What can word vectors tell us about society?

Sandeep Soni

09/10/2024

Story so far

- Tokenization: raw text to tokens
- Distributional hypothesis: Words used in similar contexts have similar meanings
- Many different ways to transform words into vectors to encode word semantics

From last class

	d_1	d_2	d_3	…	d_m
w_1	2				1
w_2		3			
w_3	1				
…					
w_n			5		

Term-Document Matrix

	w_1	w_2	w_3	…	w_m
w_1	2				1
w_2		3			
w_3	1				
…					
w_n			5		

Cooccurrence Matrix

w_1	0.1				1.06
w_2		0.14			
w_3	-0.2				
…					
w_n			0.75		

Word Embedding Matrix

Words as distribution of counts over documents

Size = $|V| \times |D|$

Sparse

Words as distribution of counts over cooccurrences

Size = $|V| \times |V|$

Sparse

Words as learned vectors from cooccurrence data

Size = $|V| \times k$

Dense

From last class

- Instead of counting, word2vec or GloVe are methods to learn vectors by predicting which words will co-occur together

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

Vocab

service
wine
cold
great

Vocab

wine

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

Vocab
service
wine
cold
great

wine

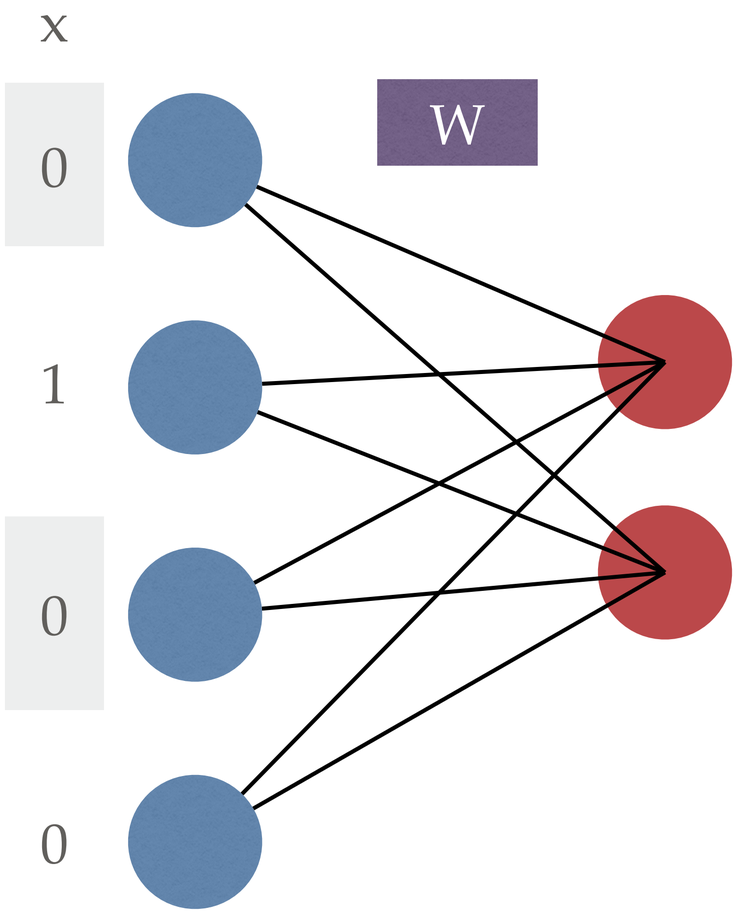
x
0
1
0
0

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

service
wine
cold
great

Vocab

wine

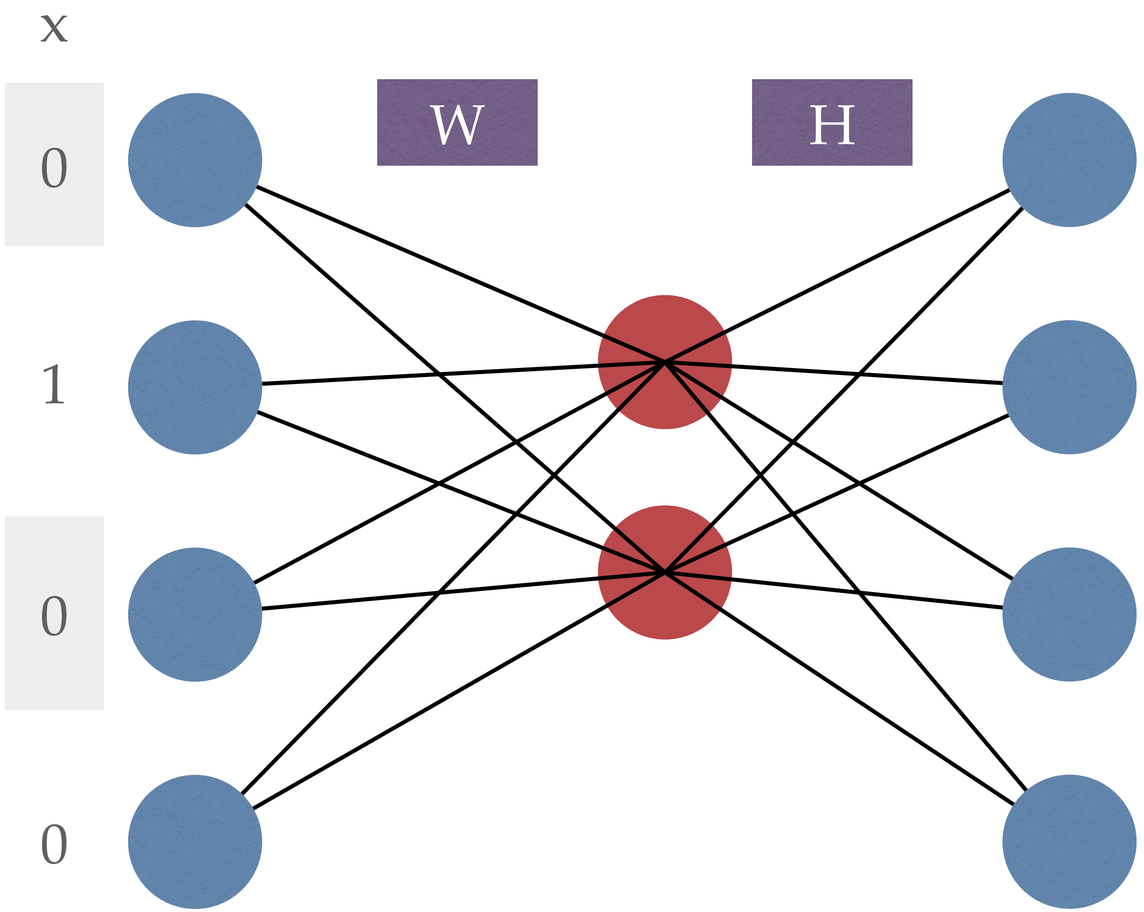


x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

service
wine
cold
great

Vocab

wine

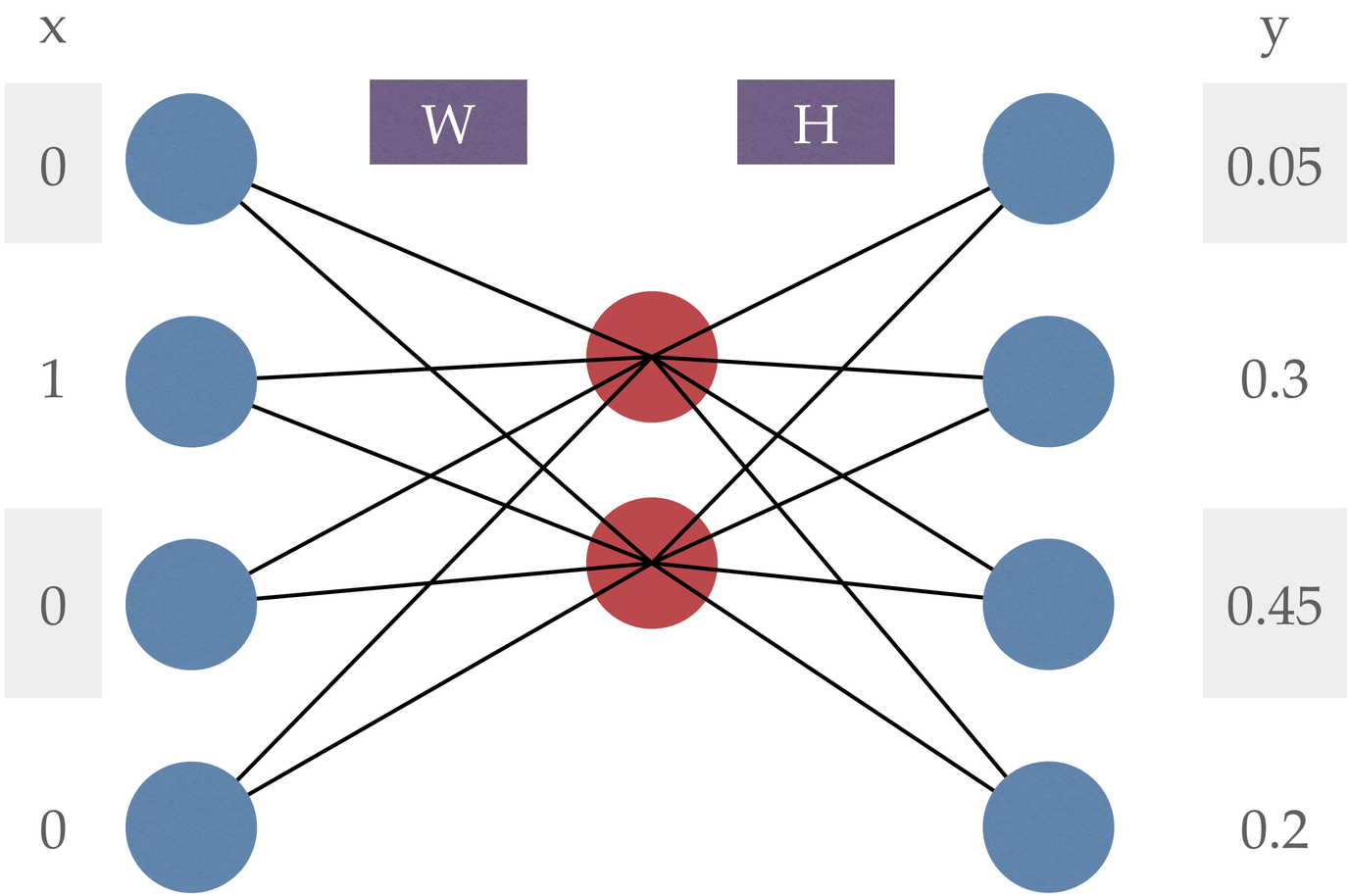


x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

service
wine
cold
great

Vocab

wine

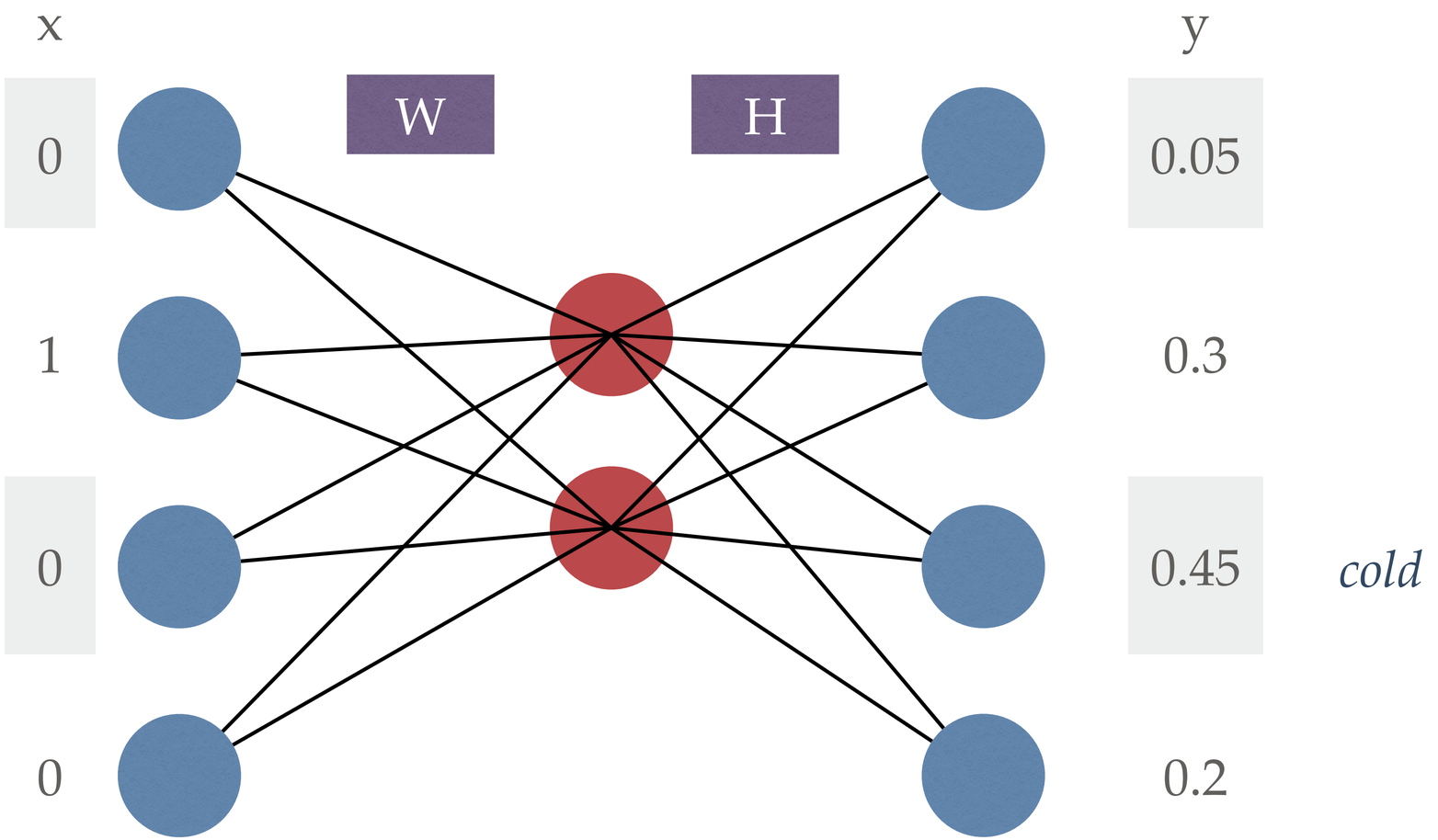


x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

service
wine
cold
great

Vocab

wine



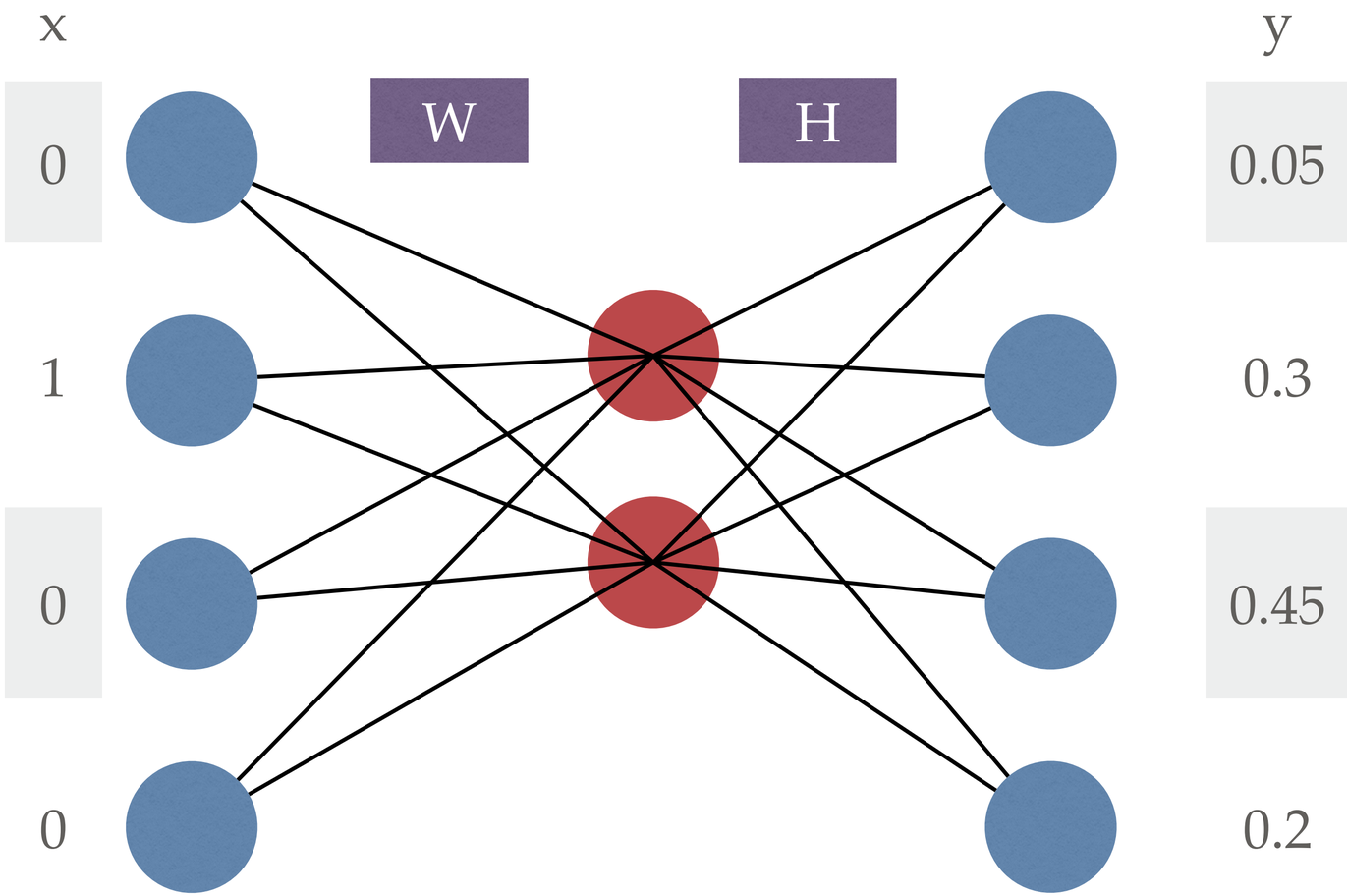
cold

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

service
wine
cold
great

Vocab

wine



cold

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

Word embeddings as columns

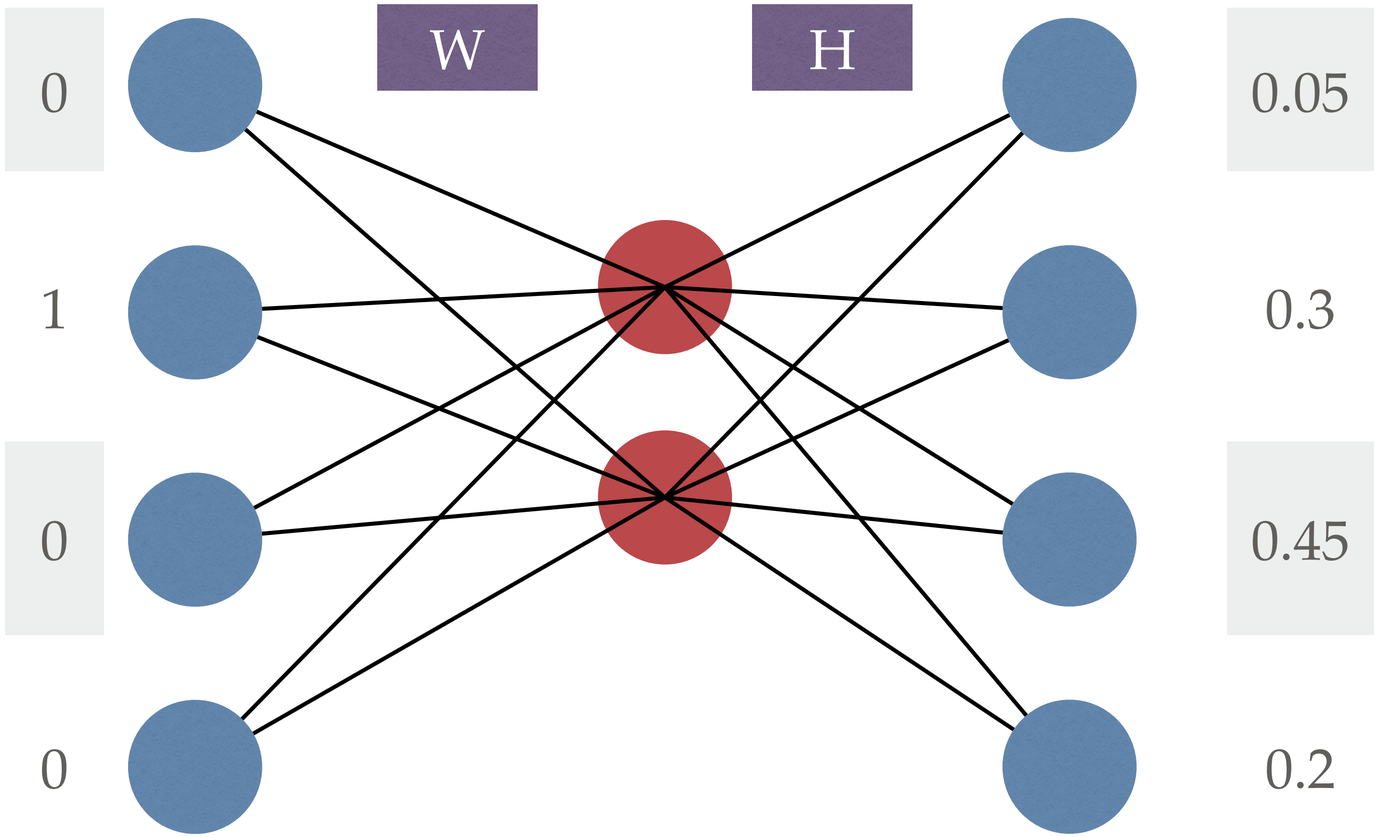
W

-0.3	1.2	0.5	-0.6
0.2	0.9	0.1	-0.4

service
wine
cold
great

Vocab

wine



cold

x	y
…	…
wine	*cold*
wine	*sweet*
wine	*spirit*
wine	*drink*
…	…

W

