Putting language in context

Social networks and discourse structures for robust NLP

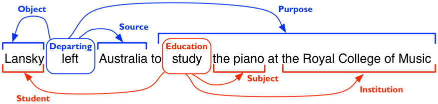
Jacob Eisenstein
@jacobeisenstein

Georgia Institute of Technology

February 15, 2017

Machine reading

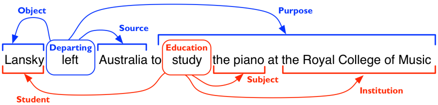
From text to structured representations.



Machine reading

From text to structured representations.

Annotate and train

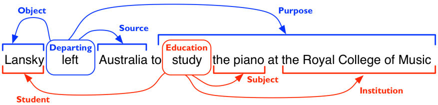


CONGRESS.GOV
United States Legislative Information

Annotate and train

Machine reading

From text to structured representations.



New domains of digitized texts offer opportunities as well as challenges.

Language data then and now

Then: news text, small set of authors, professionally edited, fixed style

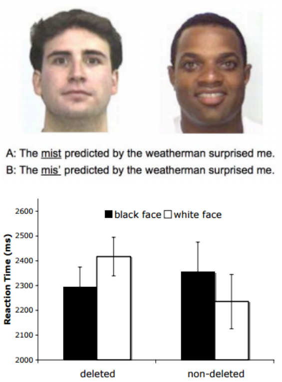
Language data then and now

Then: news text, small set of authors, professionally edited, fixed style

Now: open domain, everyone is an author, unedited, many styles

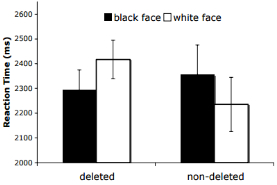
How people do it

- Listeners use **social context** in speech interpretation.
  - race (Casasanto, 2008)
  - nationality (Niedzielski, 1999)
- **Discourse context** supports the resolution of reference ambiguity (Kehler & Rohde, 2013).



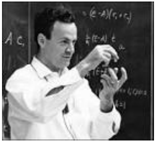
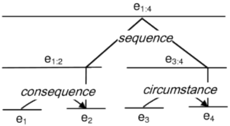
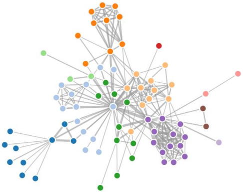
How people do it

- Listeners use **social context** in speech interpretation.
    - race (Casasanto, 2008)
    - nationality (Niedzielski, 1999)
- **Discourse context** supports the resolution of reference ambiguity (Kehler & Rohde, 2013).



Can we build software with the same ability to leverage context to resolve ambiguity?

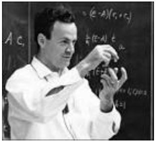
Language across contexts

Visual	(Eisenstein et al., 2008a,b; Eisenstein, 2008)	
Discourse	(Ji & Eisenstein, 2015; Bhatia et al., 2015; Ji et al., 2016)	
Social	(Eisenstein, 2013; Yang et al., 2016; Yang & Eisenstein, 2017)	

Language across contexts

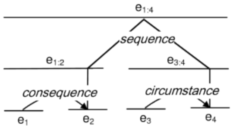
Visual

(Eisenstein et al., 2008a,b; Eisenstein, 2008)



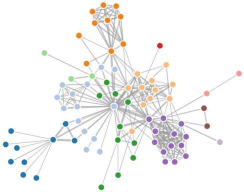
Discourse

(Ji & Eisenstein, 2015; Bhatia et al., 2015; Ji et al., 2016)



Social

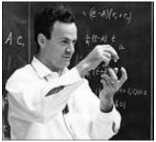
(Eisenstein, 2013; Yang et al., 2016; Yang & Eisenstein, 2017)



Language across contexts

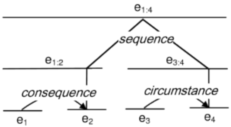
Visual

(Eisenstein et al., 2008a,b; Eisenstein, 2008)



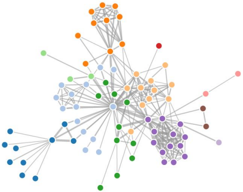
Discourse

(Ji & Eisenstein, 2015; Bhatia et al., 2015; Ji et al., 2016)



Social

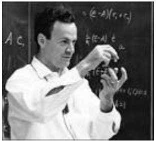
(Eisenstein, 2013; Yang et al., 2016; Yang & Eisenstein, 2017)



