



Computational linguistic analysis of dehumanization of LGBTQ groups in the media

Julia Mendelsohn
Stanford University

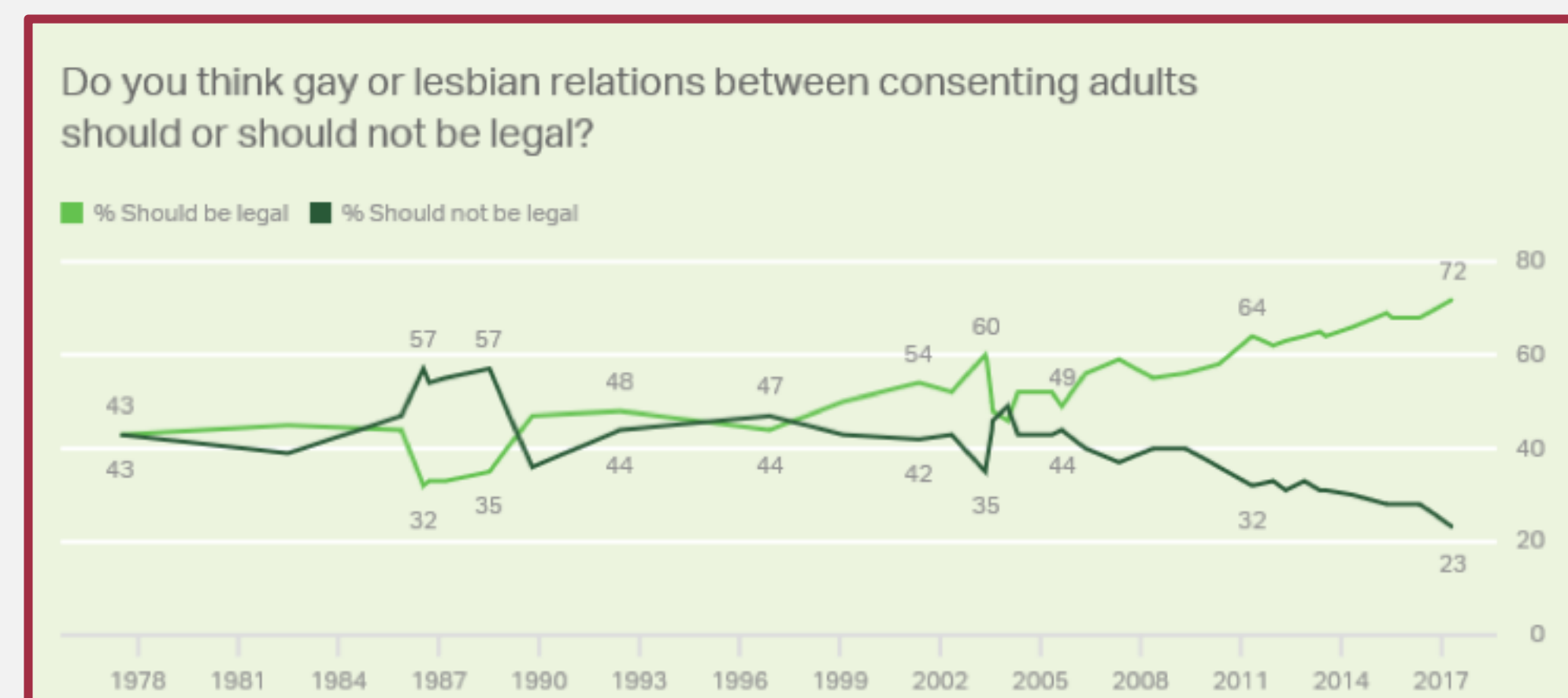


Goals

- Propose & evaluate techniques to quantify dehumanization
- Understand how LGBTQ representation has changed over time
- Explore different meanings of terms *gay* and *homosexual*

Background

- Dehumanization**: “the act of perceiving or treating people as less than human” (Haslem & Stratemeyer; 2016).
- Typically consists of multiple component psychological processes
- LGBTQ acceptance in USA has increased in recent decades (Gallup)



