Viet-An Nguyen, **Jordan Boyd-Graber**, Philip Resnik, Deborah Cai, Jennifer Midberry, and Yuanxin Wang.
Modeling Topic Control to Detect Influence in Conversations using Nonparametric Topic Models.
Machine Learning, 2014, 48 pages.

Links:

- Journal [http://dx.doi.org/10.1007/s10994-013-5417-9]
- Code [http://github.com/vietansegan/sits]
- Data [http://www.cs.umd.edu/~vietan/topicshift/crossfire.zip]

1

Modeling Topic Control to Detect Influence in Conversations using Nonparametric Topic Models

Viet-An Nguyen · Jordan Boyd-Graber ·
Philip Resnik · Deborah A. Cai ·
Jennifer E. Midberry · Yuanxin Wang

Abstract Identifying influential speakers in multi-party conversations has been the focus of research in communication, sociology, and psychology for decades. It has been long acknowledged qualitatively that controlling the topic of a conversation is a sign of influence. To capture who introduces new topics in conversations, we introduce SITS—Speaker Identity for Topic Segmentation—a nonparametric hierarchical Bayesian model that is capable of discovering (1) the topics used in a set of conversations, (2) how these topics are shared across conversations, (3) when these topics change during conversations, and (4) a speaker-specific measure of "topic control". We validate the model via evaluations using multiple datasets, including work meetings, online discussions, and political debates. Experimental results confirm the effectiveness of SITS in both intrinsic and extrinsic evaluations.

V.-A. Nguyen
Department of Computer Science, University of Maryland, College Park
E-mail: vietan@cs.umd.edu

J. Boyd-Graber
iSchool and UMIACS, University of Maryland, College Park
E-mail: jbg@umiacs.umd.edu

P. Resnik
Department of Linguistics and UMIACS, University of Maryland, College Park
E-mail: resnik@umd.edu

D. A. Cai
School of Media and Communication, Temple University
E-mail: debcai@temple.edu

J. E. Midberry
School of Media and Communication, Temple University
E-mail: jmidberry@gmail.com

Y. Wang
School of Media and Communication, Temple University
E-mail: yuanxin@gmail.com

1 Influencing Conversations by Controlling the Topic

Conversation, interactive discussion between two or more people, is one of the most essential and common forms of communication in our daily lives.[1] One of the many functions of conversations is influence: having an effect on the belief, opinions or intentions of other conversational participants. Using multi-party conversations to study and identify influencers, the people who influence others, has been the focus of researchers in communication, sociology, and psychology (Katz and Lazarsfeld, 1955; Brooke and Ng, 1986; Weimann, 1994), who have long acknowledged that there is a correlation between the conversational behaviors of a participant and how influential he or she is perceived to be by others (Reid and Ng, 2000).

In an early study on this topic, Bales (1970) argues "*To take up time speaking in a small group is to exercise power over the other members for at least the duration of the time taken, regardless of the content.*" This statement asserts that *structural patterns* such as speaking time and activeness of participation are good indicators of power and influence in a conversation. Participants who talk most during a conversation are often perceived as having more influence (Sorrentino and Boutiller, 1972; Regula and Julian, 1973; Daley et al, 1977; Ng et al, 1993), more leadership ability (Stang, 1973; Sorrentino and Boutiller, 1972), more dominance (Palmer, 1989; Mast, 2002) and more control of the conversation (Palmer, 1989). Recent work using computational methods also confirms that structural features such as number of turns and turn length are among the most discriminative features to classify whether a participant is influential or not (Rienks et al, 2006; Biran et al, 2012).

However, it is wrong to take Bales's claim too far; the person who speaks loudest and longest is not always the most powerful. In addition to structural patterns, the characteristics of language used also play an important role in establishing influence and controlling the conversation (Ng and Bradac, 1993). For example, particular linguistic choices such as message clarity, powerful and powerless language (Burrel and Koper, 1998), and language intensity (Hamilton and Hunter, 1998) in a message can increase influence. More recently, Huffaker (2010) showed that linguistic diversity expressed by lexical complexity and vocabulary richness has a strong relationship with leadership in online communities. To build a classifier to detect influencers in written online conversations, Biran et al (2012) also propose to use a set of content-based features to capture various participants' conversational behaviors, including persuasion and agreement/disagreement.

Among many studied behaviors, topic control and management is considered one of the most effective ways to control the conversation (Planalp and Tracy,

---

[1] This paper significantly revises and extends the work described in (Nguyen et al, 2012).

1980). Palmer (1989) shows that the less related a participants' utterances are to the immediate topic, the more dominant they are, and then argues, "*the ability to change topical focus, especially given strong cultural and social pressure to be relevant, means having enough interpersonal power to take charge of the agenda.*" Recent work by Rienks et al (2006) also shows that topic change, among other structural patterns discussed above, is the most robust feature in detecting influencers in small group meetings.

In this article, we introduce a new computational model capturing the role of topic control in participants' influence of conversations. Speaker Identity for Topic Segmentation (SITS), a hierarchical Bayesian nonparametric model, uses an unsupervised statistical approach which requires few resources and can be used in many domains without extensive training and annotation. More important, SITS incorporates an explicit model of speaker behavior by characterizing quantitatively individuals' tendency to exercise control over the topic of conversation (Section 3). By focusing on topic changes in conversations, we go beyond previous work on influencers in two ways:

- First, while structural statistics such as number of turns, turn length, speaking time etc are relatively easy to extract from a conversation, defining and detecting topic changes is less well understood. Topic, by itself, is a complex concept (Blei et al, 2003; Kellermann, 2004). In addition, despite the large number of techniques proposed trying to divide a document into smaller, topically coherent segments (Purver, 2011), topic segmentation is still an open research problem. Most previous computational methods for topic discovery and topic segmentation focus on content, ignoring the speaker identities. We show that we can capture conversational phenomena and influence better by explicitly modeling behaviors of participants.
- Second, the conversation is often controlled explicitly, to some extent, by a subset of participants. For example, in political debates questions come from the moderator(s), and candidates typically have a fixed time to respond. These imposed aspects of conversational structure decrease the value of more easily extracted structural statistics for a variety of conversation types; observe, for example, that similar properties of the conversation can also be observed when looking at hosts and guests in televised political discussion shows such as CNN's *Crossfire*.

Applying SITS on real-world conversations (Section 4), we show that this modeling approach is not only more effective than previous methods on traditional topic segmentation (Section 5), but also more intuitive in that it is able to capture an important behavior of individual speakers during conversations (Section 6). We then show that using SITS to model topic control improves influencer detection (Section 7). Taking quantitative and qualitative analysis together, the pattern of results suggests that our approach holds significant promise for further development; we discuss directions for future work in Section 8.

2 What is an Influencer?

2.1 Influencer Definition

In most research on persuasion and power, an influencer attempts to gain compliance from others or uses tactics to shape the opinions, attitudes, or behaviors of others (Scheer and Stern, 1992; Schlenker et al, 1976). In research on social media, such as blogs and Twitter, measurements such as the number of followers or readers serve as a proxy for influence (Alarcon-del Amo et al, 2011; Booth and Matic, 2011; Trammell and Keshelashvili, 2005). Others have studied what influencers say; Drake and Moberg (1986) demonstrated that linguistic influence differs from attempts to influence that rely on power and exchange relationships. In interactions with targets, influencers may rely more on linguistic frames and language than on resources offered, which is proposed as the requirement for influence by exchange theorists (Blau, 1964; Foa and Foa, 1972; Emerson, 1981).

We define an influencer as someone who has persuasive ability over where an interaction is headed, what topics are covered, and what positions are espoused within that interaction. In the same way that persuasion shapes, reinforces, or changes attitudes or beliefs, an influencer shapes, reinforces, or changes the direction of the interaction. An influencer within an interaction is someone who may introduce new ideas or arguments into the conversation that others pick up on and discuss (shapes new directions through topic shift), may express arguments about an existing topic that others agree to and further in the discussion (i.e., reinforces the direction), or may provide counter-arguments that others agree to and perpetuate, thereby redirecting where the topic of conversation is headed (i.e., changes the direction of the conversation).

2.2 Data Scope and Characteristics

We are interested in influence in turn-taking, multiparty discussion. This is a broad category including political debates, business meetings, online chats, discussions, conference panels, and many TV or radio talk shows. 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.[2] In each turn $t \in [1, T_c]$, a speaker $a_{c,t}$ utters $N_{c,t}$ words $\boldsymbol{w}_{c,t} = \{w_{c,t,n} \mid n \in [1, N_{c,t}]\}$. Each word is from a vocabulary of size V, and there are M distinct speakers.

3 Modeling Topic Shift

In this section, we describe SITS, a hierarchical nonparametric Bayesian model for topic segmentation that takes into consideration speaker identities, allowing us to characterize speakers' topic control behavior over the course of the

---

[2] Note the distinction from phonetic utterances, which by definition are bounded by silence.

discussion (Nguyen et al, 2012). We begin with an overview of the topic segmentation problem and some related work. We then highlight the differences between SITS and previous approaches and describe the generative process and the inference technique we use to estimate the model.

3.1 Topic Segmentation and Modeling Approaches

Whether in an informal situation or in more formal settings such as a political debate of 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 building conversational assistants (Tur et al, 2010) and developing tools that summarize (Murray et al, 2005) and display (Ehlen et al, 2007) conversational data. Understanding when and how the topics change also helps us study human conversational behaviors such as individuals' agendas (Boydstun et al, 2013), patterns of agreement and disagreement (Hawes et al, 2009; Abbott et al, 2011), relationships among conversational participants (Ireland et al, 2011), and dominance and influence among participants (Palmer, 1989; Rienks et al, 2006).

One of the most natural ways to capture conversational structure is topic segmentation — the task of "automatically dividing single long recordings or transcripts into shorter, topically coherent segments" (Purver, 2011). There are broadly two basic approaches previous work has used to tackle this problem. The first approach focuses on identifying *discourse markers* which distinguish topical boundaries in the conversations. There are certain cue phrases such as *well, now, that reminds me*, etc. that explicitly indicate the end of one topic or the beginning of another (Hirschberg and Litman, 1993; Passonneau and Litman, 1997). These markers can also serve as features for a discriminative classifier (Galley et al, 2003) or observed variables in generative model (Dowman et al, 2008). However, in practice the discourse markers that are most indicative of topic change often depend heavily on the domain of the data (Purver, 2011). This drawback makes methods solely relying on these markers difficult to adapt to new domains or settings.

