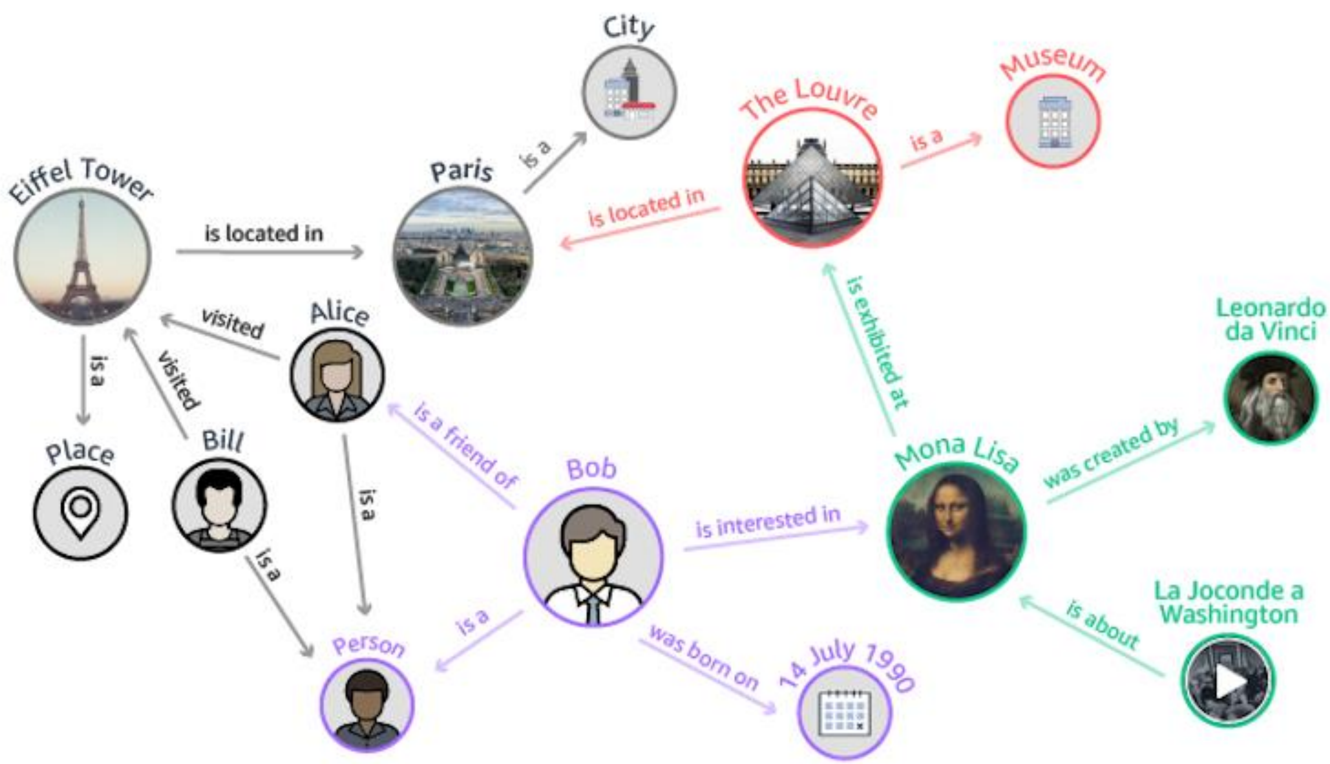
Injecting Prior Information and Multiple Modalities into Knowledge Base Embeddings

Sameer Singh

University of California, Irvine

Knowledge Graphs



What is a knowledge graph?

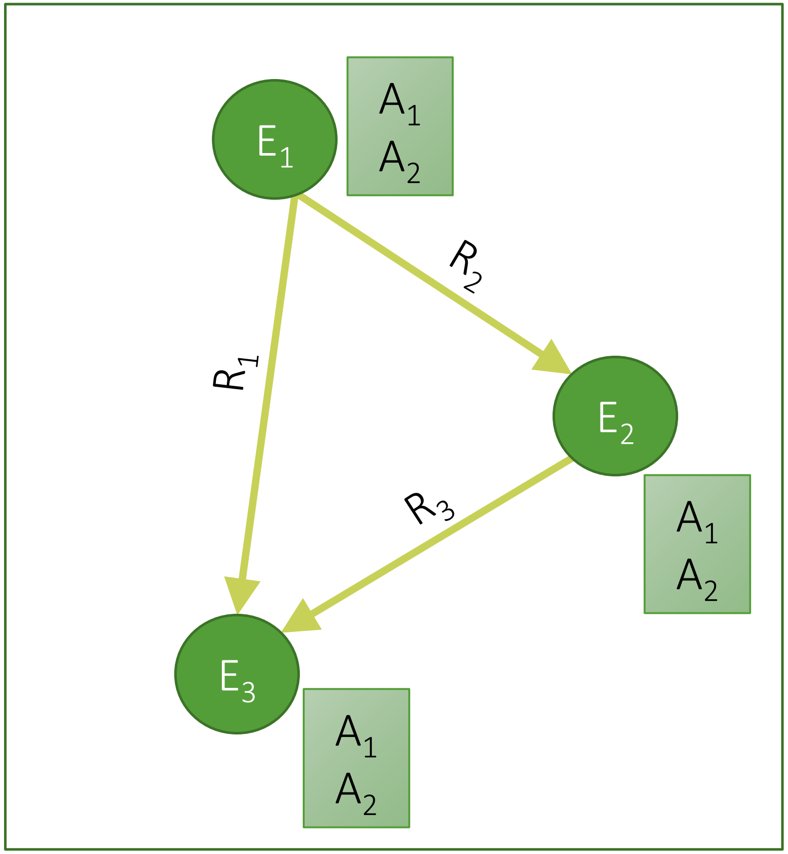
Knowledge in graph form!

Captures entities, attributes, and relationships

Nodes are entities

Nodes are labeled with attributes (e.g., types)

Typed edges between two nodes capture a relationship between entities



Why knowledge graphs?

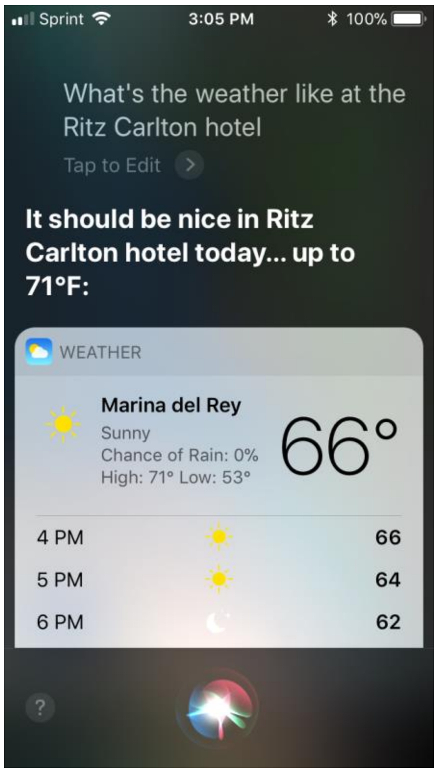
Humans:

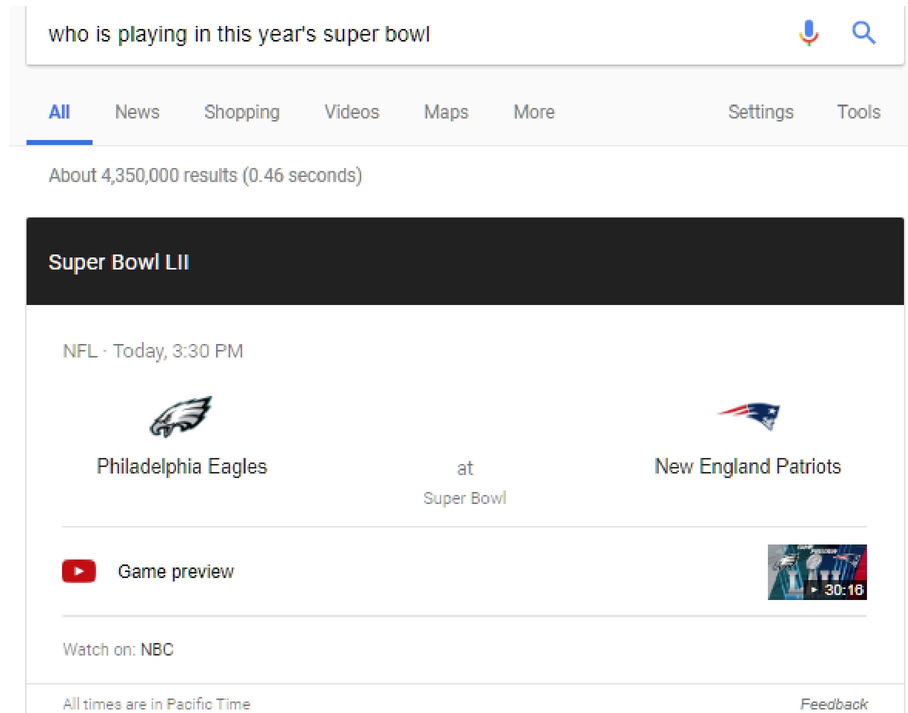
Combat information overload
Explore via intuitive structure
Tool for supporting knowledge-driven tasks

AIs:

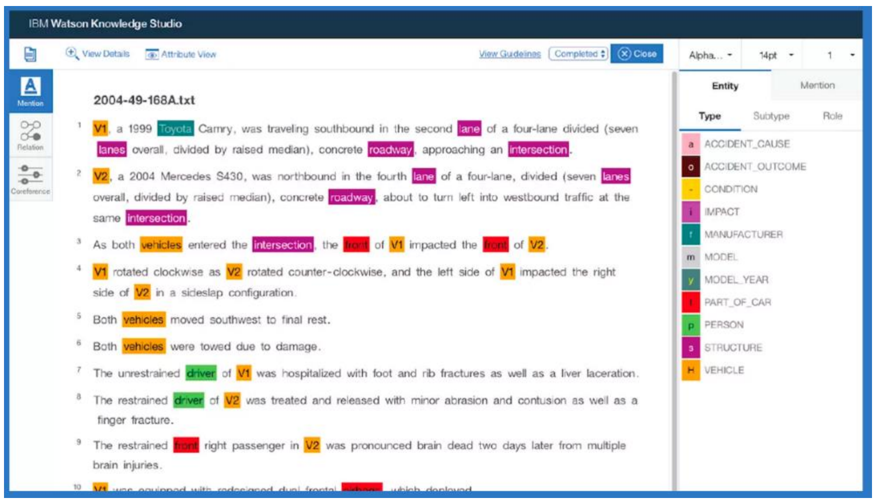
Key ingredient for many AI tasks
Bridge from data to human semantics
Use decades of work on graph analysis

