

Viet-An Nguyen, Jordan Boyd-Graber, and Philip Resnik. **SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations**. *Association for Computational Linguistics*, 2012, 10 pages.

```
@inproceedings{Nguyen:Boyd-Graber:Resnik-2012,  
Title = {{SITS}: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multi  
Author = {Viet-An Nguyen and Jordan Boyd-Graber and Philip Resnik},  
Booktitle = {Association for Computational Linguistics},  
Year = {2012},  
Location = {Jeju, South Korea},  
Url = {http://umiacs.umd.edu/~jbg//docs/acl_2012_sits.pdf},  
}
```

Links:

- Data [<http://www.cs.umd.edu/~vietan/topicshift/crossfire.zip>]
- Code [<https://github.com/vietansegan/sits>]
- Slides [[http://www.cs.umd.edu/~vietan/ACL12\\_slides.pdf](http://www.cs.umd.edu/~vietan/ACL12_slides.pdf)]
- Appendix [<http://www.cs.umd.edu/~vietan/topicshift/appendix.pdf>]

Downloaded from [http://umiacs.umd.edu/~jbg/docs/acl\\_2012\\_sits.pdf](http://umiacs.umd.edu/~jbg/docs/acl_2012_sits.pdf)

Contact Jordan Boyd-Graber ([jbg@boydgraber.org](mailto:jbg@boydgraber.org)) for questions about this paper.

# SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations

**Viet-An Nguyen**  
Department of Computer Science  
and UMIACS  
University of Maryland  
College Park, MD  
vietan@cs.umd.edu

**Jordan Boyd-Graber**  
iSchool  
and UMIACS  
University of Maryland  
College Park, MD  
jbg@umiacs.umd.edu

**Philip Resnik**  
Department of Linguistics  
and UMIACS  
University of Maryland  
College Park, MD  
resnik@umd.edu

## Abstract

One of the key tasks for analyzing conversational data is segmenting it into coherent topic segments. However, most models of topic segmentation ignore the *social* aspect of conversations, focusing only on the words used. We introduce a hierarchical Bayesian nonparametric model, Speaker Identity for Topic Segmentation (SITS), that discovers (1) the topics used in a conversation, (2) how these topics are shared across conversations, (3) when these topics shift, and (4) a person-specific tendency to introduce new topics. We evaluate against current unsupervised segmentation models to show that including person-specific information improves segmentation performance on meeting corpora and on political debates. Moreover, we provide evidence that SITS captures an individual’s tendency to introduce new topics in political contexts, via analysis of the 2008 US presidential debates and the television program Crossfire.

## 1 Topic Segmentation as a Social Process

Conversation, interactive discussion between two or more people, is one of the most essential and common forms of communication. Whether in an informal situation or in more formal settings such as a political debate or business meeting, a conversation is often not about just one thing: topics evolve and are replaced as the conversation unfolds. Discovering this hidden structure in conversations is a key problem for conversational assistants (Tur et al., 2010) and tools that summarize (Murray et al., 2005) and display (Ehlen et al., 2007) conversational data. Topic segmentation also can illuminate individuals’ agendas (Boydston et al., 2011), patterns of agreement and disagreement (Hawes et al., 2009; Abbott

et al., 2011), and relationships among conversational participants (Ireland et al., 2011).

One of the most natural ways to capture conversational structure is topic segmentation (Reynar, 1998; Purver, 2011). Topic segmentation approaches range from simple heuristic methods based on lexical similarity (Morris and Hirst, 1991; Hearst, 1997) to more intricate generative models and supervised methods (Georgescu et al., 2006; Purver et al., 2006; Gruber et al., 2007; Eisenstein and Barzilay, 2008), which have been shown to outperform the established heuristics.

However, previous computational work on conversational structure, particularly in topic discovery and topic segmentation, focuses primarily on content, ignoring the speakers. We argue that, because conversation is a *social* process, we can understand conversational phenomena better by explicitly modeling behaviors of conversational participants. In Section 2, we incorporate participant identity in a new model we call Speaker Identity for Topic Segmentation (SITS), which discovers topical structure in conversation while jointly incorporating a participant-level social component. Specifically, we explicitly model an individual’s tendency to introduce a topic. After outlining inference in Section 3 and introducing data in Section 4, we use SITS to improve state-of-the-art-topic segmentation and topic identification models in Section 5. In addition, in Section 6, we also show that the per-speaker model is able to discover individuals who shape and influence the course of a conversation. Finally, we discuss related work and conclude the paper in Section 7.

## 2 Modeling Multiparty Discussions

**Data Properties** We are interested in turn-taking, multiparty discussion. This is a broad category, in-

cluding political debates, business meetings, and on-line chats. More formally, such datasets contain  $C$  conversations. A conversation  $c$  has  $T_c$  turns, each of which is a maximal uninterrupted utterance by one speaker.<sup>1</sup> In each turn  $t \in [1, T_c]$ , a speaker  $a_{c,t}$  utters  $N$  words  $\{w_{c,t,n}\}$ . Each word is from a vocabulary of size  $V$ , and there are  $M$  distinct speakers.

**Modeling Approaches** The key insight of topic segmentation is that segments evince lexical cohesion (Galley et al., 2003; Olney and Cai, 2005). Words within a segment will look more like their neighbors than other words. This insight has been used to tune supervised methods (Hsueh et al., 2006) and inspire unsupervised models of lexical cohesion using bags of words (Purver et al., 2006) and language models (Eisenstein and Barzilay, 2008).

We too take the unsupervised statistical approach. It requires few resources and is applicable in many domains without extensive training. Like previous approaches, we consider each turn to be a bag of words generated from an admixture of topics. Topics—after the topic modeling literature (Blei and Lafferty, 2009)—are multinomial distributions over terms. These topics are part of a generative model posited to have produced a corpus.

