Variation and Change in Online Writing

Jacob Eisenstein
@jacobeisenstein

Georgia Institute of Technology

June 5, 2015

Social media in NAACL 2015

- ✓ *Soricut and Och* train skipgrams on Wikipedia.
- ✓ *Faruqui et al* test on IMDB movie reviews.
- ✗ *Krishnan and Eisenstein* analyze movie dialogues
- ✓ Tutorial on social media predictive analysis from *Volkova et al*.
- ✓ Keynote speech by *Lillian Lee* on message propagation in Twitter.

Social media in (E)ACL 2014

- ✗ *Lei et al* train and test on lots of newstext treebanks
- ✓ *Devlin et al* evaluate on Darpa BOLT Web Forums
- ✓ *Plank et al* focus on Twitter POS tagging
- ✓ *Olariu* summarizes microblogging streams

Social media in (E)ACL 2014

- ✗ *Lei et al* train and test on lots of newstext treebanks
- ✓ *Devlin et al* evaluate on Darpa BOLT Web Forums
- ✓ *Plank et al* focus on Twitter POS tagging
- ✓ *Olariu* summarizes microblogging streams

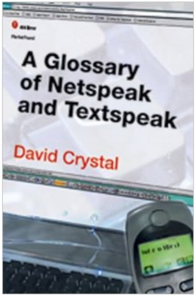
Social media won!
Now what?

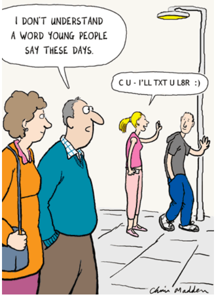
NLP tools versus social media

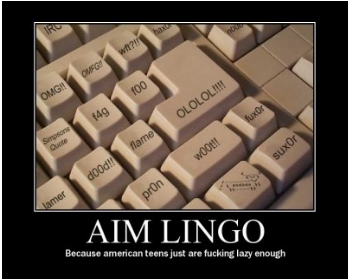
- Part-of-speech errors increase by 5x (Gimpel et al., 2011)
- Named entity recognition accuracy from 86% to 44% (Ritter et al., 2011)
- Syntactic parsing accuracy down by double-digits (Foster et al., 2011)

Why and what to do?

Some herald the birth of a new "netspeak" dialect (Thurlow, 2006).







If we build new treebanks for netspeak, will our problems be solved?

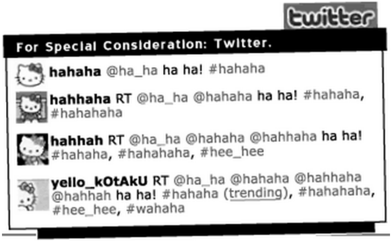
What's different in social media: who

"On the Internet, nobody knows you're a dog."

then a few authors, largely homogeneous

now millions of authors, highly diverse

What's different in social media: what



then constrained set of topics, focusing on "what's fit for print"

now unconstrained content, with emphasis on phatic communication

What's different in social media: when

then	asynchronous: write it today, read it tomorrow, few opportunities to respond
now	speech-like synchrony in written text

What's different in social media: how

then	professionalized writing process, subject to strong institutional regulation
now	diverse social contexts for writing, largely free of (traditional) institutional pressures

From netspeak to netspeak**s**: variation

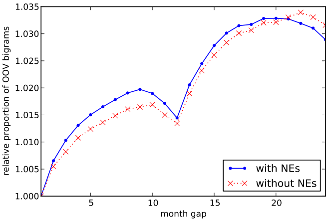
Social media is not a dialect, genre, or register.
Diversity is one of its most salient properties.

- `hubs blogged bloggers giveaway @klout`
- `kidd hubs xo =] xoxoxo muah xoxo darren`
- `(: :') xd (; /: <333 d: <33 </3 -___-`
- `nods softly sighs smiles finn laughs`
- `lmfaoo niggas ctfu lmfaooo wyd lmaoo`
- `gop dems senate unions conservative democrats`
- `/cc api ios ui portal developer e3 apple's`

(from Bamman et al., 2014)

From netspeak to netspeak**s**: change

- As social media takes on a speech-like role, new textual affordances are needed for paralinguistic information.
- Weaker language standards encourages experimentation and novelty.

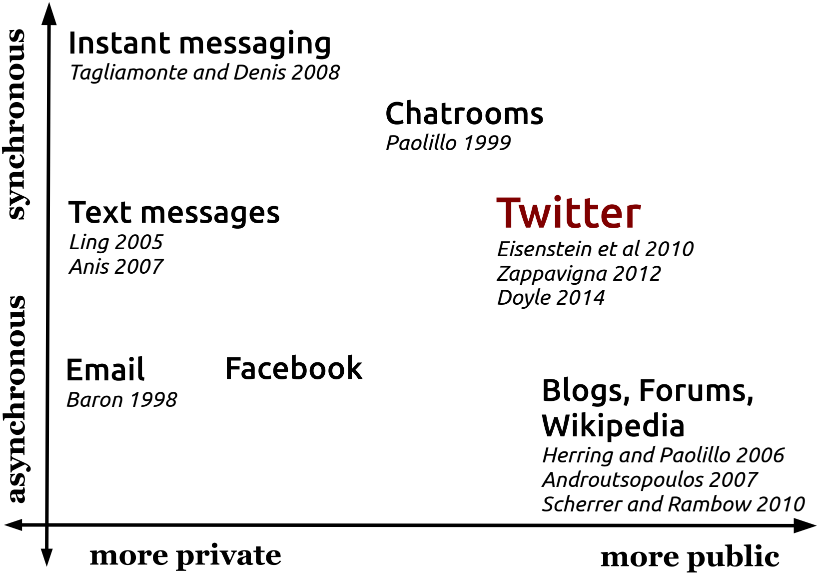


Out-**o**f-**v**ocabulary bigrams between pairs of 1M-word samples, divided by base rate (Eisenstein, 2013b).

Variation and change in social media

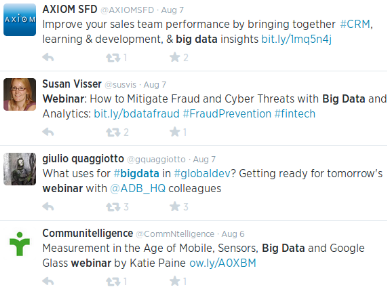
- Traditional annotation + learning approaches will not "solve" social media NLP.
- Building robust language technology for social media requires understanding variation and change.
- Sociolinguistics is dedicated to exactly these issues, but has mainly focused on small speech corpora. My goal is to apply sociolinguistic ideas to large-scale social media.

