

Learning Distributed Representations of Large-Scale Knowledge Graphs

Quan Wang, Bin Wang

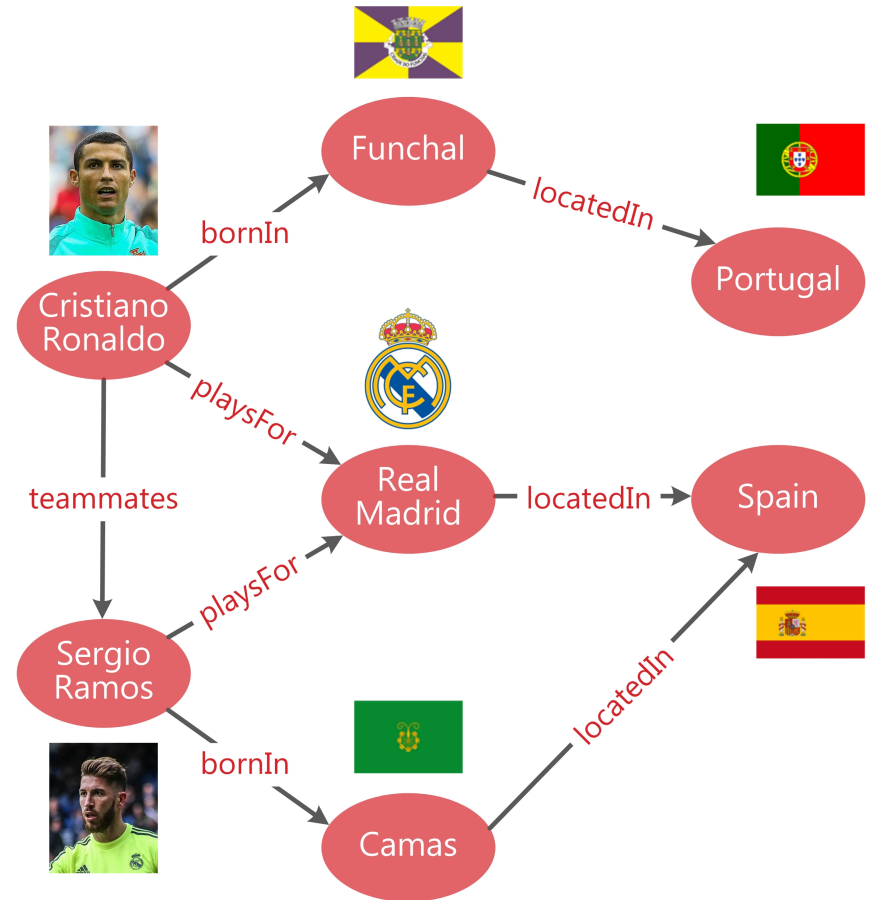
Institute of Information Engineering

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

The image shows a Google search interface for the query "New York". The search bar at the top contains the text "New York" and a microphone icon. Below the search bar, navigation tabs for "Web", "Images", "News", "Maps", "Videos", "More", and "Search tools" are visible. The search results page displays "About 716,000,000 results (0.95 seconds)".

The search results include several links:

- New York - Wikipedia, the free encyclopedia**: en.wikipedia.org/wiki/New_York. Description: New York is a state in the Northeastern and Mid-Atlantic regions of the United States. New York is the 27th-most extensive, the third-most populous, and the ... New York City - Albany - List of cities in New York - New York metropolitan area
- New York City - Wikipedia, the free encyclopedia**: en.wikipedia.org/wiki/New_York_City. Description: For other uses, see NYC (disambiguation) and New York, New ... Neighborhoods - History of New York City - Nicknames - Borough
- The Official New York City Guide to NYC Attractions, Dining ...**: www.nycgo.com/. Description: Visit NYCgo for official NYC information on travel, hotels, deals and offers like Restaurant Week, and the best restaurants, shops, clubs and cultural events. Must-See NYC - Broadway Shows & Tickets - Events - Tours and Attractions
- The New York Times - Breaking News, World News ...**: www.nytimes.com/. Description: There were no reports of survivors on a Malaysia Airlines flight that crashed on Thursday in eastern Ukraine near the Russian border, the scene of fighting ... Natalie Glance and one other person +1'd this
- New York Magazine -- NYC Guide to Restaurants, Fashion ...**: nymag.com/. Description: Daily coverage of New York's restaurants, nightlife, shopping, fashion, politics, and culture. NYMag.com is the online counterpart to New York Magazine.

