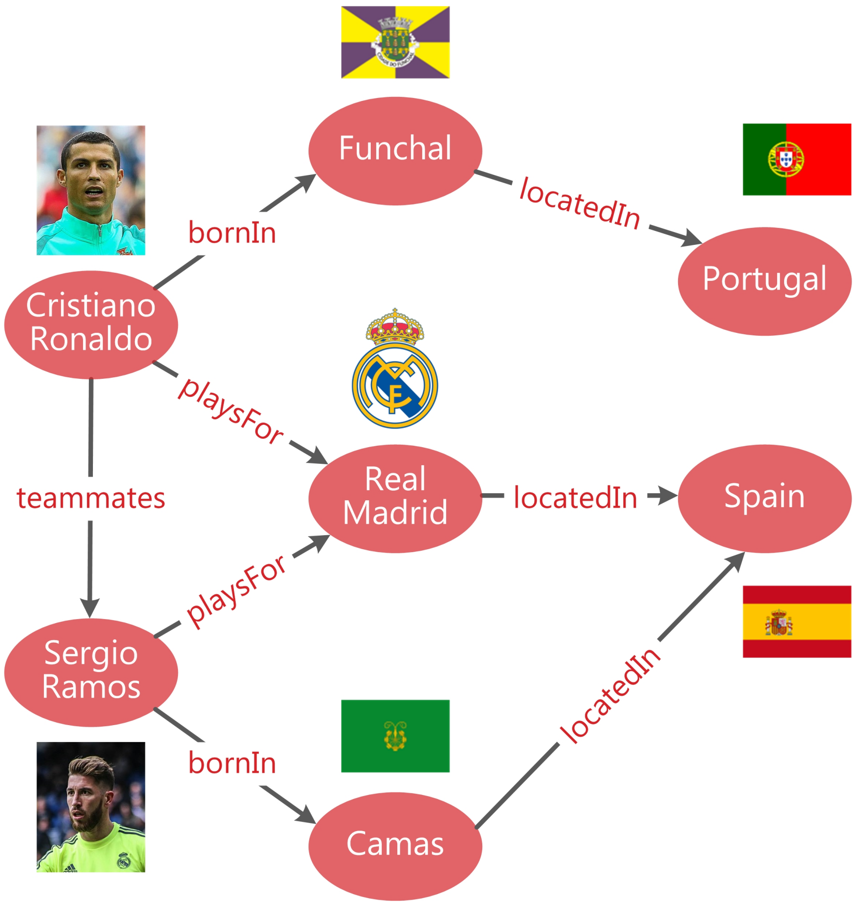
Learning Distributed Representations of Large-Scale Knowledge Graphs

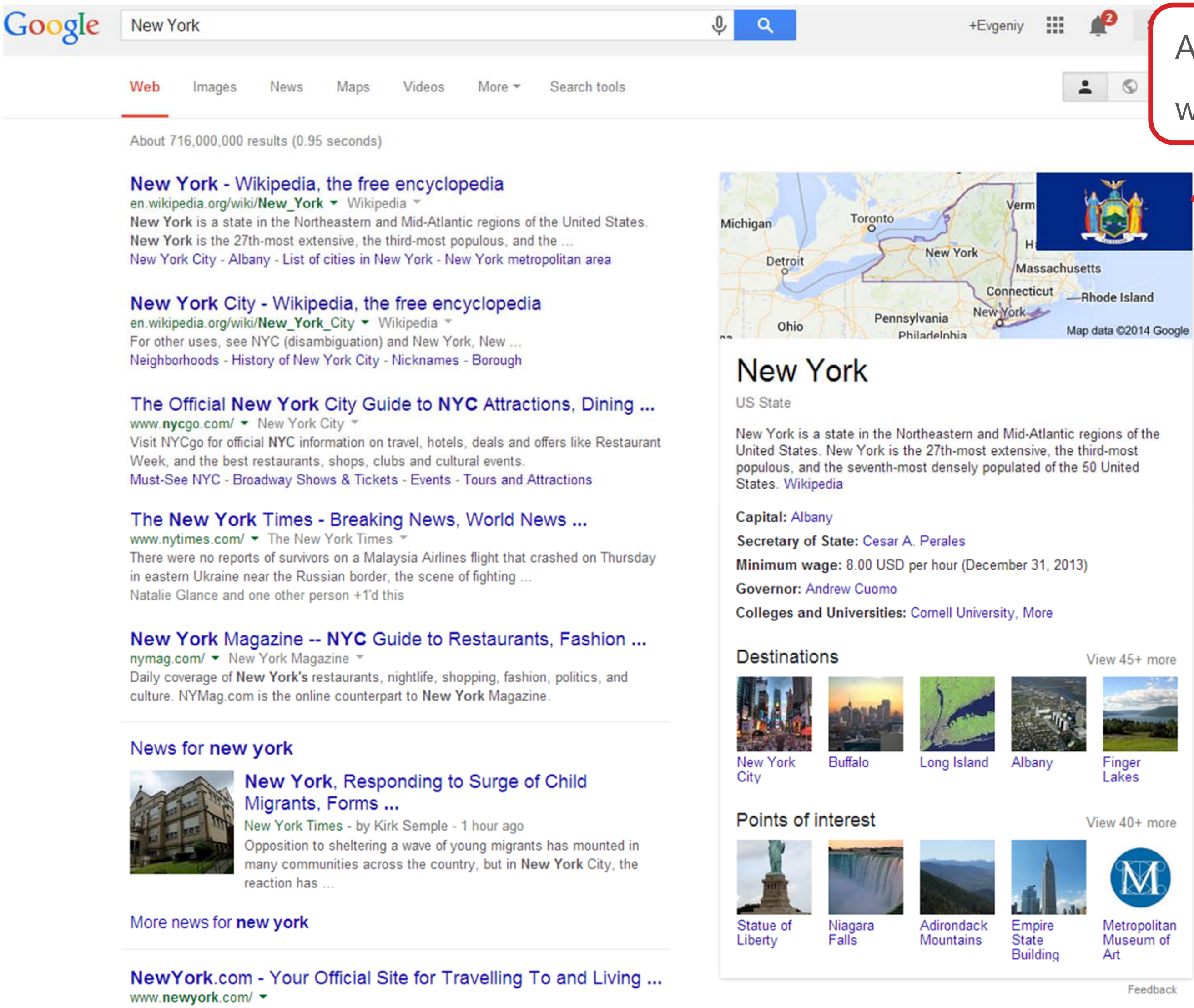
Quan Wang, Bin Wang
Institute of Information Engineering
Chinese Academy of Sciences
{wangquan,wangbin}@iie.ac.cn