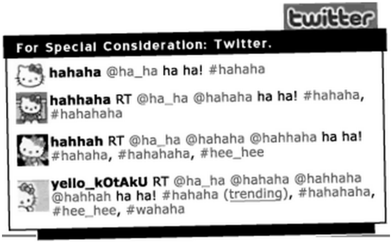
A landscape of digital communication



Twitter

- 140-character messages
- Each user has a custom *timeline* of people they've chosen to *follow*.
- Most data is publicly accessible, and social network and geographical metadata is available.





Who are these people?

	2013	2014
All internet users	*18%*	*23%**
Men	17	24*
Women	18	21
White, Non-Hispanic	16	21 *
Black, Non-Hispanic	29	27
Hispanic	16	25
18-29	31	37
30-49	19	25
50-64	9	12
65+	5	10*
High school grad or less	17	16
Some college	18	24
College+ (n=685)	18	30*
Less than $30,000/yr	17	20
$30,000-$49,999	18	21
$50,000-$74,999	15	27*
$75,000+	19	27*
Urban	18	25*
Suburban	19	23
Rural	11	17

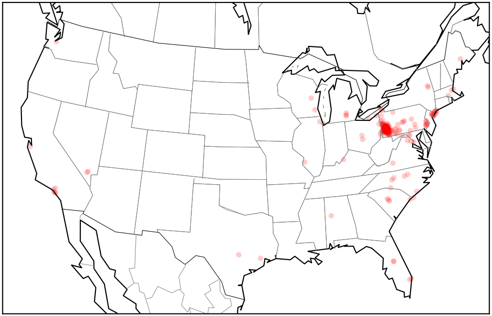
(Pew Research Center)

- % of online adults who use Twitter; per-message statistics will differ.
- Representativeness concerns are real, but there are potential solutions.
- Social media has important representativeness advantages too.

Table of Contents

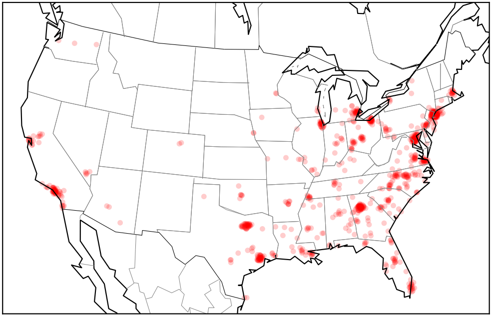
Yinz

- 2nd-person pronoun
- Western Pennsylvania
- Very rare: appears in 535 of 10^8 messages



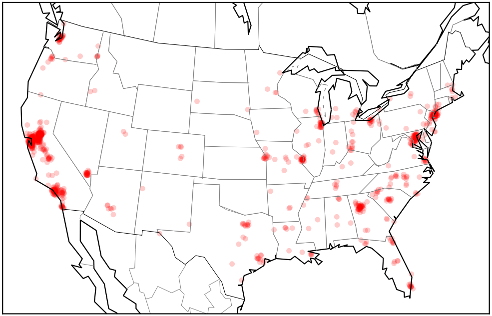
Yall

- 2nd-person pronoun
- Southeast, African-American English
- Once per 250 messages



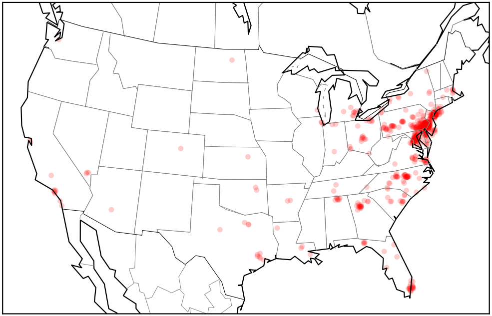
Hella

- Intensifier, e.g.
  i got hella nervous
- Northern California (Bucholtz et al., 2007)
- Once per 1000 messages



Jawn

- Noun, diffuse semantics
- Philadelphia, hiphop (Alim, 2009)
- Once per 1000 messages



  - @user ok u have heard this jawn right
  - i did wear that jawn but it was kinda warm this week

Summary of spoken dialect terms

	rate	*region*
yinz	$200,000$	mainly used in Western PA
yall	250	ubiquitous
hella	1000	ubiquitous, but more frequent in Northern California
jawn	1000	mainly used in Philadelphia

- Overall: mixed evidence for spoken language dialect variation in Twitter.
- But are these the right words?

Measuring regional specificity

Per region r,

- *Difference* in frequencies, $f_{i,r} - f_i$

  over-emphasizes frequent words

Measuring regional specificity

Per region r,

- *Difference* in frequencies, $f_{i,r} - f_i$

  over-emphasizes frequent words

- *Log-ratio* in frequencies, $\log f_{i,r} - \log f_i = \log \frac{f_{i,r}}{f_i}$

  over-emphasizes rare words

Measuring regional specificity

Per region r,

- *Difference* in frequencies, $f_{i,r} - f_i$

  over-emphasizes frequent words

- *Log-ratio* in frequencies, $\log f_{i,r} - \log f_i = \log \frac{f_{i,r}}{f_i}$

  over-emphasizes rare words

- *Regularized* log-frequency ratio, $\eta_{i,r} \approx \log f_{i,r} - \log f_i$, where $|\eta_{i,r}|$ is penalized.

$$\hat{\eta}_r = \arg\max_{\eta} \quad \log P(w|\eta; f) - \lambda|\eta|$$

λ controls the tradeoff between rare and frequent words

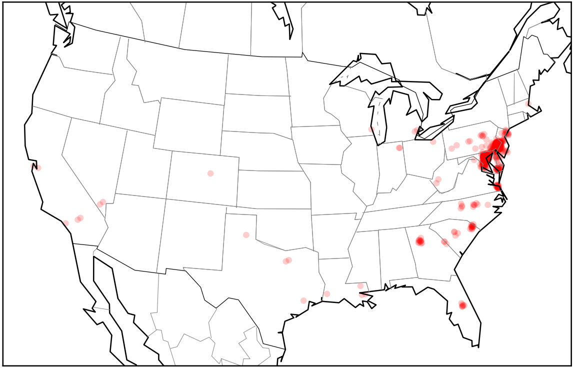
Discovered words

- **New York**: flatbush, baii, brib, bx, staten, *mta*, odee, soho, deadass, werd
- **Los Angeles**: pasadena, venice, anaheim, *dodger*, disneyland, angeles, compton, *ucla*, *dodgers*, melrose
- **Chicago**: #chicago, lbvs, chicago, *blackhawks*, *#bears*, *#bulls*, mfs, *cubs*, burbs, bogus
- **Philadelphia**: jawn, ard, *#phillies*, *sixers*, *phils*, *wawa*, *philadelphia*, delaware, *philly*, *phillies*