Below the search results, there is a section for "News for new york" featuring a thumbnail image of a building and the headline "New York, Responding to Surge of Child Migrants, Forms ...". The article is from the New York Times, by Kirk Semple, and is dated 1 hour ago. The description mentions opposition to sheltering a wave of young migrants across the country, but in New York City, the reaction has ...

At the bottom of the search results, there is a link to "NewYork.com - Your Official Site for Travelling To and Living ...".

On the right side of the search results, there is a "Knowledge panel" for "New York". The panel includes a map of New York state, the state flag, and the following information:

- New York**: US State
- Description: New York is a state in the Northeastern and Mid-Atlantic regions of the United States. New York is the 27th-most extensive, the third-most populous, and the seventh-most densely populated of the 50 United States. Wikipedia
- Capital**: Albany
- Secretary of State**: Cesar A. Peralas
- Minimum wage**: 8.00 USD per hour (December 31, 2013)
- Governor**: Andrew Cuomo
- Colleges and Universities**: Cornell University, More

Below the knowledge panel, there are sections for "Destinations" and "Points of interest".

Destinations: View 45+ more. Includes images and links for New York City, Buffalo, Long Island, Albany, and Finger Lakes.

Points of interest: View 40+ more. Includes images and links for Statue of Liberty, Niagara Falls, Adirondack Mountains, Empire State Building, and Metropolitan Museum of Art.

Augmenting the presentation with relevant facts

Applications: Question answering

Google

Barack Obama place of birth

Web News Images Videos Maps More Search tools

About 15,600,000 results (0.76 seconds)



Honolulu, HI

Barack Obama, Place of birth

Feedback

Barack Obama citizenship conspiracy theories - Wikipedia ...

en.wikipedia.org/.../Barack_Obama_citizenship_conspiracy_th... Wikipedia ▾
Birth notices for **Barack Obama** were published in the Honolulu Advertiser on ... in a biography in **place** until April 2007) which misidentified Obama's **birthplace** Barack Hussein Obama was born in Hawaii and is a natural-born **American** citizen. In a June 2012 interview at her **Kenyan** home, Sarah Obama was asked: ...
Background - Release of the birth certificates - False claims

Barack Obama - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Barack_Obama ▾ Wikipedia ▾
Barack Obama standing in front of a wooden writing desk and two flagpoles. ... the 44th and current President of the United States, and the first African **American** to hold He returned to Kenya in August 2006 for a visit to his father's **birthplace**, ...
Michelle Obama - Barack Obama, Sr. - Early life and career of Barack - Ann Dunham

Orange F 10:40

Barack Obama was born in Honolulu.

Wikipedia

Barack Obama

Barack Hussein Obama II is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he served as president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney and taught constitutional law at the University of



Applications: Social networks

The image shows a screenshot of the Harvard University Facebook page. At the top, the page header includes the Harvard University logo and name, along with navigation links for Home and Post. Below the header is a large photo of a Harvard campus scene. Underneath the photo, the page name 'Harvard University' is displayed, along with the number of likes (2,703,624) and the number of people talking about it (47,280). Below the main photo, there are two red arrows pointing outwards, one to the left and one to the right, indicating the direction of the links to the 'People who like Harvard University' and 'People who visited Harvard University' sections.

