Machine Learning with Graphs (MLG)

Link Prediction on Knowledge Graphs

Can we predict links conditioned on types?

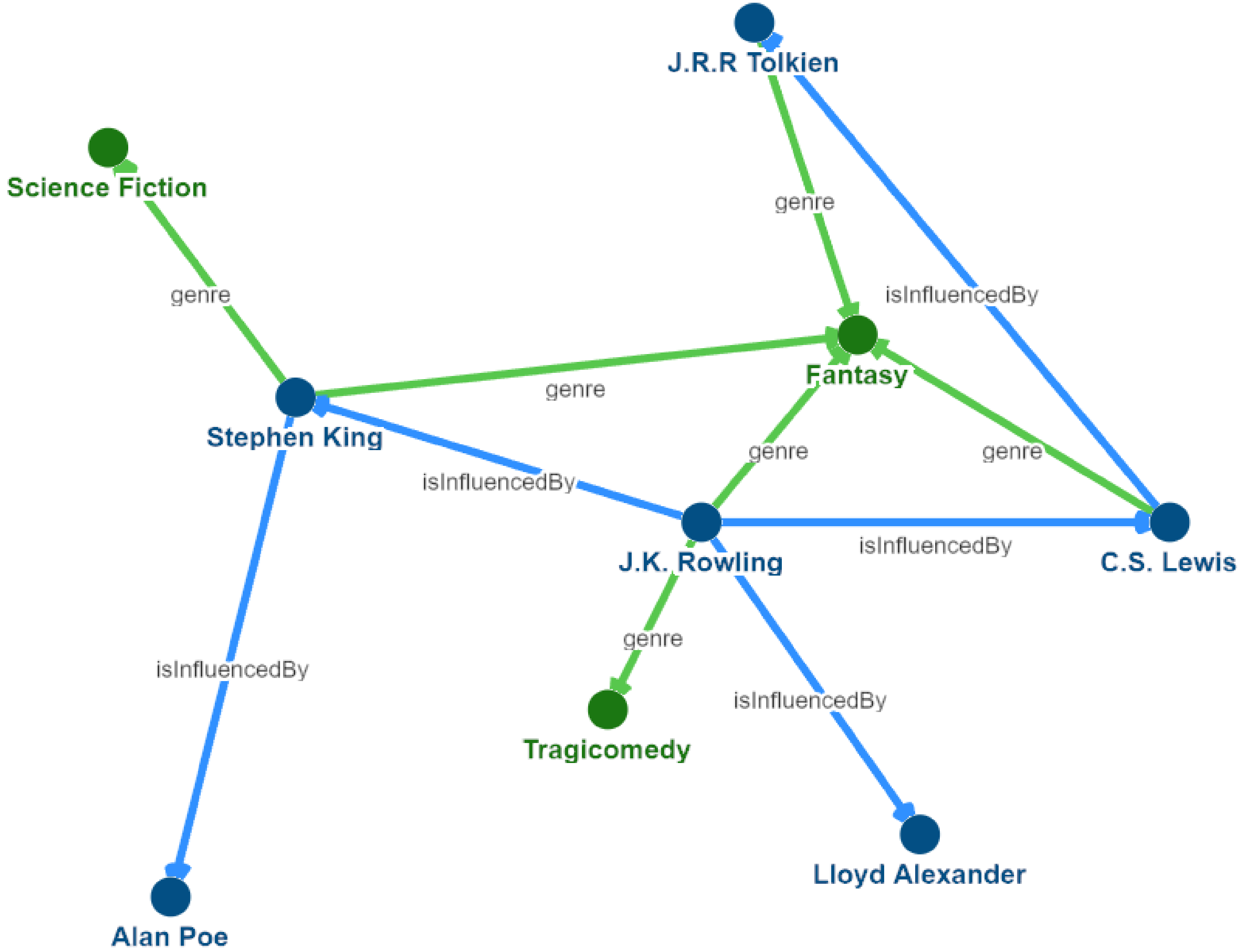
Cheng-Te Li (李政德)