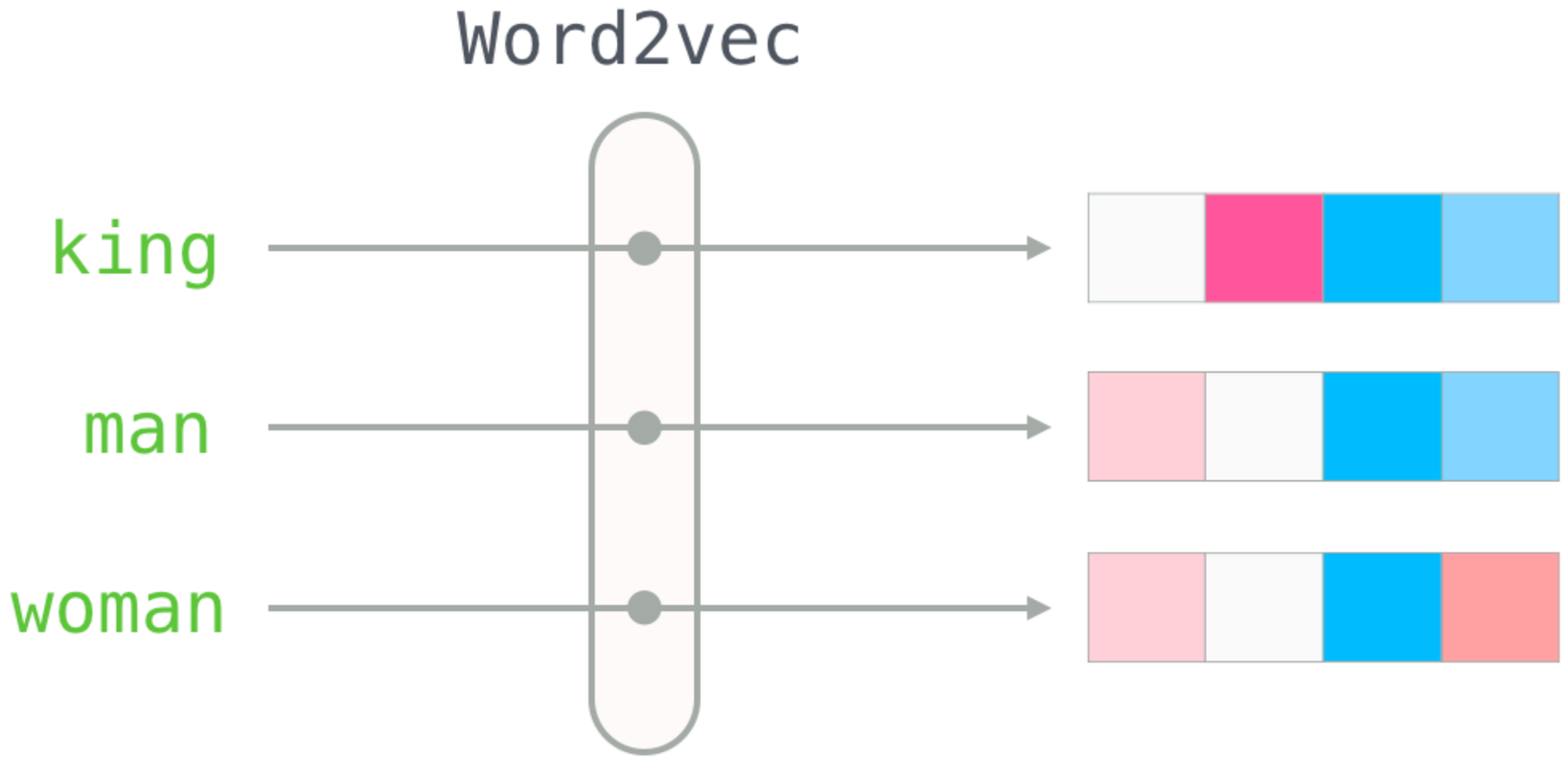
Word embeddings as columns

-0.3	1.2	0.5	-0.6
0.2	0.9	0.1	-0.4

H

Context embeddings as rows

0.1	-0.4
0.4	-0.5
0.3	-0.1
0.2	0.1



http://jalammar.github.io/illustrated-word2vec/

Things to remember

- Embedding size
- Window size
- Corpus size
- Initialization

Word2vec: few more considerations

- Pretrain then finetune
- Word2Vec variations
  - CBoW -> P(*focus_word* | *context_word*)
  - Skipgram -> P(*context_word* | *focus_word*)
- Hierarchical softmax or negative sampling to calculate probabilities

In class

- Word2Vec demo

Question for the day

"What do word embeddings encode and what can we do with them?"

Word Embeddings

- What is the geometry of word embeddings?
- What is their use as predictors?
- Can they be used to explain something about the world?

Geometry

Analogical reasoning

man:woman::king:?

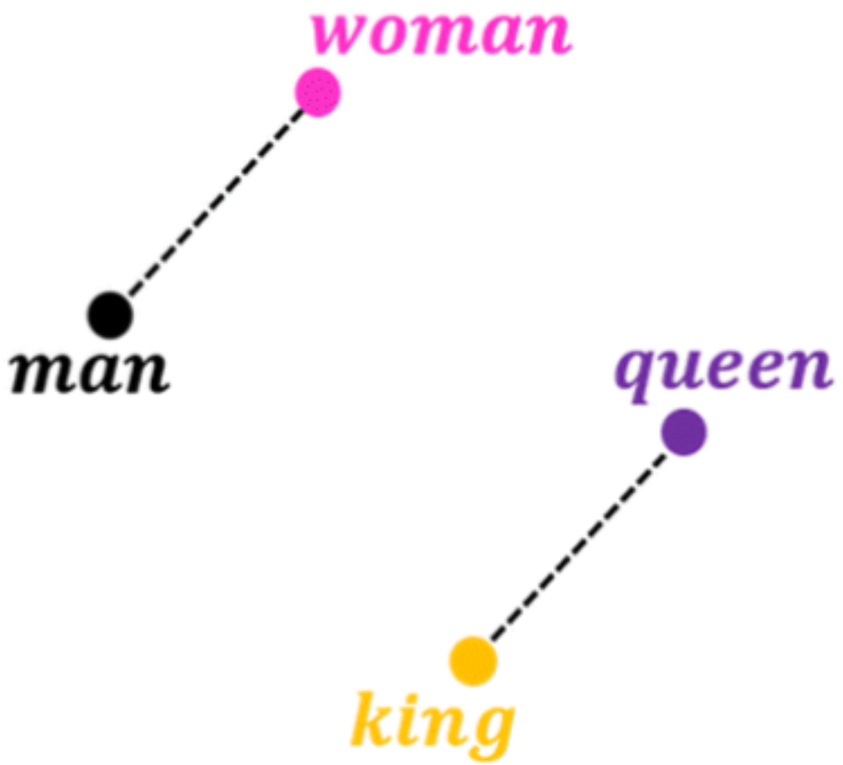
Geometry

man:woman::king:queen

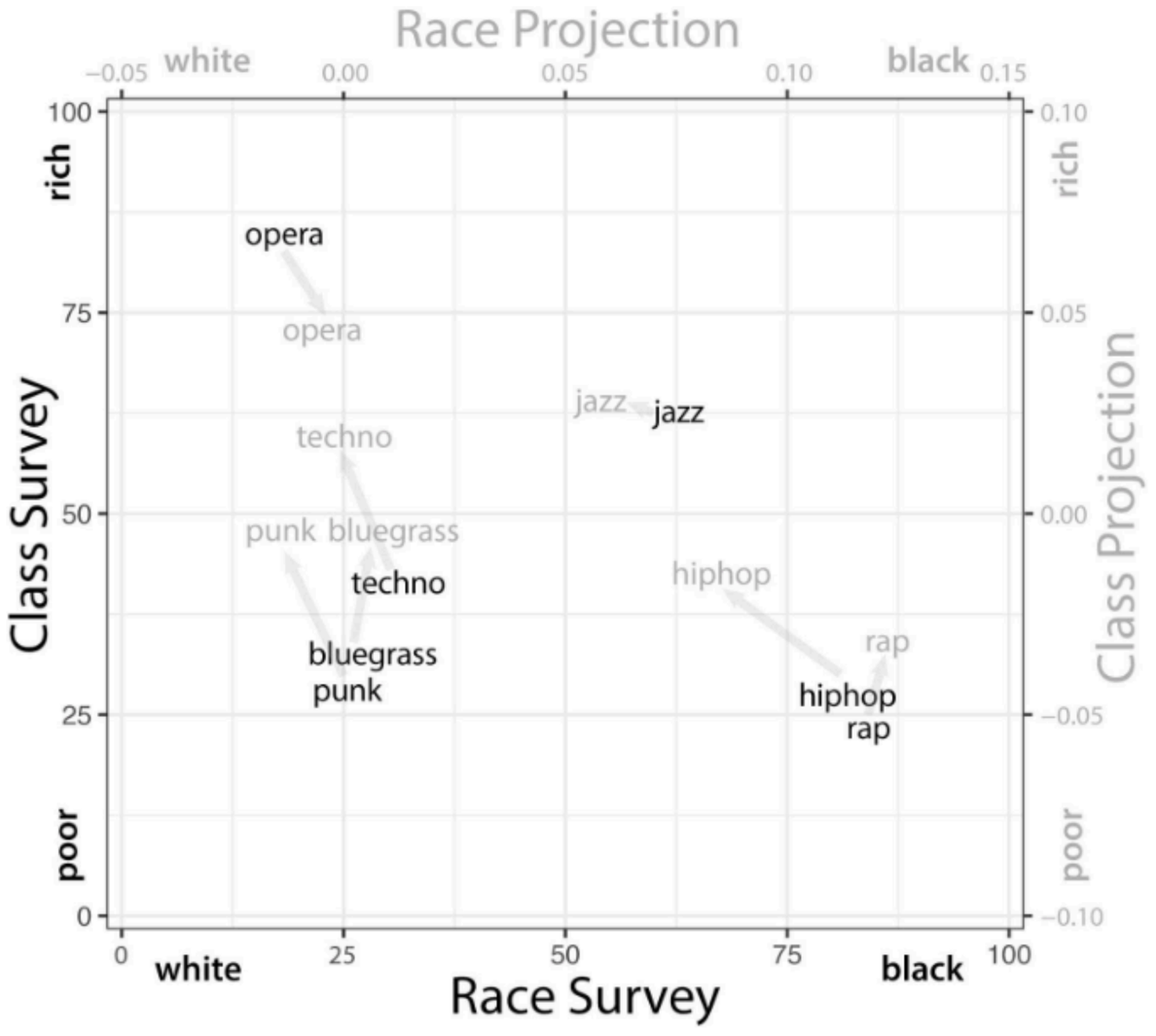
$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{king}} - \overrightarrow{\text{queen}}$$