place names *entities* words

ard

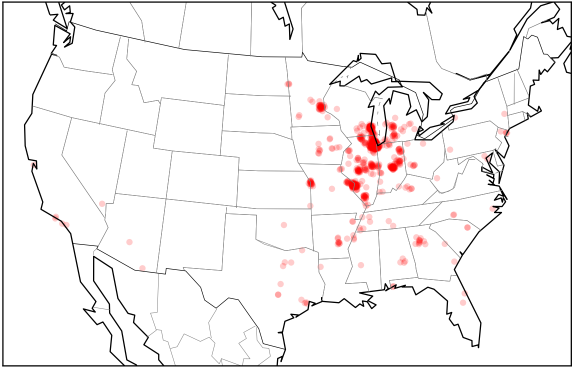
alternative spelling for alright



- @name ard let me kno
- lol u'll be ard

lbvs

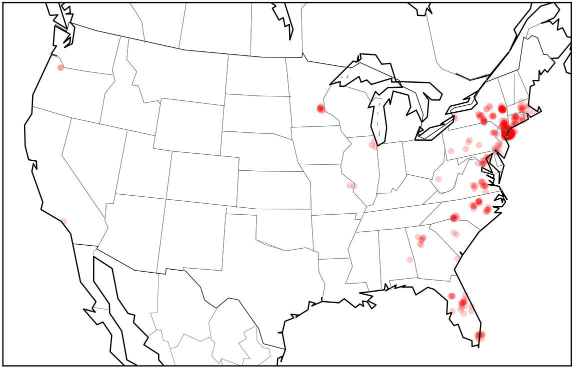
laughing but very serious



- i wanna rent a hotel room just to swim lbvs
- tell ur momma 2 buy me a car lbvs

odee

intensifier, related to overdose or overdone



- i'm odee sleepy
- she said she odee miss me
- its rainin odee :(

Table of Contents

Phonologically-motivated variables

- **-t,-d deletion** jus, ol
- **th-stopping** dis, doe
- **r-lessness** togetha, neva, lawd, yaself, shawty
- **vowels** tha (`the`), mayne (`man`), bruh, brah (`bro`)
- **relaxed pronunciations** prolly, aight
- **"allegro spellings" (Preston, 1985)** gonna, finna, fitna, bouta, tryna, iono

alternative spelling	rate	gloss	alt. freq
wanna	1,078	want to	0.642
tryna	4,073	trying to	0.444
wassup	8,336	what's up	0.499
bruh	11,423	bro	0.204
prolly	12,872	probably	0.271
doe	13,228	though	0.149
na	14,354	no	0.0263
betta	15,096	better	0.0720
holla	15,814	holler	0.918
neva	15,898	never	0.0628
aight	16,004	alright	0.373
ta	17,948	to	0.00351
bouta	21,301	about to	0.118
shawty	21,966	shorty	0.601
ion	26,196	i don't	0.0377

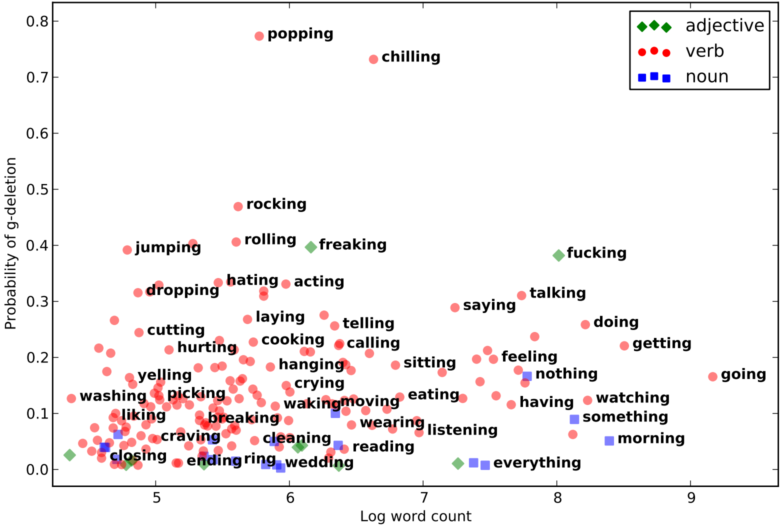
G-deletion

- In speech, "g" is deleted more often from verbs. *Does this syntactic conditioning transfer to writing?*

G-deletion

- In speech, “g” is deleted more often from verbs. *Does this syntactic conditioning transfer to writing?*
- Corpus: 120K tokens of top 200 unambiguous -ing words (ex. king, thing, sing)
- Part-of-speech tags from CMU Twitter tagger (Gimpel et al., 2011).

G-deletion: type-level analysis



(Colored by most common POS tag)

G-deletion: variable rules analysis

	Weight	Log odds	%	N
Verb	.556	.227	.200	89,173
Noun	.497	-.013	.083	18,756
Adjective	.447	-.213	.149	4,964
monosyllable	.071	-2.57	.001	108,804
Total			.178	112,893

G-deletion: variable rules analysis

	Weight	Log odds	%	N
Verb	.556	.227	.200	89,173
Noun	.497	-.013	.083	18,756
Adjective	.447	-.213	.149	4,964
monosyllable	.071	-2.57	.001	108,804
@-message	.534	.134	.205	36,974
Total			.178	112,893

G-deletion: variable rules analysis

	Weight	Log odds	%	N
Verb	.556	.227	.200	89,173
Noun	.497	-.013	.083	18,756
Adjective	.447	-.213	.149	4,964
monosyllable	.071	-2.57	.001	108,804
@-message	.534	.134	.205	36,974
High Euro-Am county	.452	-.194	.117	28,017
High Afro-Am county	.536	.145	.241	27,022
High pop density county	.514	.055	.228	27,773
Low pop density county	.496	-.017	.144	28,228
Total			.178	112,893