Knowledge graphs

- Knowledge stored in graphs
- Describing entities and their relationships
- Nodes stand for entities
- Typed edges between nodes indicate various relationships between entities



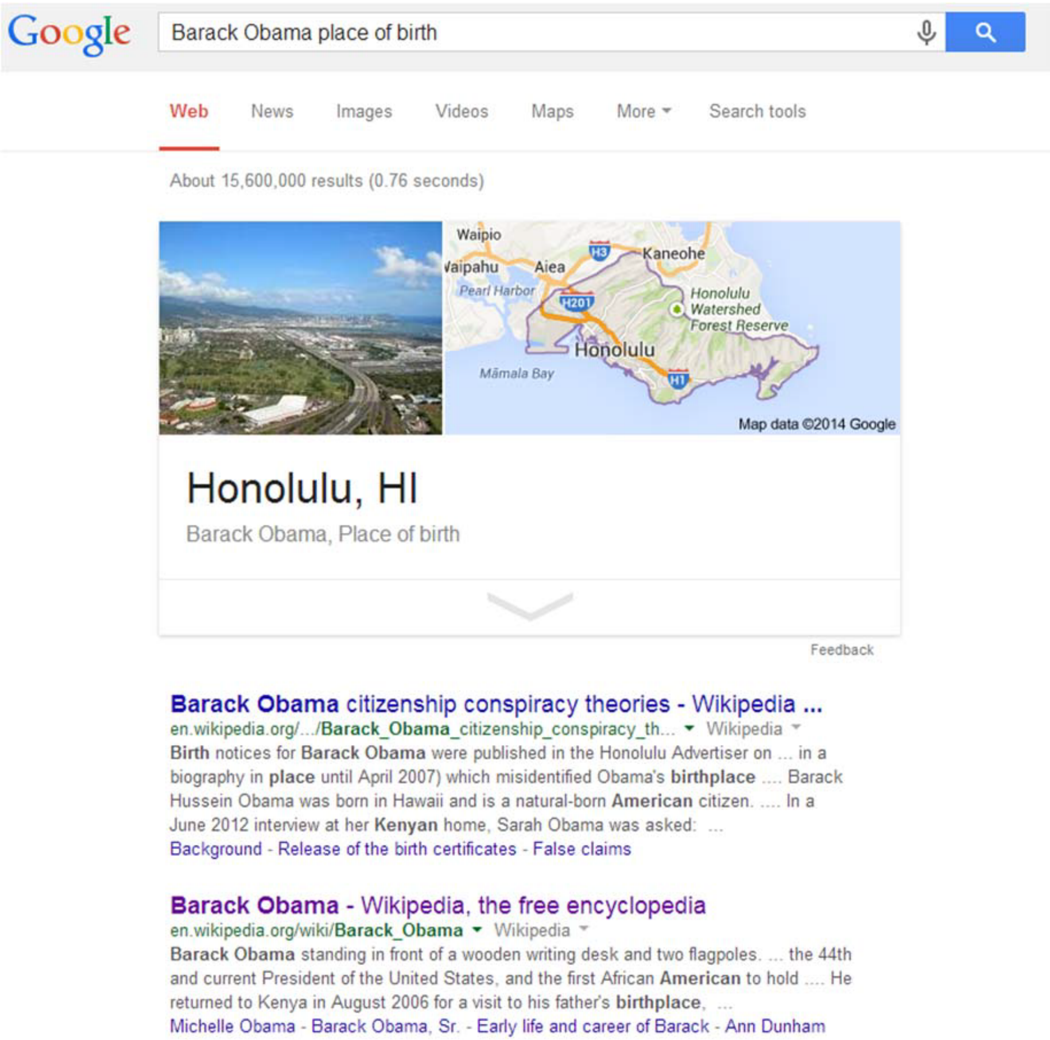
Applications: Web search

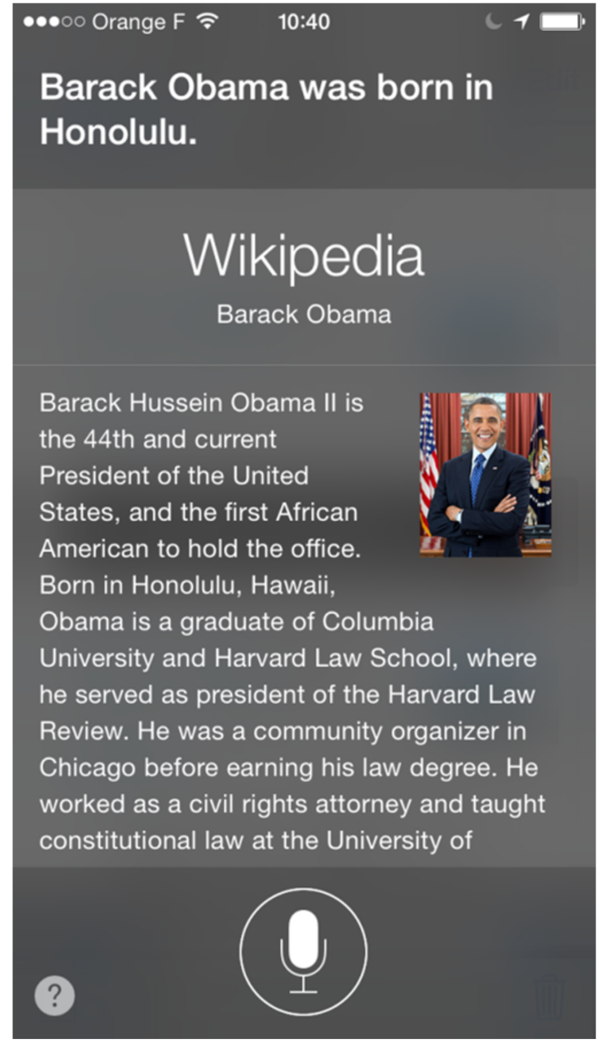


Augmenting the presentation with relevant facts

Example from: Antoine Bordes and Evgeniy Gabrilovich (2014). [Constructing and mining web-scale knowledge graphs](). KDD'14 tutorial.