However, topic models alone cannot model the dynamics of a conversation. Topic models typically do not model the temporal dynamics of individual documents, and those that do (Wang et al., 2008; Gerrish and Blei, 2010) are designed for larger documents and are not applicable here because they assume that most topics appear in every time slice.

Instead, we endow each turn with a binary latent variable  $l_{c,t}$ , called the *topic shift*. This latent variable signifies whether the speaker changed the topic of the conversation. To capture the topic-controlling behavior of the speakers across different conversations, we further associate each speaker  $m$  with a latent *topic shift tendency*,  $\pi_m$ . Informally, this variable is intended to capture the propensity of a speaker to effect a topic shift. Formally, it represents the probability that the speaker  $m$  will change the topic (distribution) of a conversation.

We take a Bayesian nonparametric approach (Müller and Quintana, 2004). Unlike

<sup>1</sup>Note the distinction with phonetic utterances, which by definition are bounded by silence.

parametric models, which *a priori* fix the number of topics, nonparametric models use a flexible number of topics to better represent data. Nonparametric distributions such as the Dirichlet process (Ferguson, 1973) share statistical strength among conversations using a hierarchical model, such as the hierarchical Dirichlet process (HDP) (Teh et al., 2006).

## 2.1 Generative Process

In this section, we develop SITS, a generative model of multiparty discourse that jointly discovers topics and speaker-specific topic shifts from an unannotated corpus (Figure 1a). As in the hierarchical Dirichlet process (Teh et al., 2006), we allow an unbounded number of topics to be shared among the turns of the corpus. Topics are drawn from a base distribution  $H$  over multinomial distributions over the vocabulary, a finite Dirichlet with symmetric prior  $\lambda$ . Unlike the HDP, where every document (here, every turn) draws a new multinomial distribution from a Dirichlet process, the social and temporal dynamics of a conversation, as specified by the binary topic shift indicator  $l_{c,t}$ , determine when new draws happen.

The full generative process is as follows:

1. For speaker  $m \in [1, M]$ , draw speaker shift probability  $\pi_m \sim \text{Beta}(\gamma)$
2. Draw global probability measure  $G_0 \sim \text{DP}(\alpha, H)$
3. For each conversation  $c \in [1, C]$ 
  - (a) Draw conversation distribution  $G_c \sim \text{DP}(\alpha_0, G_0)$
  - (b) For each turn  $t \in [1, T_c]$  with speaker  $a_{c,t}$ 
    - i. If  $t = 1$ , set the topic shift  $l_{c,t} = 1$ . Otherwise, draw  $l_{c,t} \sim \text{Bernoulli}(\pi_{a_{c,t}})$ .
    - ii. If  $l_{c,t} = 1$ , draw  $G_{c,t} \sim \text{DP}(\alpha_c, G_c)$ . Otherwise, set  $G_{c,t} \equiv G_{c,t-1}$ .
    - iii. For each word index  $n \in [1, N_{c,t}]$ 
      - Draw  $\psi_{c,t,n} \sim G_{c,t}$
      - Draw  $w_{c,t,n} \sim \text{Multinomial}(\psi_{c,t,n})$

The hierarchy of Dirichlet processes allows statistical strength to be shared across contexts; within a conversation and across conversations. The per-speaker topic shift tendency  $\pi_m$  allows speaker *identity* to influence the evolution of topics.

To make notation concrete and aligned with the topic segmentation, we introduce notation for *segments* in a conversation. A segment  $s$  of conversation  $c$  is a sequence of turns  $[\tau, \tau']$  such that  $l_{c,\tau} = l_{c,\tau'+1} = 1$  and  $l_{c,t} = 0, \forall t \in (\tau, \tau']$ . When  $l_{c,t} = 0$ ,  $G_{c,t}$  is the same as  $G_{c,t-1}$  and all topics (i.e. multinomial distributions over words)  $\{\psi_{c,t,n}\}$  that generate words in turn  $t$  and the topics  $\{\psi_{c,t-1,n}\}$  that generate words in turn  $t - 1$  come from the same

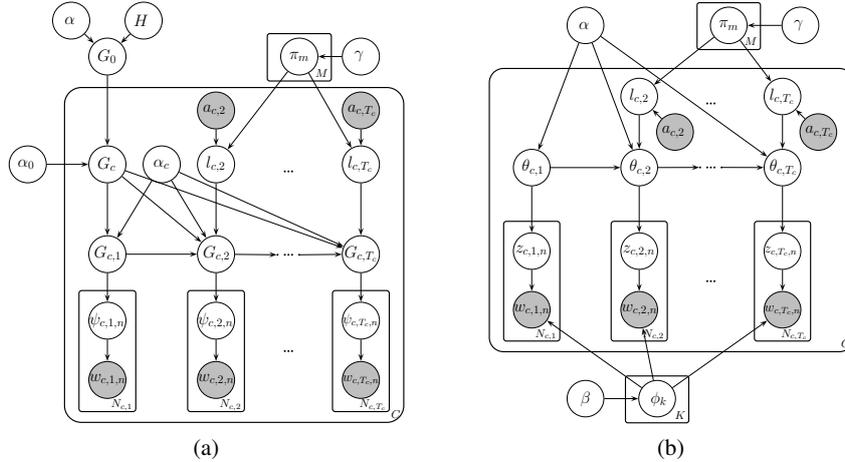


Figure 1: Graphical model representations of our proposed models: (a) the nonparametric version; (b) the parametric version. Nodes represent random variables (shaded ones are observed), lines are probabilistic dependencies. Plates represent repetition. The innermost plates are turns, grouped in conversations.

distribution. Thus all topics used in a segment  $s$  are drawn from a single distribution,  $G_{c,s}$ ,

$$G_{c,s} \mid l_{c,1}, l_{c,2}, \dots, l_{c,T_c}, \alpha_c, G_c \sim \text{DP}(\alpha_c, G_c) \quad (1)$$

For notational convenience,  $S_c$  denotes the number of segments in conversation  $c$ , and  $s_t$  denotes the segment index of turn  $t$ . We emphasize that all segment-related notations are derived from the posterior over the topic shifts  $l$  and not part of the model itself.

