Interactive Analysis of Word Vector Embeddings

Florian Heimerl and **Michael Gleicher**
Department of Computer Sciences
University of Wisconsin - Madison

Interactive Analysis of Word Vector Embeddings

Florian Heimerl and Michael Gleicher
Department of Computer Sciences
University of Wisconsin - Madison

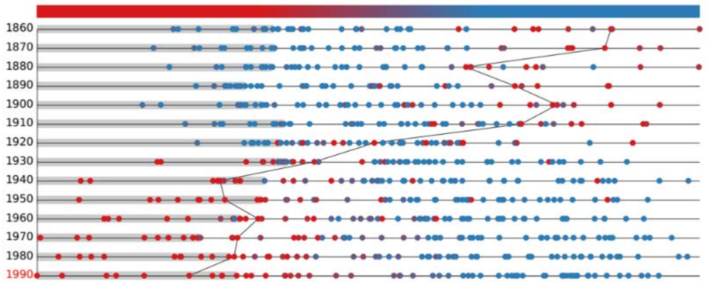
Summary:

Word vector embeddings offer unique challenges

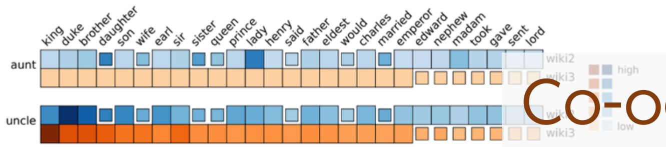
Task analysis of needs

characteristic	single target	multiple targets
similarity	(1) inspect local neighborhood	(2) compare local neighborhoods
average, offset	(3) inspect arithmetic structure	(4) compare arithmetic results
co-oc. probability	(5) analyze encoded probabilities	(6) compare probabilities
concept axis	(7) analyze vector relations	(8) compare vector relations
multiple	(9) discover interesting aspects	(10) compare interesting aspects

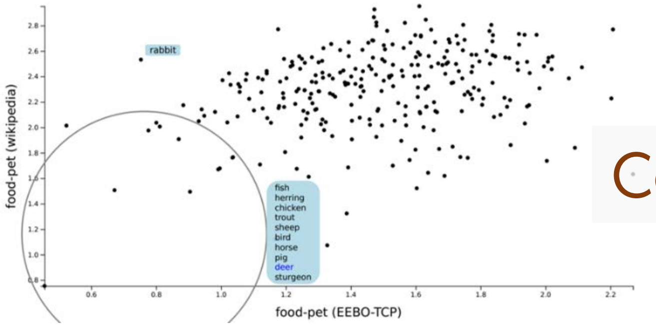
3 Designs for unmet needs



Buddy Plots



Co-occurance Matrices



Concept Axis Plots

What is an embedding?

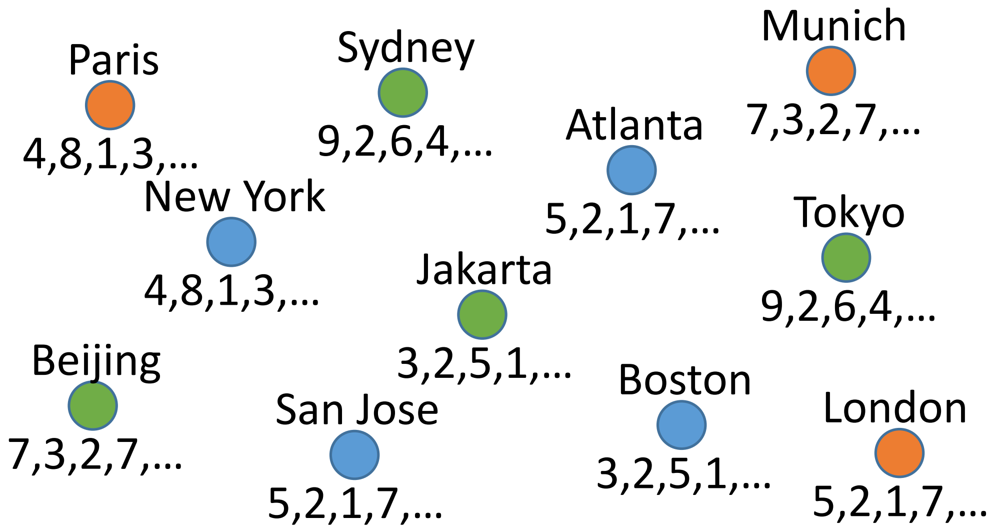
General mathematics:

Place a smaller structure into a larger structure

Computer science:

Place a discrete set of objects into a vector space

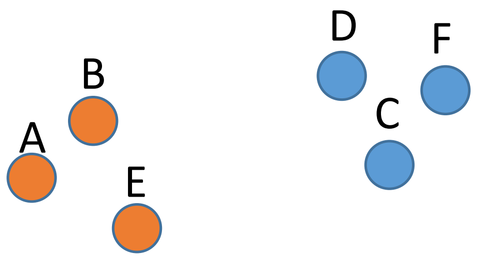
Encode relationships between objects



High Dimensional Data

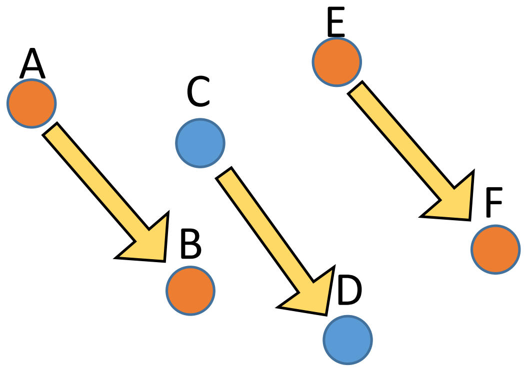
Objects ○ have associated **Vectors**

Kinds of relationships in embeddings



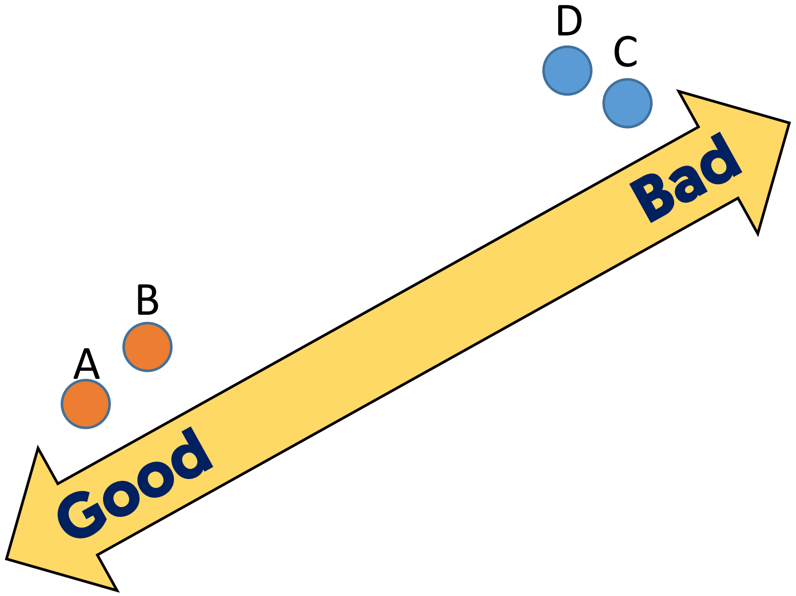
Distance

A is closer to B than to C



Linear Structure

A is to B as C is to D



Semantic Directions

A is more X than C

Relationships are interesting even if global positions are not

Word vector embeddings

Place **words** in a high-dimensional **vector** space

Words **similar in meaning** should be **close in space**

Infer similarity by **distributional semantics**:

similar context implies similar meaning

```
my pet cat is brown
my pet dog is brown
my big car is brown
```

Construct embeddings by processing a **corpus** of text

Several ways to build embeddings

Word2Vec

Skip-gram model

Neural embedding

GLoVE

Co-occurrence model

Factor matrix by optimization

Why use Word Vector Embeddings?

Learn about
Language or Corpora (Texts)

Find similar words/synonyms

Track changes of word usage

Exploring polysemy

Creating lexical resources

Evidence of bias

…

Natural Language Applications
Pre-Processing

Translation

Sentiment Analysis

Interpretation

…

Challenges of Word Vector Embeddings

Large numbers of Words
High-dimensional spaces
Complex relationships – meaningless positions

Complex processes for building embeddings
Complex downstream applications

No ground truth - subjective aspects

Why visualization?

Tasks are inherently human-centric

Variety of tasks involving interpretation
Linguistic and domain knowledge for applications

But what are those tasks?

Task Analysis: What do people do with Word Embeddings?

Literature survey

111 papers from diverse communities

Consider use cases in Linguistics, HCI, Digitial Humanities, etc.

Augment this list by extrapolation

what tasks would users want – but aren't doing yet

Use cases suggest tasks

Learn about
Language or Corpora (Texts)

Find similar words/synonyms

Track changes of word usage

Exploring polysemy

Creating lexical resources

Natural Language Applications
Pre-Processing

Translation

Sentiment Analysis

Interpretation

Interpretation

- Identify items of interest
- Probe values of interest

Evaluation

- Intrinsic (good embedding?)
- Extrinsic (applications success?)

Linguistic Tasks and Characteristics

We identified 7 distinct *linguistic tasks* within the literature

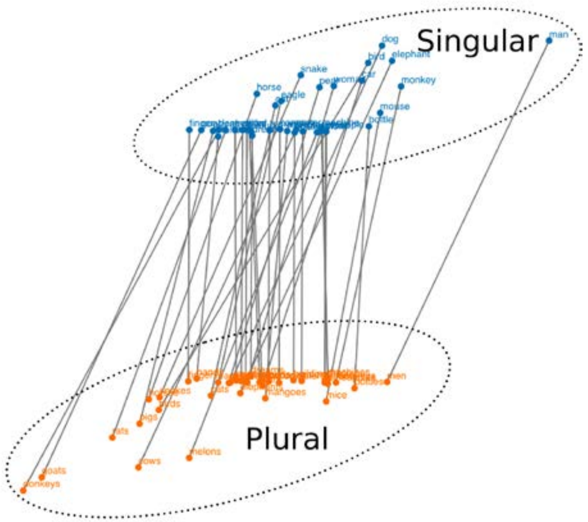
Two ways those tasks are relevant

- Automatically test embeddings through human-curated resources
- Users probe embedding

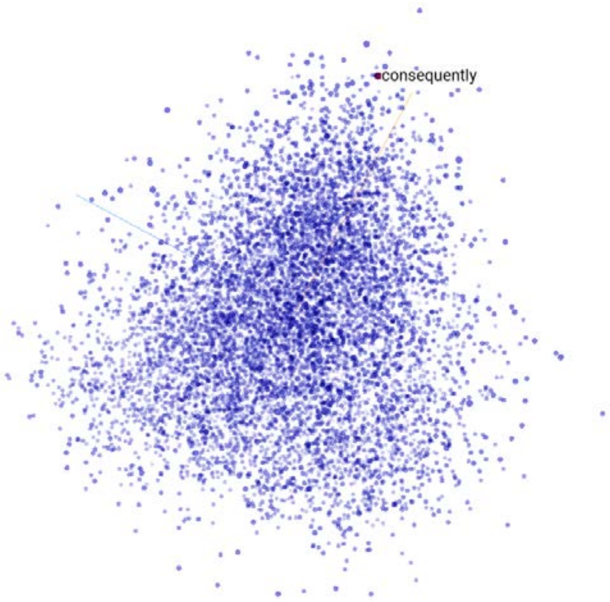
4 characteristics pertinent to those tasks: similarity, arithmetic structures, concept axis, and co-occurrences

linguistic tasks	**characteristics**	**examples**
rank word pairs	similarity	[BDK14, PSM14]
compare concepts	average, similarity	[RBS17, SLMJ15]
find analogies	offset, similarity	[SLMJ15, LG14]
view neighbors	similarity	[HLJ16, YWL* 16]
select synonyms	similarity	[BDK14, FDJ* 14]
project based on concepts	concept axis	[BCZ* 16, FRMW17]
predict contexts	co-oc. probability	[SN16, LJW* 15]