Institute of Data Science

National Cheng Kung University

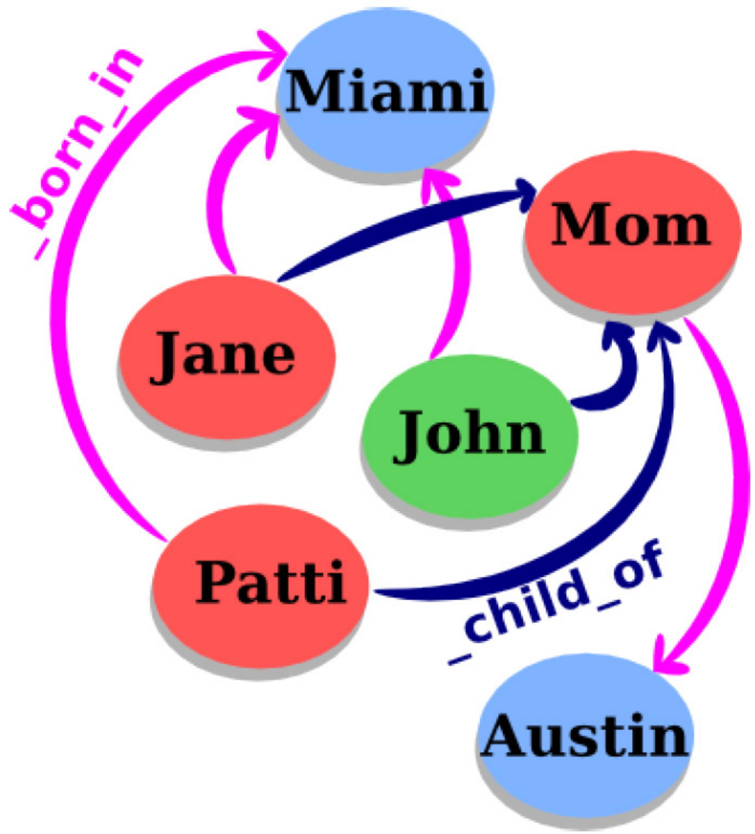
chengte@mail.ncku.edu.tw

Knowledge Graph



Knowledge Graph

- Data is structured as a graph
- Each **node** = an entity
- Each **edge** = a relation/fact
- A relation = (sub, rel, obj)
  - sub =subject
  - rel = relation type
  - obj = object
- Nodes w/o features



KG Examples

- WordNet: dictionary where each entity is a sense
  - Popular in NLP
  - 117K entities, 20 relation types, 500K facts
  - Examples:
    - (car_NN_1, has_part, wheel_NN_1)
    - (score_NN_1, is_a, rating_NN_1)
    - (score_NN_2, is_a, sheet_music_NN_1)

WORDNET

https://wordnet.princeton.edu/

- Freebase: huge collaborative knowledge base
  - Part of the Google Knowledge Graph
  - 80M entities, 20K relation types, 1.2B facts
  - Examples
    - (Barack Obama, place of birth, Hawai)
    - (Albert Einstein, follows diet, Veganism)
    - (San Francisco, contains, Telegraph Hill)

Freebase™

https://developers.google.com/freebase

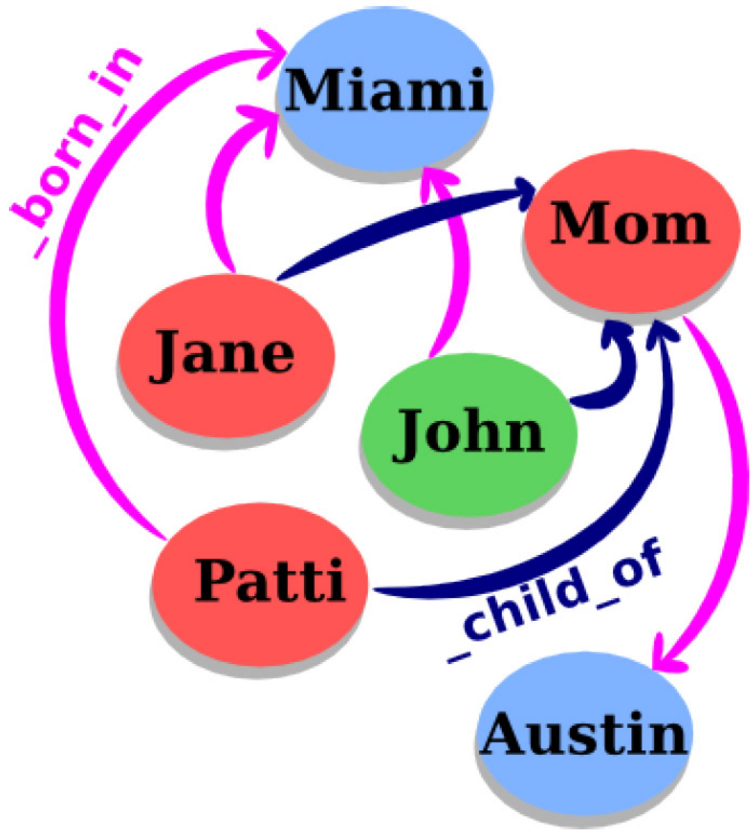
Link Prediction on Knowledge Graphs

- Add new facts without requiring extra knowledge
- From known info to access the validity of an unknown fact
  - Collective classification
  - Reasoning in embedding space
- Goal
  - We want to model, from data,

