A broad view on word vectors

Dmitrijs Milajevs
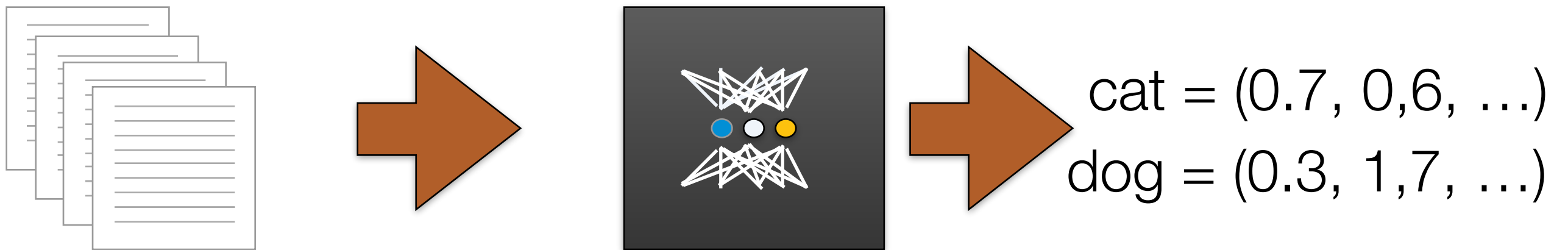
http://www.eecs.qmul.ac.uk/~dm303/bigo/