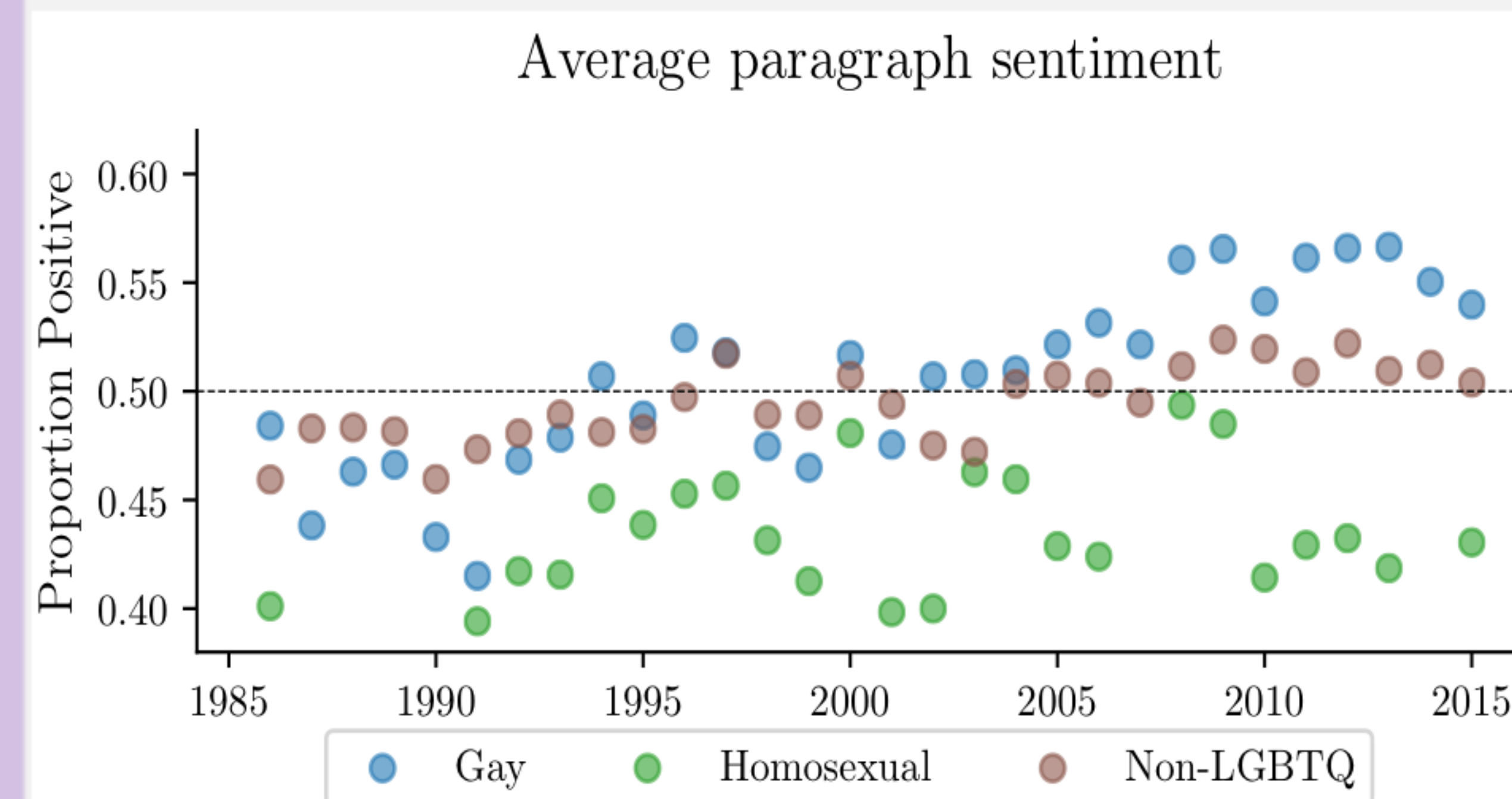
- Term *homosexual* considered outdated, clinical, derogatory (GLAAD)

Data

- Paragraphs from *New York Times* (NYT) containing LGBTQ terms
- Jan. 1986 - Apr. 2016; over 110,000 paragraphs and 7 million words

Negative Evaluation of Out-Group

- Approximated by sentiment analysis using Bing Liu's sentiment lexicon
- Proportion of positive words (relative to negative) averaged by year



Negative language is not always sign of dehumanization. More negative words are used in debates and reporting on LGBTQ-related issues