$$P\left[rel_k\left(sub_i, obj_j\right) = 1\right]$$

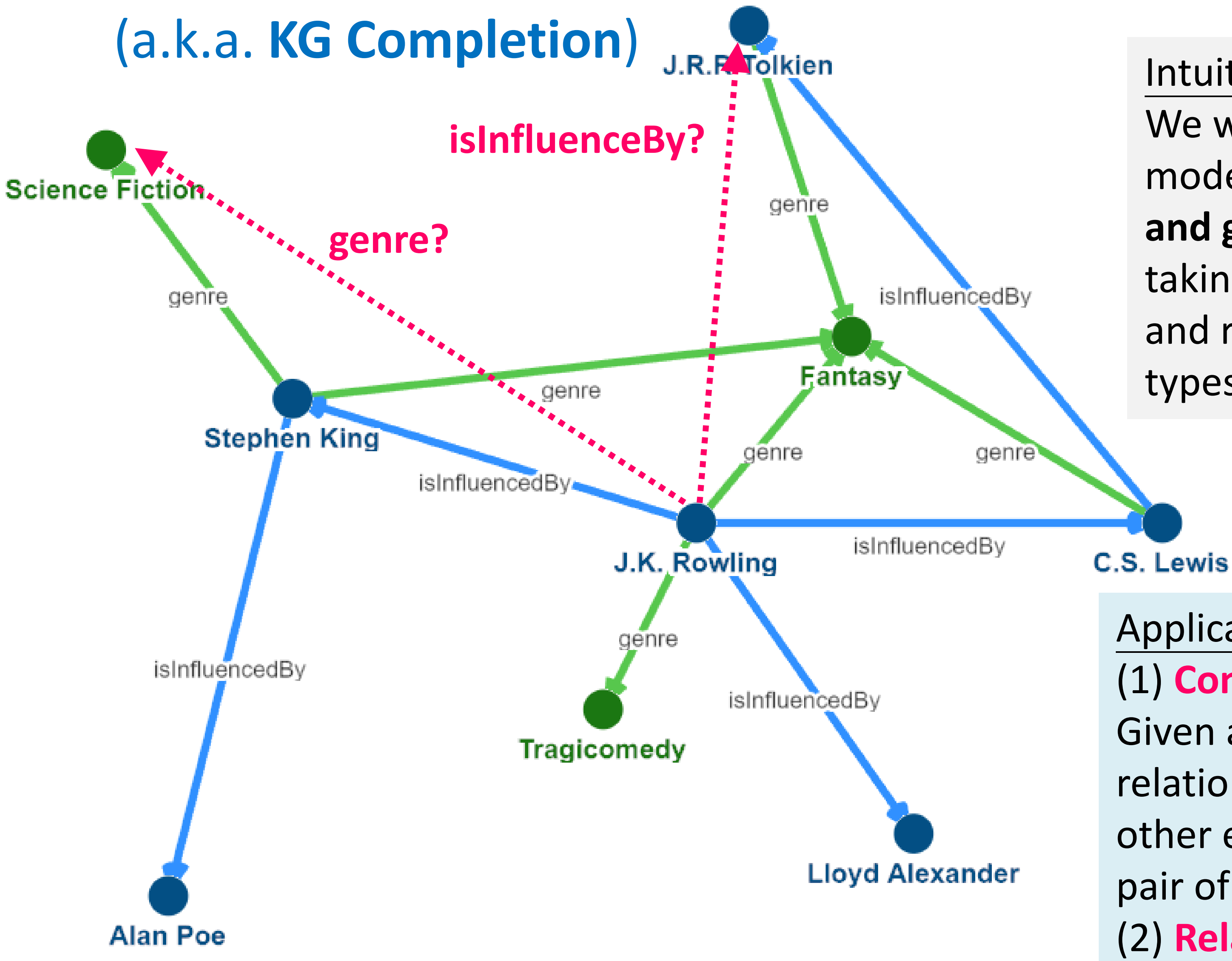
Head Relation Tail

(h) Subject (r) Object (t)



Link Prediction on Knowledge Graphs

(a.k.a. **KG Completion**)



Intuition

We want a **link prediction** model that learns from **local and global patterns** in the KG, taking into account entities and relationships of different types at the same time!

Application Tasks

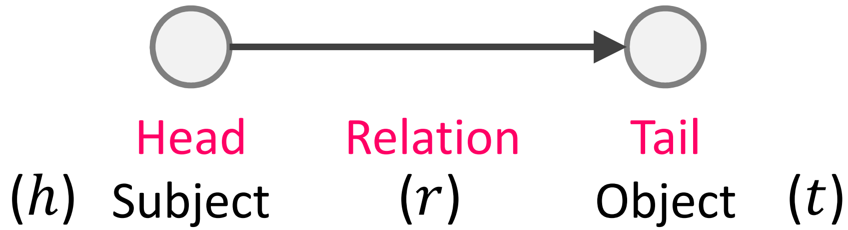
(1) **Conditional Link Prediction**

Given an subject entity and a relation, predict which of the other entity satisfies such a pair of subject and relation