**Parametric Version** SITS is a generalization of a parametric model (Figure 1b) where each turn has a multinomial distribution over  $K$  topics. In the parametric case, the number of topics  $K$  is fixed. Each topic, as before, is a multinomial distribution  $\phi_1 \dots \phi_K$ . In the parametric case, each turn  $t$  in conversation  $c$  has an explicit multinomial distribution over  $K$  topics  $\theta_{c,t}$ , identical for turns within a segment. A new topic distribution  $\theta$  is drawn from a Dirichlet distribution parameterized by  $\alpha$  when the topic shift indicator  $l$  is 1.

The parametric version does not share strength within or across conversations, unlike SITS. When applied on a single conversation without speaker identity (all speakers are identical) it is equivalent to (Purver et al., 2006). In our experiments (Section 5), we compare against both.

### 3 Inference

To find the latent variables that best explain observed data, we use Gibbs sampling, a widely used Markov chain Monte Carlo inference technique (Neal, 2000; Resnik and Hardisty, 2010). The state space is latent variables for topic indices assigned to all tokens  $\mathbf{z} = \{z_{c,t,n}\}$  and topic shifts assigned to turns  $\mathbf{l} = \{l_{c,t}\}$ . We marginalize over all other latent variables. Here, we only present the conditional sampling equations; for more details, see our supplement.<sup>2</sup>

#### 3.1 Sampling Topic Assignments

To sample  $z_{c,t,n}$ , the index of the shared topic assigned to token  $n$  of turn  $t$  in conversation  $c$ , we need to sample the *path* assigning each word token to a segment-specific topic, each segment-specific topic to a conversational topic and each conversational topic to a shared topic. For efficiency, we make use of the *minimal path assumption* (Wallach, 2008) to generate these assignments.<sup>3</sup> Under the minimal path assumption, an observation is assumed to have been generated by using a new distribution if and only if there is no existing distribution with the same value.

<sup>2</sup> <http://www.cs.umd.edu/~vietan/topicshift/appendix.pdf>

<sup>3</sup> We also investigated using the maximal assumption and fully sampling assignments. We found the minimal path assumption worked as well as explicitly sampling seating assignments and that the maximal path assumption worked less well.

We use  $N_{c,s,k}$  to denote the number of tokens in segment  $s$  in conversation  $c$  assigned topic  $k$ ;  $N_{c,k}$  denotes the total number of segment-specific topics in conversation  $c$  assigned topic  $k$  and  $N_k$  denotes the number of conversational topics assigned topic  $k$ .  $TW_{k,w}$  denotes the number of times the shared topic  $k$  is assigned to word  $w$  in the vocabulary. Marginal counts are represented with  $\cdot$  and  $*$  represents all hyperparameters. The conditional distribution for  $z_{c,t,n}$  is  $P(z_{c,t,n} = k \mid w_{c,t,n} = w, \mathbf{z}^{-c,t,n}, \mathbf{w}^{-c,t,n}, \mathbf{1}, *) \propto$

$$\frac{N_{c,s_t,k}^{-c,t,n} + \alpha_c \frac{N_{c,k}^{-c,t,n} + \alpha_0 \frac{N_k^{-c,t,n} + \frac{\alpha}{K}}{N_{c,t,n}^{-c,t,n} + \alpha_0}}{N_{c,s_t,\cdot}^{-c,t,n} + \alpha_c}}{\frac{N_{c,s_t,\cdot}^{-c,t,n} + \alpha_c}}{N_{c,s_t,\cdot}^{-c,t,n} + \alpha_c}} \times \begin{cases} \frac{TW_{k,w}^{-c,t,n} + \lambda}{TW_{k,\cdot}^{-c,t,n} + V\lambda}, \\ \frac{1}{V} & k \text{ new.} \end{cases} \quad (2)$$

Here  $V$  is the size of the vocabulary,  $K$  is the current number of shared topics and the superscript  $^{-c,t,n}$  denotes counts without considering  $w_{c,t,n}$ . In Equation 2, the first factor is proportional to the probability of sampling a path according to the minimal path assumption; the second factor is proportional to the likelihood of observing  $w$  given the sampled topic. Since an uninformed prior is used, when a new topic is sampled, all tokens are equiprobable.

### 3.2 Sampling Topic Shifts

Sampling the topic shift variable  $l_{c,t}$  requires us to consider merging or splitting segments. We use  $\mathbf{k}_{c,t}$  to denote the shared topic indices of all tokens in turn  $t$  of conversation  $c$ ;  $S_{a_{c,t},x}$  to denote the number of times speaker  $a_{c,t}$  is assigned the topic shift with value  $x \in \{0, 1\}$ ;  $J_{c,s}^x$  to denote the number of topics in segment  $s$  of conversation  $c$  if  $l_{c,t} = x$  and  $N_{c,s,j}^x$  to denote the number of tokens assigned to the segment-specific topic  $j$  when  $l_{c,t} = x$ .<sup>4</sup> Again, the superscript  $^{-c,t}$  is used to denote exclusion of turn  $t$  of conversation  $c$  in the corresponding counts.

