Entity Embeddings with Conceptual Subspaces as a Basis for Plausible Reasoning

Shoaib Jameel

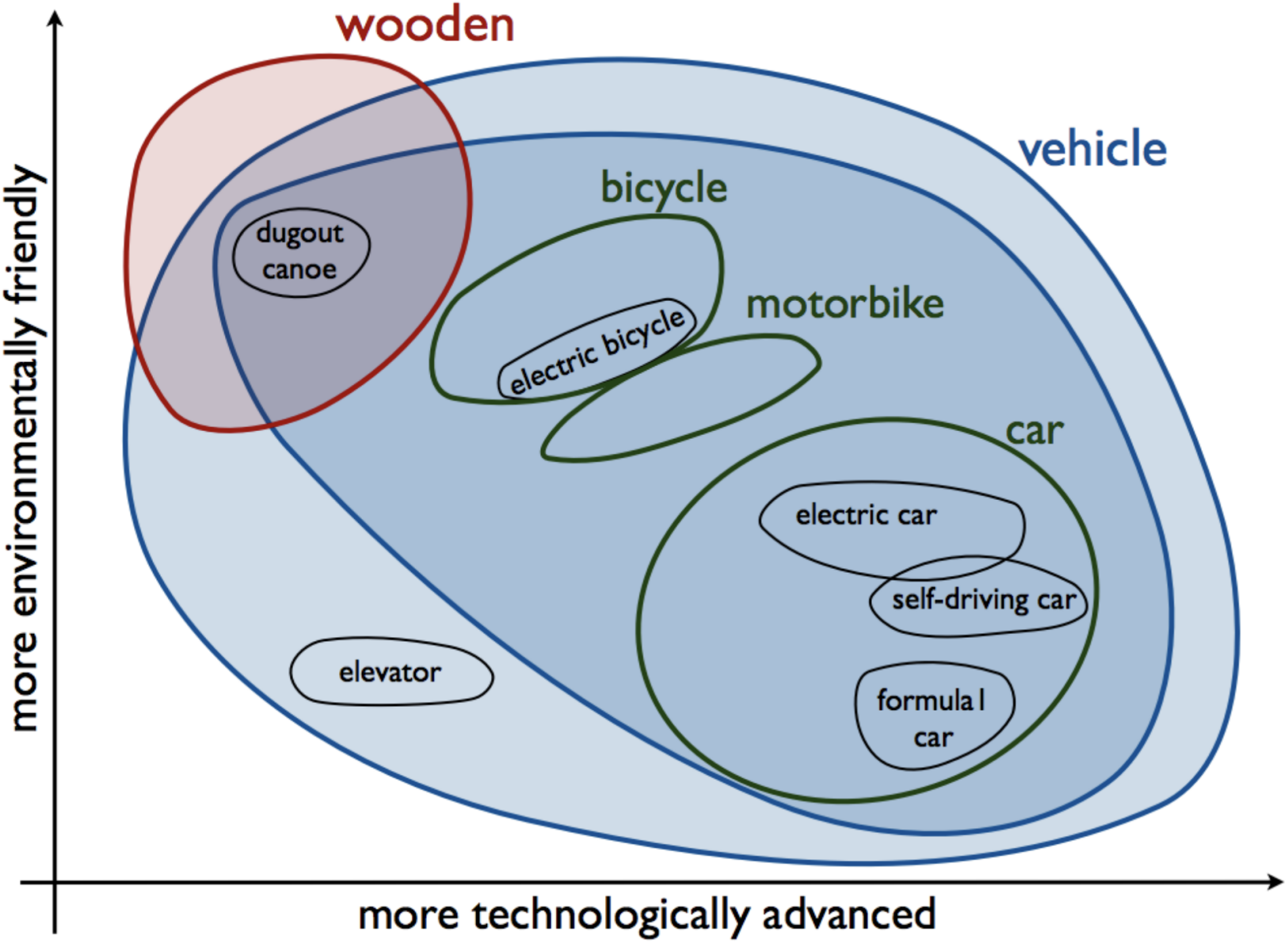
School of Computing.
University of Kent.

May 15, 2018

One Line Summary of the Work

Constructing a **domain-independent** semantic space, in which **entities** belonging to the **same semantic type** are **embedded** in **domain-independent subspaces**.

Conceptual Spaces in Psychology



Plausible Reasoning

Undergraduate students are exempt from council tax in the UK

PhD students are exempt from council tax in the UK

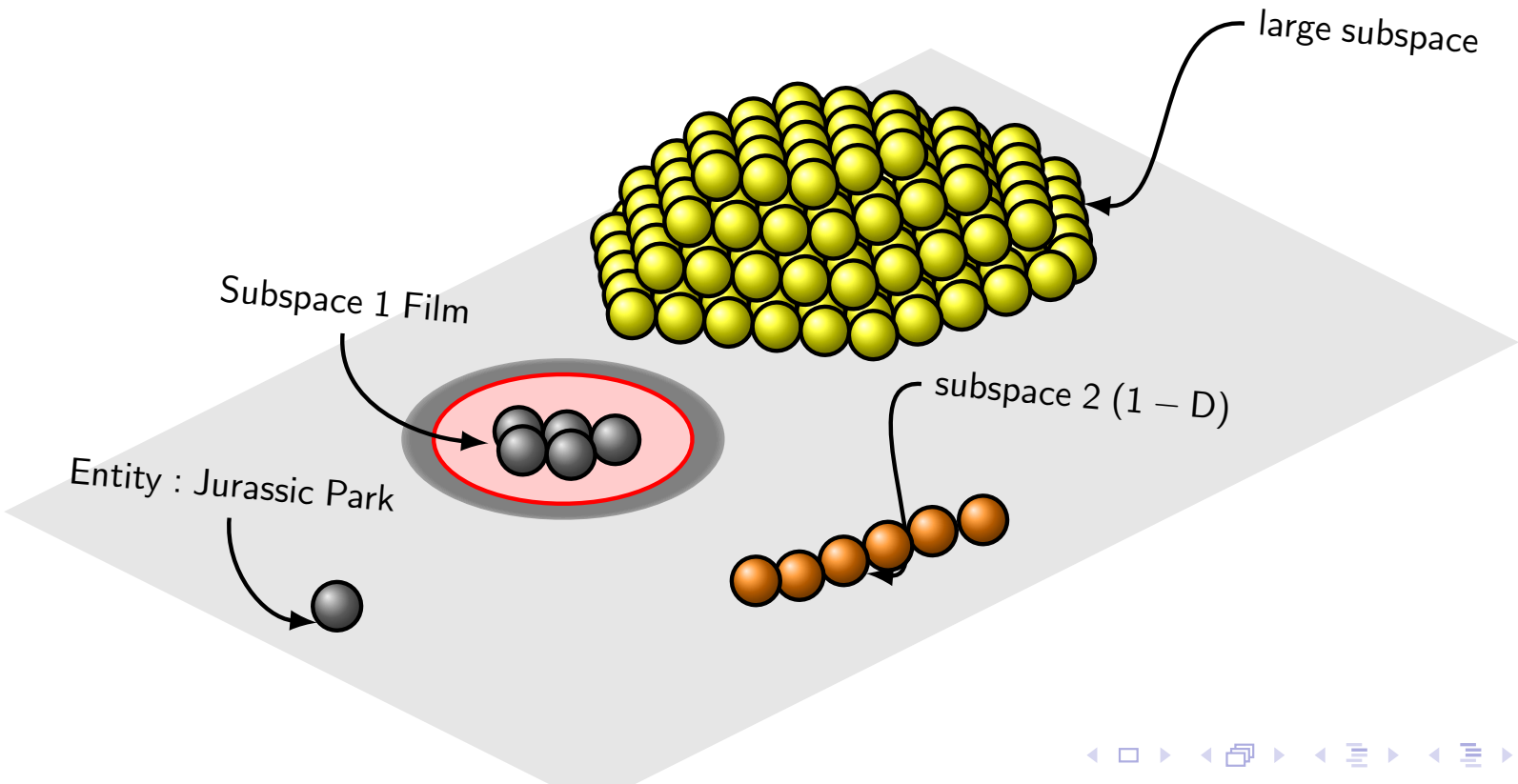
Infer via Interpolation

Master's students are exempt from council tax in the UK

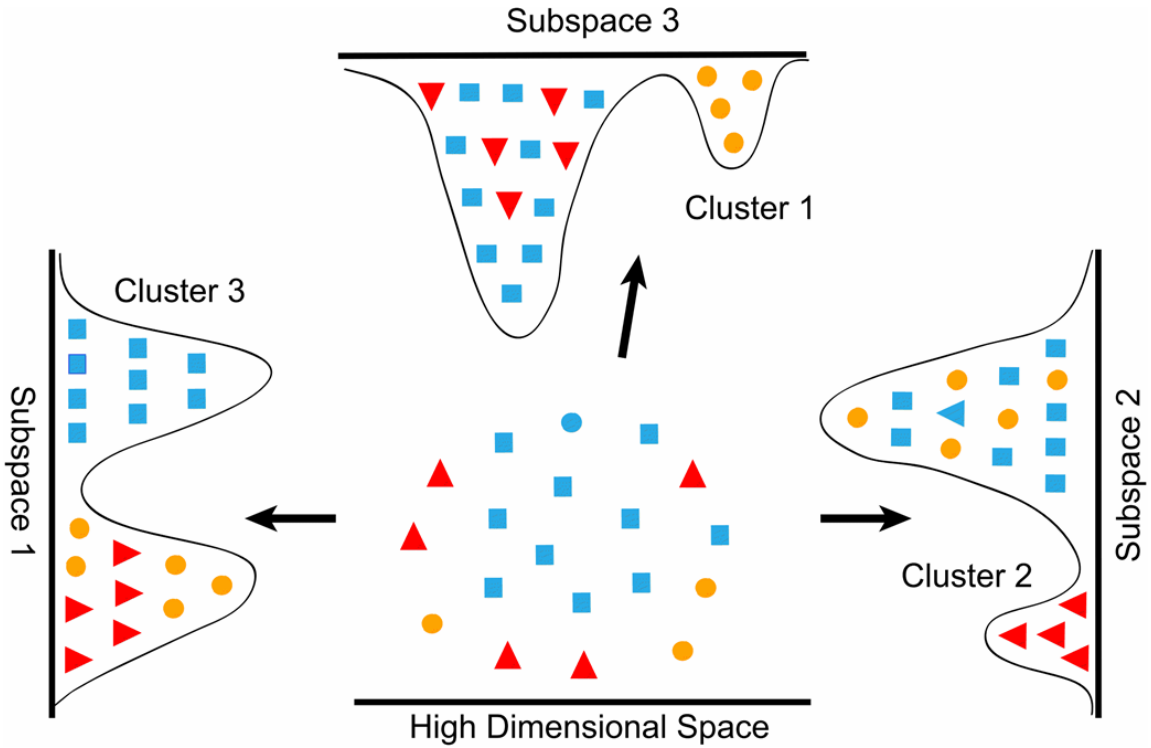
Automatically Generating Conceptual Subspaces