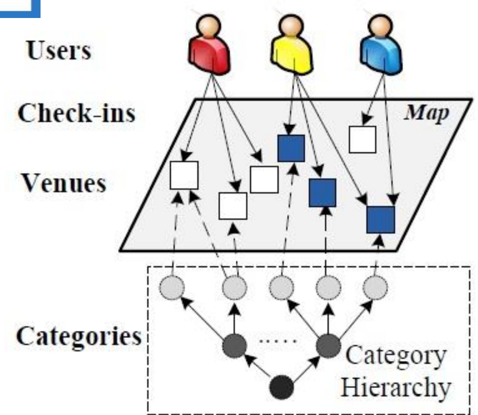
Applications 1: QA/Agents





Applications 2: Decision Support





Applications 3: Fueling Discovery

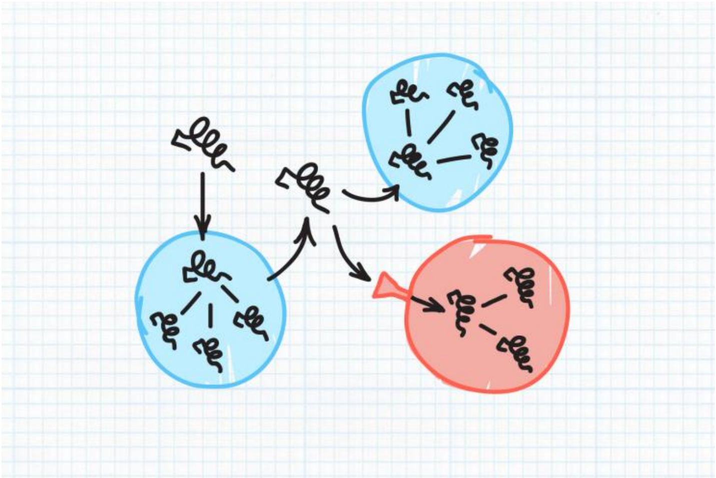
beatles (musicartist)

literal strings: BEATLES, Beatles, beatles

Help NELL Learn!

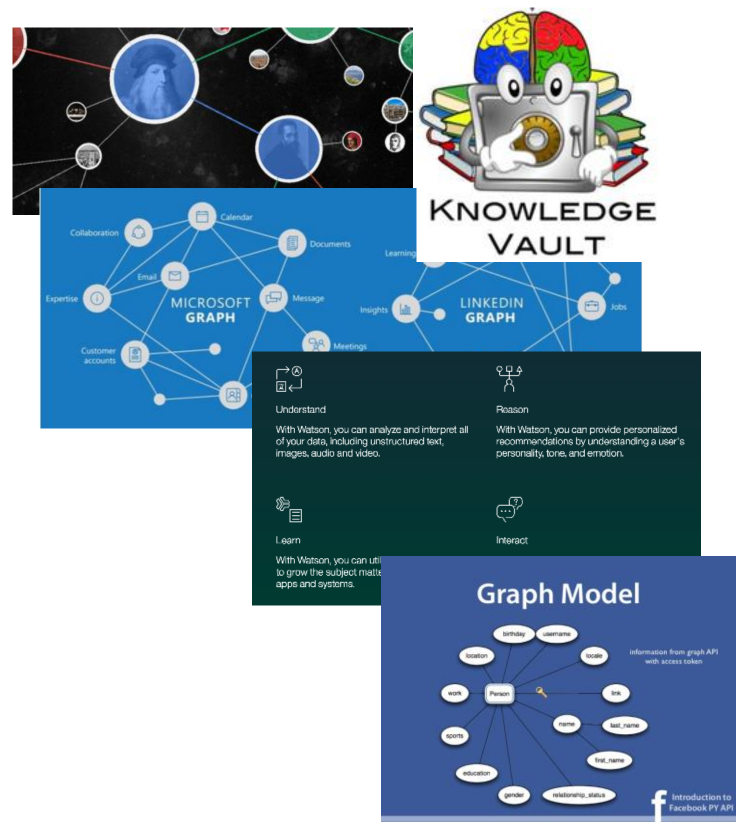
NELL wants to know if these be
If they are or ever were, click thumbs-up. O

- beatles is a musical artist
- beatles is a musician in the genre classic_pop (musicgenre)
- beatles is a musician in the genre pop (musicgenre)
- beatles is a musician in the genre rock (musicgenre)
- beatles is a musician in the genre classic_rock (musicgenre)



Knowledge Graphs & Industry

- Google Knowledge Graph
    - Google Knowledge Vault
- Amazon Product Graph
- Facebook Graph API
- IBM Watson
- Microsoft Satori
    - Project Hanover/Literome
- LinkedIn Knowledge Graph
- Yandex Object Answer
- Diffbot, GraphIQ, Maana, ParseHub, Reactor Labs, SpazioDati

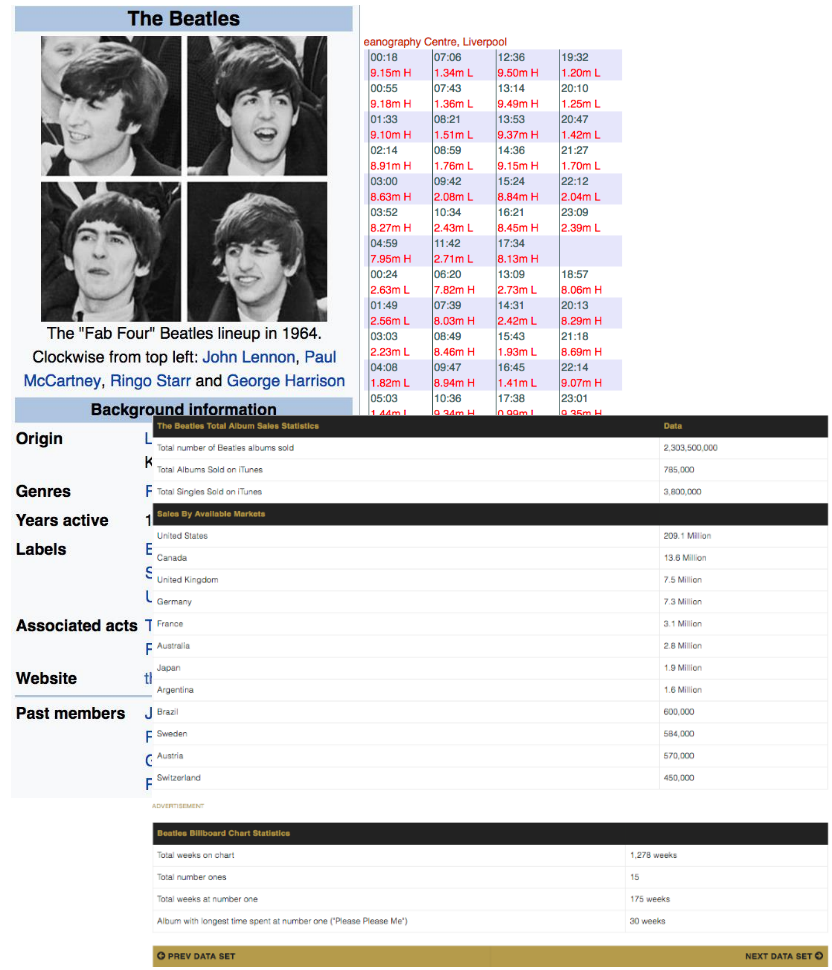


Where do knowledge graphs come from?

Where do knowledge graphs come from?

Structured Text

Wikipedia Infoboxes, tables, databases, social nets



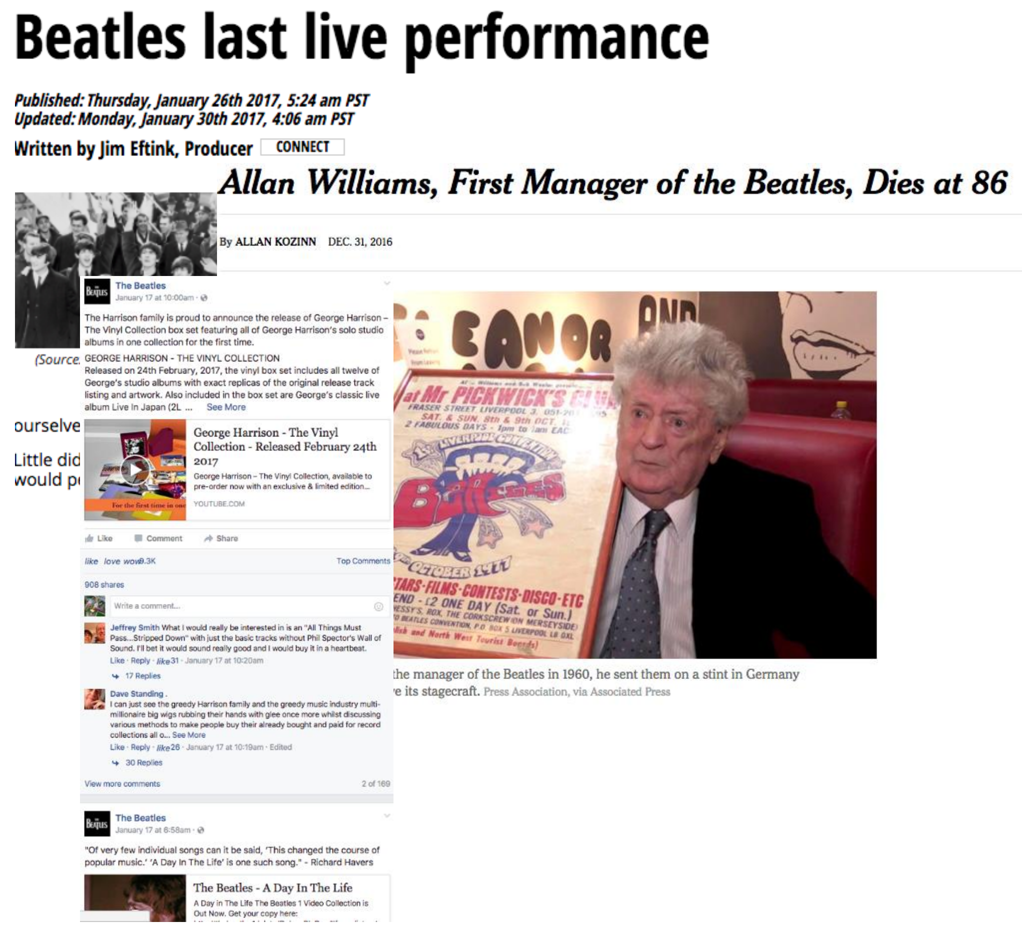
Where do knowledge graphs come from?

Structured Text

Wikipedia Infoboxes, tables, databases, social nets

Unstructured Text

WWW, news, social media, reference articles



Where do knowledge graphs come from?