Two broad categories of variables

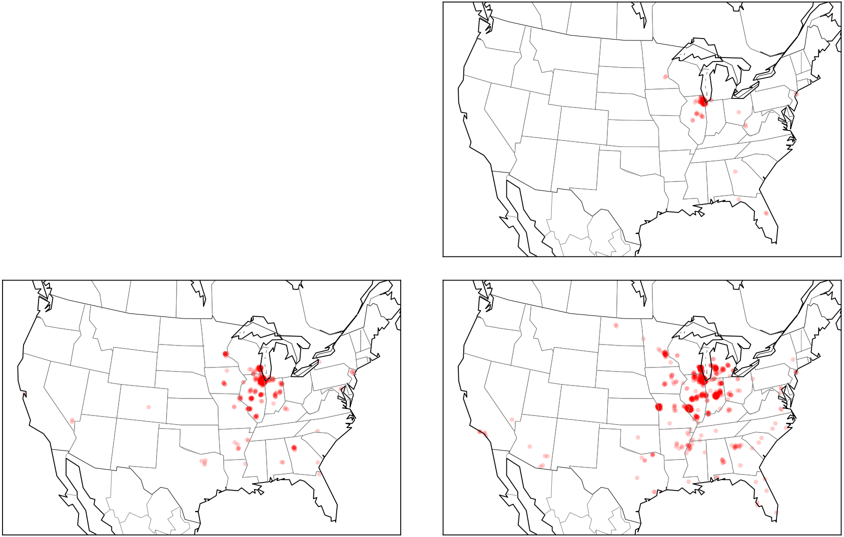
1. Imported from speech
   - Lexical variables (jawn, hella)
   - Phonologically-inspired variation (-g and -t,-d deletion)
   - These variables bring traces of their social and linguistic properties from speech.

2. Endogenous to digital writing
   - Abbreviations (lls, ctfu, asl, ...)
   - Emoticons (-_-)
   - Why should these vary with geography?
   - How stable is this form of variation?

Table of Contents

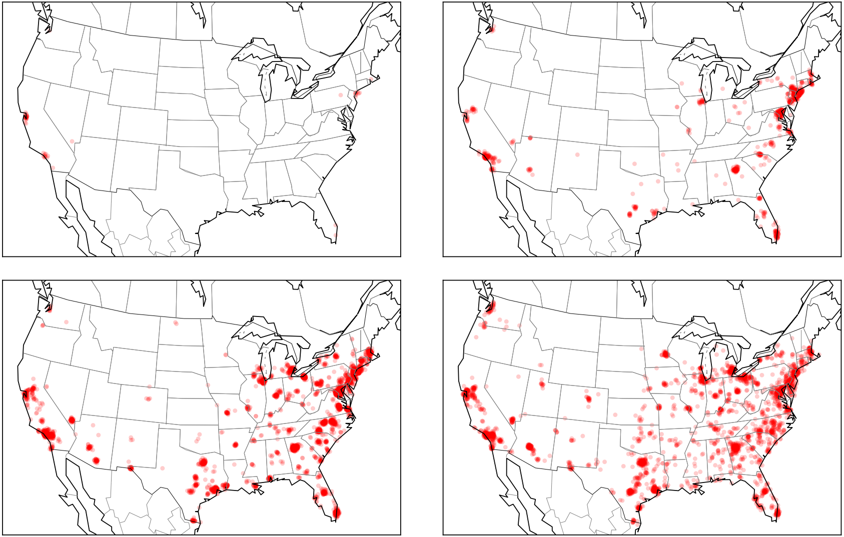
Change from 2010-2012: lbvs

tell ur momma 2 buy me a car lbvs



Change from 2009-2012: -_ _-

flight delayed -_ _- just what i need



Diffusion in social networks

Propagation of a cultural innovation requires:

1. Exposure
2. Decision to adopt it

Why is there geographical variation in netspeak?

Diffusion in social networks

Propagation of a cultural innovation requires:

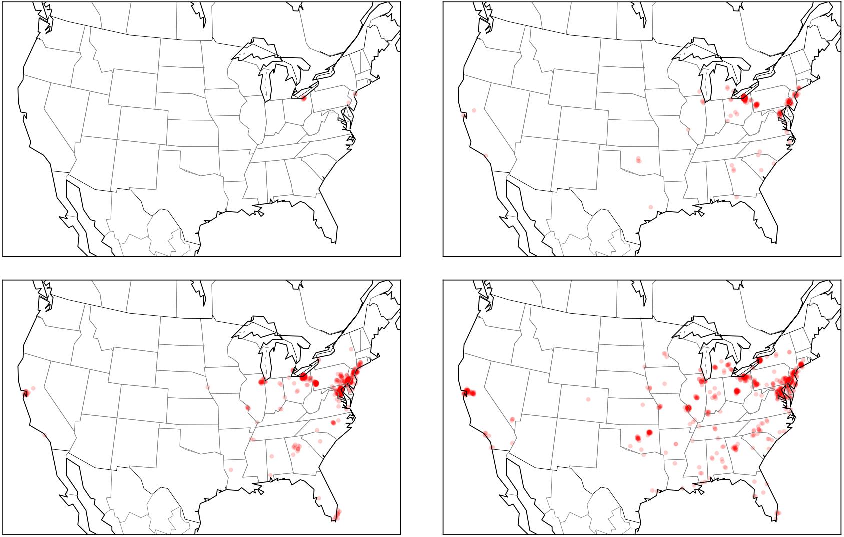
1. **Exposure**
2. Decision to adopt it

Why is there geographical variation in netspeak?

- 97% of "strong ties" (mutual @mentions) are between dyads in the same metro area.

Change from 2009-2012: ctfu

@name lmao! haahhaa ctfu!



The voyage of ctfu

2009	Cleveland
2010	Pittsburgh, Philadelphia
2011	Washington DC, Chicago, NY
2012	San Francisco, Columbus

The voyage of ctfu

2009	Cleveland
2010	Pittsburgh, Philadelphia
2011	Washington DC, Chicago, NY
2012	San Francisco, Columbus

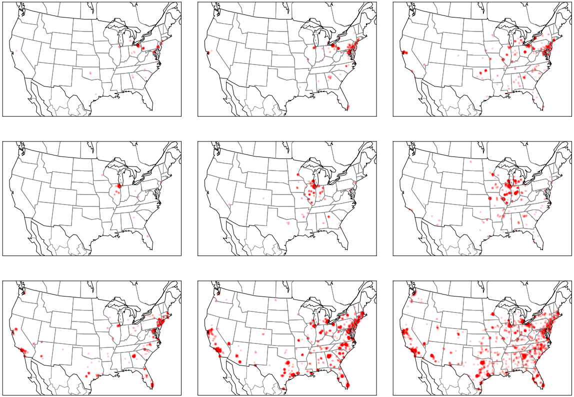
- This trajectory is hard to explain with models based only on geography or population.
- Is there a role for cultural influence? (Labov, 2011)

An aggregate model of lexical diffusion

ctfu

lbvs

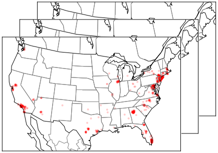
-_-



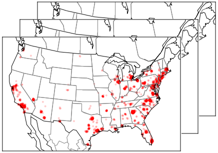
- Thousands of words have changing frequencies.
- Each spatiotemporal trajectory is idiosyncratic.
- What's the aggregate picture?