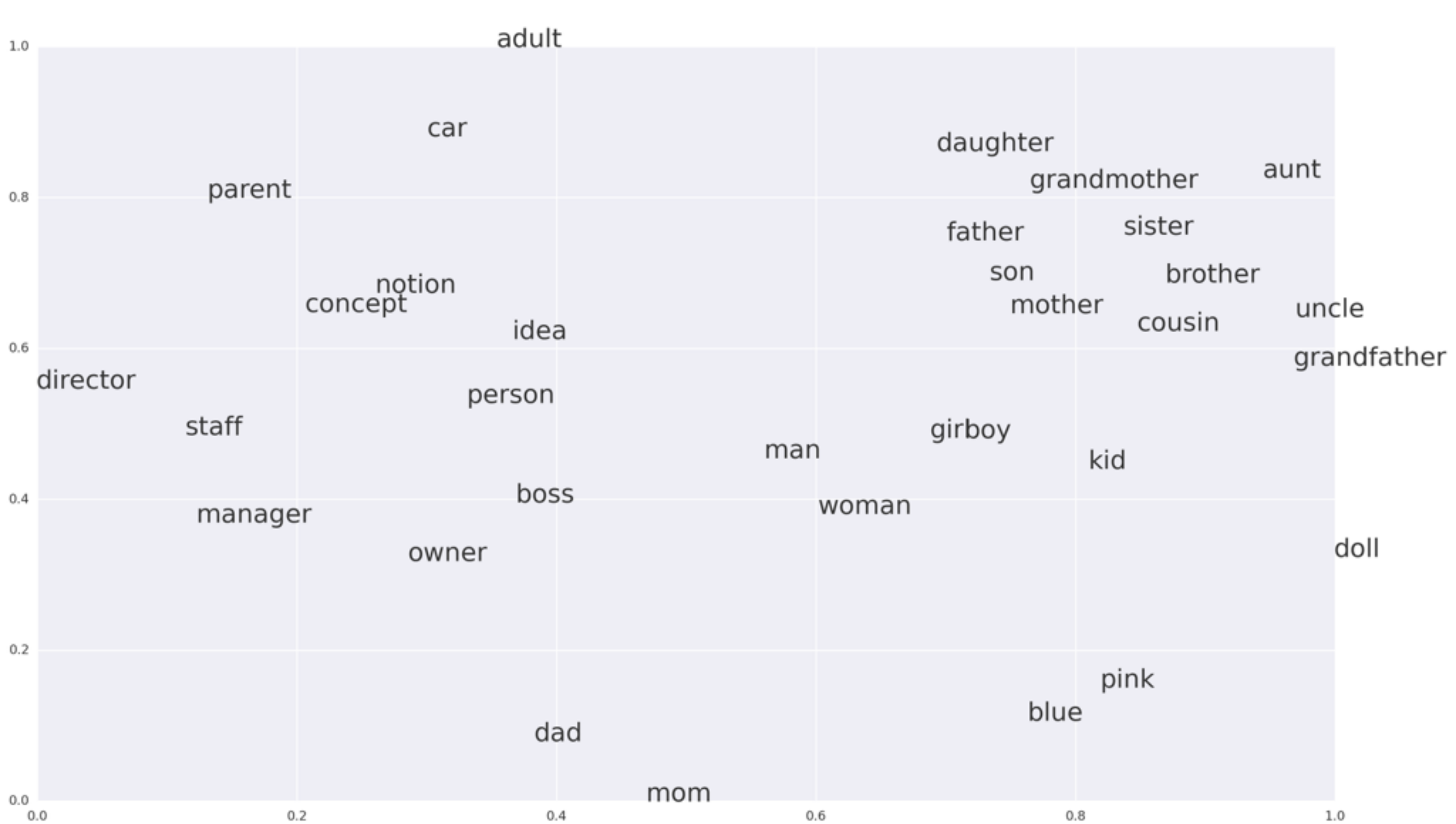
Queen Mary
University of London
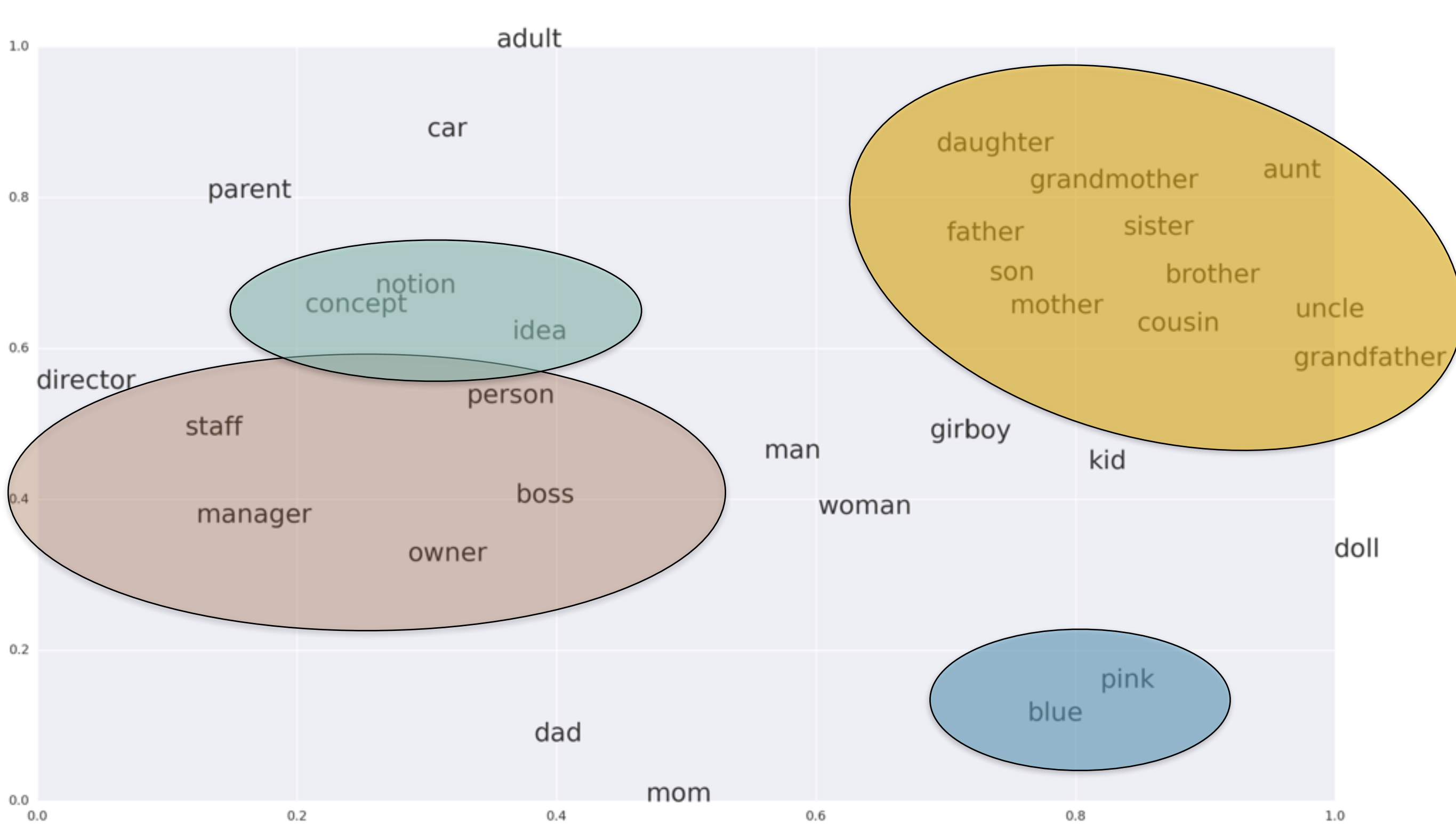
cat = (0.7, 0,6, ...)
dog = (0.3, 1,7, ...)