Structured Text

- Wikipedia Infoboxes, tables, databases, social nets

Unstructured Text

- WWW, news, social media, reference articles

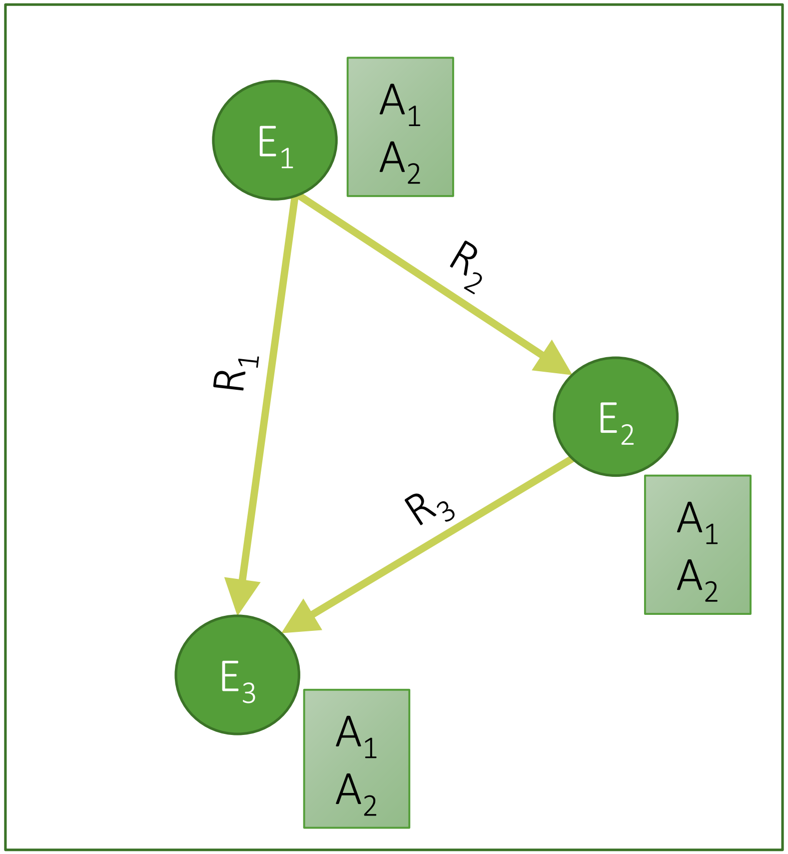
Images

Basic problems

Who are the entities (nodes) in the graph?

What are their attributes and types (labels)?

How are they related (edges)?



Outline

Knowledge Graph Embeddings

Injecting Prior Information

Injecting Multiple Modalities

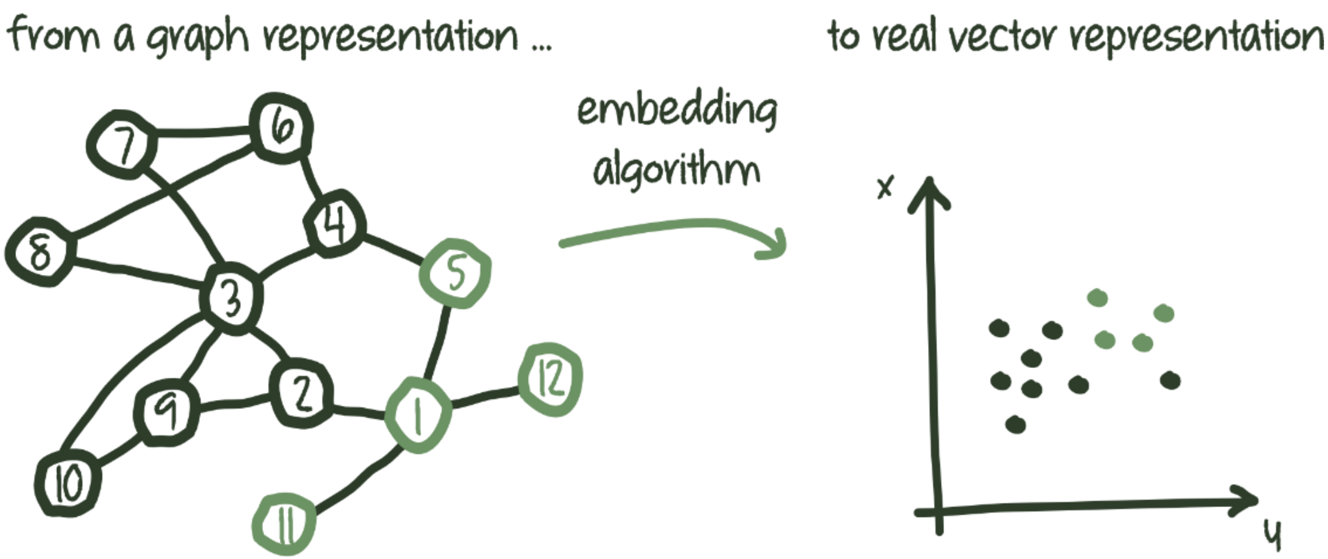
Outline

Knowledge Graph Embeddings

Injecting Prior Information

Injecting Multiple Modalities

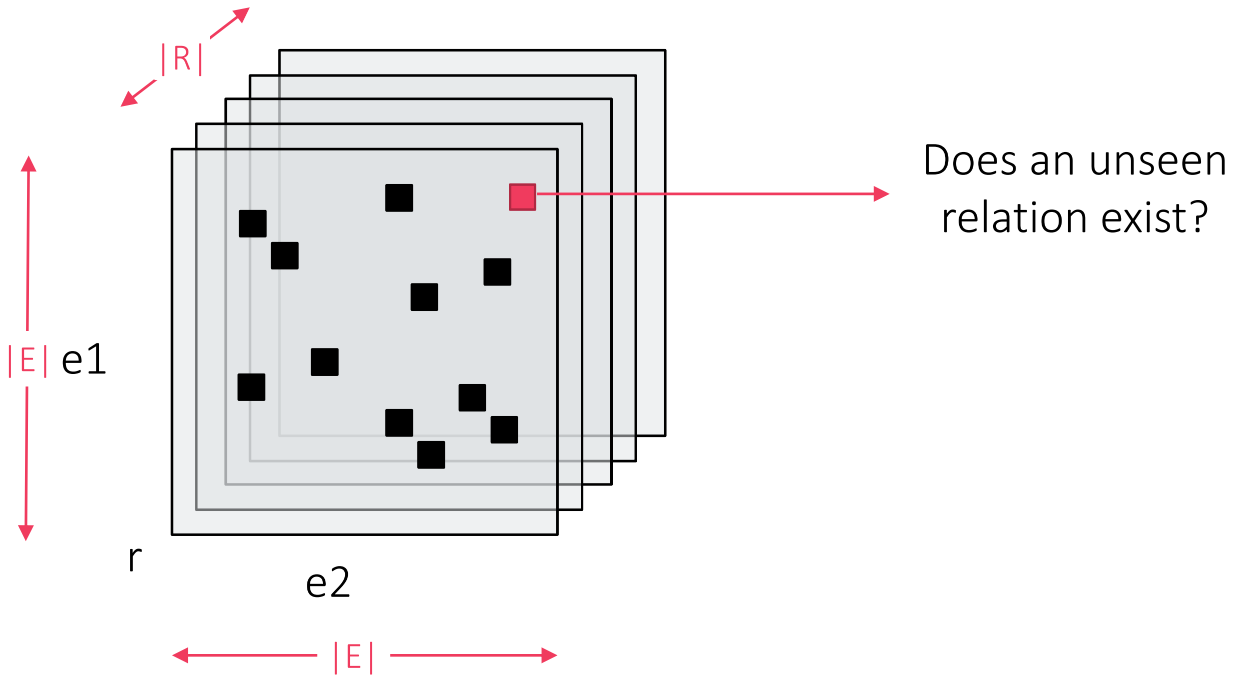
Graph Embeddings



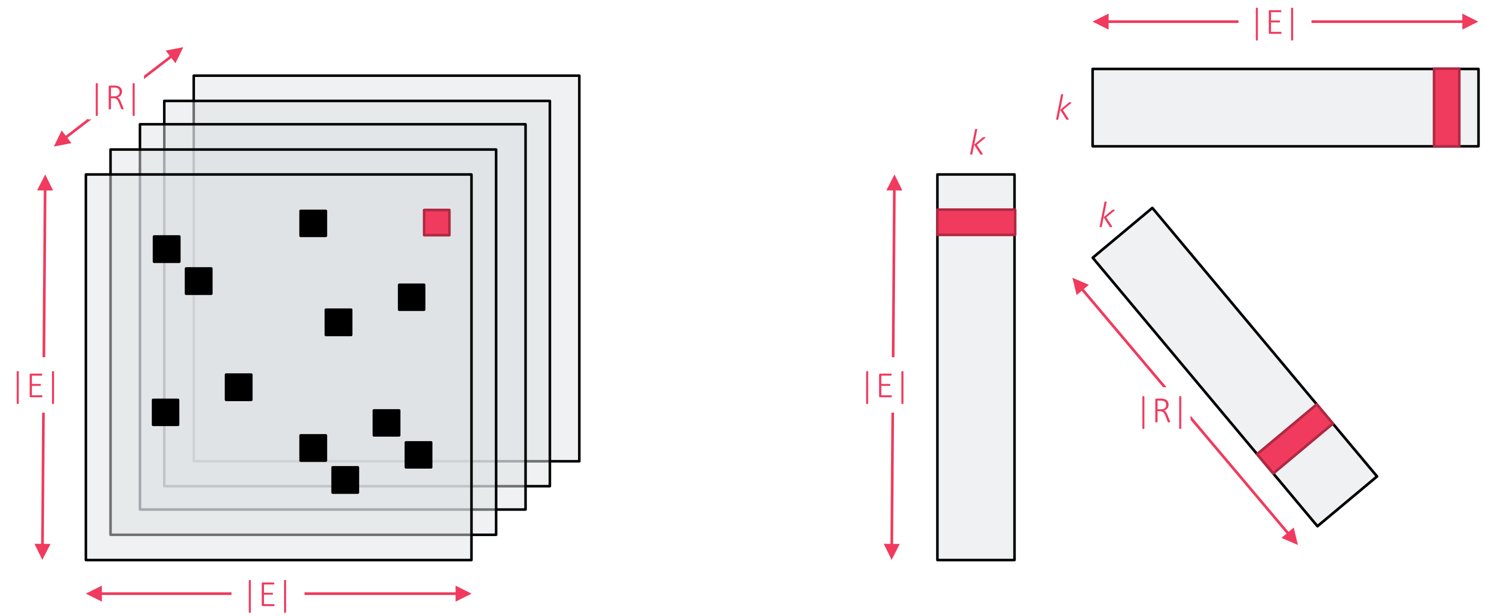
1. Encodes all the information about entities
2. Predict missing relations/facts
3. Clean up incorrect (inconsistent) information

Generate from
Graph Embeddings!

