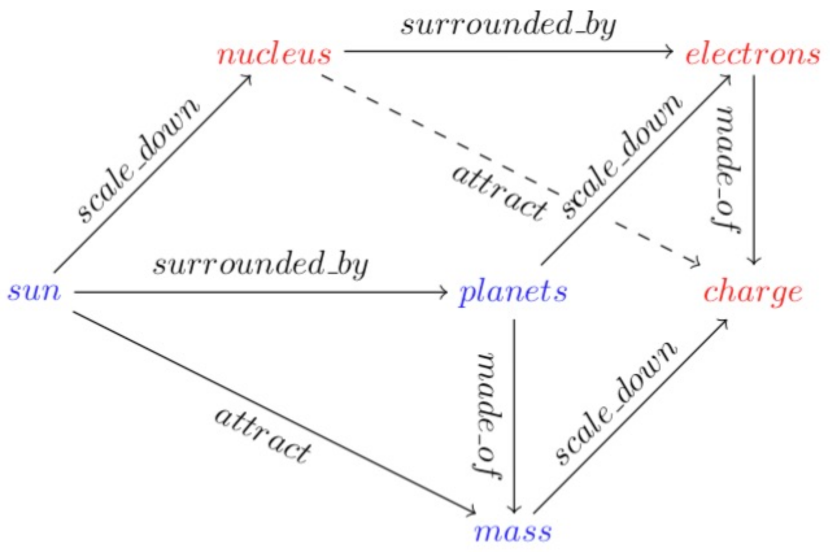


[Rocktäschel et al. 2017]
[Minervini et al. 2020]

Interplay with Other Reasoning Regimes: Analogical Reasoning

ANALOGY [Liu et al 2017]

- Models analogical structures in multi-relationa embeddings.
- “Differentiable” analogical reasoning combined with KGE models.



(Some) Open Research Questions

MORE EXPRESSIVE MODELS

Model KG regularities and dependencies while keeping runtime/space complexity low.

SUPPORT FOR MULTIMODALITY

Node and edge attributes, time-awareness still in their infancy.

ROBUSTNESS & INTERPRETABILITY

Techniques to dissect, investigate, explain, and protect from adversarial attacks.

BEYOND LINK PREDICTION

Multi-path predictions, adoption in larger differentiable architectures to inject background knowledge from graphs.

Literature

[1] G. A. Gesese, R. Biswas, M. Alam, and H. Sack, "A Survey on Knowledge Graph Embeddings with Literals: Which model links better Literal-ly?" arXiv, May 14, 2020. doi: 10.48550/arXiv.1910.12507.

[2] P. Bhardwaj, J. Kelleher, L. Costabello, and D. O'Sullivan, "Adversarial Attacks on Knowledge Graph Embeddings via Instance Attribution Methods," Nov. 2021. Accessed: Oct. 03, 2022. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv211103120B

[3] H. Liu, Y. Wu, and Y. Yang, "Analogical Inference for Multi-relational Embeddings," in *Proceedings of the 34th International Conference on Machine Learning*, Jul. 2017, pp. 2168–2178. Accessed: Oct. 03, 2022. [Online]. Available: https://proceedings.mlr.press/v70/liu17d.html

[4] E. Arakelyan, D. Daza, P. Minervini, and M. Cochez, "Complex Query Answering with Neural Link Predictors," presented at the International Conference on Learning Representations, Mar. 2022. Accessed: Oct. 03, 2022. [Online]. Available: https://openreview.net/forum?id=Mos9F9kDwkz

[5] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, "Diachronic Embedding for Temporal Knowledge Graph Completion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, Art. no. 04, Apr. 2020, doi: 10.1609/aaai.v34i04.5815.

[6] Y. Ma, V. Tresp, and E. A. Daxberger, "Embedding models for episodic knowledge graphs," *Journal of Web Semantics*, vol. 59, p. 100490, Dec. 2019, doi: 10.1016/j.websem.2018.12.008.

[7] P. Pezeshkpour, L. Chen, and S. Singh, "Embedding Multimodal Relational Data for Knowledge Base Completion," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, Oct. 2018, pp. 3208–3218. doi: 10.18653/v1/D18-1359.

[8] T. Rocktäschel and S. Riedel, "End-to-end Differentiable Proving," in *Advances in Neural Information Processing Systems*, 2017, vol. 30. Accessed: Oct. 03, 2022. [Online]. Available: https://papers.nips.cc/paper/2017/hash/b2ab001909a8a6f04b51920306046ce5-Abstract.html

[9] A. Hogan *et al.*, "Knowledge Graphs," *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–37, May 2022, doi: 10.1145/3447772.

[10] S. Pai and L. Costabello, "Learning Embeddings from Knowledge Graphs With Numeric Edge Attributes," Aug. 2021, vol. 3, pp. 2869–2875. doi: 10.24963/ijcai.2021/395.

[11] P. Minervini, S. Riedel, P. Stenetorp, E. Grefenstette, and T. Rocktäschel, "Learning reasoning strategies in end-to-end differentiable proving," in *Proceedings of the 37th International Conference on Machine Learning*, Jul. 2020, pp. 6938–6949.

[12] A. García-Durán, S. Dumančić, and M. Niepert, "Learning Sequence Encoders for Temporal Knowledge Graph Completion," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, Oct. 2018, pp. 4816–4821. doi: 10.18653/v1/D18-1516.

[13] M. Galkin, E. Denis, J. Wu, and W. L. Hamilton, "NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs," presented at the International Conference on Learning Representations, Feb. 2022. Accessed: Oct. 03, 2022. [Online]. Available: https://openreview.net/forum?id=xMJWUKJnFSw

[14] P. Bhardwaj, J. Kelleher, L. Costabello, and D. O'Sullivan, "Poisoning Knowledge Graph Embeddings via Relation Inference Patterns," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online, Aug. 2021, pp. 1875–1888. doi: 10.18653/v1/2021.acl-long.147.

[15] P. Tabacof and L. Costabello, "PROBABILITY CALIBRATION FOR KNOWLEDGE GRAPH EMBEDDING MODELS," p. 15, 2020.

[16] H. Ren, W. Hu, and J. Leskovec, "Query2box: Reasoning over Knowledge Graphs in Vector Space using Box Embeddings," Sep. 2020, Accessed: Oct. 03, 2022. [Online]. Available: https://openreview.net/forum?id=uM4k4_nnEF

[17] P. Minervini, L. Costabello, E. Muñoz, V. Nováček, and P.-Y. Vandenbussche, "Regularizing Knowledge Graph Embeddings via Equivalence and Inversion Axioms," in *Machine Learning and Knowledge Discovery in Databases*, Cham, 2017, pp. 668–683. doi: 10.1007/978-3-319-71249-9_40.