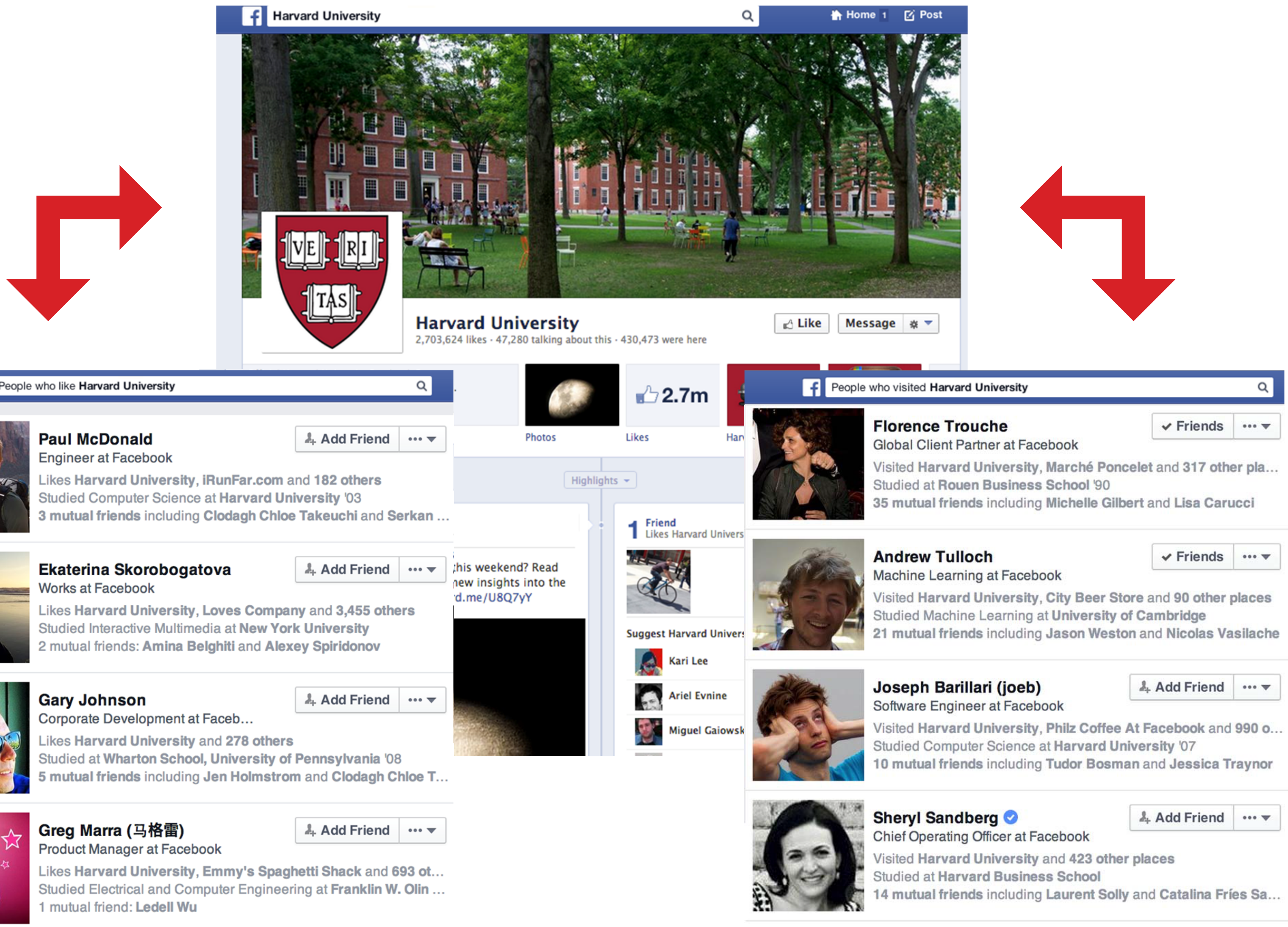
Applications: Question answering





Example from: Antoine Bordes and Evgeniy Gabrilovich (2014). [Constructing and mining web-scale knowledge graphs](). KDD'14 tutorial.

Applications: Social networks



Example from: Antoine Bordes and Evgeniy Gabrilovich (2014). [Constructing and mining web-scale knowledge graphs](). KDD'14 tutorial.

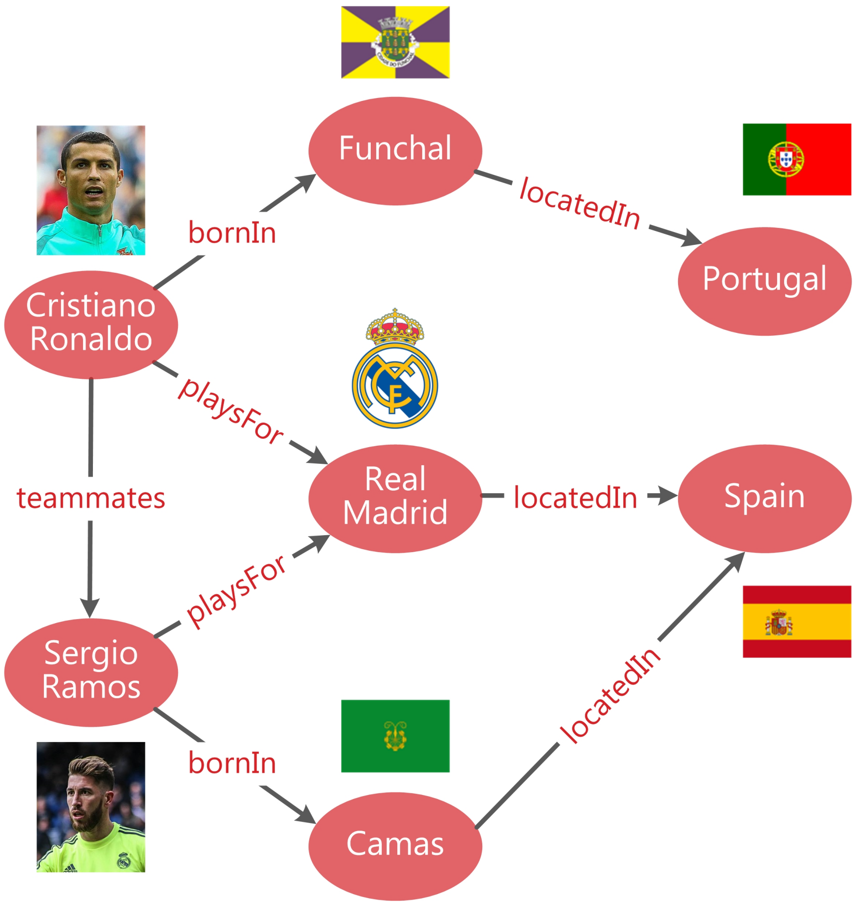
Applications: Social networks (cont.)

Enabling structured search within the graph

Example from: Antoine Bordes and Evgeniy Gabrilovich (2014). [Constructing and mining web-scale knowledge graphs](). KDD'14 tutorial.

Symbolic representation of knowledge graphs

- Most knowledge graph implementations use RDF triples
  - Each fact represented as a subject-relation-object triple



subject	relation	object
Cristiano Ronaldo	bornIn	Funchal
Cristiano Ronaldo	playsFor	Real Madrid
Cristiano Ronaldo	teammates	Sergio Ramos
Sergio Ramos	bornIn	Camas
Sergio Ramos	playsFor	Real Madrid
Funchal	locatedIn	Portugal
Real Madrid	locatedIn	Spain
Camas	locatedIn	Spain