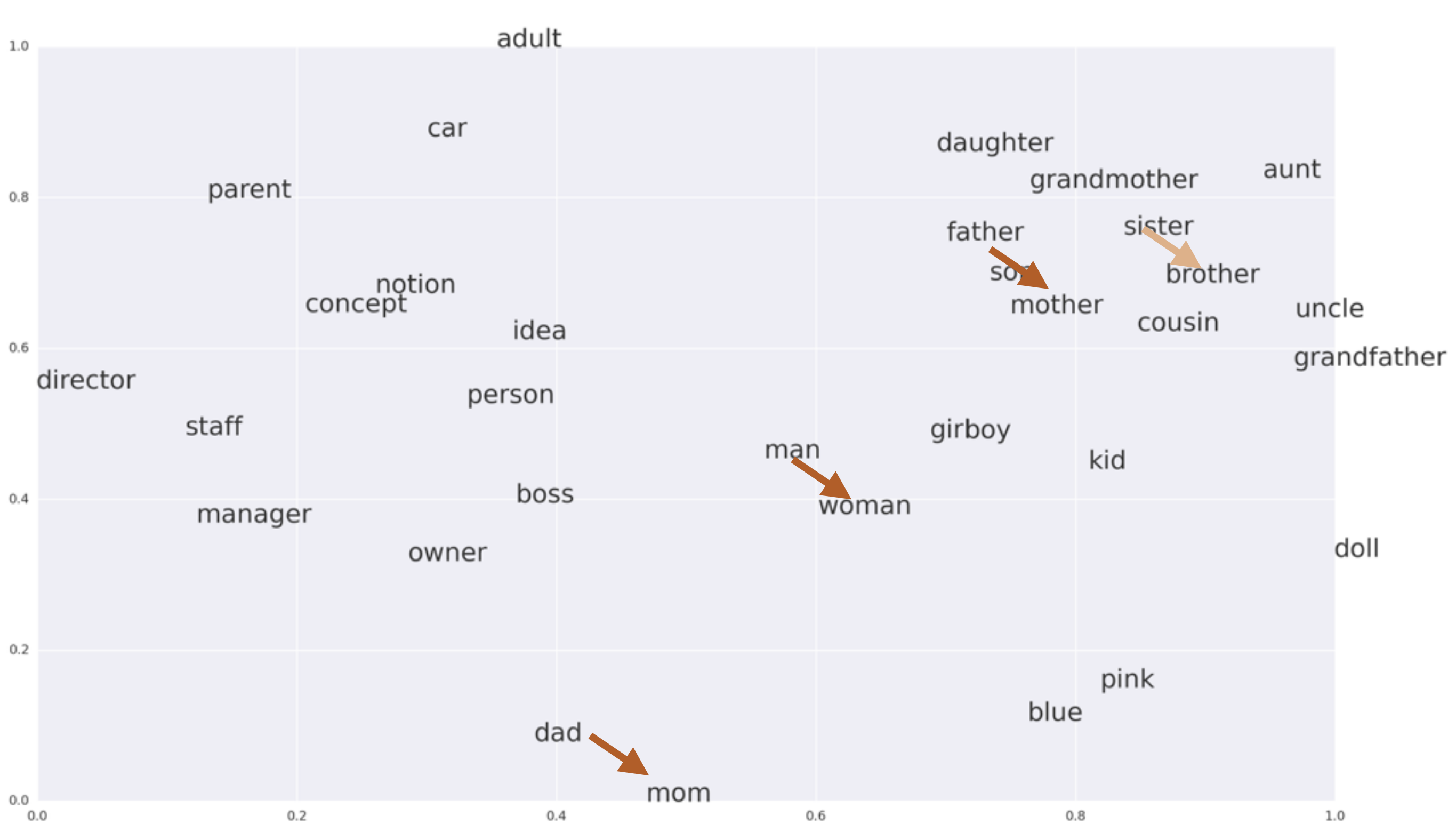Word meaning representation

word2vec in a nutshell

Locate similar words close to each other | The goal

Word clusters

Relations between words

dad to *mom* = *man* to *woman*

A lot of attention

- Deep learning

- Originated from Google, Facebook and Microsoft

- A lot of work to **understand** what's going on

- Before it was called **lexical distributional semantics**

___ - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/___
___ refers to a variety of different feelings, states, and attitudes that ranges from
interpersonal affection ("I ___ my mother") to pleasure ("I ___ that meal").
___ (disambiguation) - Falling in ___ - Biological basis - Romance

The ___ Calculator
www.___calculator.com/
Calculates the chance on a successful relationship between two people.

Urban Dictionary: ___
www.urbandictionary.com/define.php?term=___
___ is an incredibly powerful word. When you're in ___, you always want to be
together, and when you're not, you're thinking about being together because ...

5 Ways to ___ - wikiHow
www.wikihow.com › ... › Relationships › Dating › ___ and Romance
How to ___ ___ is expressed as an action and experienced as a feeling. Yet, ___
has an essence that resists defining in any single way — it encompasses ...

___ - Definition and More from the Free Merriam-Webster Dictionary
www.merriam-webster.com/dictionary/___