Tensor Formulation of KG



Factorize that Tensor



$$S(r(a,b)) = f(\mathbf{v}_r, \mathbf{v}_a, \mathbf{v}_b)$$

Knowledge Base Completion

Scoring Function

Model	Score $\psi_r(\mathbf{e}_s, \mathbf{e}_o)$
RESCAL [21]	$\mathbf{e}_s^T \mathbf{W}_r \mathbf{e}_o$
SE [3]	$\left\Vert \mathbf{W}_r^L \mathbf{e}_s - \mathbf{W}_r^R \mathbf{e}_o \right\Vert_p$
TransE [1]	$\Vert \mathbf{e}_s + \mathbf{r}_r - \mathbf{e}_o \Vert_p$
DistMult [34]	$\langle \mathbf{e}_s, \mathbf{r}_r, \mathbf{e}_o \rangle$
ComplEx [33]	$\langle \mathbf{e}_s, \mathbf{r}_r, \mathbf{e}_o \rangle$
ConvE	$f\left(\mathrm{vec}\left(f\left([\overline{\mathbf{e}_s}; \overline{\mathbf{r}_r}] * \omega\right)\right) \mathbf{W}\right) \mathbf{e}_o$

Table from Dettmers, et al. (2017)

Many Different Factorizations

CANDECOMP/PARAFAC-Decomposition

$$S\left(r(a,b)\right) = \sum_k R_{r,k} \cdot e_{a,k} \cdot e_{b,k}$$

Tucker2 and RESCAL Decompositions

$$S\left(r(a,b)\right) = \left(\mathbf{R}_r \times \mathbf{e}_a\right) \times \mathbf{e}_b$$

Model E

$$S\left(r(a,b)\right) = \mathbf{R}_{r,1} \cdot \mathbf{e}_a + \mathbf{R}_{r,2} \cdot \mathbf{e}_b$$

Holographic Embeddings

$$S(r(a,b)) = \mathbf{R}_r \times \left(\mathbf{e}_a \star \mathbf{e}_b\right)$$

Not tensor factorization (per se)

HOLE: Nickel et al, AAAI (2016), Model E: Riedel et al, NAACL (2013), RESCAL: Nickel et al, WWW (2012), CP: Harshman (1970), Tucker2: Tucker (1966)

Graph Completion

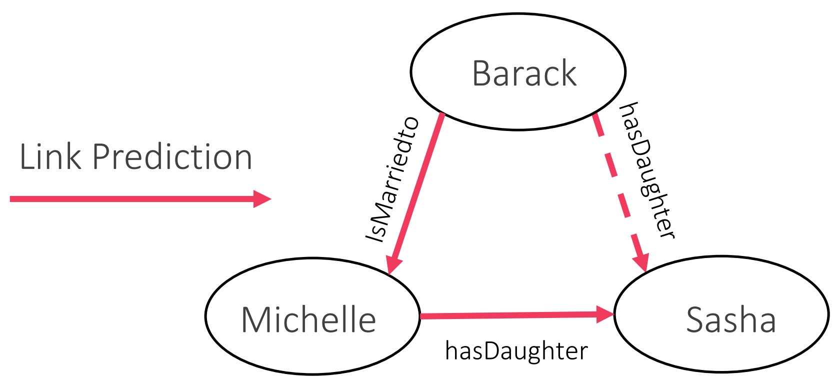
Entity Prediction

Link Prediction

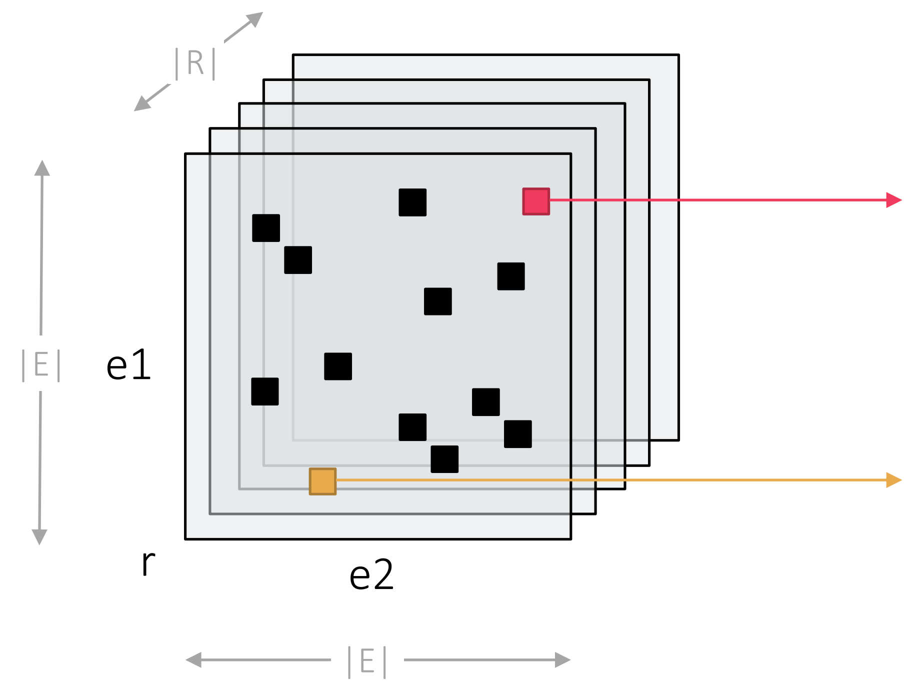
<Barack, IsMarriedto, Michelle>

<Barack, hasDaughter, Sasha>

< Michelle, hasDaughter, ?>



Parameter Estimation



Observed cell: increase score

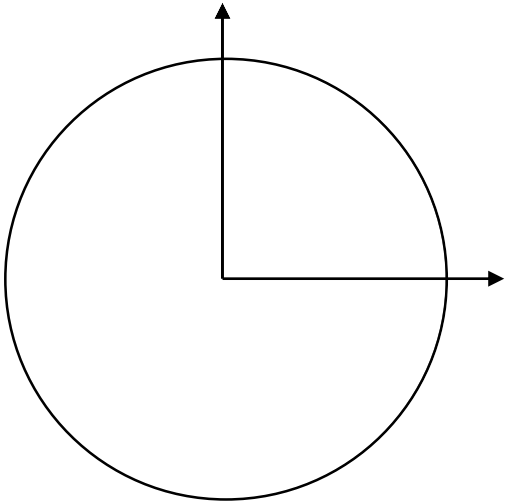
$$S\left(r(a,b)\right)$$

Unobserved cell: decrease score

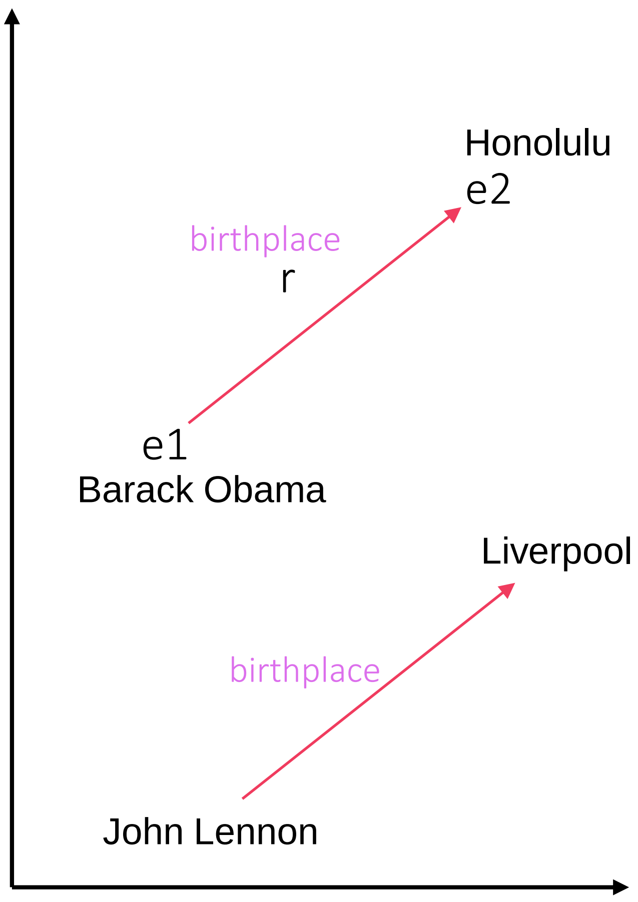
$$S\left(r'(x,y)\right)$$

Why do they work?

How can they remember one spouse from million possible ones?



Translation Embeddings



TransE

$$S\left(r(a,b)\right) = -\|\mathbf{e}_a + \mathbf{R}_r - \mathbf{e}_b\|_2^2$$

TransH

$$S\left(r(a,b)\right) = -\|\mathbf{e}_a^{\perp} + \mathbf{R}_r - \mathbf{e}_b^{\perp}\|_2^2$$

$$\mathbf{e}_a^{\perp} = \mathbf{e}_a - \mathbf{w}_r^T \mathbf{e}_a \mathbf{w}_r$$

TransR

$$S\left(r(a,b)\right) = -\|\mathbf{e}_a \mathbf{M}_r + \mathbf{R}_r - \mathbf{e}_b \mathbf{M}_r\|_2^2$$

TransE: Bordes et al. XXX (2011), TransH: Bordes et al. XXX (2011), TransR: Bordes et al. XXX (2011)

Outline

Knowledge Graph Embeddings

Injecting Prior Information

Injecting Multiple Modalities

Injecting Knowledge

Most people are married to one person.
“is native to” is same as birthplace relation.

I don’t understand. Give me labeled data.

Link Predictor

Sigh… okay.
Barack, spouseOf, Michelle ✓
Barack, spouseOf, Ann Dunham ✗

How can we make it easy for users to inject prior knowledge?