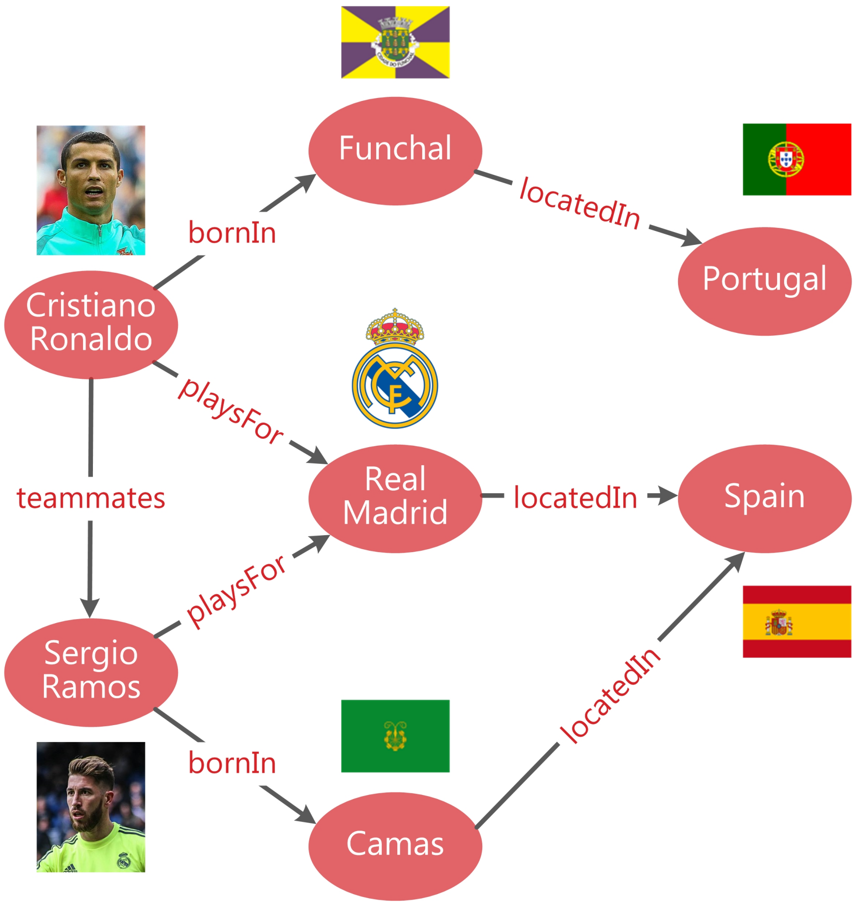
Limitations of symbolic representation

- Hard to carry out computation and inference on the graph
  - Is **Spain** more similar to **Camas** (a municipality located in Spain) or **Portugal** (both Portugal and Spain are European countries)?
  - What is the relationship between **Cristiano Ronaldo** and **Portugal**?
  - What is the **nationality** of **Sergio Ramos**?
  - …
- Computational complexity of algorithms
  - Complexity depends on explicit dimensionality (in size of data)
  - Query-time inference is sometimes NP-hard
  - Not trivial to parallelize, or use GPUs

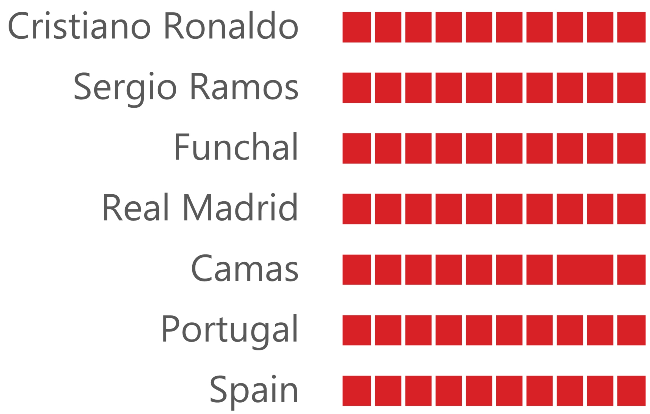
Material based on: Jay Pujara and Sameer Singh (2017). Mining knowledge graphs from text. KDD'17 tutorial.

Distributed representation of knowledge graphs

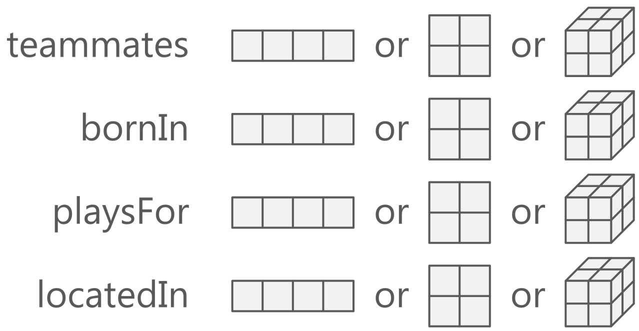
- Represent entities and relations in continuous vector spaces
  - Efficient task-independent feature learning for entities and relations



Entities as points in vector spaces (vectors)

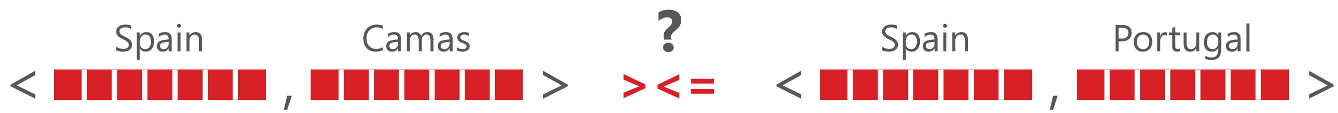


Relations as operations between entities (vectors/matrices/tensors)