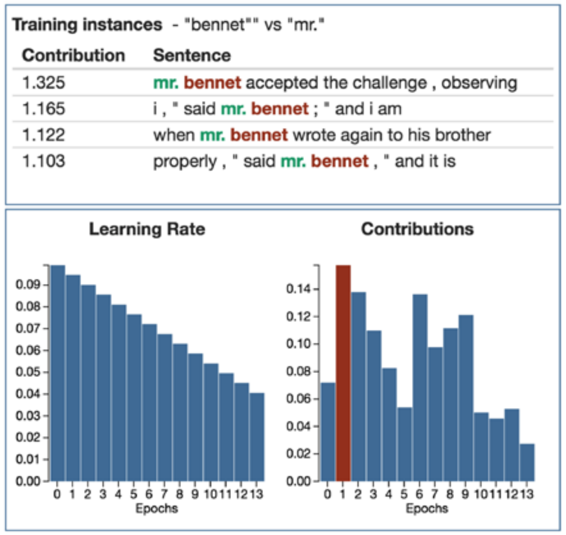
Prior Work:
Word Vector Embedding Visualizations



Liu et al., InfoVis 2017
Analogy Relationships



Smilkov et al., 2016
Distance Relationships



Rong and Adar, 2016
Training Process

Tasks and Characteristics vs. Visualizations

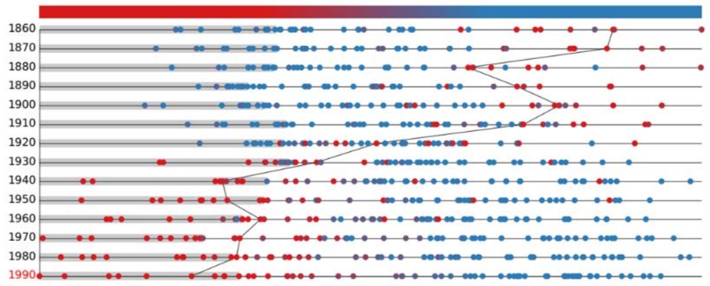
Rank word pairs	similarity
View neighbors	similarity ← **Smilkov, et al 2016**
Select synonyms	similarity
Compare concepts	average, similarity
Find analogies	offset, similarity ← **Liu, et al 2017**
Project on Concept	concept axis
Predict Contexts	co-oc. probability

Tasks and Characteristics vs. Visualizations

Rank word pairs	similarity	
View neighbors	**similarity**	Smilkov, et al 2016
Select synonyms	**similarity**	
Compare concepts	**average, similarity**	
Find analogies	offset, similarity	Liu, et al 2017
Project on Concept	**concept axis**	
Predict Contexts	**co-oc. probability**	

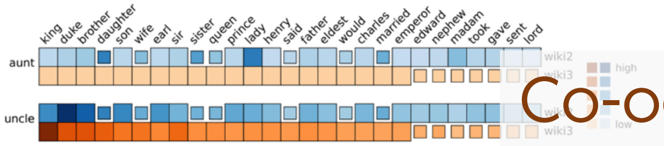
Tasks we seek designs for in this paper

1. Similarities (local distances)



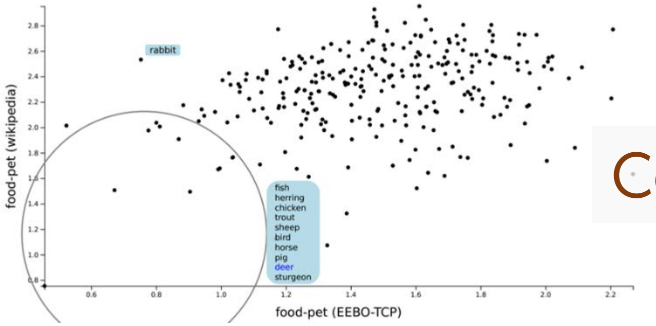
Buddy Plots

2. Co-Occurrances



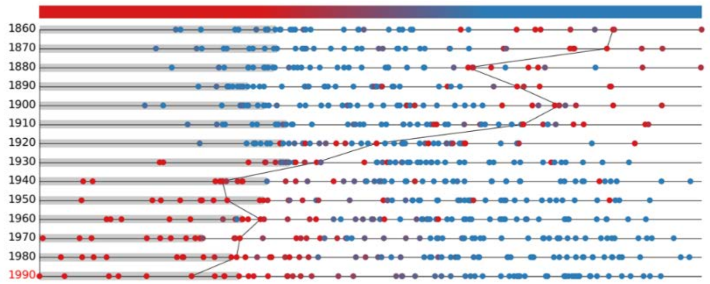
Co-occurance Matrices

3. Concept Axes



Concept Axis Plots

1. Similarities (local distances)



Buddy Plots

2. Co-Occurrances

Co-occurance Matrices

3. Concept Axes

Concept Axis Plots

Similarities: Understanding local distances

Distances are meaningful

even if absolute values are not

What is close to a word?

Are there groups of words that are similar?

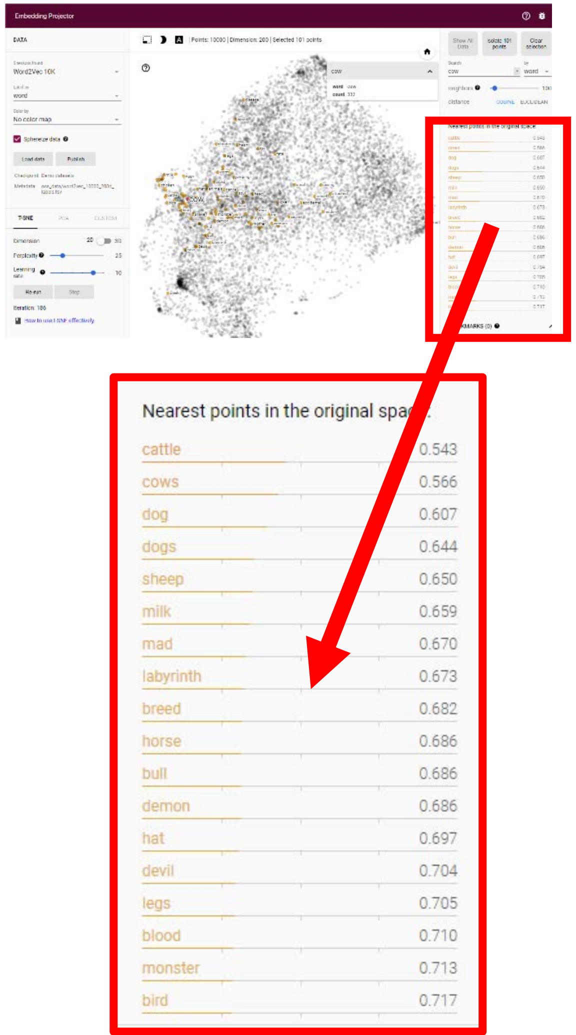
Ordered lists are useful

Density (how many can you show)

Sense of relative distances

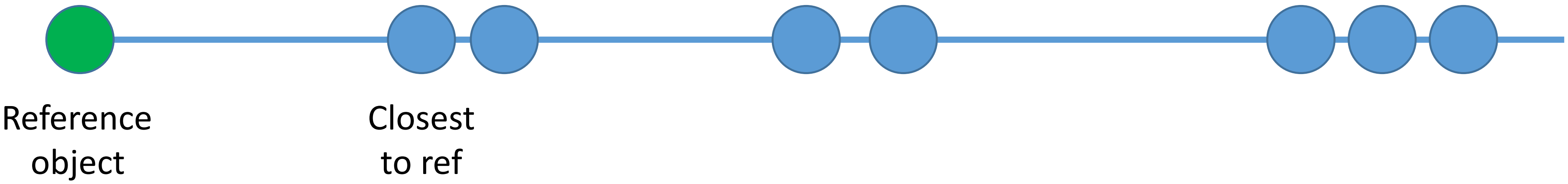
Comparison between words

Embedding Projector
Smilkov, et al. 2016



Buddy Plots (1D lists)

Map distance (to selected reference) to horizontal axis

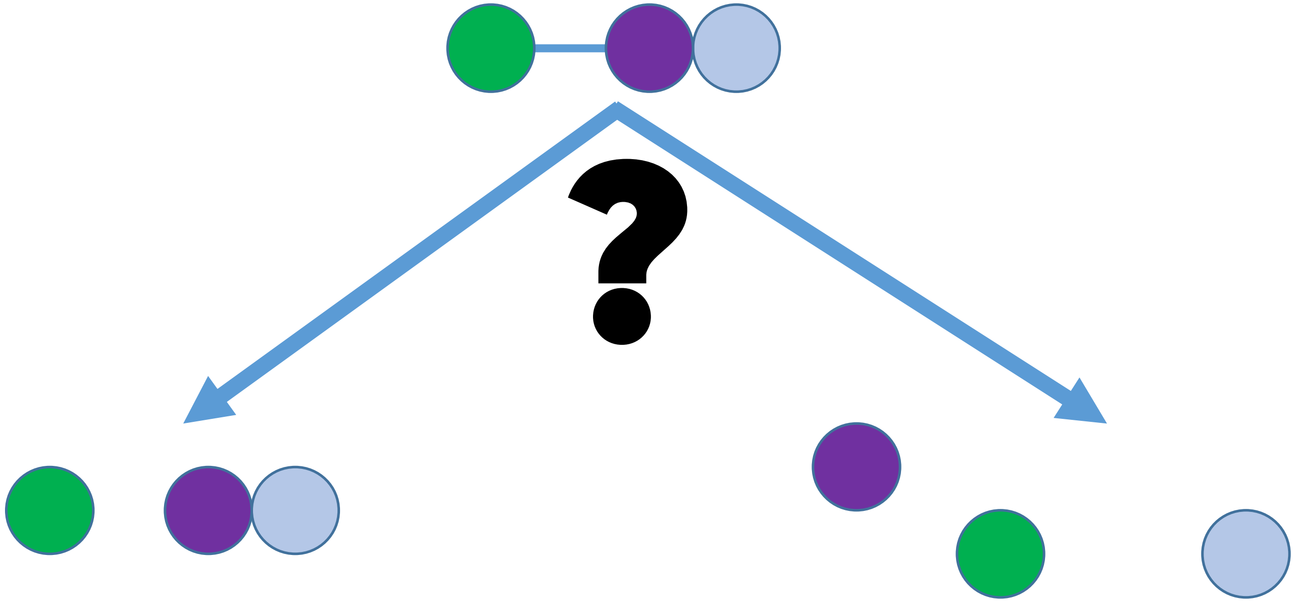


Alexander and Gleicher, 2016 – for Topic Models

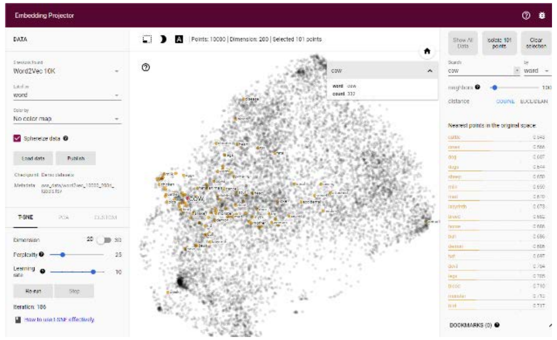
Not dimensionality reduction?