(2) **Relation Prediction**

Given a pair of entities (sub. & obj.), predict their relation

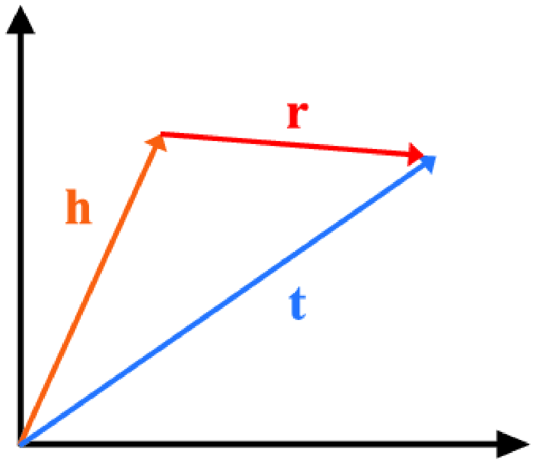
Relation Patterns

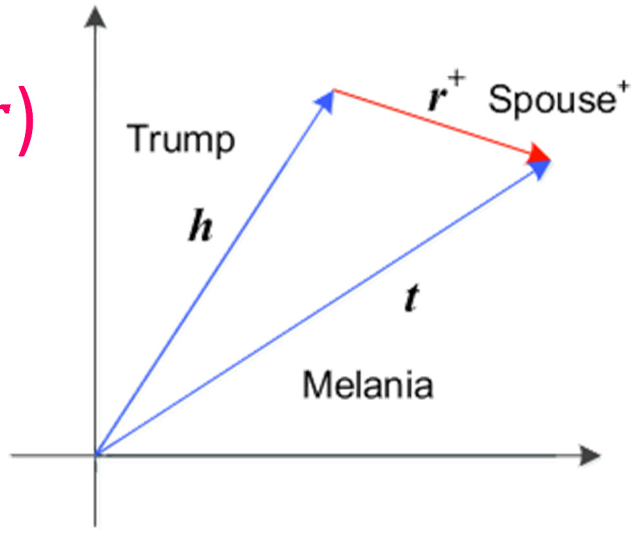


- **Symmetric** Relations
  - $r(h,t) \Leftrightarrow r(t,h)\ \ \forall h,t$
  - E.g., family, roommate
- **Composition** Relations
  - $r_1(x,y) \wedge r_2(y,z) \Rightarrow r_3(x,z)\ \ \forall x,y,z$
  - E.g., my mother's husband is my father
- **1-to-N**, **N-to-1** Relations
  - $r(h,t_1), r(h,t_2), \dots, r(h,t_n)$ are all TRUE
  - E.g., r is "StudentsOf"

KG Representation Learning

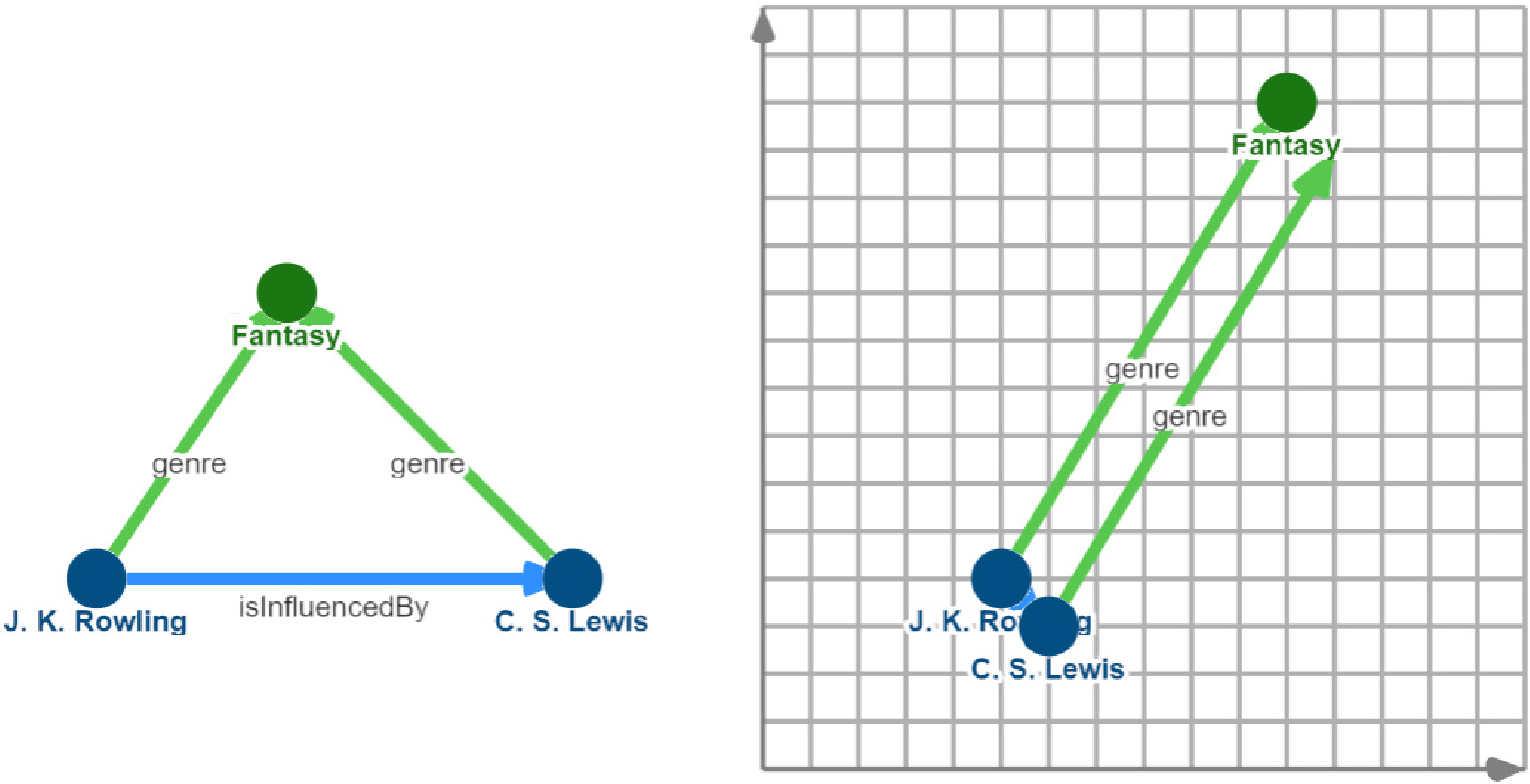
- Edges in KG are represented as triples (h, r, t)
  - Head (h) has relation (r) with tail (t)
- Main idea
  - Model entities and relations in the embedding/vector space $\mathbb{R}^d$
  - Given a true triple (h, r, t), the goal is that the embedding of (h, r) **<u>should be close to</u>** embedding of t
    - Q1: How to embed (h, r)?
    - Q2: How to define closeness?