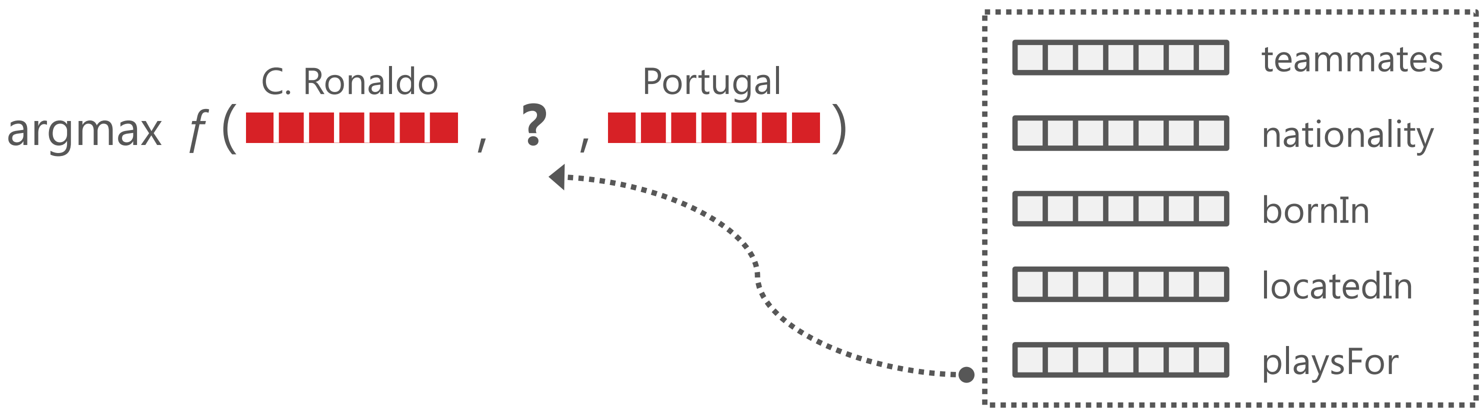


Benefits of distributed representation

- Easy computation and inference on the graph
  - Is **Spain** more similar to **Camas** (a municipality located in Spain) or **Portugal** (both Portugal and Spain are European countries)?

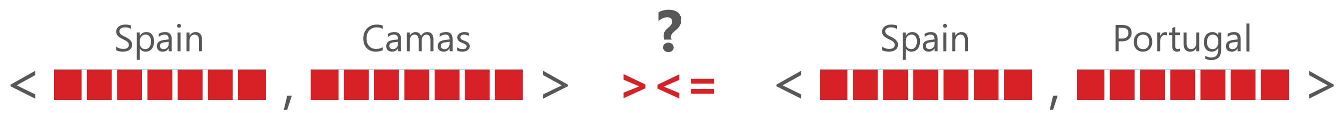


  - What is the relationship between **Cristiano Ronaldo** and **Portugal**?

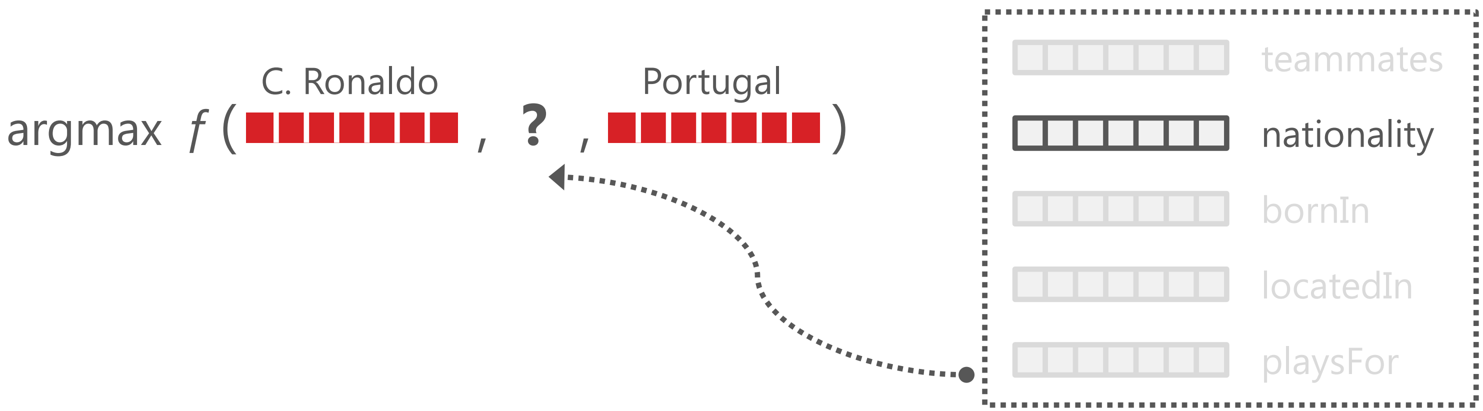


Benefits of distributed representation

- Easy computation and inference on the graph
  - Is **<u>Spain</u>** more similar to **<u>Camas</u>** (a municipality located in Spain) or **<u>Portugal</u>** (both Portugal and Spain are European countries)?