Conceptual Spaces

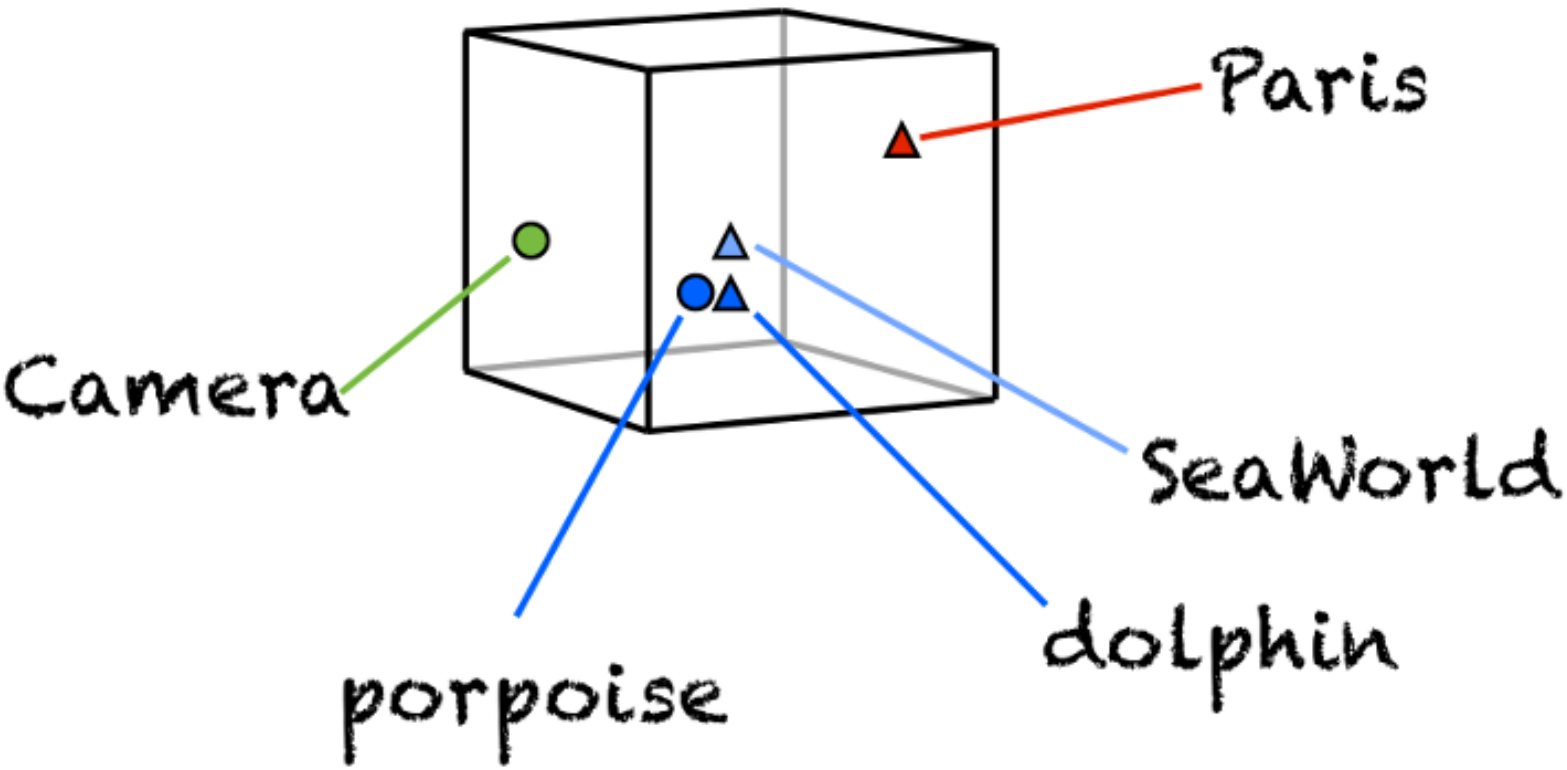
- Geometric representations of conceptual knowledge, in which,
  - Entities are represented as points. **Jurassic Park**
  - Natural properties correspond to convex regions. **Film**
  - Dimensions of the space correspond to features. **Scary**



What it all Means?



Entities in Low Dimensions - Word Embeddings



Building Blocks of Our Model

Jurassic Park (Q167726)

1993 science fiction adventure movie edit

In more languages Configure

Language	Label	Description	Also known as
English	Jurassic Park	1993 science fiction adventure movie	
Scots	Jurassic Park	No description defined	
ਪੰਜਾਬੀ	No label defined	No description defined	
Welsh	Jurassic Park	No description defined	

All entered languages

Statements

followed by — The Lost World: Jurassic Park — edit
0 references
add reference
add

instance of — film — edit
1 reference
add

director — Steven Spielberg — edit
9 references
add

producer — Kathleen Kennedy — edit

film (Q11424)

sequence of images that give the impression of movement edit
movie | motion picture | cinematic work | films

In more languages Configure

Language	Label	Description	Also known as
English	film	sequence of images that give the impression of movement	movie motion picture cinematic work films
Scots	film	No description defined	
ਪੰਜਾਬੀ	No label defined	No description defined	
Welsh	ffilm	No description defined	

All entered languages

Statements

topic's main Wikimedia portal — Portal:Film — edit
0 references
add reference
add

image — Hapci-fr.gif — edit
0 references
add reference
add

subclass of — work of art — edit
0 references

Building Blocks of Our Model

work of art (Q838948)

aesthetic physical item or artistic creation
artwork | piece of art

edit

In more languages Configure

Language	Label	Description	Also known as
English	work of art	aesthetic physical item or artistic creation	artwork piece of art
Scots	work o airt	No description defined	
ਪੰਜਾਬੀ	No label defined	No description defined	
Welsh	gwaith celf	No description defined	celfwaith

All entered languages

Statements

OpenStreetMap tag or key	Tag:tourism=artwork 0 references add reference	edit
	add	
part of	art 0 references add reference	edit
	add	
subclass of	work 0 references add reference	edit

work (Q386724)

distinct intellectual or artistic creation
creation

edit

In more languages Configure

Language	Label	Description	Also known as
English	work	distinct intellectual or artistic creation	creation
Scots	creative wirk	No description defined	
ਪੰਜਾਬੀ	No label defined	No description defined	
Welsh	gwaith	No description defined	

All entered languages

Statements

subclass of	artificial entity 0 references add reference	edit
	add	
topic's main category	Category:Works 0 references add reference	edit
	add	
equivalent class	http://dbpedia.org/ontology/Work described at URL: http://mappings.dbpedia.org/index.php/OntologyClass:Work retrieved: 16 June 2015	edit