Language change as an autoregressive process

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



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



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

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

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

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

Inference

$$P(\text{words}; \text{influence}) \triangleq P(c; a)$$

$$= \sum_z P(c, z; a) = \sum_z \overbrace{P(c \mid z)}^{\text{emission}} \overbrace{P(z; a)}^{\text{transition}}$$

(z represents "activation")

Inference

$$
\begin{aligned}
P(\text{words; influence}) &\triangleq P(c; a) \\
&= \sum_z P(c, z; a) = \sum_z \overbrace{P(c \mid z)}^{\text{emission}} \overbrace{P(z; a)}^{\text{transition}} \\
&\qquad\qquad (z \text{ represents "activation"}) \\
&= \int P(c \mid z) P(z; a) dz \qquad\qquad (\text{uh oh...})
\end{aligned}
$$

Inference

$$
\begin{aligned}
P(\text{words}; \text{influence}) &\triangleq P(c; a) \\
&= \sum_z P(c, z; a) = \sum_z \overbrace{P(c \mid z)}^{\text{emission}} \overbrace{P(z; a)}^{\text{transition}}
\end{aligned}
$$

(z represents "activation")

$$
= \int P(c \mid z) P(z; a) dz
$$

(uh oh...)

$$
\begin{aligned}
&\rightarrow z^{(k)}, k \in \{1, 2, \ldots, K\} \\
&\approx \sum_k P(c \mid z^{(k)}) P(z^{(k)}; a)
\end{aligned}
$$

(Monte Carlo approximation to the rescue!)

Inference

$$P(\text{words}; \text{influence}) \triangleq P(c; a)$$

$$= \sum_z P(c, z; a) = \sum_z \overbrace{P(c \mid z)}^{\text{emission}} \overbrace{P(z; a)}^{\text{transition}}$$

(z represents "activation")

$$= \int P(c \mid z) P(z; a) dz$$

(uh oh...)

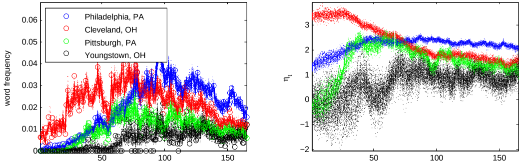
$$\rightarrow z^{(k)}, k \in \{1, 2, \ldots, K\}$$

$$\approx \sum_k P(c \mid z^{(k)}) P(z^{(k)}; a)$$

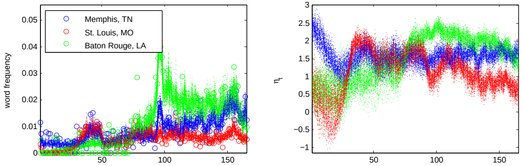
(Monte Carlo approximation to the rescue!)

$$\hat{a} = \arg\max_a \sum_k P(c \mid z^{(k)}) P(z^{(k)}; a)$$

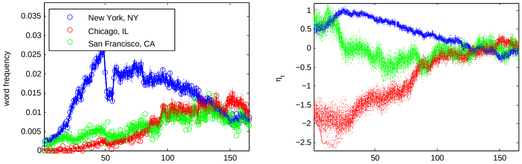
ctfu


Philadelphia, PA
Cleveland, OH
Pittsburgh, PA
Youngstown, OH
word frequency
η_t

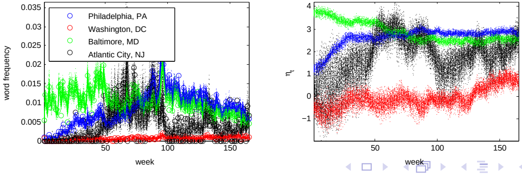
ion


Memphis, TN
St. Louis, MO
Baton Rouge, LA
word frequency
η_t

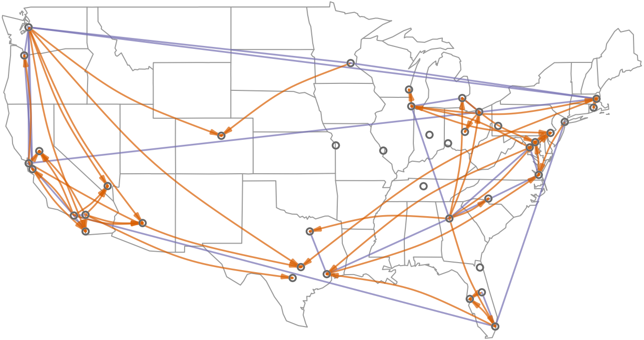
-_-


New York, NY
Chicago, IL
San Francisco, CA
word frequency
η_t

ard


Philadelphia, PA
Washington, DC
Baltimore, MD
Atlantic City, NJ
word frequency
η_t
week
week

Aggregating region-to-region influence



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

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.

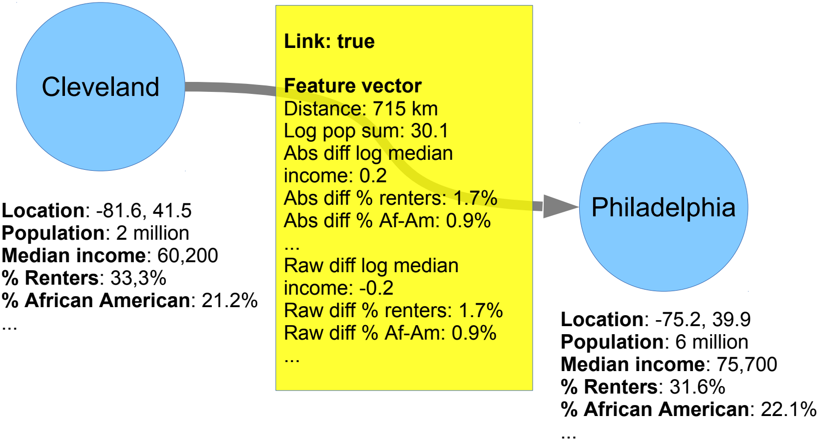
- 2010 US Census gives detailed demographics for each city.
- Are there types of demographic relationships that are especially frequent among linked cities?