Trump + Spouse+ = Melania

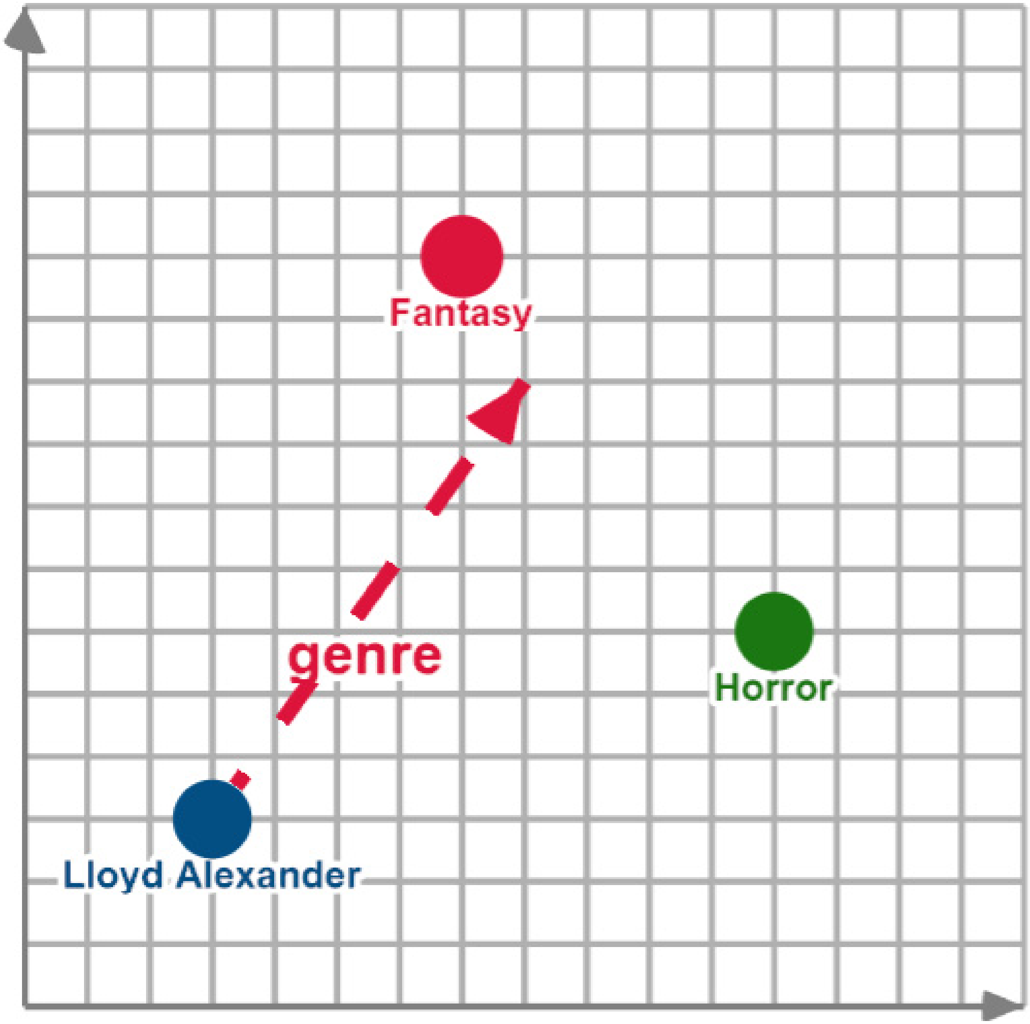
Illustration of KG Representation Learning



http://pyvandenbussche.info/2017/translating-embeddings-transe/

Applying TransE to QA

Q: What is the literature genre of "Lloyd Alexander"?



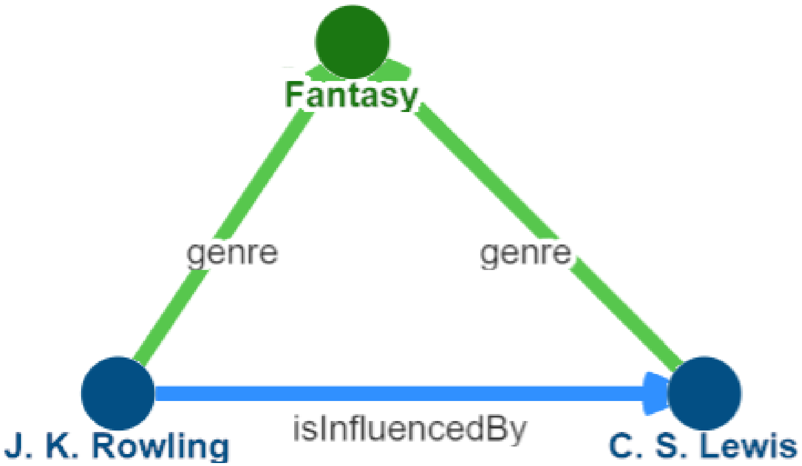
http://pyvandenbussche.info/2017/translating-embeddings-transe/

