Social Meanings in Social Networks

Jacob Eisenstein
@jacobeisenstein

Georgia Institute of Technology

December 8, 2016

Language is socially situated (weak)

What we **say** depends on

- who we are
- who we are talking to.

Example 1: Shared context

- Me to nerd friends: I'm giving a talk at the LTI.

Example 1: Shared context

- Me to nerd friends: I'm giving a talk at the LTI.
- Me to normal people: I'm giving a research presentation at Carnegie Mellon.

Example 2: Formality

- Me to my brother: Okay, I'll do this for you, but you owe me.

Example 2: Formality

- Me to my brother: Okay, I'll do this for you, but you owe me.
- Me to my chair: Dear Annie, I've had some time to consider your request, and I would be happy to assist the department in this important matter.

Language is socially situated (strong)

What we **mean** depends on

- who we are
- who we are talking to.

Example 3: Semantic variation

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

Example 3: Semantic 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."

Example 4: Forms of Address

- **Jacob**: I'm going for a bike ride.
- **Jacob's dad**: Have fun.

- **Jacob**: Thanks, Dad.

Example 4: Forms of Address

- **Jacob**: I'm going for a bike ride.
- **Jacob's friend Jeff**: Don't forget your helmet.
- **Jacob**: Thanks, Dad.

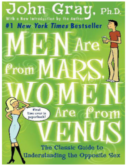
Lessons

- Understanding social meaning is part of natural language understanding (e.g., thanks dad).
- Even propositional content depends on social factors (e.g., this sick beat) and context (e.g., a talk at LTI).
- Language analysis can reveal latent properties of social relationships (e.g., Dear Annie . . .).

Social variables, from macro to micro

Early **computational sociolinguistics** (Nguyen et al., 2016) modeled relationship between linguistic features and broad social categories, e.g.

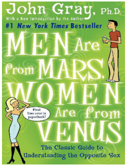
- age (Nguyen et al., 2013);
- gender (Argamon et al., 2007);
- geography (Eisenstein et al., 2010).



Social variables, from macro to micro

Early **computational sociolinguistics** (Nguyen et al., 2016) modeled relationship between linguistic features and broad social categories, e.g.

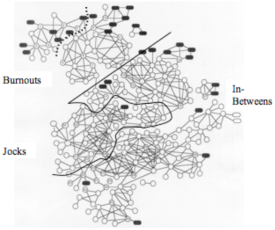
- age (Nguyen et al., 2013);
- gender (Argamon et al., 2007);
- geography (Eisenstein et al., 2010).



But many interesting social distinctions are not discrete, not macro-scale constructs, not available from metadata.

Social networks as social variables

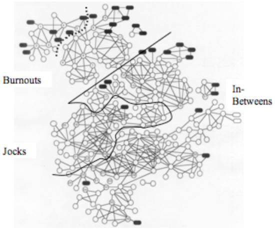
- Social networks are structured representations of interactions between individuals.
- Sociology and network science offer a rich array of theoretical models and analytic tools.



(Eckert, 2000)

Social networks as social variables

- Social networks are structured representations of interactions between individuals.
- Sociology and network science offer a rich array of theoretical models and analytic tools.



(Eckert, 2000)

This talk: machine learning models linking social network theory and linguistic analysis.

Overview

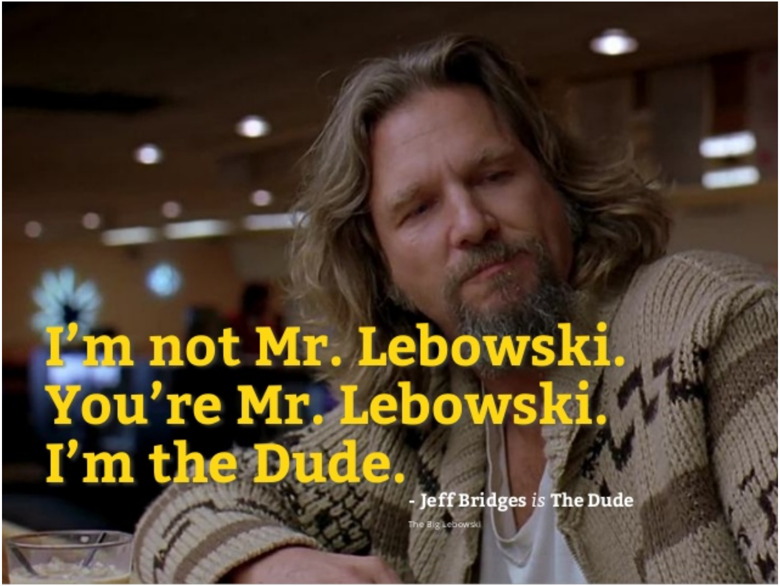
Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	homophily	deep learning
Language change	diffusion of innovations	Hawkes process

Overview

Language	*Social*	*ML*
Formality of address	**structural balance**	**latent variables**
Entity linking and sentiment	homophily	deep learning
Language change	diffusion of innovations	Hawkes process

Linguistic Signals of Social Relationships

You're Mr. Lebowski, I'm the Dude: Inducing Address Term Formality in Signed Social Networks
Krishnan & Eisenstein (2015).


I'm not Mr. Lebowski.
You're Mr. Lebowski.
I'm the Dude.
- Jeff Bridges *is* The Dude
The Big Lebowski

Modeling formality of address

- What is the nature of the relationships between actors in a social network?
- Are there regular structures that emerge across signed networks?
- How does language reflect and reproduce social relationships?

Modeling formality of address

- What is the nature of the relationships between actors in a social network?

  We induce a clustering of edges by formality.

- Are there regular structures that emerge across signed networks?

- How does language reflect and reproduce social relationships?

Modeling formality of address

- What is the nature of the relationships between actors in a social network?

  We induce a clustering of edges by formality.

- Are there regular structures that emerge across signed networks?

  We estimate a prior distribution over signed network structures.

- How does language reflect and reproduce social relationships?

Modeling formality of address

- What is the nature of the relationships between actors in a social network?

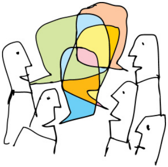
  We induce a clustering of edges by formality.

- Are there regular structures that emerge across signed networks?

  We estimate a prior distribution over signed network structures.

- How does language reflect and reproduce social relationships?

  We estimate a likelihood distribution over address terms given formality.



Related work

Signed social networks Hassan et al. (2012) and West et al. (2014) induce signs from sentiment analysis;

Formality of address Faruqui & Padó (2011) project formality from German into English;

Power in language Several researchers link linguistic features to annotations of power relations (Bramsen et al., 2011; Prabhakaran et al., 2012; Danescu-Niculescu-Mizil et al., 2012).

Key difference: we assume no labels, just text on a social network.

A generative model of networked content

- Undirected graph $G = \langle i, j \rangle$, $i < j$

A generative model of networked content

- Undirected graph $G = \langle i, j \rangle$, $i < j$
- All edges have labels $y_{ij} \in \mathcal{Y}$

A generative model of networked content

- Undirected graph $G = \langle i, j \rangle$, $i < j$
- All edges have labels $y_{ij} \in \mathcal{Y}$
- Content vectors $\boldsymbol{x}_{i \to j}, \boldsymbol{x}_{j \leftarrow i} \in \mathbb{N}^V$, where V is the size of the vocabulary

A generative model of networked content

- Undirected graph $G = \langle i, j \rangle$, $i < j$
- All edges have labels $y_{ij} \in \mathcal{Y}$
- Content vectors $\boldsymbol{x}_{i \rightarrow j}, \boldsymbol{x}_{j \leftarrow i} \in \mathbb{N}^V$, where V is the size of the vocabulary
- Each edge label has **two** associated distributions, so that:

$$
\begin{aligned}
\boldsymbol{x}_{i \rightarrow j} \mid y_{ij} &\sim \text{Multinomial}(\boldsymbol{\theta}^{\rightarrow}_{y_{ij}}) \\
\boldsymbol{x}_{i \leftarrow j} \mid y_{ij} &\sim \text{Multinomial}(\boldsymbol{\theta}^{\leftarrow}_{y_{ij}}).
\end{aligned}
$$

A generative model of networked content

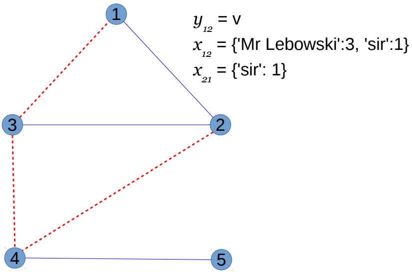
- Undirected graph $G = \langle i, j \rangle$, $i < j$
- All edges have labels $y_{ij} \in \mathcal{Y}$
- Content vectors $\boldsymbol{x}_{i \to j}, \boldsymbol{x}_{j \leftarrow i} \in \mathbb{N}^V$, where V is the size of the vocabulary
- Each edge label has **two** associated distributions, so that:

$$\begin{aligned}
\boldsymbol{x}_{i \to j} \mid y_{ij} \sim & \text{Multinomial}(\boldsymbol{\theta}^{\to}_{y_{ij}}) \\
\boldsymbol{x}_{i \leftarrow j} \mid y_{ij} \sim & \text{Multinomial}(\boldsymbol{\theta}^{\leftarrow}_{y_{ij}}).
\end{aligned}$$

- We add the constraint $\boldsymbol{\theta}^{\to}_{y_{ij}} = \boldsymbol{\theta}^{\leftarrow}_{y_{ij}} = \boldsymbol{\theta}_{y_{ij}}$; without it, we can generalize to directed signs.