Language across contexts

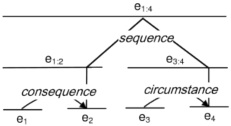
Visual

(Eisenstein et al., 2008a,b; Eisenstein, 2008)



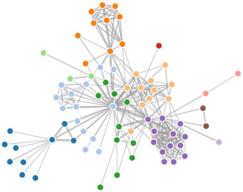
Discourse

(Ji & Eisenstein, 2015; Bhatia et al., 2015; Ji et al., 2016)



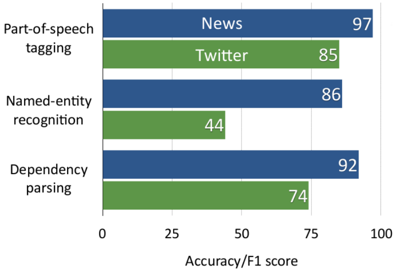
Social

(Eisenstein, 2013; Yang et al., 2016; Yang & Eisenstein, 2017)



Natural language processing in the wild

Social media has forced NLP to confront the challenge of missing social context (Eisenstein, 2013):

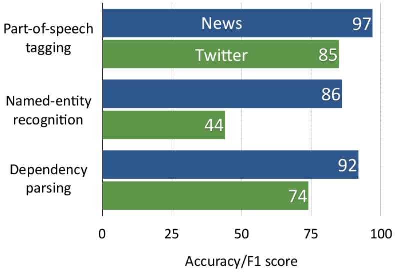


(Gimpel et al., 2011)
(Ritter et al., 2011)
(Foster et al., 2011)

Natural language processing in the wild

Social media has forced NLP to confront the challenge of missing social context (Eisenstein, 2013):

- tacit assumptions about audience knowledge
- language variation across social groups

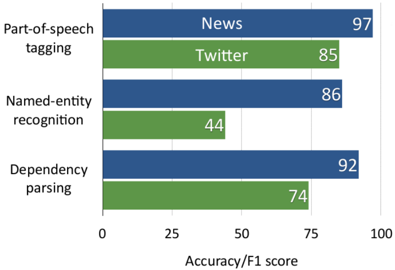


(Gimpel et al., 2011)
(Ritter et al., 2011)
(Foster et al., 2011)

Natural language processing in the wild

Social media has forced NLP to confront the challenge of missing social context (Eisenstein, 2013):

- tacit assumptions about audience knowledge
- language variation across social groups



(Gimpel et al., 2011)
(Ritter et al., 2011)
(Foster et al., 2011)

Shea Serrano
@SheaSerrano

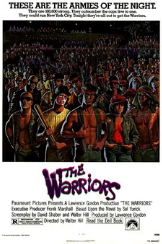
Follow

an absolutely perfect response by the warriors

Shea Serrano
@SheaSerrano

Follow

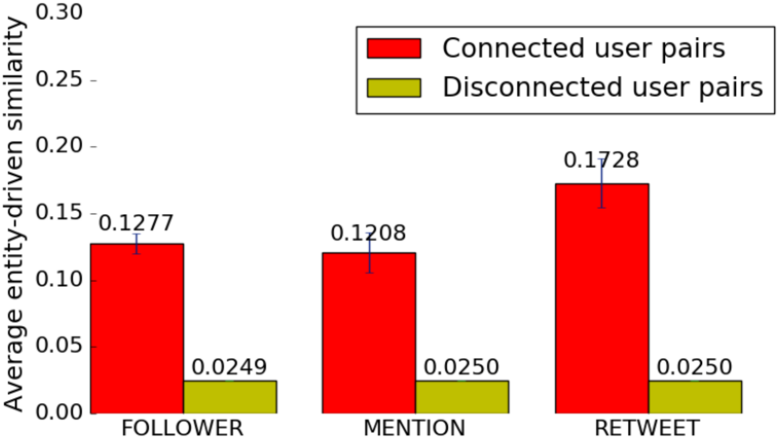
an absolutely perfect response by $\boxed{\text{the warriors}}$



Finding tacit context in the social network

- Social media texts lack context, because it is implicit between the writer and the reader.
- **Homophily**: socially connected individuals tend to share traits.