Our method follows the second general approach, which relies on the insight that topical segments evince *lexical cohesion* (Halliday and Hasan, 1976). Intuitively, words within a segment will look more like their neighbors than like words in other segments. This has been a key idea in previous work. Morris and Hirst (1991) try to determine the structure of text by finding "lexical chains" which consists of units of text that are about the same thing. The often used text segmentation algorithm TextTiling (Hearst, 1997) exploits this insight to compute the lexical similarity between adjacent sentences. More recent improvements to this approach include using different lexical similarity metrics like LSA (Choi et al, 2001; Olney and Cai, 2005) and improving feature extraction for supervised methods (Hsueh et al, 2006). It also inspires unsupervised models using bags of words (Purver et al, 2006),

language models (Eisenstein and Barzilay, 2008), and shared structure across documents (Chen et al, 2009).

We also use lexical cohesion using a probabilistic topic modeling method (Blei et al, 2003; Blei, 2012). The approach we take is unsupervised, so it requires few resources and is applicable in many domains without extensive training. Following the literature on topic modeling, we define each topic as a multinomial distribution over the vocabulary. Like previous generative models proposed for topic segmentation (Purver et al, 2006), each turn is considered a bag of words generated from an admixture of topics and topics are shared across different turns within a conversation or across different conversations.[3] In addition, we take a Bayesian nonparametric approach (Müller and Quintana, 2004) to allow the number of topics to be unbounded, in order to better represent the observed data.

The settings described above are still consistent with those in popular topic models such as latent Dirichlet allocation (LDA) (Blei et al, 2003) or hierarchical Dirichlet processes (HDP) (Teh et al, 2006), in which turns in a conversation are considered independent. In practice, however, this is not the case. Obviously the topics of a turn at time t are highly correlated with those of the turn at $t+1$. To address this issue, there have been several recent attempts trying to capture the temporal dynamics within a document. Du et al (2010) propose Sequential LDA to study how topics within a document evolve over its structure. It uses the nested two-parameter Poisson Dirichlet process (PDP) to model the progressive dependency between consecutive part of a document, which can capture the continuity of topical flow in a document nicely but does not capture the topic change explicitly. Fox et al (2008) proposed Sticky HDP-HMM, which is an extension of HDP-HMM (Teh et al, 2006) for the problem of speaker diarization involving segmenting an audio recording into intervals associated with individual speakers. Applying to the conversational setting, Sticky HDP-HMM associates each turn with a single topic; this is a strong assumption since people tend to talk about more than one thing in a turn, especially in political debates. We will, however, use it as one of the baselines in our topic segmentation experiment (Section 5). A related problem is to discover how topics themselves change over time (Blei and Lafferty, 2006; Wang et al, 2008; Ren et al, 2008; Ahmed and Xing, 2008, 2010), e.g., documents that talk about "physics" in 1900 will use very different terms than "physics" in 2000. These models assume documents are much longer and that topics evolve much more slowly than in a conversation.

Moreover, many of the above methods do not explicitly model the changes of the topics within a document or conversation. In order to address this, we endow each turn with a binary latent variable $l_{c,t}$, called the *topic shift indicator* (Purver et al, 2006). This latent variable signifies whether in this turn the speaker changed the topic of the conversation. In addition, to capture

---

[3] The "bag of words" treatment of linguistic utterances is widely used, but of course a gross simplification. In other research, we have investigated nonparametric models capturing arbitrary-length phrases (Hardisty et al, 2010) and syntactic topic models (Boyd-Graber and Blei, 2008); integrating linguistically richer models with SITS is a topic for future work.

the topic-controlling behavior of the speakers across different conversations, we further associate each speaker m with a latent *topic shift tendency* denoted by π_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. In the remainder of this section, we will describe the model in more detail together with the inference technique we use.
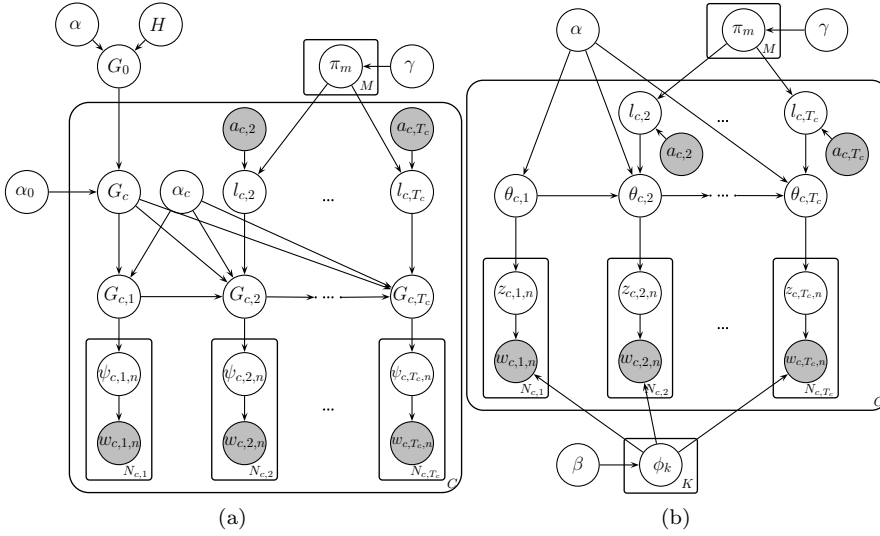
3.2 Generative Process of SITS

Fig. 1: Plate diagrams of our proposed models: (a) nonparametric SITS; (b) parametric SITS. Nodes represent random variables (shaded nodes are observed); lines are probabilistic dependencies. Plates represent repetition. The innermost plates are turns, grouped in conversations.

SITS is 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 of size V; H is a finite Dirichlet distribution with symmetric prior λ. Unlike the HDP, where every document (here, every turn) independently 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.

Generative process: The formal generative process is as follows:

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

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 π_m allows speaker *identity* to influence the evolution of topics.

Intuitively, SITS generates a conversation as follows: At the beginning of a conversation c, the first speaker $a_{c,1}$ draws a distribution over topics $G_{c,1}$ from the base distribution, and uses that topic distribution to draw a topic $\psi_{c,1,n}$ for each token $w_{c,1,n}$. Subsequently, at turn t, speaker $a_{c,t}$ will first flip a speaker-specific biased coin $\pi_{a_{c,t}}$ to decide whether $a_{c,t}$ will change the topic of the conversation. If the coin comes up tails ($l_{c,t} = 0$), $a_{c,t}$ will not change the conversation topic and uses the previous turn's topic distribution $G_{c,t-1}$ to generate turn t's tokens. If, on the other hand, the coin comes up heads ($l_{c,t} = 1$), $a_{c,t}$ will change the topic by drawing a new topic distribution $G_{c,t}$ from the conversation-specific collection of topics $DP(\alpha_c, G_c)$.

Segmentation Notation: To make notation more concrete and to connect our model with topic segmentation, we introduce the notion of *segments* in a conversation. A segment s of conversation c is a sequence of turns $[\tau, \tau']$ such that

$$\begin{cases} l_{c,\tau} = l_{c,\tau'+1} = 1 \\ l_{c,t} = 0, \forall t \in [\tau+1, \tau'] \end{cases}$$

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) $\{\psi_{c,t,n} \mid n \in [1, N_{c,t}]\}$ that generate words in turn t and the topics $\{\psi_{c,t-1,n'} \mid n' \in [1, N_{c,t-1}]\}$ that generate words in turn $t-1$ come from the same distribution. Thus, all topics used in a segment s are drawn from a single segment-specific probability measure $G_{c,s}$,

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

A visual illustration of these notations can be found in Figure 2. 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.

Topic shift indicators $\qquad$ $l_{c,\tau} = 1 \quad l_{c,\tau+1} = 0 \quad \cdots\cdots \quad l_{c,\tau'} = 0 \quad l_{c,\tau'+1} = 1$

Turns $\quad$ ..... $\boxed{\tau - 1} \quad \boxed{\tau} \quad \boxed{\tau + 1} \quad ..... \quad \boxed{\tau'} \quad \boxed{\tau' + 1} \quad .....$

Speakers $\quad a_{c,\tau-1} \quad a_{c,\tau} \quad a_{c,\tau+1} \quad ..... \quad a_{c,\tau'} \quad a_{c,\tau'+1}$

Turn probability measures $\quad G_{c,\tau-1} \quad G_{c,\tau} \quad G_{c,\tau+1} \quad ..... \quad G_{c,\tau'} \quad G_{c,\tau'+1}$

Segment probability measures $\quad \equiv G_{c,s-1} \qquad\qquad \equiv G_{c,s} \qquad\qquad \equiv G_{c,s+1}$
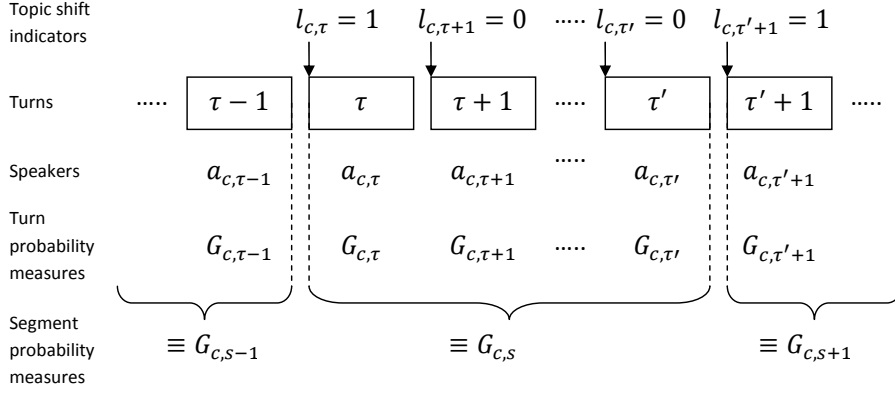
Fig. 2: Diagram of notation for topic shift indicators and conversation segments: Each turn is associated with a latent binary variable *topic shift indicator l* specifying whether the topic of the turn is shifted. In this example, topic shifts occur in turns τ and $\tau' + 1$. As a result, the topic shift indicators of turn τ and $\tau' + 1$ are equal to 1 (i.e. $l_{c,\tau} = l_{c,\tau'+1} = 1$) and the topic shift indicators of all turns in between are 0 (i.e. $l_{c,t} = 0, \forall t \in [\tau + 1, \tau']$). Turns $[\tau, \tau']$ form a segment s in which all topic distributions $G_{c,\tau}, G_{c,\tau+1}, \cdots, G_{c,\tau'}$ are the same and are denoted collectively as $G_{c,s}$.