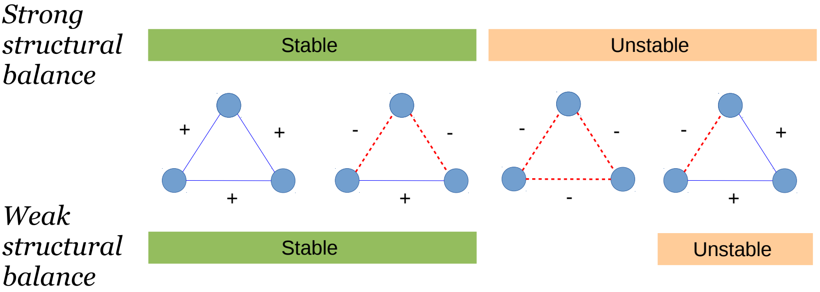
Example

- At inference time, we observe x but not y.
- Inferring y gives a labeling over edges.
- Estimating θ gives the distribution over addresses for each edge type.



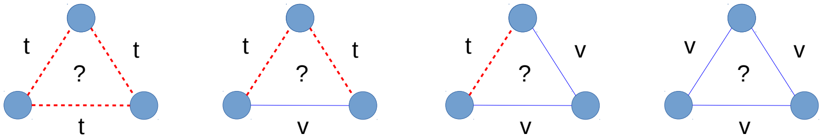
Adding social theory

- So far, this is just a mixture model over dyads.
- But social theory may tell us that not all label configurations are equally likely.
- Ex: **structural balance theory** describes networks of friend/enemy links, where signed triads may be stable or unstable:



Social theories with unknown parameters

- West et al (2014) incorporate structural balance theory by preferring stable triads.
- But what if the magnitude, and even the direction of the effect of each triad type is a priori unknown?



- We assume a triadic form, but make no assumptions about the specifics.

Social theory in a prior distribution

Assume the prior factors over dyads and triads.

$$P(\boldsymbol{y}; G, \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)} \times \exp \sum_{\langle i,j \rangle \in G} \boldsymbol{\eta}^\top \boldsymbol{f}(y_{ij}, i, j, G)$$
$$\times \exp \sum_{\langle i,j,k \rangle \in \mathcal{T}(G)} \beta_{y_{ij}, y_{jk}, y_{ik}},$$

where,

- $Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)$ is a normalizing constant;
- $\boldsymbol{f}(y_{ij}, i, j, G)$ is a set of dyad features, with associated weights $\boldsymbol{\eta}$;
- $\mathcal{T}(G)$ is the set of triads in the graph G;
- $\beta_{y_{ij}, y_{jk}, y_{ik}}$ scores the stability of a triad type.

Social theory in a prior distribution

Assume the prior factors over dyads and triads.

$$P(\boldsymbol{y}; G, \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)} \times \exp \sum_{\langle i,j \rangle \in G} \boldsymbol{\eta}^\top \boldsymbol{f}(y_{ij}, i, j, G) \times \exp \sum_{\langle i,j,k \rangle \in \mathcal{T}(G)} \beta_{y_{ij}, y_{jk}, y_{ik}},$$

where,

- $Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)$ is a normalizing constant;
- $\boldsymbol{f}(y_{ij}, i, j, G)$ is a set of dyad features, with associated weights $\boldsymbol{\eta}$;
- $\mathcal{T}(G)$ is the set of triads in the graph G;
- $\beta_{y_{ij}, y_{jk}, y_{ik}}$ scores the stability of a triad type.

Social theory in a prior distribution

Assume the prior factors over dyads and triads.

$$P(\boldsymbol{y}; G, \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)} \times \exp \sum_{\langle i,j \rangle \in G} \boldsymbol{\eta}^\top \boldsymbol{f}(y_{ij}, i, j, G) \times \exp \sum_{\langle i,j,k \rangle \in \mathcal{T}(G)} \beta_{y_{ij}, y_{jk}, y_{ik}},$$

where,

- $Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)$ is a normalizing constant;
- $\boldsymbol{f}(y_{ij}, i, j, G)$ is a set of dyad features, with associated weights $\boldsymbol{\eta}$;
- $\mathcal{T}(G)$ is the set of triads in the graph G;
- $\beta_{y_{ij}, y_{jk}, y_{ik}}$ scores the stability of a triad type.

Social theory in a prior distribution

Assume the prior factors over dyads and triads.

$$P(\boldsymbol{y}; G, \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)} \times \exp \sum_{\langle i,j \rangle \in G} \boldsymbol{\eta}^\top \boldsymbol{f}(y_{ij}, i, j, G)$$
$$\times \exp \sum_{\langle i,j,k \rangle \in \mathcal{T}(G)} \beta_{y_{ij}, y_{jk}, y_{ik}},$$

where,

- $Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)$ is a normalizing constant;
- $\boldsymbol{f}(y_{ij}, i, j, G)$ is a set of dyad features, with associated weights $\boldsymbol{\eta}$;
- $\mathcal{T}(G)$ is the set of triads in the graph G;
- $\beta_{y_{ij}, y_{jk}, y_{ik}}$ scores the stability of a triad type.

Social theory in a prior distribution

Assume the prior factors over dyads and triads.

$$P(\boldsymbol{y}; G, \boldsymbol{\eta}, \boldsymbol{\beta}) = \frac{1}{Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)} \times \exp \sum_{\langle i,j \rangle \in G} \boldsymbol{\eta}^\top \boldsymbol{f}(y_{ij}, i, j, G)$$
$$\times \exp \sum_{\langle i,j,k \rangle \in \mathcal{T}(G)} \beta_{y_{ij}, y_{jk}, y_{ik}},$$

where,

- $Z(\boldsymbol{\eta}, \boldsymbol{\beta}; G)$ is a normalizing constant;
- $\boldsymbol{f}(y_{ij}, i, j, G)$ is a set of dyad features, with associated weights $\boldsymbol{\eta}$;
- $\mathcal{T}(G)$ is the set of triads in the graph G;
- $\beta_{y_{ij}, y_{jk}, y_{ik}}$ scores the stability of a triad type.

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

- The likelihood factors across dyads;
- The prior factors across dyads and triads.

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

- The **likelihood** factors across dyads;
- The prior factors across dyads and triads.

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

- The likelihood factors across dyads;
- The **prior** factors across dyads and triads.

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

Bayesian inference answers several questions:

1. What is the label (formality) of each dyad?
2. How is formality expressed in language?
3. Are there structural regularities across networks?

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

Bayesian inference answers several questions:

1. What is the label (formality) of each dyad?
2. How is formality expressed in language?
3. Are there structural regularities across networks?

Complete model specification

$$P(\boldsymbol{y}, \boldsymbol{x} \mid G; \Theta, \boldsymbol{\beta}, \boldsymbol{\eta}) = P(\boldsymbol{x} \mid \boldsymbol{y}; \Theta) P(\boldsymbol{y} \mid G; \boldsymbol{\beta}, \boldsymbol{\eta})$$

Bayesian inference answers several questions:

1. What is the label (formality) of each dyad?
2. How is formality expressed in language?
3. Are there structural regularities across networks?

Intractability of Inference

- The normalizing constant z requires summing across all labelings.

  The number of labelings is $\mathcal{O}(\#|\mathcal{Y}|^N)$.