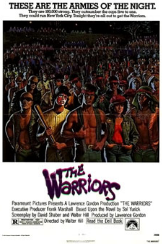
Assortativity of entity references



Shea Serrano
@SheaSerrano

Follow

an absolutely perfect response by $\boxed{\text{the warriors}}$



GOLDEN STATE WARRIORS

/r/NBA
@NBA_Reddit

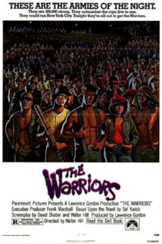
Lana Berry
@Lana

Michael Lee
@MrMichaelLee

Shea Serrano
@SheaSerrano

Follow

an absolutely perfect response by the warriors



GOLDEN STATE WARRIORS

/r/NBA
@NBA_Reddit

Lana Berry
@Lana

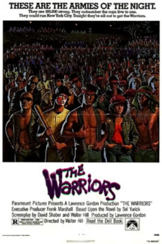
Michael Lee
@MrMichaelLee

The return of Clutch Dirk Nowitzki is one of the more exciting, unexpected developments in an already bonkers NBA season

Shea Serrano
@SheaSerrano

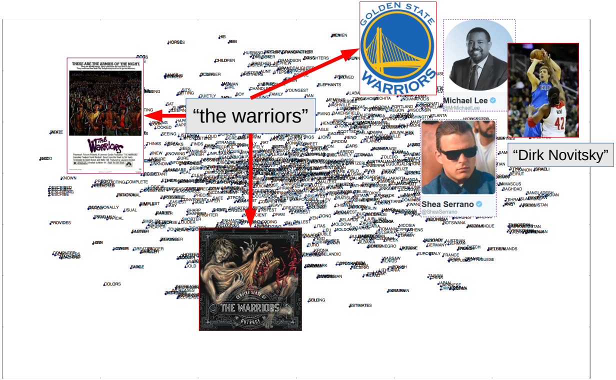
Follow

an absolutely perfect response by the warriors



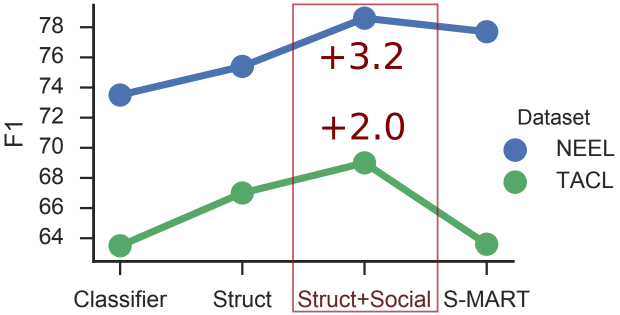
Projecting into a joint semantic space

We project embeddings for entities, words, and authors into a shared semantic space.



Inner products in this space indicate compatibility.

Results

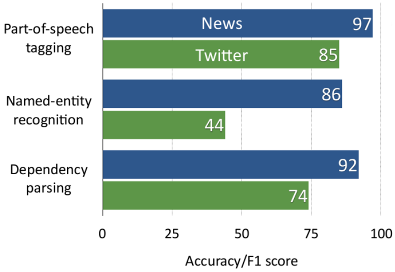


- Structure prediction improves accuracy.
- Social context yields further improvements.
- S-MART is the prior state-of-the-art (Yang & Chang, 2015).

Natural language processing in the wild

Social media has forced NLP to confront the challenge of missing social context (Eisenstein, 2013):

- tacit assumptions about audience knowledge
- language variation across social groups

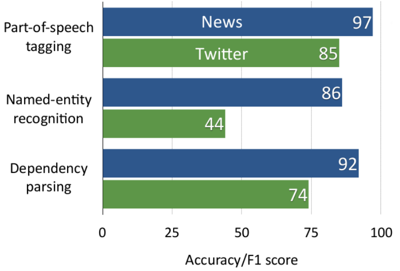


(Gimpel et al., 2011)
(Ritter et al., 2011)
(Foster et al., 2011)

Natural language processing in the wild

Social media has forced NLP to confront the challenge of missing social context (Eisenstein, 2013):

- tacit assumptions about audience knowledge
- language variation across social groups



(Gimpel et al., 2011)
(Ritter et al., 2011)
(Foster et al., 2011)

Language variation

"I would like to believe he's **sick** rather than just mean and evil."

Language variation

“I would like to believe he’s **sick** rather than just mean and evil.”

“You could’ve been getting down to this **sick** beat.”