Map distance to word to horizontal axis

Focus on a single point – other relations not preserved

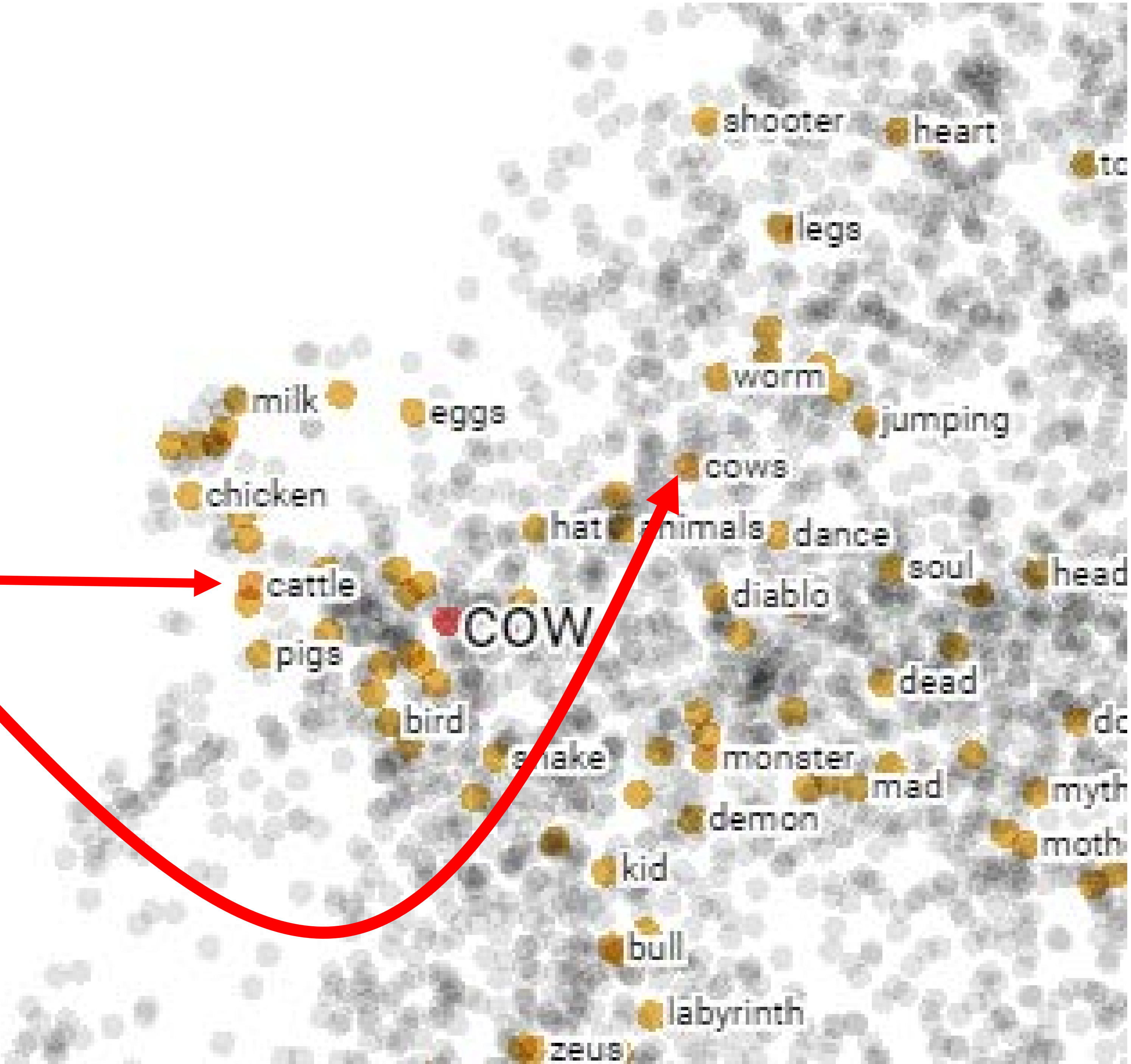


Embedding Projector
Smilkov, et al. 2016



Nearest points in the original space:

cattle	0.543
cows	0.566
dog	0.607
dogs	0.644
sheep	0.650
milk	0.659
mad	0.670
labyrinth	0.673
breed	0.682
horse	0.686
bull	0.686

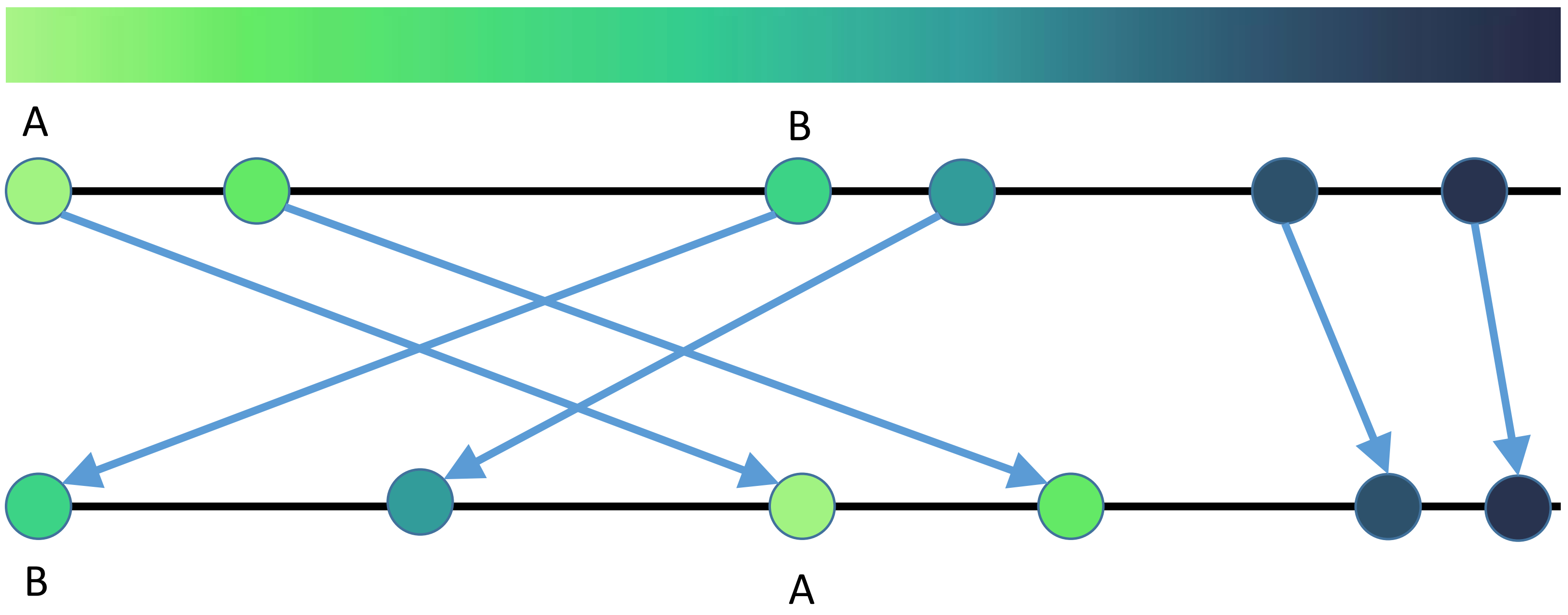


Stacked/Chained buddy plots

Use color to encode distance

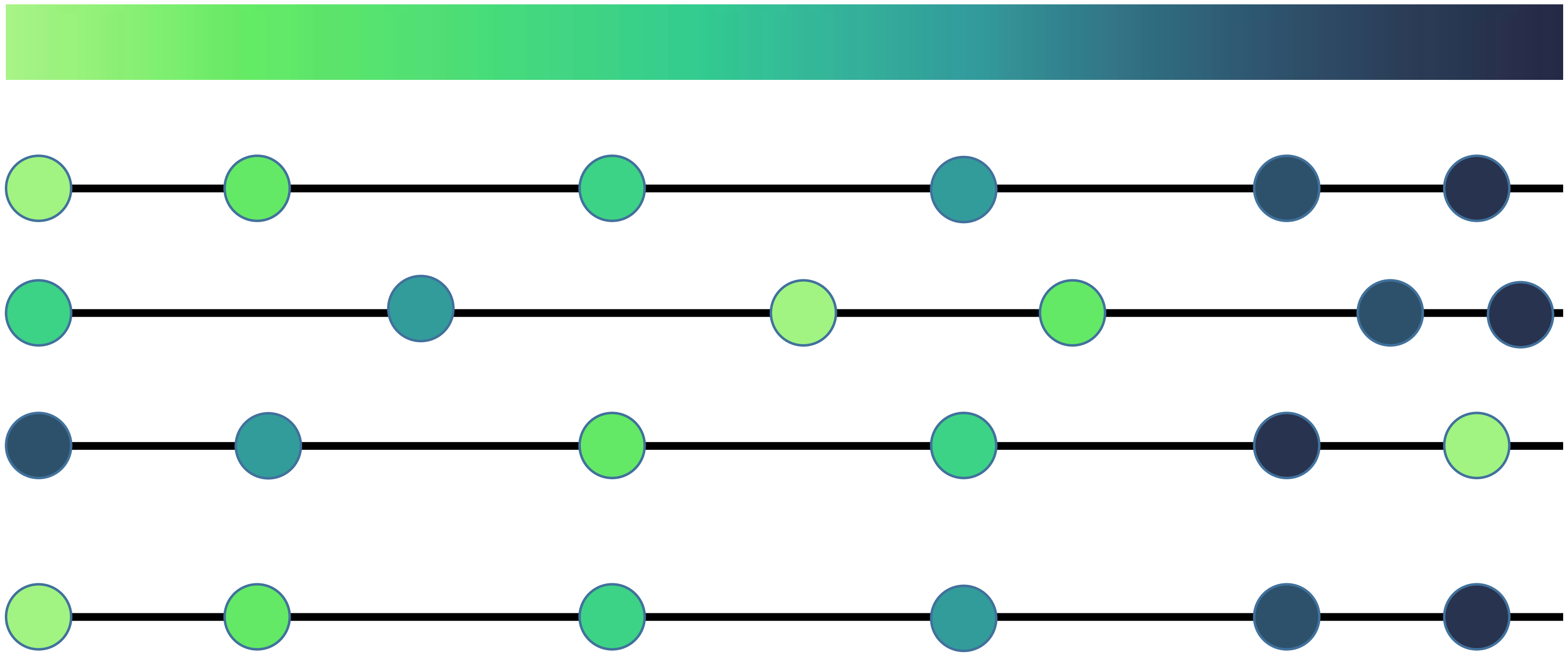
Stacked/Chained buddy plots

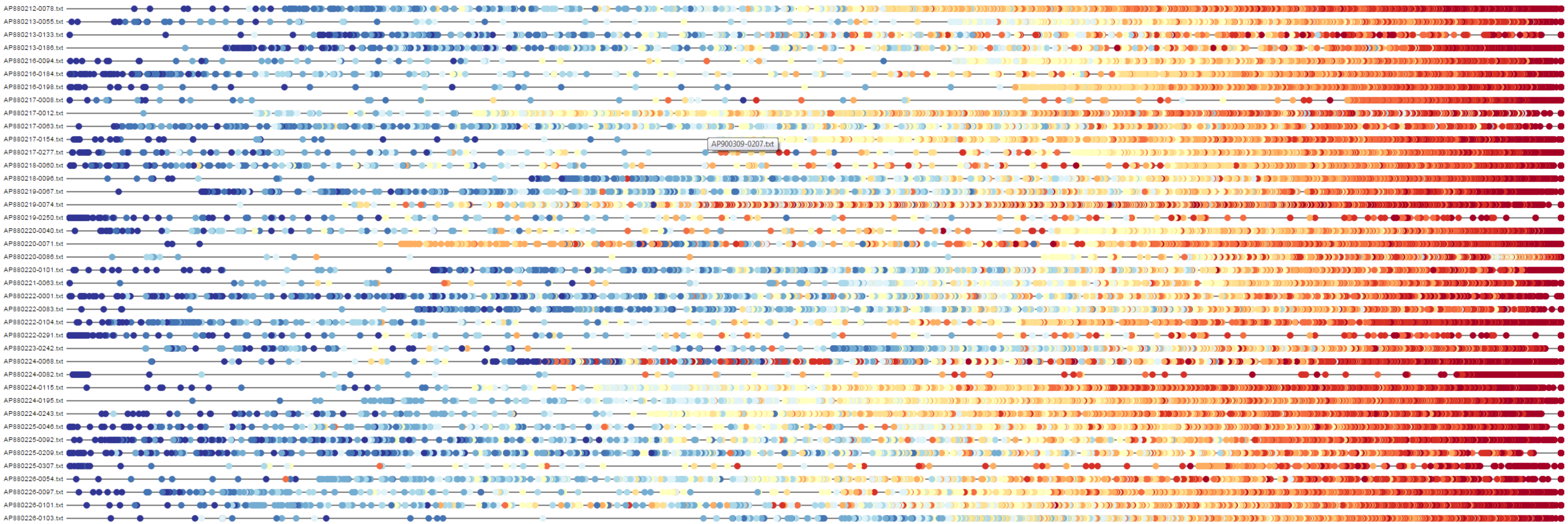
Use color to encode distance **in the reference row**