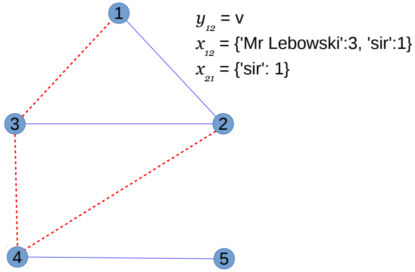
- West et al (2014) show that optimizing an objective over dyads and triads is NP-hard.

  Even exact posterior decoding of $\boldsymbol{y}$ is not tractable, given point estimates for parameters $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\beta}$.

- We therefore apply a variational approximation.
  - We make point estimates of the parameters $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\beta}$ (empirical Bayes).
  - We represent uncertainty over $\boldsymbol{y}$ through a factorized distribution $Q(\boldsymbol{y}) = \prod_{\langle i,j \rangle \in G} q_{ij}(\boldsymbol{y}_{ij})$.

Let's do it!

- The Cornell Movie Dialogue corpus offers 300K conversational turns between 10K dyads, in 617 movies.
- All we need are the address terms...
- But no such resource appears to exist!



Forms of address

Names such as Barack, Barack Hussein Obama.

Titles such as Ms., Dr., Private, Reverend.
Titles can be used for address either by preceding a name (e.g., Colonel Kurtz), or in isolation (e.g., Yes, Colonel.).

Placeholder names such as dude, bro, brother, sweetie, cousin, and asshole.
These terms can be used for address only in isolation.

Subtasks

- Build a vocabulary of titles.
- Build a vocabulary of placeholder names.
- Distinguish address **tokens**:

  His/O name/O is/O **Lebowski/O** *?/O*
  That's/O your/O name/O, **Dude/ADDR**

- There is surprisingly little prior work on these problems, including lexicons.

Automatic address annotations

Text:	You	're	Mr.	Lebowski	.
POS:	PRP	VBP	NNP	NNP	.
Address:	O	O	B-ADDR	L-ADDR	O

1. Look for character names (mined from rotten tomatoes).
2. Identify NNP tag sequences including those names.
3. Automatically label those sequences as entity spans.

Automatic address annotations

Text:	You	're	Mr.	Lebowski	.
POS:	PRP	VBP	NNP	NNP	.
Address:	O	O	B-ADDR	L-ADDR	O

1. Look for character names (mined from rotten tomatoes).
2. Identify NNP tag sequences including those names.
3. Automatically label those sequences as entity spans.

Automatic address annotations

Text:	You	're	Mr.	Lebowski	.
POS:	PRP	VBP	NNP	NNP	.
Address:	O	O	B-ADDR	L-ADDR	O

1. Look for character names (mined from rotten tomatoes).
2. Identify NNP tag sequences including those names.
3. Automatically label those sequences as entity spans.

Automatic address annotations

Text:	You	're	Mr.	Lebowski	.
POS:	PRP	VBP	NNP	NNP	.
Address:	O	O	B-ADDR	L-ADDR	O

1. Look for character names (mined from rotten tomatoes).
2. Identify NNP tag sequences including those names.
3. Automatically label those sequences as entity spans.

Address tagger: features

Feature	Description
Lexical	The word to be tagged, and its two predecessors and successors, $w_{i-2:i+2}$
POS	The part-of-speech of the token to be tagged, and its neighbors
Case	The case of the word to be tagged, and its neighbors.
Constituency parse	First non-NNP ancestor node of the word w_i in the constituent parse tree, and all leaf node siblings in the tree.
Dependency	All dependency relations involving w_i.
Location	Distance of w_i from the start and the end of the sentence or turn.
Punctuation	All punctuation symbols occurring before and after w_i.
Second person pronoun	All forms of the second person pronoun within the sentence.

Address tagger: accuracy

Class	F-measure	Total Instances
I-ADDR	0.58	53
B-ADDR	0.800	483
U-ADDR	0.987	1864
L-ADDR	0.813	535
O-ADDR	0.993	35975

Lexicon induction: titles

- Run the tagger, find terms that frequently appear at the beginning of address spans containing the character's name.
- We then manually filter out 17 of 34 candidates, obtaining:

agent, aunt, captain, colonel, commander, cousin, deputy, detective, dr, herr, inspector, judge, lord, master, mayor, miss, mister, miz, monsieur, mr, mrs, ms, professor, queen, reverend, sergeant, uncle

Lexicon induction: placeholder names

- Remove the CURRENT-WORD feature from the tagger model, then find terms that are frequently tagged as the unique element in an address span.
- After manually filter out 41 of 96 candidates, we obtain:

asshole, babe, baby, boss, boy, bro, bud, buddy, cocksucker, convict, cousin, cowboy, cunt, dad, darling, dear, detective, doll, dude, dummy, father, fella, gal, ho, hon, honey, kid, lad, lady, lover, ma, madam, madame, man, mate, mister, mon, moron, motherfucker, pal, papa, partner, peanut, pet, pilgrim, pop, president, punk, shithead, sir, sire, son, sonny, sport, sucker, sugar, sweetheart, sweetie, tiger

Feature vector construction

Content features

- Addressee name, including any title in lexicon
  You're **Mr. Lebowski** → Mr. LastName
- Any element in the placeholder name lexicon, if tagged as the unique element in address span
  Thanks, dude → DUDE

Dyad feature: Adamic-Adar metric (normalized mutual friends) for each dyad

Model comparison

	Text	Dyad Feature	Signed triads	Predictive Log-likelihood
M1	✓			
M2	✓		✓	
M3	✓	✓		
M4	✓	✓	✓	

Model comparison

	Text	Dyad Feature	Signed triads	Predictive Log-likelihood
M1	✓			-2133.28
M2	✓		✓	-2018.21
M3	✓	✓		-1884.02
M4	✓	✓	✓	-1582.43

Predictive likelihood is evaluated on held-out address terms for a 10% test fold.

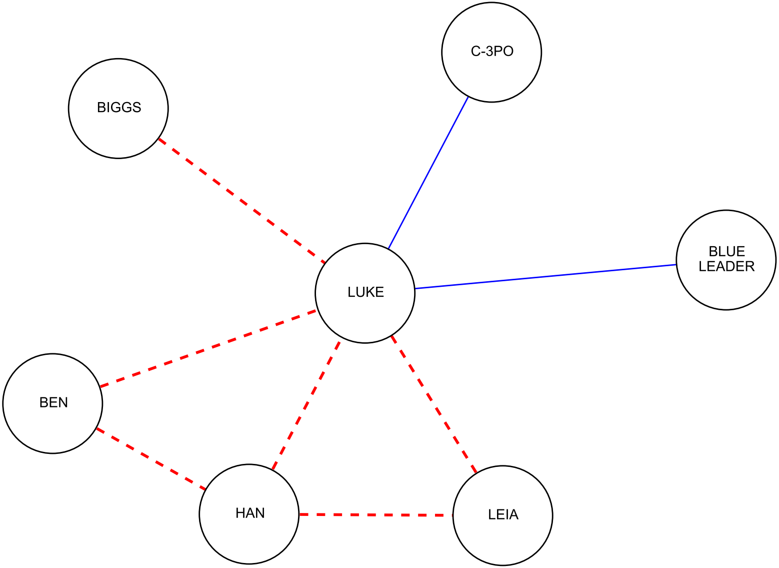
Cluster coherence

V-cluster	T-cluster
sir	FIRSTNAME
mr+LASTNAME	man
mr+FIRSTNAME	baby
mr	honey
miss+LASTNAME	darling
son	sweetheart
mister+FIRSTNAME	buddy
mrs	sweetie

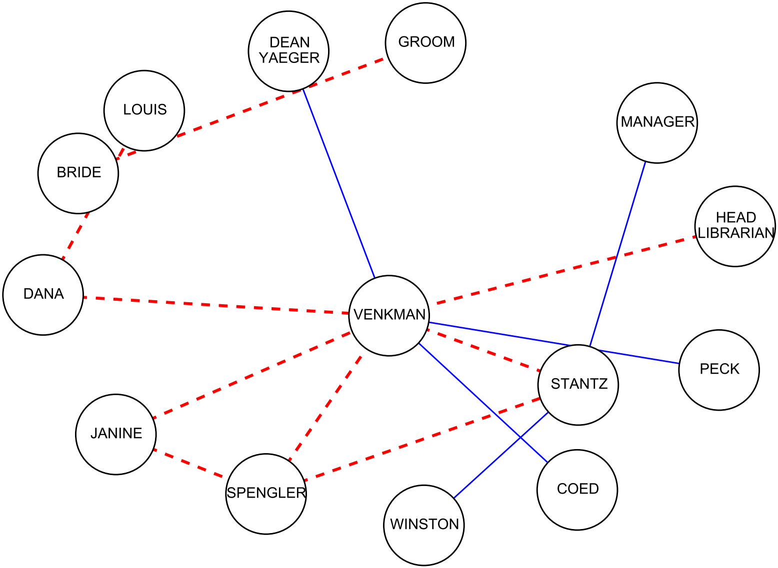
On an intrusion task

- Raters found the intruder in 73% of cases for the full model (M4).
- ... versus 52% in the text-only model (M1).