People who like Harvard University

- Paul McDonald**
Engineer at Facebook
Likes Harvard University, iRunFar.com and 182 others
Studied Computer Science at Harvard University '03
3 mutual friends including Clodagh Chloe Takeuchi and Serkan ...
- Ekaterina Skorobogatova**
Works at Facebook
Likes Harvard University, Loves Company and 3,455 others
Studied Interactive Multimedia at New York University
2 mutual friends: Amina Belghiti and Alexey Spiridonov
- Gary Johnson**
Corporate Development at Facebook
Likes Harvard University and 278 others
Studied at Wharton School, University of Pennsylvania '08
5 mutual friends including Jen Holmstrom and Clodagh Chloe T...
- Greg Marra (马格雷)**
Product Manager at Facebook
Likes Harvard University, Emmy's Spaghetti Shack and 693 others
Studied Electrical and Computer Engineering at Franklin W. Olin ...
1 mutual friend: Ledell Wu

People who visited Harvard University

- Florence Trouche**
Global Client Partner at Facebook
Visited Harvard University, Marché Poncelet and 317 other places
Studied at Rouen Business School '90
35 mutual friends including Michelle Gilbert and Lisa Carucci
- Andrew Tulloch**
Machine Learning at Facebook
Visited Harvard University, City Beer Store and 90 other places
Studied Machine Learning at University of Cambridge
21 mutual friends including Jason Weston and Nicolas Vasilache
- Joseph Barillari (joeb)**
Software Engineer at Facebook
Visited Harvard University, Philz Coffee At Facebook and 990 others
Studied Computer Science at Harvard University '07
10 mutual friends including Tudor Bosman and Jessica Traynor
- Sheryl Sandberg**
Chief Operating Officer at Facebook
Visited Harvard University and 423 other places
Studied at Harvard Business School
14 mutual friends including Laurent Solly and Catalina Fries Sa...

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

Applications: Social networks (cont.)

People who like **Harvard University** and **Basketball** and work at **Facebook**

Groups
Apps
Events

Mike Vernal
VP Engineering at Facebook
Likes **Harvard University**, **Harvard Crimson** and **F@ceb00k Su...**
Studied Computer Science at **Harvard University '02**
10 mutual friends including **Keith Adams** and **Philip Bohannon**

Jared Morgenstern
Product Manager / Ninja - Games ...
Likes **Harvard University**, **F@ceb00k Summer Basketball Leag...**
Studied Computer Science at **Harvard University**
5 mutual friends including **Clodagh Chloe Takeuchi** and **Pierre ...**

Florin Rațiu
Software Engineer at Facebook
Likes **Harvard School of Public Health**, **Stanford 6th Man** and **ba...**
Studied Management Science and Engineering at **Stanford Univ...**
3 mutual friends including **Alexey Spiridonov** and **Serkan Piantino**

Ning Zhang (张宁)
Software Engineer at Facebook
Likes **Harvard University**, **Basketball** and **314 others**
Studied Computer Science at **University of Waterloo '06**
5 mutual friends including **Tudor Bosman** and **Ves Stoyanov**

Zhongyuan Xu (徐重远)
Software Engineer at Facebook
Likes **Harvard University**, **Basketball** and **364 others**
Studied at **Stony Brook University**
1 mutual friend: **Ledell Wu**