Logical Statements as Supervision

If you see "was a native of", it means birthplace

X was native of Y => birthplace(X,Y)

If a founder of the company is employed by the company, he's the CEO

X is the founder of Y $\wedge$ employee(X,Y) => ceoOf(X,Y)

Everyone is married to at most one person

spouse(X, Y) => $\forall$Y' $\neg$spouse(X,Y')

Logic Representation of Relations

Relations are binary predicates

$$\text{bornIn}(a, b) = \top \text{ or } \bot$$
$$\text{was-born-in}(a, b) = \top \text{ or } \bot$$
$$\text{where } a, b \in \{\text{“Bernie Sanders”}, \text{“Brooklyn”}, \text{“Michelle Obama”}, \ldots\}$$

Facts/Triples are ground atoms:

$$\mathcal{F} = \begin{cases} \text{bornIn}(\text{Bernie Sanders,Brooklyn}) \\ \text{was-born-in}(\text{Bernie Sanders,Brooklyn}) \\ \text{spouse}(\text{Barack Obama,Michelle Obama}) \\ \vdots \end{cases}$$

Models maximize the probability of ground atoms

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_{f \in \mathcal{F}} \log \mathcal{P}_\theta(f)$$

Model's belief in a formula f

For facts, we know this belief: $\mathcal{P}_\theta(f)$

Otherwise, recurse…

Can be any model!

$$\mathcal{P}_\theta(f) = \begin{cases} R(a,b) & \text{then} \quad \text{compute directly} \\ \neg f' & \text{then} \quad 1 - \mathcal{P}_\theta(f') \\ f_1 \wedge f_2 & \text{then} \quad \mathcal{P}_\theta(f_1)\mathcal{P}_\theta(f_2) \\ \forall_e f(e) & \text{then} \quad \prod_e \mathcal{P}_\theta(f(e)) \end{cases}$$

$$\mathcal{P}_\theta\left(\forall_{a,b}\ \text{was-born-in}(a,b) \Rightarrow \text{bornIn}(a,b)\right) =$$

$$\prod_{a,b} 1 - \mathcal{P}_\theta(\text{was-born-in}(a,b))\,(1 - \mathcal{P}_\theta(\text{bornIn}(a,b)))$$

$\text{was-born-in}(a,b)$ $\quad$ $\neg\text{bornIn}(a,b)$

$\text{was-born-in}(a,b) \Rightarrow \text{bornIn}(a,b)$

$\forall_{a,b}\ \text{was-born-in}(a,b) \Rightarrow \text{bornIn}(a,b)$

Updating the Embeddings

Our model is maximizing probability of ground atoms

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_{f \in \mathcal{F}} \log \mathcal{P}_\theta(f)$$

But now we have a set of formulae, ground or otherwise

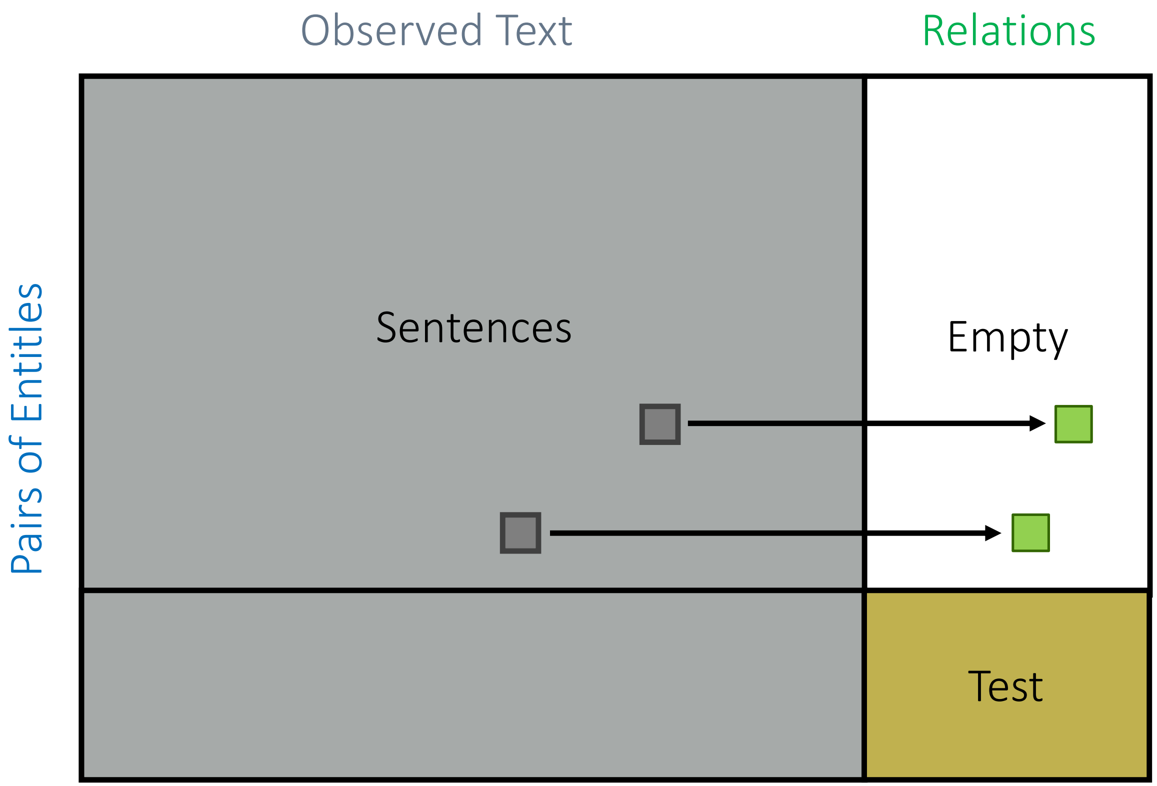
$$\mathcal{F} = \begin{cases} \text{bornIn}(\text{Bernie Sanders}, \text{Brooklyn}) \\ \text{was-born-in}(\text{Bernie Sanders}, \text{Brooklyn}) \\ \text{spouse}(\text{Barack Obama}, \text{Michelle Obama}) \\ \forall_{a,b}\ \text{was-born-in}(a, b) \Rightarrow \text{bornIn}(a, b) \\ \vdots \end{cases}$$

Still maximizing the probability: $\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_{f \in \mathcal{F}} \log \mathcal{P}_\theta(f)$

Optimized using gradient descent

works for most models!

Zero-Shot Learning



~2 million docs

~400,000 Entity pairs

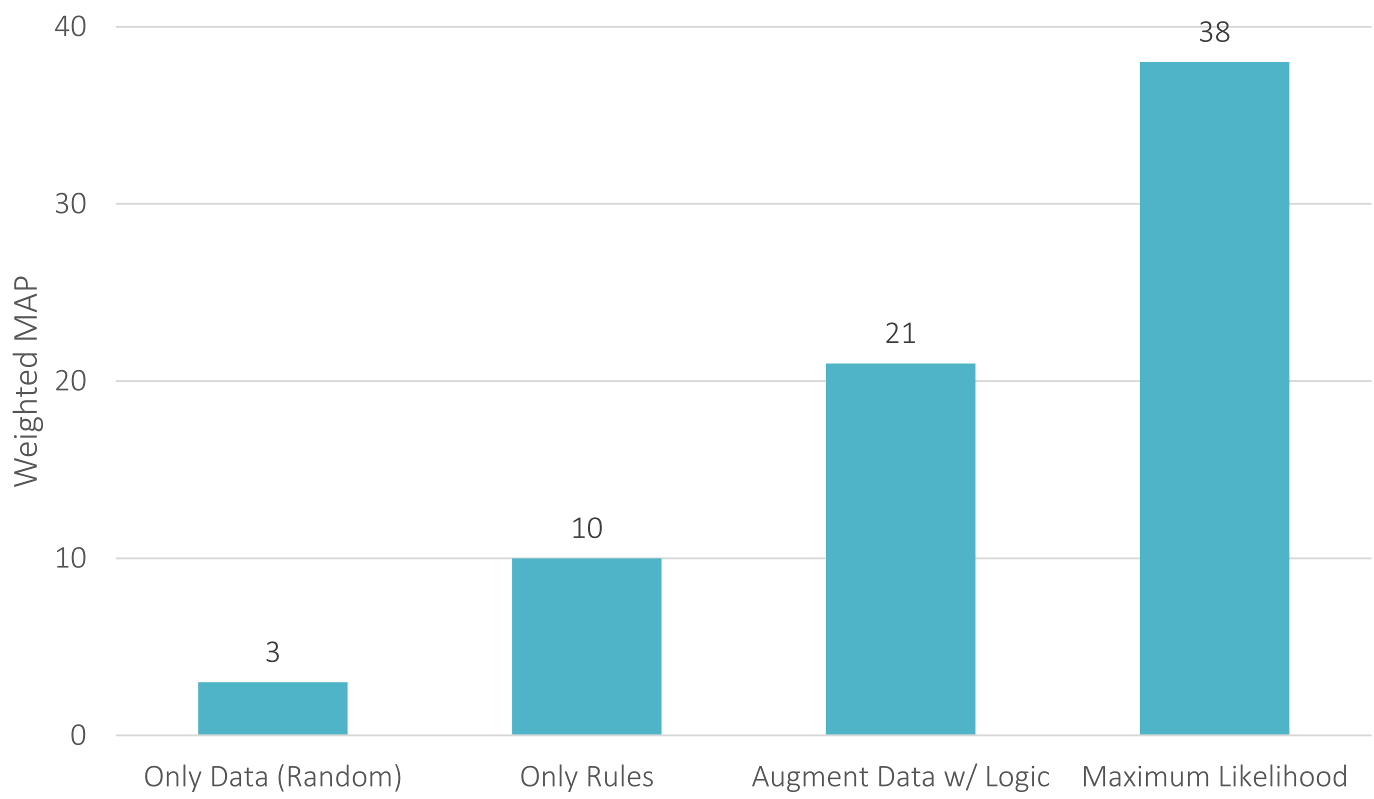
~4000 columns

~50 Relations of interest

30 Logical Implications

We're evaluating whether formulae can be used instead of labeled data.

Zero-Shot Learning



Outline

Knowledge Graph Embeddings

Injecting Prior Information

Injecting Multiple Modalities

Carles_Puyol
isAffiliatedTo Spain_national_under-18_football_team
isAffiliatedTo Spain_national_under-21_football_team
isAffiliatedTo Spain_national_under-23_football_team
isAffiliatedTo Catalonia_national_football_team

Carles_Puyol, isAffiliatedTo, ??

Spain_national_football_team Italy_national_football_team

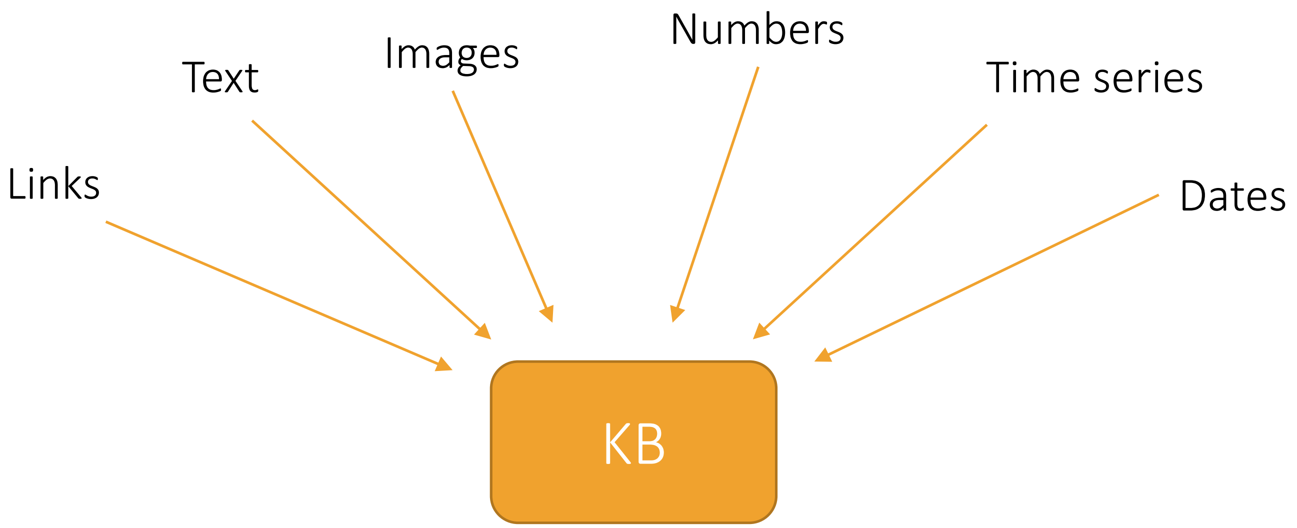
Carles_Puyol
isAffiliatedTo Spain_national_under-18_football_team
isAffiliatedTo Spain_national_under-21_football_team
isAffiliatedTo Spain_national_under-23_football_team
isAffiliatedTo Catalonia_national_football_team

Carles_Puyol, playsFor, ??

