The Bayesian Echo Chamber: Modeling Influence in Conversations

Richard (Fangjian) Guo, Charles Blundell, Hanna Wallach and Katherine A. Heller
Duke University, UCL Gatsby Unit, Microsoft Research and Duke University
Contact: guo@cs.duke.edu

Motivation: mapping influence from linguistic accommodation

Question: Who influences whom in a conversation?

- The influence relations are usually **implicit** in social interactions. Can we infer them from *"who said what at what time"*? (availability of many transcript data)
- Existing approach — modeling the dynamics of turn-taking behavior, i.e. *"who speaks next?"* (e.g. you are influencing me if I respond quickly to what you said)
- Yet, content information should tell even more. Can we infer influence from **content + dynamics**?
- **Linguistic accommodation** phenomenon in *sociolinguistics*
  - When two people interact, the use of a word by one person can **increase** the other person's **probability** of **subsequently using that word**.
  - The language used by a less powerful person will drift further so as to more closely resemble or "**accommodate**" the language used by **more powerful** people.

Influence via Linguistic Accommodation: our model

The Bayesian Echo Chamber specifies a **Markovian generative process** for the words that occur in a set of utterances $\{\mathcal{W}^{(p)}\}_{p=1}^{P}$ made by P people, conditioned on the utterance start times and durations. Our model is inspired by both dynamic Bayesian language modeling and multivariate Hawkes processes.

- Each utterance made by p consists of $L_n^{(p)}$ word tokens, i.e., $\mathcal{W}^{(p)} = \{\{w_{l,n}^{(p)}\}_{l=1}^{L_n^{(p)}}\}_{n=1}^{N^{(p)}(T)}$. Each token is an instance of one of V unique word types.
- The l^{th} token in the n^{th} utterance made by person p is drawn from a Dirichlet-Multinomial distribution.

$$w_{l,n}^{(p)} \sim \text{Categorical}\,(\boldsymbol{\phi}_n^{(p)}), \quad \boldsymbol{\phi}_n^{(p)} \sim \text{Dirichlet}\,(\alpha^{(p)}, \boldsymbol{B}_n^{(p)})$$

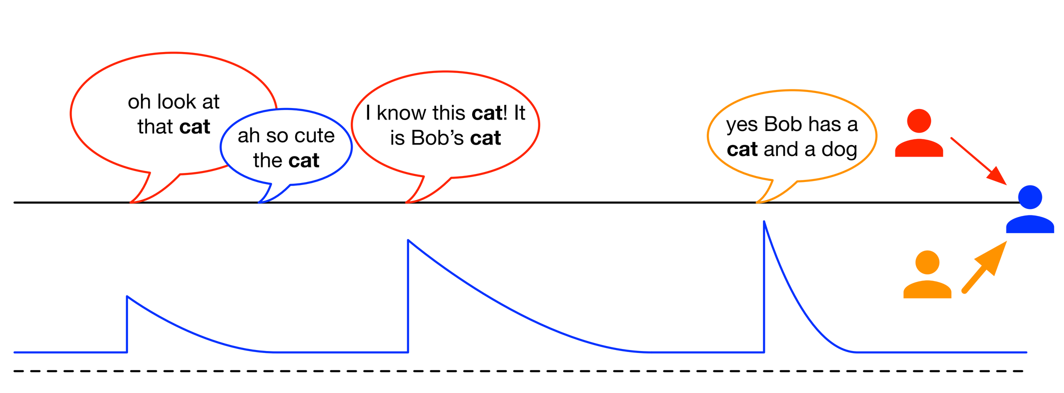
- The utterance-specific $\boldsymbol{B}_n^{(p)}$ is a V-dimensional discrete probability vector that determines the mean of the distribution and satisfies