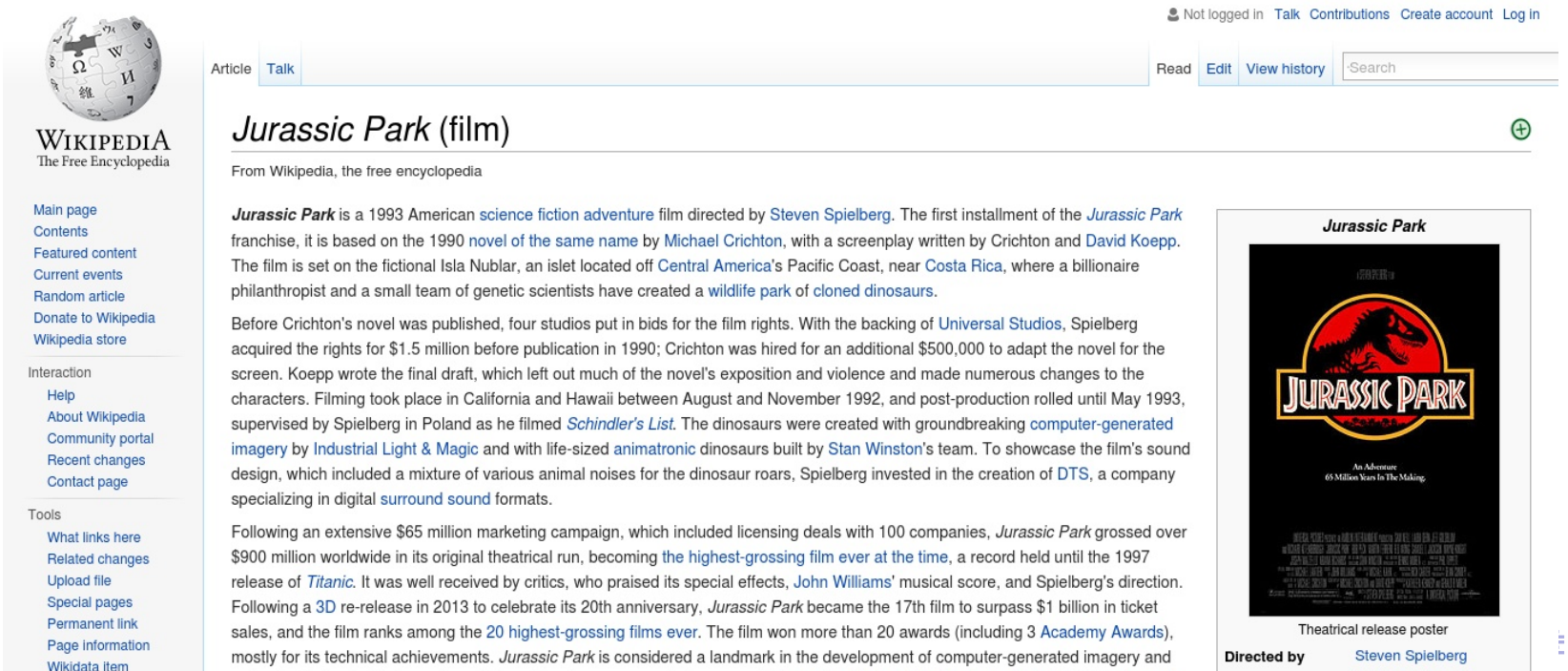
The Input - Text Data from Wikipedia

Raw Context Information to Build Co-occurrence Statistics

...his bioengineering company, InGen, have created a theme park called {**Jurassic Park**} *on Isla Nublar, a tropical Costa Rican island populated with...*

entity

In addition,



Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item

Jurassic Park (film)

From Wikipedia, the free encyclopedia

Jurassic Park is a 1993 American science fiction adventure film directed by Steven Spielberg. The first installment of the *Jurassic Park* franchise, it is based on the 1990 novel of the same name by Michael Crichton, with a screenplay written by Crichton and David Koepp. The film is set on the fictional Isla Nublar, an islet located off Central America's Pacific Coast, near Costa Rica, where a billionaire philanthropist and a small team of genetic scientists have created a wildlife park of cloned dinosaurs.

Before Crichton's novel was published, four studios put in bids for the film rights. With the backing of Universal Studios, Spielberg acquired the rights for $1.5 million before publication in 1990; Crichton was hired for an additional $500,000 to adapt the novel for the screen. Koepp wrote the final draft, which left out much of the novel's exposition and violence and made numerous changes to the characters. Filming took place in California and Hawaii between August and November 1992, and post-production rolled until May 1993, supervised by Spielberg in Poland as he filmed *Schindler's List*. The dinosaurs were created with groundbreaking computer-generated imagery by Industrial Light & Magic and with life-sized animatronic dinosaurs built by Stan Winston's team. To showcase the film's sound design, which included a mixture of various animal noises for the dinosaur roars, Spielberg invested in the creation of DTS, a company specializing in digital surround sound formats.

Following an extensive $65 million marketing campaign, which included licensing deals with 100 companies, *Jurassic Park* grossed over $900 million worldwide in its original theatrical run, becoming the highest-grossing film ever at the time, a record held until the 1997 release of *Titanic*. It was well received by critics, who praised its special effects, John Williams' musical score, and Spielberg's direction. Following a 3D re-release in 2013 to celebrate its 20th anniversary, *Jurassic Park* became the 17th film to surpass $1 billion in ticket sales, and the film ranks among the 20 highest-grossing films ever. The film won more than 20 awards (including 3 Academy Awards), mostly for its technical achievements. *Jurassic Park* is considered a landmark in the development of computer-generated imagery and

Jurassic Park

Theatrical release poster

Directed by Steven Spielberg

Symbols [edit]

Mascot [edit]

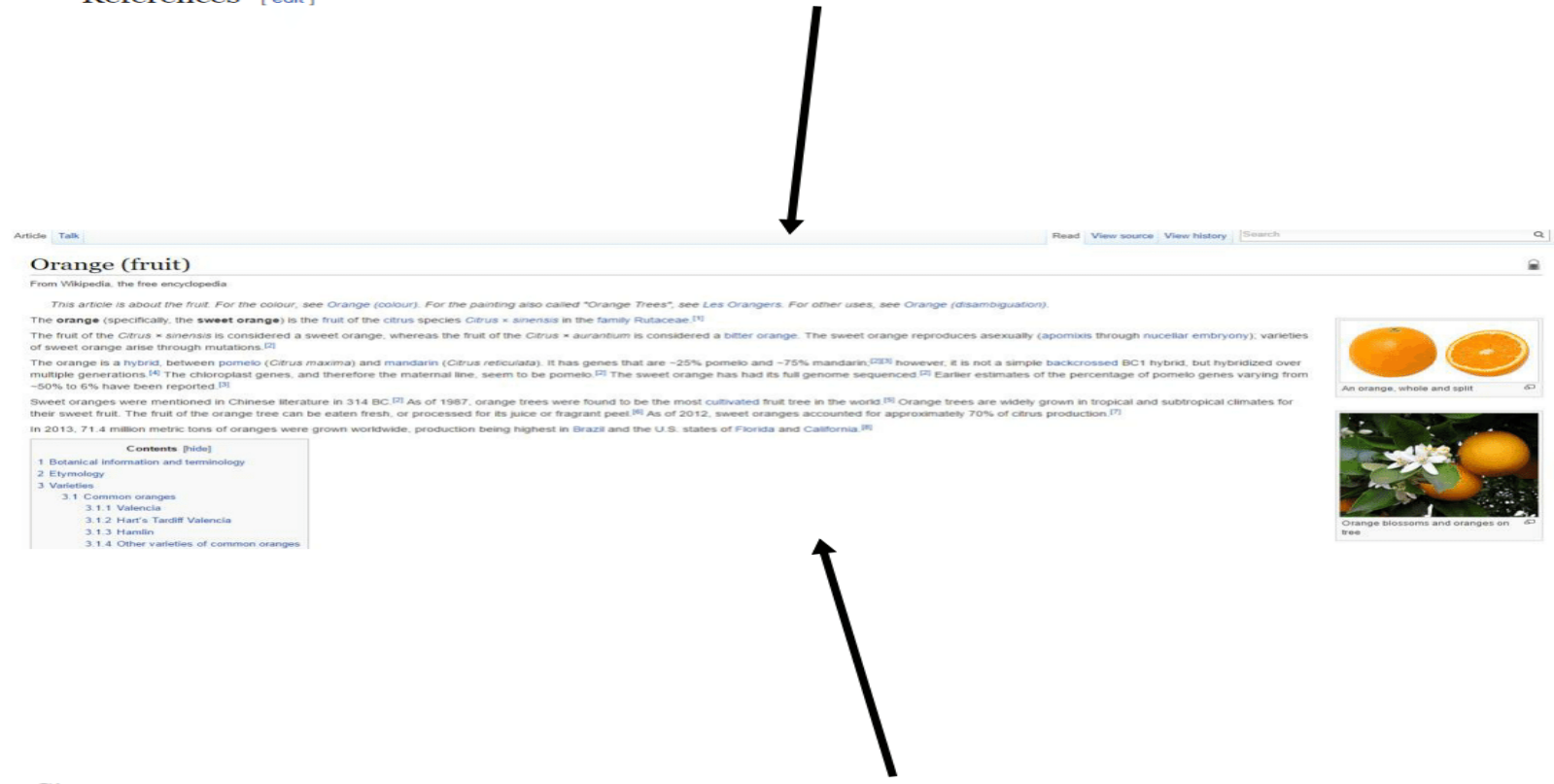
The official mascot of this World Cup was ***Naranjito***, an anthropomorphised orange, a typical fruit in Spain, wearing the kit of the host's national team. I

Football in action (*fútbol en acción*) was the name of the animated series first performed in 1982 on public broadcaster RTVE. Chapters had a duratic adventures and World Cup of 82. *Naranjito* was accompanied by other characters, as his girlfriend *Clementine*, his friend *citronio* and *Imarchi* the robot

Match ball [edit]

The match ball for 1982 World Cup, manufactured by Adidas, was the Tango España.

References [edit]



Article Talk Read View source View history Search

Orange (fruit)

From Wikipedia, the free encyclopedia

This article is about the fruit. For the colour, see Orange (colour). For the painting also called "Orange Trees", see Les Orangers. For other uses, see Orange (disambiguation).