(Yang & Eisenstein, 2017)

Personalization by ensemble

- Goal: personalized conditional likelihood, $P(y \mid \boldsymbol{x}, a)$, where $\boldsymbol{x}$ is the text and a is the author.

Personalization by ensemble

- Goal: personalized conditional likelihood, $P(y \mid \boldsymbol{x}, a)$, where $\boldsymbol{x}$ is the text and a is the author.
- **Personalization ensemble**

$$P(y \mid \boldsymbol{x}, a) = \sum_k P_k(y \mid \boldsymbol{x}) \pi_a(k)$$

  - $P_k(y \mid \boldsymbol{x})$ is a basis model
  - $\boldsymbol{\pi}_a(\cdot)$ are the ensemble weights for author a

Personalization by ensemble

- Goal: personalized conditional likelihood, $P(y \mid \mathbf{x}, a)$, where $\mathbf{x}$ is the text and a is the author.
- **Personalization ensemble**

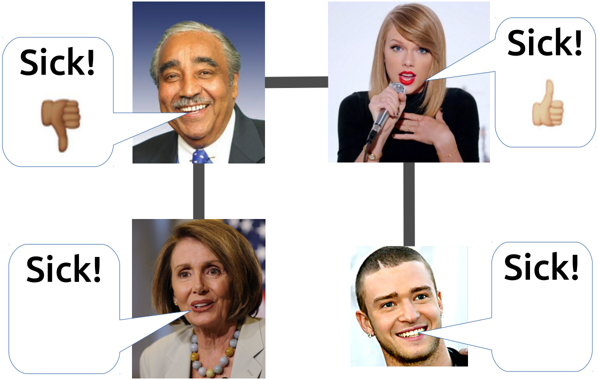
$$P(y \mid \mathbf{x}, a) = \sum_k P_k(y \mid \mathbf{x})\pi_a(k)$$

  - $P_k(y \mid \mathbf{x})$ is a basis model
  - $\boldsymbol{\pi}_a(\cdot)$ are the ensemble weights for author a
- Problem: We have labeled examples for only a few authors.

Homophily to the rescue?

Labeled data

Unlabeled data



Homophily again: are language styles assortative on the social network?

Evidence for linguistic homophily

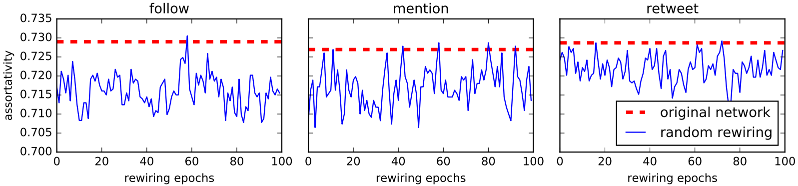
Pilot study: is classifier accuracy **assortative** on the Twitter social network?

$$\text{assort}(G) = \frac{1}{\#|G|} \sum_{(i,j)\in G} \delta(y_i = \hat{y}_i)\delta(y_j = \hat{y}_j) + \delta(y_i \neq \hat{y}_i)\delta(y_j \neq \hat{y}_j)$$

Evidence for linguistic homophily

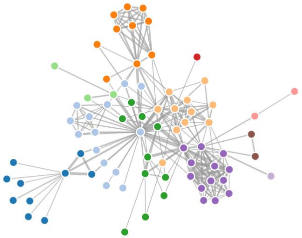
Pilot study: is classifier accuracy **assortative** on the Twitter social network?

$$\text{assort}(G) = \frac{1}{\#|G|} \sum_{(i,j)\in G} \delta(y_i = \hat{y}_i)\delta(y_j = \hat{y}_j) + \delta(y_i \neq \hat{y}_i)\delta(y_j \neq \hat{y}_j)$$



Network-driven personalization

- For each author, estimate a **node embedding** $\boldsymbol{e}_a$ (Tang et al., 2015).
- Nodes who share neighbors get similar embeddings.



$$\boldsymbol{\pi}_a = \mathrm{SoftMax}(f(\boldsymbol{e}_a))$$

$$P(y \mid \boldsymbol{x}, a) = \sum_{k=1}^{K} P_k(y \mid \boldsymbol{x}) \pi_a(k)$$