Future Work

- Improve methods for precise, robust, and interpretable measurements
- Incorporate other dehumanization components (e.g. denial of agency)
- Analyze more groups, different newspapers, and social media

Experiment Overview

Dehumanization Component	Linguistic/Quantitative Correlate
Lexical association exploration	Log-odds & word embeddings
Disgust	Vector-similarity disgust analysis
Vermin Metaphor	Vector-similarity metaphor analysis
Negative evaluation of out-group	Sentiment analysis
Psychological distance	Definite plurals
Essentialization	Pejorative nominalization

Lexical Associations

Log-odds: find words overrepresented in one corpus relative to another

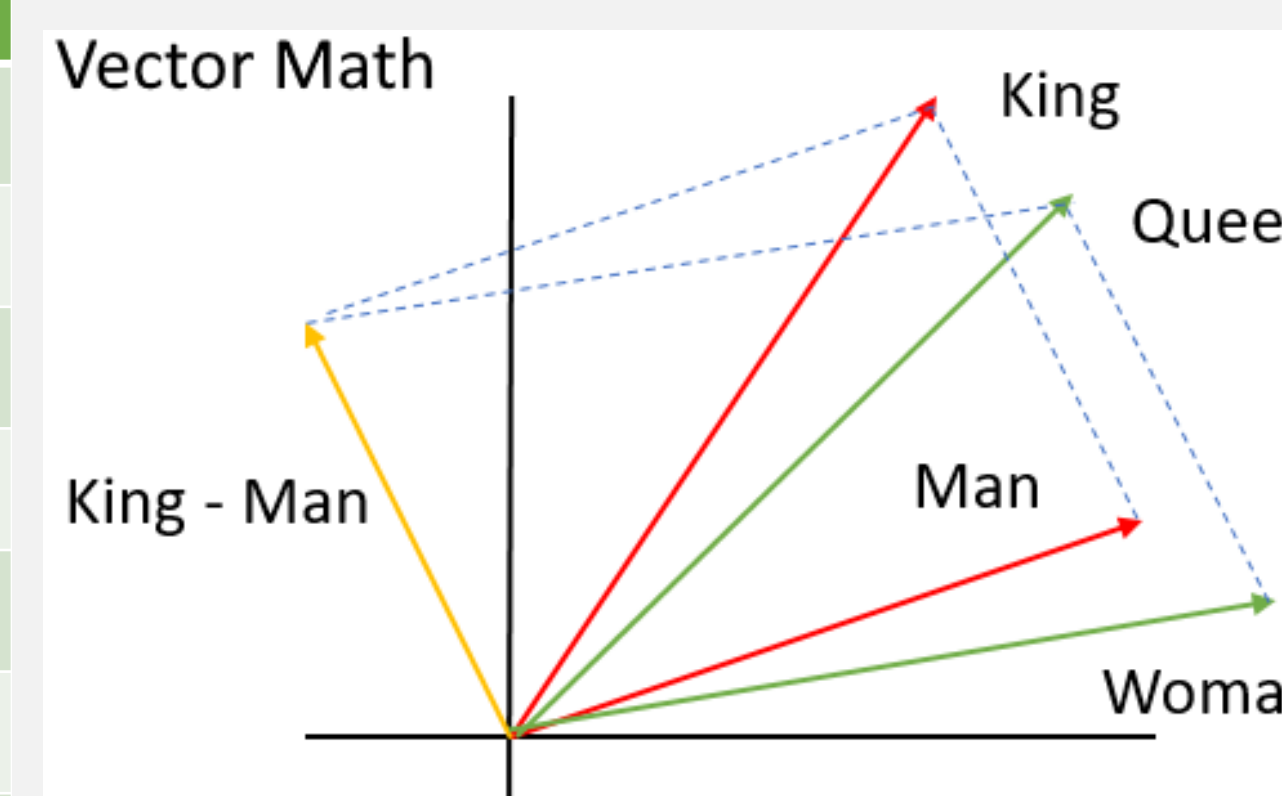
Calculated log-odds ratios of words in earliest decade relative to latest decade

Most personal pronouns (*you, I, my, her*, etc.) in top 50 words associated with later years. Personal stories & experiences in later years → less denial of subjectivity

Early Years	Later Years
homosexuals	marriage
homosexual	same-sex
aids	couples
drug	to
don't	you
clinton	on
parade	marriages
disease	supreme
virus	unions
military	state
groups	married

Word embeddings: represent words as points in vector space based on word co-occurrences. Geometric relationships between word vectors capture aspects of meaning, semantic change, and stereotypes.