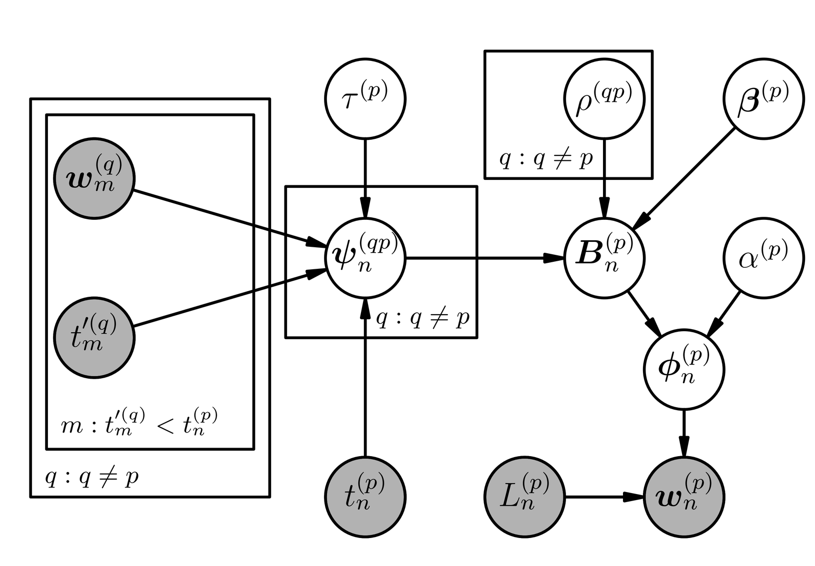
$$B_{v,n}^{(p)} \propto \beta_v^{(p)} + \sum_{q \neq p} \rho^{(qp)} \psi_{v,n}^{(qp)}, \quad \sum_{v=1}^{V} B_{v,n}^{(p)} = 1.$$

- V-dimensional vector $\boldsymbol{\beta}^{(p)} \in \mathbb{R}_+^V$ characterizes person p's inherent language usage.
- Non-negative parameter $\rho^{(qp)}$ controls the degree of linguistic excitation from person q to person p, thus **defining the weighted influence network**. Gamma(1,2) prior is specified to encourage sparsity. Self-excitation is prohibited by enforcing $\rho^{(pp)} = 0$.
- $\boldsymbol{\psi}_n^{(qp)} \in \mathbb{R}_+^V$ is a V-dimensional vector of decayed excitation pseudocounts, constructed from all the previous utterances made by person q.

$$\psi_{v,n}^{(qp)} = \sum_{m:t'^{(q)}_m < t_n^{(p)}} \left(\sum_{l=1}^{L_m^{(q)}} \mathbf{1}(w_{l,m}^{(q)} = v) \right) \exp\left(- \frac{t_n^{(p)} - t'^{(q)}_m}{\tau_L^{(p)}} \right).$$

- $\tau_L^{(p)}$ is a time decay parameter specific to person p that characterizes how fast excitation decays.
- **Inference**: collapsed slice-with-Gibbs MCMC sampling.

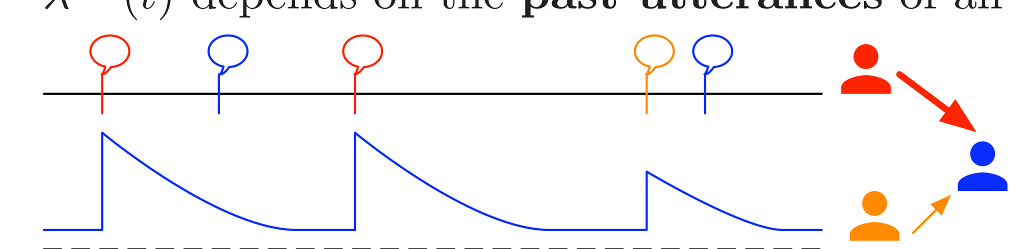




Influence via Turn-Taking: Blundell et al.'s model

Blundell et al.'s model specifies a probabilistic generative process for the time stamps $\{\mathcal{T}^{(p)}\}_{p=1}^{P}$ associated with a set of utterances made by P people.

- The whole conversation is observed in $(0, T]$. Person p's utterance times $\mathcal{T}^{(p)} = \{t_n^{(p)}\}_{n=1}^{N^{(p)}(T)}$. The n-th utterance ends at $t'^{(p)}_n = t_n^{(p)} + \Delta t_n^{(p)}$, assuming the length $\Delta t_n^{(p)}$ is given.
- A **multivariate Hawkes process** model to describe influence as *excitation*.
  - Each person p is associated with a Hawkes process, which is an inhomogeneous Poisson process with a **conditional stochastic rate function** $\lambda^{(p)}(t)$.
  - $\lambda^{(p)}(t)$ depends on the **past utterances** of all the other people.



$$\lambda^{(p)}(t) = \lambda_0^{(p)} + \sum_{q \neq p} \sum_{n: t'^{(q)}_n < t} g^{(qp)}(t, t'^{(q)}_n)$$