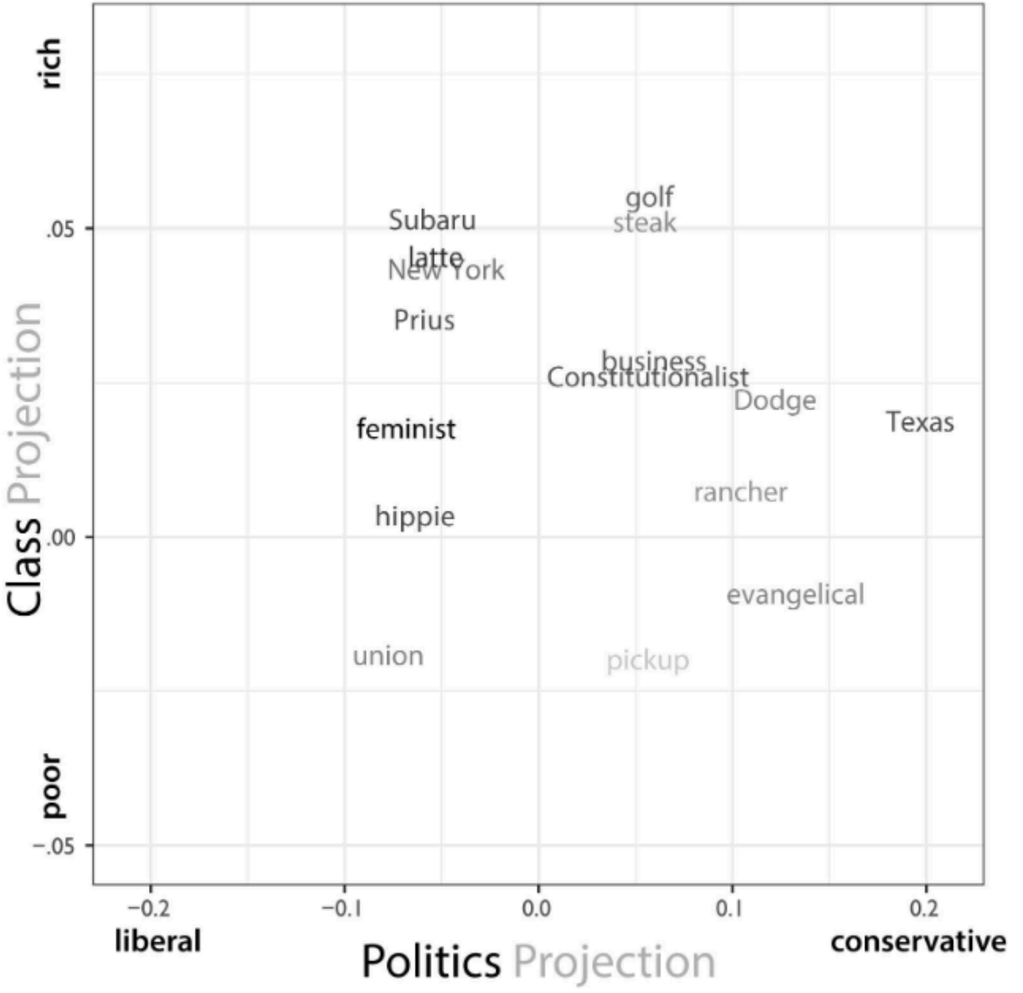
$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} \approx \overrightarrow{\text{queen}}$$

$$\overrightarrow{\text{king}} - \overrightarrow{\text{man}} + \overrightarrow{\text{woman}} = ?$$



Mikolov et. al. 2013 Linguistic Regularities in Continuous Space Word Representations


Race Projection
white
black
−0.05
0.00
0.05
0.10
0.15
rich
poor
Class Survey
Race Survey
Class Projection
opera
jazz
techno
punk
bluegrass
hiphop
rap
0
25
50
75
100
−0.10
−0.05
0.00
0.05
0.10


rich
poor
Class Projection
Politics Projection
liberal
conservative
−0.2
−0.1
0.0
0.1
0.2
.05
.00
−.05
golf
steak
Subaru
latte
New York
Prius
business
Constitutionalist
Dodge
Texas
feminist
rancher
hippie
evangelical
union
pickup

Prediction

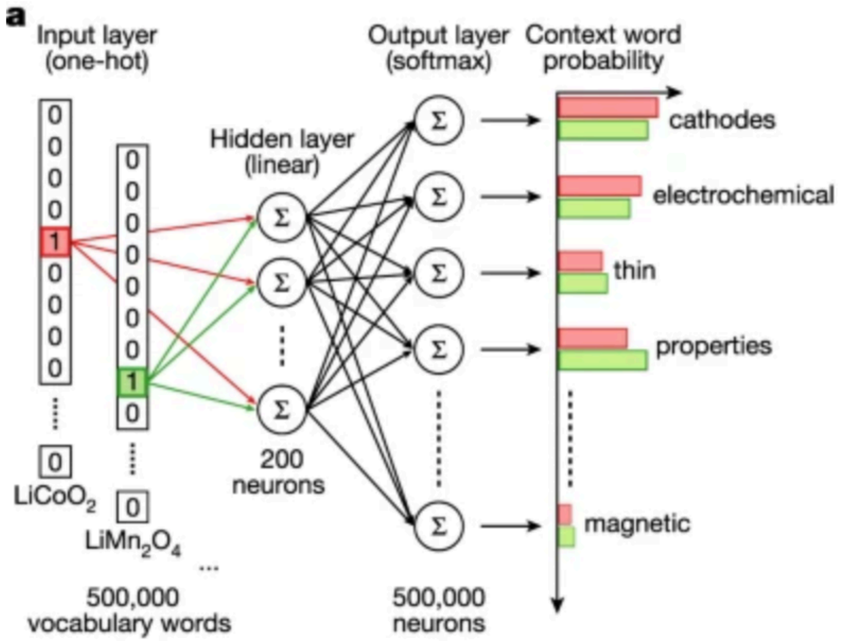
Prediction

$$\text{ferromagnetic} - \text{NiFe} + \text{IrMn} \approx \text{antiferromagnetic}$$

Oxides $\quad Zr - ZrO_2 \approx Cr - Cr_2O_3 \approx Ni - NiO$

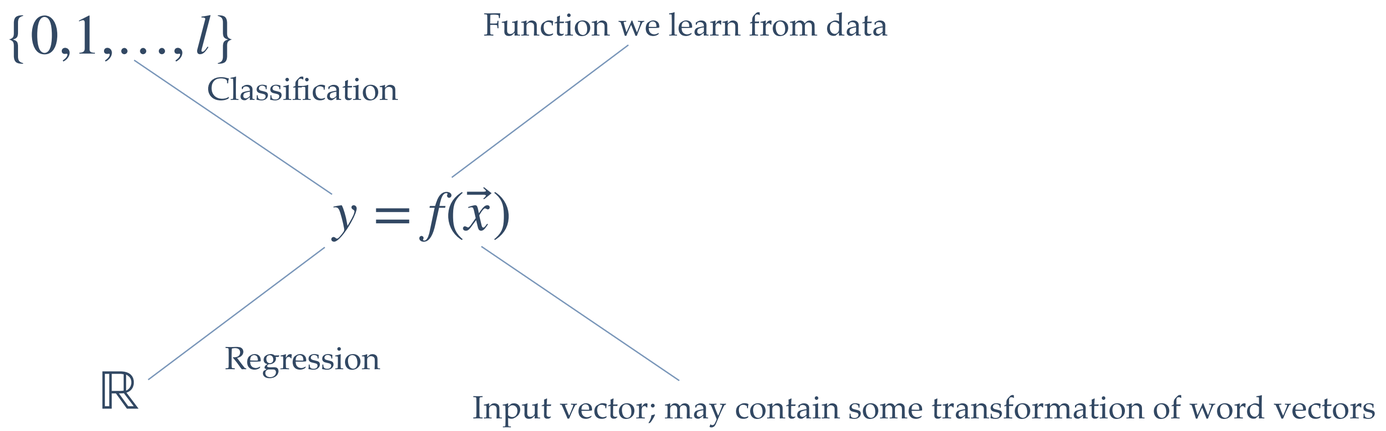
Structure $\quad Zr - HCP \approx Cr - BCC \approx Ni - FCC$

Embeddings can be used to construct knowledge bases that can lead to new discoveries



Tshitoyan, Vahe, et al. "Unsupervised word embeddings capture latent knowledge from materials science literature." *Nature* 571.7763 (2019): 95-98.

Prediction


$\{0,1,\ldots,l\}$
Classification
Function we learn from data
$y = f(\vec{x})$
$\mathbb{R}$
Regression
Input vector; may contain some transformation of word vectors

Prediction

Are Word Embedding-based Features Useful for Sarcasm Detection?

Aditya Joshi[1,2,3] **Vaibhav Tripathi**[1] **Kevin Patel**[1]
Pushpak Bhattacharyya[1] **Mark Carman**[2]
[1]Indian Institute of Technology Bombay, India
[2]Monash University, Australia
[3]IITB-Monash Research Academy, India
{adityaj,kevin.patel,pb}@cse.iitb.ac.in, mark.carman@monash.edu

	Word2Vec	LSA	GloVe	Dep. Wt.
+S	0.835	0.86	0.918	**0.978**
+WS	**1.411**	0.255	0.192	1.372
+S+WS	**1.182**	0.24	0.845	0.795

Table 4: Average gain in F-Scores obtained by using intersection of the four word embeddings, for three word embedding feature-types, augmented to four prior works; Dep. Wt. indicates vectors learned from dependency-based weights

Word Embedding	Average F-score Gain
LSA	0.452
Glove	0.651
Dependency	1.048
Word2Vec	1.143

Table 5: Average gain in F-scores for the four types of word embeddings; These values are computed for a subset of these embeddings consisting of words common to all four

Bias

man:woman::king:?

man:woman::waiter:?

man:woman::doctor:?

man:woman::king:?	queen	okay
man:woman::waiter:?	waitress	okay
man:woman::doctor:?	nurse	huh

man:woman::king:?	queen	okay
man:woman::waiter:?	waitress	okay
man:woman::doctor:?	nurse	huh

RESEARCH

REPORT

COGNITIVE SCIENCE

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan,[1]* Joanna J. Bryson,[1,2]* Arvind Narayanan[1]*

Machine learning is a means to derive artificial intelligence by discovering patterns in existing data. Here, we show that applying machine learning to ordinary human language results in human-like semantic biases. We replicated a spectrum of known biases, as measured by the Implicit Association Test, using a widely used, purely statistical machine-learning model trained on a standard corpus of text from the World Wide Web. Our results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral as toward insects or flowers, problematic as toward race or gender, or even simply veridical, reflecting the status quo distribution of gender with respect to careers or first names. Our methods hold promise for identifying and addressing sources of bias in culture, including technology.