Stacked/Chained buddy plots

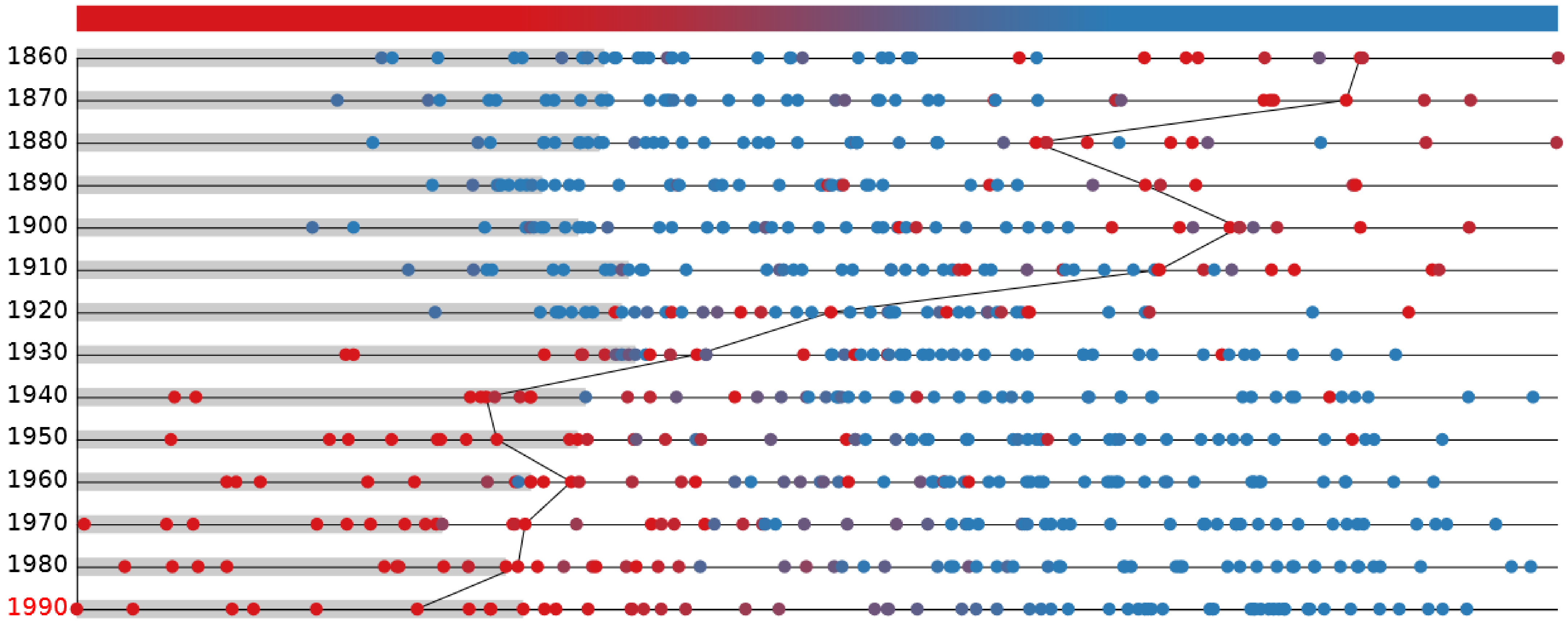
Use color to encode distance **in the reference**





AP880212-0078.txt
AP880213-0055.txt
AP880213-0133.txt
AP880213-0186.txt
AP880216-0094.txt
AP880216-0184.txt
AP880216-0198.txt
AP880217-0008.txt
AP880217-0012.txt
AP880217-0063.txt
AP880217-0154.txt
AP880217-0277.txt
AP880218-0060.txt
AP880218-0096.txt
AP880219-0067.txt
AP880219-0074.txt
AP880219-0250.txt
AP880220-0040.txt
AP880220-0071.txt
AP880220-0086.txt
AP880220-0101.txt
AP880221-0063.txt
AP880222-0001.txt
AP880222-0083.txt
AP880222-0104.txt
AP880222-0291.txt
AP880223-0242.txt
AP880224-0068.txt
AP880224-0082.txt
AP880224-0115.txt
AP880224-0195.txt
AP880224-0243.txt
AP880225-0046.txt
AP880225-0092.txt
AP880225-0209.txt
AP880225-0307.txt
AP880226-0054.txt
AP880226-0097.txt
AP880226-0101.txt
AP880226-0103.txt
AP900309-0207.txt

Same word... different embedding

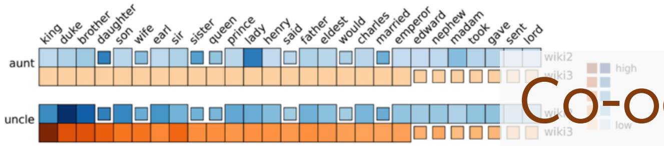


Broadcast

1. Similarities (local distances)

Buddy Plots

2. Co-Occurrances



Co-occurance Matrices

3. Concept Axes

Concept Axis Plots

Why are words similar?
Understandng word co-occurence

Similarity based on co-occurrence

  count how often one word occurs near another

Co-occurrence matrix

  main form of input data

  many models approximate the matrix (reconstruct)

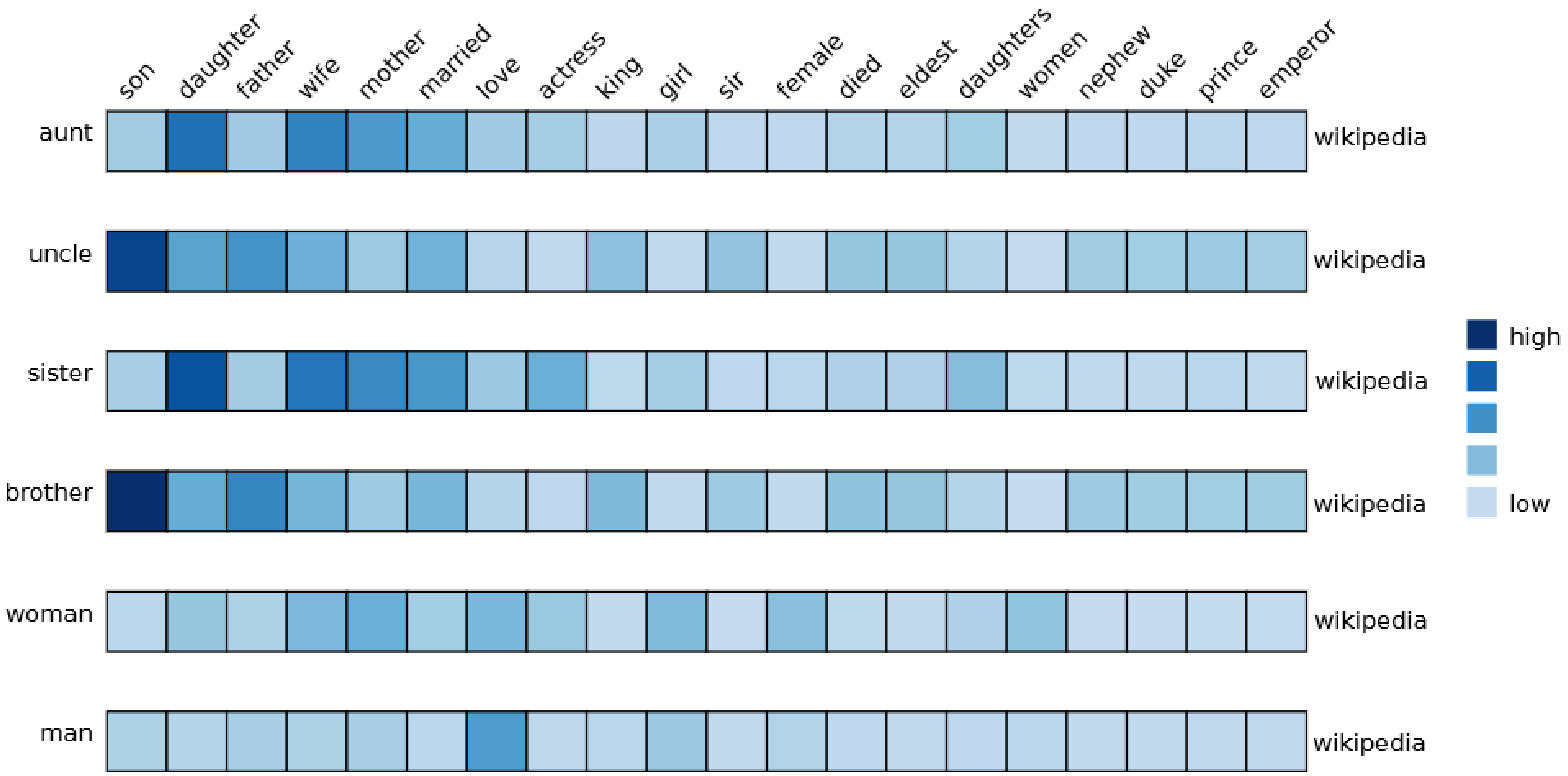
Useful for understanding and diagnosing models

Co-occurrence view

How do we view the massive matrix?

