Variation and Change in Online Writing

Jacob Eisenstein
@jacobbeisenstein

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

then a few authors, largely homogeneous
now millions of authors, highly diverse
What’s different in social media: what

then a 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 netspeaks: 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)
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-of-vocabulary 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

- Instant messaging
  - Tagliamonte and Denis 2008
- Chatrooms
  - Paolillo 1999
- Text messages
  - Ling 2005
  - Anis 2007
- Twitter
  - Eisenstein et al 2010
  - Zappavigna 2012
  - Doyle 2014
- Facebook
- Bloggs, forums, Wikipedia
  - Herring and Paolillo 2006
  - Androutsopoulos 2007
  - Scherrer and Rambow 2010
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?

- % 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.

(Pew Research Center)
Table of Contents

Lexical variation

Orthographic variation

Language change as sociocultural influence

Language change in social networks
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

<table>
<thead>
<tr>
<th>Term</th>
<th>Rate</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>yinz</td>
<td>200,000</td>
<td>mainly used in Western PA</td>
</tr>
<tr>
<td>yall</td>
<td>250</td>
<td>ubiquitous</td>
</tr>
<tr>
<td>hella</td>
<td>1000</td>
<td>ubiquitous, but more frequent in Northern California</td>
</tr>
<tr>
<td>jawn</td>
<td>1000</td>
<td>mainly used in Philadelphia</td>
</tr>
</tbody>
</table>

- 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.

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

  \( \lambda \) 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, philis, wawa, philadelphia, delaware, philly, phillies
ard

alternative spelling for alright

- @name ard let me kno
- lol u’ll be ard
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 :(  

{i'm odee sleepy
she said she odee miss me
its rainin odee :(
Table of Contents

Lexical variation

Orthographic variation

Language change as sociocultural influence

Language change in social networks
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  proly, aight
“allegro spellings” (Preston, 1985)  gonna, finna,
fitna, bouta, tryna, iono
<table>
<thead>
<tr>
<th>alternative spelling</th>
<th>rate</th>
<th>gloss</th>
<th>alt. freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>wanna</td>
<td>1,078</td>
<td>want to</td>
<td>0.642</td>
</tr>
<tr>
<td>tryna</td>
<td>4,073</td>
<td>trying to</td>
<td>0.444</td>
</tr>
<tr>
<td>wassup</td>
<td>8,336</td>
<td>what’s up</td>
<td>0.499</td>
</tr>
<tr>
<td>bruh</td>
<td>11,423</td>
<td>bro</td>
<td>0.204</td>
</tr>
<tr>
<td>proly</td>
<td>12,872</td>
<td>probably</td>
<td>0.271</td>
</tr>
<tr>
<td>doe</td>
<td>13,228</td>
<td>though</td>
<td>0.149</td>
</tr>
<tr>
<td>na</td>
<td>14,354</td>
<td>no</td>
<td>0.0263</td>
</tr>
<tr>
<td>betta</td>
<td>15,096</td>
<td>better</td>
<td>0.0720</td>
</tr>
<tr>
<td>holla</td>
<td>15,814</td>
<td>holler</td>
<td>0.918</td>
</tr>
<tr>
<td>neva</td>
<td>15,898</td>
<td>never</td>
<td>0.0628</td>
</tr>
<tr>
<td>aight</td>
<td>16,004</td>
<td>alright</td>
<td>0.373</td>
</tr>
<tr>
<td>ta</td>
<td>17,948</td>
<td>to</td>
<td>0.00351</td>
</tr>
<tr>
<td>bouta</td>
<td>21,301</td>
<td>about to</td>
<td>0.118</td>
</tr>
<tr>
<td>shawty</td>
<td>21,966</td>
<td>shorty</td>
<td>0.601</td>
</tr>
<tr>
<td>ion</td>
<td>26,196</td>
<td>i don’t</td>
<td>0.0377</td>
</tr>
</tbody>
</table>
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

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

@-message

High Euro-Am county

High Afro-Am county

High pop density county

Low pop density county

Total
G-deletion: variable rules analysis

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Log odds</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>.556</td>
<td>.227</td>
<td>.200</td>
<td>89,173</td>
</tr>
<tr>
<td>Noun</td>
<td>.497</td>
<td>-.013</td>
<td>.083</td>
<td>18,756</td>
</tr>
<tr>
<td>Adjective</td>
<td>.447</td>
<td>-.213</td>
<td>.149</td>
<td>4,964</td>
</tr>
<tr>
<td>monosyllable</td>
<td>.071</td>
<td>-2.57</td>
<td>.001</td>
<td>108,804</td>
</tr>
<tr>
<td>@-message</td>
<td>.534</td>
<td>.134</td>
<td>.205</td>
<td>36,974</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>.178</strong></td>
<td></td>
<td></td>
<td><strong>112,893</strong></td>
</tr>
</tbody>
</table>
G-deletion: variable rules analysis