FC_Barcelona

Real_Madrid_CF

Information is in many modalities



Maybe we should be reasoning about all of these?

Time is ripe for doing multimodal stuff

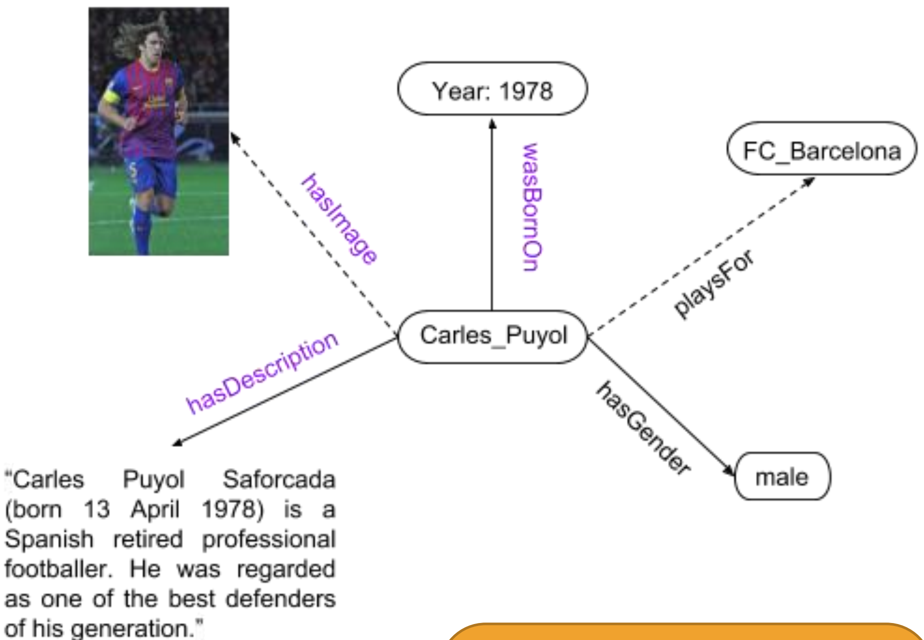
Link Prediction with Text/Images

Knowledge
Graphs

Text/Images

Entity and Relation
Embeddings

Multimodal Knowledge Graph



How do we get embeddings for these new relations and "objects"?

Entity

Text

Images

Numbers, etc.

Multimodal KB Embeddings

Scoring
Function

Object

Encoder

Everything else remains the same!

Multimodal KB Embeddings

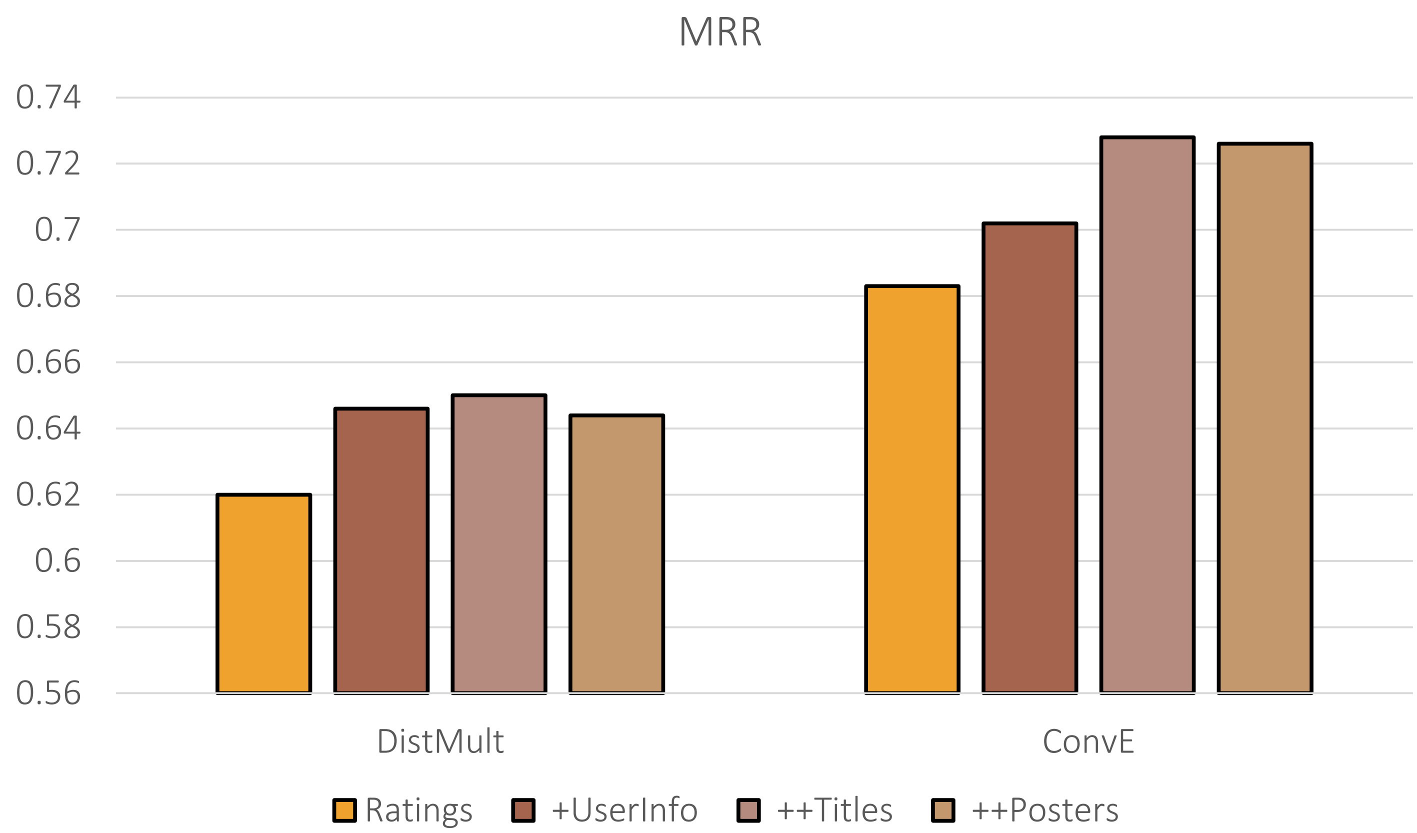
Object	Encoder
Entity	Lookup
Images	CNN
Text	LSTM
Numbers, etc.	FeedFwd

Augmenting Existing Graphs

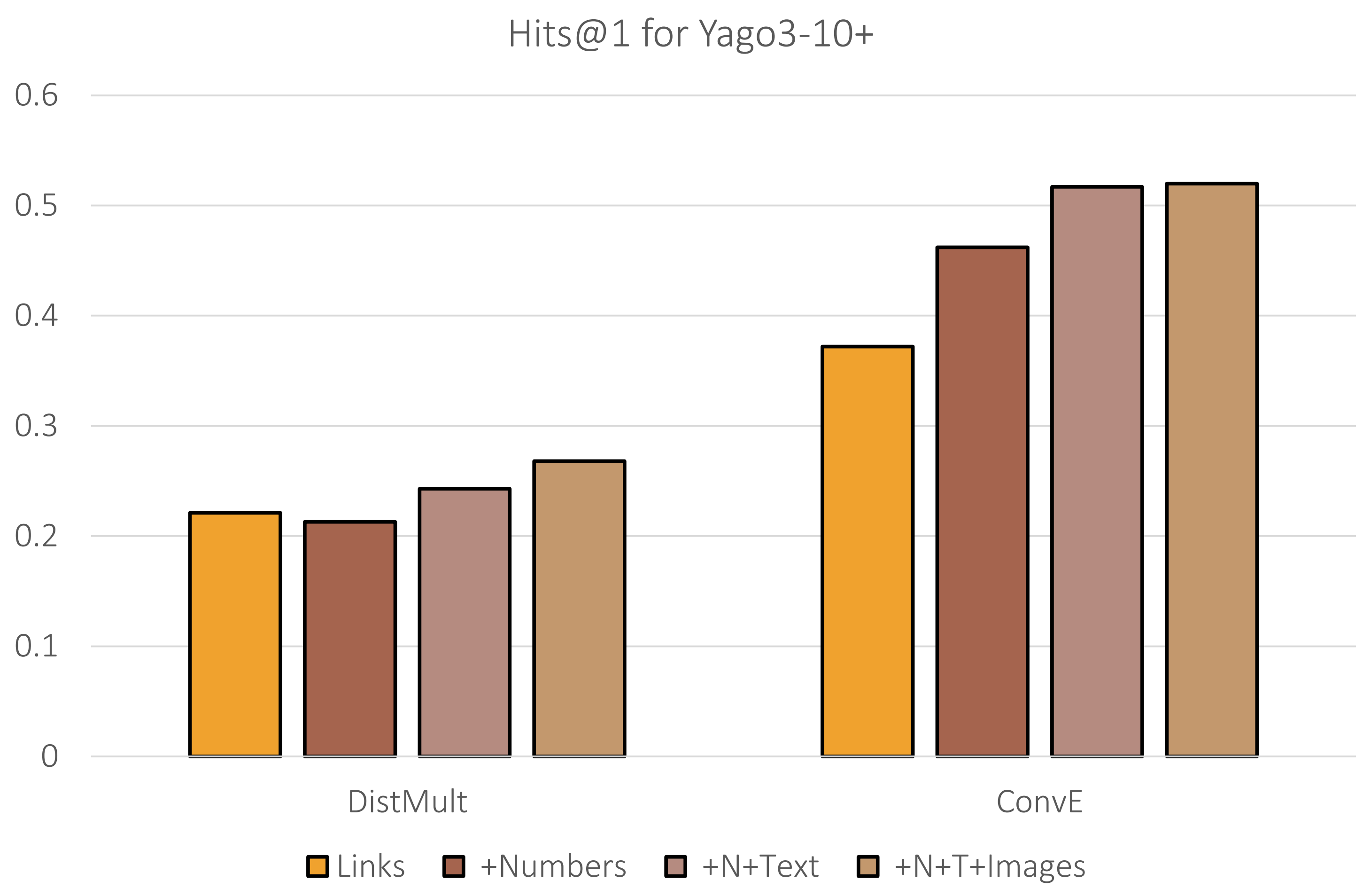
MovieLens-100k	
Relations	13
Users	943
Movies	1682
Posters	1651
Ratings	100,000