3.3 Inference for SITS

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 in our Gibbs sampler consists of the latent variables for topic indices assigned to all tokens $\boldsymbol{z} = \{z_{c,t,n}\}$ and topic shifts assigned to turns $\boldsymbol{l} = \{l_{c,t}\}$. We marginalize over all other latent variables. For each iteration of the sampling process, we loop over each turn in each conversation. For a given turn t in conversation c, we first sample the topic shift indicator variable $l_{c,t}$ (Section 3.3.2) and then sample the topic assignment $z_{c,t,n}$ for each token in the turn (Section 3.3.1). Here, we only present the conditional sampling equations; for details on how these are derived, see the Appendix A.

3.3.1 Sampling Topic Assignments

In Bayesian nonparametrics, the Chinese restaurant process (CRP) metaphor is often used to explain the clustering effect of the Dirichlet process (Ferguson, 1973). The CRP is an exchangeable distribution over partitions of integers, which facilitates Gibbs sampling (Neal, 2000) (as we will see in Equation 2). When used in topic models, each Chinese restaurant consists of infinite number of tables, each of which corresponds to a topic. Customers, each of which
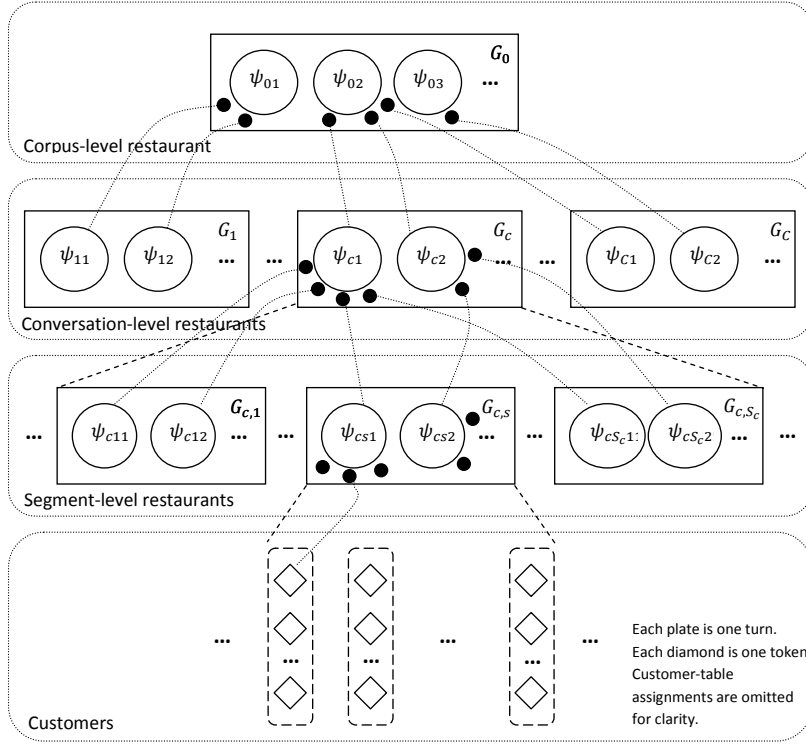
Fig. 3: Illustration of topic assignments in our inference algorithm. Each solid rectangle represents a restaurant (i.e., a topic distribution) and each circle represents a table (i.e., a topic). To assign token n of turn t in conversation c to a table $z_{c,t,n}$ in the corpus-level restaurant, we need to sample a path assigning the token to a segment-level table, the segment-level table to a conversation-level table and the conversation-level table to a globally shared corpus-level table.

corresponds to a token, are assigned to tables and if two tokens are assigned to the same table: they share the same topic.

The CRP has a "rich get richer" property, which means that tables with many customers will attract yet more customers—a new customer will sit at an existing table with probability proportional to the number of customers currently at the table. The CRP has no limit on the number of tables; when a customer needs to be seated, there is always a probability—proportional to the Dirichlet parameter α—that it will be seated at a new table. When a new table is formed, it is assigned a "dish"; this is a draw from the Dirichlet process's base distribution. In a topic model, this atom associated with a new table is a multinomial distribution over word types. In a standard, non-hierarchical CRP, this multinomial distribution comes from a Dirichlet distribution.

But it doesn't have to—hierarchical nonparametric models extend the metaphor further by introducing a hierarchy of restaurants (Teh et al, 2006; Teh, 2006), where the base distribution of one restaurant can be another restaurant. This is where things can get tricky. Instead of having a seating assignment, a customer now has a seating *path* and is potentially responsible for spawning new tables in every restaurant. In SITS there are restaurants for the current segment, the conversation, and the entire corpus, as shown in Figure 3.

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-level table, each segment-level table to a conversation-level table and each conversation-level table to a shared dish. Before describing the sampling equations, we introduce notation denoting the counts:

– $N_{c,s,k}$: number of tokens in segment s in conversation c assigned to dish k
– $N_{c,k}$: number of segment-level tables in conversations c assigned to dish k
– N_k: number of conversation-level tables assigned to dish k

Note that we use k to index the global topics shared across the corpus, each of which corresponds to a dish in the corpus-level restaurant. In general, computing the exact values of these counts makes bookkeeping rather complicated. Since there might be multiple tables at a lower-level restaurant assigned to the same table at the higher-level restaurant, to compute the correct counts, we need to sum the number of customers over all these tables. For example, in Figure 3, since both $\psi_{c,1}$ and $\psi_{c,2}$ are assigned to $\psi_{0,2}$ (i.e., $k = 2$), to compute $N_{c,k}$ we have to sum over the number of customers currently assigned to $\psi_{c,1}$ and $\psi_{c,2}$ (which are 4 and 2 respectively in this example).

To mitigate this problem of bookkeeping and to speed up the sampling process, we use the *minimal path assumption* (Cowans, 2006; Wallach, 2008) to generate the path assignments.[4] Under the minimal path assumption, a new table in a restaurant is created only when there is no table already serving the dish. In other words in a restaurant, there is at most one table serving a given dish. A more detailed example of the minimal path assumption is illustrated in Figure 4. Using this assumption, in the example shown in Figure 3, $\psi_{c,1}$ and $\psi_{c,2}$ will be merged together since they are both assigned to $\psi_{0,2}$.

Now that we have introduced our notations, the conditional distribution for $z_{c,t,n}$ is

$P(z_{c,t,n} \mid w_{c,t,n}, \boldsymbol{z}^{-c,t,n}, \boldsymbol{w}^{-c,t,n}, \boldsymbol{l}, *) \propto$

$$P(z_{c,t,n} \mid \boldsymbol{z}^{-c,t,n})P(w_{c,t,n} \mid z_{c,t,n}, \boldsymbol{w}^{-c,t,n}, \boldsymbol{l}, *) \tag{2}$$

---

[4] 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. Another, more complicated, sampling method is to sample the counts $N_{c,k}$ and N_k according to their corresponding Antoniak distributions (Antoniak, 1974), similar to the direct assignment sampling method described in Teh et al (2006).
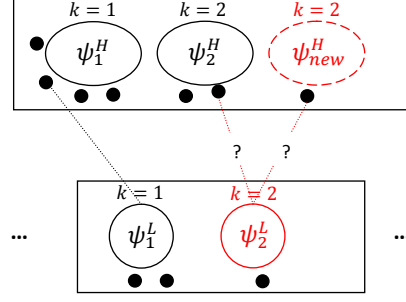
Fig. 4: Illustration of minimal path assumption. This figure shows an example of the seating assignments in a hierarchy of Chinese restaurants of a higher-level restaurant and a lower-level restaurant. Each table in the lower restaurant is assigned to a table in the higher restaurant and tables on the same path serve the same dish k. When sampling the assignment for table ψ_2^L in the lower restaurant, given that dish $k = 2$ is assigned to this table, there are two options for how the table in the higher restaurant could be selected. It could be an existing table ψ_2^H or a new table ψ_{new}^H, both serving dish $k = 2$. Under the minimal path assumption, it is always assigned to an existing table (if possible) and only assigned to a new table if there is no table with the given dish. In this case, the minimal path assumption will assign ψ_2^L to ψ_2^H.

The first factor is the prior probability of assigning to a path according to the minimum path assumption (Wallach, 2006, p. 60),

$$P(z_{c,t,n} = k \mid \boldsymbol{z}^{-c,t,n}) \propto \frac{N_{c,s_t,k}^{-c,t,n} + \alpha_c \dfrac{N_{c,k}^{-c,t,n} + \alpha_0 \dfrac{N_k^{-c,t,n} + \alpha \dfrac{1}{K^+}}{N_{\cdot}^{-c,t,n} + \alpha}}{N_{c,\cdot}^{-c,t,n} + \alpha_0}}{N_{c,s_t,\cdot}^{-c,t,n} + \alpha_c}, \quad (3)$$

where K^+ is the current number of shared topics.[5] Intuitively, Equation 3 computes the probability of token $w_{c,t,n}$ being generated from a shared topic k. This probability is proportional to $N_{c,s_t,k}$—the number of customers sitting at table serving dish k at restaurant G_{c,s_t}, smoothed by the probability of generating this token from the table serving dish k at the higher-level restaurant (i.e., restaurant G_c). This smoothing probability is computed in the same hierarchical manner until the top restaurant is reached, where the base distribution over topics is uniform and the probability of picking a topic is equal to $1/K^+$. Equation 3 also captures the case where a table is empty; when the number of customers on that table is zero, the probability of generating the token from the corresponding topic relies entirely on the smoothing probability from the higher-level restaurant's table.

---

[5] The superscript $^+$ is to denote that this number is unbounded and varies during the sampling process.

The second factor is the data likelihood. After integrating out all ψ's, we have

$$P(w_{c,t,n} = w \mid z_{c,t,n} = k, \boldsymbol{w}^{-c,t,n}, \boldsymbol{l}, *) \propto \begin{cases} \dfrac{M_{k,w}^{-c,t,n} + \lambda}{M_{k,\cdot}^{-c,t,n} + V\lambda}, & \text{if } k \text{ exists;} \\ \dfrac{1}{V}, & \text{if } k \text{ is new.} \end{cases} \tag{4}$$

Here, $M_{k,w}$ denotes the number of times word type w in the vocabulary is assigned to topic k; marginal counts are represented with $\cdot$ and $*$ represents all hyperparameters; V is the size of the vocabulary, and the superscript $^{-c,t,n}$ denotes the same counts excluding $w_{c,t,n}$.