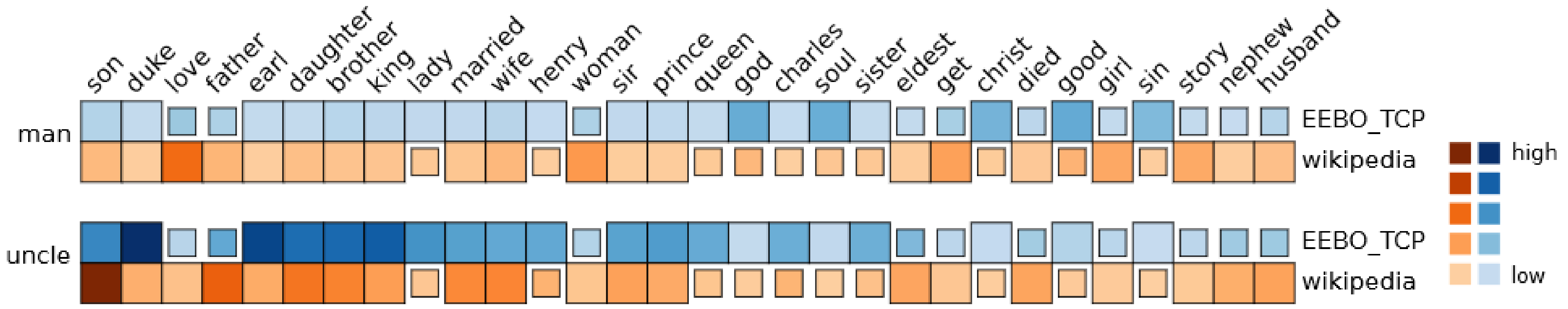
- color encoding (heat map) – density, relative values
- select rows (specify words of interest)
- select columns (metrics of interestingness – given rows)
    - highest values
    - highest variance

Co-occurrence matrix view



Matrix comparison view

High variance words in one embedding may be low in the other



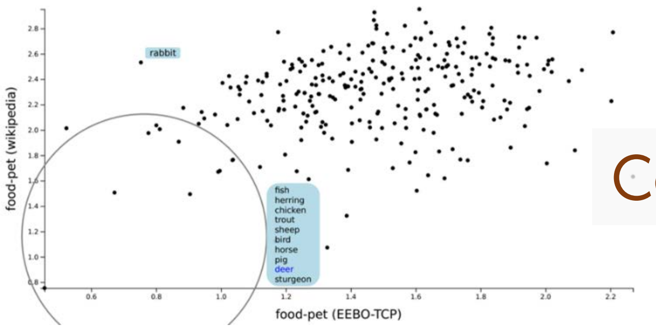
1. Similarities (local distances)

Buddy Plots

2. Co-Occurrances

Co-occurance Matrices

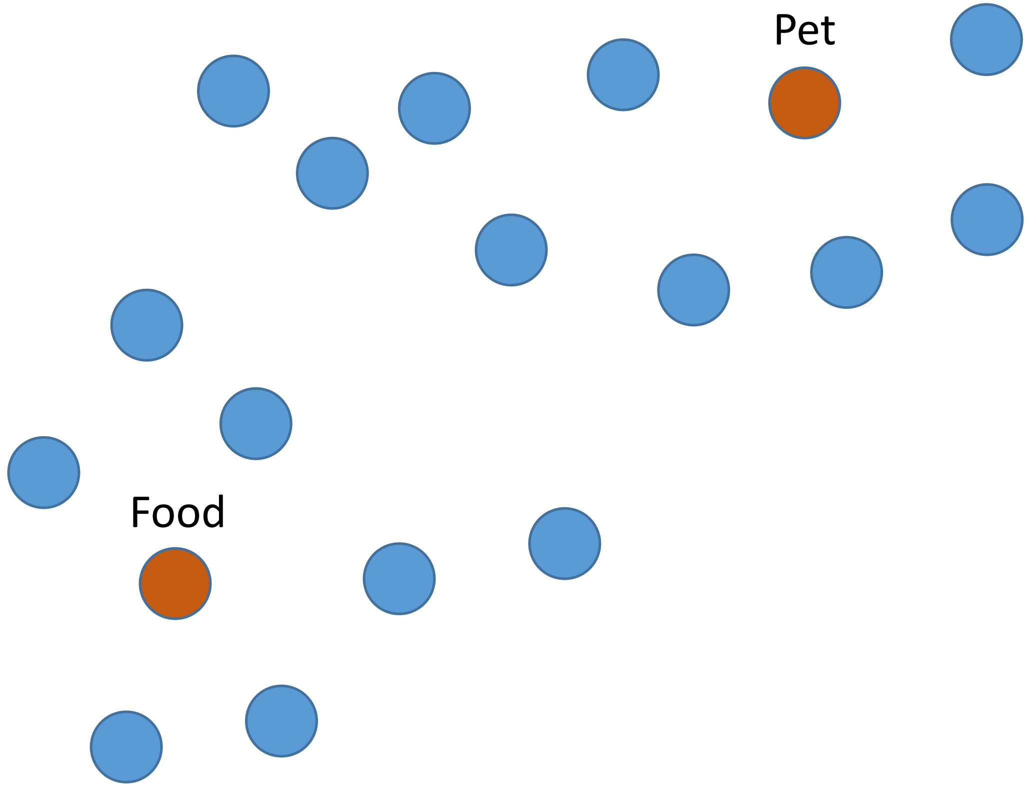
3. Concept Axes



Concept Axis Plots

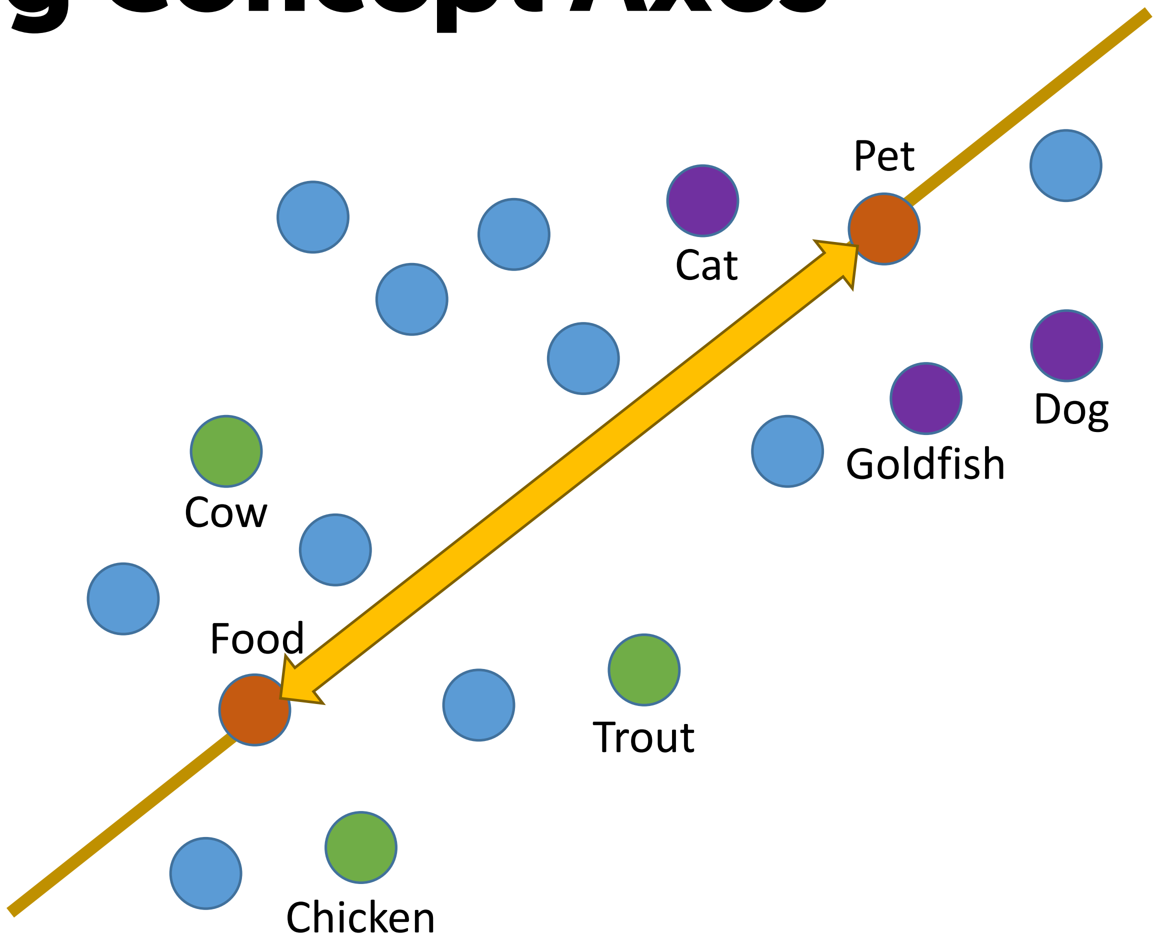
Concept Axes: Understanding Semantic Directions

Opposing concepts make an axis



Concept Axes: Understanding Concept Axes

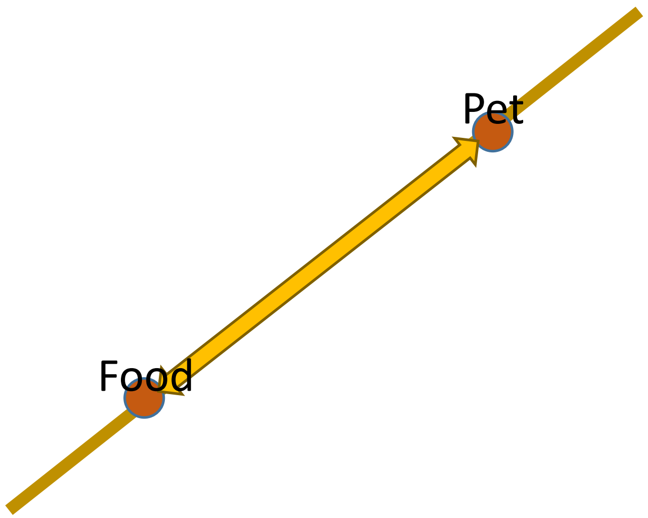
Define an axis from one concept to another