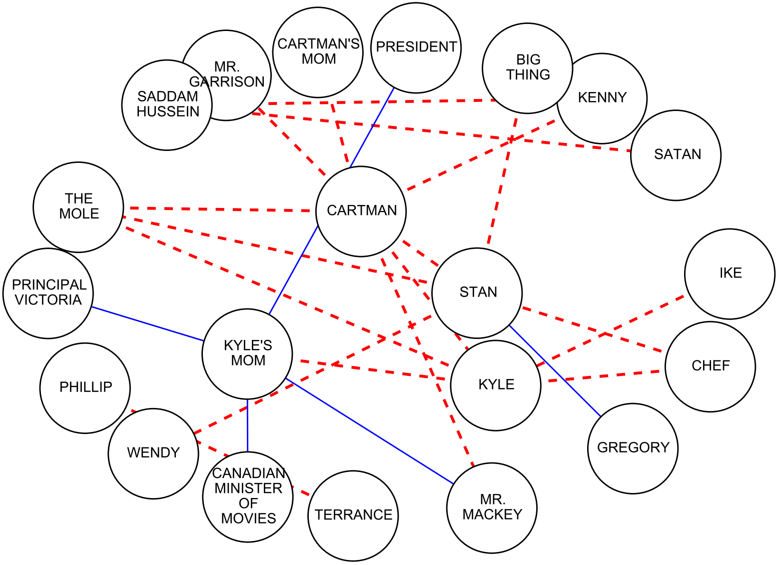
Star Wars



Ghostbusters

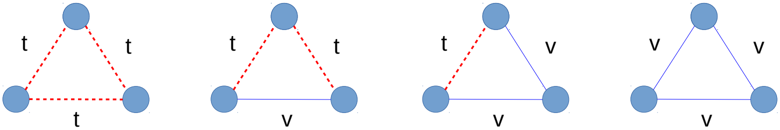


South Park



MR. GARRISON
CARTMAN'S MOM
PRESIDENT
BIG THING
KENNY
SADDAM HUSSEIN
SATAN
THE MOLE
CARTMAN
IKE
PRINCIPAL VICTORIA
STAN
KYLE'S MOM
CHEF
PHILLIP
KYLE
WENDY
GREGORY
CANADIAN MINISTER OF MOVIES
TERRANCE
MR. MACKEY

Network features



+3.73 -6.48 -1.05 +1.23

Overview

Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	homophily	deep learning
Language change	diffusion of innovations	Hawkes process

Overview

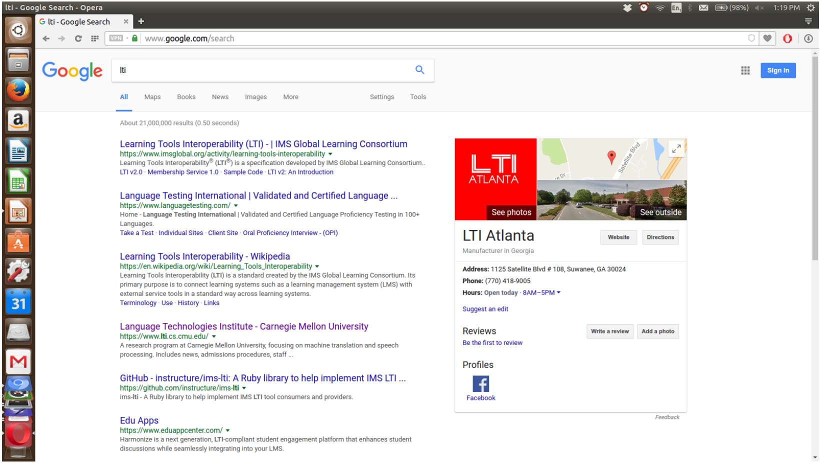
Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	**homophily**	**deep learning**
Language change	diffusion of innovations	Hawkes process

Social Networks for Better NLP

- *Putting Things in Context: Community-specific Embedding Projections for Sentiment Analysis* (Yang et al., 2016).
- *Sentiment Analysis with Social Attention* (Yang & Eisenstein, 2016)

Example 1: Shared context

- Me to nerd friends: I'm giving a talk at the LTI.


lti - Google Search - Opera
1:19 PM
lti - Google Search
www.google.com/search
Google
lti
All
Maps
Books
News
Images
More
Settings
Tools
Sign in
About 21,000,000 results (0.50 seconds)
Learning Tools Interoperability (LTI) - | IMS Global Learning Consortium
https://www.imsglobal.org/activity/learning-tools-interoperability
Learning Tools Interoperability® (LTI®) is a specification developed by IMS Global Learning Consortium..
LTI v2.0 · Membership Service 1.0 · Sample Code · LTI v2: An Introduction
Language Testing International | Validated and Certified Language ...
https://www.languagetesting.com/
Home - Language Testing International | Validated and Certified Language Proficiency Testing in 100+ Languages.
Take a Test · Individual Sites · Client Site · Oral Proficiency Interview - (OPI)
Learning Tools Interoperability - Wikipedia
https://en.wikipedia.org/wiki/Learning_Tools_Interoperability
Learning Tools Interoperability (LTI) is a standard created by the IMS Global Learning Consortium. Its primary purpose is to connect learning systems such as a learning management system (LMS) with external service tools in a standard way across learning systems.
Terminology · Use · History · Links
Language Technologies Institute - Carnegie Mellon University
https://www.lti.cs.cmu.edu/
A research program at Carnegie Mellon University, focusing on machine translation and speech processing. Includes news, admissions procedures, staff ...
GitHub - instructure/ims-lti: A Ruby library to help implement IMS LTI ...
https://github.com/instructure/ims-lti
ims-lti - A Ruby library to help implement IMS LTI tool consumers and providers.
Edu Apps
https://www.eduappcenter.com/
Harmonize is a next generation, LTI-compliant student engagement platform that enhances student discussions while seamlessly integrating into your LMS.
LTI
ATLANTA
See photos
See outside
LTI Atlanta
Website
Directions
Manufacturer in Georgia
Address: 1125 Satellite Blvd # 108, Suwanee, GA 30024
Phone: (770) 418-9005
Hours: Open today · 8AM–5PM
Suggest an edit
Reviews
Write a review
Add a photo
Be the first to review
Profiles
Facebook
Feedback

Label homophily

LTI should be linked to the canonical entity Language Technologies Institute. How?

Label homophily

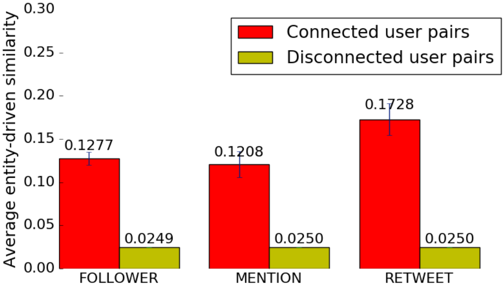
LTI should be linked to the canonical entity LANGUAGE TECHNOLOGIES INSTITUTE. How?

- Jacob's friends say: CLSP, JHU, GaTech, . . .
- Many of these mentions are unambiguous.
- In an **entity embedding space**, $e(\text{CLSP}) \approx e(\text{Language Tech Inst})$, etc.
- If entity mentions are **assortative**, then LANGUAGE TECHNOLOGIES INSTITUTE is the most likely referent.

(McPherson et al., 2001; Thelwall, 2009)

Testing Entity Homophily

- Entity-driven similarity between authors
  - Cosine similarity of entity vectors mentioned by authors.