[18] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor Decompositions for Temporal Knowledge Base Completion," presented at the International Conference on Learning Representations, Apr. 2020. Accessed: Oct. 03, 2022. [Online]. Available: https://openreview.net/forum?id=rke2P1BFwS

[19] T. Jiang *et al.*, "Towards Time-Aware Knowledge Graph Completion," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, Osaka, Japan, Dec. 2016, pp. 1715–1724. Accessed: Oct. 03, 2022. [Online]. Available: https://aclanthology.org/C16-1161

Outline

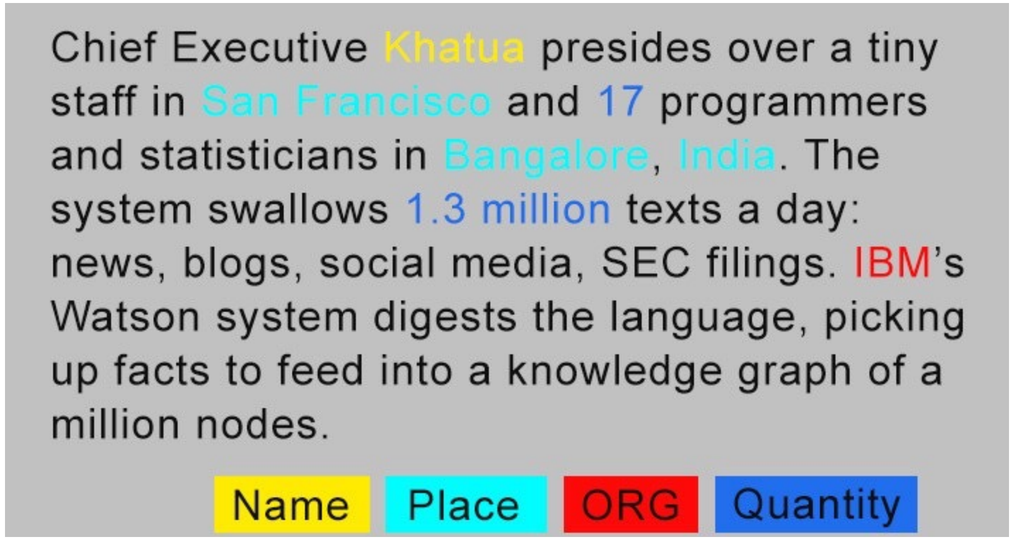
Foundations	40m
Advanced KGE Topics	20m
KGs for NLP	**30m**
Break	30m
Live Q&A	10m
Applications & Software Ecosystem	20m
Hands-on Session	45m
Live Q&A	10m

Mapping from Natural Language Text to KG

- Semantic gap: ambiguity
  - Named entity recognition
  - Relationship extraction and Linking
  - Entity Linking (Named Entity Disambiguation)
  - Schema linking

Mapping from Natural Language Text to KG

- Semantic gap: ambiguity
  - Named entity recognition
  - Relationship extraction and Linking
  - Entity Linking (Named Entity Disambiguation)
  - Schema linking

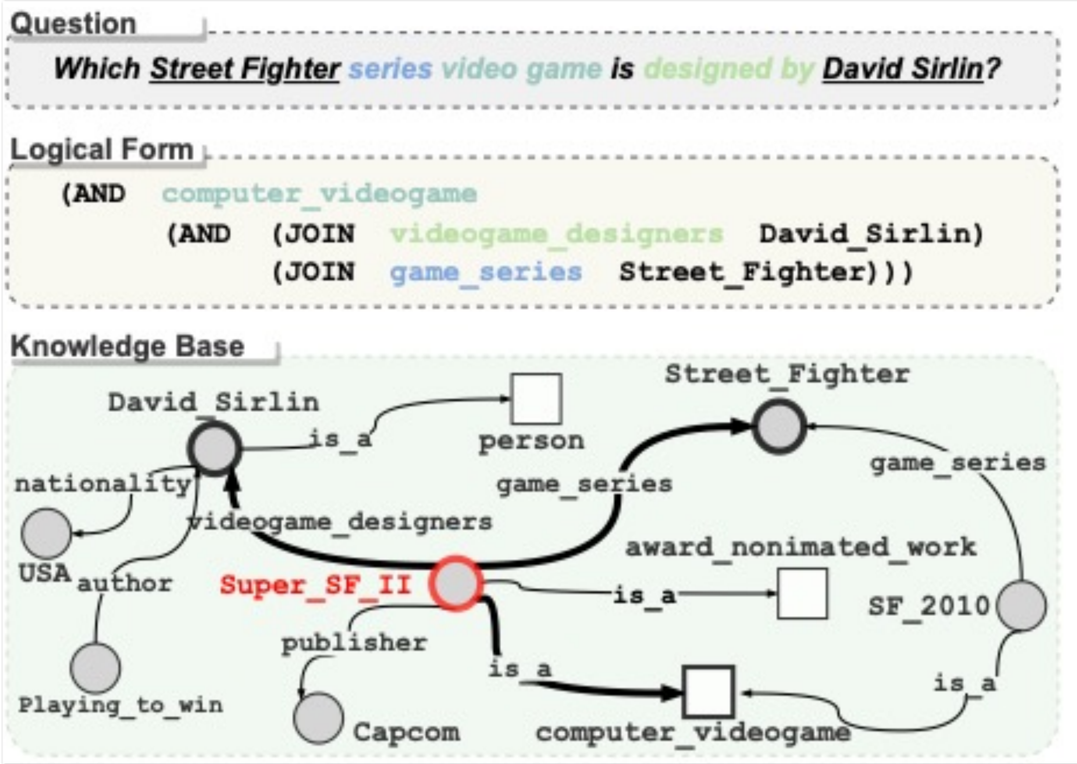


https://medium.com/cogitotech/how-does-named-entity-recognition-work-ner-methods-f23201a69648

Natural Language Question Answering over Knowledge Graphs

- Given NLQ, locate the answer entity in KG

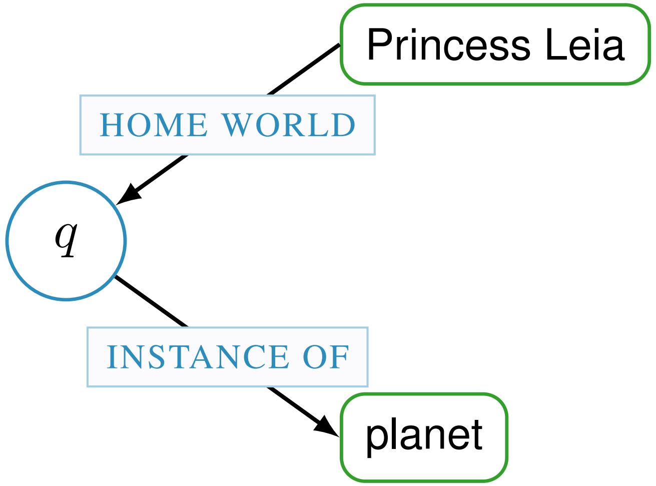
- **Semantic parsing**
  - mapping the NLQ onto logic form



[GU, PAHUJA, CHENG & SU, 2022]

Natural Language Question Answering over Knowledge Graphs

- Given NLQ, locate the answer entity in KG
- **Semantic parsing**
  - mapping the NLQ onto logic form
- **Semantic graph** from the input NLQ
  - consisting of the queried entities, relations from the KG



Semantic graph for question
"What is Princess Leia's home planet?"

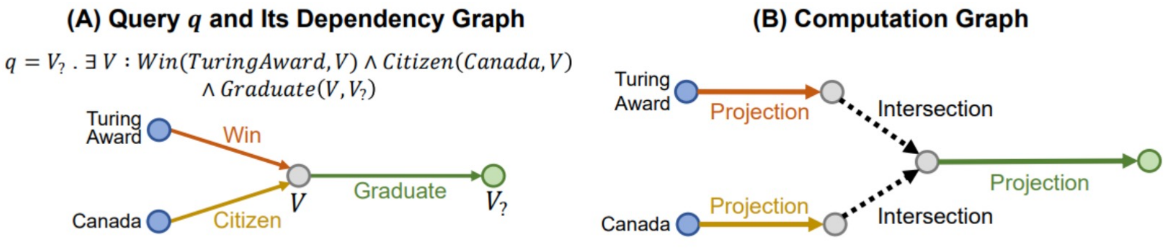
[Sorokin & Gurevych, 2018]

Complex Logical Query Answering

First-Order Logic (FOL) queries on KGs, involving logic operations

- existential quantifier (∃)
- conjunction (∧)
- disjunction (∨)
- negation (¬)