The SemEval social network

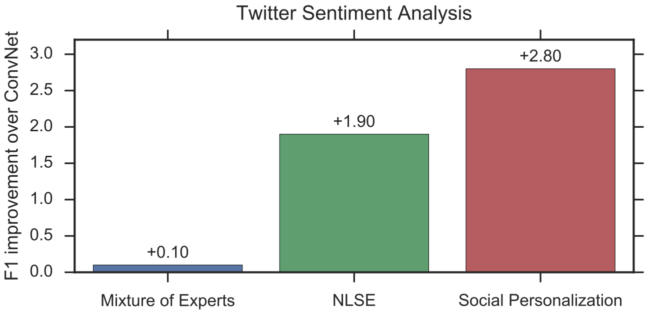
SemEval 2013-2015 is a standard benchmark for Twitter Sentiment Analysis.
(Nakov et al., 2013; Rosenthal et al., 2015).

	# Author	# Relations	# Isolates
original	14,087	40,110	3,633
expanded	17,417	1,050,369	689

Table: Follower network statistics

We "densify" the social network by adding individuals who are followed by at least 100 SemEval authors.

Results



Improvements over ConvNet baseline:

- +2.8% on Twitter Sentiment Analysis
- +2.7% on Ciao Product Reviews

NLSE is prior state-of-the-art (Astudillo et al., 2015).

Variable sentiment words

	More positive	More negative
1	banging loss fever broken fucking	dear like god yeah wow
2	chilling cold ill sick suck	satisfy trust wealth strong lmao
3	ass damn piss bitch shit	talent honestly voting win clever
4	insane bawling fever weird cry	lmao super lol haha hahaha
5	ruin silly bad boring dreadful	*lovatics* wish *beliebers arianators kendall*

Variable sentiment words

	More positive	More negative
1	banging loss fever broken fucking	dear like god yeah wow
2	chilling cold ill sick suck	satisfy trust wealth strong **lmao**
3	ass damn piss bitch shit	talent honestly voting win clever
4	insane bawling fever weird cry	**lmao** super **lol** haha hahaha
5	ruin silly bad boring dreadful	**lovatics** wish **beliebers ari-anators** kendall

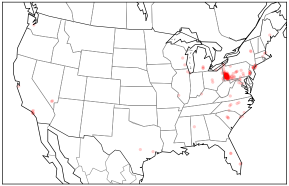
Language as social scientific evidence

- Propagating neologisms like lovatics and lmao requires:
  1. **Exposure**
  2. **Decision** to adopt (Rogers, 1962).

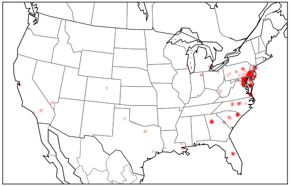
- By tracking the spread of these words, it is possible to reconstruct "deep networks" of social affinity and influence.
  (Eisenstein et al., 2014; Goel et al., 2016).

The important of place

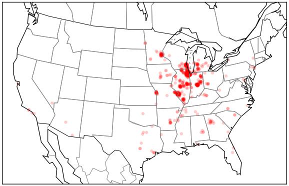
yinz



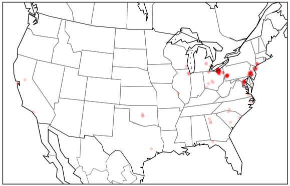
ard ("alright")



lbvs ("laughing but very serious")

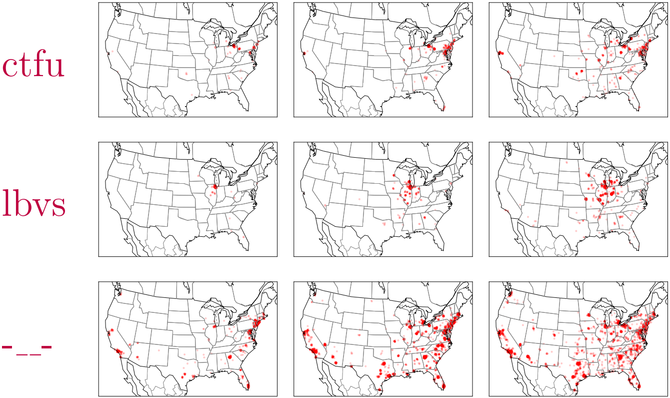


ctfu ("cracking the fuck up")



(Eisenstein et al., 2010, 2014)

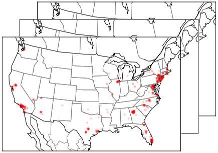
An aggregate model of lexical diffusion



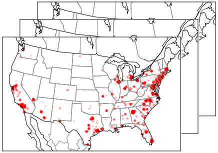
- Thousands of words have changing frequencies.
- Each spatiotemporal trajectory is idiosyncratic.
- What's the aggregate picture?