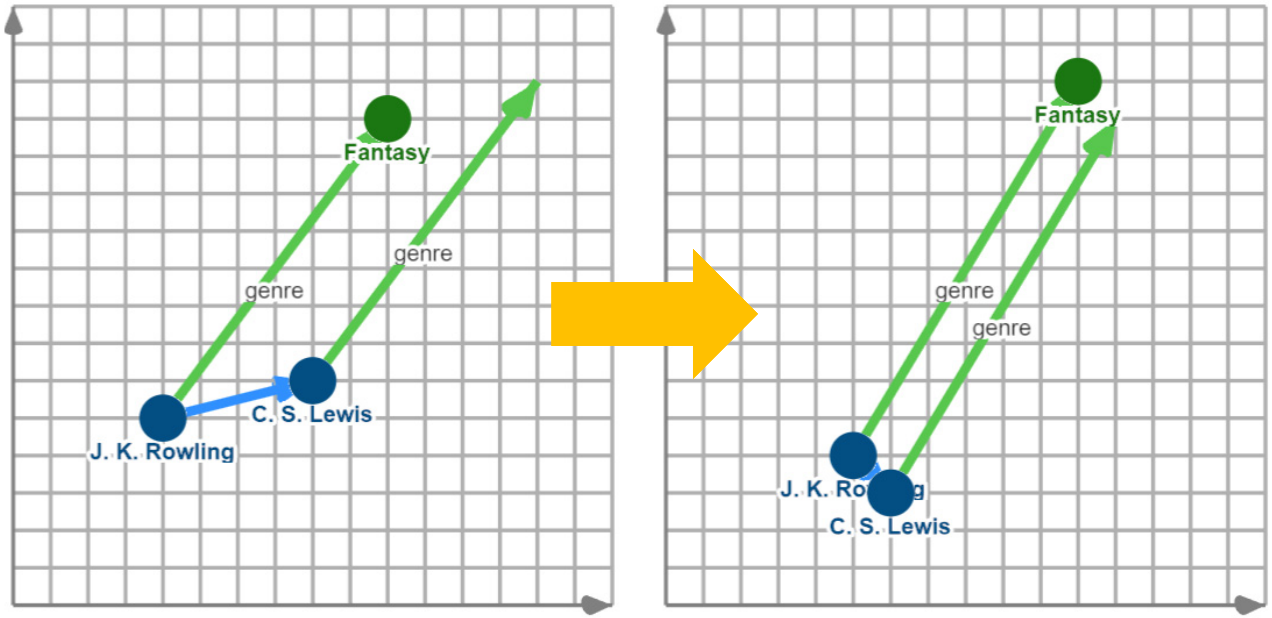
TransE: Modeling Relations as Translations

- **Translation**: given a triple (h, r, t), learn their embedding vectors $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$, such that: $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$
- Define the distance function:

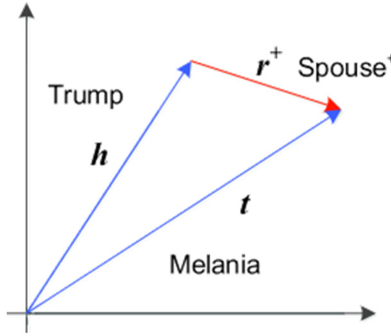
$$d(sub, rel, obj) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2^2$$

$\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ are learnable embedding vectors





Learning TransE



- **Translation**: given a triple (h, r, t), learn their embedding vectors $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^d$, such that: $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$
- Minimizing **the max-margin ranking loss**

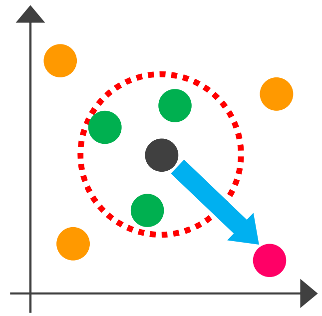
$$\mathcal{L} = \underbrace{\sum_{(h,\ell,t)\in S}}_{\text{positive samples}} \underbrace{\sum_{(h',\ell,t')\in S'_{(h,\ell,t)}}}_{\text{negative samples}} \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$$

$$\gamma + \|\mathbf{h} + \mathbf{l} - \mathbf{t}\|_2^2 - \|\mathbf{h} + \mathbf{l} - \mathbf{t}\|_2^2$$

$$S'_{(h,\ell,t)} = \{(h', \ell, t) | h' \in E\} \cup \{(h, \ell, t') | t' \in E\}$$

Positive triple (h, l, t):
(Trump, Spouse, Melania)

Negative triples:
(h', l, t) = (Obama, Spouse, Melania)
(h, l, t') = (Trump, Spouse, Michelle)



γ is the margin, i.e., the smallest distance tolerated by the model between a positive triple and a negative one

TransE Learning Algorithm

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b

$T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

for $(h, \ell, t) \in S_{batch}$ **do**

$(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

end for

Update embeddings w.r.t. $$\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla\big[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\big]_+$$

positive samples

negative samples

$$\gamma + \|\mathbf{h} + \mathbf{l} - \mathbf{t}\|_2^2 - \|\boldsymbol{h}' + \mathbf{l} - \boldsymbol{t}'\|_2^2$$