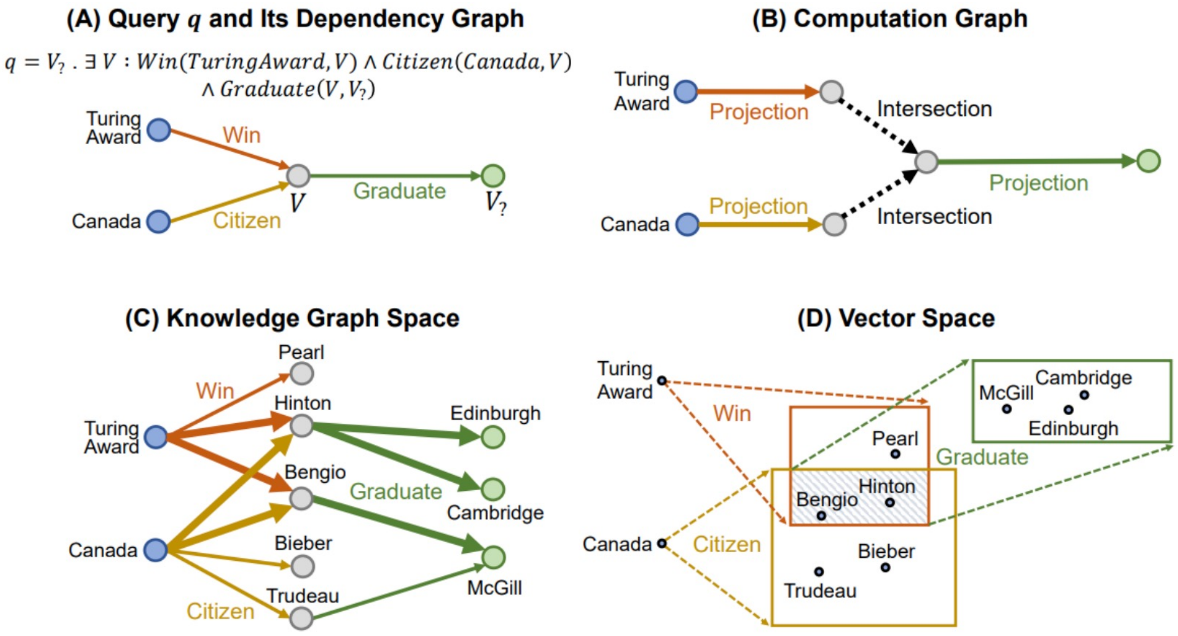
[Ren et al. 2020]



Complex Logical Query Answering over incomplete KG

First-Order Logic (FOL) queries on KGs, involving logic operations

- existential quantifier (∃)
- conjunction (∧)
- disjunction (∨)
- negation (¬)



[Ren et al. 2020]

Query2box: reasoning over Knowledge Graphs in a vector space using *box embeddings* to answer complex queries.

Complex Logical Query Answering over incomplete KG

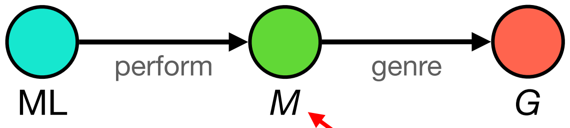
Continuous Query Decomposition (CQD)

- Compositional framework: pretrained KGE + optimization on logic query
  - Out-of-distribution generalization
  - Explainability

[Arakelyan et al., 2020]

"In what genres of movies did Martin Lawrence appear?"

$?G : \exists M \,.\, \text{perform}(\text{ML}, M) \wedge \text{genre}(M, G)$



Query: $?G : \exists M.\text{perform}(\text{ML}, M) \wedge \text{genre}(M, G)$

M	G	**Rank**	**Correctness**
Do the Right Thing	Drama	1	✓
	Comedy	4	✓
	Crime Fiction	7	✓
National Security	Action	2	✓
	Thriller	3	✓
	Crime Fiction	5	✓
The Nutty Professor	Comedy	6	✓
	Romantic Com.	8	✗
	Romance Film	9	✗

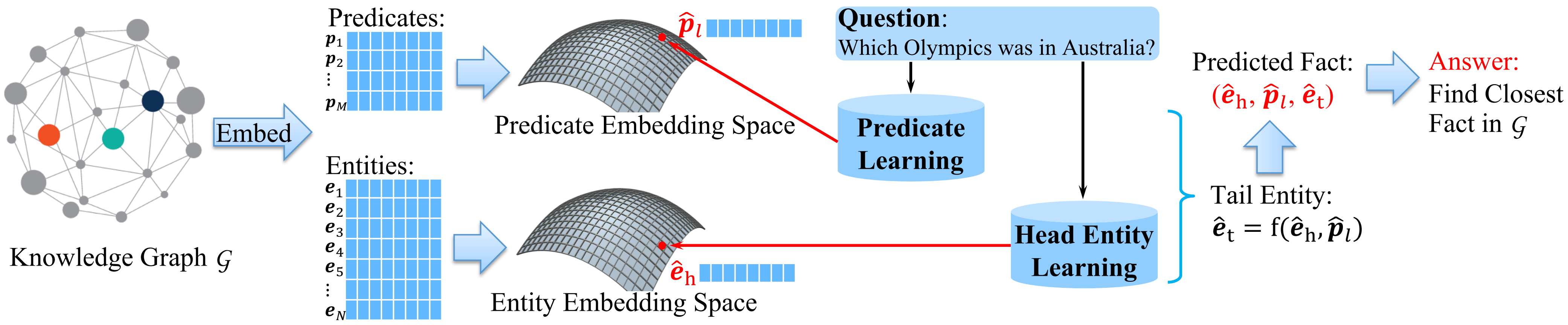
Intermediate variable assignment

KG Embedding Learning for NLQA

Knowledge graph embedding based question answering (KEQA)

- Given a simple NLQ
  return corresponding head entity and predicate embedding
- Joint representation search on embedding spaces
- Scoring function of the KGE model used to predict tail entity embedding

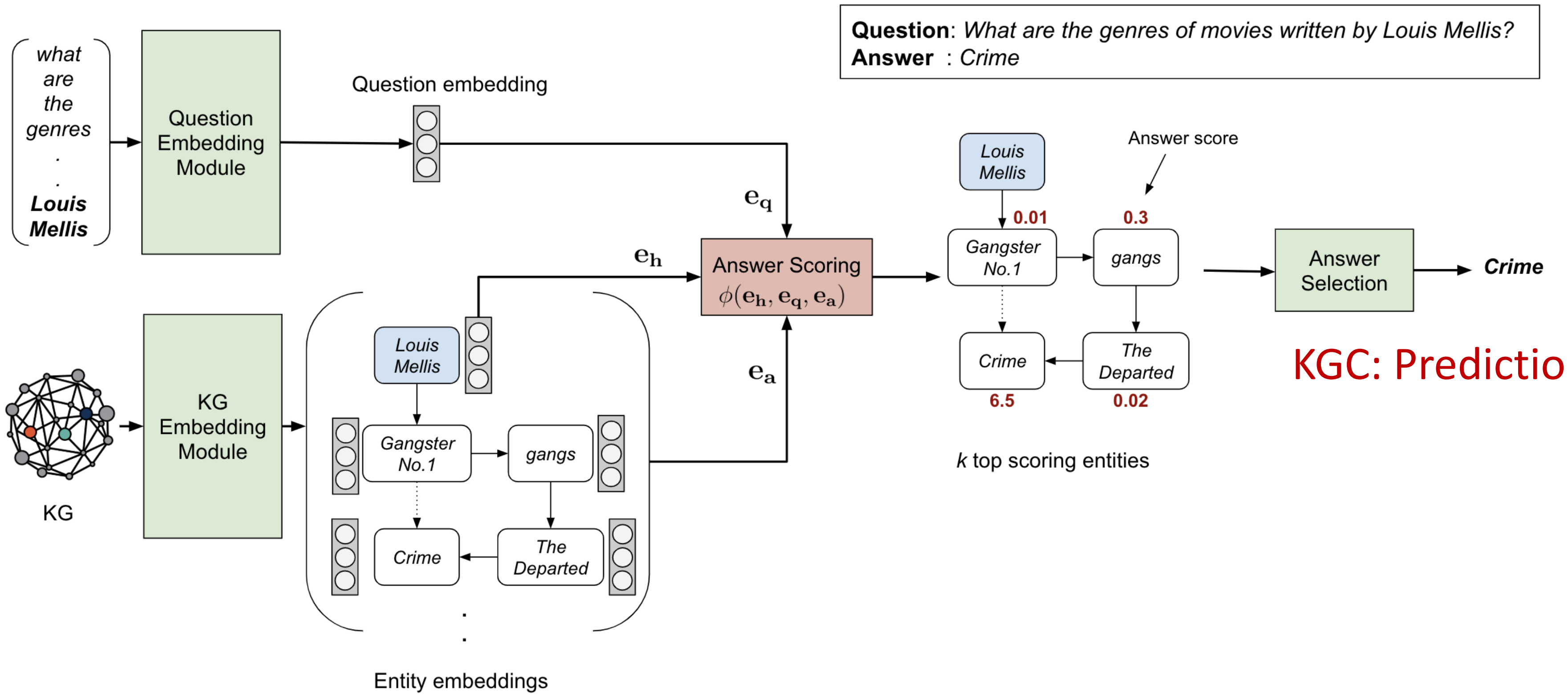
[Huang et al., 2019]