Gay	Homosexual
sex	cohabiting
queers	indissolubility
marriage	excommunication
interracial	indissoluble
couples	bestiality
homophobia	abomination
closeted	mormons
openly	immoral
ordain	religious
mormons	sin
polygamy	sex
cohabiting	irja
equality	adultery



(Above) Vector arithmetic reveals relationships between words, such as *man : king :: woman : queen* analogy (L. Shure; blogs.mathworks.com)

(Left) words closest to *gay* & *homosexual* in Word2Vec model trained on 2015 New York Times data

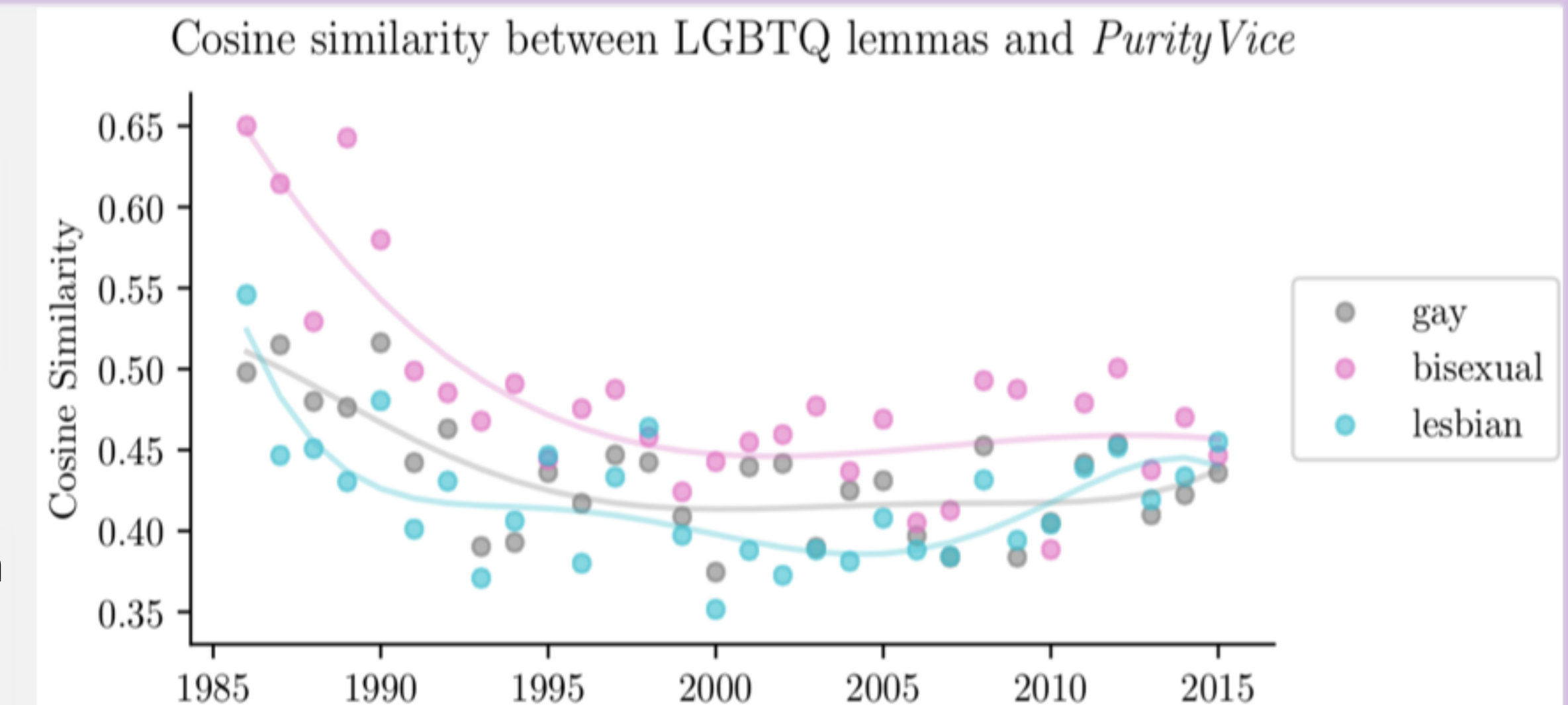
Conclusions

- LGBTQ individuals have been increasingly humanized in the *NYT*
- Homosexual* is used in more dehumanizing contexts than *gay*
- Implications for detecting harmful bias and abusive language online
- NLP tools for understanding aspects of language variation and change