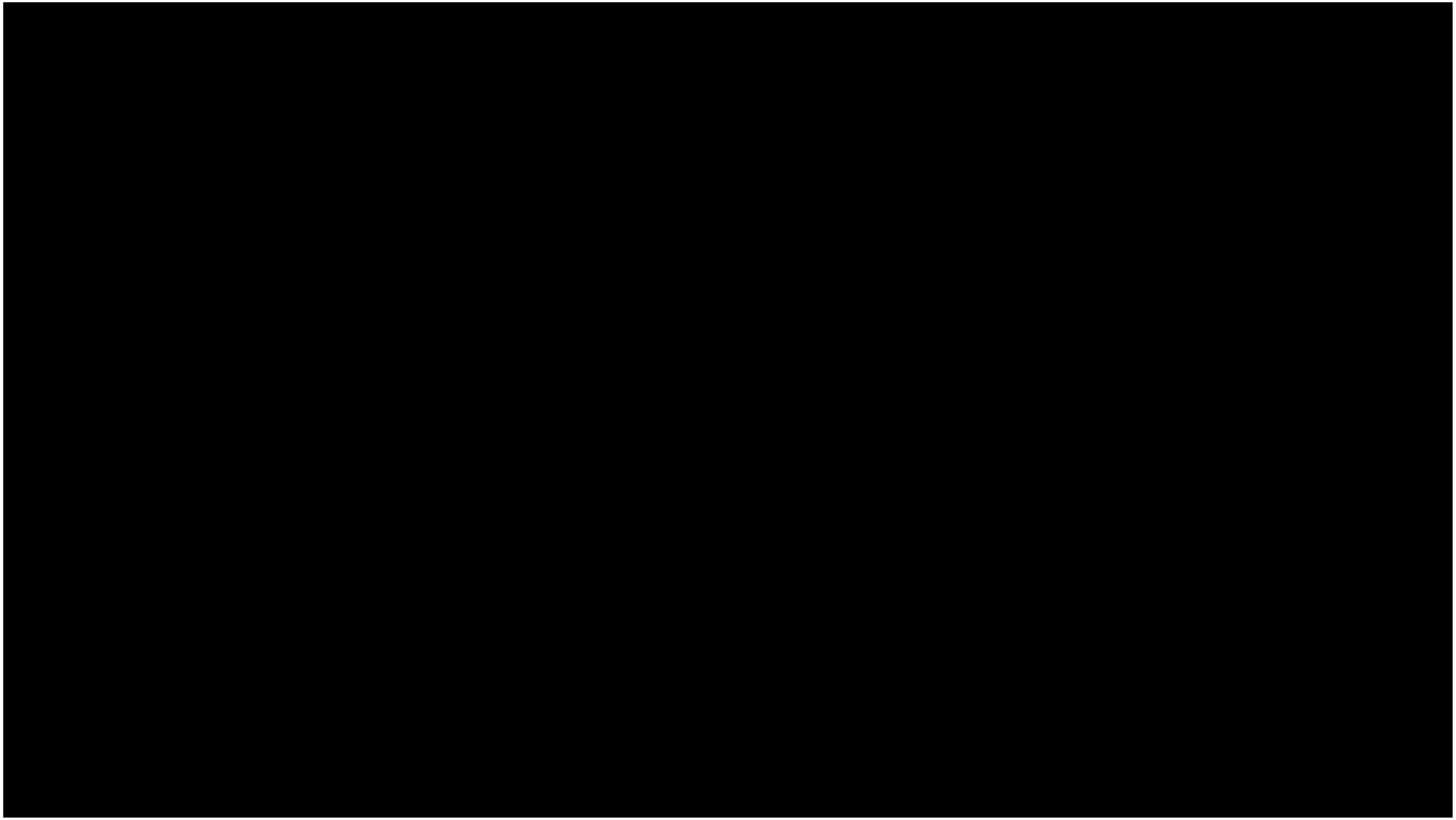
end loop

Entities and relations are initialized uniformly, and normalized

Negative sampling with triplet that does not appear in the KG

Comparative loss: favors lower distance values for valid triplets, high distance values for corrupted ones