3.3.2 Sampling Topic Shift Indicators

Sampling the topic shift variable $l_{c,t}$ requires us to consider merging or splitting segments. We define the following notation:

- $\boldsymbol{k}_{c,t}$: the shared topic indices of all tokens in turn t of conversation c.
- $S_{a_{c,t},x}$: the number of times speaker $a_{c,t}$ is assigned the topic shift with value $x \in \{0, 1\}$.
- $J_{c,s}^x$: the number of topics in segment s of conversation c if $l_{c,t} = x$
- $N_{c,s,j}^x$: the number of tokens assigned to the segment-level topic j when $l_{c,t} = x$.[6]

Again, the superscript $^{-c,t}$ is used to denote the 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 "no shift" case, i.e. when the topic distribution is identical to that of the previous turn. We sample this assignment with the following probability:

$P(l_{c,t} = 0 \mid \boldsymbol{l}^{-c,t}, \boldsymbol{w}, \boldsymbol{k}, \boldsymbol{a}, *) \propto$

$$\frac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \times \frac{\alpha_c^{J_{c,s_t}^0} \prod_{j=1}^{J_{c,s_t}^0} (N_{c,s_t,j}^0 - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^0} (x - 1 + \alpha_c)} \tag{5}$$

In Equation 5, 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 $l_{c,t} = 0$.[7]

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. The probability of sampling this assignment is:

---

[6] 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 out in full detail in the Appendix.

[7] Refer to (Gershman and Blei, 2012) for a detailed derivation of this joint probability.

$$P(l_{c,t} = 1 \mid \boldsymbol{l}^{-c,t}, \boldsymbol{w}, \boldsymbol{k}, \boldsymbol{a}, *) \propto$$

$$\frac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \times \left(\frac{\alpha_c^{J_{c,(s_t-1)}^1} \prod_{j=1}^{J_{c,(s_t-1)}^1} (N_{c,(s_t-1),j}^1 - 1)!}{\prod_{x=1}^{N_{c,(s_t-1),\cdot}^1} (x - 1 + \alpha_c)} \quad \frac{\alpha_c^{J_{c,s_t}^1} \prod_{j=1}^{J_{c,s_t}^1} (N_{c,s_tj}^1 - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^1} (x - 1 + \alpha_c)} \right) \tag{6}$$

As above, the first factor in Equation 6 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 Data and Annotations

Datasets	Speakers	Conversations	Annotations	Content
ICSI Meetings	60	75	topics	engineering
2008 Debates	9	4	topics	politics
2012 Debates	40	9	none	politics
Crossfire†	2567	1134	influencer	politics
Wikipedia discussions†	604	1991	influencer	varied

Table 1: Summary of datasets detailing how many distinct speakers are present, how many distinct conversations are in the corpus, the annotations available, and the general content of the dataset. † denotes datasets we annotated.

We validate our approach using five different datasets (Table 1). In this section, we describe the properties of each of the datasets and what information is available from the data. The datasets with interesting existing annotations typically are small and specialized. After validating our approach on simpler datasets, we move to larger datasets that we can explore qualitatively or by annotating them ourselves.

4.1 Datasets

We first describe the datasets that we use in our experiments. For all datasets, we tokenize texts using OpenNLP's tokenizer and remove common stopwords.[8] After that, we remove turns that are very short since they do not contain much information content-wise and most likely there is no topic shift during these turns. We empirically remove turns that have fewer than 5 tokens after removing stopwords.

---

[8] http://opennlp.apache.org/

Speaker	Type	Turn clauses	T_Q	T_R
Brokaw	Q	Sen. Obama, time for a discussion. I'm going to begin with you. Are you saying to Mr. Clark and to the other members of the American television audience 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, deal with critical issues like health care and energy, and we're going to have to change the culture in Washington so that lobbyists and special interests aren't driving the process and your voices aren't being drowned out.	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 2: Example turns from the annotated 2008 election debates (Boydstun et al, 2013). Each clause in a turn is coded with a *Question Topic Code* (T_Q) and a *Response Topic Code* (T_R). The topic codes (T_Q and T_R) are from the Policy Agendas Topics Codebook. In this example, the following topic codes are used: Macroeconomics (1), Housing & Community Development (14), Government Operations (20).

The ICSI Meeting Corpus: The ICSI Meeting Corpus consists of 75 transcribed meetings at the International Computer Science Institute in Berkeley, California (Janin et al, 2003). Among these, 25 meetings were annotated with reference segmentations (Galley et al, 2003). Segmentations are *binary*, i.e., each point in the document is either a segment boundary or not, and on average each meeting has 8 segment boundaries. We use this dataset for evaluating topic segmentation (Section 5). 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 between Barack Obama and John McCain and a vice presidential debate between Joe Biden and Sarah Palin (Boydstun et al, 2013). 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 Code* (T_Q) and a *Response Topic Code* (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 Codebook, a widely used inventory

containing codes for 19 major topics and 225 subtopics.[9] Table 2 shows an example annotation.

To obtain reference segmentations in debates, 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 topic codes not appearing in the previous turn; for response-typed turns, the score is the fraction of clause topic codes 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 labeling a turn as a segment boundary if the computed score is 1. This threshold is chosen to include only true segment boundaries. After preprocessing, this dataset contains 9 unique speakers and the vocabulary contains 1,761 non-stopword tokens.

The 2012 Republican Primary Debates: We also downloaded nine transcripts in the 2012 Republican Party presidential debates, whose information is shown in Table 3. Since the transcripts are pulled from different sources, we perform a simple entity resolution step using edit distance to merge duplicate participants' names. For example, "ROMNEY", "MITT ROMNEY" are resolved into "ROMNEY"; "PAUL", "REP. PAUL", "REPRESENTATIVE RON PAUL R-TX" are resolved into "PAUL" etc. We also merge anonymous participants such as "UNIDENTIFIED FEMALE", "UNIDENTIFIED MALE", "QUESTION", "UNKNOWN" etc into a single participant named "AUDIENCE". After preprocessing, there are 40 unique participants in these 9 debates including candidates, moderators and audience members. This dataset is not annotated and we only use it for qualitative evaluation.

Date	Place	Sponsor	Participants
13 Jun. 2011	Goffstown, NH	CNN	Bachmann, Cain, Gingrich, Paul, Pawlenty, Romney, Santorum
12 Sep. 2011	Tampa, FL	CNN	Bachmann, Cain, Gingrich, Huntsman, Paul, Perry, Romney, Santorum
18 Oct. 2011	Las Vegas, NV	CNN	Bachmann, Cain, Gingrich, Paul, Perry, Romney, Santorum
09 Nov. 2011	Rochester, MI	CNBC	Bachmann, Cain, Gingrich, Huntsman, Paul, Perry, Romney, Santorum
22 Nov. 2011	Washington, DC	CNN	Bachmann, Cain, Gingrich, Huntsman, Paul, Perry, Romney, Santorum
19 Jan. 2012	Charleston, SC	CNN	Gingrich, Paul, Romney, Santorum
23 Jan. 2012	Tampa, FL	NBC	Gingrich, Paul, Romney, Santorum
26 Jan. 2012	Jacksonville, FL	CNN	Gingrich, Paul, Romney, Santorum
22 Feb. 2012	Mesa, AZ	CNN	Gingrich, Paul, Romney, Santorum

Table 3: List of the 9 Republican Party presidential debates used.

---

[9] http://www.policyagendas.org/page/topic-codebook

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.[10] There are 2567 unique speakers and the vocabulary size is 16,791. Unlike the previous two datasets, *Crossfire* does not have explicit topic segmentations, so we use it to explore speaker-specific characteristics (Section 6.2).

A:	The current lead sentence has been agreed upon by many - I know, I was embroiled in the huge debate that developed into the current lead. However, the sentence is still kinda awkward - even though it captures the broader essence of evolutionary theory. I would like to propose an alternate (below), because there is a problem with the way that the term change is used, as Kirk J. Fitzhugh has noted: "Change is not the pertinent quality of interest in evolution". Hence: Evolution is the gradual departure across successive generations in the constituency of the inherited characteristics of organisms in biological populations.
B:	No thank you, this is just more obscurantism.
A:	It's wp:V, not obscurantism, consistent with the history of the science. Not much thought goes into conceiving that "Evolution is change", but if you are asked to think past this and call it obscurantism in your critique, it is a strange response. Obscurantism: "is the practice of deliberately preventing the facts or the full details of some matter from becoming known" - ironic that this applies more aptly to your rejection.
B:	Your obsession with providing the most scientifically accurate and current definition of evolution prevents the average reader from having a chance at understanding this article. That is obscurantism. It is not WPV, because that definition is not by a longshot the most commonly used, and specifically it is entirely unsuited for works meant to be read by lay readers.
C:	This is a general encyclopedia, not a graduate level evolutionary biology course. Keeping it simple so that people can understand what we write without having an advanced degree is a good thing. So no, let's keep the lead as is.

Table 4: Example of a Wikipedia discussion in our dataset.

Wikipedia Discussions: Each article on Wikipedia has a related discussion page so that the individuals writing and editing the article can discuss the content, editorial decisions, and the application of Wikipedia policies (Butler et al, 2008). Unlike the other situations, Wikipedia discussions are not spoken conversations that have been transcribed. Instead, these conversations are written asynchronously.

However, Wikipedia discussions have much of the same properties as our other corpora. Contributors have different levels of responsibility and prestige, and many contributors are actively working to persuade the group to accept their proposed policies (for an example, see Table 4), other contributors are attempting to maintain civility, and other contributors are attacking their ostensible collaborators.

---

[10] http://www.cs.umd.edu/~vietan/topicshift/crossfire.zip

Unlike spoken conversations, Wikipedia discussions lack social norms that prevent an individual from writing as often or as much as they want. This makes common techniques such as counting turns or turn lengths less helpful measures to discover who influencers are.