Ritter et al. (2011)
Cano et al. (2014)

Representations

- Surface features $\phi(\mathbf{x}, y_t, t)$
  - 37 dense features from NER, Wikipedia statistics, etc. (Yang and Chang, 2015)

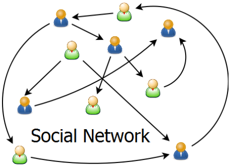
Surface Feature
Entity popularity
Textual similarity
…

Representations

- Surface features $\phi(\mathbf{x}, y_t, t)$
  - 37 dense features from NER, Wikipedia statistics, etc. (Yang and Chang, 2015)

Surface Feature
Entity popularity
Textual similarity
…

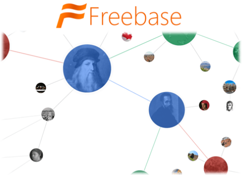
- Distributed representations



Author embeddings

$$\overrightarrow{\text{“Red Sox”}} = \frac{\overrightarrow{\text{“Red”}} + \overrightarrow{\text{“Sox”}}}{2}$$

Mention embeddings



Entity embeddings

Socially-Infused Entity Linking

$$s(\underbrace{\boxed{\mathbf{x}}}_{\text{tweet}}, \underbrace{\boxed{\mathbf{y}}}_{\text{entity assignments}}, \underbrace{\boxed{u}}_{\text{author}}) = g_1(\mathbf{x}, y_t, t) + g_2(\mathbf{x}, y_t, u, t)$$

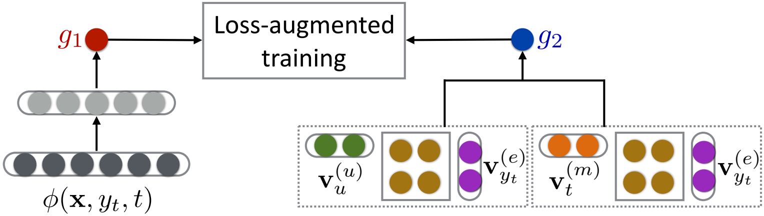
- g_1 is employed to model surface features.
- g_2 is used to capture two assumptions:
  - Entity homophily
  - Semantically related mentions tend to refer similar entities

Socially-Infused Entity Linking

$$g_1(\mathbf{x}, y_t, t; \Theta_1) = \boldsymbol{\beta}^\top \tanh(\mathbf{W}\phi(\mathbf{x}, y_t, t) + \mathbf{b}) + b$$

$$g_2(\mathbf{x}, y_t, u, t; \Theta_2) = \boxed{\mathbf{v}_u^{(u)}}^\top \mathbf{W}^{(u,e)} \boxed{\mathbf{v}_{y_t}^{(e)}} + \boxed{\mathbf{v}_t^{(m)}}^\top \mathbf{W}^{(m,e)} \boxed{\mathbf{v}_{y_t}^{(e)}}$$

author embedding mention embedding entity embedding



Inference

- Non-overlapping structure

Tweet: thought u were a **Red Sox** fan

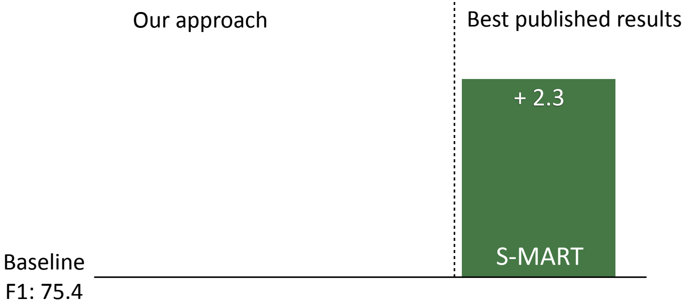
Overlapped mentions: 'Red' 'Sox' **'Red Sox'**

Entities: **Nil**

In order to link 'Red Sox' to a real entity, 'Red' and 'Sox' should be linked to Nil.

Entity Linking Results

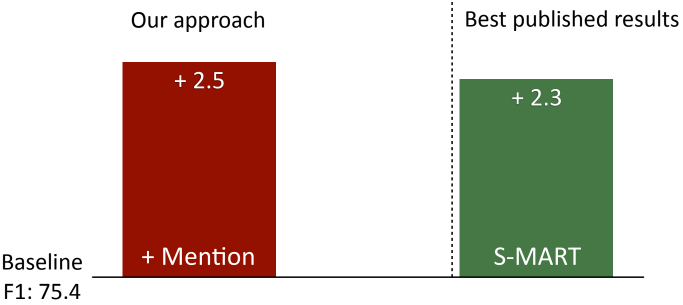
- Dataset: Named Entity Extraction & Linking (NEEL) Challenge



Cano et al. (2014)
Yang and Chang (2015)

Entity Linking Results

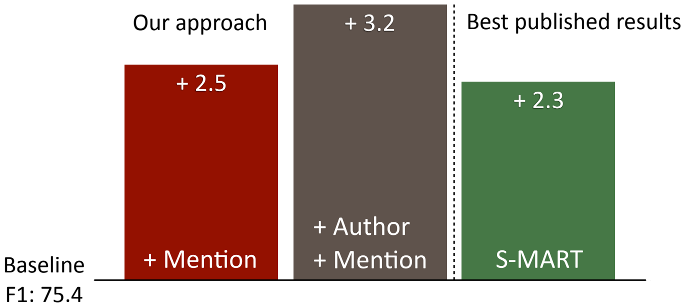
- Dataset: Named Entity Extraction & Linking (NEEL) Challenge



Cano et al. (2014)
Yang and Chang (2015)

Entity Linking Results

- Dataset: Named Entity Extraction & Linking (NEEL) Challenge



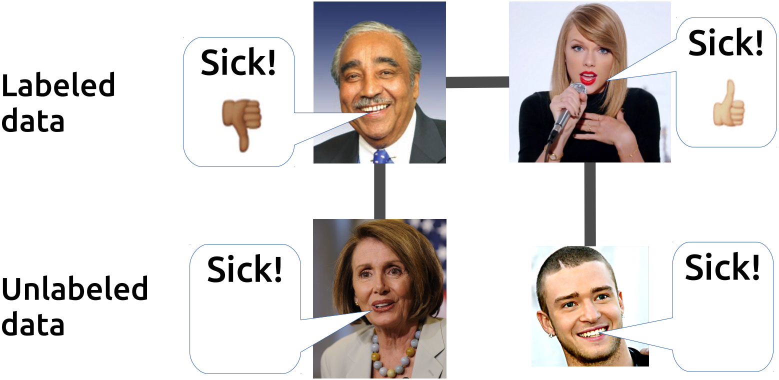
Cano et al. (2014)
Yang and Chang (2015)

Example 3: Semantic 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."

Homophily to the rescue?



Linguistic homophily: linguistic meaning is aligned with the social network.

The SemEval Social Network

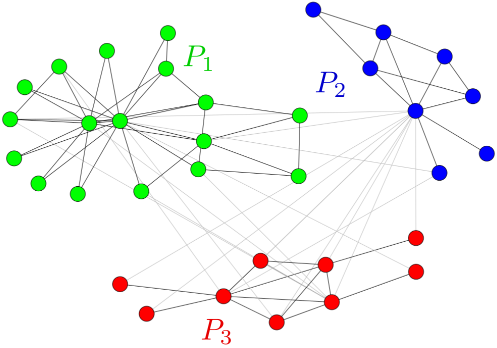
- SemEval 2013-2015 is a standard benchmark for Twitter Sentiment Analysis (Nakov et al., 2013; Rosenthal et al., 2015).
- Follower network statistics:

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

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

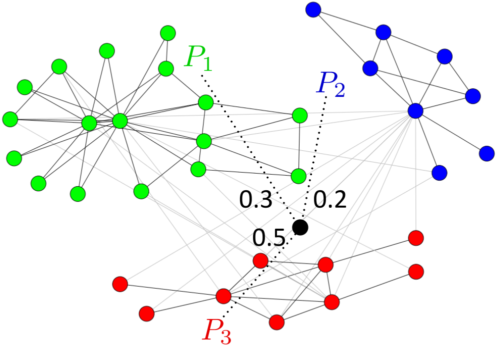
Social Attention

- Ensemble method based on social information
  - Basis models focus on different regions.



Social Attention

- Ensemble method based on social information
  - Basis models focus on different regions.
  - Mixture densities are given by the distances to the regions.