Ways to define axes

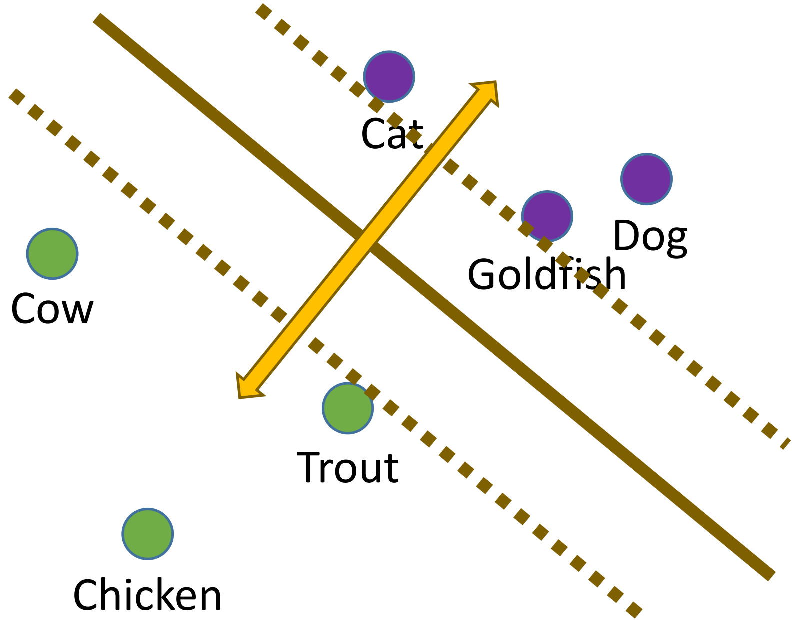
Vector between two concepts

Interaxis - Kim et al., 2015



Classifier between two groups

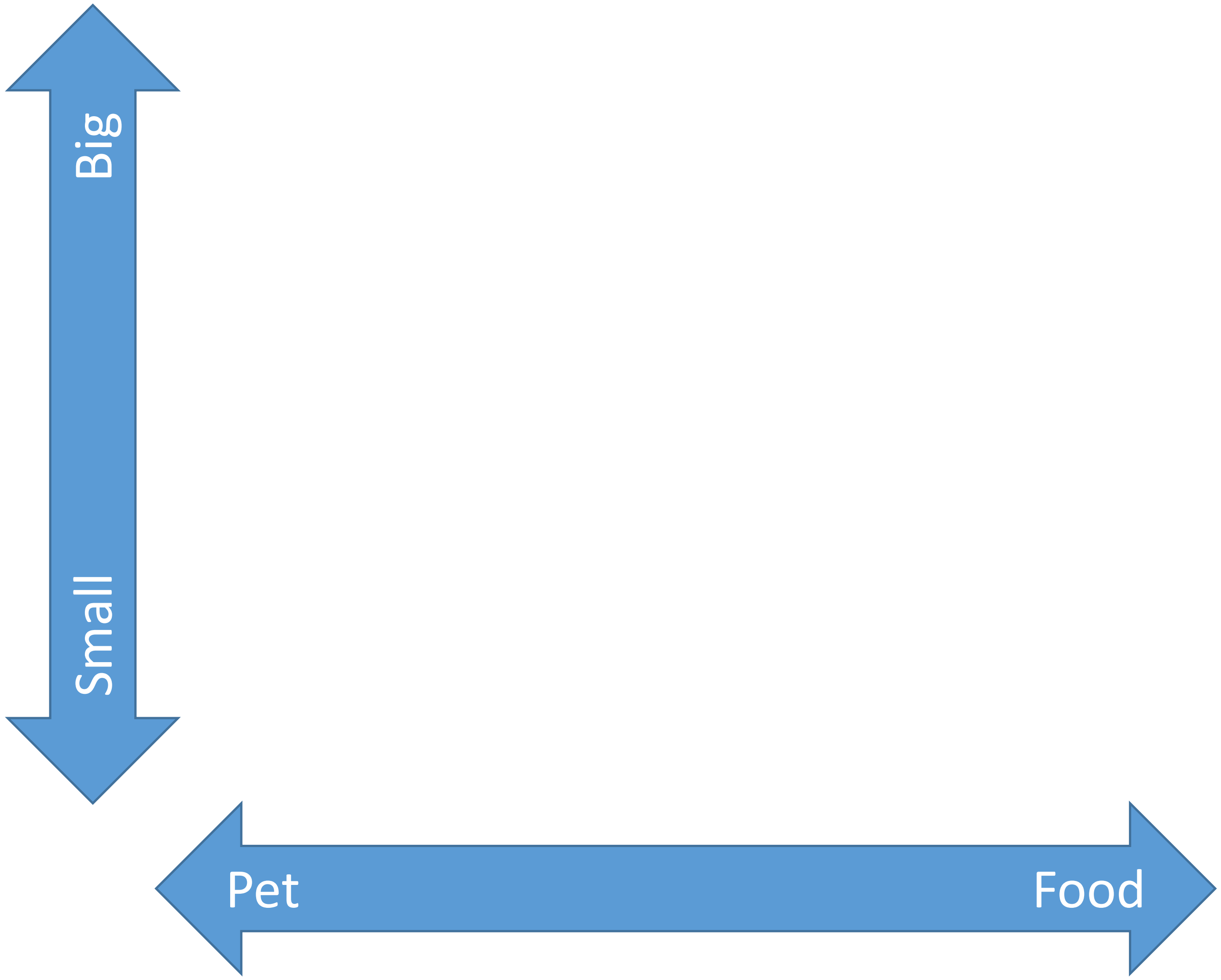
Explainers – Gleicher, 2013



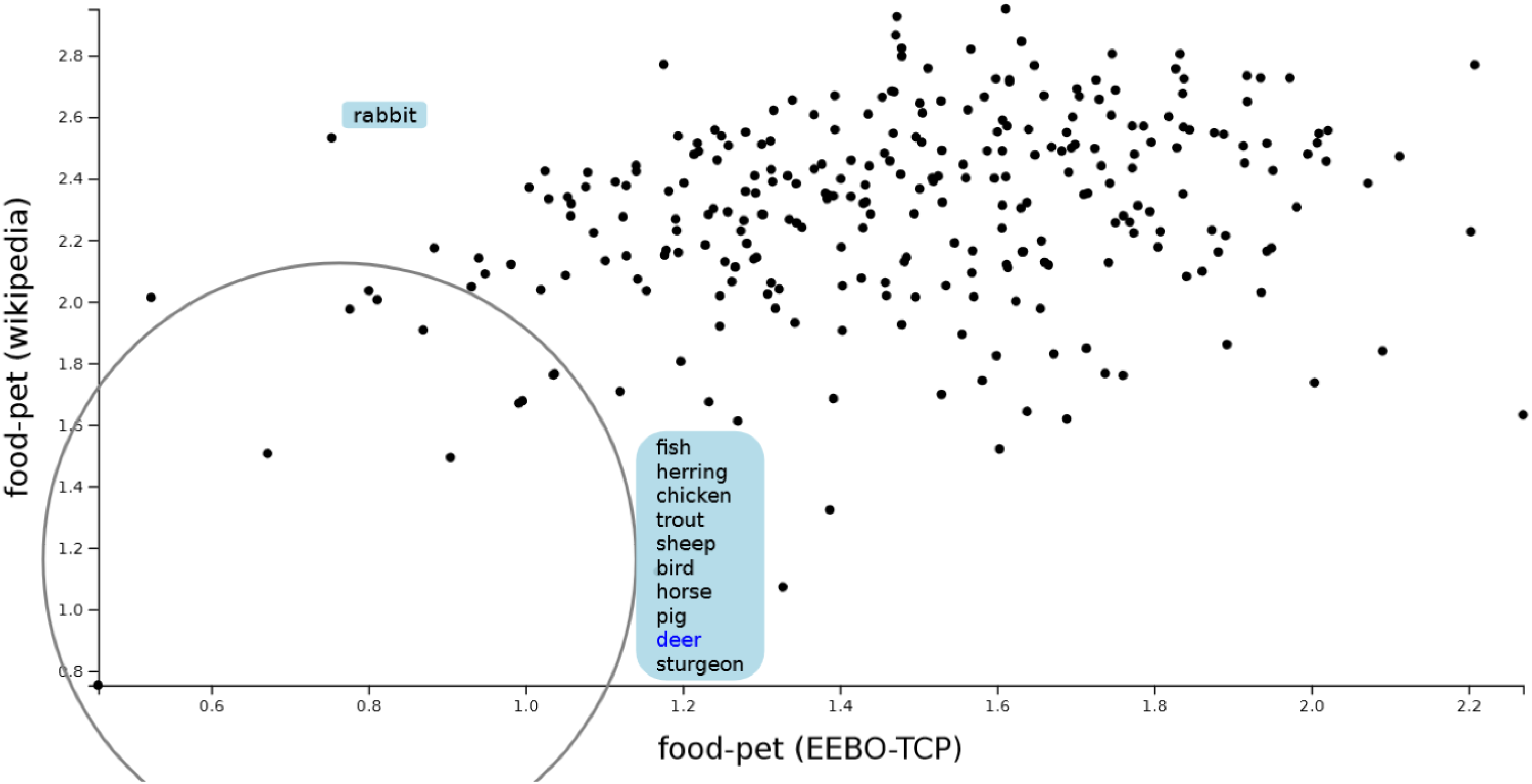
Multiple Concept Axes

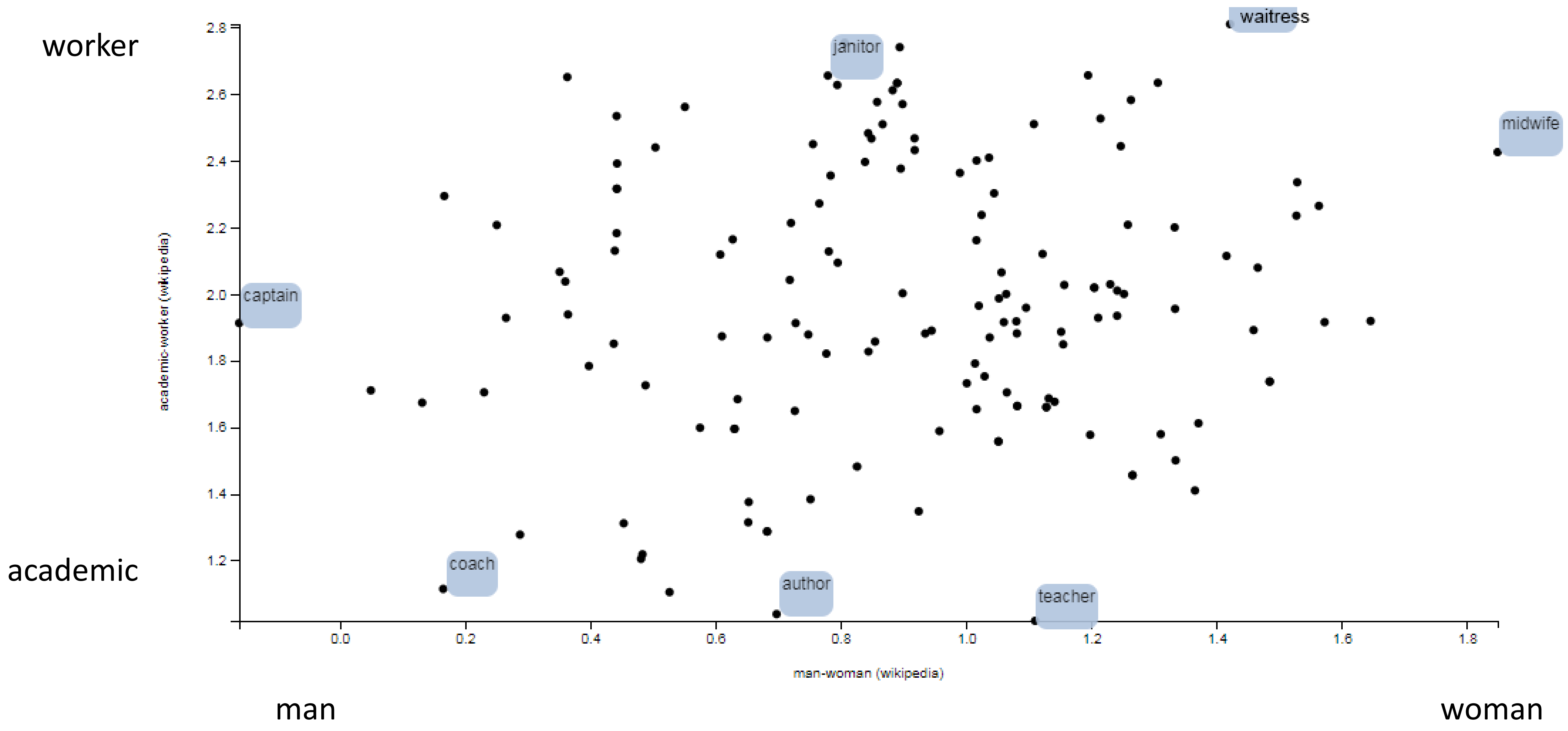
Use multi-variate plots

2D = Scatterplot

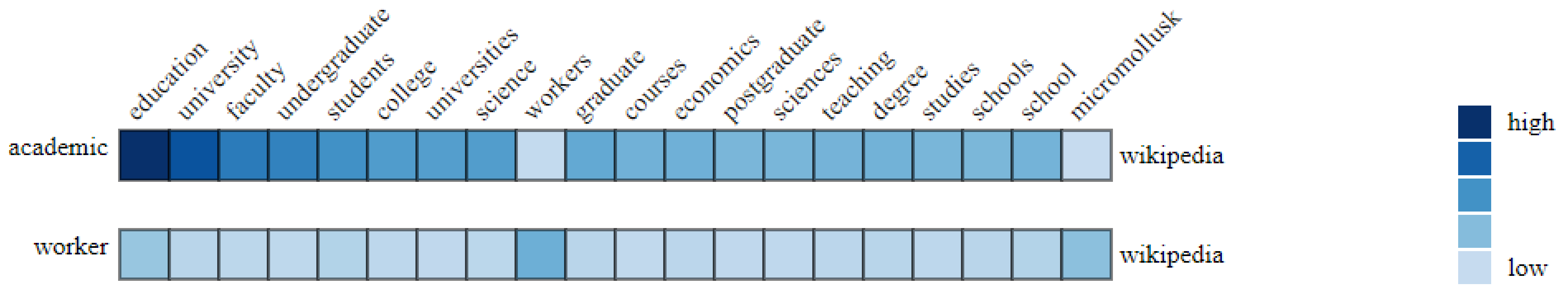


Scatterplot Interaction

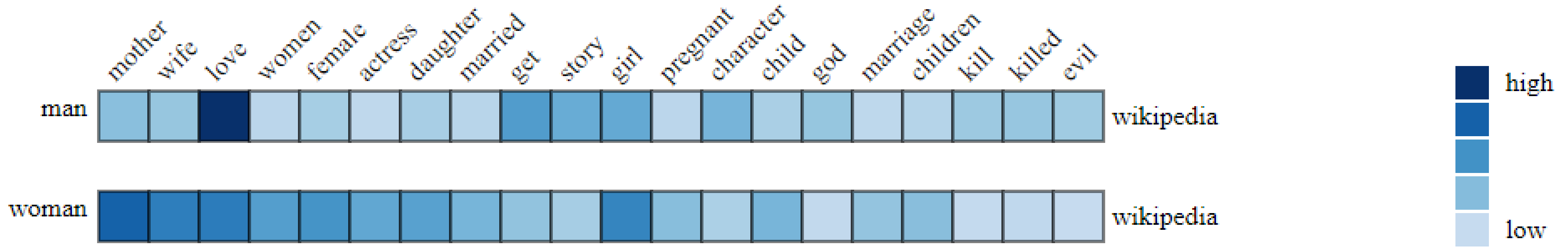




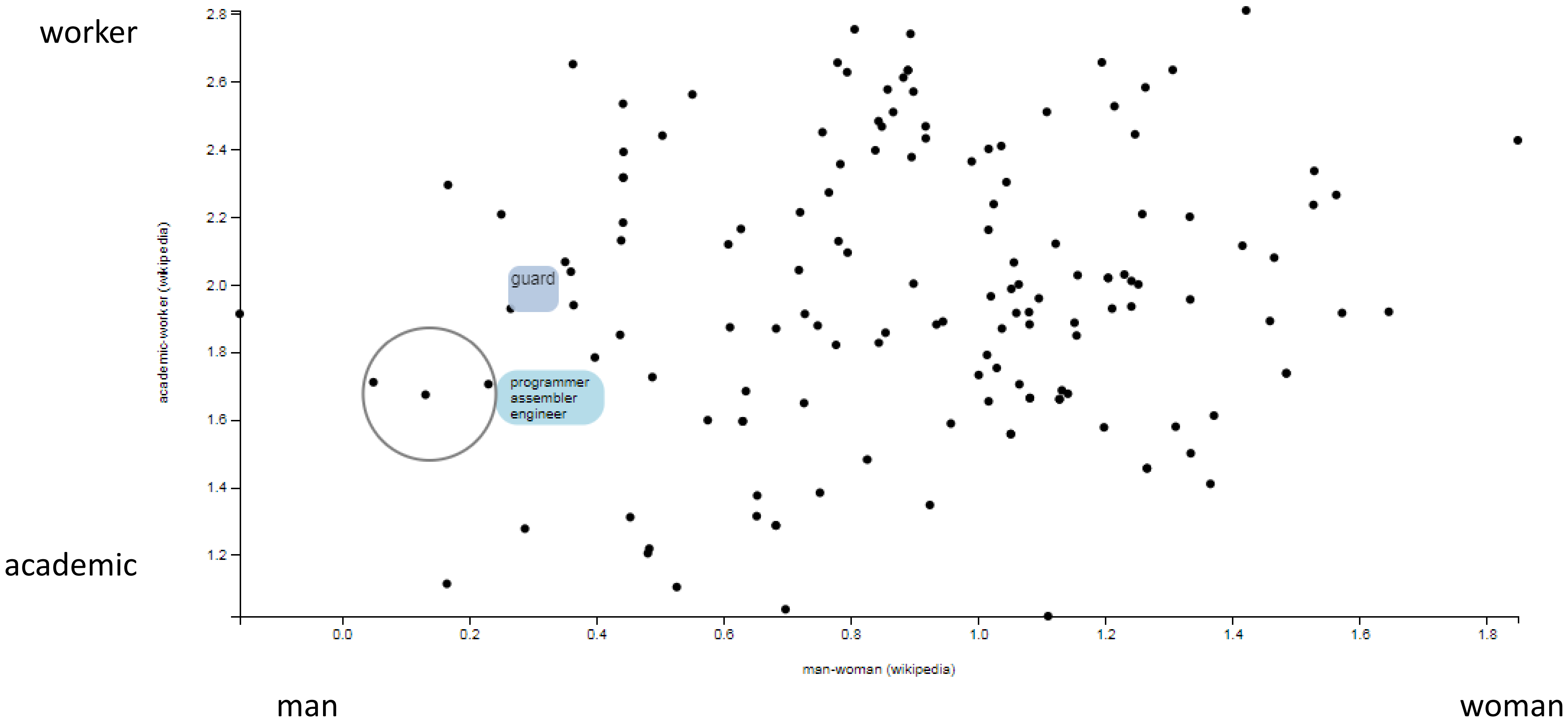
worker
academic
man
woman
academic-worker (wikipedia)
man-woman (wikipedia)