Recall that the topic shift is a binary variable. We use 0 to represent the case that the topic distribution is identical to the previous turn. We sample this assignment  $P(l_{c,t} = 0 \mid \mathbf{I}^{-c,t}, \mathbf{w}, \mathbf{k}, \mathbf{a}, *) \propto$

$$\frac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \times \frac{\alpha_c^{J_{c,s_t}^0} \prod_{j=1}^{J_{c,s_t}^0} (N_{c,s_t,j}^0 - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^0} (x - 1 + \alpha_c)}. \quad (3)$$

<sup>4</sup>Deterministically knowing the path assignments is the primary efficiency motivation for using the minimal path assumption. The alternative is to explicitly sample the path assignments, which is more complicated (for both notation and computation). This option is spelled in full detail in the supplementary material.

In Equation 3, the first factor is proportional to the probability of assigning a topic shift of value 0 to speaker  $a_{c,t}$  and the second factor is proportional to the joint probability of all topics in segment  $s_t$  of conversation  $c$  when  $l_{c,t} = 0$ .

The other alternative is for the topic shift to be 1, which represents the introduction of a new distribution over topics *inside* an existing segment. We sample this as  $P(l_{c,t} = 1 \mid \mathbf{I}^{-c,t}, \mathbf{w}, \mathbf{k}, \mathbf{a}, *) \propto$

$$\frac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \times \left( \frac{\alpha_c^{J_{c,(s_t-1)}^1} \prod_{j=1}^{J_{c,(s_t-1)}^1} (N_{c,(s_t-1),j}^1 - 1)!}{\prod_{x=1}^{N_{c,(s_t-1),\cdot}^1} (x - 1 + \alpha_c)} \frac{\alpha_c^{J_{c,s_t}^1} \prod_{j=1}^{J_{c,s_t}^1} (N_{c,s_t,j}^1 - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^1} (x - 1 + \alpha_c)} \right). \quad (4)$$

As above, the first factor in Equation 4 is proportional to the probability of assigning a topic shift of value 1 to speaker  $a_{c,t}$ ; the second factor in the big bracket is proportional to the joint distribution of the topics in segments  $s_t - 1$  and  $s_t$ . In this case  $l_{c,t} = 1$  means splitting the current segment, which results in two joint probabilities for two segments.

## 4 Datasets

This section introduces the three corpora we use. We preprocess the data to remove stopwords and remove turns containing fewer than five tokens.

**The ICSI Meeting Corpus:** The *ICSI Meeting Corpus* (Janin et al., 2003) is 75 transcribed meetings. For evaluation, we used a standard set of reference segmentations (Galley et al., 2003) of 25 meetings. Segmentations are *binary*, i.e., each point of the document is either a segment boundary or not, and on average each meeting has 8 segment boundaries. After preprocessing, there are 60 unique speakers and the vocabulary contains 3346 non-stopword tokens.

**The 2008 Presidential Election Debates** Our second dataset contains three annotated presidential debates (Boydston et al., 2011) between Barack Obama and John McCain and a vice presidential debate between Joe Biden and Sarah Palin. Each turn is one of two types: *questions* ( $Q$ ) from the moderator or *responses* ( $R$ ) from a candidate. Each clause in a turn is coded with a *Question Topic* ( $T_Q$ ) and a *Response Topic* ( $T_R$ ). Thus, a turn has a list of  $T_Q$ 's and  $T_R$ 's both of length equal to the number of clauses in the turn. Topics are from the Policy Agendas Topics

Speaker	Type	Turn clauses	$T_Q$	$T_R$
Brokaw	$Q$	Sen. Obama, [...] Are you saying [...] that the American economy is going to get much worse before it gets better and they ought to be prepared for that?	1	N/A
Obama	$R$	No, I am confident about the American economy.	1	1
		[...] But most importantly, we're going to have to help ordinary families be able to stay in their homes, make sure that they can pay their bills [...]	1	14
Brokaw	$Q$	Sen. McCain, in all candor, do you think the economy is going to get worse before it gets better?	1	N/A
McCain	$R$	[...] I think if we act effectively, if we stabilize the housing market—which I believe we can,	1	14
		if we go out and buy up these bad loans, so that people can have a new mortgage at the new value of their home	1	14
		I think if we get rid of the cronyism and special interest influence in Washington so we can act more effectively. [...]	1	20

Table 1: Example turns from the annotated 2008 election debates. The topics ( $T_Q$  and  $T_R$ ) are from the Policy Agendas Topics Codebook which contains the following codes of topic: Macroeconomics (1), Housing & Community Development (14), Government Operations (20).

Codebook, a manual inventory of 19 major topics and 225 subtopics.<sup>5</sup> Table 1 shows an example annotation.

To get reference segmentations, we assign each turn a real value from 0 to 1 indicating how much a turn changes the topic. For a question-typed turn, the score is the fraction of clause topics not appearing in the previous turn; for response-typed turns, the score is the fraction of clause topics that do not appear in the corresponding question. This results in a set of *non-binary* reference segmentations. For evaluation metrics that require binary segmentations, we create a binary segmentation by setting a turn as a segment boundary if the computed score is 1. This threshold is chosen to include only true segment boundaries.

**CNN’s Crossfire** *Crossfire* was a weekly U.S. television “talking heads” program engineered to incite heated arguments (hence the name). Each episode features two recurring hosts, two guests, and clips from the week’s news. Our *Crossfire* dataset contains 1134 transcribed episodes aired between 2000 and 2004.<sup>6</sup> There are 2567 unique speakers. Unlike the previous two datasets, *Crossfire* does not have explicit topic segmentations, so we use it to explore speaker-specific characteristics (Section 6).

## 5 Topic Segmentation Experiments

In this section, we examine how well SITS can replicate annotations of when new topics are introduced.