Social Attention

$$p(\boxed{y} \mid \boxed{\mathbf{x}}, \boxed{u}) = \sum_{k=1}^{K} p(z = k \mid u) p(y \mid \mathbf{x}, \boxed{z} = k)$$

sentiment label

tweet

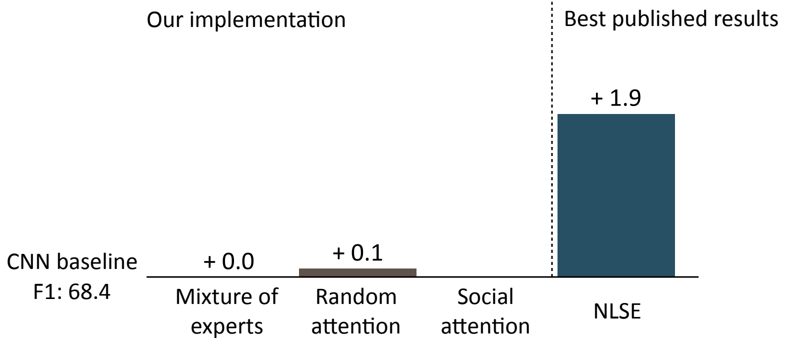
author

latent variable
indicating basis models

- $p(z = k \mid u)$ is a softmax classifier based on user embeddings.
  - Socially linked users share similar attention weights.
- $p(y \mid \mathbf{x}, z = k)$ is a basis convolutional neural network model.
  - Basis models should focus on different network regions.

Sentiment Analysis Results

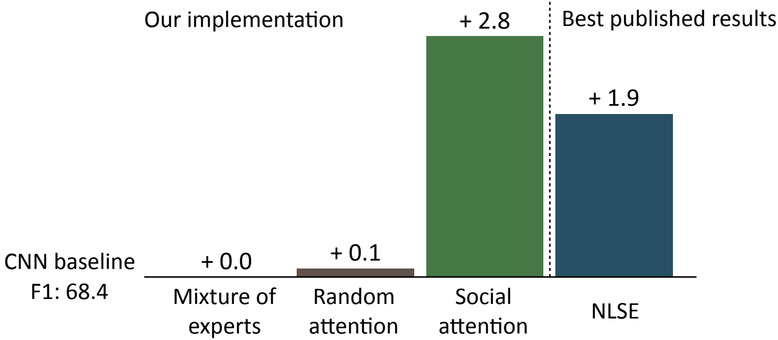
- Dataset: 2013-2015 SemEval Twitter sentiment analysis datasets



Rosenthal et al. (2015); Astudillo et al. (2015)

Sentiment Analysis Results

- Dataset: 2013-2015 SemEval Twitter sentiment analysis datasets



Rosenthal et al. (2015); Astudillo et al. (2015)

Overview

Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	homophily	deep learning
Language change	diffusion of innovations	Hawkes process

Overview

Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	homophily	deep learning
Language change	**diffusion of innovations**	**Hawkes process**

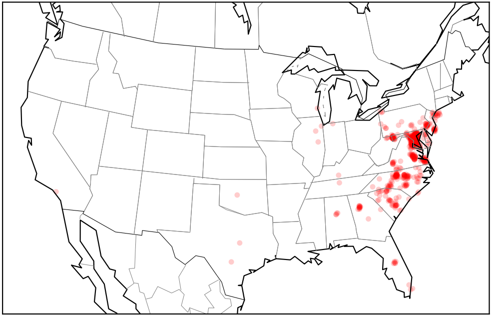
Language change in social networks

- *The Social Dynamics of Language Change in Online Networks*. Rahul Goel, Sandeep Soni, Naman Goyal, John Paparrizos, Hanna Wallach, Fernando Diaz, Jacob Eisenstein. (Goel et al., 2016)

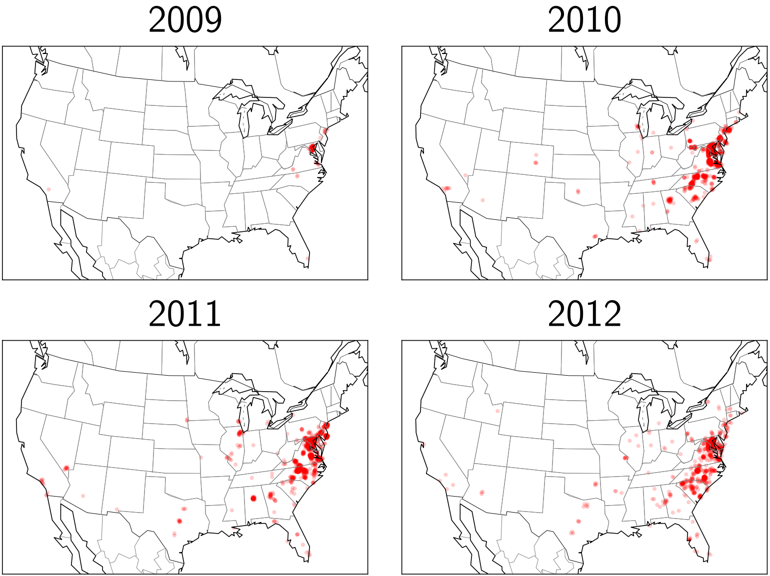
Linguistic innovations in social media

lls (laughing like shit)

- `@user lls, we need to skate homie`
- `@user lls chill`



Change over time



Diffusion of innovations

Adoption of new ideas and practices is a five-step process (Rogers, 1962):

1. Knowledge
2. Persuasion
3. Decision
4. Implementation
5. Confirmation

Social theories of language change

- **Cultural capital**: dialects are a form of social differentiation, which language change helps to maintain (Bourdieu, 1984).
- **Covert prestige**: stigmatized linguistic forms can convey "covert" social advantages, leading to resistance to change (Trudgill, 1972).
- **Indexicality**: non-standard forms "index" various social attributes, and speakers creatively draw from these social associations to craft distinct personal styles (Eckert, 2008).

Hypotheses

This work uses large-scale Twitter data to test three hypotheses about language change.

- **H1**: language change is transmitted across social networks that are visible from metadata in online social media platforms.
- **H2**: geographically local social network ties are better conduits of language change.
- **H3**: strong ties are better conduits of innovations.

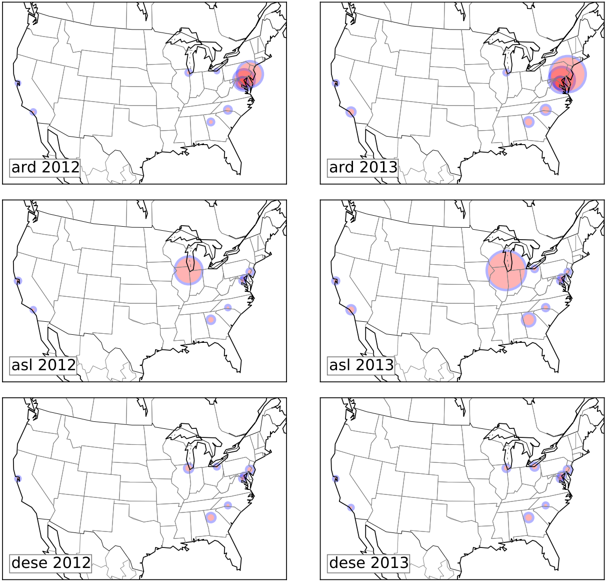
Dataset

- Twitter analysis is usually conducted on a **sample** from the streaming API (e.g., Eisenstein et al., 2010, 2014).
- Modeling the fine structure of language change requires **complete data**, because random samples miss most of the co-occurrences that reveal sociolinguistic influence.
- This work: a dataset of all public tweets between 2011 and 2014, with 4.35 million unique user accounts.