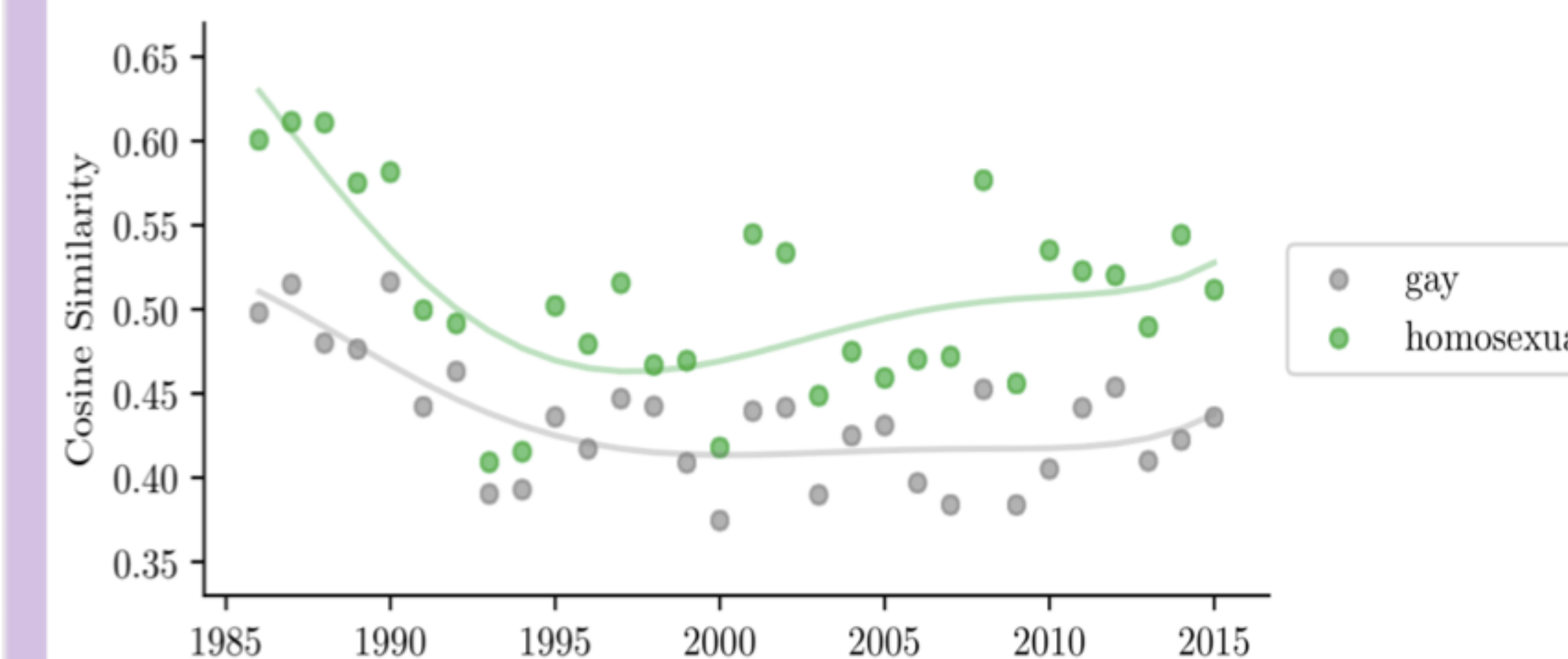
Moral Disgust

Similarities between LGBTQ vectors and moral disgust vector in Word2Vec captures extent of association