Language change as an autoregressive process

Word counts are binned into 200 metro areas and 165 weeks.



$\eta_2 \sim N(A\eta_1, \Sigma)$



$\eta_3 \sim N(A\eta_2, \Sigma)$

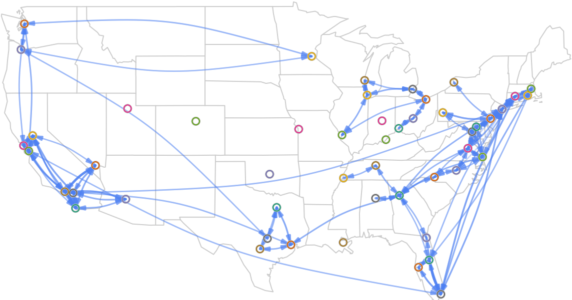
$$
\begin{aligned}
c_{\text{ctfu},1} &\sim \text{Binomial}(f(\eta_{\text{ctfu},1}), N_1) \\
c_{\text{hella},1} &\sim \text{Binomial}(f(\eta_{\text{hella},1}), N_1) \\
&\ldots
\end{aligned}
$$

$$
\begin{aligned}
c_{\text{ctfu},2} &\sim \text{Binomial}(f(\eta_{\text{ctfu},2}), N_2) \\
c_{\text{hella},2} &\sim \text{Binomial}(f(\eta_{\text{hella},2}), N_2) \\
&\ldots
\end{aligned}
$$

Estimating parameters of this autoregressive process reveals geographic pathways of diffusion across thousands of words (Eisenstein et al., 2014).

Aggregating city-to-city influence

Highly-confident pathways of diffusion
(from autoregressive parameter A).



Is geography the whole story?

Possible roles for demographics

- **Assortativity**: similar cities evolve together.
- **Influence**: certain types of cities tend to lead, others follow.

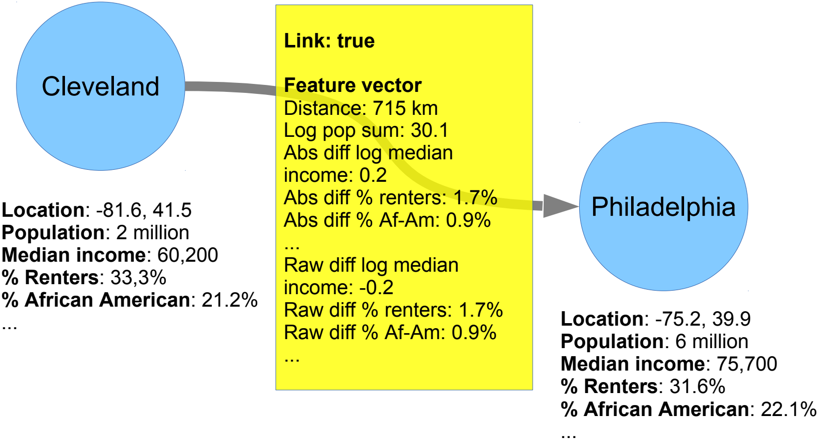
Possible roles for demographics

- **Assortativity**: similar cities evolve together.
- **Influence**: certain types of cities tend to lead, others follow.

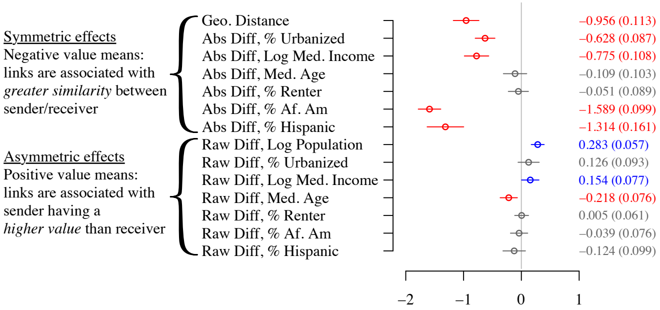
- 2010 US Census gives detailed demographics for each city.
- Are there types of demographic relationships that are especially frequent among linked cities?