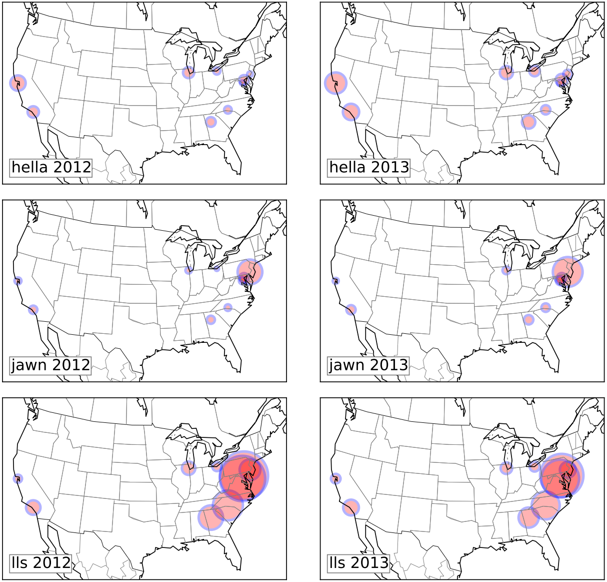
Cities and distinctive features

- Atlanta: ain, dese, yeen
- Baltimore: ard, inna, lls, phony
- Charlotte: cookout
- Chicago: asl, mfs
- Los Angeles: graffiti, tfti
- Philadelphia: ard, ctfuu, jawn
- San Francisco: hella
- Washington, DC: inna, lls, stamp

Linguistic variables



Linguistic variables



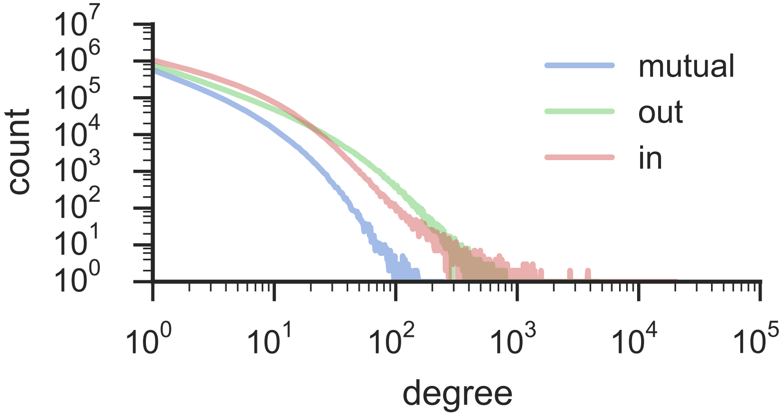
Social network

Two users are considered to have a social network tie if they have each mentioned each other in a message, e.g.

- User1: @user2 salut
- User2: @user1 what's up?

This *mention network* is be more socially meaningful than the *articulated network* of follower-followee links (Huberman et al., 2008; Puniyani et al., 2010).

Social network



The symmetrized ("mutual") mention network yields a more credible degree distribution.

Summary of data

Social network

Bart	Lisa
Bart	Milhouse
Lisa	Homer
Homer	Barney
...	...

Locations

Bart	Los Angeles
Milhouse	Los Angeles
Lisa	Atlanta
Homer	Chicago
...	...

Language

Bart	jawn	Feb 1, 2013, 13:45
Milhouse	jawn	Feb 1, 2013, 13:50
Homer	hella	Feb 1, 2013, 18:15
Bart	lls	Feb 2, 2013, 07:30
Milhouse	lls	Feb 2, 2013, 07:40
...	...	...

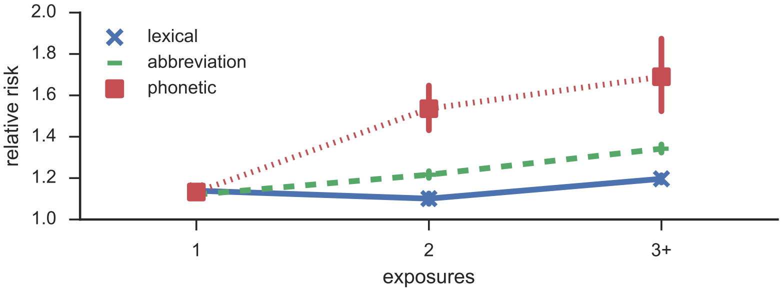
Hypotheses

- **H1**: Language change is transmitted across social networks that are visible from metadata in online social media platforms.
- **H2**: Geographically local social network ties are better conduits of language change.
- **H3**: Strong ties are better conduits of language change.

Hypotheses

- **H1**: Language change is transmitted across social networks that are visible from metadata in online social media platforms.
- **H2**: Geographically local social network ties are better conduits of language change.
- **H3**: Strong ties are better conduits of language change.

Diffusion of innovations



- Relative risk: likelihood of infection given exposure, normalized against rate in randomly-rewired network.
- Rel. risk > 1: evidence of non-random contagion.
- For phonetic variables, risk increases with multiple exposures, a characteristic of **complex contagion**.

Hypotheses

- **H1**: Language change is transmitted across social networks that are visible from metadata in online social media platforms.
- **H2**: Geographically local social network ties are better conduits of language change.
- **H3**: Strong ties are better conduits of language change.

Hypotheses

- **H1**: Language change is transmitted across social networks that are visible from metadata in online social media platforms.
- **H2**: Geographically local social network ties are better conduits of language change.
- **H3**: Strong ties are better conduits of language change.

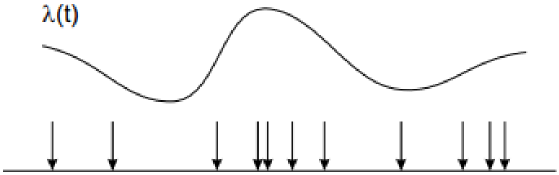
The Poisson process

- Suppose we have a cascade of event times, $\{t_n\}_{n \in 1 \ldots N}$.
- Let $y(t_1, t_2)$ be the count of events between times t_1 and t_2. Then,

$$y(t_1, t_2) \sim \text{Poisson}(\Lambda(t_1, t_2)) \tag{1}$$

$$\Lambda(t_1, t_2) = \int_{t_1}^{t_2} \lambda(t) dt. \tag{2}$$

The Poisson process



For example:

- $y(t_1, t_2)$ is the count of the word lls between 2013 and 2014
- $\lambda(t)$ is the (continuously varying) intensity function.

Hawkes process

A Poisson process in which the intensity function depends on the history (Hawkes, 1971)

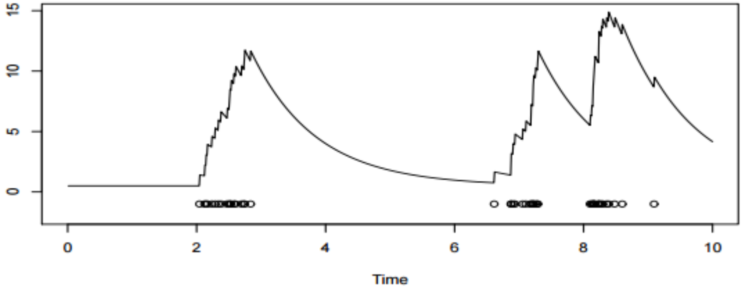
$$\lambda(t) = \mu + \alpha \sum_{t_n < t} \kappa(t - t_n), \tag{3}$$

where the time kernel κ is typically defined as,

$$\kappa(\Delta t) = e^{-\gamma \Delta t}. \tag{4}$$

- μ is the **base rate**;
- α captures the degree of self-excitation;
- γ is the time scale.

Hawkes process



For example:

- $y(t_1, y_2)$ is the count of the word lls
- α captures the tendency of usages of lls to "excite" other usages.

Multivariate Hawkes process

Now suppose each event has some *source* m.

- The cascade is $\{(t_n, m_n)\}_{n \in 1 \ldots N}$.
- The intensity for source m is,

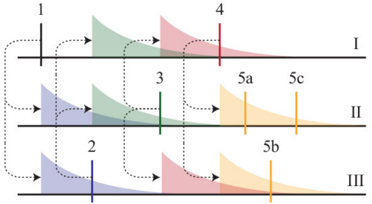
$$\lambda_m(t) = \mu_m + \sum_{t_n < t} \alpha_{m_n \to m} \kappa(t - t_n), \tag{5}$$

where $\alpha_{m_n \to m}$ is the excitation exerted by events with source m_n on source m.

Multivariate Hawkes process

For example:

- Each source m corresponds to an individual social media user.
- $y_m(t_1, t_2)$ is the count of usages of lls by user m between t_1 and t_2.
- $\alpha_{m_1 \to m_2}$ captures the influence of m_1 on m_2.



See Blundell et al. (2012) for another application to language data.

Maximum likelihood estimation

$$\mathcal{L}(\{(t_n, m_n)\}_{n\in 1\ldots N}) = \sum_{n=1}^{N} \log \lambda_m(t_n) - \sum_{m=1}^{M} \Lambda_m(0, T) \qquad (6)$$

$$= \sum_{n=1}^{N} \log \lambda_m(t_n) - \sum_{m=1}^{M} \int_0^T \lambda_m(t) dt \qquad (7)$$

Estimation: maximum likelihood s.t. $\alpha > 0, \mu > 0$.

- Convex in the parameters α and μ
- Linear complexity in the number of parameters and the number events.