Disgust words from Moral Foundation Theory *PurityVice* dictionary



Cosine similarity between LGBTQ lemmas and *PurityVice*



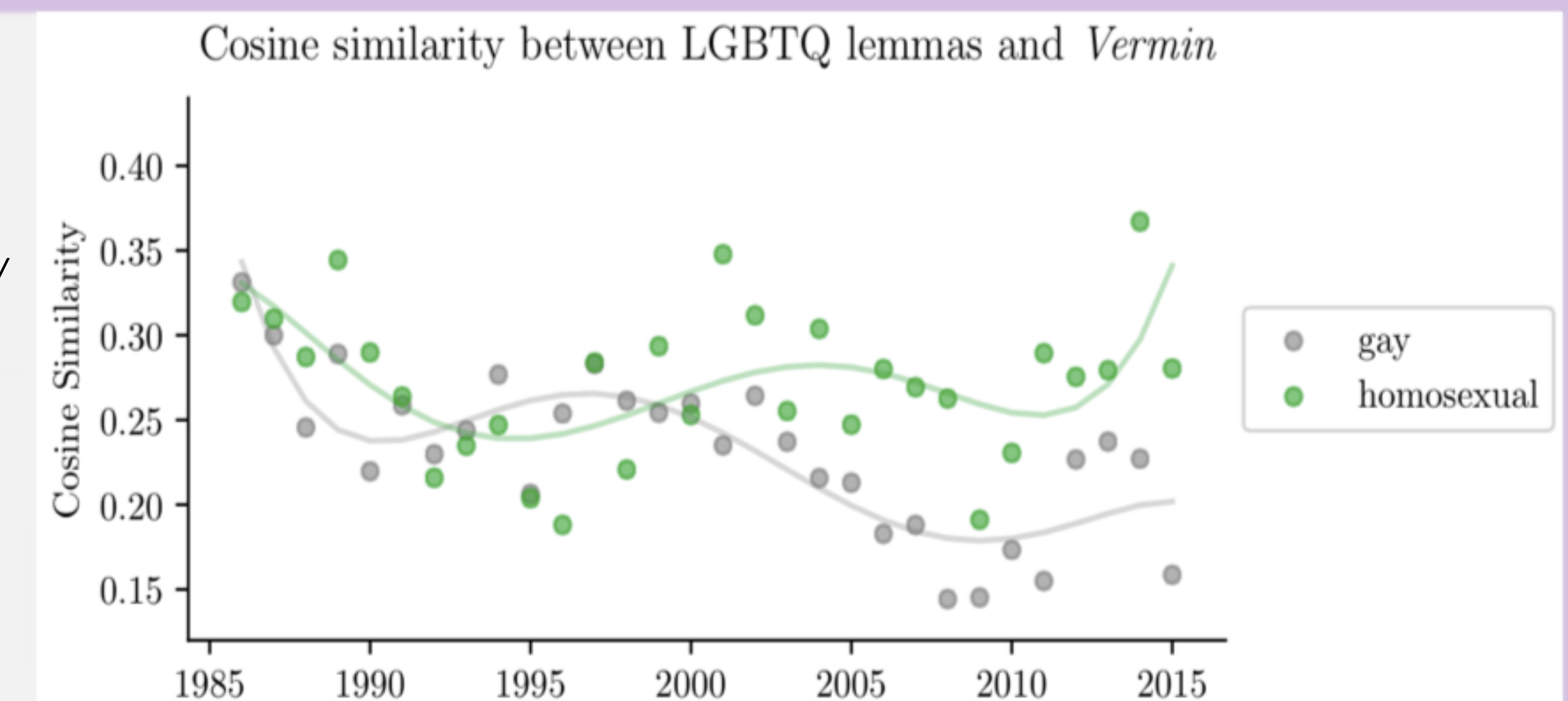
Disgust vector = weighted average of *PurityVice* words' vectors

Higher cosine similarity →
Vectors closer to each other →
Closer association with moral disgust