Logistic regression

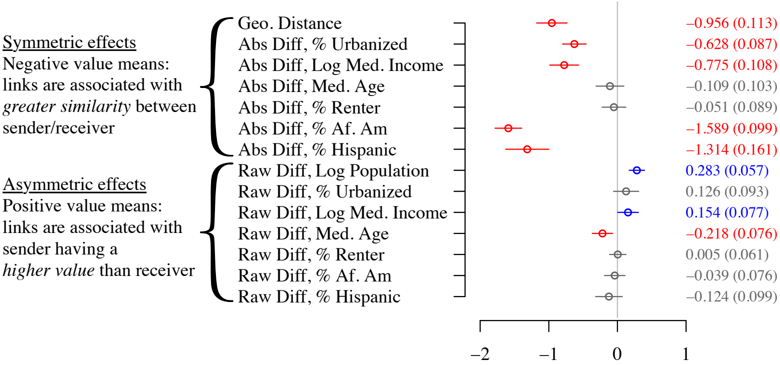
Location: -81.6, 41.5
Population: 2 million
Median income: 60,200
% Renters: 33,3%
% African American: 21.2%
...

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

Logistic regression



Regression coefficients



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

Diffusion in social networks

Propagation of a cultural innovation requires:

1. **Exposure**
2. Decision to adopt it

Why is there geographical variation in netspeak?

- 97% of "strong ties" (mutual @mentions) are between dyads in the same metro area.

Diffusion in social networks

Propagation of a cultural innovation requires:

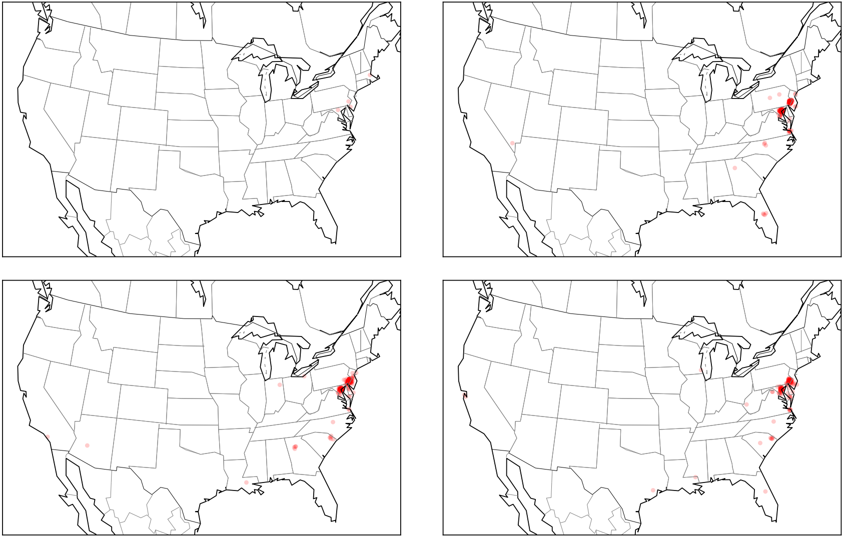
1. Exposure
2. **Decision** to adopt it

Why is there geographical variation in netspeak?

- 97% of "strong ties" (mutual @mentions) are between dyads in the same metro area.
- Diffusion depends on sociocultural affinity and influence, not just geography and population.

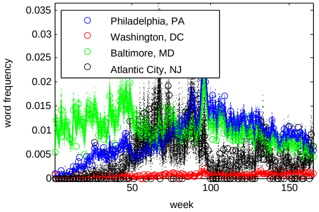
One more example: ard

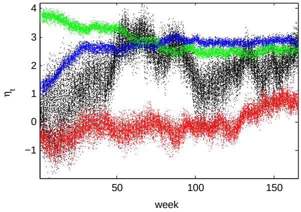
lol u'll be ard



Stable variation

ard





- In three years, ard never gets from Baltimore to DC! (It gets to Philadelphia within a year.)
- The connection to spoken variation is tenuous.
- So what explains this stability?

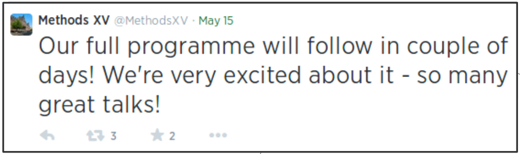
Table of Contents

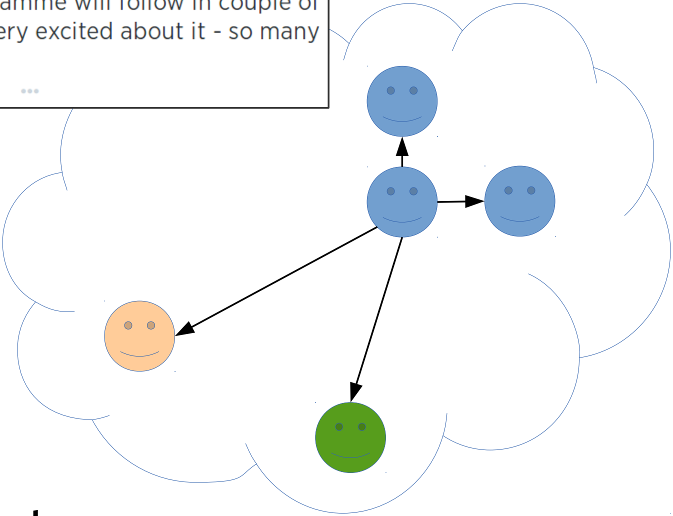
From macro to micro

Macro-level variation and change must ground out in individual linguistic decisions.

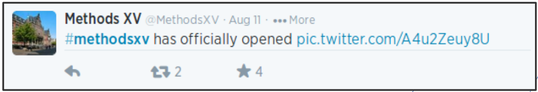
- With social media data, we can distinguish the *contexts* in which feature counts appear.
- One way to define context is by the *intended audience*.
- Variables that are used for smaller, more local audiences may be more persistent.