<sup>5</sup> <http://www.policyagendas.org/page/topic-codebook>

<sup>6</sup> <http://www.cs.umd.edu/~vietan/topicshift/crossfire.zip>

We discuss metrics for evaluating an algorithm’s segmentation against a gold annotation, describe our experimental setup, and report those results.

**Evaluation Metrics** To evaluate segmentations, we use  $P_k$  (Beeferman et al., 1999) and WindowDiff (WD) (Pevzner and Hearst, 2002). Both metrics measure the probability that two points in a document will be incorrectly separated by a segment boundary. Both techniques consider all spans of length  $k$  in the document and count whether the two endpoints of the window are (im)properly segmented against the gold segmentation.

However, these metrics have drawbacks. First, they require both hypothesized and reference segmentations to be binary. Many algorithms (e.g., probabilistic approaches) give *non-binary* segmentations where candidate boundaries have real-valued scores (e.g., probability or confidence). Thus, evaluation requires arbitrary thresholding to binarize soft scores. To be fair, thresholds are set so the number of segments are equal to a predefined value (Purver et al., 2006; Galley et al., 2003).

To overcome these limitations, we also use *Earth Mover’s Distance* (EMD) (Rubner et al., 2000), a metric that measures the distance between two distributions. The EMD is the minimal cost to transform one distribution into the other. Each segmentation can be considered a multi-dimensional distribution where each candidate boundary is a dimension. In EMD, a distance function across features allows partial credit for “near miss” segment boundaries. In

addition, because EMD operates on distributions, we can compute the distance between non-binary hypothesized segmentations with binary or real-valued reference segmentations. We use the FastEMD implementation (Pele and Werman, 2009).

**Experimental Methods** We applied the following methods to discover topic segmentations in a document:

- **TextTiling** (Hearst, 1997) is one of the earliest general-purpose topic segmentation algorithms, sliding a fixed-width window to detect major changes in lexical similarity.
- **P-NoSpeaker-S**: parametric version without speaker identity run on each conversation (Purver et al., 2006)
- **P-NoSpeaker-M**: parametric version without speaker identity run on all conversations
- **P-SITS**: the parametric version of SITS with speaker identity run on all conversations
- **NP-HMM**: the HMM-based nonparametric model which a single topic per turn. This model can be considered a Sticky HDP-HMM (Fox et al., 2008) with speaker identity.
- **NP-SITS**: the nonparametric version of SITS with speaker identity run on all conversations.

**Parameter Settings and Implementations** In our experiment, all parameters of TextTiling are the same as in (Hearst, 1997). For statistical models, Gibbs sampling with 10 randomly initialized chains is used. Initial hyperparameter values are sampled from  $U(0, 1)$  to favor sparsity; statistics are collected after 500 burn-in iterations with a lag of 25 iterations over a total of 5000 iterations; and slice sampling (Neal, 2003) optimizes hyperparameters.

**Results and Analysis** Table 2 shows the performance of various models on the topic segmentation problem, using the ICSI corpus and the 2008 debates.

Consistent with previous results, probabilistic models outperform TextTiling. In addition, among the probabilistic models, the models that had access to speaker information consistently segment better than those lacking such information, supporting our assertion that there is benefit to modeling conversation as a *social* process. Furthermore, NP-SITS outperforms NP-HMM in both experiments, suggesting that using a distribution over topics to turns is better than using a single topic. This is consistent with parametric results reported in (Purver et al., 2006).

The contribution of speaker identity seems more valuable in the debate setting. Debates are characterized by strong rewards for setting the agenda; dodging a question or moving the debate toward an oppo-

nent’s weakness can be useful strategies (Boydston et al., 2011). In contrast, meetings (particularly low-stakes ICSI meetings) are characterized by pragmatic rather than strategic topic shifts. Second, agenda-setting roles are clearer in formal debates; a moderator is tasked with setting the agenda and ensuring the conversation does not wander too much.

The nonparametric model does best on the smaller debate dataset. We suspect that an evaluation that directly accessed the topic quality, either via prediction (Teh et al., 2006) or interpretability (Chang et al., 2009) would favor the nonparametric model more.

## 6 Evaluating Topic Shift Tendency

In this section, we focus on the ability of SITS to capture speaker-level attributes. Recall that SITS associates with each speaker a topic shift tendency  $\pi$  that represents the probability of asserting a new topic in the conversation. While topic segmentation is a well studied problem, there are no established quantitative measurements of an individual’s ability to control a conversation. To evaluate whether the tendency is capturing meaningful characteristics of speakers, we compare our inferred tendencies against insights from political science.

**2008 Elections** To obtain a posterior estimate of  $\pi$  (Figure 3) we create 10 chains with hyperparameters sampled from the uniform distribution  $U(0, 1)$  and averaged  $\pi$  over 10 chains (as described in Section 5). In these debates, Ifill is the moderator of the debate between Biden and Palin; Brokaw, Lehrer and Schieffer are the three moderators of three debates between Obama and McCain. Here “Question” denotes questions from audiences in “town hall” debate. The role of this “speaker” can be considered equivalent to the debate moderator.