KG Embedding Learning for NLQA and KGC

Multi-hop Question Answering using KGE (EmbedKGE)

[Saxena et al., 2020]



NLQA

KGC: Prediction of missing links

Pre-trained LM for KG completion

- Fine-tune LM with textual description of the entities and relations

- KG-BERT [Yao etl al., 2019]
  - Link prediction using scoring function of the KG-BERT language model
  - Generalize representations of unseen entities during training

Inductive Link Prediction

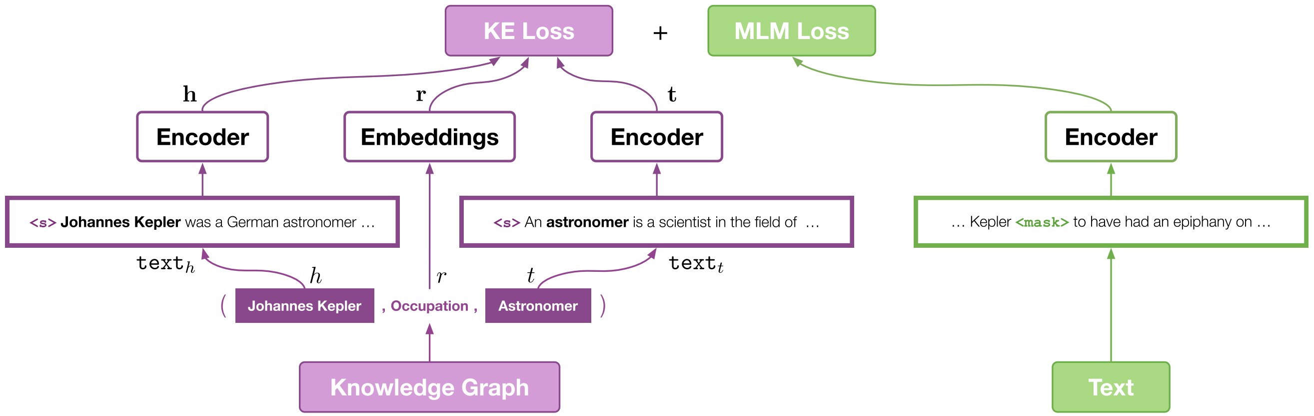
Pre-trained LM for KG completion

- Fine-tune BERT with textual description of the entities and relations

- KG-BERT [Yao etl al., 2019]
  - Link prediction using scoring function of the KG-BERT language model
  - Generalize representations of unseen entities during training

Inductive Link Prediction

- KEPLER [Wang et al., 2021]
  - jointly optimizing pretrained LM and KGE model



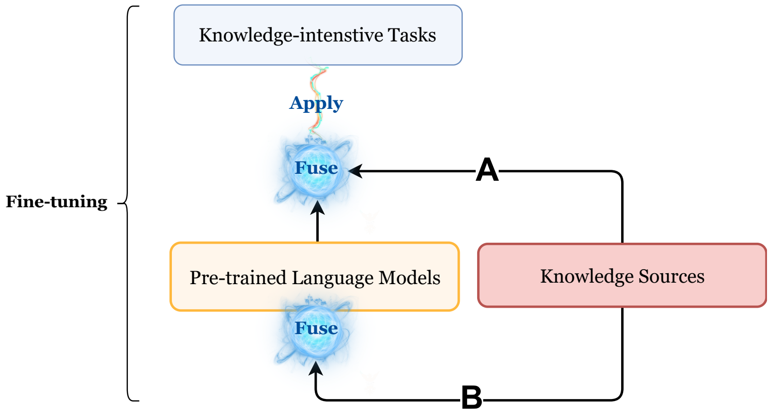
NLP tasks
+
Inductive Link Prediction

Knowledge Intensive Tasks

- Pretrained LMs
  - Embodies implicit broad world knowledge
  - BUT lack of explicit structured knowledge

Knowledge Intensive Tasks

- Pretrained LMs
  - Embodies implicit broad world knowledge
  - BUT lack of explicit structured knowledge
- Knowledge intensive NLP tasks
  - Entity linking
  - NLQA → question answering systems, dialog systems
  - Fact verification
- KG enhanced LMs: Infusing structured world knowledge into pretrained LM through KG
  - Infusing can take place in
    - Pretraining stage → pre-fusion
    - Fine-tuning stage → post-fusion >> joint reasoning
    - Hybrid



[Yin et al., 2022]

Joint reasoning with pretrained LMs and KGs

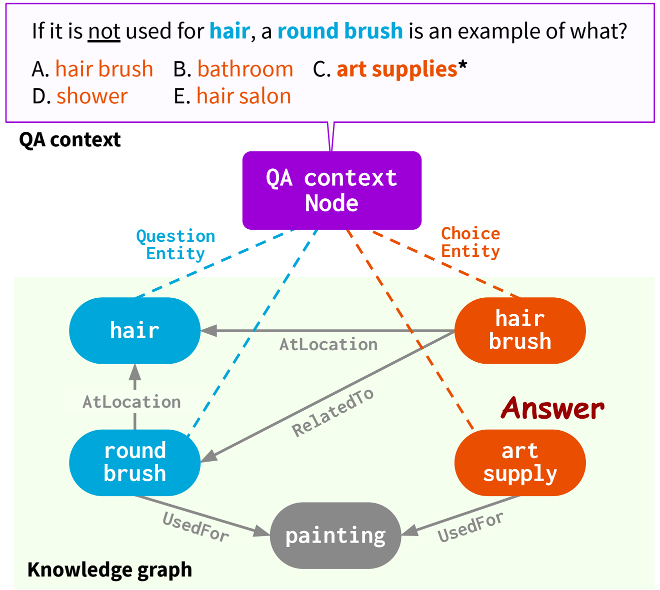
- Combining LMs and KGs

LMs	KGs
✓ Broad coverage	Lack of coverage, noisy, sparse
Lack of structured reasoning	✓ Explainability due to structure

- Subproblems
  1. How to identify relevant knowledge in the KG
  2. How to perform the joint reasoning