2147 ms

Enabling structured search within the graph

Symbolic representation of knowledge graphs

❑ Most knowledge graph implementations use RDF triples

- Each fact represented as a **subject-relation-object** triple



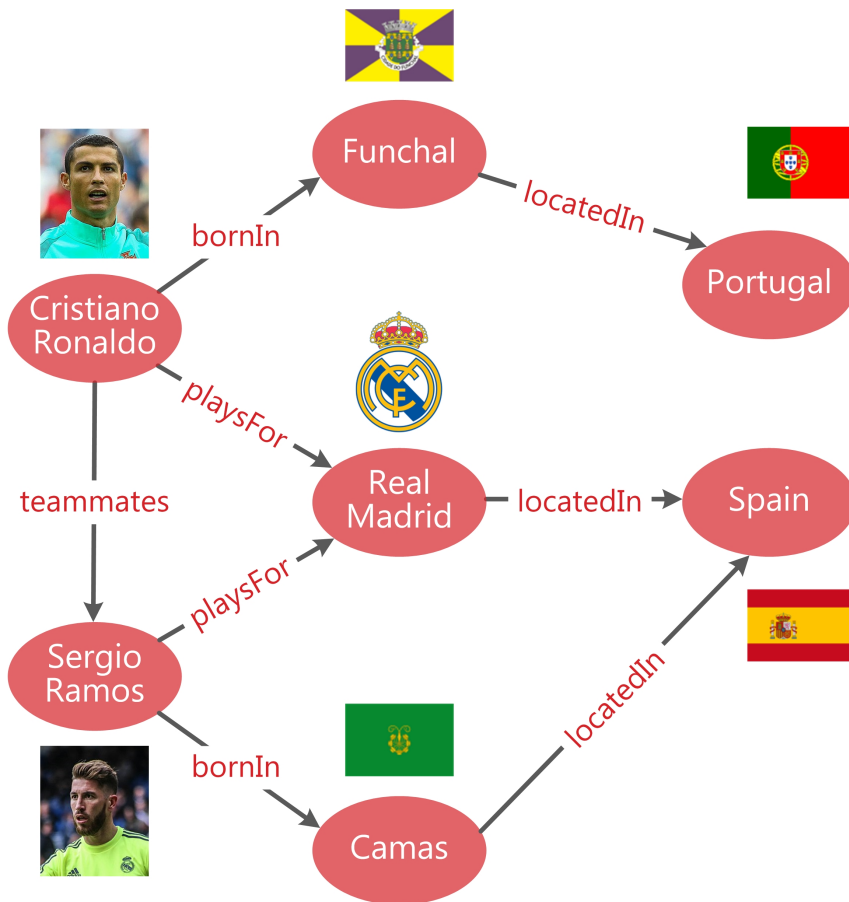
subject	relation	object
Cristiano Ronaldo	bornIn	Funchal
Cristiano Ronaldo	playsFor	Real Madrid
Cristiano Ronaldo	teammates	Sergio Ramos
Sergio Ramos	bornIn	Camas
Sergio Ramos	playsFor	Real Madrid
Funchal	locatedIn	Portugal
Real Madrid	locatedIn	Spain
Camas	locatedIn	Spain

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

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)

Cristiano Ronaldo	■■■■■■■■■■
Sergio Ramos	■■■■■■■■■■
Funchal	■■■■■■■■■■
Real Madrid	■■■■■■■■■■
Camas	■■■■■■■■■■
Portugal	■■■■■■■■■■
Spain	■■■■■■■■■■

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

teammates	□□□□	or	□□	or	□□□□
bornIn	□□□□	or	□□	or	□□□□
playsFor	□□□□	or	□□	or	□□□□
locatedIn	□□□□	or	□□	or	□□□□

Benefits of distributed representation

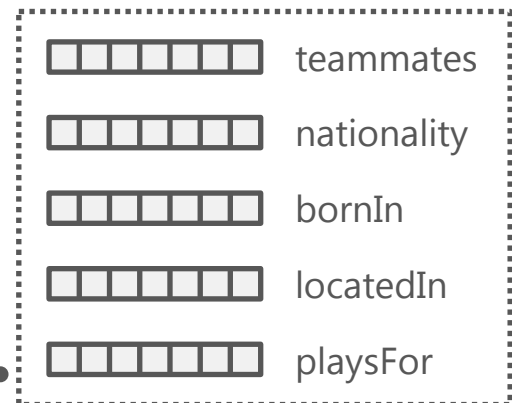
□ Easy computation and inference on the graph

- Is **Spain** more similar to **Camas** (a municipality located in Spain) or **Portugal** (both Portugal and Spain are European countries)?