The **orange** (specifically, the **sweet orange**) is the fruit of the citrus species *Citrus × sinensis* in the family Rutaceae.[1]

The fruit of the *Citrus × sinensis* is considered a sweet orange, whereas the fruit of the *Citrus × aurantium* is considered a bitter orange. The sweet orange reproduces asexually (apomixis through nucellar embryony); varieties of sweet orange arise through mutations.[2]

The orange is a hybrid, between pomelo (*Citrus maxima*) and mandarin (*Citrus reticulata*). It has genes that are ~25% pomelo and ~75% mandarin,[2][3] however, it is not a simple backcrossed BC1 hybrid, but hybridized over multiple generations.[4] The chloroplast genes, and therefore the maternal line, seem to be pomelo.[2] The sweet orange has had its full genome sequenced.[2] Earlier estimates of the percentage of pomelo genes varying from ~50% to 6% have been reported.[3]

Sweet oranges were mentioned in Chinese literature in 314 BC.[2] As of 1987, orange trees were found to be the most cultivated fruit tree in the world.[5] Orange trees are widely grown in tropical and subtropical climates for their sweet fruit. The fruit of the orange tree can be eaten fresh, or processed for its juice or fragrant peel.[6] As of 2012, sweet oranges accounted for approximately 70% of citrus production.[7]

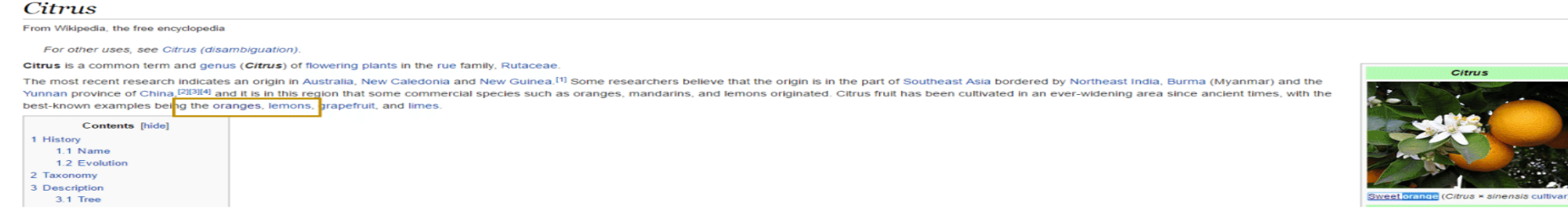
In 2013, 71.4 million metric tons of oranges were grown worldwide, production being highest in Brazil and the U.S. states of Florida and California.[8]

Contents [hide]

1 Botanical information and terminology
2 Etymology
3 Varieties
  3.1 Common oranges
    3.1.1 Valencia
    3.1.2 Hart's Tardiff Valencia
    3.1.3 Hamlin
    3.1.4 Other varieties of common oranges

An orange, whole and split

Orange blossoms and oranges on tree



Citrus

From Wikipedia, the free encyclopedia

For other uses, see Citrus (disambiguation).

Citrus is a common term and genus (***Citrus***) of flowering plants in the rue family, Rutaceae.

The most recent research indicates an origin in Australia, New Caledonia and New Guinea.[1] Some researchers believe that the origin is in the part of Southeast Asia bordered by Northeast India, Burma (Myanmar) and the Yunnan province of China,[2][3][4] and it is in this region that some commercial species such as oranges, mandarins, and lemons originated. Citrus fruit has been cultivated in an ever-widening area since ancient times, with the best-known examples being the oranges, lemons, grapefruit, and limes.

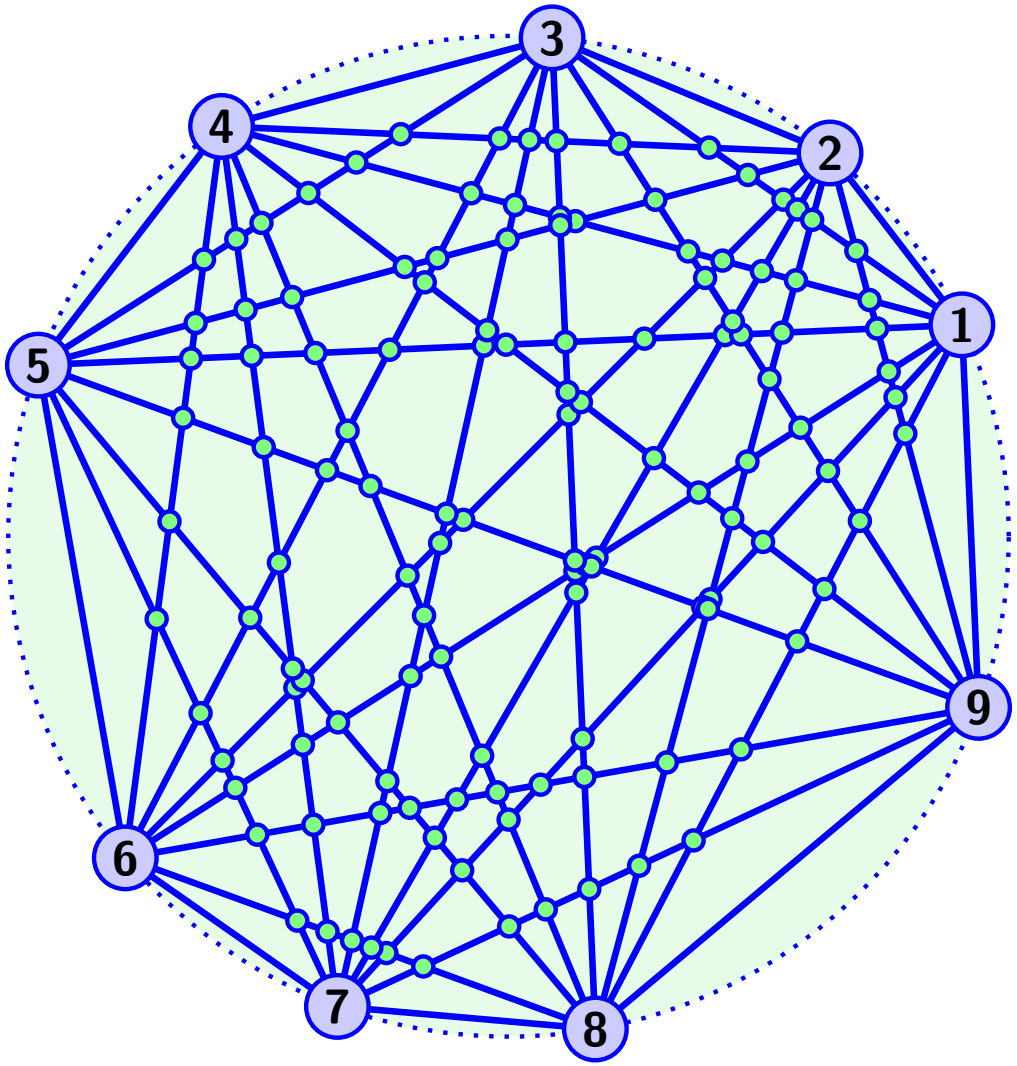
Contents [hide]

1 History
  1.1 Name
  1.2 Evolution
2 Taxonomy
3 Description
  3.1 Tree

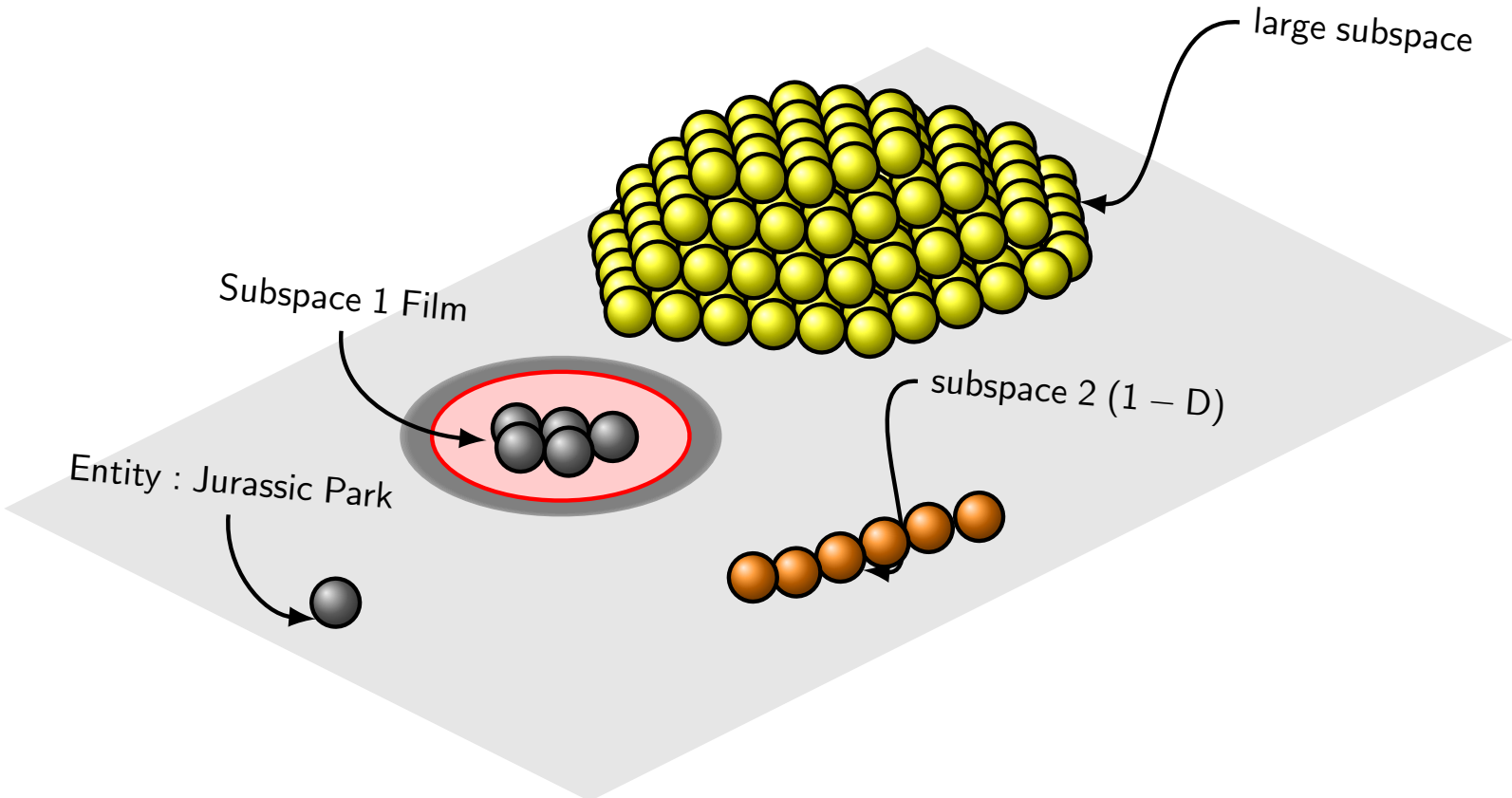
Citrus

Sweet orange (*Citrus × sinensis* cultivar

Graph Structure of Wikipedia



Output

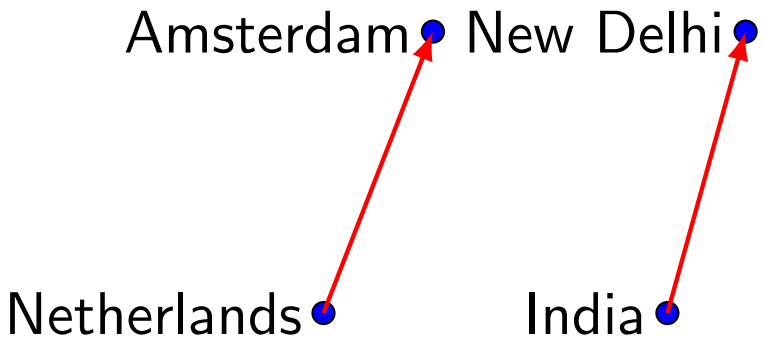

large subspace
Subspace 1 Film
subspace 2 (1 – D)
Entity : Jurassic Park

What Others Have Done So Far?