The topic shift tendencies of moderators are much higher than for candidates. In the three debates between Obama and McCain, the moderators—Brokaw, Lehrer and Schieffer—have significantly higher scores than both candidates. This is a useful reality check, since in a debate the moderators are the ones asking questions and literally controlling the topical focus. Interestingly, in the vice-presidential debate, the score of moderator Ifill is only slightly higher than those of Palin and Biden; this is consistent with media commentary characterizing her as a

	Model	EMD	$P_k$			WindowDiff		
			$k = 5$	10	15	$k = 5$	10	15
ICSI Dataset	TextTiling	2.507	.289	.388	.451	.318	.477	.561
	P-NoSpeaker-S	1.949	.222	.283	.342	.269	.393	.485
	P-NoSpeaker-M	1.935	<b>.207</b>	.279	.335	<b>.253</b>	.371	.468
	P-SITS	<b>1.807</b>	.211	<b>.251</b>	.289	.256	<b>.363</b>	<b>.434</b>
	NP-HMM	2.189	.232	.257	.263	.267	.377	.444
	NP-SITS	2.126	.228	.253	<b>.259</b>	.262	.372	.440
Debates Dataset	TextTiling	2.821	.433	.548	.633	.534	.674	.760
	P-NoSpeaker-S	2.822	.426	.543	.653	.482	.650	.756
	P-NoSpeaker-M	2.712	.411	.522	.589	.479	.644	.745
	P-SITS	2.269	.380	.405	.402	.482	.625	.719
	NP-HMM	2.132	.362	.348	.323	.486	.629	.723
	NP-SITS	<b>1.813</b>	<b>.332</b>	<b>.269</b>	<b>.231</b>	<b>.470</b>	<b>.600</b>	<b>.692</b>

Table 2: Results on the topic segmentation task. Lower is better. The parameter  $k$  is the window size of the metrics  $P_k$  and WindowDiff chosen to replicate previous results.

weak moderator.<sup>7</sup> Similarly, the “Question” speaker had a relatively high variance, consistent with an amalgamation of many distinct speakers.

These topic shift tendencies suggest that all candidates manage to succeed at some points in setting and controlling the debate topics. Our model gives Obama a slightly higher score than McCain, consistent with social science claims (Boydston et al., 2011) that Obama had the lead in setting the agenda over McCain. Table 4 shows of SITS-detected topic shifts.

**Crossfire** Crossfire, unlike the debates, has many speakers. This allows us to examine more closely what we can learn about speakers’ topic shift tendency. We verified that SITS can segment topics, and assuming that changing the topic is useful for a speaker, how can we characterize who does so effectively? We examine the relationship between topic shift tendency, social roles, and political ideology.

To focus on frequent speakers, we filter out speakers with fewer than 30 turns. Most speakers have relatively small  $\pi$ , with the mode around 0.3. There are, however, speakers with very high topic shift tendencies. Table 5 shows the speakers having the highest values according to SITS.

We find that there are three general patterns for who influences the course of a conversation in *Crossfire*. First, there are structural “speakers” the show uses to frame and propose new topics. These are

<sup>7</sup> <http://harpers.org/archive/2008/10/hbc-90003659>

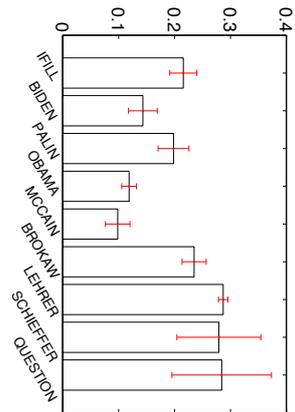


Table 3: Topic shift tendency  $\pi$  of speakers in the 2008 Presidential Election Debates (larger means greater tendency)

audience questions, news clips (e.g. many of Gore’s and Bush’s turns from 2000), and voice overs. That SITS is able to recover these is reassuring. Second, the stable of regular hosts receives high topic shift tendencies, which is reasonable given their experience with the format and ostensible moderation roles (in practice they also stoke lively discussion).

The remaining class is more interesting. The remaining non-hosts with high topic shift tendency are relative moderates on the political spectrum:

- John Kasich, one of few Republicans to support the assault weapons ban and now governor of Ohio, a swing state
- Christine Todd Whitman, former Republican governor of New Jersey, a very Democratic state
- John McCain, who before 2008 was known as a “maverick” for working with Democrats (e.g. Russ Feingold)

This suggests that, despite *Crossfire*’s tendency to create highly partisan debates, those who are able to work across the political spectrum may best be able to influence the topic under discussion in highly polarized contexts. Table 4 shows detected topic shifts from these speakers; two of these examples (McCain and Whitman) show disagreement of Republicans with President Bush. In the other, Kasich is defending a Republican plan (school vouchers) popular with traditional Democratic constituencies.

## 7 Related and Future Work

In the realm of statistical models, a number of techniques incorporate social connections and identity to explain content in social networks (Chang and Blei,

	Previous turn	Turn detected as shifting topic
Debates Dataset	PALIN: Your question to him was whether he supported <i>gay marriage</i> and my answer is the same as his and it is that I do not.	IFILL: Wonderful. You agree. On that note, let’s move to <i>foreign policy</i> . You both have sons who are in Iraq or on their way to Iraq. You, Governor Palin, have said that you would like to see a real clear plan for an exit strategy. [...]
	MCCAIN: I think that Joe Biden is qualified in many respects. ...	SCHIEFFER: [...] Let’s talk about <i>energy</i> and <i>climate control</i> . Every president since Nixon has said what both of you [...]
	IFILL: So, Governor, as vice president, there’s nothing that you have promised [...] that you wouldn’t take off the table because of this <i>financial crisis</i> we’re in?	BIDEN: Again, let me—let’s talk about those tax breaks. [Obama] voted for an energy bill because, for the first time, it had real support for alternative energy. [...] on eliminating the tax breaks for the oil companies, Barack Obama voted to eliminate them. [...]
Cross fire Dataset	PRESS: But what do you say, governor, to Governor Bush and [...] your party who would let politicians and not medical scientists decide what <i>drugs</i> are distributed [...]	WHITMAN: Well I disagree with them on this particular issues [...] that’s important to me that George Bush stands for education of our children [...] I care about tax policy, I care about the environment. I care about all the issues where he has a proven record in Texas [...]
	WEXLER: [...] They need a Medicare prescription drug plan [...] Talk about schools, [...] Al Gore has got a real plan. George Bush offers us vouchers. Talk about the environment. [...] Al Gore is right on in terms of the majority of Americans, but George Bush [...]	KASICH: [...] I want to talk about choice. [...] George Bush believes that, if schools fail, parents ought to have a right to get their kids out of those schools and give them a chance and an opportunity for success. Gore says “no way” [...] Social Security. George Bush says [...] direct it the way federal employees do [...] Al Gore says “No way” [...] That’s real choice. That’s real bottom-up, not a bureaucratic approach, the way we run this country.
	PRESS: Senator, Senator Breaux mentioned that it’s President Bush’s aim to start on education [...] [McCain] [...] said he was going to do introduce the legislation the first day of the first week of the new administration. [...]	MCCAIN: After one of closest elections in our nation’s history, there is one thing the American people are unanimous about They want their government back. We can do that by ridding politics of large, unregulated contributions that give special interests a seat at the table while average Americans are stuck in the back of the room.