$$\begin{matrix} \text{Spain} & & \text{Camas} & & ? & & \text{Spain} & & \text{Portugal} \\ < \text{■■■■■■■} , \text{■■■■■■■} > > < = < \text{■■■■■■■} , \text{■■■■■■■} > \end{matrix}$$

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

$$\text{argmax } f(\begin{matrix} \text{C. Ronaldo} \\ \text{■■■■■■■} \end{matrix}, \begin{matrix} ? \\ \text{■■■■■■■} \end{matrix}, \begin{matrix} \text{Portugal} \\ \text{■■■■■■■} \end{matrix})$$



Benefits of distributed representation

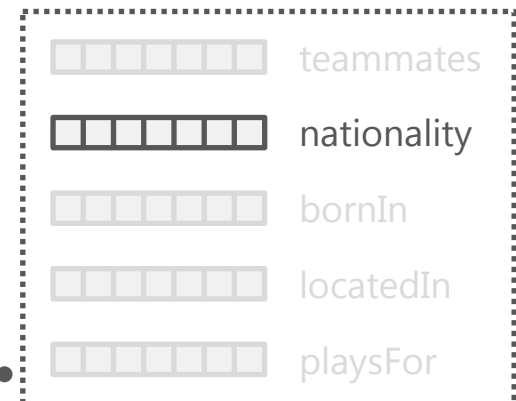
□ Easy computation and inference on the graph

- Is **Spain** more similar to **Camas** (a municipality located in Spain) or **Portugal** (both Portugal and Spain are European countries)?

$$\begin{matrix} \text{Spain} & & \text{Camas} & & ? & & \text{Spain} & & \text{Portugal} \\ < \text{■■■■■■} & , & \text{■■■■■■} & > > < = < \text{■■■■■■} & , & \text{■■■■■■} & > \end{matrix}$$

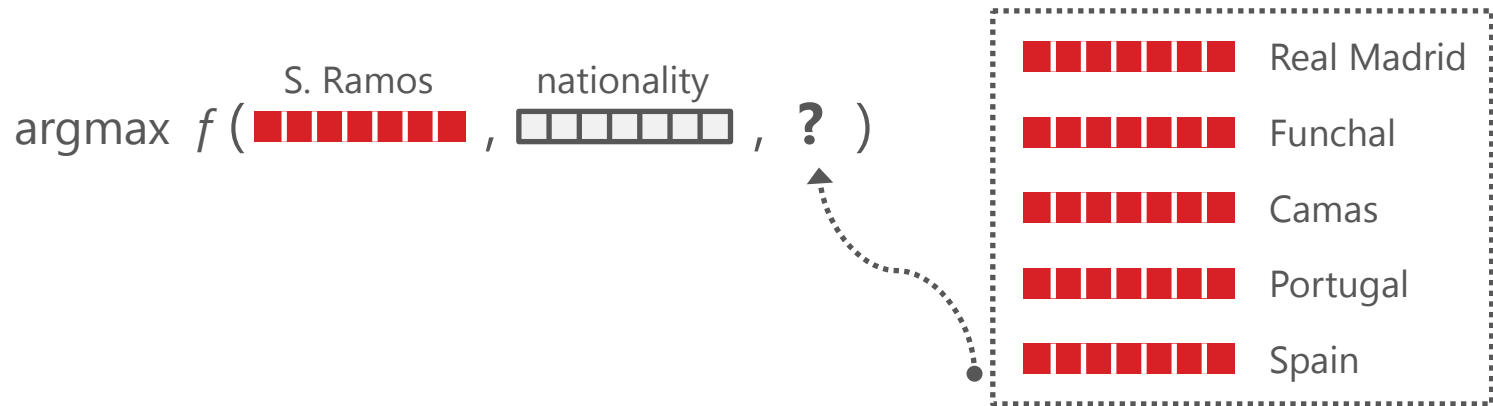
- What is the relationship between **Cristiano Ronaldo** and **Portugal**?