Joint reasoning (LMs + GNNs)

- QA-GNN [Yasunaga et al., 2021]
  - Encode QA context using LM
  - Retrieve a KG subgraph
  - Relevancy scoring of KG entities wrt QA context—solves (1)

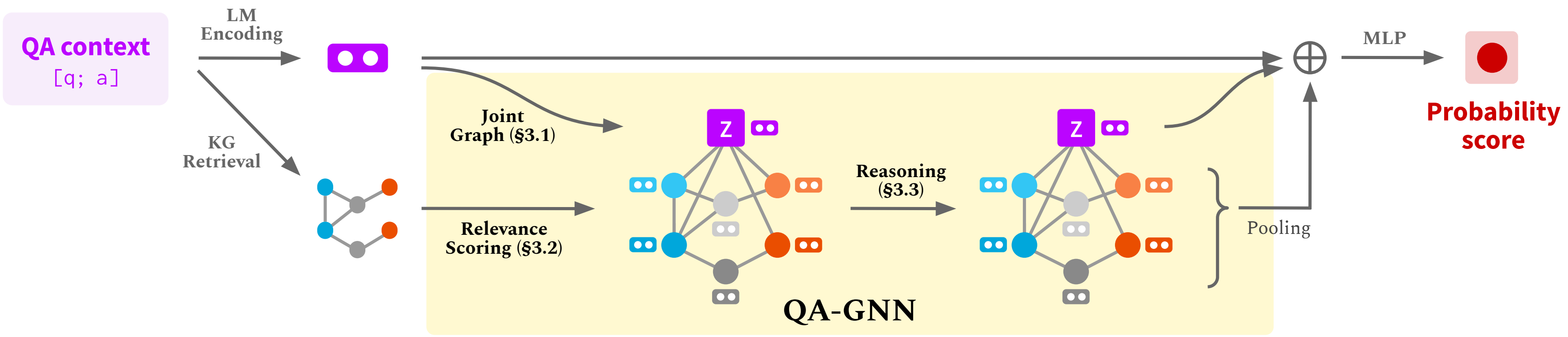


question entities in blue
answer choice entities in red

- Subproblems
  1. How to identify relevant knowledge in the KG
  2. How to perform the joint reasoning

Joint reasoning (LMs + GNNs)

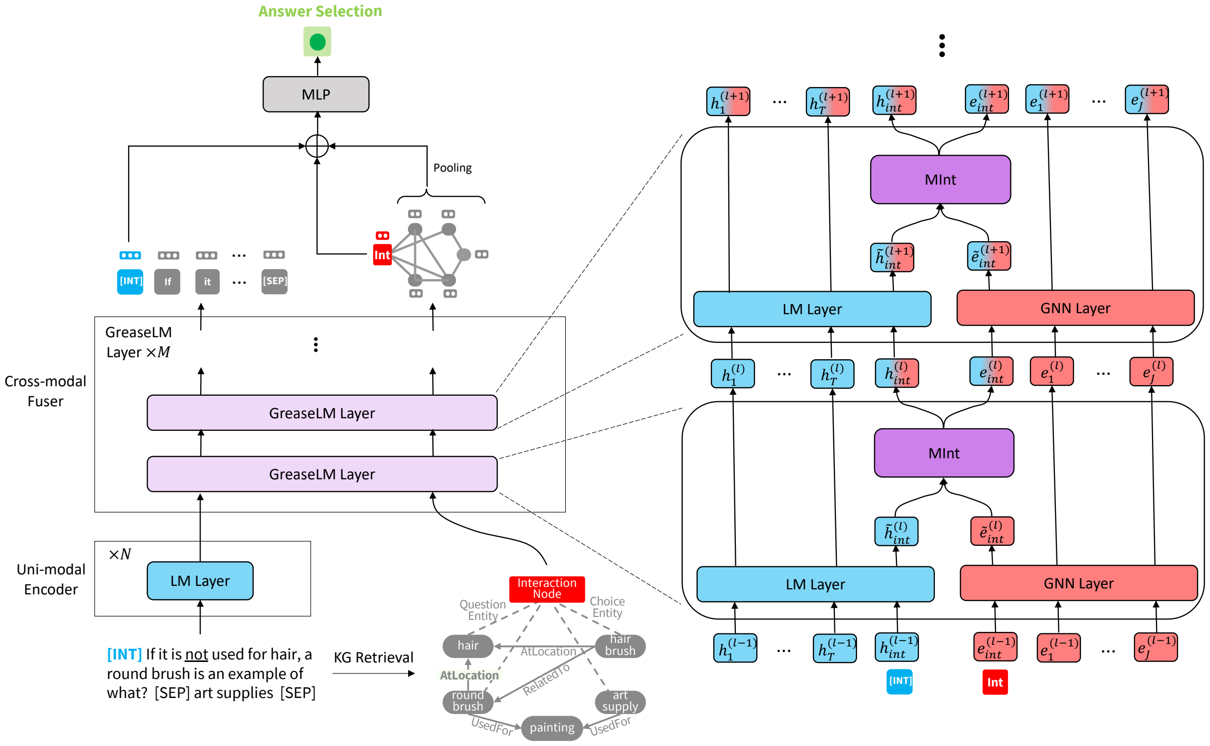
- QA-GNN [Yasunaga et al., 2021]
  - Encode QA context using LM
  - Retrieve a KG subgraph
  - Relevancy scoring of KG entities wrt QA context—solves (1)
  - GNN updating the graph representation — solves (2)



question entities in blue
answer choice entities in red

Joint reasoning (LMs + GNNs)

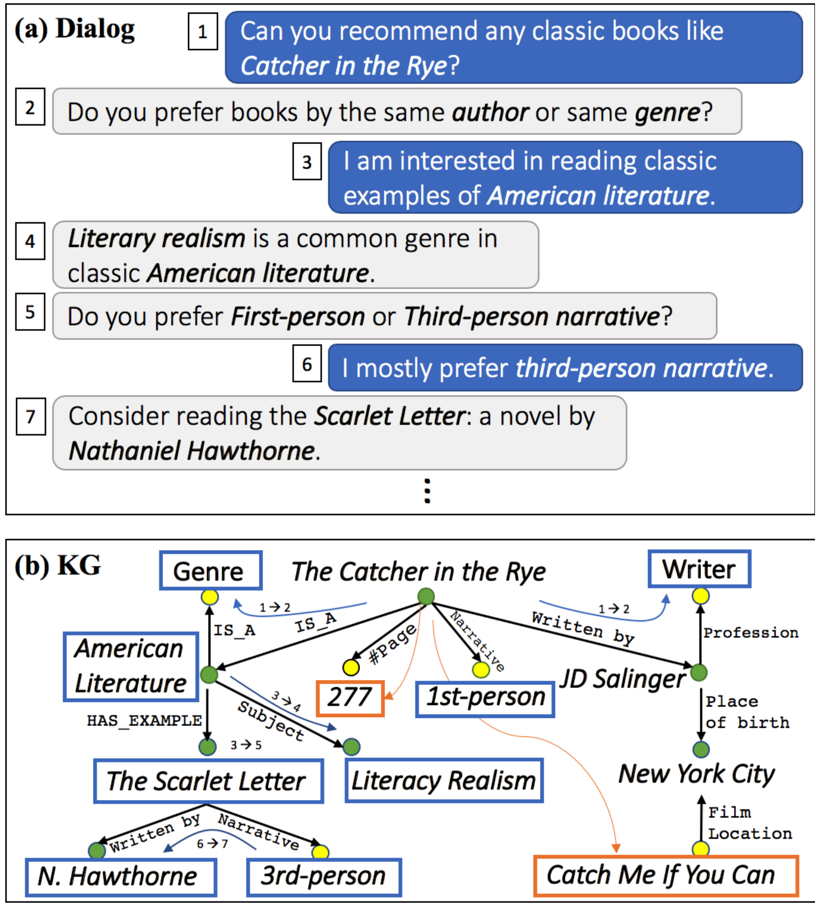
- GreaseLM: Graph Reasoning Enhanced Language Model [Zhang et al., 2022]
  - Representation fusion layers
  - Textual representation is jointly updated with graph representation.