Bias

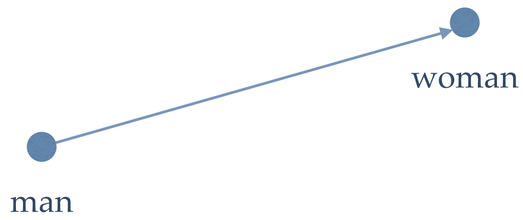
- Allocation harms: Unfair allocation of resources
- Representational harms: Unfair misrepresentation

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of “Bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

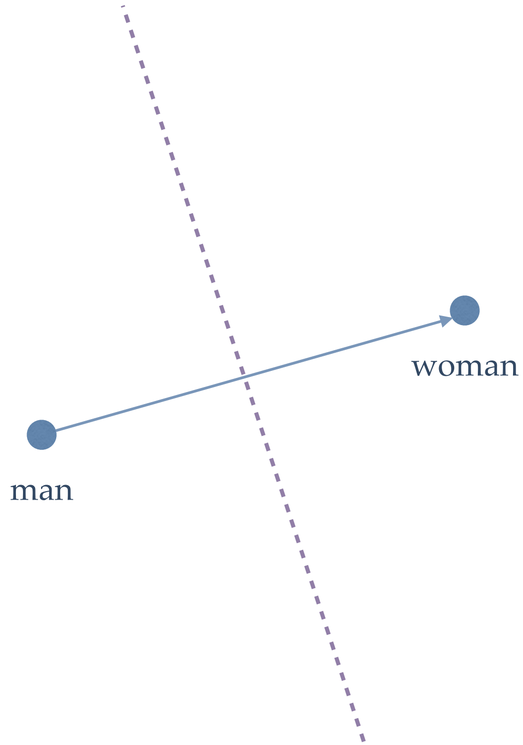
Calculating bias

Calculating bias

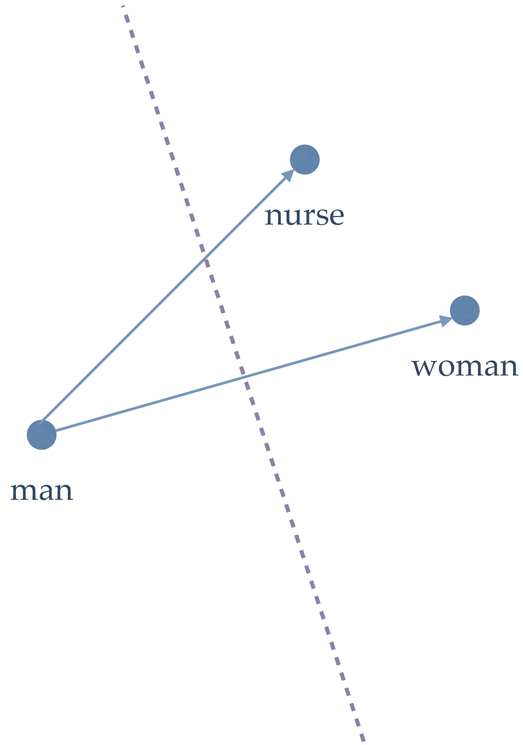
Calculating bias



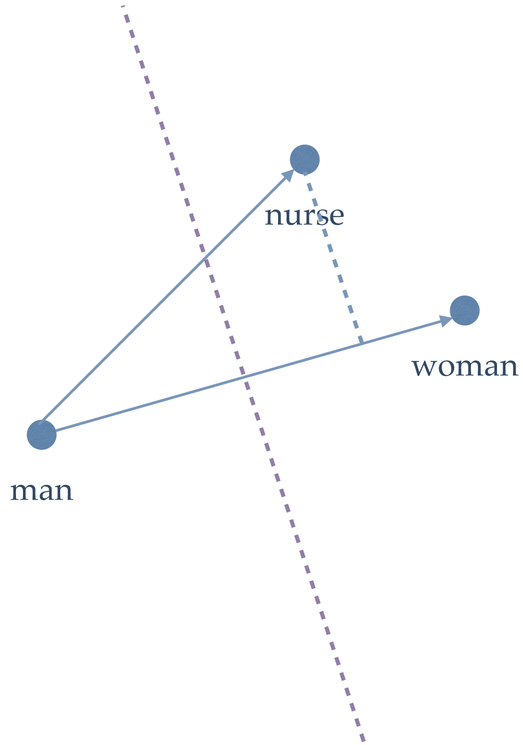
Calculating bias



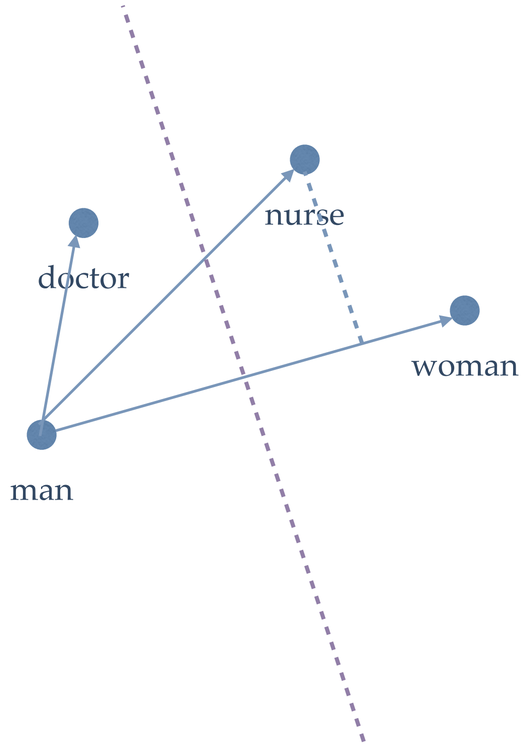
Calculating bias



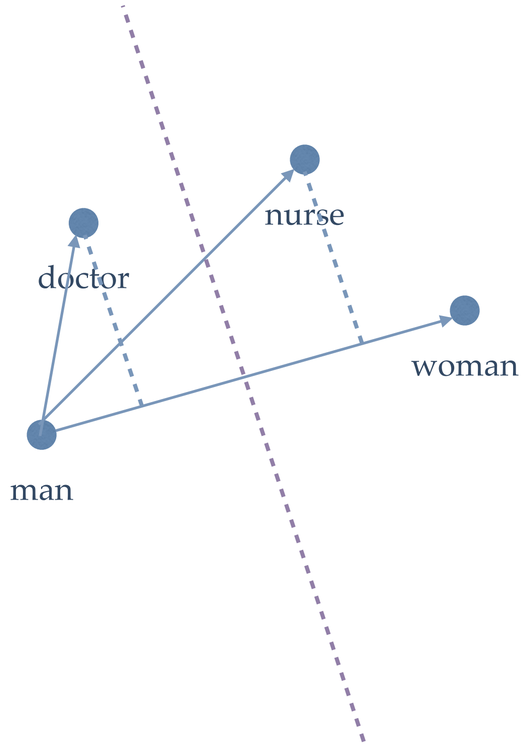
Calculating bias



Calculating bias



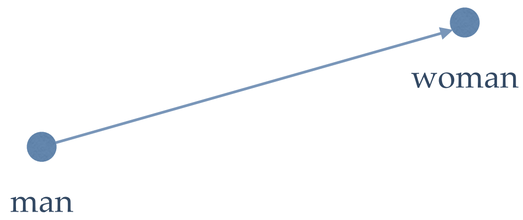
Calculating bias



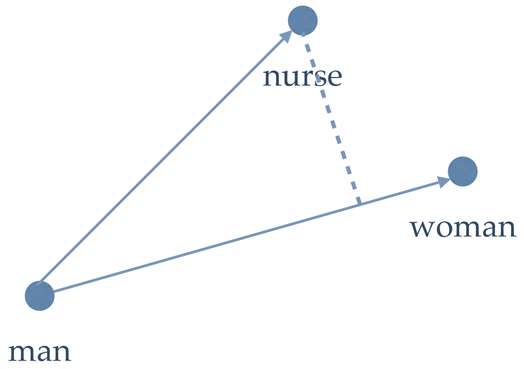
Calculating bias

Calculating bias

$$\vec{b} = \overrightarrow{\text{man}} - \overrightarrow{\text{woman}}$$



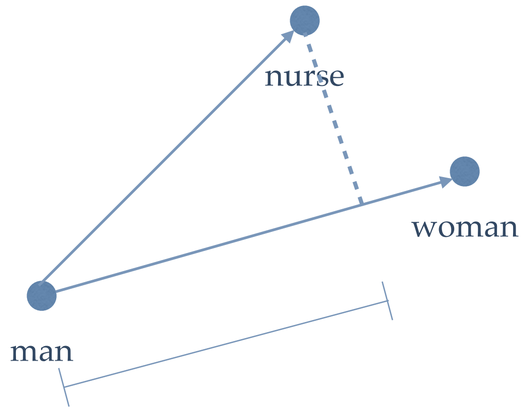
Calculating bias



$$\vec{b} = \overrightarrow{\text{man}} - \overrightarrow{\text{woman}}$$

$$\vec{v} = \overrightarrow{\text{nurse}}$$