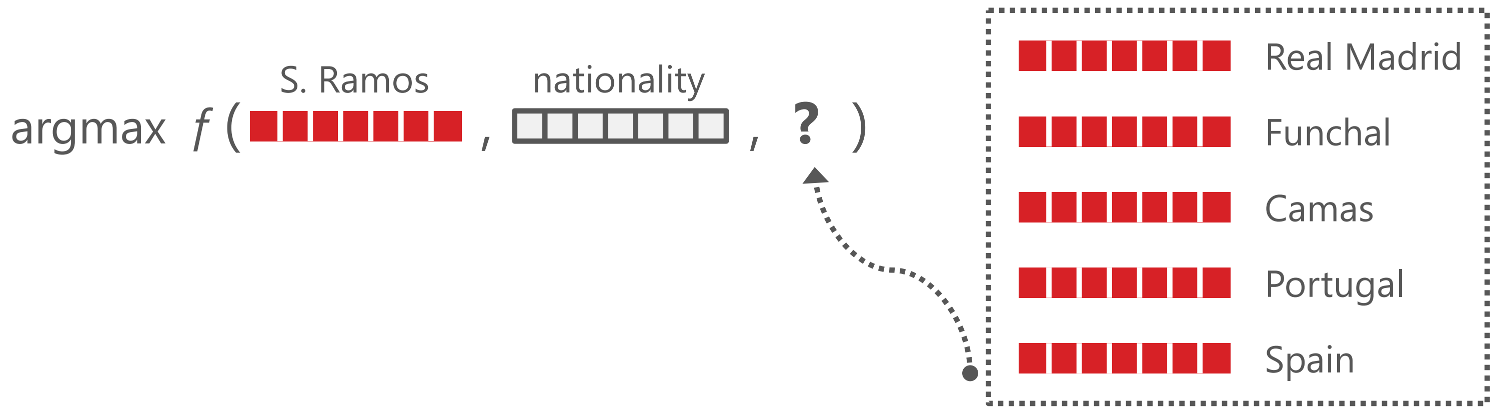


  - What is the relationship between **<u>Cristiano Ronaldo</u>** and **<u>Portugal</u>**?



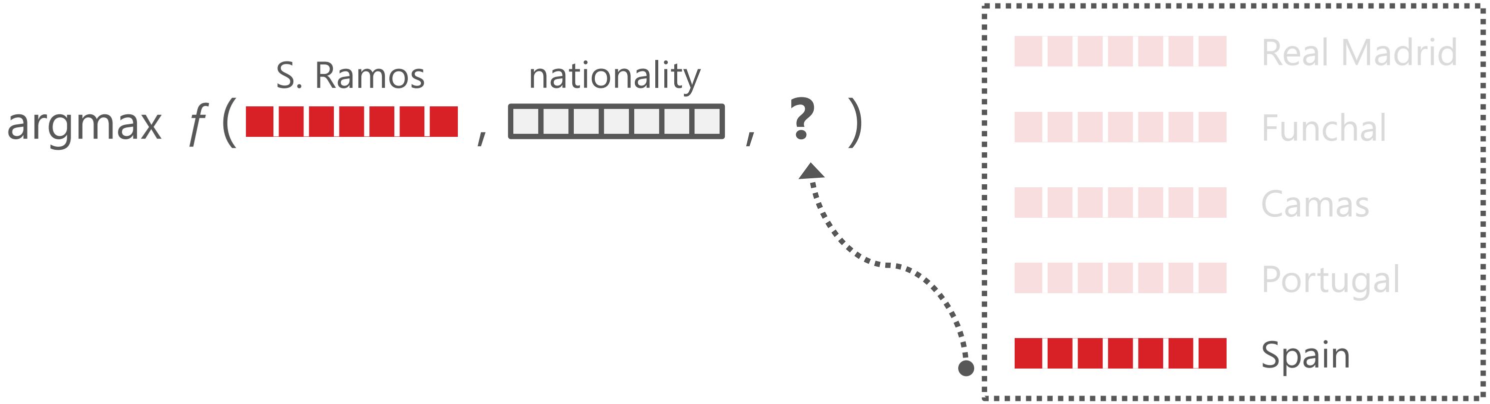
Benefits of distributed representation (cont.)

- Easy computation and inference on the graph
  - What is the **nationality** of **Sergio Ramos**?



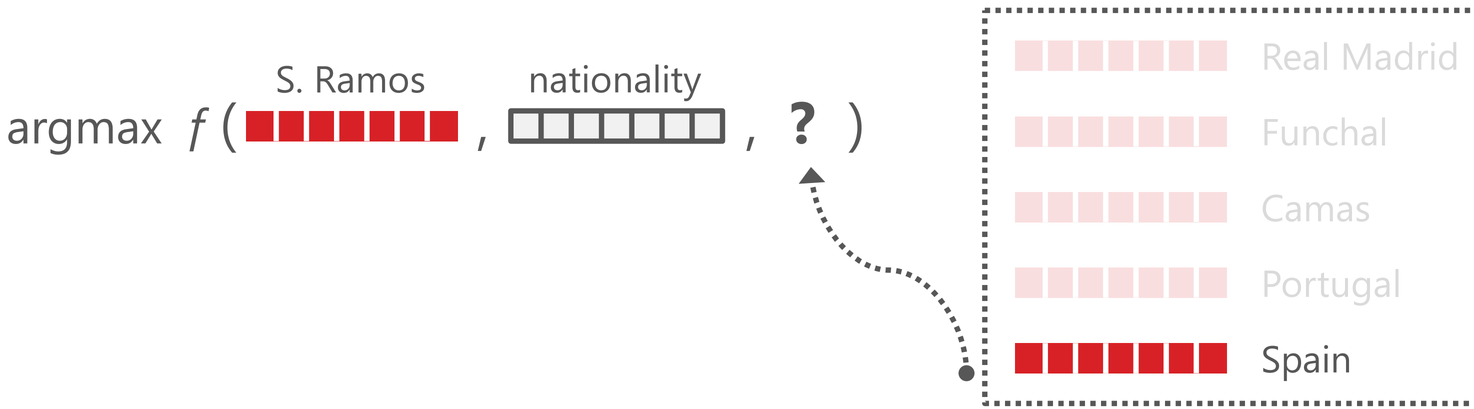
Benefits of distributed representation (cont.)