$$\text{argmax } f(\begin{matrix} \text{C. Ronaldo} \\ \text{■■■■■■} \end{matrix}, \begin{matrix} ? \\ \text{■■■■■■} \end{matrix}, \begin{matrix} \text{Portugal} \\ \text{■■■■■■} \end{matrix})$$



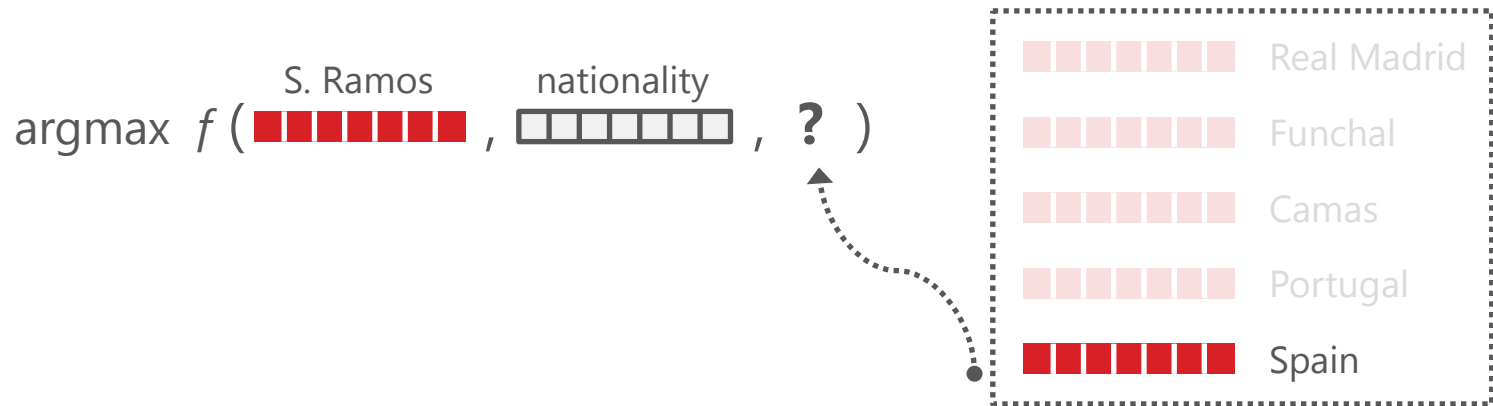
Benefits of distributed representation (cont.)

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



Benefits of distributed representation (cont.)

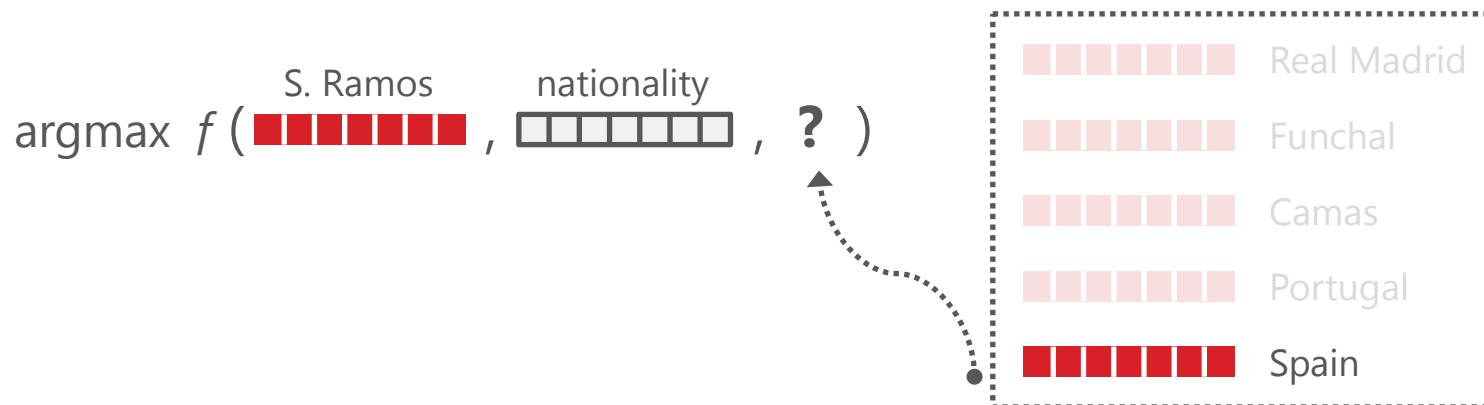
- Easy computation and inference on the graph
 - What is the **nationality** of **Sergio Ramos**?