Maximum likelihood estimation

$$\mathcal{L}(\{(t_n, m_n)\}_{n\in 1\ldots N}) = \sum_{n=1}^{N} \log \lambda_m(t_n) - \sum_{m=1}^{M} \Lambda_m(0, T) \tag{6}$$

$$= \sum_{n=1}^{N} \log \lambda_m(t_n) - \sum_{m=1}^{M} \int_0^T \lambda_m(t) dt \tag{7}$$

Estimation: maximum likelihood s.t. $\alpha > 0, \mu > 0$.

- Convex in the parameters α and μ
- Linear complexity in the number of parameters and the number events.

But! The number of parameters is **quadratic** in the number of sources.

Parametric Hawkes process

Let's make the infection parameters a function of shared features of each pair of individuals,

$$\alpha_{m_1 \to m_2} = \boldsymbol{\theta}^\top \boldsymbol{f}(m_1 \to m_2). \tag{8}$$

- We now need estimate only $\#|\theta|$ parameters, rather than M^2.
- Because α is an affine function of θ, convexity is preserved.
- Given binary features, non-negativity constraints on the weights $\theta_i \geq 0$ ensure that $\alpha_{m_1,m_2} \geq 0$.

Features

Self-excitation $f_1(m_1 \rightarrow m_2) = 1$ if $m_1 = m_2$, zero otherwise

Social network $f_2(m_1 \rightarrow m_2) = 1$ if there is an edge between m_1 and m_2 in the articulated social network, $(m_1, m_2) \in E$.

Locality $f_3(m_1 \rightarrow m_2) = 1$ if $(m_1, m_2) \in E$ **and** m_1 and m_2 are geolocated to the same metropolitan area.

Tie strength $f_4(m_1 \rightarrow m_2) = 1$ if $(m_1, m_2) \in E$ **and** m_1 and m_2 is a densely embedded tie.

Measuring tie strength

Mutual friends

$$mf(i,j) = \#|\{k : k \in \Gamma(i) \cap \Gamma(j)\}| \tag{9}$$

Measuring tie strength

Mutual friends

$$mf(i,j) = \#|\{k : k \in \Gamma(i) \cap \Gamma(j)\}| \qquad (9)$$

Adamic & Adar (2003): reweight each mutual friend by its degree:

$$aa(i,j) = \sum_{k \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log \#|\Gamma(k)|} \qquad (10)$$

We set $f_4(m_1, m_2) = 1$ if $aa(m_1, m_2)$ is in the 90th percentile.

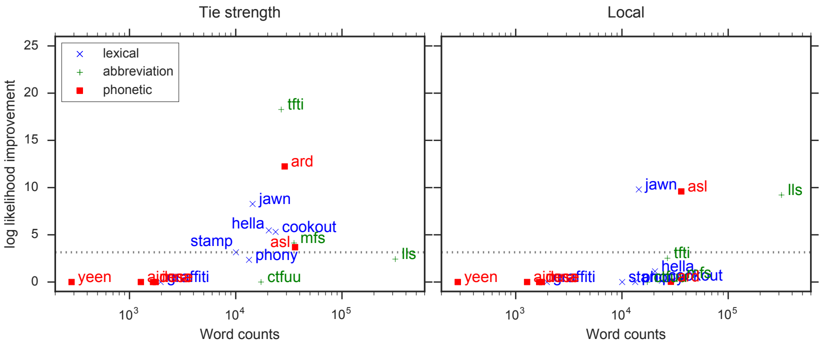
Hypothesis testing

We compare a series of **nested models**.

- $F2 + F1$ **vs** $F1$: is language change transmitted across the social network?
- **All features vs** $F1 + F2 + F4$: are local ties better conduits of language change?
- **All features vs** $F1 + F2 + F3$: are densely embedded ties better conduits of language chnage?

Each comparison is performed using a likelihood ratio test, with correction for multiple comparisons (Benjamini & Hochberg, 1995).

Goodness of fit



- Tie strength feature improves model fit for many words, evidence for H2.
- Geography locality rarely improves model fit, contradicting H3.

Overview

Language	*Social*	*ML*
Formality of address	structural balance	latent variables
Entity linking and sentiment	homophily	deep learning
Language change	diffusion of innovations	Hawkes process

Summary

- Social science theories can be formalized, tested, and exploited using custom machine learning architectures.

Summary

- Social science theories can be formalized, tested, and exploited using custom machine learning architectures.
- **Social meaning** adds depth to computational social science, which typically focuses on network structure, metadata, diffusion of hashtags and URLs.

Summary

- Social science theories can be formalized, tested, and exploited using custom machine learning architectures.
- **Social meaning** adds depth to computational social science, which typically focuses on network structure, metadata, diffusion of hashtags and URLs.
- **Social structure** adds robustness to language technology, which typically treats all authors and speakers as the same.

Acknowledgments

- **Collaborators**: Ming-Wei Chang, Fernando Diaz, Naman Goyal, Rahul Goel, Vinodh Krishnan, John Paparrizos, Umashanthi Pavalanathan, Sandeep Soni, Hanna Wallach, Yi Yang
- **Sponsors**: National Science Foundation, Air Force Office of Scientific Research, National Institutes for Health, Microsoft Research.

References I

Adamic, L. A. & Adar, E. (2003). Friends and neighbors on the web. *Social networks*, *25*(3), 211–230.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday*, *12*(9).

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Blundell, C., Beck, J., & Heller, K. A. (2012). Modelling reciprocating relationships with hawkes processes. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 2600–2608). Curran Associates, Inc.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Harvard University Press.

Bramsen, P., Escobar-Molano, M., Patel, A., & Alonso, R. (2011). Extracting social power relationships from natural language. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 773–782)., Portland, OR.

Cano, A. E., Rizzo, G., Varga, A., Rowe, M., Stankovic, M., & Dadzie, A.-S. (2014). microposts2014 neel challenge: Measuring the performance of entity linking systems in social streams. *Proc. of the Microposts2014 NEEL Challenge*.

Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the Conference on World-Wide Web (WWW)*, (pp. 699–708)., Lyon, France.

Eckert, P. (2000). *Linguistic variation as social practice*. Blackwell.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476.

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*.

Faruqui, M. & Padó, S. (2011). "i thou thee, thou traitor": Predicting formal vs. informal address in english literature. In *Proceedings of the Association for Computational Linguistics (ACL)*, (pp. 467–472)., 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)*.

Hassan, A., Abu-Jbara, A., & Radev, D. (2012). Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, (pp. 6–14). Association for Computational Linguistics.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90.

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

Krishnan, V. & Eisenstein, J. (2015). "You're Mr. Lebowski, I'm The Dude": Inducing address term formality in signed social networks. In *NAACL*.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*, 415–444.

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*.

Nguyen, D., Doğruöz, A. S., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, *42*(3), 537–593.

Nguyen, D., Gravel, R., Trieschnigg, D., & Meder, T. (2013). "how old do you think i am?" a study of language and age in twitter. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*.

Prabhakaran, V., Rambow, O., & Diab, M. (2012). Predicting overt display of power in written dialogs. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, (pp. 518–522).

Puniyani, K., Eisenstein, J., Cohen, S., & Xing, E. P. (2010). Social links from latent topics in microblogs. In *Proceedings of NAACL Workshop on Social Media*, Los Angeles.

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

References III

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*.

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).

Thelwall, M. (2009). Homophily in MySpace. *J. Am. Soc. Inf. Sci.*, *60*(2), 219–231.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban british english of norwich. *Language in Society*, *1*(2), 179–195.

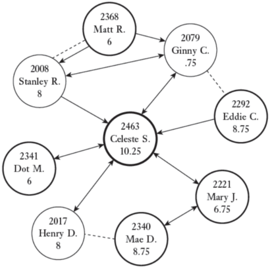
West, R., Paskov, H., Leskovec, J., & Potts, C. (2014). Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, *2*, 297–310.

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. (2016). Sentiment analysis with social attention. *Conditionally Accepted to the Transactions of the Association for Computational Linguistics (TACL)*.

Sociolinguistic approaches

- Sociolinguistics relies heavily on the method of **apparent time** to understand change.
- Social networks are recovered by snowball sampling and interviews.
- These methods have yielded many insights, but generalization is limited by the high cost of data acquisition.



Large-scale study of language change

Social media analysis offers several advantages:

- **Scale**: by studying millions of speakers, it is possible to make more confident generalizations and to investigate more rare phenomena.
- **Speed**: language change in online social media is currently so rapid that real time investigation is practical.
- **Metadata**: explicit records of social interactions make it far more feasible to link language with social structures.