- Easy computation and inference on the graph
  - What is the **nationality** of **Sergio Ramos**?



Benefits of distributed representation (cont.)

- Easy computation and inference on the graph
  - What is the **nationality** of **Sergio Ramos**?

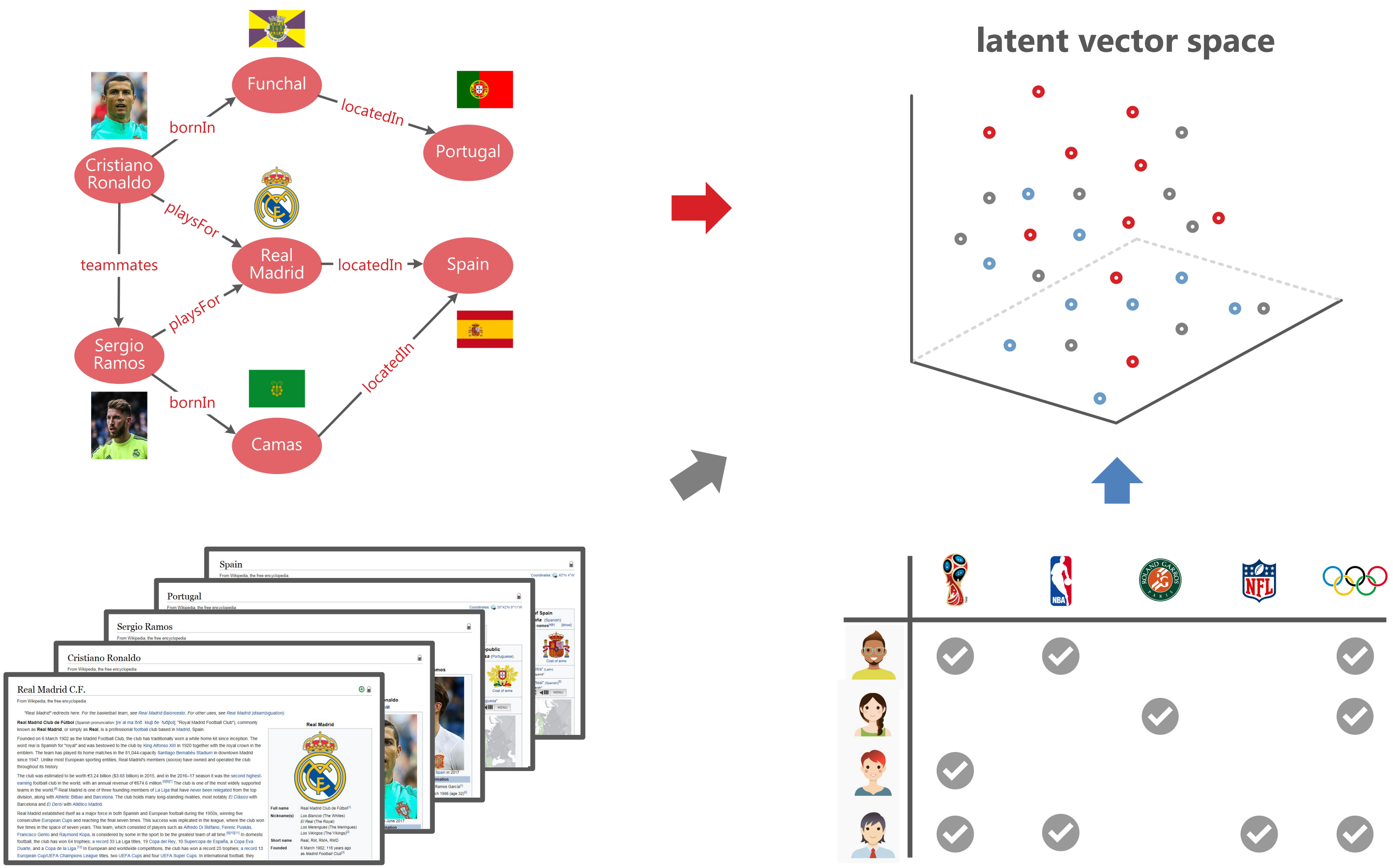


- Computational complexity of algorithms
  - Complexity depends on latent dimensions
  - Querying is often cheap
  - GPU-parallelism friendly

Material based on: Jay Pujara and Sameer Singh (2017). Mining knowledge graphs from text. KDD'17 tutorial.

Benefits of distributed representation (cont.)

- Better support for data fusion, facilitating application of knowledge graphs to various domains



This talk

- Part I: Models based on RDF triples
  - Learning distributed representations from RDF triples
- Part II: Models based on extra information
  - Learning with extra information beyond RDF triples
- Part III: Applications
  - Applying distributed representations to downstream tasks
- Part IV: Summary and concluding remarks