Benefits of distributed representation (cont.)

□ Easy computation and inference on the graph

- What is the **nationality** of **Sergio Ramos**?

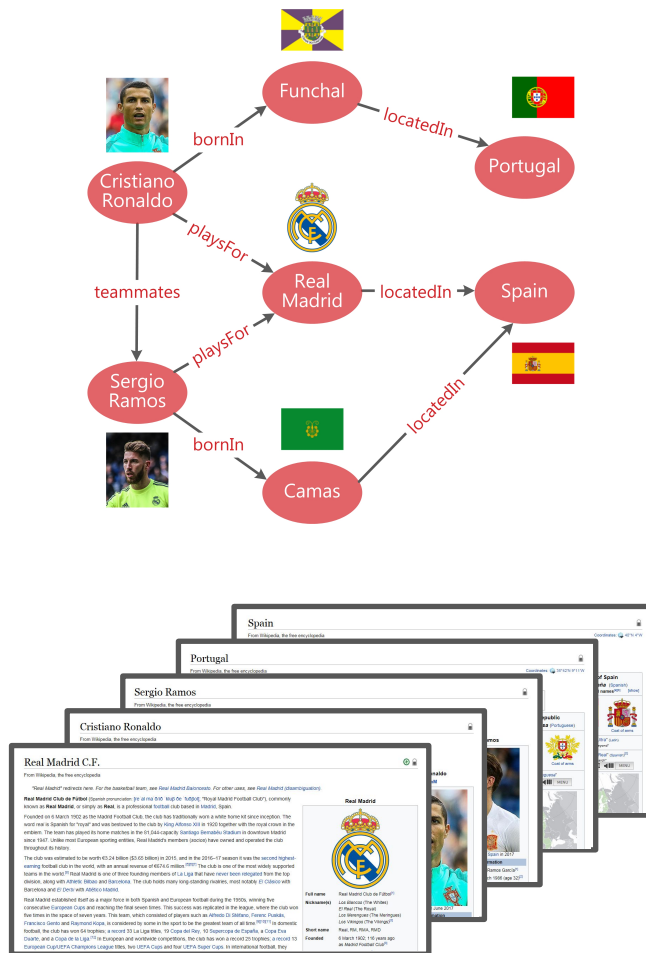


□ Computational complexity of algorithms

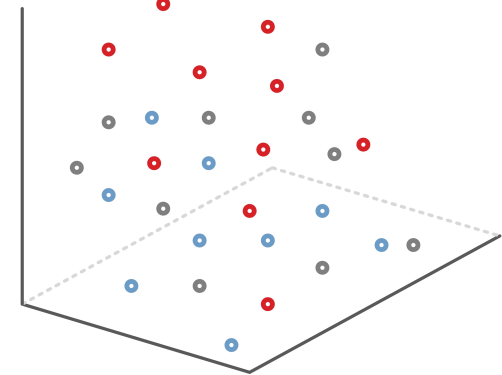
- Complexity depends on latent dimensions
- Querying is often cheap
- GPU-parallelism friendly

Benefits of distributed representation (cont.)

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



latent vector space



					
	✓	✓			✓
			✓		✓
	✓				
	✓	✓		✓	✓

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