4.2 Influencer Annotation

Our goal is to discover who are the influencers in these discussions. To assess our ability to discover influencers, we annotated randomly selected documents from both the Wikipedia and *Crossfire* datasets. This process proceeded as follows. First, we followed the annotation guidelines for influencers proposed by Bender et al (2011) for Wikipedia discussion. In our guideline, we use the following definition to identify influencers: A discussant is considered an influencer if he or she initiated a topic shift that steered the conversation in a different direction, convinced others to agree to a certain viewpoint, or used an authoritative voice that caused others to defer to or reference that person's expertise. A discussant is not identified as an influencer if he or she merely initiated a topic at the start of a conversation, did not garner any support from others for the points he or she made, or was not recognized by others as speaking with authority. After annotating an initial set of documents, we revised our annotation guidelines and retrained two independent annotators until we reached an intercoder reliability Cohen's Kappa (Artstein and Poesio, 2008) of .80.[11]

Wikipedia Discussions: Coders first learned to annotate transcripts using Wikipedia discussion data.The two coders annotated over 400 English Wikipedia discussion transcripts for influencer in batches of 20 to 30 transcripts each week. For the English transcripts, each coder annotated the transcripts independently, then annotations were compared for agreement; any discrepancies in the annotations were resolved through discussion of how to apply the coding scheme. After the first four sets of 20 to 30 transcripts, the coders were able to code the transcripts with acceptable intercoder reliability (Cohen's Kappa > .80). Once the coders reached acceptable intercoder reliability for two sets of English data in a row, the coders began independently coding the remaining set of transcripts. Intercoder reliability was maintained at an acceptable level (Cohen's Kappa > .80) for the English transcripts over the subsequent weeks of coding.

---

[11] Kappa was measured based on whether the two annotators agreed on (a) whether there was an influencer, (b) who the primary influencer was, and (c) if there was a secondary influencer. When discrepancies occurred between the annotators, they were resolved through discussion between the annotators and with the supervising researcher. So decisions were not "yes or no" about each speaker; instead, they were about whether or not there was an influencer in each overall interaction, and if so, who the primary and secondary influencers were in a particular interaction.

Crossfire: We then turned our attention to the *Crossfire* dataset. We split each *Crossfire* episode into smaller segments using the "COMMERCIAL_BREAK" tags and use each segment as a unit of conversation. The same two coders annotated the *Crossfire* data. To prepare for annotating the *Crossfire* interactions, the coders both annotated the same set of 20 interactions. First the intercoder reliability Cohen's Kappa was calculated for the agreement between the coders, then any disagreements between the coders were resolved through discussion about the discrepant annotations. The first set of 20 transcripts was coded with a Cohen's Kappa of .65 (before discussion). This procedure was repeated twice; each time the coders jointly annotated 20 transcripts, reliability was calculated, and any discrepancies were resolved through discussion. The third set achieved an acceptable Cohen's Kappa of .80. The remaining transcripts were then split and annotated separately by the two coders. In all, 105 *Crossfire* episode segments were annotated. An annotation guideline for *Crossfire* is included in the Appendix B.

5 Evaluating Topic Segmentation

In this section, we examine how well SITS can identify when new topics are introduced, i.e., how well it can segment conversations. We discuss metrics for evaluating an algorithm's segmentation relative to a gold annotation, describe our experimental setup, and report those results.

5.1 Experiment Setups

Evaluation Metrics: To evaluate the performance on topic segmentation, 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 windows of size k in the document and count whether the two endpoints of the window are (im)properly segmented against the gold segmentation. More formally, given a reference segmentation $\mathcal{R}$ and a hypothesized segmentation $\mathcal{H}$, the value of P_k for a given window size k is defined as follow:

$$P_k = \frac{\sum_{i=1}^{N-k} \delta_{\mathcal{H}}(i, i+k) \oplus \delta_{\mathcal{R}}(i, i+k)}{N - k} \tag{7}$$

where $\delta_{\mathcal{X}}(i, j)$ is 1 if the segmentation $\mathcal{X}$ assigns i and j to the same segment and 0 otherwise; $\oplus$ denotes the XOR operator; N is the number of candidate boundaries.

WD improves P_k by considering how many boundaries lie between two points in the document, instead of just looking at whether the two points are separated or not. WD of size k between two segmentations $\mathcal{H}$ and $\mathcal{R}$ is defined as:

$$\text{WD} = \frac{\sum_{i=1}^{N-k} [|b_{\mathcal{H}}(i, i+k) - b_{\mathcal{R}}(i, i+k)| > 0]}{N - k} \tag{8}$$

	Model	$\widehat{EMD}$	P_k			WindowDiff		
			$k = 5$	10	15	$k = 5$	10	15
ICSI	TextTiling	2.507	.289	.388	.451	.318	.477	.561
	P-NoSpeaker-single	1.949	.222	.283	.342	.269	.393	.485
	P-NoSpeaker-all	1.935	**.207**	.279	.335	**.253**	.371	.468
	P-SITS	**1.807**	.211	**.251**	.289	.256	**.363**	**.434**
	NP-HMM	2.189	.232	.257	.263	.267	.377	.444
	NP-SITS	2.126	.228	.253	**.259**	.262	.372	.440
2008 Debates	TextTiling	2.821	.433	.548	.633	.534	.674	.760
	P-NoSpeaker-single	2.822	.426	.543	.653	.482	.650	.756
	P-NoSpeaker-all	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	**1.813**	**.332**	**.269**	**.231**	**.470**	**.600**	**.692**

Table 5: 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.

where $b_{\mathcal{X}}(i, j)$ counts the number of boundaries that the segmentation $\mathcal{X}$ puts between two points i and j.

However, these metrics have a major drawback. 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. In previous work, to be fair for all methods, thresholds are usually set so that the number of segments is equal to a predefined value (Purver et al, 2006; Galley et al, 2003). In practice, this value is usually unknown.

To overcome these limitations, we also use $\widehat{EMD}$ (Pele and Werman, 2008), a variant of the *Earth Mover's Distance* (*EMD*). Originally proposed by Rubner et al (2000), *EMD* is a metric that measures the distance between two normalized histograms. Intuitively, it measures the minimal cost that must be paid to transform one histogram into the other. *EMD* is a true metric only when the two histograms are normalized (e.g., two probability distributions). $\widehat{EMD}$ relaxes this restriction to define a metric for non-normalized histograms by adding or subtracting masses so that both histograms are of equal size.

Applied to our segmentation problem, each segmentation can be considered a histogram where each candidate boundary point corresponds to a bin. The probability of each point being a boundary is the mass of the corresponding bin. We use $|i - j|$ as the ground distance between two points i and j.[12] To compute $\widehat{EMD}$ we use the FastEMD implementation (Pele and Werman, 2009).

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

---

[12] The ground distance is the distance between two bins in a histogram. Please refer to (Pele and Werman, 2008) for a more formal definition of $\widehat{EMD}$.

- **TextTiling** (Hearst, 1997) is one of the earliest and most widely used general-purpose topic segmentation algorithms, sliding a fixed-width window to detect major changes in lexical similarity.
- **P-NoSpeaker-single**: parametric version of SITS without speaker identity, run individually on each single conversation (Purver et al, 2006).
- **P-NoSpeaker-all**: parametric version of SITS 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 with speaker identity. This model uses the same assumption as the Sticky HDP-HMM (Fox et al, 2008), where a single topic is associated with each turn.
- **NP-SITS**: the nonparametric version of SITS with speaker identity run on all conversations.

Parameter Settings and Implementation: 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. Parametric models are run with 25, 50 and 100 topics and the best results (averaged over 10 chains) are reported.

5.2 Results and Analysis

Table 5 shows the performance of various models on the topic segmentation problem, using the ICSI corpus and the 2008 debates.

Consistent with previous results in the literature, 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. Furthermore, NP-SITS outperforms NP-HMM in both experiments, suggesting that using a distribution over topics for turns is better than using a single topic. This is consistent with parametric models 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 opponent's weakness can be useful strategies (Boydstun et al, 2013). In contrast, meetings (particularly low-stakes ICSI meetings, technical discussions in R&D group) tend to have pragmatic rather than strategic topic shifts. In addition, 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, 2009b) would favor the nonparametric model more.

6 Evaluating Topic Control

In this section, we focus on the ability of SITS to capture the extent to which individual speakers affect topic shifts in conversations. Recall that SITS associates with each speaker a topic shift tendency π that represents the probability of changing the topic in the conversation. While topic segmentation is a well studied problem, hence the evaluation in Section 5, 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 look qualitatively at the behavior of the model.
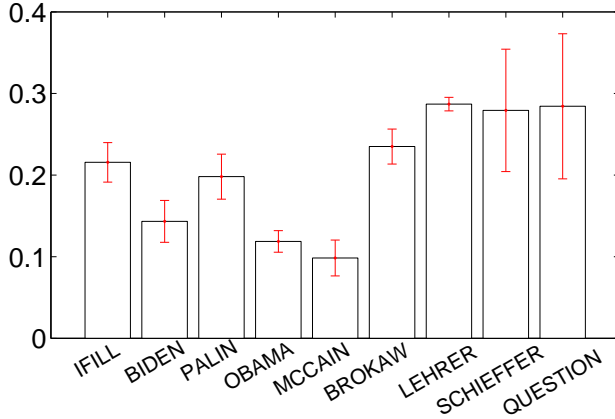
6.1 2008 Election Debates

Fig. 5: Topic shift tendency π of speakers in the 2008 Presidential Election Debates (larger means greater tendency). IFILL was the moderator in the vice presidential debate between BIDEN and PALIN; BROKAW, LEHRER and SCHIEFFER were the moderators in the three presidential debates between OBAMA and MCCAIN; QUESTION collectively refers to questions from the audiences.

To obtain a posterior estimate of π (Figure 5) we create 10 chains with hyperparameters sampled from the uniform distribution $U(0, 1)$ and average π over 10 chains (as described in Section 5.1). In these debates, Ifill is the moderator of the debate between Biden and Palin; Brokaw, Lehrer and Schieffer

are the three moderators of the 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 generally 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. Similarly, observe that the "Question" speaker had a relatively high variance, consistent with that "participant" in the model actually being an amalgamation of many distinct speakers.