Location: -81.6, 41.5
Population: 2 million
Median income: 60,200
% Renters: 33,3%
% African American: 21.2%
...

Location: -75.2, 39.9
Population: 6 million
Median income: 75,700
% Renters: 31.6%
% African American: 22.1%
...


Cleveland
Location: -81.6, 41.5
Population: 2 million
Median income: 60,200
% Renters: 33,3%
% African American: 21.2%
...
Link: true
Feature vector
Distance: 715 km
Log pop sum: 30.1
Abs diff log median income: 0.2
Abs diff % renters: 1.7%
Abs diff % Af-Am: 0.9%
...
Raw diff log median income: -0.2
Raw diff % renters: 1.7%
Raw diff % Af-Am: 0.9%
...
Philadelphia
Location: -75.2, 39.9
Population: 6 million
Median income: 75,700
% Renters: 31.6%
% African American: 22.1%
...

Regression coefficients

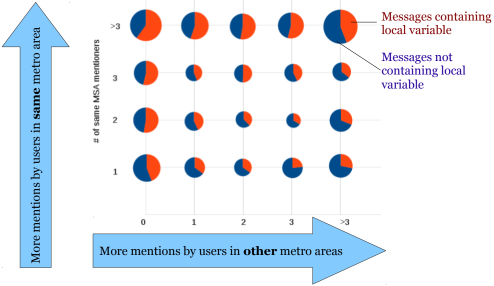


- Assortativity by race (of cities!) even more important than geography.
- Asymmetric effects are weaker, but bigger, younger metros tend to lead.

Why is geography so important?

Pavalanathan & Eisenstein (2015): People are more likely to use non-standard words when addressing audiences that are

1. small
2. local

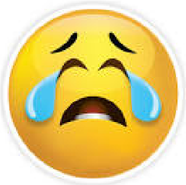


Some other results on language change

- Goel et al. (2016): neologisms are best transmitted through strong ties.
- Stewart et al. (2017): change is accelerated by censorship.
- Pavalanathan & Eisenstein (2016): emojis are driving down the frequency of non-standard words and spellings.

Some other results on language change

- Goel et al. (2016): neologisms are best transmitted through strong ties.
- Stewart et al. (2017): change is accelerated by censorship.
- Pavalanathan & Eisenstein (2016): emojis are driving down the frequency of non-standard words and spellings.



(but emojis raise a bunch of new interesting research questions ☺)

Summary and some next steps

- **Social context** plays a crucial role in disambiguating language, enabling robust text processing in domains like social media.

- Language data offers a new window on social phenomena such as **cultural affinity and influence**.

Summary and some next steps

- **Social context** plays a crucial role in disambiguating language, enabling robust text processing in domains like social media.

  What context will we need to build more effective language technology for other domains, such as electronic health records?

- Language data offers a new window on social phenomena such as **cultural affinity and influence**.

Summary and some next steps

- **Social context** plays a crucial role in disambiguating language, enabling robust text processing in domains like social media.

  What context will we need to build more effective language technology for other domains, such as electronic health records?

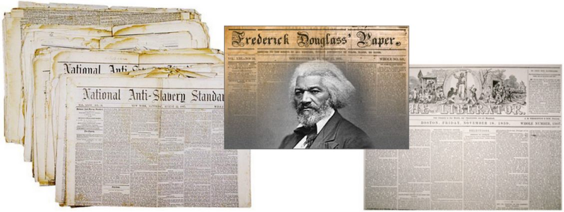
- Language data offers a new window on social phenomena such as **cultural affinity and influence**.

  How can we move computational social science beyond the bag-of-words?

Computational social science beyond bag of words

Proposal: model the transmission of **meaning** and **argumentation** in historical social networks

- Peldszus & Stede (2013) propose formal models of argumentation structure.
- We plan to automatically parse argumentation in historical anti-slavery newspapers.
- How did the argumentation around the abolition of slavery evolve in the period before the US civil war?



Acknowledgments

- **Students**: *Yangfeng Ji*, *Yi Yang*, Umashanthi Pavalanathan, Sandeep Soni, Ian Stewart, Yuval Pinter
- **Collaborators**: Stergos Afantenos, Ming-Wei Chang, Munmun De Choudhury, Eric Gilbert, Scott Kiesling, Lauren F. Klein, Dong Nguyen, Brendan O'Connor, Noah A. Smith, Manfred Stede, Eric P. Xing
- **Sponsors**: NSF, AFOSR, NIH, DTRA, NEH, Google

References I

Astudillo, R. F., Amir, S., Lin, W., Silva, M., & Trancoso, I. (2015). Learning word representations from scarce and noisy data with embedding sub-spaces. In *Proceedings of the Association for Computational Linguistics (ACL)*, Beijing.