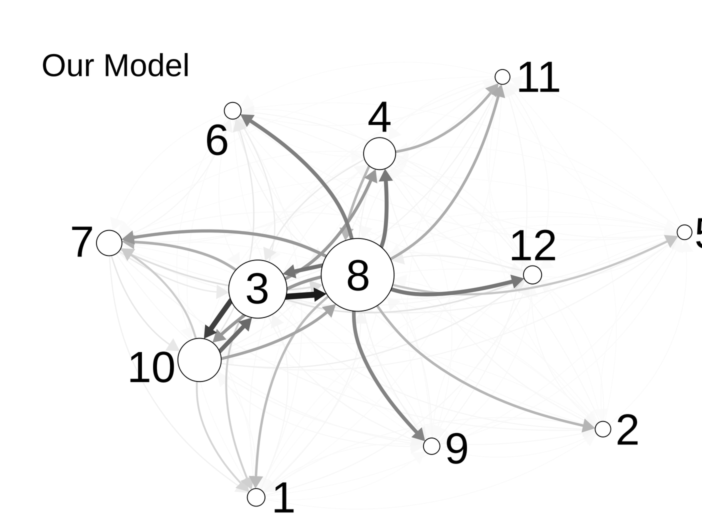
  - Choosing an exponential kernel $g^{(qp)}(t, u) = \nu^{(qp)} \exp\left(-(t-u)\,/\,\tau_T^{(p)}\right)$, then $\nu^{(qp)}$ defines the **influence network**.

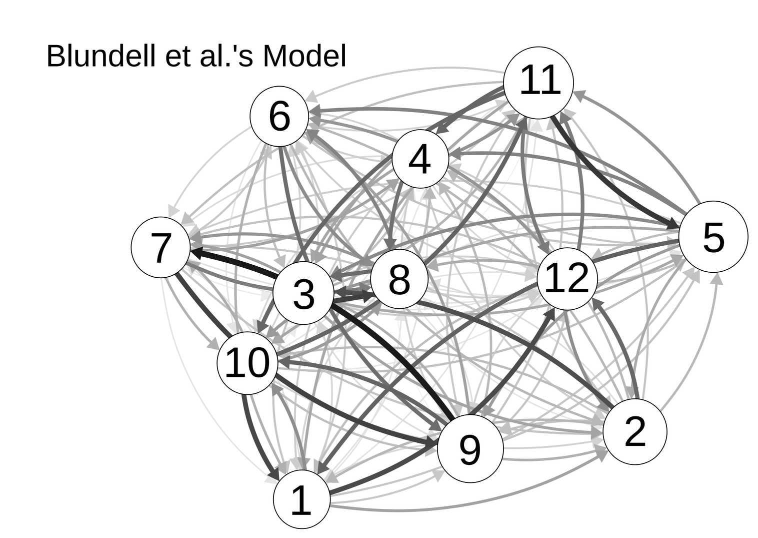
Experiments

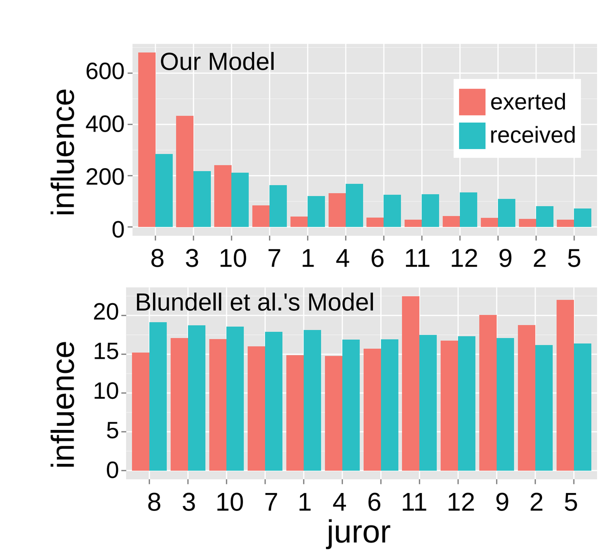
Table 1: Predictive Log Probabilities of Held-Out Data

Dataset	10% test set			20% test set		
	Our Model	Unigram	DTM	Our Model	Unigram	DTM
Synthetic	**-4292.97**±0.02	-4297.92±0.04	-4364.81	**-8702.92**±0.04	-8717.77±0.08	-8948.07
DC v. Heller	**-7383.45**±0.12	-7794.25±0.21	-7533.58	**-12404.21**±0.15	-13126.73±0.26	-12744.73
L&G v. Texas	**-6663.33**±0.12	-6937.66±0.18	-6759.06	**-10248.80**±0.21	-10791.25±0.23	-10459.87
Citizens United v. FEC	**-5770.12**±0.14	-6120.67±0.18	-5851.224	**-16370.7**±0.95	-17157.21±0.40	-16400.46
12 Angry Men	**-4667.47**±0.24	-4920.21±0.14	-4691.11	**-8722.97**±0.27	-9222.99±0.25	-8787.35