The interaction token and node pulled, concatenated, and passed through a modality interaction (MInt) unit

Dialog Systems using KGs

- Conversational context is captured using pre-trained LM



- OpenDialKG [Moon et el., 2019]
- AttnIO [Jung et al., 2020]
- Neural Path Hunter [Dziri et al., 2021]

References

[GU, PAHUJA, CHENG & SU, 2022] Gu, Yu, Vardaan Pahuja, Gong Cheng, and Yu Su. "Knowledge Base Question Answering: A Semantic Parsing Perspective." arXiv preprint arXiv:2209.04994 (2022).
[Sorokin & Gurevych, 2018] Sorokin, Daniil, and Iryna Gurevych. "Modeling semantics with gated graph neural networks for knowledge base question answering." arXiv preprint arXiv:1808.04126 (2018).
[Ren et al. 2020] Ren, Hongyu, Weihua Hu, and Jure Leskovec. "Query2box: Reasoning over Knowledge Graphs in Vector Space Using Box Embeddings." In International Conference on Learning Representations. 2019.
[Arakelyan et al., 2020] Arakelyan, Erik, Daniel Daza, Pasquale Minervini, and Michael Cochez. "Complex Query Answering with Neural Link Predictors." In International Conference on Learning Representations. 2020.
[Huang et al., 2019] Huang, Xiao, Jingyuan Zhang, Dingcheng Li, and Ping Li. "Knowledge graph embedding based question answering." In Proceedings of the twelfth ACM international conference on web search and data mining, pp. 105-113. 2019.
[Saxena et al., 2020] Saxena, Apoorv, Aditay Tripathi, and Partha Talukdar. "Improving multi-hop question answering over knowledge graphs using knowledge base embeddings." In Proceedings of the 58th annual meeting of the association for computational linguistics, pp. 4498-4507. 2020.
[Yao etl al., 2019] Yao, Liang, Chengsheng Mao, and Yuan Luo. "KG-BERT: BERT for knowledge graph completion." arXiv preprint arXiv:1909.03193 (2019).
[Wang et al., 2021] Wang, Xiaozhi, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. "KEPLER: A unified model for knowledge embedding and pre-trained language representation." Transactions of the Association for Computational Linguistics 9 (2021): 176-194.
[Yin et al., 2022] Yin, Da, Li Dong, Hao Cheng, Xiaodong Liu, Kai-Wei Chang, Furu Wei, and Jianfeng Gao. "A survey of knowledge-intensive nlp with pre-trained language models." arXiv preprint arXiv:2202.08772 (2022).
[Yasunaga et al., 2021] Yasunaga, Michihiro, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. "QA-GNN: Reasoning with Language Models and Knowledge Graphs for Question Answering." In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 535-546. 2021.
[Zhang et al., 2022] Zhang, X., A. Bosselut, M. Yasunaga, H. Ren, P. Liang, C. Manning, and J. Leskovec. "GreaseLM: Graph REASoning Enhanced Language Models for Question Answering." In International Conference on Representation Learning (ICLR). 2022.
[Moon et el., 2019] Moon, Seungwhan, Pararth Shah, Anuj Kumar, and Rajen Subba. "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 845-854. 2019.
[Jung et al., 2020] Jung, Jaehun, Bokyung Son, and Sungwon Lyu. "Attnio: Knowledge graph exploration with in-and-out attention flow for knowledge-grounded dialogue." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 3484-3497. 2020.
[Dziri et al., 2021] Dziri, Nouha, Andrea Madotto, Osmar Zaiane, and Avishek Joey Bose. "Neural path hunter: Reducing hallucination in dialogue systems via path grounding." arXiv preprint arXiv:2104.08455 (2021).

Outline

Foundations	40m
Advanced KGE Topics	20m
KGs for NLP	30m
Break	30m
Live Q&A	10m
Applications & Software Ecosystem	**20m**
Hands-on Session	45m
Live Q&A	10m

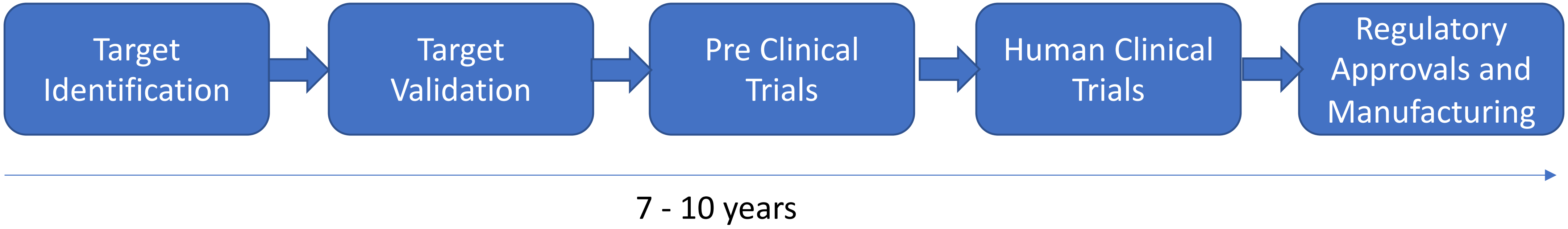
Industrial applications:

Pharmaceutical Industry
Drug Side-effects Prediction

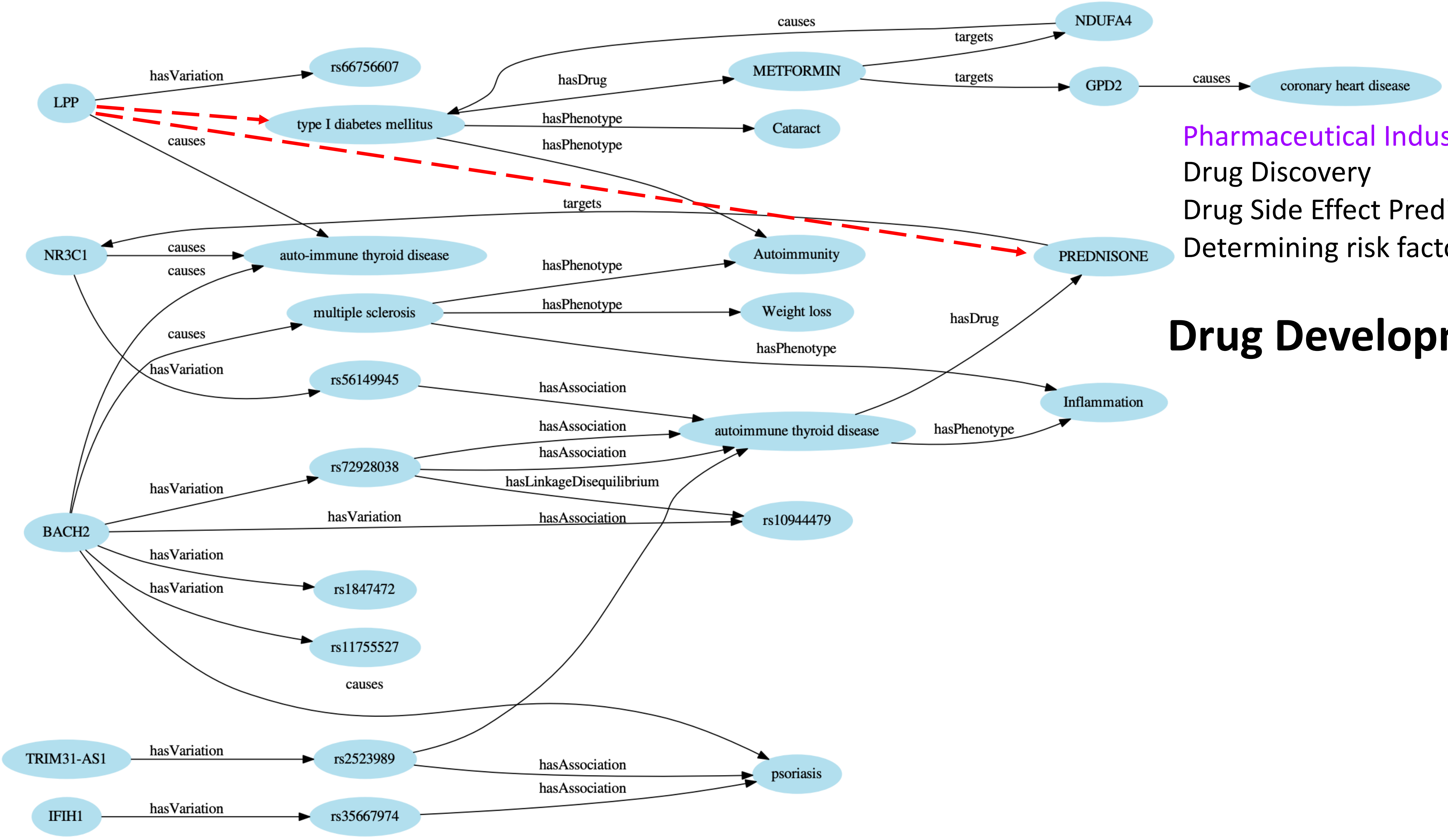
Oncology
Lung cancer patients relapse prediction