Bhatia, P., Ji, Y., & Eisenstein, J. (2015). Better document-level sentiment analysis from rst discourse parsing. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Casasanto, L. S. (2008). Does social information influence sentence processing? In *Proceedings of the Cognitive Science Society*, volume 30.

Eisenstein, J. (2008). *Gesture in Automatic Discourse Processing*. PhD thesis, Massachusetts Institute of Technology.

Eisenstein, J. (2013). What to do about bad language on the internet. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, (pp. 359–369).

Eisenstein, J., Barzilay, R., & Davis, R. (2008a). Gestural cohesion for discourse segmentation. In *Proceedings of the Association for Computational Linguistics (ACL)*, Columbus, OH.

Eisenstein, J., Barzilay, R., & Davis, R. (2008b). Modeling gesture salience as a hidden variable for coreference resolution and keyframe extraction. *Journal of Artificial Intelligence Research*, *31*, 353–398.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010). A latent variable model for geographic lexical variation. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, (pp. 1277–1287).

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLoS ONE*, *9*.

Foster, J., Cetinoglu, O., Wagner, J., Le Roux, J., Nivre, J., Hogan, D., & van Genabith, J. (2011). From news to comment: Resources and benchmarks for parsing the language of web 2.0. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, (pp. 893–901)., Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., & Smith, N. A. (2011). Part-of-speech tagging for Twitter: annotation, features, and experiments. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 42–47)., Portland, OR.

References II

Goel, R., Soni, S., Goyal, N., Paparrizos, J., Wallach, H., Diaz, F., & Eisenstein, J. (2016). The social dynamics of language change in online networks. In *The International Conference on Social Informatics (SocInfo)*.

Ji, Y. & Eisenstein, J. (2015). One vector is not enough: Entity-augmented distributional semantics for discourse relations. *Transactions of the Association for Computational Linguistics (TACL)*.

Ji, Y., Haffari, G., & Eisenstein, J. (2016). A latent variable recurrent neural network for discourse relation language models. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, San Diego, CA.

Kehler, A. & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*(1-2), 1–37.

Nakov, P., Kozareva, Z., Ritter, A., Rosenthal, S., Stoyanov, V., & Wilson, T. (2013). Semeval-2013 task 2: Sentiment analysis in twitter. In *Proceedings of the 7th International Workshop on Semantic Evaluation, SemEval*.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology*, *18*(1), 62–85.

Pavalanathan, U. & Eisenstein, J. (2015). Audience-modulated variation in online social media. *American Speech*, *90*(2).

Pavalanathan, U. & Eisenstein, J. (2016). More emojis, less :) The competition for paralinguistic functions in microblog writing. *First Monday*, *22*(11).

Peldszus, A. & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *7*(1), 1–31.

Ritter, A., Clark, S., Mausam, & Etzioni, O. (2011). Named entity recognition in tweets: an experimental study. In *Proceedings of EMNLP*.

Rogers, E. M. (1962). *Diffusion of innovations*. New York: The Free Press.

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S. M., Ritter, A., & Stoyanov, V. (2015). Semeval-2015 task 10: Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*.

References III

Stewart, I., Chancellor, S., Chouhury, M. D., & Eisenstein, J. (2017). #anorexia, #anarexia, #anarexyia: Characterizing evolving online community practices with spelling variation. In *submitted to ICWSM*.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). Line: Large-scale information network embedding. In *Proceedings of the Conference on World-Wide Web (WWW)*, (pp. 1067–1077).

Yang, Y. & Chang, M.-W. (2015). S-mart: Novel tree-based structured learning algorithms applied to tweet entity linking. In *Proceedings of the Association for Computational Linguistics (ACL)*, Beijing.

Yang, Y., Chang, M.-W., & Eisenstein, J. (2016). Toward socially-infused information extraction: Embedding authors, mentions, and entities. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Yang, Y. & Eisenstein, J. (2017). Overcoming language variation in sentiment analysis with social attention. *Transactions of the Association for Computational Linguistics (TACL)*.