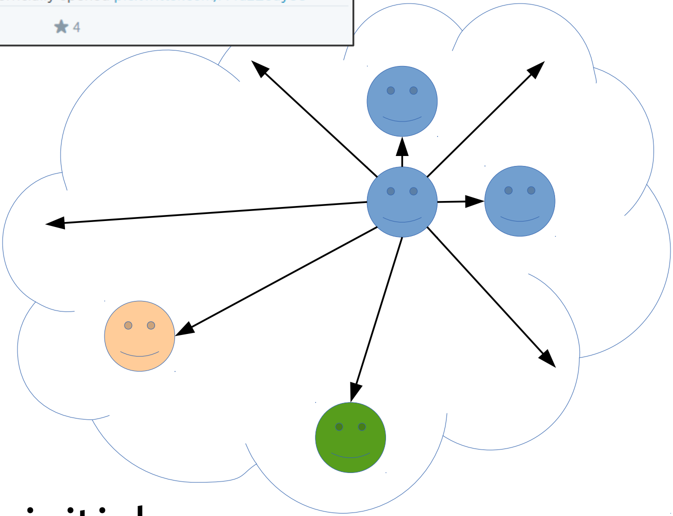
(Pavalanathan & Eisenstein, 2015)


Methods XV @MethodsXV · May 15
Our full programme will follow in couple of days! We're very excited about it - so many great talks!
3
2

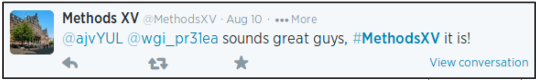


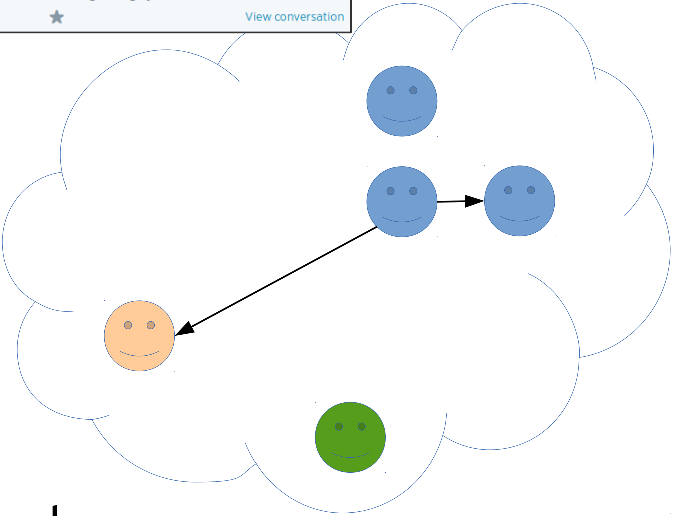
Broadcast





Hashtag-initial


Methods XV @MethodsXV · Aug 10 · ••• More
@ajvYUL @wgi_pr31ea sounds great guys, #MethodsXV it is!
View conversation

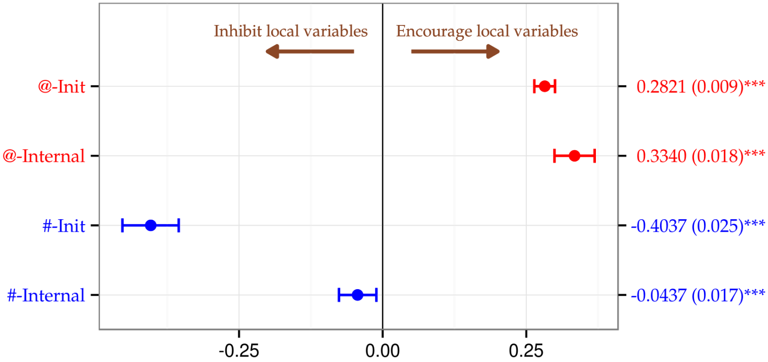


Addressed

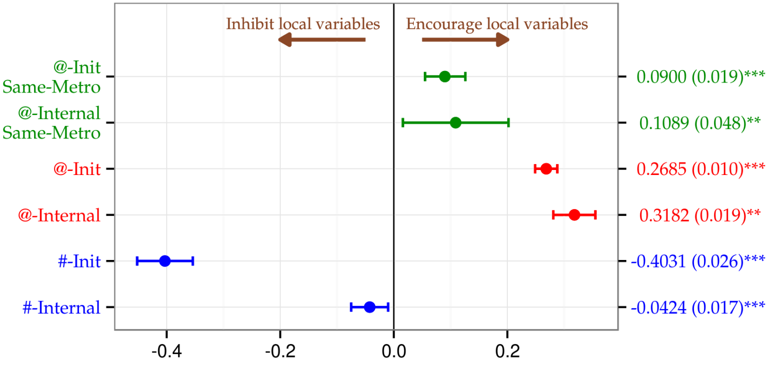
Logistic regression

- *Dependent variable*: does the tweet contain a local word (e.g., lbvs, hella, jawn)
- *Predictors*
  - **Message type**: broadcast, addressed, #-initial
  - **Controls**: message length, author statistics

Small audience → less standard language



Local audience → less standard language



Diffusion in social networks

Propagation of a cultural innovation requires:

1. Exposure
2. **Decision** to adopt it

Why is there geographical variation in netspeak?

- 97% of “strong ties” (mutual @mentions) are between dyads in the same metro area.
- Diffusion depends on sociocultural affinity and influence, not just geography and population.

Diffusion in social networks

Propagation of a cultural innovation requires:

1. **Exposure**
2. Decision to adopt it

Why is there geographical variation in netspeak?

- 97% of "strong ties" (mutual @mentions) are between dyads in the same metro area.
- Diffusion depends on sociocultural affinity and influence, not just geography and population.
- Non-standard features are more likely to be transmitted along strong, local ties.

Summary

- Social media is transforming written language!
- Social media writing is *variable* and *dynamic*, but not noisy: there is always an underlying sociolinguistic structure.
- Recovering this structure promises new insights for both linguistics and language technology.
- Next steps:
  - modeling individual linguistic decisions
  - applying these results to build more robust language technology

Thanks!

To my collaborators:

- David Bamman (CMU)
- Fernando Diaz (MSR)
- Naman Goyal (Georgia Tech)
- Brendan O'Connor (UMass)
- Ioannis Paparrizos (Columbia)
- Umashanthi Pavalanathan (Georgia Tech)
- Tyler Schnoebelen (Stanford and Idibon)
- Noah A. Smith (University of Washington)
- Hanna Wallach (MSR and UMass)
- Eric P. Xing (CMU)

And to the National Science Foundation.

Alim, H. S. (2009). Hip hop nation language. In A. Duranti (Ed.), *Linguistic Anthropology: A Reader* (pp. 272–289). Malden, MA: Wiley-Blackwell.

Anis, J. (2007). Neography: Unconventional spelling in French SMS text messages. In B. Danet & S. C. Herring (Eds.), *The Multilingual Internet: Language, Culture, and Communication Online* (pp. 87–115). Oxford University Press.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160.

Bucholtz, M., Bermudez, N., Fung, V., Edwards, L., & Vargas, R. (2007). Hella nor cal or totally so cal? the perceptual dialectology of california. *Journal of English Linguistics*, *35*(4), 325–352.