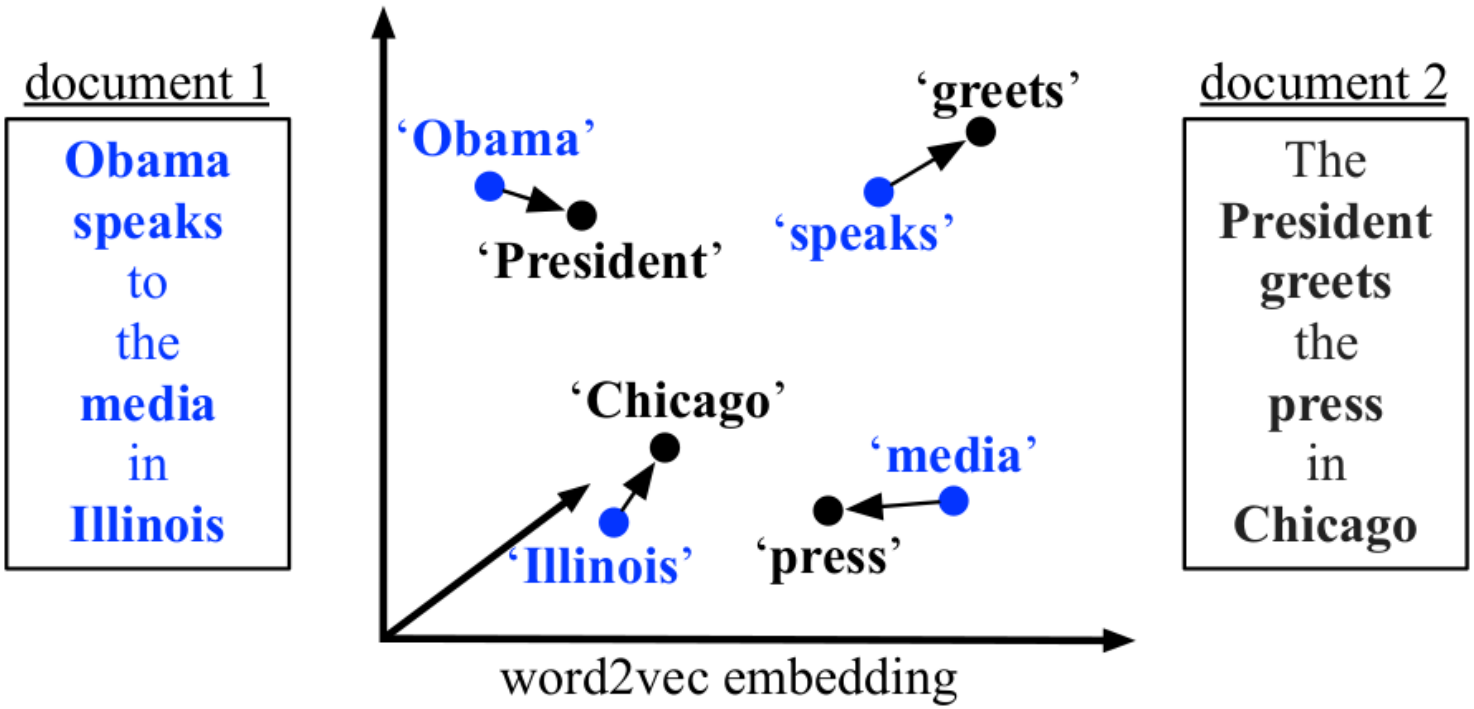
Word Embedding Models

Word embeddings are vector space representations which are used to model the meaning of words.

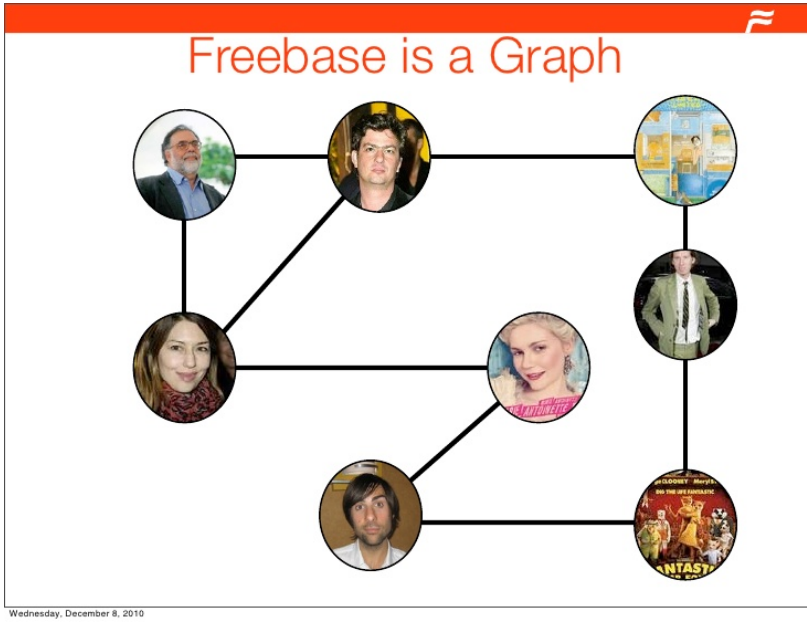
1. Skip-gram model [1] tries to find word vectors that can be used to predict the probability of seeing a context word, given an occurrence of the word being modelled
2. The continuous bag-or-words (CBOW) model focuses on the probability of seeing the word being modelled, given the occurrence of a context word.
3. The GloVe model [2], we will discuss in detail.

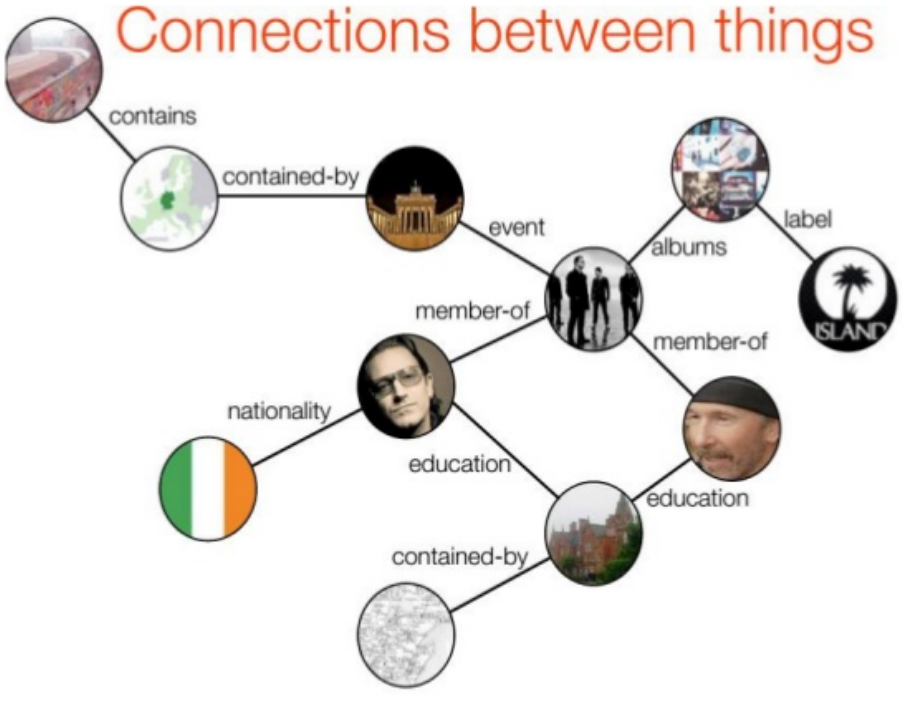


Word Embeddings Example



Knowledge Graphs



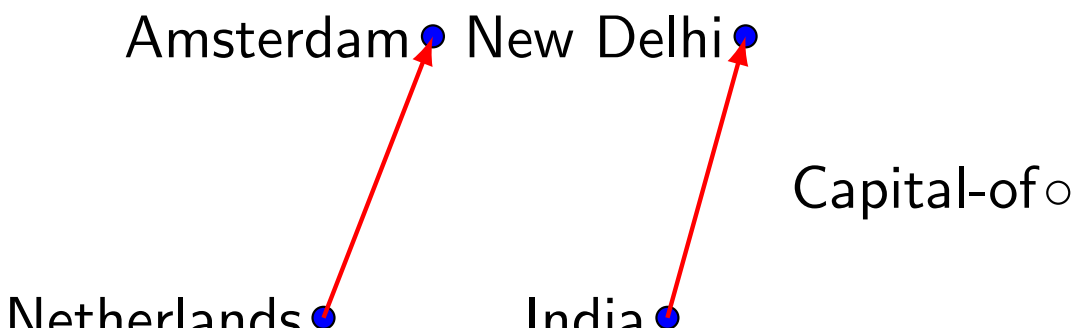


What Others Have Done So Far?