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
waitress
janitor
midwife
captain
coach
author
teacher



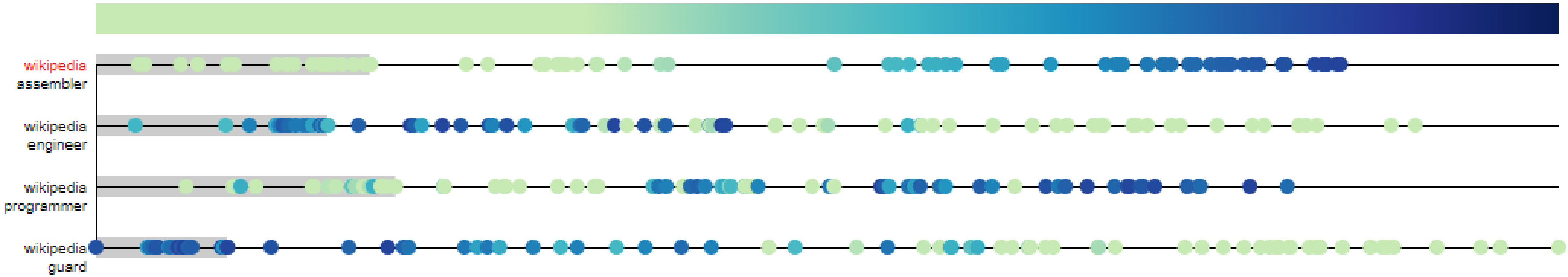
education
university
faculty
undergraduate
students
college
universities
science
workers
graduate
courses
economics
postgraduate
sciences
teaching
degree
studies
schools
school
micromollusk
academic
worker
wikipedia
wikipedia
high
low

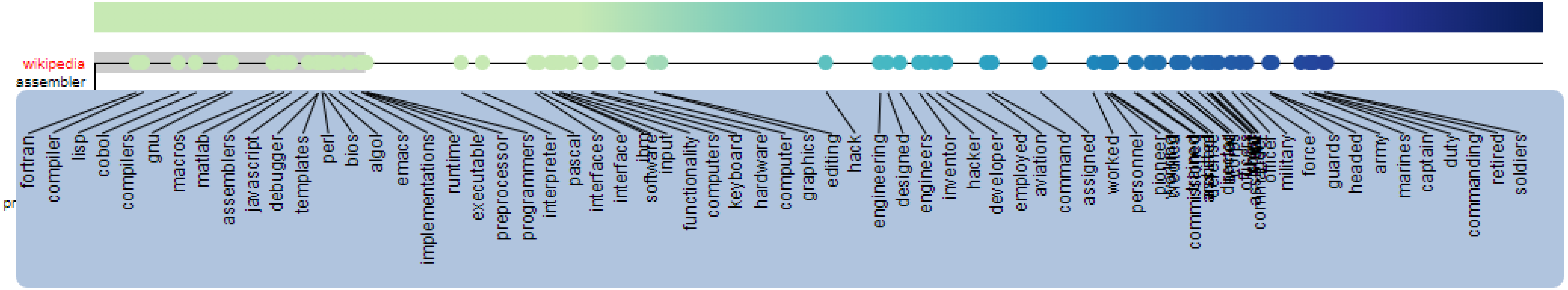


mother
wife
love
women
female
actress
daughter
married
get
story
girl
pregnant
character
child
god
marriage
children
kill
killed
evil
man
woman
wikipedia
wikipedia
high
low