Calculating bias

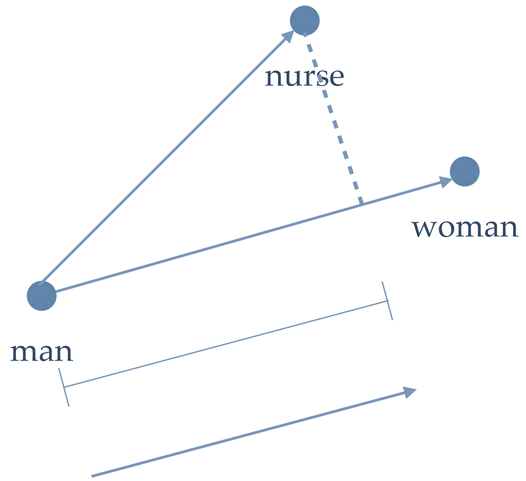


$$\vec{b} = \overrightarrow{\text{man}} - \overrightarrow{\text{woman}}$$

$$\vec{v} = \overrightarrow{\text{nurse}}$$

$$\text{Proj. length} = \vec{b} \cdot \vec{v}$$

Calculating bias



$$\vec{b} = \overrightarrow{\text{man}} - \overrightarrow{\text{woman}}$$

$$\vec{v} = \overrightarrow{\text{nurse}}$$

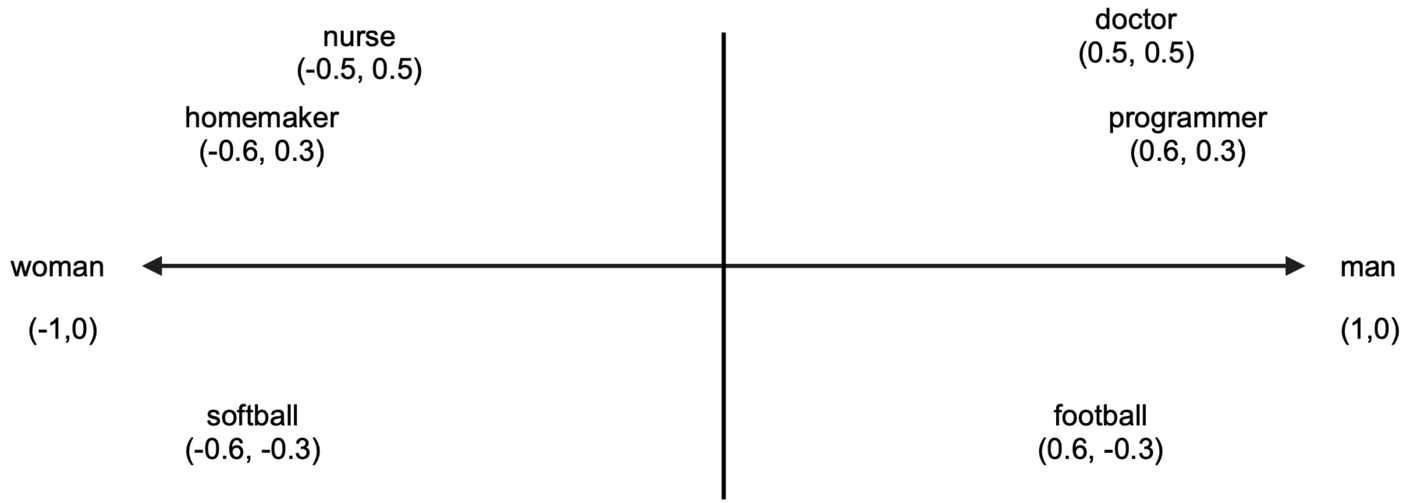
Proj. length $= \vec{b} \cdot \vec{v}$

Proj. vector $= (\vec{b} \cdot \vec{v})\ \vec{b}$

Projection of embeddings on a subspace

$$\vec{b} = \overrightarrow{\text{man}} - \overrightarrow{\text{woman}}$$

$$\vec{p} = (\vec{v} \cdot \vec{b})\vec{b}$$



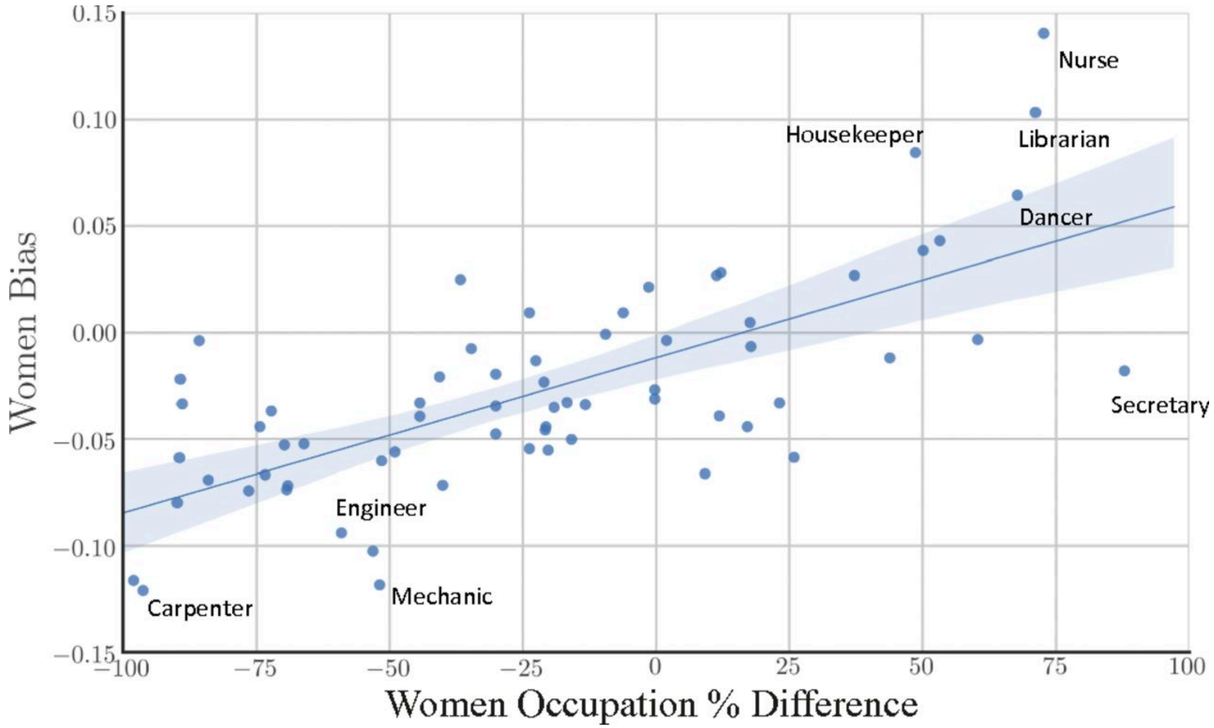


Image from Garg et. al. (2018)

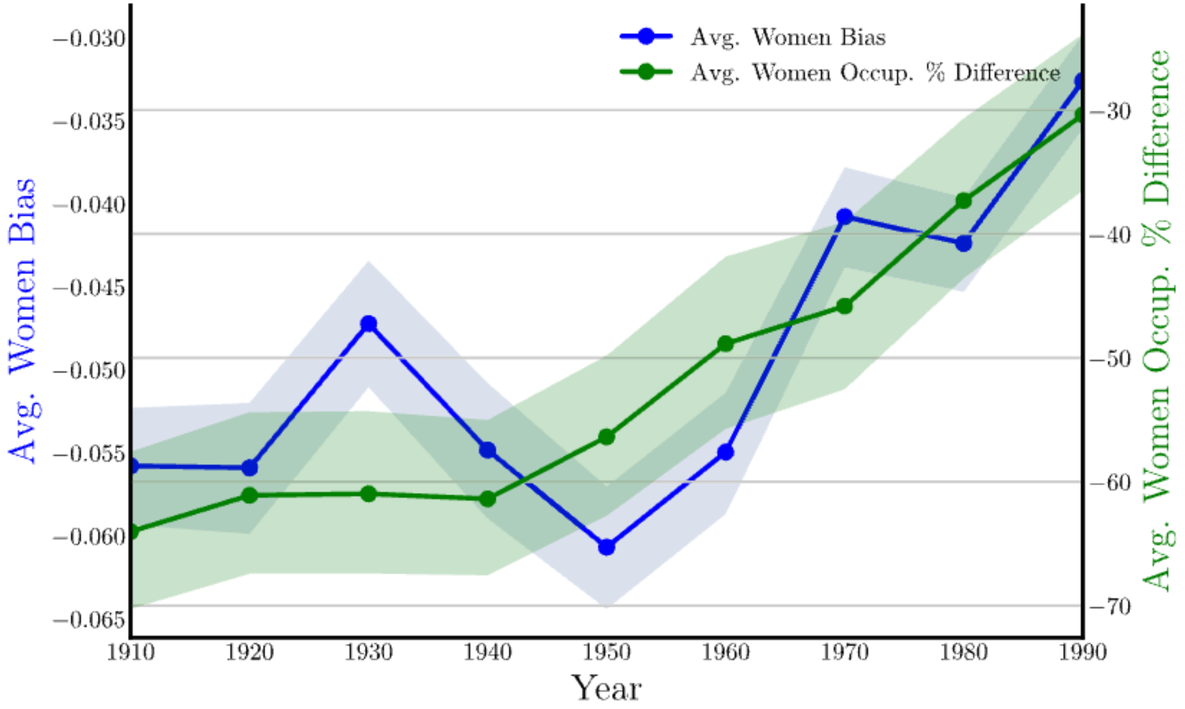


Image from Garg et. al. (2018)

Change and Variation

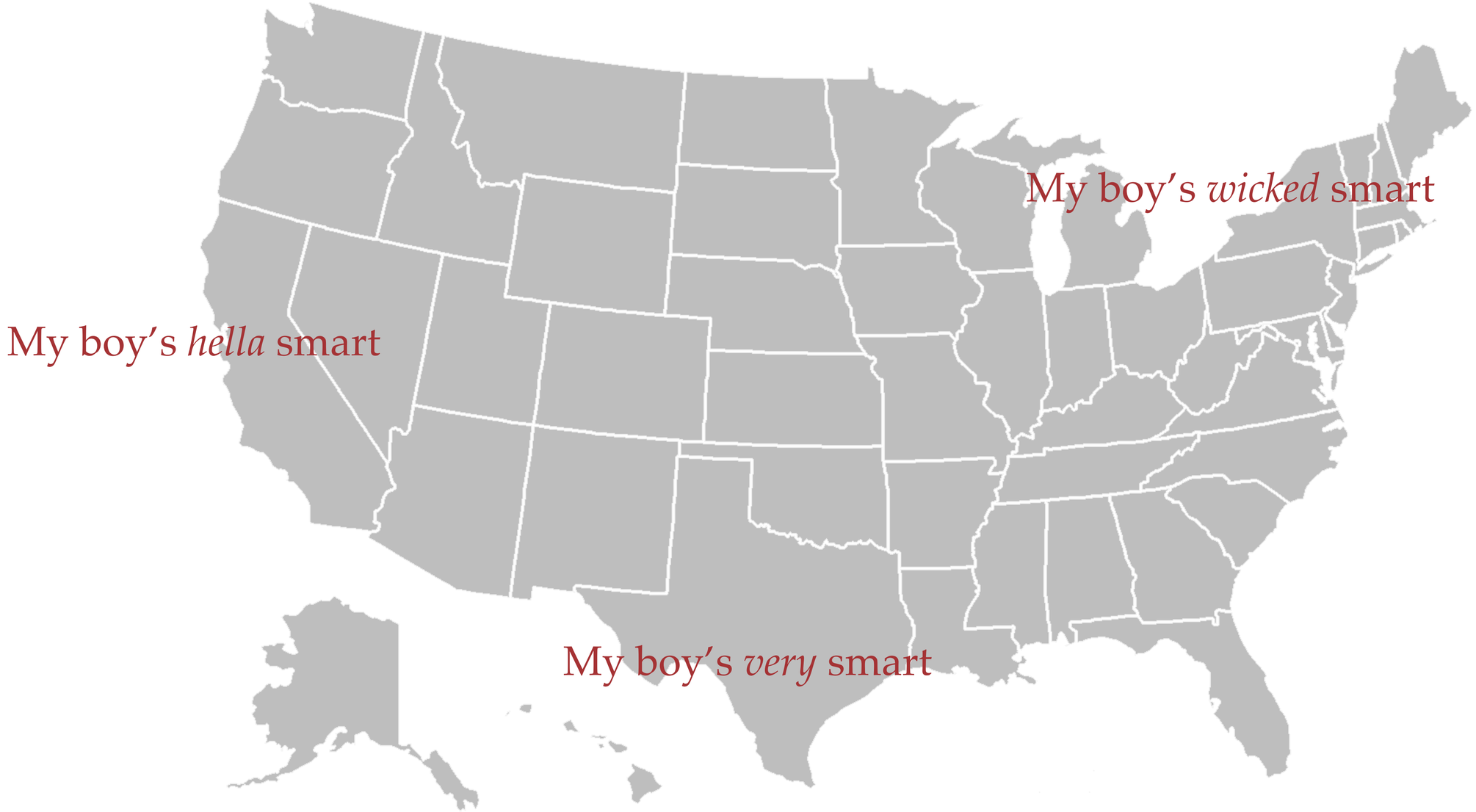
Change and Variation

- Since language is situational, one can learn embeddings that depend on time, geography or other social contexts