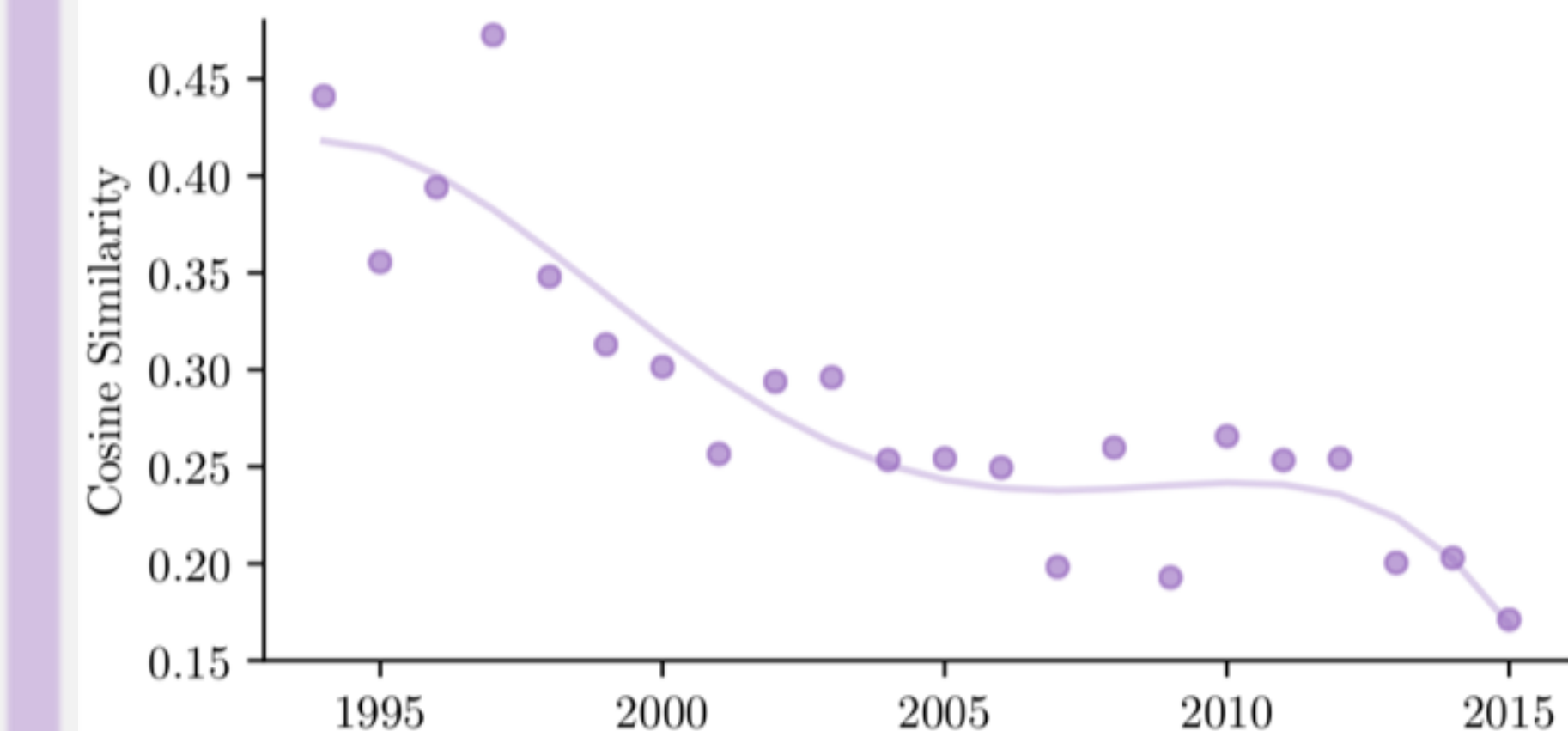
Vermin Metaphor

Dangerous, extreme, and rarely directly invoked, but exists in subtle forms

Vermin vector = Weighted average of vermin words (rats, cockroaches, etc.)



Cosine similarity of *transgender* lemmas relative to *Vermin*



Cosine similarity to measure associations between LGBTQ labels and vermin representation.

Higher cosine similarity →
Closer association with vermin

Acknowledgments

- Thanks to Dan Jurafsky, Yulia Tsvetkov, Chris Potts, and Stanford Linguistics Department for advice and feedback
- Thanks to Ethan Fast for supplying the *New York Times* data

References

- Haslam, N. & M. Stratemeyer. 2016. Recent research on dehumanization. *Current Opinion in Psychology*.
- Monroe, B.L., M.P. Colaresi & K.M. Quinn. 2008. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*.
- Garg, N., L. Schiebinger, D. Jurafsky & J. Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*.
- Hu, M. & B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining*.
- Graham, J., J. Haidt & B.A. Nosek. 2009. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*.

Contact: jmendels@stanford.edu