Knowldege Graph Embedding Models

Knowledge graphs such as Freebase and Wikidata exist as (subject, predicate, object) collection.

1. SE model [4] entities are represented as vectors. Relation using two matrices and finding least distance between entities and relations i.e. triples.
2. TransE model [3] represented relation as a vector and is suitable for one-to-one-relation.
3. TransH model [5] hyper-plane and the relation vector is associated with each type.
4. pTransE model [6] uses text content and a knowledge graph.



Knowledge Graph Embedding Example

Knowledge Graph Embedding Example

The Big Bang Theory - Season 10 - CBS.com
https://www.cbs.com/shows/big_bang_theory/
The **Big Bang Theory** - Monday nights on CBS. Watch full episodes of The **Big Bang Theory**, view video clips and browse photos on CBS.com.

The Big Bang Theory (TV Series 2007–) - IMDb
www.imdb.com/title/tt0898266/
Comedy · A woman who moves into an apartment across the hall from two brilliant but socially awkward physicists shows them how little they know about life outside of the laboratory.

The Big Bang Theory - All 4
www.channel4.com/programmes/the-big-bang-theory
5 days ago - Leonard and Sheldon know everything about physics, but are clueless about people...

The Big Bang Theory (@bigbangtheory) | Twitter
https://twitter.com/bigbangtheory?lang=bg
6 321 туита • 2 465 снимки/видеоклипа • 4,77 млн. последователя. Виж последните туитове на The **Big Bang Theory** (**@bigbangtheory**)

Things you might not know about the ladies of 'The Big Bang Theory ...
www.foxnews.com/.../2018/.../things-might-not-know-about-ladies-big-bang-theory.ht...
1 day ago - You may know the entire speedy "**Big Bang Theory**" theme song by heart, but that doesn't mean you know everything there is to know about the lovely ladies from the series. Here are some surprising facts and things you might have missed about the female stars on CBS' hit show:

Cast

View 10+ more


Kaley Cuoco
Penny


Jim Parsons
Sheldon Cooper


Johnny Galecki
Leonard Hofstadter


Kunal Nayyar
Raj Koothrappali


Simon Helberg
Howard Wolowitz

Profiles

Twitter YouTube Facebook Instagram

People also search for

View 15+ more

Young Sheldon
Since 2017

Two and a Half Men
2003 – 2015

Friends
1994 – 2004

2 Broke Girls
2011 – 2017

How I Met Your Mother
2005 – 2014

Feedback

Computational Model

Our model is expressed as:

Basic Formulation - A Joint Optimization Model

$$J = \alpha J_{\text{text}} + (1 - \alpha)(J_{\text{type}} + J_{\text{rel}}) + \beta J_{\text{reg}}$$

Details

- $\alpha \in [0, 1]$
- $\beta \in [0, +\infty[$
- J_{text} control the representation of entities based on their textual representation
- J_{type} impose that entities of the same type belong to the same subspace
- J_{rel} control alignment between subspaces
- J_{reg} regularization component to automatically discover the low-dimensional subspace representation

GloVe Model [2]

Formulation

$$J_{\text{text}} = \sum_{e \in E} \sum_{t_j \in W_{e_i}} f(y_{ji})(p_{e_i}.w_j + b_i + b_j - \log(y_{ij}))^2$$

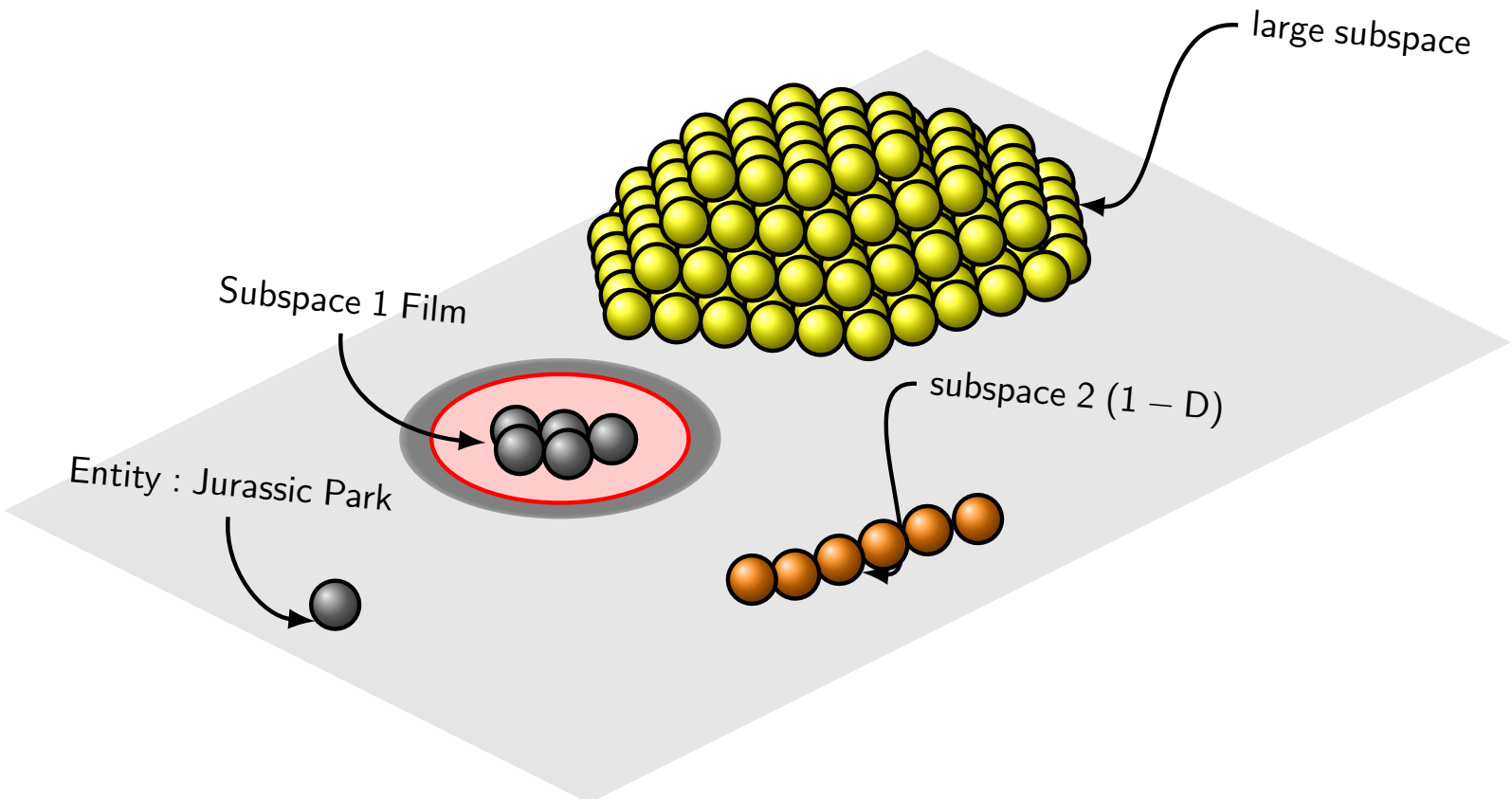
Decay Function

$$\begin{cases} f(y_{ij}) = \left(\frac{y_{ij}}{y_{\max}}\right)^{\alpha} & \text{if } y_{ij} < y_{\max} \\ 1 & \text{otherwise} \end{cases}$$

What it does?

- This component helps our model position similar entities close to each other based on the word-word statistical co-occurrence information.
- Discovers salient features as directions in space

GloVe Constraints



Subspace Constraints

What it does?

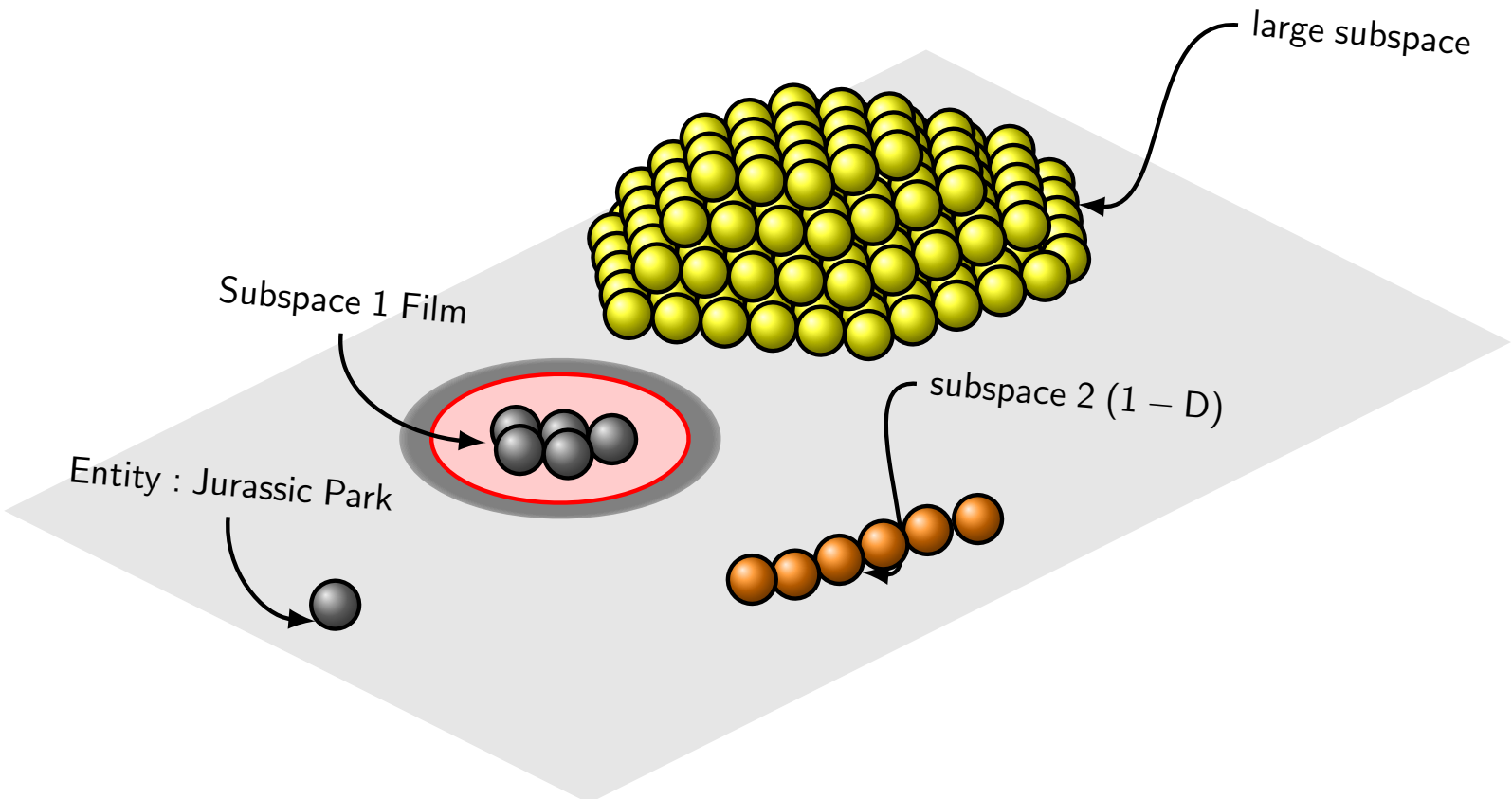
- All entities of a given type belong to the same subspace.