Products
Customer Segmentation

Drug Development



Time consuming & Expensive process

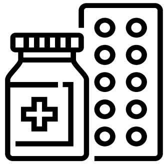


Pharmaceutical Industry:

Drug Discovery

Drug Side Effect Prediction

Determining risk factors



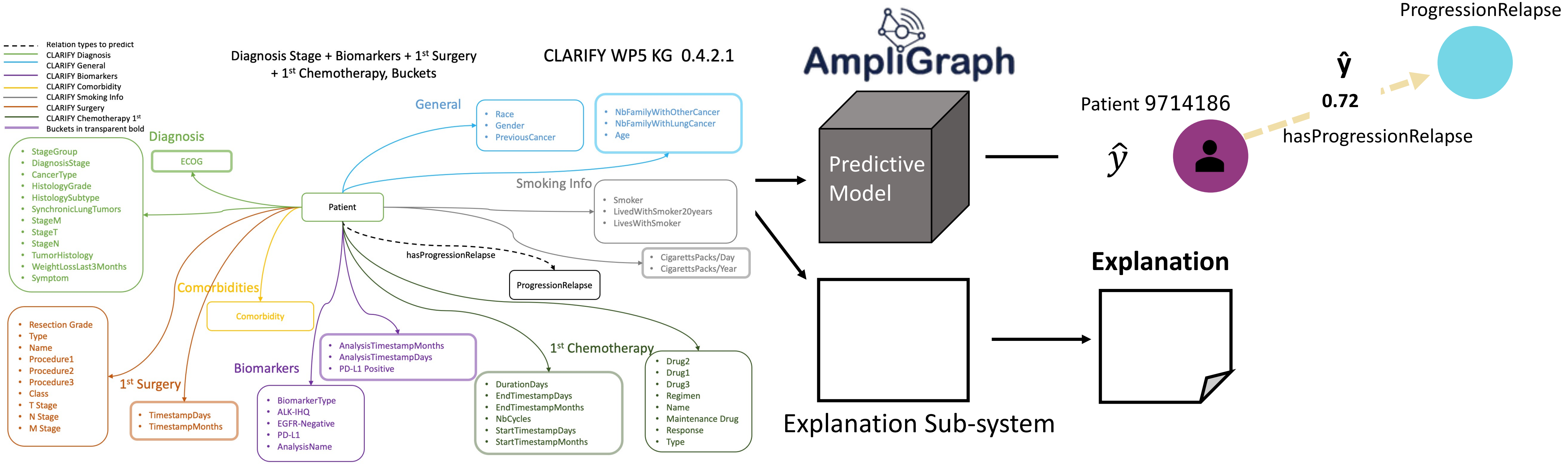
Drug Development

https://www.clarify2020.eu/

CLARIFY: Cancer Long Survivor Artificial Intelligence Follow-up

Oncology

Lung cancer patients relapse prediction



Explanation

We predict that **Patient 9714186** will relapse with **72%** of chances.

As evidence for the above prediction, we bring to your attention **1 similar patient that also relapsed** (Patient 7105830).

Features **in common** between Patient 9714186 and other similar relapsed patients (Patient 7105830):

Diagnosis

- **Tumor Histology: Squamous**
- **Stage Group: early stage**
- **Stage M: M0**
- **Diagnosis Stage: IIB**
- **Cancer Type: Non small cell lung cancer**

Comorbidities

- **[...]**

Nevertheless, some features of the similar relapsed patients (Patient 7105830) **do not occur in Patient 9714186:**

Diagnosis

- Stage T: T2b
- Symptom: asymptomatic
- Stage N: N1

Comorbidities [...]

Beer Customers Marketing Segmentation

[Pai et al. , 2022]

Challenge

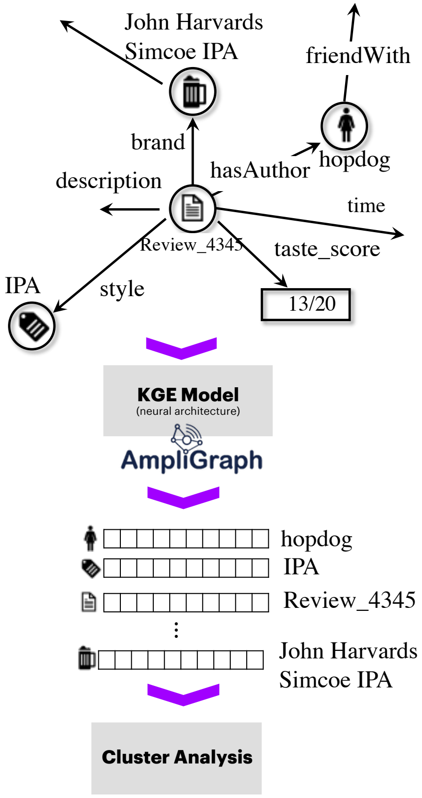
Traditional customer segmentation is **expensive, time-consuming** and **require labelled datasets** difficult to acquire.