Doyle, G. (2014). Mapping dialectal variation by querying social media. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, (pp. 98–106)., Stroudsburg, Pennsylvania. Association for Computational Linguistics.

Eisenstein, J. (2013a). Phonological factors in social media writing. In *Proceedings of the Workshop on Language Analysis in Social Media*, (pp. 11–19)., Atlanta.

Eisenstein, J. (2013b). 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)., Stroudsburg, Pennsylvania. Association for Computational Linguistics.

Eisenstein, J. (2015a). Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, *19*, 161–188.

Eisenstein, J. (2015b). Written dialect variation in online social media. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *Handbook of Dialectology*. Wiley.

Eisenstein, J., Ahmed, A., & Xing, E. P. (2011). Sparse additive generative models of text. In *Proceedings of the International Conference on Machine Learning (ICML)*, (pp. 1041–1048)., Seattle, WA.

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)., Stroudsburg, Pennsylvania. Association for Computational Linguistics.

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

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.

Herring, S. C. & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, *10*(4), 439–459.

Huberman, B., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. *First Monday*, *14*(1).

Labov, W. (2011). *Principles of Linguistic Change*, volume 3: Cognitive and Cultural Factors. Wiley-Blackwell.

Paolillo, J. C. (1999). The virtual speech community: Social network and language variation on irc. *J. Computer-Mediated Communication*, *4*(4), 0.

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

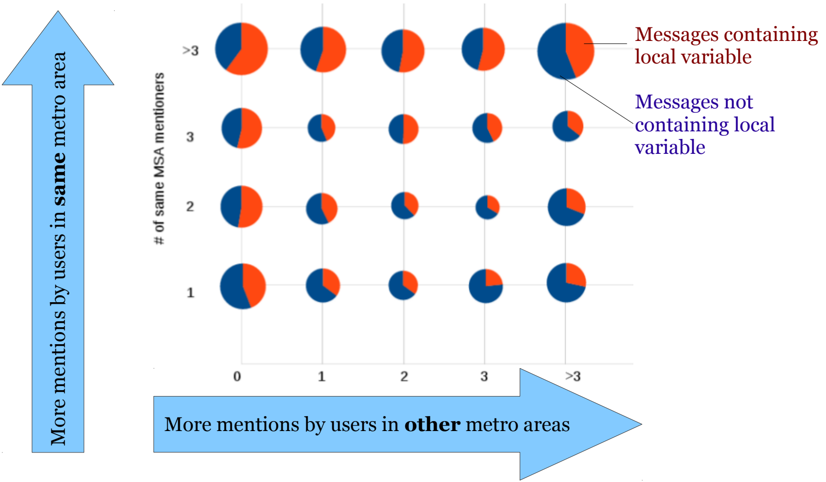
Preston, D. R. (1985). The Li'l Abner syndrome: Written Representations of Speech. *American Speech*, *60*(4), 328–336.

Tagliamonte, S. A. & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech*, *83*(1), 3–34.

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of twitter networks. *Social networks*, *34*(1), 73–81.

Thurlow, C. (2006). From statistical panic to moral panic: The metadiscursive construction and popular exaggeration of new media language in the print media. *Journal of Computer-Mediated Communication*, 667–701.

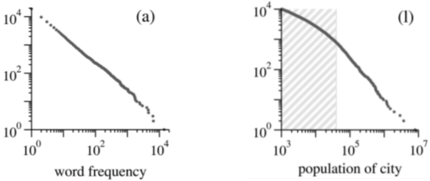
Local audience → less standard language



Why raw word counts won't work

We observe counts $c_{w,r,t}$ for word w in region r at time t. How does $c_{w,r,t}$ influence $c_{w,r',t+1}$?

- Both word counts and city sizes follow power law distributions, with lots of zero counts.



- Exogenous events such as pop culture and weather introduce global temporal effects.
- Twitter's sampling rate is inconsistent, both spatially and temporally.

Latent activation model

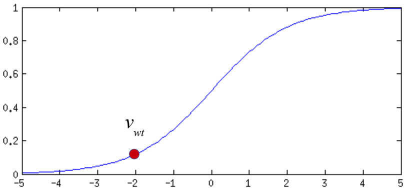
$$c_{w,r,t} \sim \mathrm{Binomial}(\beta_{w,r,t}, s_{r,t})$$

Latent activation model

$$
\begin{aligned}
c_{w,r,t} \sim & \mathrm{Binomial}(\beta_{w,r,t}, s_{r,t}) \\
\beta_{w,r,t} = & \mathrm{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t})
\end{aligned}
$$

Latent activation model

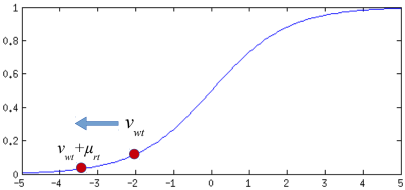
$$c_{w,r,t} \sim \text{Binomial}(\beta_{w,r,t}, s_{r,t})$$
$$\beta_{w,r,t} = \text{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t})$$



- Base word log-probability

Latent activation model

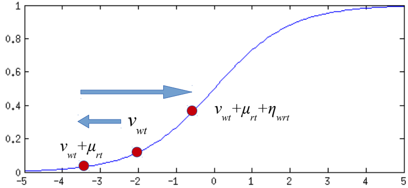
$$c_{w,r,t} \sim \text{Binomial}(\beta_{w,r,t}, s_{r,t})$$
$$\beta_{w,r,t} = \text{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t})$$



- Base word log-probability
- City-specific "verbosity"

Latent activation model

$$c_{w,r,t} \sim \text{Binomial}(\beta_{w,r,t}, s_{r,t})$$
$$\beta_{w,r,t} = \text{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t})$$



- Base word log-probability
- City-specific "verbosity"
- *Spatio-temporal activation*

Dynamics model

$$
\begin{aligned}
c_{w,r,t} \sim & \mathrm{Binomial}(\beta_{w,r,t}, s_{r,t}) \\
\beta_{w,r,t} = & \mathrm{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t}) \\
\eta_{w,r,t} \sim & \mathrm{Normal}(\sum_{r'} a_{r' \to r} \eta_{w,r',t-1}, \gamma_{w,r})
\end{aligned}
$$

- $a_{i \to j}$ captures the linguistic "influence" of city i on city j.
- If $\eta_{j,t+1} = \eta_{i,t}$, then $a_{i \to j} = 1$, and $a_{i \to j} = 0$.
- If η_j and η_i co-evolve smoothly, then $a_{i,j} > 0$ and $a_{j,i} > 0$.