Interestingly, however, in the vice-presidential debate, the score of moderator Ifill is higher than the candidates' scores only by a small margin, and it is indistinguishable from the degree of topic control displayed by Palin. Qualitatively, the assessment of the model is consistent with widespread perceptions and media commentary at the time that characterized Ifill as a weak moderator. For example, *Harper's Magazine*'s Horton (2008) discusses the context of the vice-presidential debate, in particular the McCain campaign's characterization of Ifill as a biased moderator because she "was about to publish a book entitled *The Breakthrough* that discusses Barack Obama, and a number of other black politicians, achieving national prominence". According to Horton:

> First, the charges against Ifill would lead to her being extremely passive in her questioning of Palin and permissive in her moderating the debate. Second, the charge of bias against Ifill would enable Palin to simply skirt any questions she felt uncomfortable answering and go directly to a pre-rehearsed and nonresponsive talking point. This strategy succeeded on both points.

Similarly, Fallows (2008) of *The Atlantic* included the following in his "quick guide" remarks on the debate:

> **Ifill**, moderator: *Terrible.* Yes, she was constrained by the agreed debate rules. But she gave not the slightest sign of chafing against them or looking for ways to follow up the many unanswered questions or self-contradictory answers. This was the big news of the evening . . .
> **Palin**: *"Beat expectations."* In every single answer, she was obviously trying to fit the talking points she had learned to the air time she had to fill, knowing she could do so with impunity from the moderator.

That said, our quantitative modeling of topic shift tendency suggests that all candidates managed to succeed at some points in setting and controlling the topic of conversation in the debates. In the presidential debates, our model gives Obama a slightly higher score than McCain, consistent with social science claims that Obama had the lead in setting the agenda over McCain (Boydstun et al, 2013). Table 6 shows some examples of SITS-detected topic shifts.

	Previous turn	Turn detected as shifting topic
2008 Debates Dataset	BIDEN: Well, mortgage-holders didn't pay the price [...] Barack Obama pointed out two years ago that there was a *subprime mort-gage* [...]	PALIN: That is not so, but because that's just a quick answer, I want to talk about, again, my record on *energy* ... When we talk about energy, we need to consider the need to do all that we can to allow this nation to become energy independent [...]
	PALIN: Your question to him was whether he supported *gay mar-riage* 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 *foreign policy*. 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 *energy* and *climate control*. 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 *financial crisis* 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. [...]
Crossfire Dataset	PRESS: But what do you say, governor, to Governor Bush and [...] your party who would let politicians and not medical scientists decide what *drugs* 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 6: 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 π. Some keywords are manually italicized to highlight the topics discussed.

6.2 *Crossfire*

The *Crossfire* dataset has many more speakers than the presidential and vice-presidential debates. This allows us to examine more closely what we can learn about speakers' topic shift tendency and ask additional questions; for example, assuming that changing the topic is useful for a speaker, how can we characterize who does so effectively? In our analysis, we take advantage of properties of the *Crossfire* data to examine the relationship between topic shift tendency, social roles, and political ideology.

In order to focus on frequent speakers, we filter out speakers with fewer than 30 turns. Most speakers have relatively small π, with the mode around

Rank	Speaker	π	Rank	Speaker	π
1	ANNOUNCER	.884	10	JOHN KASICH	.570
2	MALE	.876	11	JAMES CARVILLE[†]	.550
3	QUESTION	.755	12	TUCKER CARLSON[†]	.550
4	GEORGE W. BUSH[‡]	.751	13	PAUL BEGALA[†]	.545
5	BILL PRESS[†]	.651	14	CHRISTINE T. WHITMAN	.533
6	FEMALE	.650	15	TERRY MCAULIFFE	.529
7	AL GORE[‡]	.650	16	MARY MATALIN[†]	.527
8	NARRATOR[‡]	.642	17	JOHN MCCAIN	.524
9	ROBERT NOVAK[†]	.587	18	ARI FLEISCHER	.522

Table 7: Top speakers by topic shift tendencies from our *Crossfire* dataset. We mark hosts (†) and "speakers" who often (but not always) appeared in video clips (‡). ANNOUNCER makes announcements at the beginning and at the end of each show; NARRATOR narrates video clips; MALE and FEMALE refer to unidentified male and female respectively; QUESTION collectively refers to questions from the audience across different shows. Apart from those groups, speakers with the highest tendency were political moderates.

0.3. There are, however, speakers with very high topic shift tendencies. Table 7 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" that the show uses to frame and propose new topics. These are audience questions, news clips (e.g. many of Gore's and Bush's turns from 2000), and voiceovers. That SITS is able to recover these is reassuring, similar to what it has to say about moderators in the 2008 debates. Second, the stable of regular hosts receives high topic shift tendencies, which is again reasonable given their experience with the format and ostensible moderation roles (though in practice they also stoke lively discussion).

The third category is more interesting. The remaining non-hosts with high topic shift tendency appear to be relative moderates on the political spectrum:

- John Kasich, one of few Republicans to support the assault weapons ban and who was elected in 2010 as the 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)

Although these observations are at best preliminary and require further investigation, we would conjecture that in *Crossfire*'s highly polarized context, it was the political moderates who pushed back, exerting more control over the agenda of the discussion, rather than going along with the topical progression and framing as posed by the show's organizers. Table 6 shows several detected topic shifts from these speakers. In two of these examples, McCain and Whitman are Republicans disagreeing with President Bush. In

the other, Kasich is defending a Republican plan (school vouchers) popular with traditional Democratic constituencies.
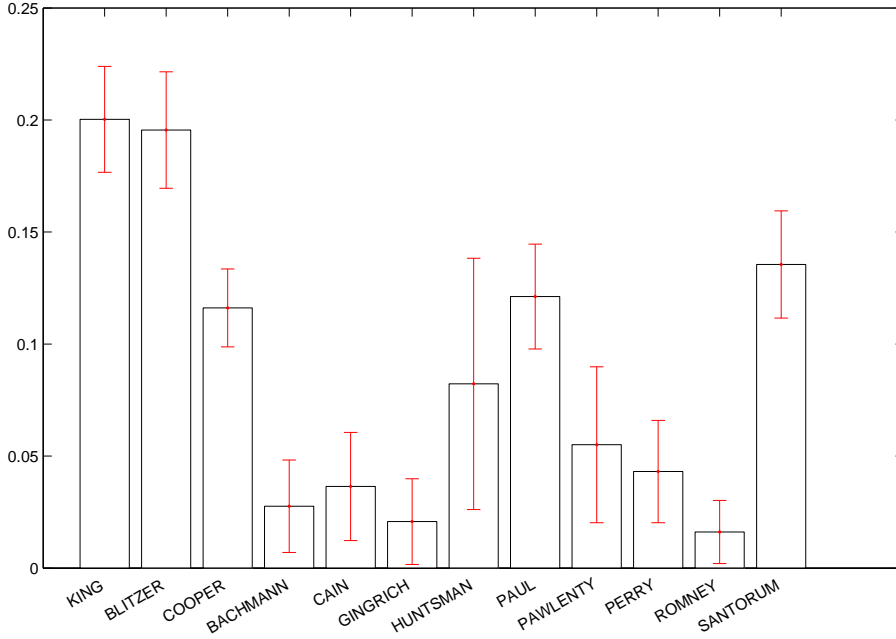
6.3 2012 Republican Primary Debates

Fig. 6: Topic shift tendency π of speakers in the 2012 Republican Primary Debates (larger means greater tendency). KING, BLITZER and COOPER are moderators in these debates; the rest are candidates.

As another qualitative data point, we include in Figure 6 the model's topic shift tendency scores for a subset of nine 2012 Republican primary debates. Although we do not have objective measures to compare against, nor clearly stated contemporary commentary as in the case of Ifill's performance as moderator, we would argue that the model displays quite reasonable face validity in the context of the Republican race.

For example, among the Republican candidates, Ron Paul is known for tight focus on a discrete set of arguments associated with his position that "the proper role for government in America is to provide national defense, a court system for civil disputes, a criminal justice system for acts of force and fraud, and little else" (Paul, 2007), often regardless of the specific question that was asked. Similarly, Rick Santorum's performance in the primary debates tended to include strong rhetoric on social issues. In contrast, Mitt Romney tended

to be less aggressive in his responses, arguably playing things safer in a way that was consistent with his general position throughout the primaries as the front-runner.

7 Detecting Influencers in Conversations

7.1 Computational Methods for Influencer Detection

In this section, we turn to the direct application and validation of the model in detecting influencers in conversations. Even though influence in conversations has been studied for decades in communication and social psychology, computational methods have only emerged in recent years, thanks to improvements in both quantity and quality of conversational data. As one example, an early computational model to quantify influence between conversational participants (Basu et al, 2001) modeled interactions among a conversational group in a multi-sensor lounge room where people played interactive debating games. In these games, each participant can be in two states: speaker or silent. The model equates each participant with a Markov model. Each participant is allowed to be in either speaking state or silent state at each time step and the transition from one state to another of an individual is influenced by other participants' states. This allows the model to capture pair-wise interactions among participants in the conversation. Zhang et al (2005) then extended the work by proposing a model with two-level structure: the participant level, representing the actions of individual participants, and the group level, representing group-level actions. In this setting, the influence of each participant on the actions of the whole group is explicitly captured by the model. These models use expensive features such as prosody and visual cues.

Another popular approach is to treat influencer detection as a supervised classification problem that separates influential individuals from non-influential ones. Rienks and Heylen (2005) focus on extracting a set of structural features that can predict participants' involvement using Support Vector Machines (SVMs) (Cortes and Vapnik, 1995). Later, Rienks et al (2006) improved their previous work by extending the set of features to include features capturing topic changes as well as those derived from audio and speech. Again, we do not use any features extracted from audio or visual data, which makes our approach more generalizable. The two most relevant and most useful features extracted from the meeting textual transcripts are number of turns and length of turns, which we use as the baseline in our experiments described in Section 7.2. Biran et al (2012) also follow a similar approach to detecting influencers in written online conversations by extracting features to capture different conversational behaviors such as persuasion, agreement/disagreement and dialog patterns.

In this paper, we are interested in determining who are the influencers in a conversation using only the conversation transcripts. We tackle this problem by using an unsupervised ranking approach. It is worth mentioning that, even though we are focused on studying how conversational influence expressed in

textual data, there has also been a body of work approaching this problem by studying audio data (Hung et al, 2011), visual data (Otsuka et al, 2006) and both audio-visual activity cues (Jayagopi et al, 2009; Aran and Gatica-Perez, 2010).