My boy's *hella* smart

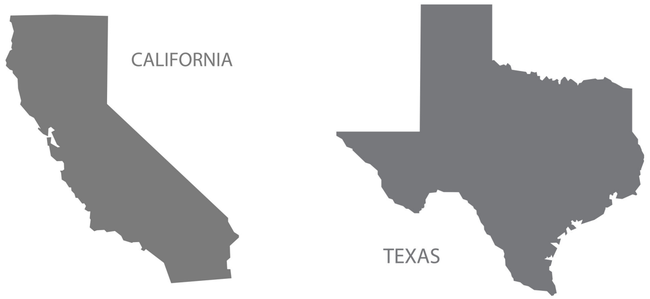
My boy's *wicked* smart

My boy's *very* smart



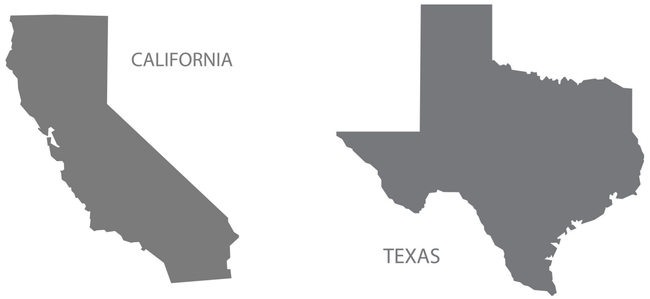
- David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.

David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	

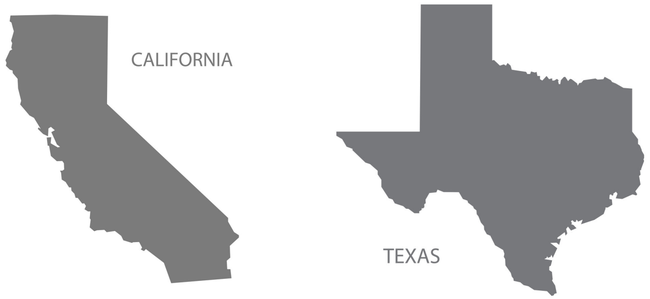
David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



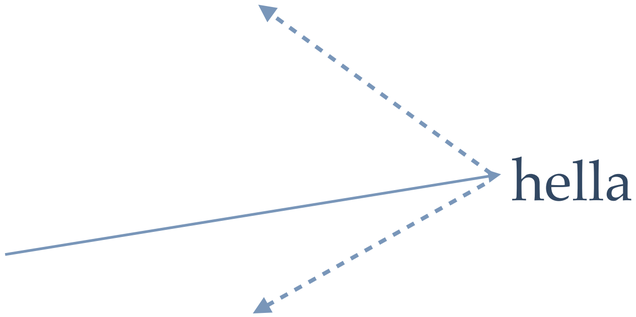
x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	

hella

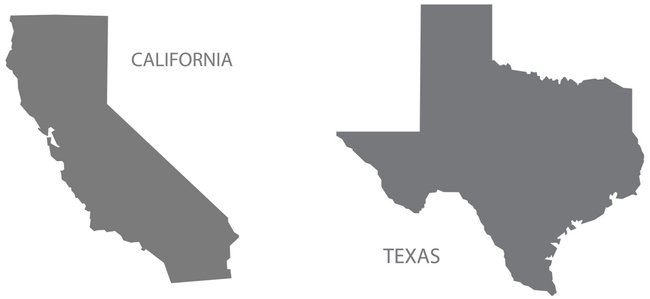
David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



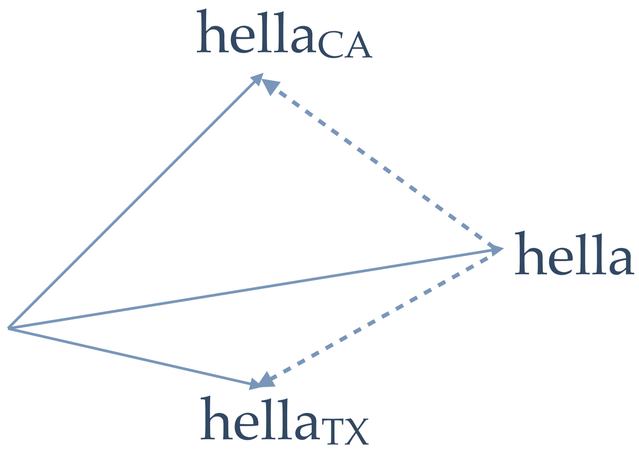
x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	



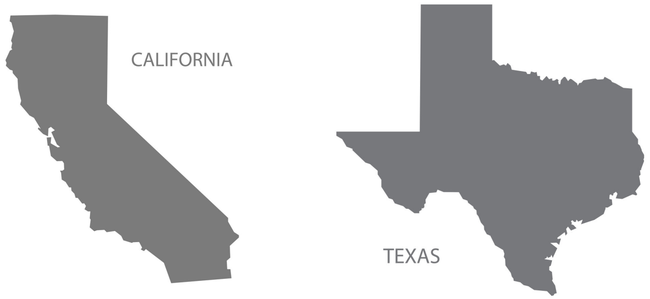
David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



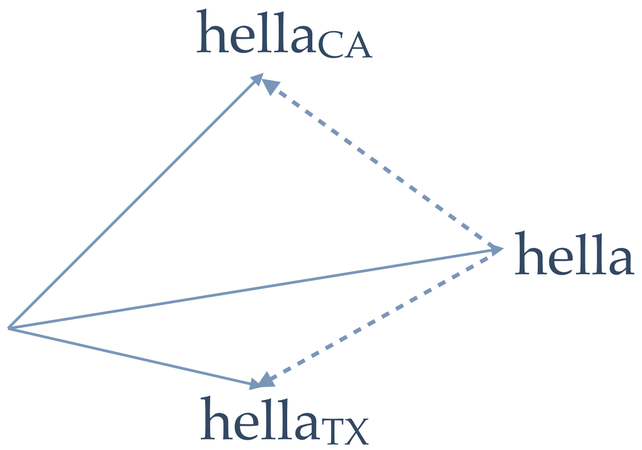
x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	



David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.

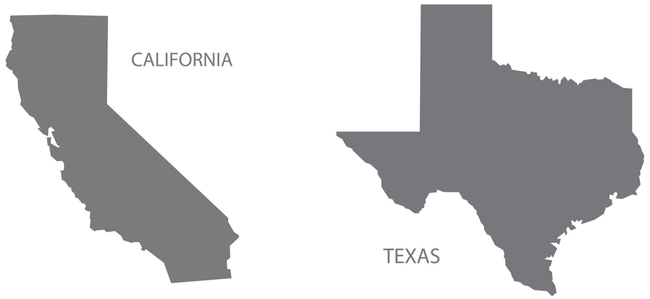


x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	

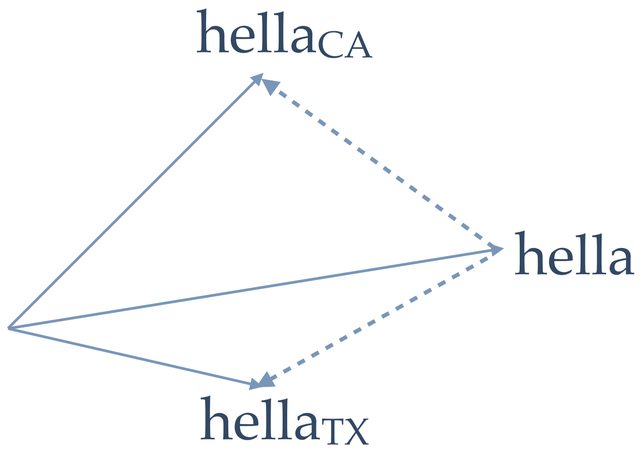


$$\overrightarrow{\text{hella}_{CA}} \stackrel{?}{\approx} \overrightarrow{\text{hella}_{TX}}$$

David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



x	y	State
…	…	…
wine	*cold*	*AZ*
wine	*sweet*	*GA*
hella	*smart*	*CA*
very	*smart*	*TX*
…	…	



$$\overrightarrow{\text{hella}_{CA}} \stackrel{?}{\approx} \overrightarrow{\text{hella}_{TX}}$$

$$\overrightarrow{\text{hella}_{CA}} \stackrel{?}{\approx} \overrightarrow{\text{very}_{TX}}$$

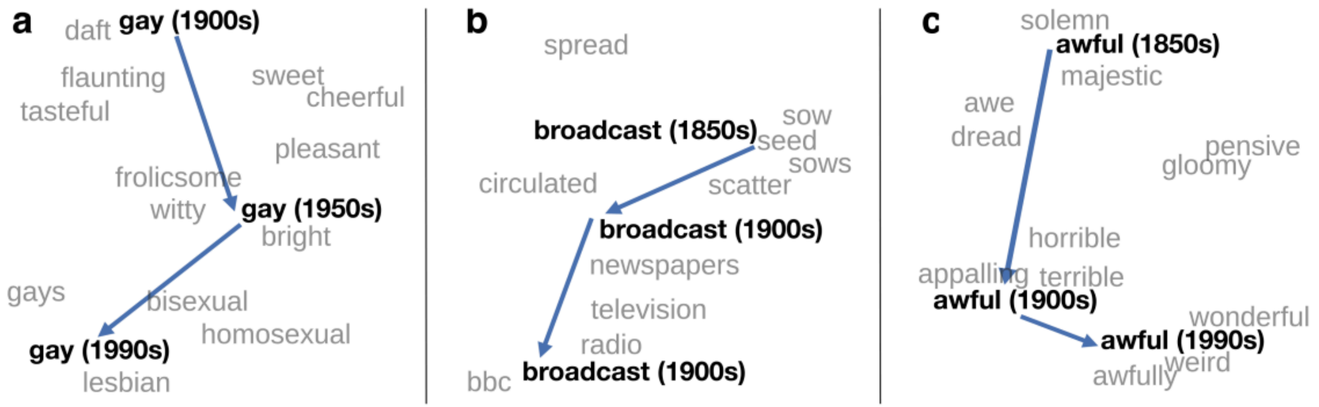
David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.

- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

x	y	Year
…	…	…
wine	*cold*	*1990*
wine	*sweet*	*1990*
gay	*time*	*1990*
gay	*person*	*1900*
…	…	

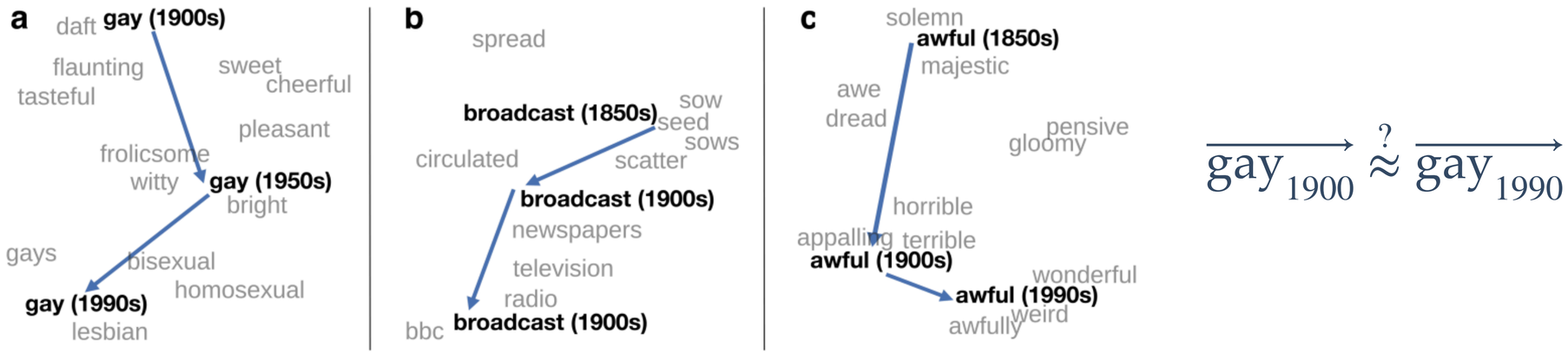
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

x	y	Year
…	…	…
wine	*cold*	*1990*
wine	*sweet*	*1990*
gay	*time*	*1990*
gay	*person*	*1900*
…	…	



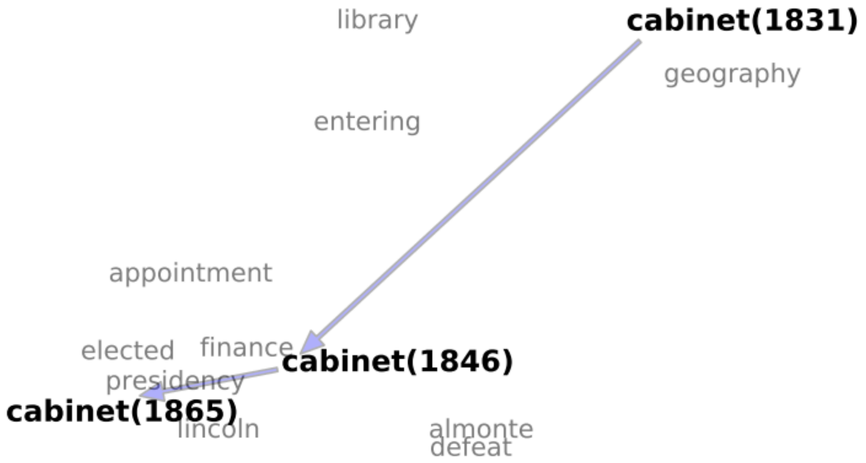
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

x	y	Year
…	…	…
wine	*cold*	*1990*
wine	*sweet*	*1990*
gay	*time*	*1990*
gay	*person*	*1900*
…	…	

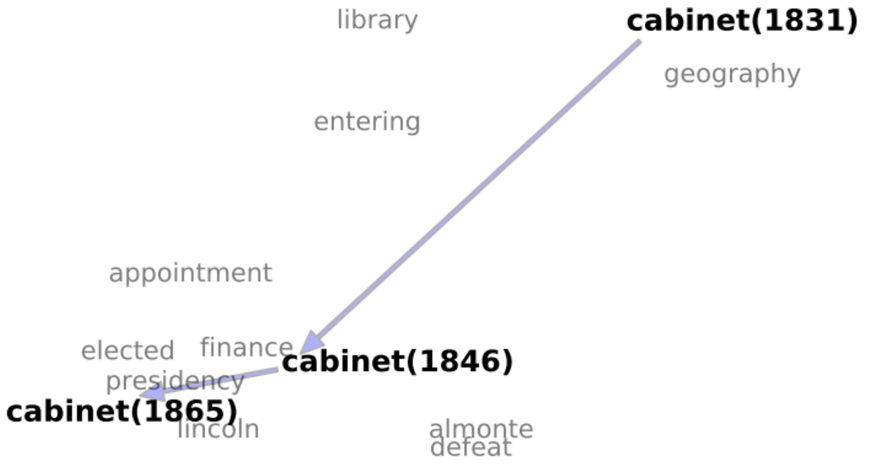


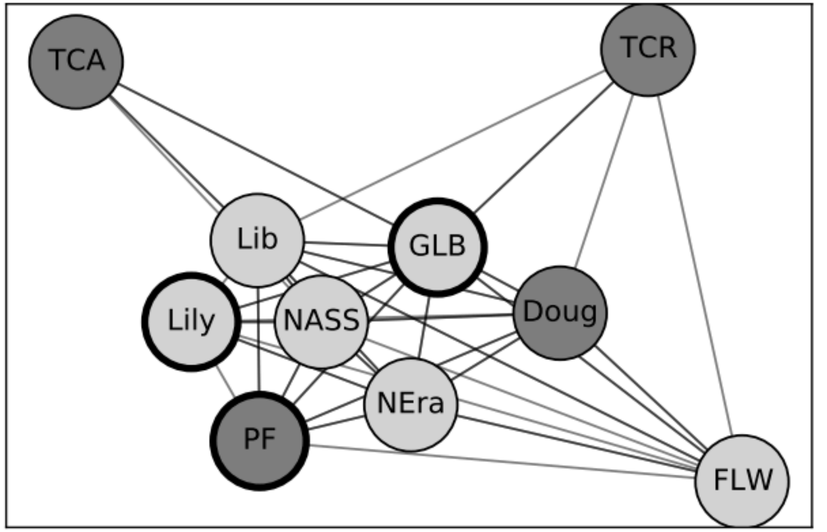
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Soni, Sandeep, Lauren Klein, and Jacob Eisenstein. "Correcting whitespace errors in digitized historical texts." *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2019.

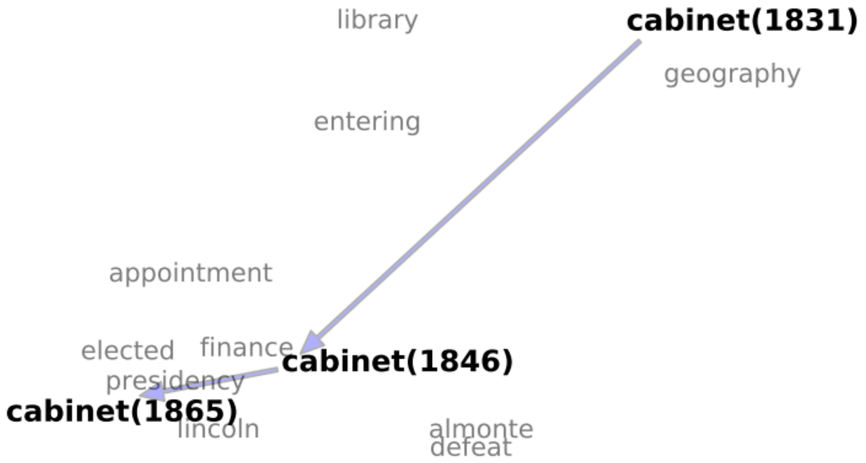


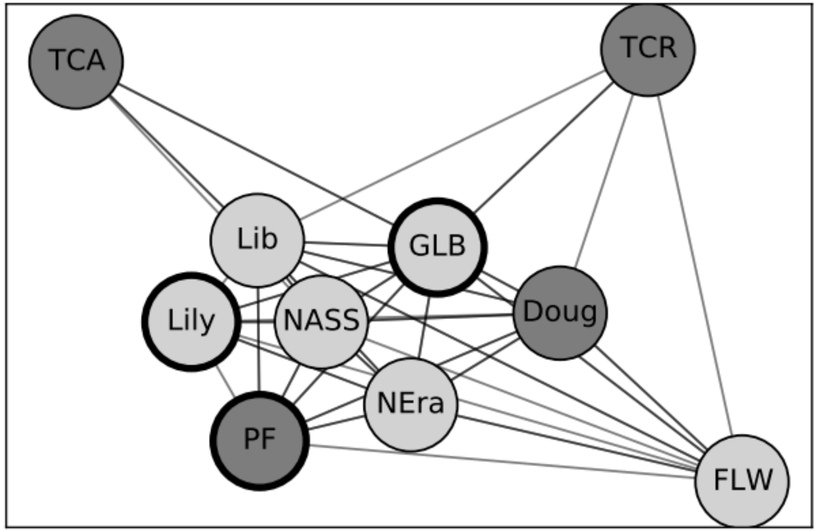
Soni, Sandeep, Lauren Klein, and Jacob Eisenstein. "Correcting whitespace errors in digitized historical texts." *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2019.





Soni, Sandeep, Lauren Klein, and Jacob Eisenstein. "Correcting whitespace errors in digitized historical texts." *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2019.





$$\text{Lead (PF, Lib, cabinet)} = \frac{\overrightarrow{\text{cabinet}_{1831,PF}} \cdot \overrightarrow{\text{cabinet}_{1865,Lib}}}{\overrightarrow{\text{cabinet}_{1831,Lib}} \cdot \overrightarrow{\text{cabinet}_{1865,Lib}}}$$

Soni, Sandeep, Lauren Klein, and Jacob Eisenstein. "Correcting whitespace errors in digitized historical texts." *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 2019.

Soni, Sandeep, Lauren F. Klein, and Jacob Eisenstein. "Abolitionist Networks: Modeling Language Change in Nineteenth-Century Activist Newspapers." *Journal of Cultural Analytics* 6.1 (2021).

In class

- Word2Vec demo