Table 4: Example of turns designated as a topic shift by SITS. Turns were chosen with speakers to give examples of those with high topic shift tendency  $\pi$ .

Rank	Speaker	$\pi$	Rank	Speaker	$\pi$
1	Announcer	.884	10	Kasich	.570
2	Male	.876	11	Carville <sup>†</sup>	.550
3	Question	.755	12	Carlson <sup>†</sup>	.550
4	G. W. Bush <sup>‡</sup>	.751	13	Begala <sup>†</sup>	.545
5	Press <sup>†</sup>	.651	14	Whitman	.533
6	Female	.650	15	McAuliffe	.529
7	Gore <sup>‡</sup>	.650	16	Matalin <sup>†</sup>	.527
8	Narrator	.642	17	McCain	.524
9	Novak <sup>†</sup>	.587	18	Fleischer	.522

Table 5: Top speakers by topic shift tendencies. We mark hosts (<sup>†</sup>) and “speakers” who often (but not always) appeared in clips (<sup>‡</sup>). Apart from those groups, speakers with the highest tendency were political moderates.

2009) and scientific corpora (Rosen-Zvi et al., 2004). However, these models ignore the temporal evolution of content, treating documents as static.

Models that do investigate the evolution of topics over time typically ignore the identify of the speaker. For example: models having sticky topics over n-grams (Johnson, 2010), sticky HDP-HMM (Fox et al., 2008); models that are an amalgam of sequential models and topic models (Griffiths et al., 2005; Wal-

lach, 2006; Gruber et al., 2007; Ahmed and Xing, 2008; Boyd-Graber and Blei, 2008; Du et al., 2010); or explicit models of time or other relevant features as a distinct latent variable (Wang and McCallum, 2006; Eisenstein et al., 2010).

In contrast, SITS jointly models topic and individuals’ tendency to control a conversation. Not only does SITS outperform other models using standard computational linguistics baselines, but it also proposes intriguing hypotheses for social scientists.

Associating each speaker with a scalar that models their tendency to change the topic does improve performance on standard tasks, but it’s inadequate to fully describe an individual. Modeling individuals’ perspective (Paul and Girju, 2010), “side” (Thomas et al., 2006), or personal preferences for topics (Grimmer, 2009) would enrich the model and better illuminate the interaction of influence and topic.

Statistical analysis of political discourse can help discover patterns that political scientists, who often work via a “close reading,” might otherwise miss. We plan to work with social scientists to validate our implicit hypothesis that our topic shift tendency correlates well with intuitive measures of “influence.”

## Acknowledgments

This research was funded in part by the Army Research Laboratory through ARL Cooperative Agreement W911NF-09-2-0072 and by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory. Jordan Boyd-Graber and Philip Resnik are also supported by US National Science Foundation Grant NSF grant #1018625. Any opinions, findings, conclusions, or recommendations expressed are the authors' and do not necessarily reflect those of the sponsors.

## References

- [Abbott et al., 2011] Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., and King, J. (2011). How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pages 2–11.
- [Ahmed and Xing, 2008] Ahmed, A. and Xing, E. P. (2008). Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering. In *SDM*, pages 219–230.
- [Beeferman et al., 1999] Beeferman, D., Berger, A., and Lafferty, J. (1999). Statistical models for text segmentation. *Mach. Learn.*, 34:177–210.
- [Blei and Lafferty, 2009] Blei, D. M. and Lafferty, J. (2009). *Text Mining: Theory and Applications*, chapter Topic Models. Taylor and Francis, London.
- [Boyd-Graber and Blei, 2008] Boyd-Graber, J. and Blei, D. M. (2008). Syntactic topic models. In *Proceedings of Advances in Neural Information Processing Systems*.
- [Boydston et al., 2011] Boydston, A. E., Phillips, C., and Glazier, R. A. (2011). It's the economy again, stupid: Agenda control in the 2008 presidential debates. *Forthcoming*.
- [Chang and Blei, 2009] Chang, J. and Blei, D. M. (2009). Relational topic models for document networks. In *Proceedings of Artificial Intelligence and Statistics*.
- [Chang et al., 2009] Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*.
- [Du et al., 2010] Du, L., Buntine, W., and Jin, H. (2010). Sequential latent dirichlet allocation: Discover underlying topic structures within a document. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 148–157.
- [Ehlen et al., 2007] Ehlen, P., Purver, M., and Niekrasz, J. (2007). A meeting browser that learns. In *In: Proceedings of the AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants*.
- [Eisenstein and Barzilay, 2008] Eisenstein, J. and Barzilay, R. (2008). Bayesian unsupervised topic segmentation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Proceedings of Empirical Methods in Natural Language Processing.
- [Eisenstein et al., 2010] Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2010). A latent variable model for geographic lexical variation. In *EMNLP'10*, pages 1277–1287.
- [Ferguson, 1973] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.
- [Fox et al., 2008] Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2008). An hdp-hmm for systems with state persistence. In *Proceedings of International Conference of Machine Learning*.
- [Galley et al., 2003] Galley, M., McKeown, K., Fosler-Lussier, E., and Jing, H. (2003). Discourse segmentation of multi-party conversation. In *Proceedings of the Association for Computational Linguistics*.
- [Georgescul et al., 2006] Georgescul, M., Clark, A., and Armstrong, S. (2006). Word distributions for thematic segmentation in a support vector machine approach. In *Conference on Computational Natural Language Learning*.
- [Gerrish and Blei, 2010] Gerrish, S. and Blei, D. M. (2010). A language-based approach to measuring scholarly impact. In *Proceedings of International Conference of Machine Learning*.
- [Griffiths et al., 2005] Griffiths, T. L., Steyvers, M., Blei, D. M., and Tenenbaum, J. B. (2005). Integrating topics and syntax. In *Proceedings of Advances in Neural Information Processing Systems*.
- [Grimmer, 2009] Grimmer, J. (2009). A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, 18:1–35.
- [Gruber et al., 2007] Gruber, A., Rosen-Zvi, M., and Weiss, Y. (2007). Hidden topic Markov models. In *Artificial Intelligence and Statistics*.
- [Hawes et al., 2009] Hawes, T., Lin, J., and Resnik, P. (2009). Elements of a computational model for multi-party discourse: The turn-taking behavior of Supreme Court justices. *Journal of the American Society for Information Science and Technology*, 60(8):1607–1615.
- [Hearst, 1997] Hearst, M. A. (1997). TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.