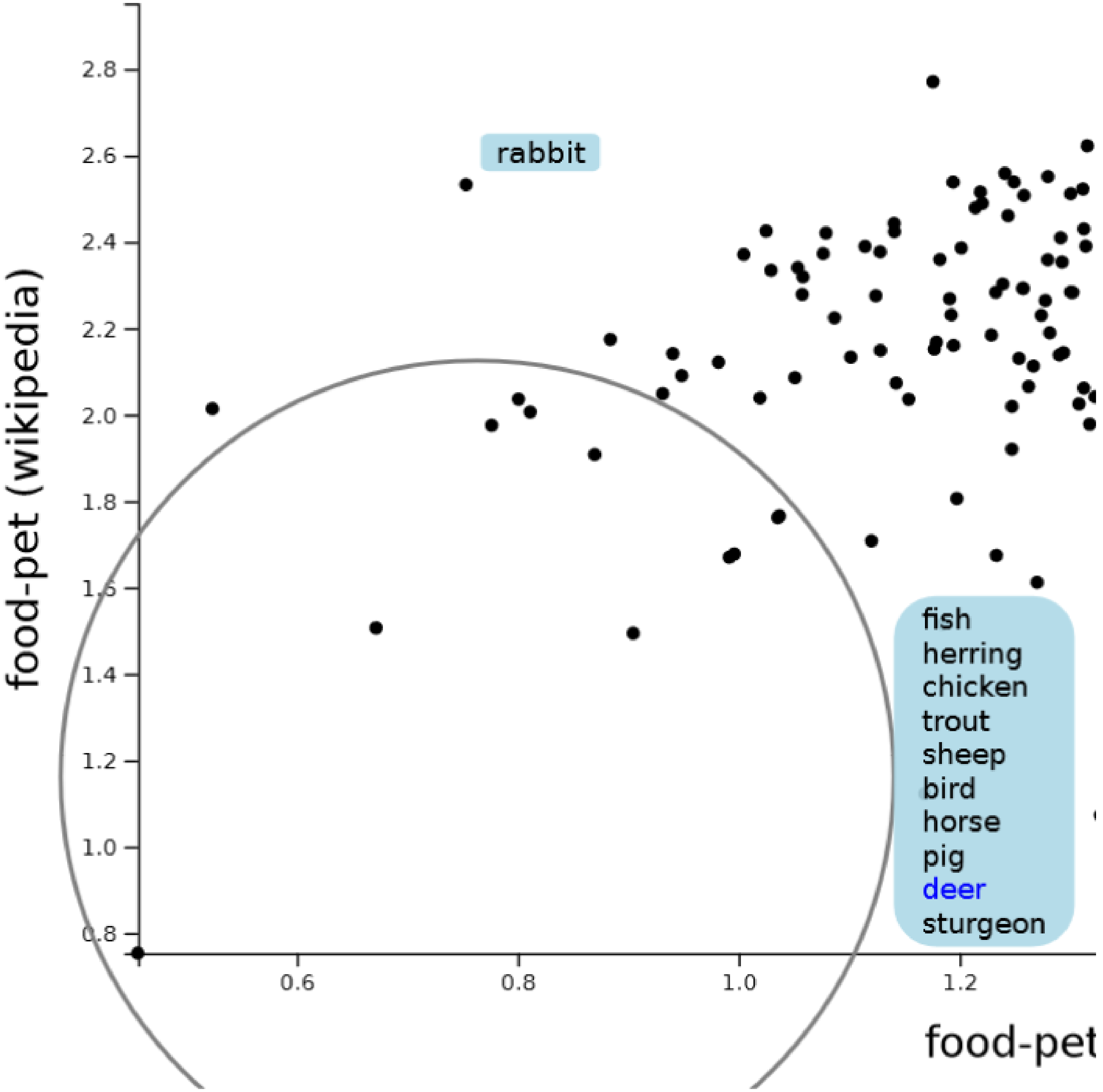


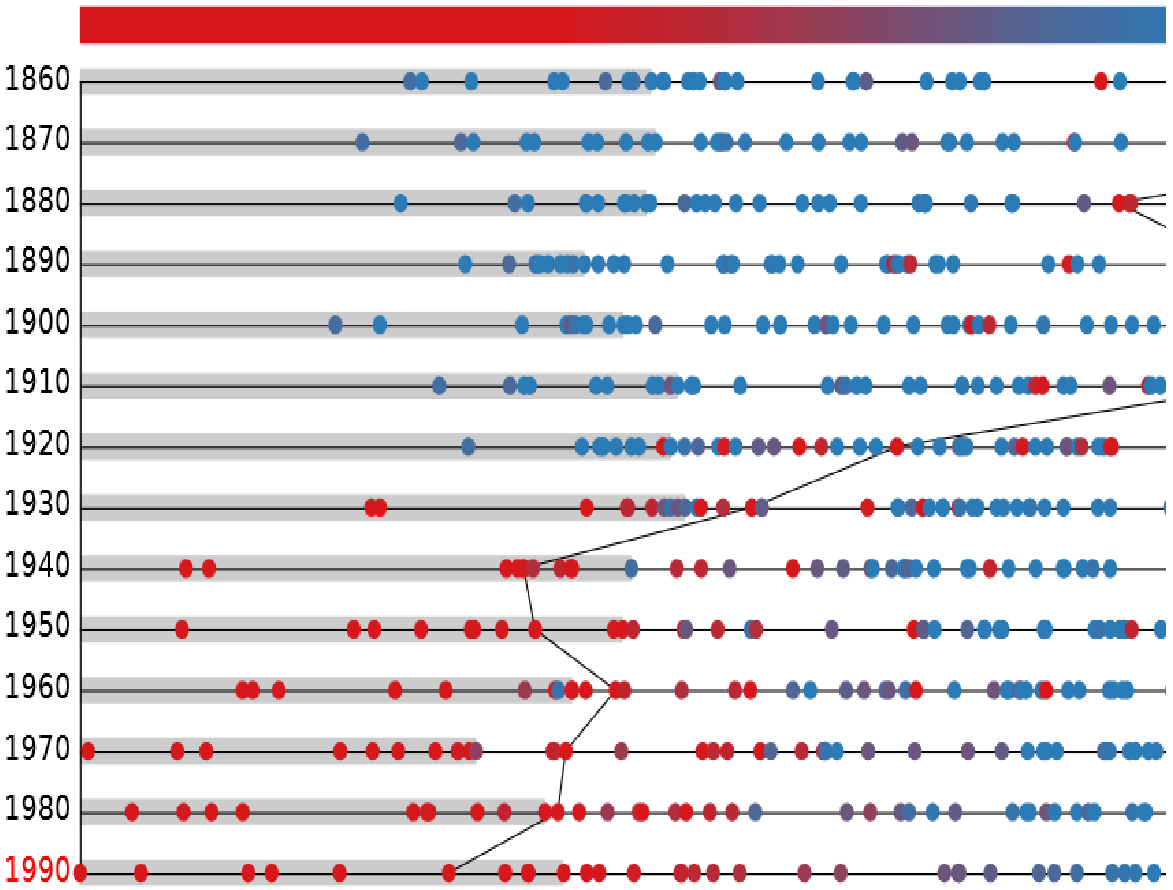
worker
academic
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
academic-worker (wikipedia)
guard
programmer
assembler
engineer
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
man-woman (wikipedia)
man
woman


wikipedia
assembler
wikipedia
engineer
wikipedia
programmer
wikipedia
guard

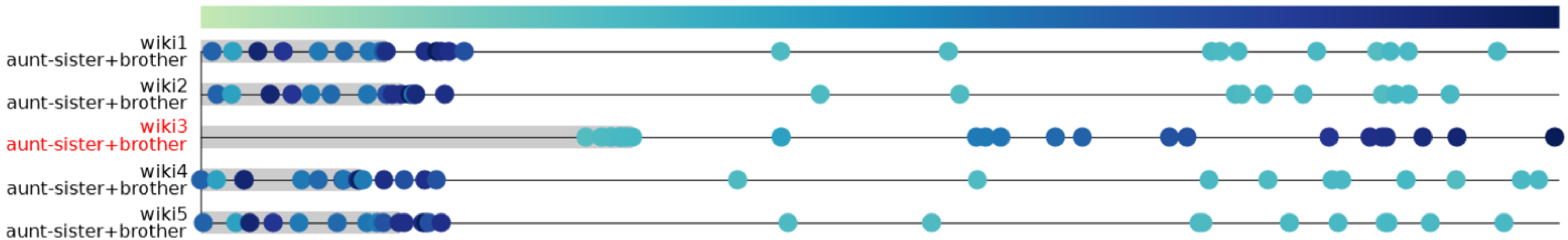

wikipedia
assembler
fortran
compiler
lisp
cobol
compilers
gnu
macros
matlab
assemblers
javascript
debugger
templates
perl
bios
algol
emacs
implementations
runtime
executable
preprocessor
programmers
interpreter
pascal
interfaces
interface
software
input
functionality
computers
keyboard
hardware
computer
graphics
editing
hack
engineering
designed
engineers
inventor
hacker
developer
employed
aviation
command
assigned
worked
personnel
pioneer
commissioned
officer
military
force
guards
headed
army
marines
captain
duty
commanding
retired
soldiers

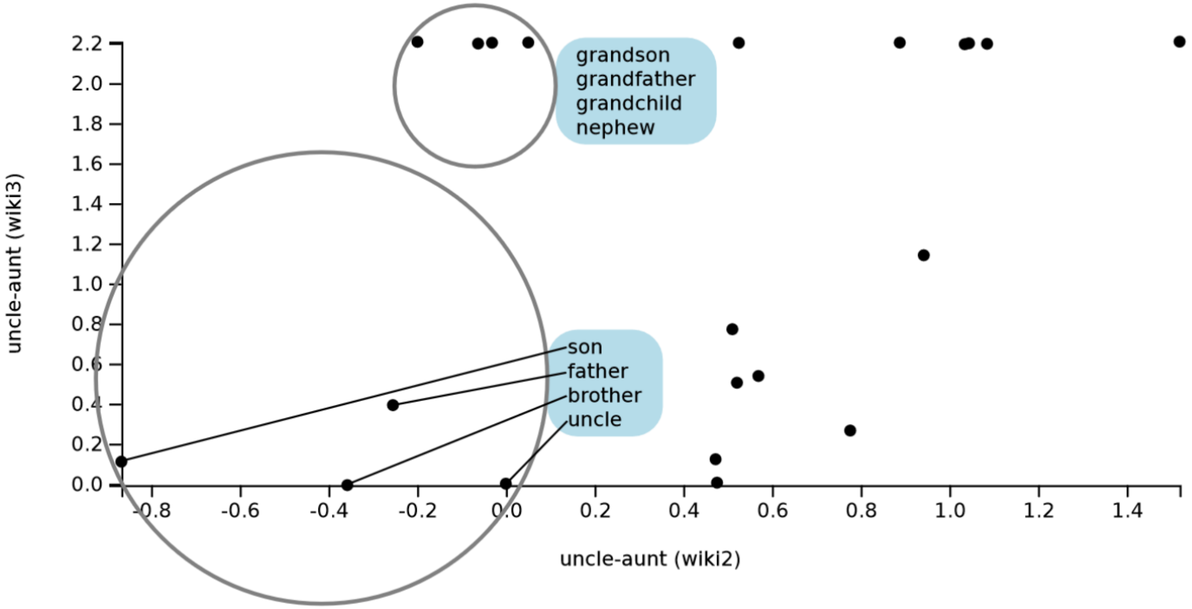
Application: Word meaning change





Application: Stability Assessment





Implementation

http://graphics.cs.wisc.edu/Vis/EmbVis/

Everything runs on line (simplified interfaces)

cloud version uses small models

Python backend / D3 frontend

Limitations

Implementation	Usability and Scalability
Effectiveness	Evaluation (of designs)
Completeness	More Tasks Identifying Probes Explicit Comparison Connection to (model) evaluation Feedback to model building

Summary:

Word vector embeddings offer unique challenges

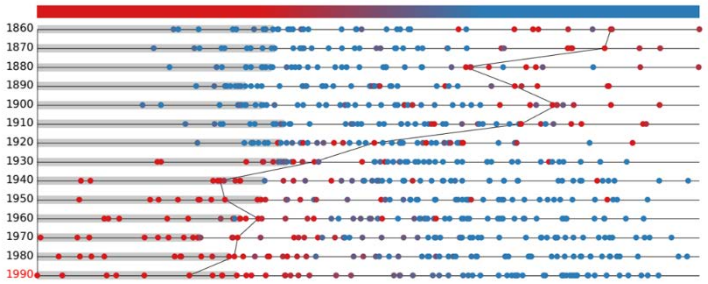
Task analysis of needs

characteristic	single target	multiple targets
similarity	(1) inspect local neighborhood	(2) compare local neighborhoods
average, offset	(3) inspect arithmetic structure	(4) compare arithmetic results
co-oc. probability	(5) analyze encoded probabilities	(6) compare probabilities
concept axis	(7) analyze vector relations	(8) compare vector relations
multiple	(9) discover interesting aspects	(10) compare interesting aspects

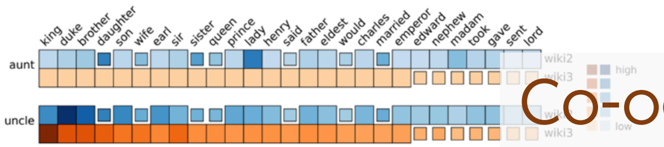
http://graphics.cs.wisc.edu/Vis/EmbVis/

Acknowledgments
This work is funded in part by NSF 1162037 and DARPA FA8750-17-2-0107

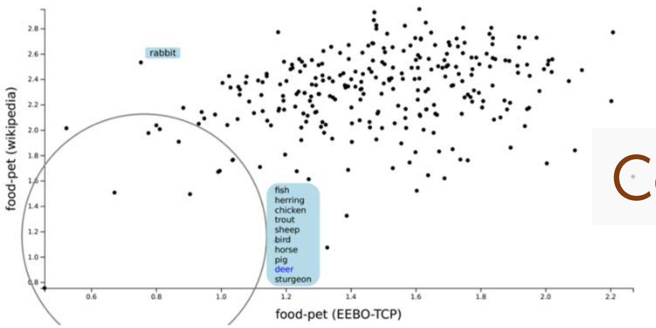
Designs for 3 unmet needs



Buddy Plots



Co-occurance Matrices



Concept Axis Plots