YAGO3-10	
Relations	37 → 45
Entities	123,182
Structure Triples	1,079,040
Numbers (Years)	1651
Descriptions	107,326
Images	61,246

MovieLens “Link Prediction”



YAGO Link Prediction Results



YAGO Relation Breakdown:

Relations	Links	+Numbers	+Text	+Images
isAffiliatedTo	0.401	0.467	**0.481**	0.478
playsFor	0.413	0.471	**0.486**	0.476
hasGender	0.596	0.599	0.627	**0.683**
isConnectedTo	0.367	0.379	**0.384**	0.372
isMarriedTo	0.207	0.221	0.296	**0.326**

Generation and Link Prediction

Knowledge
Graphs

Text/Images

Entity and Relation
Embeddings

Generating Multimodal Information

Neural Regressor	Numbers, etc.
Conditional Text GAN	Text
Conditional Image GAN	Images

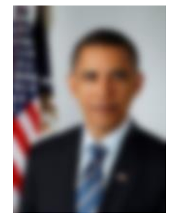
Conditional GAN Structure

Generator

Discriminator

Conditional GAN Structure

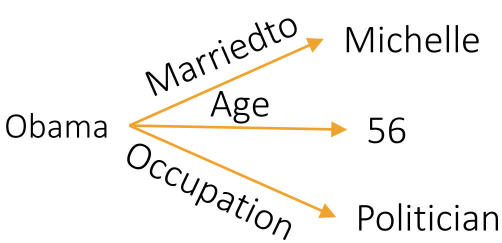
Generator



Barack Hussein Obama II is an American politician who served as the 44th President of the United States from 2009 to 2017.

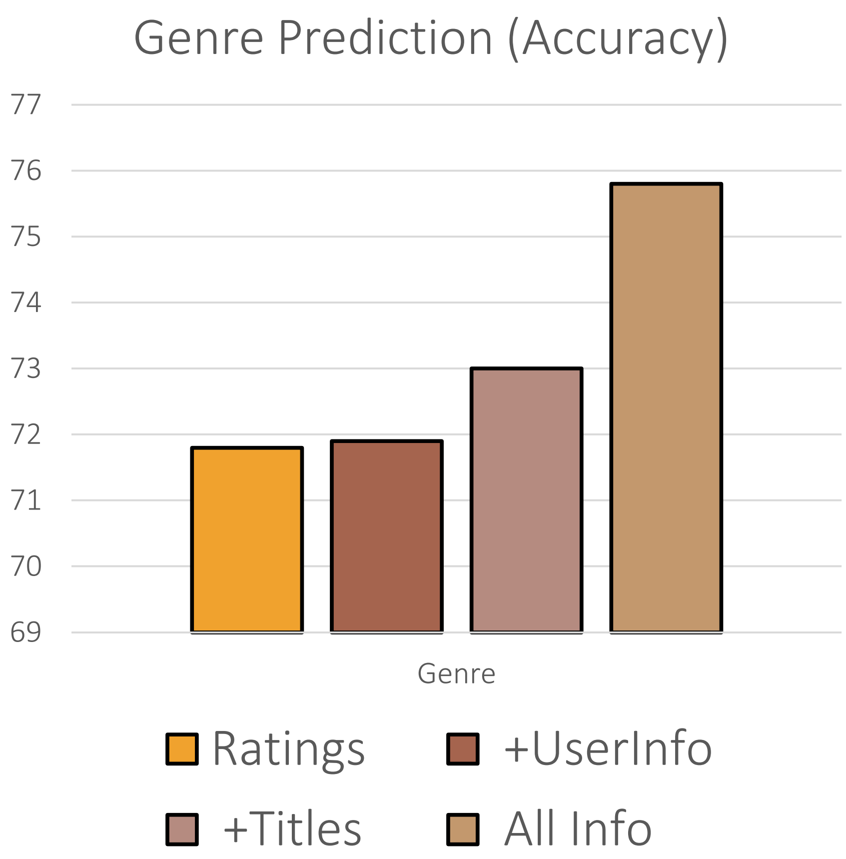
Barack Hussein Obama II is an American politician who served as the 44th President of the United States from 2009 to 2017.

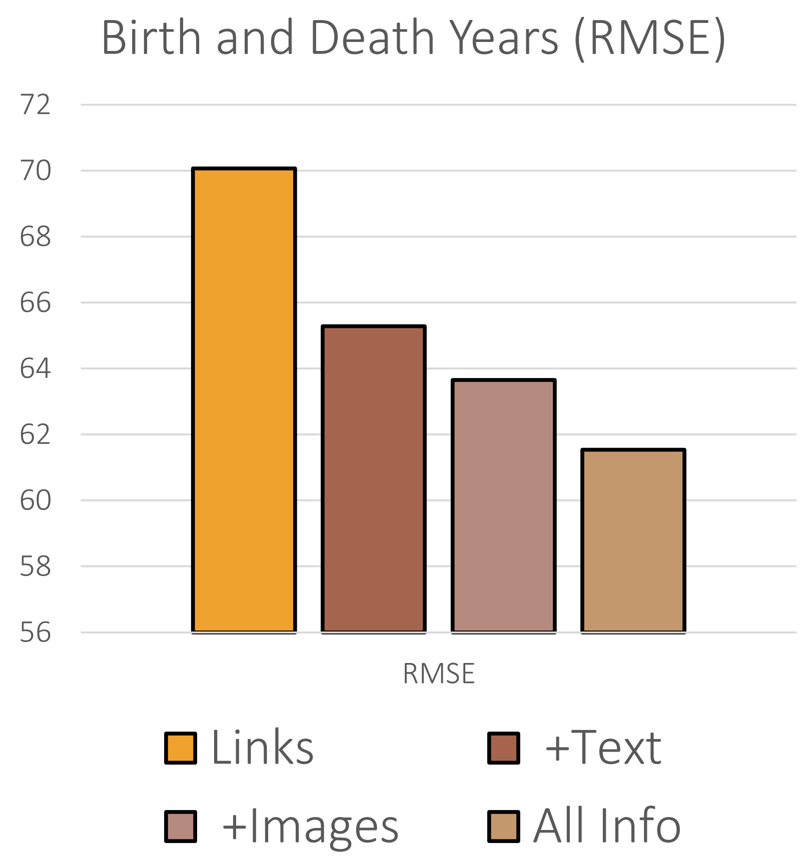
Obama



Discriminator

"Generating" Attributes





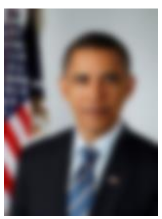
Generated Movie Titles

Reference	From Embeddings
Amityville 3-D (Horror)	Creatures
The Gay Divorcee (Romance/Musical)	Taste Condition
Jury Duty (Comedy)	Nixon World
Turbulence (Thriller)	Assignment
Mortal Kombat: Annihilation (Action/thriller)	The Cop Witness
Balto (Children's/Comedy)	Innocent Army
Jason's Lyric (Crime/Drama)	Wooden Beast

Generated Entity Descriptions

Reference	From Embeddings
Dean Sinclair (born 17 December 1984) is an English professional footballer who plays as a midfielder for Hampton & Richmond Borough.	Dean Sinclair (born 19 January 1981) is a professional footballer who plays as a left midfielder for <oov> in the England of England B.
Kelly LeBrock (born March 24, 1960) is an American actress and model.	Kelly LeBrock (born May 5, 1953) is an American composer music actress and singer.
The Lawnmower Man is a 1992 American science fiction action horror film directed by Brett Leonard and written by Brett Leonard and Gimel Everett.	The Lawnmower Man (born 10 October <oov> 1966) is a British science fiction and voice artist who had <oov> California.
Kungälv Municipality is a municipality in Västra Götaland County in western Sweden.	Kungälv Municipality is a city in Parish, Texas, Valley and Quebec. County.

What do people think?

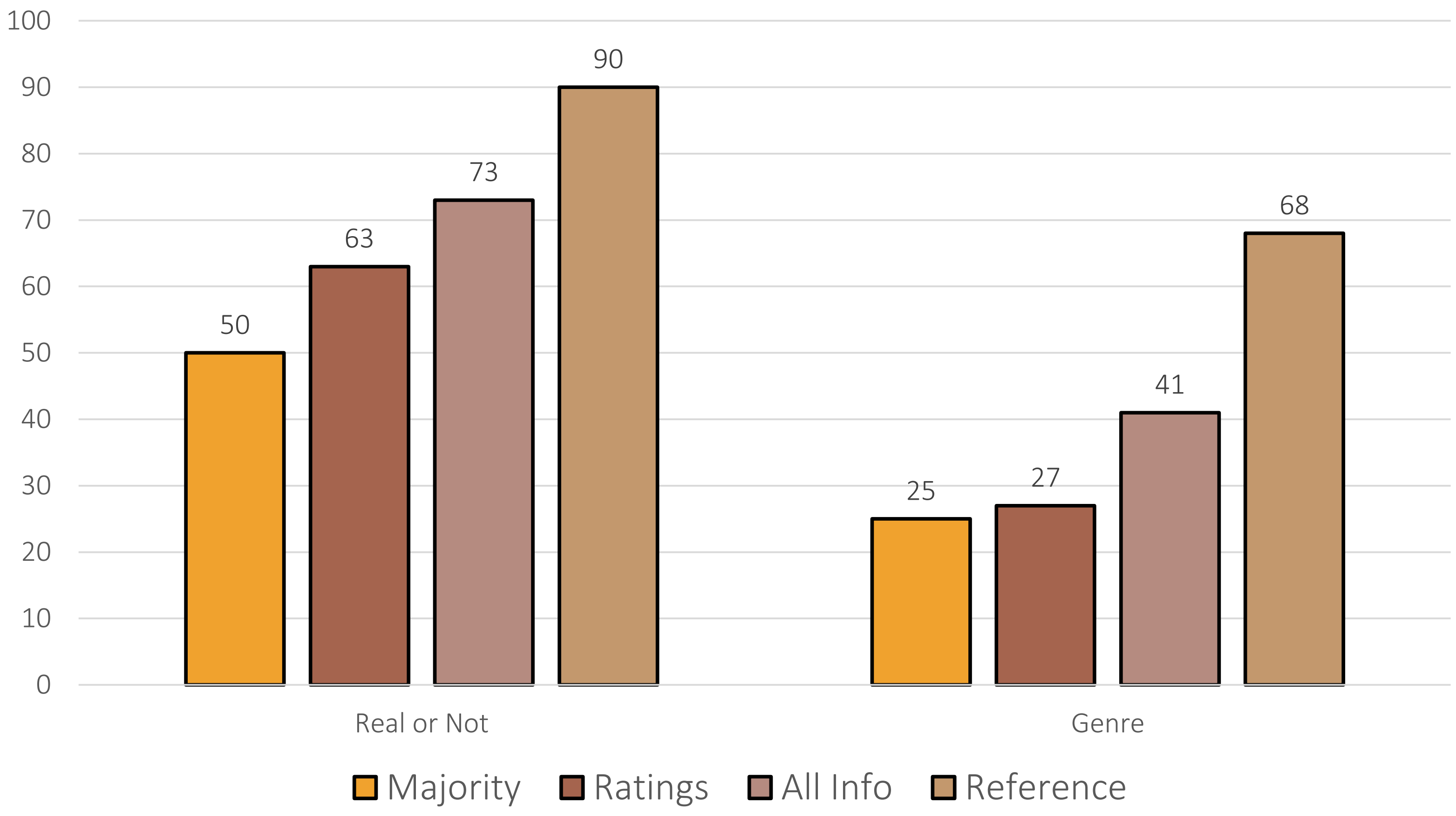


Barack Hussein Obama II is an American politician who served as the 44th President of the United States from 2009 to 2017.