Formulation

$$J_{\text{type}} = \sum_{s \in S} \sum_{e \in E_s} ||p_e - \sum_{j=0}^{n} \lambda_j^{e,s} p_j^s||^2$$

- S is the total number of semantic types
- E_s entities in that semantic type
- p_e point representation of an entity
- λ is the coefficient
- p_j^s matrix of other points
- n dimension of the space

Subspace Constraints



Relation Constraints

Example: Steven Spielberg is the **director of** Jurassic Park

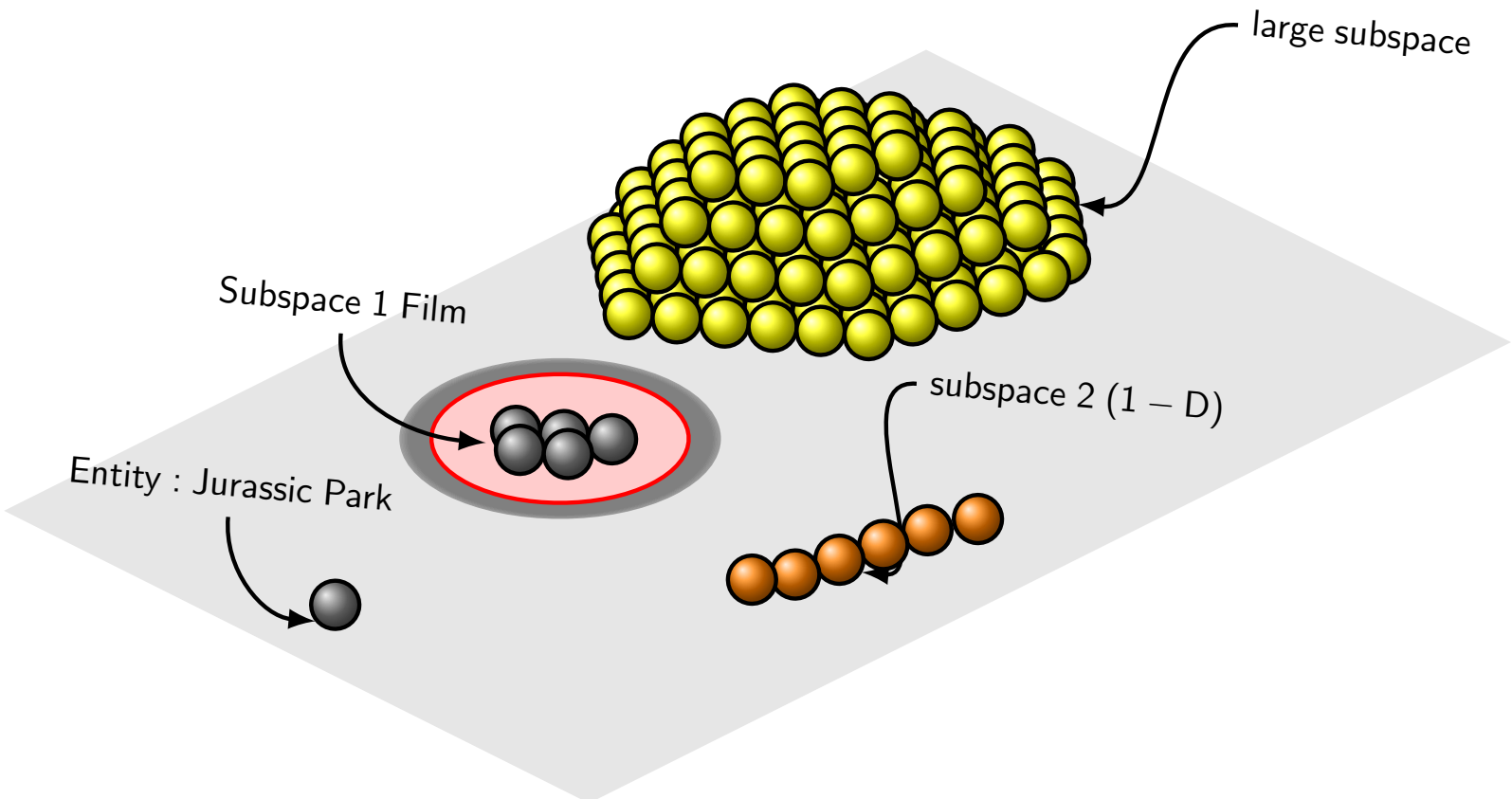
What it does?

- Align subspaces corresponding to different types

Formulation

$$J_{\text{rel}} = \sum_{k \in R} \sum_{p \in P_{e,k}} ||p - \sum_{j=0}^{n} \mu_j^{e,k} q_k^{e,k}||^2 + \sum_{p \in P_{k,f}} ||p - \sum_{j=0}^{n} \mu_j^{k,f} q_j^{k,f}||^2$$

Relation Constraints



Nuclear Norm Regularization

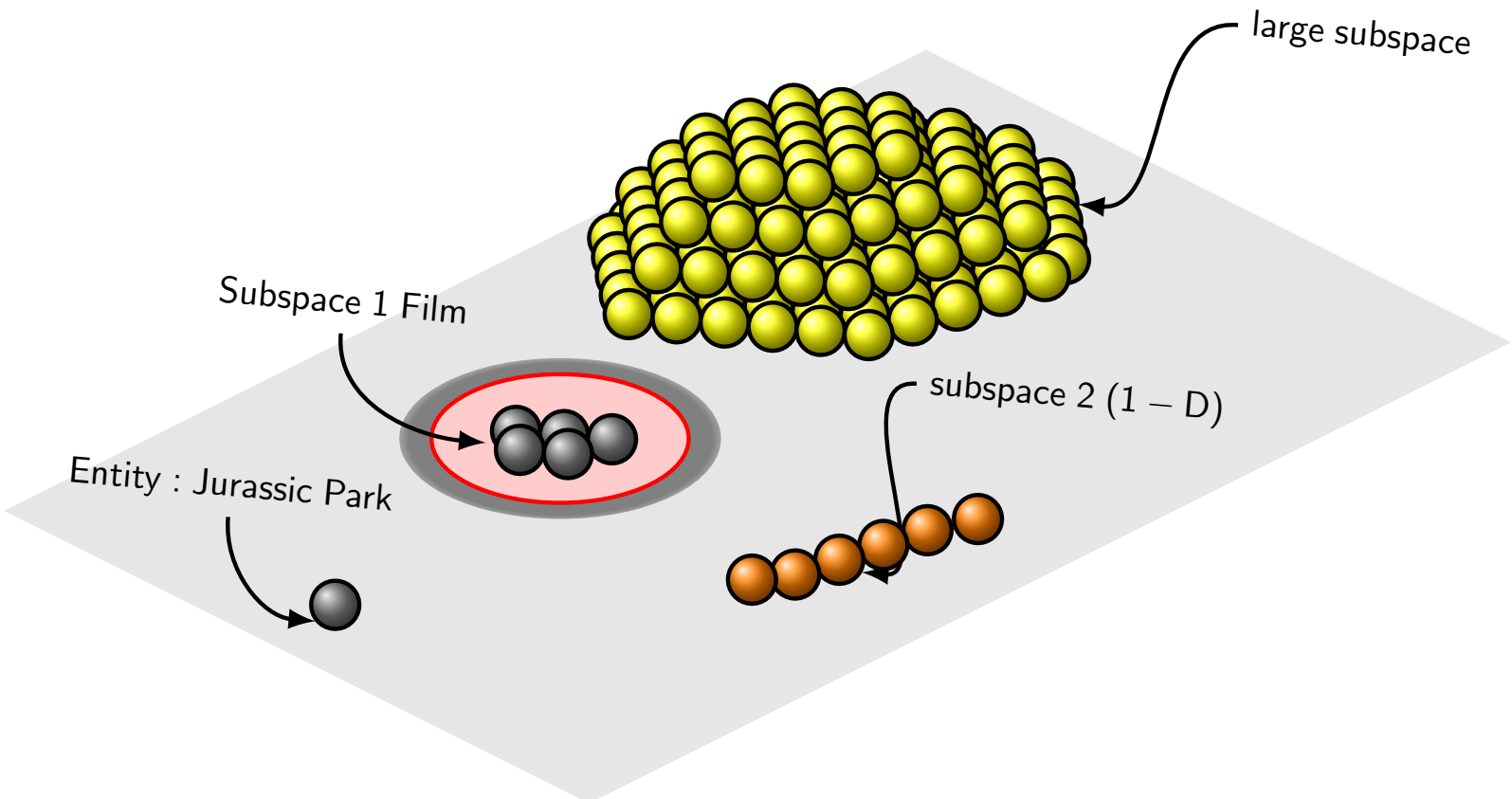
What it does?