Our main purpose in this experimentation is to assess how effective SITS can be in detecting influencers in conversations, especially in comparison with methods based on structural patterns of conversations. In order to do so, we focus on the *influencer detection problem*, which can be stated as follows: *given a speaker in a multi-party conversation, predict whether the speaker is influential*. In the remaining of this section, we describe in details the approach we take, the experimental setups, and the results.

7.2 Influencer Detection Problem

The influencer detection problem can be tackled using different methods that can be broadly classified into *classification* and *ranking* approaches. Most previous work follows the classification approach, in which different sets of features are proposed and a classifier is used (Rienks and Heylen, 2005; Rienks et al, 2006; Biran et al, 2012). In this paper, we follow the ranking approach.

The ranking approach allows us to focus on individual functions that take a set of individuals and produce an ordering over those individuals from most influential to least influential. The function that produces this ordering is called a *ranking method*. More specifically, given a speaker a in a conversation c, each ranking method will provide an *influence score* $\mathcal{I}_{a,c}$ that indicates how influential speaker a is in conversation c. We emphasize that, unlike most classification approaches (Rienks and Heylen, 2005; Rienks et al, 2006; Biran et al, 2012), the ranking approach we are focusing on is entirely unsupervised and thus requires no training data.

The ranking approach has a straightforward connection to the classification approach, as each ranking function can be turned into a feature in the supervised classification framework. However, viewing the ranking methods (features) independently allows us to compare and interpret the effectiveness of each feature in isolation. This is useful as an evaluation method because it is independent of the choice of classifier and is less sensitive to the size of training data, which is often a limiting factor in computational social science

We consider two sets of ranking methods: (1) structure-based methods, which use structural features and (2) topic-change-based methods, which use features extracted from the outputs of SITS.

Structure-based methods score each instance based on features extracted from the structures of the conversation. As defined in Section 2, we use T_c to denote the number of turns in conversation c; $a_{c,t}$ to denote the speaker that utters turn t in conversation c; and $N_{c,t}$ to denote the number of tokens in turn t in conversation c.

1. *Number of turns*: assumes that the more turns a speaker has during a conversation, the more influential he or she is. The influence score of this method is

$$\mathcal{I}_{a,c} = |\{t \in [1, T_c] \mid a_{c,t} = a\}| \tag{9}$$

2. *Total turn lengths*: instead of the number of turns, this method uses the total length of turns uttered by the speaker.

$$\mathcal{I}_{a,c} = \sum_{t \in [1, T_c] \mid a_{c,t} = a} N_{c,t} \tag{10}$$

The two structural features used here capture the activeness of the speakers during a conversation and have been shown to be among the most effective features to detect influencers. These two structure-based methods are appropriate baselines in our experiment since, although being simple, they have been proven to be very effective in detecting influencers, both qualitatively (Bales, 1970) and quantitatively (Rienks et al, 2006; Biran et al, 2012).

Topic-change-based methods score each instance based on features extracted from the posterior distributions of SITS.

1. *Total topic shifts* attempts to quantify the expected number of topic shifts speaker a makes in conversation c. Recall that in SITS, each turn t in conversation c is associated with a binary latent variable $l_{c,t}$, which indicates whether the topic of turn t is changed or not (these latent variables are introduced in Section 3). This expectation is computed through the empirical average of samples from the Gibbs sampler, $\bar{l}_{c,t}$, after a burn-in period.[13] Intuitively, the higher $\bar{l}_{c,t}$ is, the more successful the speaker $a_{c,t}$ is in changing the topic of the conversation at this turn t.

$$\mathcal{I}_{a,c} = \sum_{t \in [1, T_c] \mid a_{c,t} = a} \bar{l}_{c,t} \tag{11}$$

2. *Weighted topic shifts* also quantify the topic changes a speaker makes by using the average topic shift indicator $\bar{l}_{c,t}$ but weighted by $(1 - \pi_a)$, where π_a is the topic shift tendency score of the speaker a. The basic idea here is that not all topic shifts should be counted equally. A successful topic shift by a speaker with small topic shift tendency score should be weighted higher than a successful topic by a speaker with high topic shift tendency score. The influence score of this ranking method is defined as

$$\mathcal{I}_{a,c} = (1 - \pi_a) \cdot \sum_{t \in [1, T_c] \mid a_{c,t} = a} \bar{l}_{c,t} \tag{12}$$

---

[13] For more details on how to compute this value, refer to Section 3 of (Resnik and Hardisty, 2010)

7.3 Experimental Setup

Datasets: In this experiment, we use two datasets annotated for influencers: *Crossfire* and Wikipedia discussion pages. These two datasets and the annotation procedures are described in detail in Section 4. Table 8 shows dataset statistics.

Statistics	*Crossfire*	Wikipedia
No. conversations	3391	604
No. unique speakers	2381	1991
Avg no. turns per conversation	38.2	12.8
Avg no. speakers per conversation	5	7
No. conversations annotated	85	48
No. positive instances	197	57
No. negative instances	182	338

Table 8: Statistics of the two datasets *Crossfire* and Wikipedia discussions that we annotated influencers. We use these two datasets to evaluate SITS on influencer detection.

Parameter settings and implementation: As before, inference is conducted using Gibbs sampling with 10 randomly initialized chains is used. Initial hyperparameter values are sampled from $U(0,1)$ and statistics are collected after 200 burn-in iterations with a lag of 20 iterations over a total of 1000 iterations. Slice sampling is also used to optimize the hyperparameters.

Evaluation measurements: In order to evaluate the effectiveness of each ranking method in detecting the influencers, we use three standard evaluation measurements. The first measurement is F_1, the harmonic mean of precision and recall,

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

Even though F_1 is widely used, an important disadvantage it has is that it only examines a subset of top instances with highest scores, which might be the "easiest" cases. This phenomenon might lead to biased results when comparing the performance of different ranking methods. To overcome this problem, we also use AUC-ROC and AUC-PR, which measure the area under the Receiver-Operating-Characteristic (ROC) curve and the Precision-Recall (PR) curve. Using these two measurements, we are able to compare the performances of ranking methods using the full ranked lists. Davis and Goadrich (2006) point out that, the PR curve might be better than the ROC curve for skewed datasets.

	Ranking methods	F_1	AUC-ROC	AUC-PR
Crossfire	Num. of turns ◇	.736	.795	.726
	Total turn lengths ◇	.716	.782	.730
	Total topic shifts ⋆	$.806 \pm .0122$	$.858 \pm .0068$	$.865 \pm .0063$
	Weighted topic shifts ⋆	$.828 \pm .0100$	$.869 \pm .0078$	$.873 \pm .0057$
Wikipedia	Num. of turns ◇	.367	.730	.291
	Total turn lengths ◇	.306	.732	.281
	Total topic shifts ⋆	$.552 \pm .0353$	$.752 \pm .0144$	$.377 \pm .0284$
	Weighted topic shifts ⋆	$.488 \pm .0295$	$.749 \pm .0149$	$.379 \pm .0307$

Table 9: Influencer detection results on *Crossfire* and Wikipedia discussion pages. For both datasets, topic-change-based methods (⋆) outperform structure-based methods (◇) by large margins. For all evaluation measuments, higher is better.

7.4 Results and Analysis

Table 9 shows the results of the four ranking methods using *Crossfire* and Wikipedia discussion datasets. Since we run our Gibbs samplers multiple times, the results of the two topic-change-based methods are reported with standard deviations (across different chains).

For both datasets, the two topic-change-based methods outperform the two structure-based methods by a large margin for all three evaluation measurements. The standard deviations in all three measurements of the two topic-change-based methods are relatively small. This shows the effectiveness of features based on topic changes in detecting influencers in conversations. In addition, the weighted topic shifts ranking method generally performs better than the total topic shifts method. This provides strong evidence that SITS is capable of capturing the speakers' propensity to change the topic. The improvement (if any) in the performance of the weighted topic shifts ranking method over the total topic shifts method is more obvious in the *Crossfire* dataset than in Wikipedia discussions. We argue that this is because conversations in Wikipedia discussion pages are generally shorter and contain more speakers than those in *Crossfire* debates. This leaves less evidence about the topic change behavior of the speakers in Wikipedia dataset and thus makes it more difficult for SITS to capture the speakers' behaviors well.

8 Conclusions and Future Work

In this article, we have introduced SITS, a nonparametric hierarchical Bayesian model that jointly captures topics, topic shifts, and individuals' tendency to control the topic in conversations. SITS takes a nonparametric topic modeling approach, representing each turn in a conversation as a distribution over topics and consecutive turns' topic distributions as dependent on each other.

Crucially, SITS also models speaker-specific properties. As such, it improves performance on practical tasks such as unsupervised segmentation, but it also is

attractive philosophically. Accurately modeling *individuals* is part of a broader research agenda that seeks to understand individuals' values (Fleischmann et al, 2011), interpersonal relationships (Chang et al, 2009a), and perspective (Hardisty et al, 2010), which can help create a better understanding of what people think based on what they write or say (Pang and Lee, 2008). One particularly interesting direction is to extend the model to capture how language is coordinated during the conversation and how it correlates with influence (Giles et al, 1991; Danescu-Niculescu-Mizil et al, 2012).

The problem of finding influencers in conversation has been studied for decades by researchers in communication, sociology and psychology, who have long acknowledged qualitatively the correlation between the ability of a participant to control conversational topic and his or her influence on other participants during the conversation. With SITS, we now introduce a computational technique for modeling more formally who is controlling the conversation. Empirical results on the two datasets we annotated (*Crossfire* TV show and Wikipedia discussion pages) show that methods based on SITS outperform previous methods that used conversational structure patterns in detecting influencers.

Using an unsupervised statistical model for detecting influencers is an appealing choice because it extends easily to other languages and to corpora that are multilingual (Mimno et al, 2009; Boyd-Graber and Blei, 2009). Moreover, topic models offer opportunities for exploring large corpora (Zhai et al, 2012) in a wide range of domains including political science (Grimmer, 2009), music (Hoffman et al, 2009), programming source code (Andrzejewski et al, 2007) or even household archeology (Mimno, 2011). Recent work has created frameworks for interacting with statistical models (Hu et al, 2011) to improve the quality of the latent space (Chang et al, 2009b), understand relationships with other variables (Gardner et al, 2010), and allow the model to take advantage of expert knowledge (Andrzejewski et al, 2009) or knowledge resources (Boyd-Graber et al, 2007).