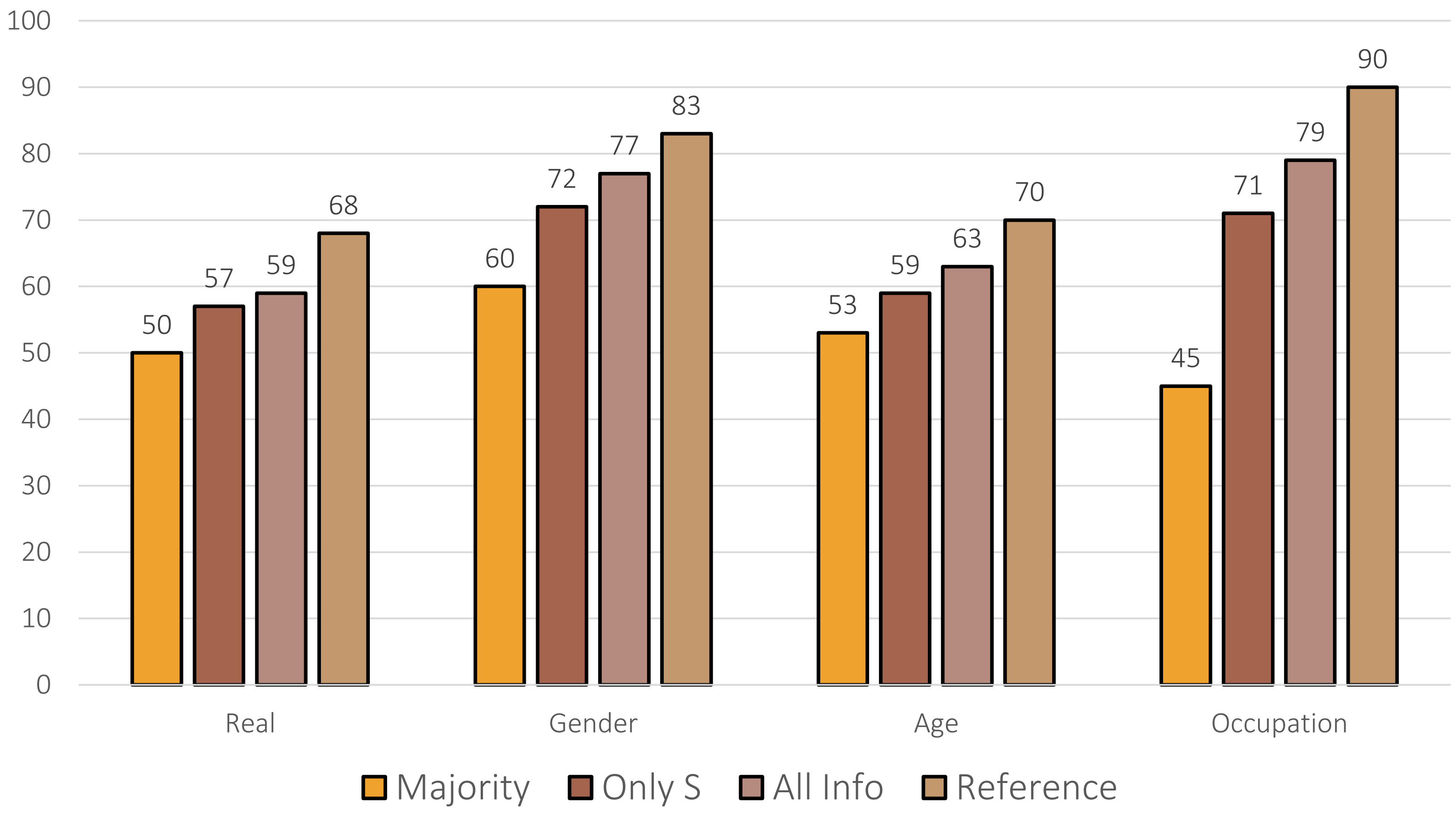
Do you think it is real or artificially generated?

Can you guess the age, gender, occupation, etc.?

Evaluation on MovieLens Titles



Evaluation on YAGO Descriptions



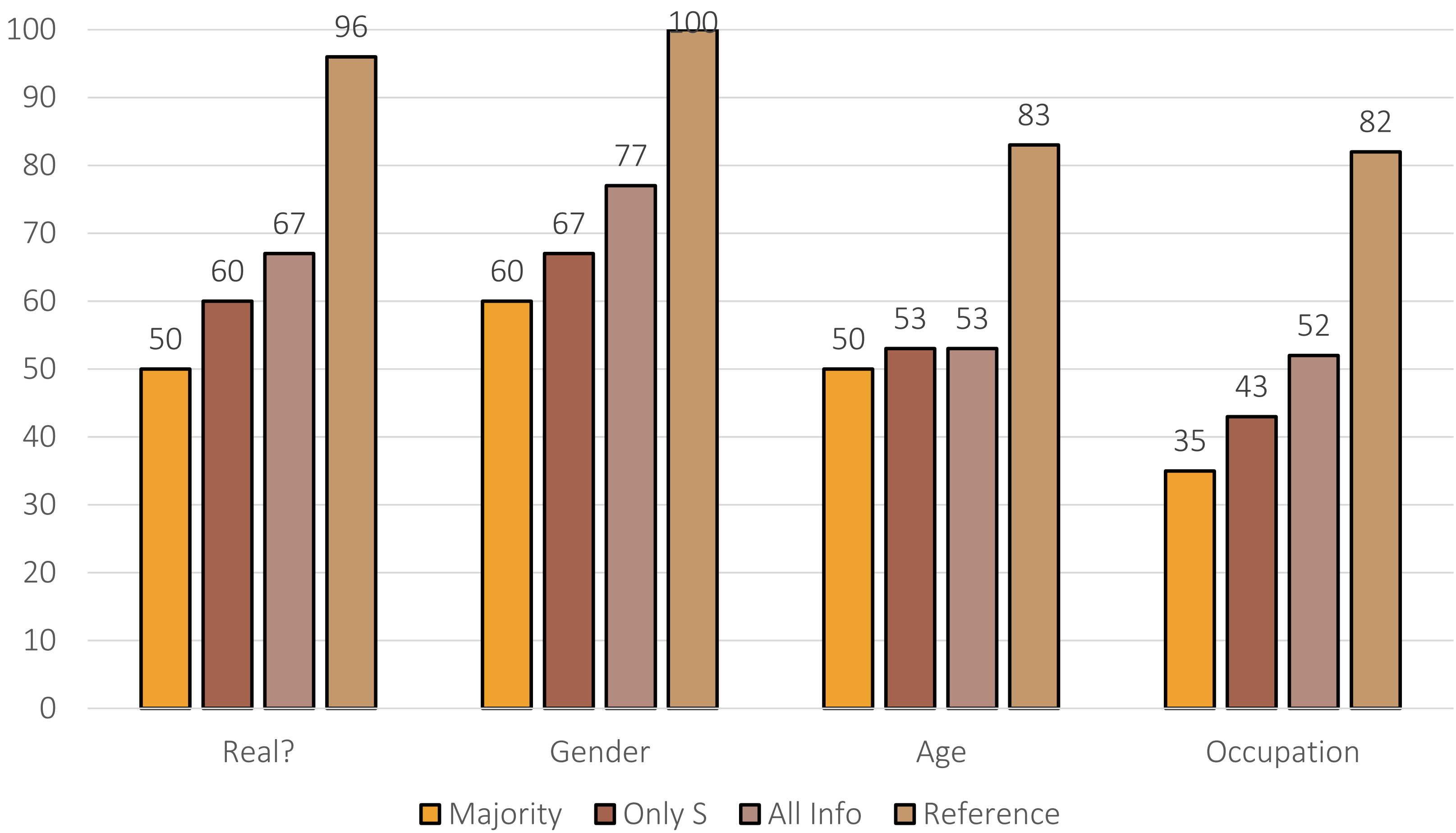
Generated Images for YAGO

Sports

Male Celebrity

Female Celebrity

Human Evaluation for Images



Multimodal Attribute Extraction

Gray Vinyl Barstool
This sleek dual purpose stool easily adjusts from counter to bar height. The backless design is casual and contemporary which allow it to seamlessly accent any area in the home. The easy to clean vinyl upholstery is perfect when being used on a regular basis. The height adjustable swivel seat adjusts from counter to bar height with the handle located below the seat….

Color Finish	Gray
Style	Contemporary
Adjustable Height	Yes
Frame Material	Metal

MAE Dataset

Cleaned up crawl of retail products in the Diffbot Knowledge Graph

Number of Entities	2.25 million
Number of Images	4.172 million
Number of unique Attributes	2,114
Number of unique Values	15,380
Number of Attribute-Value Pairs	7.671 million

Multimodal Attribute Extraction

Task: Given text and images about an entity, extract attributes

Dataset: Massive, diverse, open-domain dataset

Evaluation: Curated, small, held-out dataset

Baseline: Shows the challenge, and promise, of the task

https://rloganiv.github.io/mae/

Take-aways

- Knowledge Graphs are a useful representation
  - But, are incomplete and noisy is a problem
- Knowledge Graph Embeddings
  - Dense representations of entities and relations
  - Easy to learn, and very powerful
- Injecting Prior Knowledge
  - Use domain information to train more efficiently
- Injecting Multiple Modalities
  - Use all types of available information: Images, text, numbers

Work with **Matt Gardner** and me

as part of

The Allen Institute for Artificial Intelligence in **Irvine**, CA

All levels: pre-docs, PhD interns, postdocs, and research scientists!

Thank you!

sameersingh.org
sameer@uci.edu
@sameer_

In collaboration with Jay Pujara, Pouya Pezeshkpour, Mike Tung, Liyan Chen, Tim Rocktaschel, Samuel Humeau, Sebastian Riedel, and Robert Logan