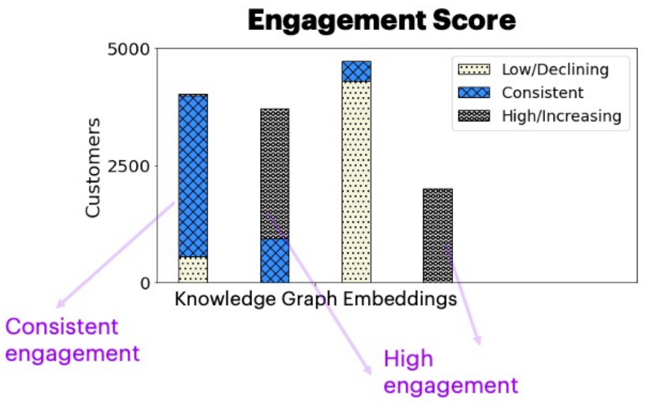
Qualitative surveys **fail to consider *consumption behavior***

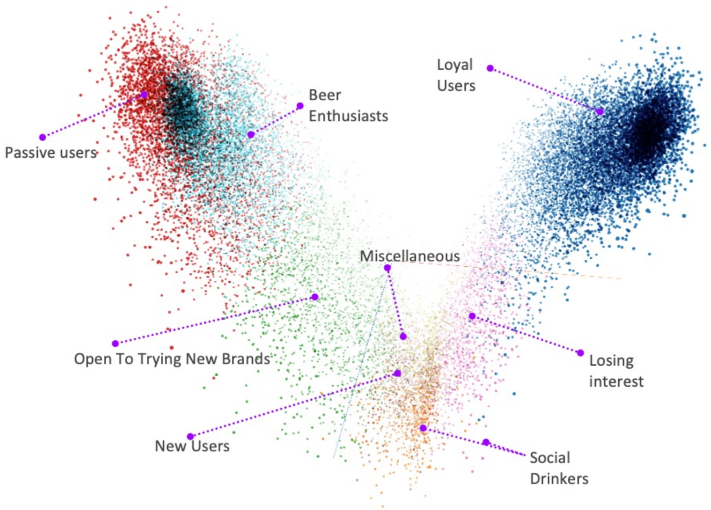
Can we learn meaningful consumer segments by only processing behavioral consumption sequences without human labeling and with no demographics or consumer-survey information?

Solution

- Hypothesis: sequences of consumed beers hold enough information to tell customers apart.
- **Unsupervised customer segmentation with graph + AmpliGraph knowledge graph embeddings**
- We learn meaningful clusters. No need for customer surveys or expensive data annotation campaigns.





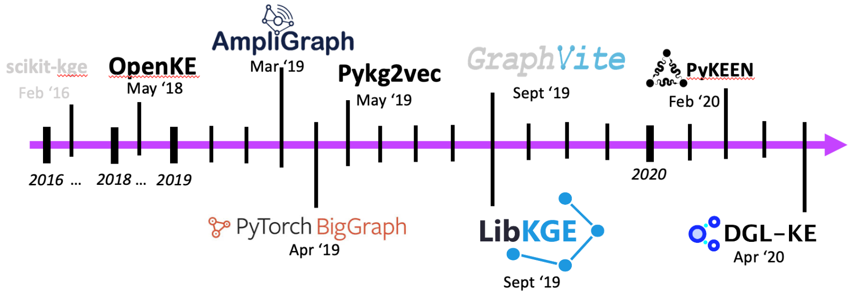


Software Ecosystem

- OpenKE
  - C++ implementation.
- AmpliGraph
  - Benchmarking Aid and pre-processing.
  - Formats: rdf, csv, ntriples.
  - Knowledge discovery API.
  - Visualization.
  - Model selection API.
  - Slack Channel.
  - Colab Tutorials.

Hands-on Session

- PyTorch-BigGraph
  - High-level operators.
  - Scalability (partitioning, experimental GPU).
- DGL-KE
  - Scalability (partitioning with METIS, faster than GraphVite and PBG).
- Pykeen
  - Incorporating multi-modal information.
  - Extensibility (wide range of interchangeable components 44 models, 15 losses and more...).
  - Multiple tracking options.
  - Hyperparameters support (Optuna).
  - 36 datasets + 5 inductive datasets.
  - Pipelines.
  - Ablation Studies.
  - Nodepiece.



[Costabello et al. ECAI 2020]

- Pykg2vec
  - Metrics summary plots.
  - Automatic discovery for hyperparameters.
  - Interactive results inspector.
- Libkge
  - Hyper param support (includes Bayesian Optimization).
  - Resuming training.
  - Configuration via yaml.
- Graphvite
  - Command line interface.
  - Visualization.
  - Configuration via yaml.
  - Auto-deduction of hyperparameters.
  - Scalability (GPU-CPU hybrid).
  - Node Embedding API.
  - Input data parser.
- muKG 2022
  - Multi-source knowledge graphs.
  - Multi process, multi GPU and more
- scikit-kge (discontinued)
- ...

Not an exhaustive list

References

1. AmpliGraph: a Library for Representation Learning on Knowledge Graphs, Luca Costabello and Sumit Pai and Chan Le Van and Rory McGrath and Nicholas McCarthy and Pedro Tabacof, 2019, https://doi.org/10.5281/zenodo.2595043
2. Ali, Mehdi, Max Berrendorf, Charles Tapley Hoyt, Laurent Vermue, Sahand Sharifzadeh, Volker Tresp, and Jens Lehmann. "PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings." *ArXiv:2007.14175 [Cs, Stat]*, July 30, 2020. http://arxiv.org/abs/2007.14175.
3. Han, Xu, Shulin Cao, Xin Lv, Yankai Lin, Zhiyuan Liu, Maosong Sun, and Juanzi Li. "OpenKE: An Open Toolkit for Knowledge Embedding," n.d., 6.
4. Lerer, Adam, Ledell Wu, Jiajun Shen, Timothee Lacroix, Luca Wehrstedt, Abhijit Bose, and Alex Peysakhovich. "PyTorch-BigGraph: A Large-Scale Graph Embedding System," n.d., 12.
5. Nickel, Maximilian, Lorenzo Rosasco, and Tomaso Poggio. "Holographic Embeddings of Knowledge Graphs." *ArXiv:1510.04935 [Cs, Stat]*, December 7, 2015. http://arxiv.org/abs/1510.04935.
6. Ruffinelli, Daniel, Samuel Broscheit, and Rainer Gemulla. "LibKGE You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings," 2019. https://openreview.net/forum?id=BkxSmlBFvr.
7. Yu, Shih Yuan, Sujit Rokka Chhetri, Arquimedes Canedo, Palash Goyal, and Mohammad Abdullah Al Faruque. "Pykg2vec: A Python Library for Knowledge Graph Embedding." *ArXiv:1906.04239 [Cs, Stat]*, June 4, 2019. http://arxiv.org/abs/1906.04239.
8. Zheng, Da, Xiang Song, Chao Ma, Zeyuan Tan, Zihao Ye, Jin Dong, Hao Xiong, Zheng Zhang, and George Karypis. "DGL-KE: Training Knowledge Graph Embeddings at Scale." *ArXiv:2004.08532 [Cs]* Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (April 18, 2020). http://arxiv.org/abs/2004.08532.
9. Zhu, Zhaocheng, Shizhen Xu, Meng Qu, and Jian Tang. "GraphVite: A High-Performance CPU-GPU Hybrid System for Node Embedding." *The World Wide Web Conference on - WWW '19*, 2019, 2494–2504. https://doi.org/10.1145/3308558.3313508.
10. Bianchi, Federico, Gaetano Rossiello, Luca Costabello, Matteo Palmonari, and Pasquale Minervini. "Knowledge Graph Embeddings and Explainable AI." *ArXiv:2004.14843 [Cs]*, April 30, 2020. https://doi.org/10.3233/SSW200011.

Thank you!

The work was partially funded by the European Union (EU) Horizon 2020 research and innovation program through the Clarify project (875160).
https://www.clarify2020.eu

Outline

Foundations	**40m**
• Introduction	
• Anatomy of a Knowledge Graph Embedding Model	
• Evaluation Protocol and Metrics	
Advanced KGE Topics	**20m**
• Advanced KGE Topics	
• Open Research Questions	
KGs for NLP	**30m**
Break	**30m**
Live Q&A	**10m**
Applications & Software Ecosystem	**20m**
Hands-on Session [bit.ly/kge4nlp-tutorial]	**45m**
Live Q&A	**10m**