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

Lexical variation

Orthographic variation

Language change as sociocultural influence

Language change in social networks
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

<table>
<thead>
<tr>
<th>Year</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>Cleveland</td>
</tr>
<tr>
<td>2010</td>
<td>Pittsburgh, Philadelphia</td>
</tr>
<tr>
<td>2011</td>
<td>Washington DC, Chicago, NY</td>
</tr>
<tr>
<td>2012</td>
<td>San Francisco, Columbus</td>
</tr>
</tbody>
</table>

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

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.

\[ \eta_2 \sim N(A\eta_1, \Sigma) \]

\[ c_{ctfu,1} \sim \text{Binomial}(f(\eta_{ctfu,1}), N_1) \]
\[ c_{hella,1} \sim \text{Binomial}(f(\eta_{hella,1}), N_1) \]
\[ \ldots \]

\[ c_{ctfu,2} \sim \text{Binomial}(f(\eta_{ctfu,2}), N_2) \]
\[ c_{hella,2} \sim \text{Binomial}(f(\eta_{hella,2}), N_2) \]
\[ \ldots \]

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

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

\[ = \sum_z P(c, z; a) = \sum_z \overbrace{P(c | z)}^{\text{emission}} \overbrace{P(z; a)}^{\text{transition}} \]

\( z \) represents “activation”
Inference

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

\[ = \sum_z P(c, z; a) = \sum_z P(c \mid z) P(z; a) \]

(z represents “activation”)

\[ = \int P(c \mid z) P(z; a) dz \quad \text{ (uh oh...)} \]
Inference

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

\[ = \sum_z P(c, z; a) = \sum_z P(c \mid z) P(z; a) \]

\( (z \text{ represents “activation”}) \)

\[ = \int P(c \mid z) P(z; a) \, dz \quad \text{(uh oh...)} \]

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

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

\( (\text{Monte Carlo approximation to the rescue!}) \)
Inference

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

\[ = \sum_z P(c, z; a) = \sum_z \underbrace{P(c \mid z)}_{\text{emission}} \underbrace{P(z; a)}_{\text{transition}} \]

\[ (z \text{ represents "activation"}) \]

\[ = \int P(c \mid z) P(z; a) \, dz \quad \text{(uh oh...)} \]

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

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

\[ (\text{Monte Carlo approximation to the rescue!}) \]

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

Cleveland
Location: -81.6, 41.5
Population: 2 million
Median income: 60,200
% Renters: 33.3%
% African American: 21.2%
...

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

Cleveland

Location: -81.6, 41.5
Population: 2 million
Median income: 60,200
% Renters: 33.3%
% African American: 21.2%
...

Philadelphia

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

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%
...
Regression coefficients

Symmetric effects
Negatives value means: links are associated with greater similarity between sender/receiver

Asymmetric effects
Positive value means: links are associated with sender having a higher value than receiver

-0.956 (0.113)
-0.628 (0.087)
-0.775 (0.108)
-0.109 (0.103)
-0.051 (0.089)
-1.589 (0.099)
-1.314 (0.161)
0.283 (0.057)
0.126 (0.093)
0.154 (0.077)
-0.218 (0.076)
0.005 (0.061)
-0.039 (0.076)
-0.124 (0.099)

- ▶ 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

- 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

Lexical variation

Orthographic variation

Language change as sociocultural influence

Language change in social networks
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)
Our full programme will follow in couple of days! We’re very excited about it - so many great talks!
#methodsxv has officially opened pic.twitter.com/A4u2Zeuy8U
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.
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.


Local audience → less standard language

More mentions by users in same metro area

More mentions by users in other metro areas

Messages containing local variable

Messages not containing local variable
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

\[ c_{w,r,t} \sim \text{Binomial}(\beta_{w,r,t}, s_{r,t}) \]
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})
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

\[ c_{w,r,t} \sim \text{Binomial}(\beta_{w,r,t}, s_{r,t}) \]
\[ \beta_{w,r,t} = \text{Logistic}(\nu_{w,t} + \mu_{r,t} + \eta_{w,r,t}) \]
\[ \eta_{w,r,t} \sim \text{Normal}(\sum_{r'} a_{r' \rightarrow r} \eta_{w,r',t-1}, \gamma_{w,r}) \]

- \( a_{i \rightarrow j} \) captures the linguistic “influence” of city \( i \) on city \( j \).
- If \( \eta_{j,t+1} = \eta_{i,t} \), then \( a_{i \rightarrow j} = 1 \), and \( a_{i \rightarrow j} = 0 \).
- If \( \eta_j \) and \( \eta_i \) co-evolve smoothly, then \( a_{i,j} > 0 \) and \( a_{j,i} > 0 \).