a feeling of strong or constant affection for a person. : attraction that includes sexual
desire : the strong affection felt by people who have a romantic relationship.
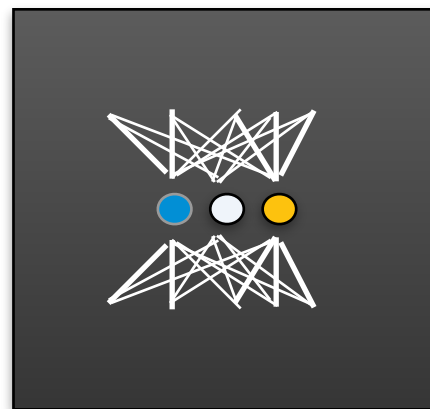
Guess the word I've blurred!

You shall know a word by the company it keeps

Text to context

A lorry might **carry** *sweet apples*

Positive contexts

carry, lorry
carry, sweet
carry, apples

carry, future
carry, tree
carry, lonely
carry, orange
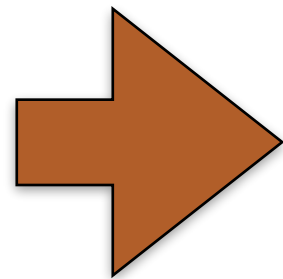
Negative contexts

P(carry|lorry) is high

P(carry|lonely) is low

Understanding what wor2vec does

- Before word2vec matrix factorisation (read SVD) was the tool to get dense vectors from a co-occurrence matrix

	might	sweet	red	...
carry	10	3	0	...
boy	43	0	12	...
potato	4	100	10	...

	a	b
carry	3543	675
boy	946	353
potato	34535	737

	might	sweet	red	...
a	-948759	34234	6547	...
b	34234	-34324	65462	...

	might	sweet	red	...
carry	12	3	0	...
boy	4	0	12	...
potato	4	10	10	...

	might	sweet	red	...
carry	12	3	0	...
boy	4	0	12	...
potato	4	10	10	...

	might	sweet	red	...
carry	10	3	0	...
boy	43	0	12	...
potato	4	100	10	...

Matrix factorisation and word2vec are related

Levy & Goldberg

Applications

- Compositional semantics:
  Nouns are vectors, adjectives are matrices

- Multimodal:
  image tagging

- Multilingual
  French words close to English *car*

- Question answering

http://www.eecs.qmul.ac.uk/~dm303/bigo/

@dimazest

dimazest@gmail.com