"12 Angry Men" Transcripts

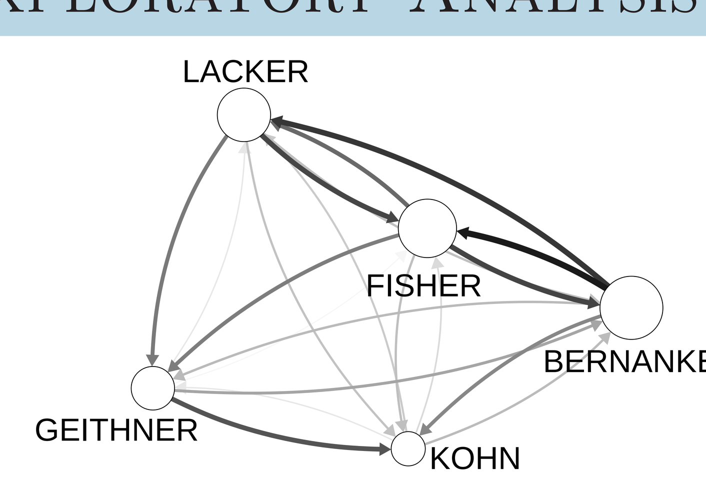




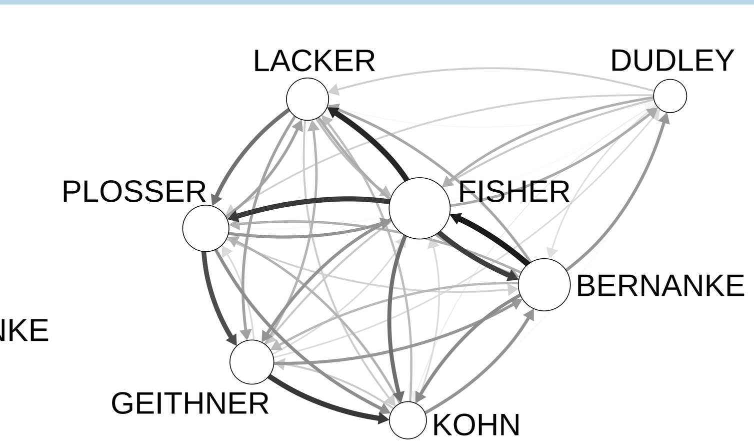


Juror 8: the protagonist of the movie, and initially casts the only "not guilty" vote. The other jurors ultimately change their votes to match his; **Juror 3, 4 and 10**: the last three jurors to change his vote.

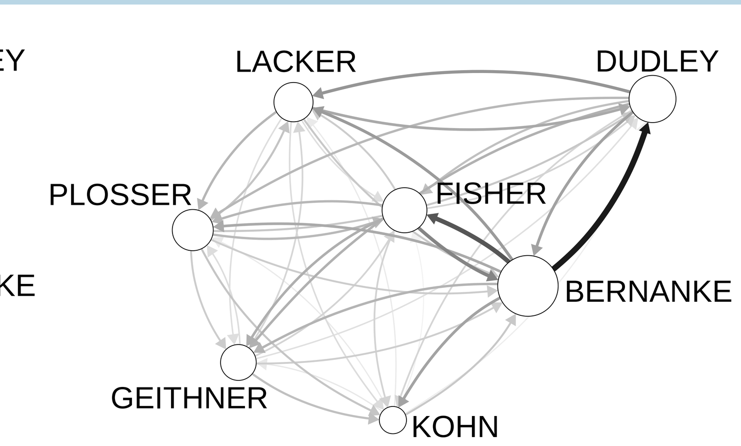
Exploratory Analysis of FOMC Meetings



Pre-crash I: March - June 2006



Pre-crash II: Aug 2006 - Aug 2007



Post-crash: Aug 2007 - Dec 2008

Pre-crash I: Bernanke, Fisher, and Lacker play the biggest roles with Bernanke, the chair, exerting the most influence over others. Unlike Bernanke, Fisher and Lacker are both "hawks" and thus generally in favor of tightening monetary policy; the meetings in this subset all resulted in an outcome of tightening.
Pre-crash II: Bernanke, Fisher, and Lacker all continue to play significant roles, but the network is much less sparse, with both hawks and "doves" (those generally in favor of easing monetary policy) exerting influence over others.
Post-crash: Bernanke (the chair and a dove) still plays a major role, while Fisher and Lacker's roles are significantly diminished. Instead, Dudley, also a dove and a close ally of Bernanke, plays a much greater role, especially in his relationship with Bernanke. These meetings all resulted in monetary policy easing, a strategy generally favored by doves and opposed by hawks.