- [Hsueh et al., 2006] Hsueh, P.-y., Moore, J. D., and Renals, S. (2006). Automatic segmentation of multiparty dialogue. In *Proceedings of the European Chapter of the Association for Computational Linguistics*.
- [Ireland et al., 2011] Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., and Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, 22(1):39–44.
- [Janin et al., 2003] Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., and Wooters, C. (2003). The ICSI meeting corpus. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*.
- [Johnson, 2010] Johnson, M. (2010). PCFGs, topic models, adaptor grammars and learning topical collocations and the structure of proper names. In *Proceedings of the Association for Computational Linguistics*.
- [Morris and Hirst, 1991] Morris, J. and Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17:21–48.
- [Müller and Quintana, 2004] Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19(1):95–110.
- [Murray et al., 2005] Murray, G., Renals, S., and Carletta, J. (2005). Extractive summarization of meeting recordings. In *European Conference on Speech Communication and Technology*.
- [Neal, 2000] Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.
- [Neal, 2003] Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31:705–767.
- [Olney and Cai, 2005] Olney, A. and Cai, Z. (2005). An orthonormal basis for topic segmentation in tutorial dialogue. In *Proceedings of the Human Language Technology Conference*.
- [Paul and Girju, 2010] Paul, M. and Girju, R. (2010). A two-dimensional topic-aspect model for discovering multi-faceted topics. In *Association for the Advancement of Artificial Intelligence*.
- [Pele and Werman, 2009] Pele, O. and Werman, M. (2009). Fast and robust earth mover’s distances. In *International Conference on Computer Vision*.
- [Pevzner and Hearst, 2002] Pevzner, L. and Hearst, M. A. (2002). A critique and improvement of an evaluation metric for text segmentation. *Computational Linguistics*, 28.
- [Purver, 2011] Purver, M. (2011). Topic segmentation. In Tur, G. and de Mori, R., editors, *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, pages 291–317. Wiley.
- [Purver et al., 2006] Purver, M., Körding, K., Griffiths, T. L., and Tenenbaum, J. (2006). Unsupervised topic modelling for multi-party spoken discourse. In *Proceedings of the Association for Computational Linguistics*.
- [Resnik and Hardisty, 2010] Resnik, P. and Hardisty, E. (2010). Gibbs sampling for the uninitiated. Technical Report UMIACS-TR-2010-04, University of Maryland. <http://www.lib.umd.edu/drum/handle/1903/10058>.
- [Reynar, 1998] Reynar, J. C. (1998). *Topic Segmentation: Algorithms and Applications*. PhD thesis, University of Pennsylvania.
- [Rosen-Zvi et al., 2004] Rosen-Zvi, M., Griffiths, T. L., Steyvers, M., and Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of Uncertainty in Artificial Intelligence*.
- [Rubner et al., 2000] Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121.
- [Teh et al., 2006] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581.
- [Thomas et al., 2006] Thomas, M., Pang, B., and Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of Empirical Methods in Natural Language Processing*.
- [Tur et al., 2010] Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tür, D., Dowding, J., Favre, B., Fernández, R., Frampton, M., Frandsen, M., Frederickson, C., Gra-ciarena, M., Kintzing, D., Leveque, K., Mason, S., Niekrasz, J., Purver, M., Riedhammer, K., Shriberg, E., Tien, J., Vergyri, D., and Yang, F. (2010). The CALO meeting assistant system. *Trans. Audio, Speech and Lang. Proc.*, 18:1601–1611.
- [Wallach, 2006] Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In *Proceedings of International Conference of Machine Learning*.
- [Wallach, 2008] Wallach, H. M. (2008). *Structured Topic Models for Language*. PhD thesis, University of Cambridge.
- [Wang et al., 2008] Wang, C., Blei, D. M., and Heckerman, D. (2008). Continuous time dynamic topic models. In *Proceedings of Uncertainty in Artificial Intelligence*.
- [Wang and McCallum, 2006] Wang, X. and McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. In *Knowledge Discovery and Data Mining*, Knowledge Discovery and Data Mining.