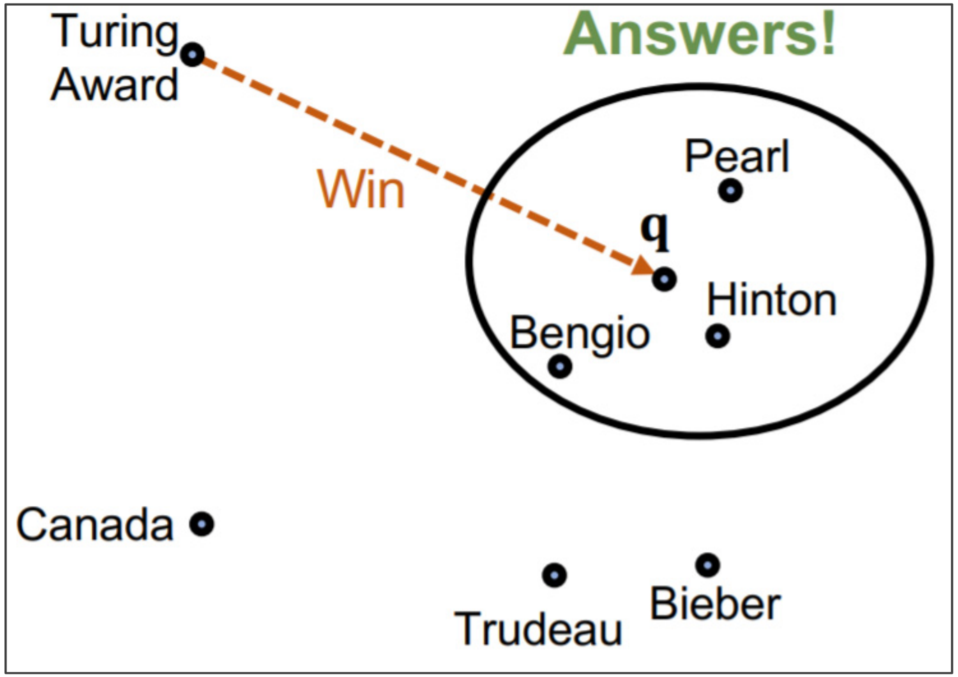
Process of TransE Training

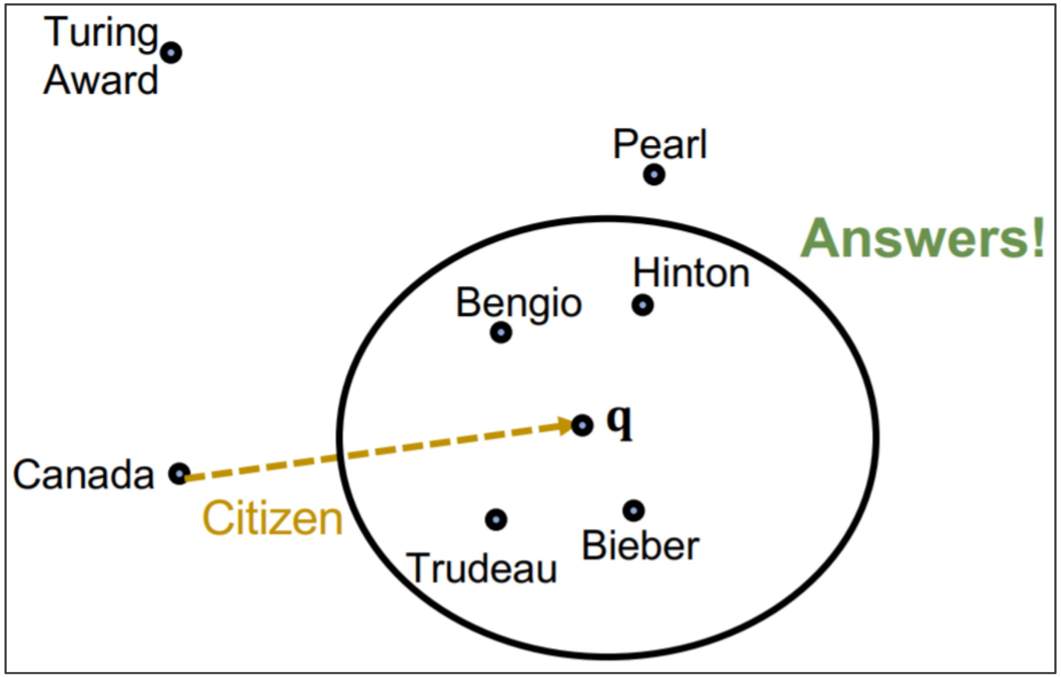


http://pyvandenbussche.info/2017/translating-embeddings-transe/

Link Prediction in KG using TransE

- Who has won the Turing award?
- Who is a Canadian citizen?

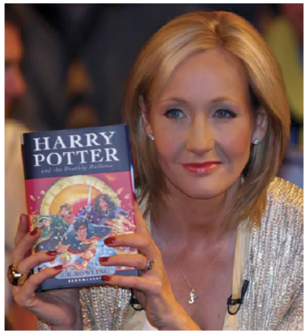




Applications on QA

"Who influenced by J. K. Rowling?"

J. K. Rowling _influenced_by G. K. Chesterton
J. R. R. Tolkien
C. S. Lewis
Lloyd Alexander
Terry Pratchett
Roald Dahl
Jorge Luis Borges
Stephen King
Ian Fleming



"Which genre is the movie WALL-E?"

WALL-E _has_genre Animation
Computer animation
Comedy film
Adventure film
Science Fiction
Fantasy
Stop motion
Satire
Drama

Resources on Knowledge Graphs

- Knowledge Graph Embeddings
  - https://github.com/thunlp/KB2E
  - https://github.com/thunlp/OpenKE
- TransE: Translating Embeddings for Modeling Multi-Relational Data (NIPS 2013)
  - Paper: https://papers.nips.cc/paper/5071-translating-embeddings-for-modeling-multi-relational-data.pdf
  - PyTorch: https://github.com/jimmywangheng/knowledge_representation_pytorch
- Knowledge Graph Embedding: A Survey of Approaches and Applications
  - https://persagen.com/files/misc/Wang2017Knowledge.pdf
- A Survey on Knowledge Graphs: Representation, Acquisition and Applications
  - https://arxiv.org/pdf/2002.00388.pdf