- Penalize high-rank subspace matrices
- Finds a low-rank solution with features that are sufficient to describe that subspace

Formulation

$$J_{\text{reg}} = \sum_{s \in S} ||M_s||$$

Regularizer Constraints



Evaluation Tasks - Objective

What we wish to attain?

- **Ranking:** To what extent important features of a given semantic type can indeed be modelled as directions in the associated subspace
- **Induction:** Assesses to what extent we can use these representations to find new instances of a given concept, given only a few example instances
- **Analogy Making:** Evaluating how well the different subspaces are aligned
- **Knowledge graph Embedding:** Our main aim in this task is to evaluate how well different subspaces are aligned

Evaluation Tasks

Description

- **Ranking:** We obtain numerical features, suitable for ranking from Wikidata. *These numerical values were not considered in learning the space*
- **Induction:** Given a number of entities of the same type which have some property in common, the task we consider is to identify other entities that also have this property
- **Word Analogy:** "a is to b what c is to ...", which is a standard evaluation task for word embeddings
- **Link Prediction:** Given an entity e and a relation r, the aim is either to find an entity f such that (e, r, f) or to find an entity f such that (f, r, e)
- **Triple Classification:** Judge whether a given triplet (e, r, f) is correct or not, i.e. whether entities e and f are in relation r with each other

	Ranking	Induction			Analogy
	ρ	MAP	P@5	MRR	Acc.
Skip-gram	0.155	0.176	0.356	0.505	0.184
CBOW	0.159	0.182	0.350	0.500	0.213
RESCAL	0.081	0.020	0.189	0.423	0.371
TransE	0.110	0.060	0.200	0.451	0.382
TransH	0.142	0.072	0.210	0.415	0.382
TransR	0.100	0.102	0.302	0.489	0.378
CTransR	0.122	0.132	0.323	0.499	0.402
$\text{pTransE}_{\text{anch}}$	0.099	0.101	0.301	0.488	0.476
$\text{pTransE}_{\text{art}}$	0.202	0.218	0.475	0.751	0.512
$\text{pTransE}_{\text{full}}$	0.213	0.224	0.490	0.756	0.532
$\text{EECS}_{\text{full}}$	**0.319**	**0.231**	**0.609**	**0.883**	0.591
$\text{EECS}_{\text{no rel}}$	0.301	0.229	0.588	0.868	0.552
$\text{EECS}_{\text{no type}}$	0.266	0.225	0.585	0.854	0.549
$\text{EECS}_{\text{no NN}}$	0.258	0.220	0.581	0.843	0.545
$\text{EECS}_{\text{text}}$	0.254	0.218	0.579	0.831	0.540
$\text{EECS}_{\text{type-comb}}$	0.312	**0.231**	0.601	**0.883**	**0.595**
$\text{EECS}_{\text{type-dist}}$	0.295	**0.231**	0.585	0.858	0.550
$\text{EECS}_{\text{rel-dim}}$	0.309	0.225	0.585	0.859	0.551
$\text{EECS}_{\text{rel-dist}}$	0.299	0.225	0.585	0.855	0.549

Qualitative Results

Table: Five lowest ranked entities for a number of ranking problem instances.

Population	Inception	Date of Birth
Malta	General Electric	Valmiki
Bermuda	IBM	Jesus Christ
Monaco	Hewlett Packard	Cleopatra
San Marino	Microsoft	Ptolemy
Barbados	Oracle Corporation	Plato

Table: Five highest ranked entities for a number of ranking problem instances.

Population	Inception	Date of Birth
China	Alphabet Inc.	Prince George of Cambridge
India	Tencent Holdings	Isabela Moner
USA	Facebook, Inc.	Justin Bieber
Soviet Union	Uber	Lionel Messi
Brazil	Amazon.com	Kim Kardashian

Ordering Entities

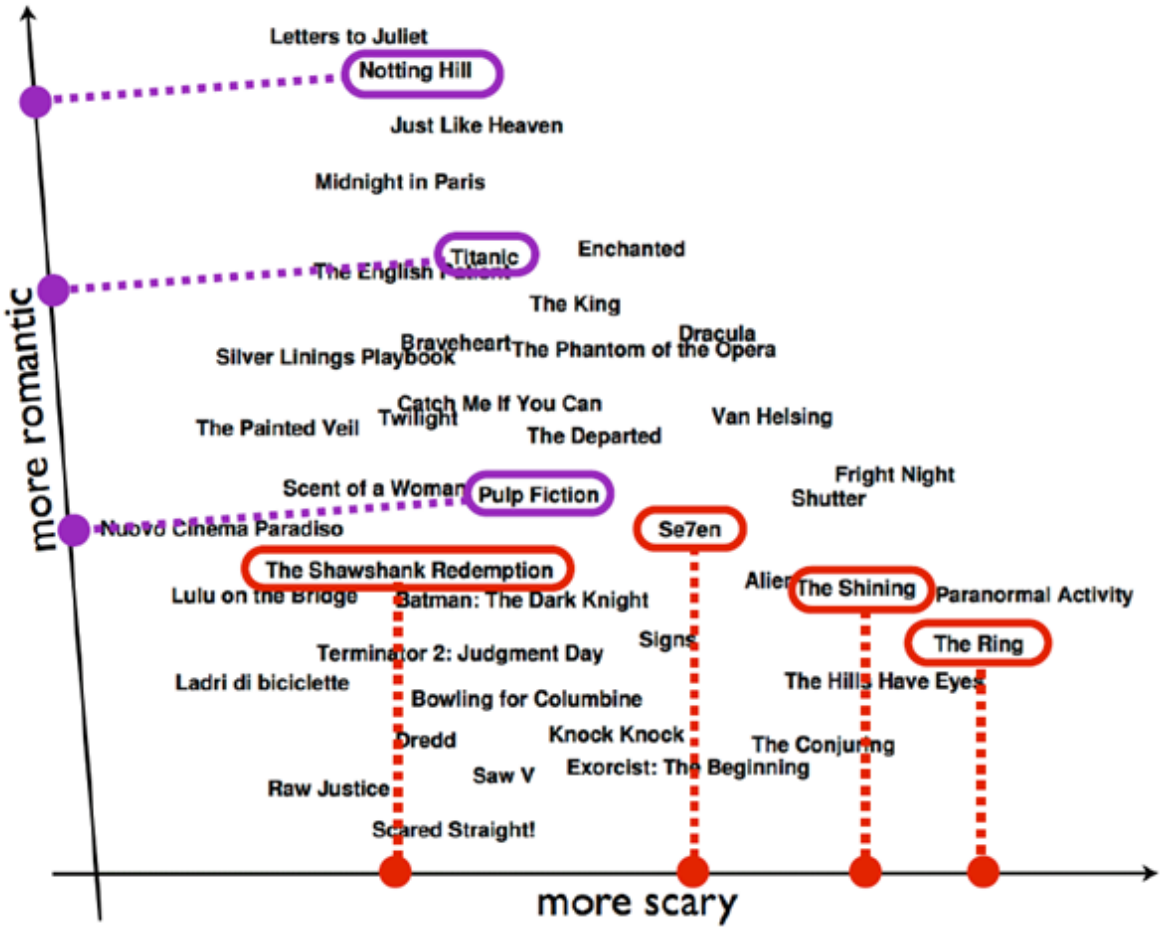


Table: Link prediction and Triple classification results.

Models	Link Prediction (FB15k)		Triple Classification	
	Mean Rank	HITS@10	FB13	FB15k
RESCAL	683	44.1	65.3	71.6
TransE	125	47.1	81.5	79.8
TransH	87	64.4	**83.3**	79.9
TransR	77	68.7	82.5	82.1
CTransR	75	70.2	-	84.3
$pTransE_{anch}$	58	84.6	73.3	74.3
$pTransE_{art}$	55	85.3	75.8	75.5
$pTransE_{full}$	51	86.4	76.3	77.4
$EECS_{full}$	48	89.7	83.1	89.6
$EECS_{no\ type}$	56	84.7	71.2	82.1
$EECS_{no\ NN}$	59	82.7	70.1	81.4
$EECS_{type\text{-}comb}$	**47**	**89.9**	**83.3**	**89.9**
$EECS_{type\text{-}dist}$	54	83.2	81.1	82.1
$EECS_{rel\text{-}dim}$	54	85.1	79.3	88.2
$EECS_{rel\text{-}dist}$	52	85.3	78.8	87.1

References I

[1] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119, 2013.

[2] J. Pennington, R. Socher, and C. D. Manning. GloVe: Global vectors for word representation. In *EMNLP*, 12:1532–1543, 2014.

[3] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

[4] A. Bordes, J. Weston, R. Collobert, and Y. Bengio. Learning structured embeddings of knowledge bases. In *AAAI*, 2011.

[5] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI* pp. 1112–1119, (2014).

[6] Z. Wang, J. Zhang, J. Feng, and Z. Chen. Knowledge graph and text jointly embedding. In *EMNLP*, pp. 1591–1601, (2014).