This work has opened many possible future directions. First, even though associating each speaker with a scalar that models their tendency to change the topic does improve performance on both topic segmentation and influencer detection tasks, it is obviously an impoverished representation of an individual's conversational behaviors and could be enriched. For example, instead of just using a fixed parameter π for each conversational participant, one could extend the model to capture evolving topic shift tendencies of participants during the conversation. Modeling individuals' perspective (Paul and Girju, 2010), "side" (Thomas et al, 2006), or personal preferences for topics (Grimmer, 2009) would also enrich the model and better illuminate the interaction of influence and topic.

Another important future direction is to extend the model to capture more explicitly the distinction between agenda setting and interaction influence. For example, in many cases, questions or comments from the moderators during a political debate are just to shape the agenda of the debate and have little influence over how the candidates would respond. Agenda setting does not

have a direct effect on the views or opinions of others, and it does not try to sway the attitudes and beliefs of others. Agenda setting focuses generally on the topics that will be addressed, determining what those topics will be from the outset (McCombs and Reynolds, 2009). It is during an interaction that an influencer is able to shape the discussion by shifting the interaction from one topic to another or providing evidence or expertise that can shape the opinions and judgments about the topics. To be identified as an influencer, however, others in the interaction must acknowledge or recognize the value of the expertise or agree with the opinion and viewpoints that have been offered. Thus, adding modules to find topic expertise (Marin et al, 2010) or agreement/disagreement (Galley et al, 2004) during the conversation might help SITS improve its performance in detecting influencers.

Understanding how individuals use language to influence others goes beyond conversational turn taking and topic control, however. In addition to *what* is said, often *how* something is expressed is nearly as important (Greene and Resnik, 2009). Combining SITS with a model that can discover syntactic patterns (Sayeed et al, 2012) or multi-word expressions (Johnson, 2010) associated with those attempting to influence a conversation would allow us to better understand how individuals use word choice and rhetorical strategies to persuade (Cialdini, 2000) or coordinate with (Danescu-Niculescu-Mizil et al, 2012) others. Such systems could have a significant social impact, as they could identify, quantify, and measure attempts to spin or influence at a large scale. Models for automatic analysis of influence could lead to more transparent public conversations, ultimately improving our ability to achieve more considered and rational discussion of important topics, particularly in the political sphere.

References

Abbott R, Walker M, Anand P, Fox Tree JE, Bowmani R, 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)

Ahmed A, Xing EP (2008) Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering. In: Proceedings of SIAM International Conference on Data Mining

Ahmed A, Xing EP (2010) Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream. In: Proceedings of Uncertainty in Artificial Intelligence

Alarcon-del Amo M, Lorenzo-Romero C, Gomez-Borja M (2011) Classifying and profiling social networking site users: a latent segmentation approach. Cyberpsychology, Behavior, and Social Networking 14(9)

Andrzejewski D, Mulhern A, Liblit B, Zhu X (2007) Statistical debugging using latent topic models. In: Proceedings of European Conference of Machine Learning

Andrzejewski D, Zhu X, Craven M (2009) Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In: Proceedings of the International Conference of Machine Learning

Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics 2(6):1152–1174

Aran O, Gatica-Perez D (2010) Fusing audio-visual nonverbal cues to detect dominant people in group conversations. In: Proceedings of the International Conference on Pattern Recognition (ICPR)

Artstein R, Poesio M (2008) Inter-coder agreement for computational linguistics. Computational Linguistics 34(4)

Bales RF (1970) Personality and interpersonal behavior. Holt, Rinehart, and Winston

Basu S, Choudhury T, Clarkson B, Pentland AS (2001) Learning human interactions with the influence mdoel. Tech. Rep. 539, MIT Media Laboratory

Beeferman D, Berger A, Lafferty J (1999) Statistical models for text segmentation. Machine Learning 34

Bender EM, Morgan JT, Oxley M, Zachry M, Hutchinson B, Marin A, Zhang B, Ostendorf M (2011) Annotating social acts: authority claims and alignment moves in wikipedia talk pages. In: Proceedings of the Workshop on Languages in Social Media (LSM)

Biran O, Rosenthal S, Andreas J, McKeown K, Rambow O (2012) Detecting influencers in written online conversations. In: Proceedings of the Workshop on Language in Social Media (LSM)

Blau P (1964) Exchange and power in social life. Sociology political science, Transaction Books

Blei DM (2012) Probabilistic topic models. Communications of the ACM 55(4)

Blei DM, Lafferty JD (2006) Dynamic topic models. In: Proceedings of the International Conference of Machine Learning

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. Journal of Machine Learning Research 3

Booth N, Matic A (2011) Mapping and leveraging influencers in social media to shape corporate brand perceptions. Corporate Communications 16(3)

Boyd-Graber J, Blei DM (2008) Syntactic topic models. In: Proceedings of Advances in Neural Information Processing Systems

Boyd-Graber J, Blei DM (2009) Multilingual topic models for unaligned text. In: Proceedings of Uncertainty in Artificial Intelligence

Boyd-Graber J, Blei DM, Zhu X (2007) A topic model for word sense disambiguation. In: Proceedings of Emperical Methods in Natural Language Processing

Boydstun AE, Glazier RA, Phillips C (2013) Agenda control in the 2008 presidential debates. American Politics Research

Brooke ME, Ng SH (1986) Language and Social Influence in Small Conversational Groups. Journal of Language and Social Psychology 5(3)

Burrel NA, Koper RJ (1998) The efficacy of powerful/powerless language on attitudes and source credibility. In: Persuasion: Advances through meta-analysis

Butler B, Joyce E, Pike J (2008) Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In: International Conference on Human Factors in Computing Systems

Chang J, Boyd-Graber J, Blei DM (2009a) Connections between the lines: Augmenting social networks with text. In: Knowledge Discovery and Data Mining

Chang J, Boyd-Graber J, Wang C, Gerrish S, Blei DM (2009b) Reading tea leaves: How humans interpret topic models. In: Proceedings of Advances in Neural Information Processing Systems

Chen H, Branavan SRK, Barzilay R, Karger DR (2009) Global models of document structure using latent permutations. In: Conference of the North American Chapter of the Association for Computational Linguistics

Choi FYY, Wiemer-Hastings P, Moore J (2001) Latent semantic analysis for text segmentation. In: Proceedings of Emperical Methods in Natural Language Processing

Cialdini RB (2000) Influence: Science and Practice (4th Edition). Allyn & Bacon

Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3)

Cowans PJ (2006) Probabilistic document modelling. PhD thesis, University of Cambridge

Daley JA, McCroskey JC, Richmond VP (1977) Relationships between vocal activity and perception of communicators in small group interaction. Western Journal of Speech Communicaiton 41(3)

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 World Wide Web Conference, pp 699–708

Davis J, Goadrich M (2006) The relationship between precision-recall and ROC curves. In: Proceedings of the International Conference of Machine Learning

Dowman M, Savova V, Griffiths TL, Kording KP, Tenenbaum JB, Purver M (2008) A probabilistic model of meetings that combines words and discourse features. Transactions on Audio, Speech, and Language Processing 16(7)

Drake BH, Moberg DJ (1986) Communicating influence attempts in dyads: Linguistic sedatives and palliatives. The Academy of Management Review 11(3)

Du L, Buntine WL, Jin H (2010) Sequential latent Dirichlet allocation: Discover underlying topic structures within a document. In: International Conference

on Data Mining

Ehlen P, Purver M, Niekrasz J (2007) A meeting browser that learns. In: Proceedings of the AAAI Spring Symposium on Interaction Challenges for Intelligent Assistants

Eisenstein J, Barzilay R (2008) Bayesian unsupervised topic segmentation. In: Proceedings of Emperical Methods in Natural Language Processing

Emerson RM (1981) Social exchange theory. In: Rosenberg M, Turner RH (eds) Social Psychology: Sociological Perspectives, Basic Books, New York, pp 30–65

Fallows J (2008) Your VP debate wrapup in four bullet points. The Atlantic Http://www.theatlantic.com/technology/archive/2008/10/your-vp-debate-wrapup-in-four-bullet-points/8887/

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. The Annals of Statistics 1(2)

Fleischmann KR, Templeton TC, Boyd-Graber J (2011) Modeling diverse standpoints in text classification: Learning to be human by modeling human values. In: iConference

Foa UG, Foa EB (1972) Resource exchange: Toward a structural theory of interpersonal communication. In: Siegman AW, Pope B (eds) Studies in dyadic communication, Pergamon Press, pp 291–325

Fox EB, Sudderth EB, Jordan MI, Willsky AS (2008) An HDP-HMM for systems with state persistence. In: Proceedings of the International Conference of Machine Learning

Galley M, McKeown K, Fosler-Lussier E, Jing H (2003) Discourse segmentation of multi-party conversation. In: Proceedings of the Association for Computational Linguistics

Galley M, McKeown K, Hirschberg J, Shriberg E (2004) Identifying agreement and disagreement in conversational speech: use of bayesian networks to model pragmatic dependencies. In: Proceedings of the Association for Computational Linguistics

Gardner M, Lutes J, Lund J, Hansen J, Walker D, Ringger E, Seppi K (2010) The topic browser: An interactive tool for browsing topic models. In: Proceedings of Advances in Neural Information Processing Systems

Gershman SJ, Blei DM (2012) A tutorial on Bayesian nonparametric models. Journal of Mathematical Psychology 56(1)

Giles H, Coupland N, Coupland J (1991) Accommodation theory: Communication, context, and consequence. In: Giles H, Coupland N, Coupland J (eds) Contexts of accommodation: Developments in applied socio-linguistics, Cambridge University Press, Cambridge, pp 1–68

Greene S, Resnik P (2009) More than words: Syntactic packaging and implicit sentiment. In: Conference of the North American Chapter of the Association for Computational Linguistics

Grimmer J (2009) A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. Political Analysis 18(1)

Halliday M, Hasan R (1976) Cohesion in English. Longman, New York

Hamilton MA, Hunter JE (1998) The effect of language intensity on receiver evaluations of message, source, and topic. In: Persuasion: Advances through meta-analysis

Hardisty E, Boyd-Graber J, Resnik P (2010) Modeling perspective using adaptor grammars. In: Proceedings of Emperical Methods in Natural Language Processing

Hawes T, Lin J, 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)

Hearst MA (1997) TextTiling: Segmenting text into multi-paragraph subtopic passages. Computational Linguistics 23(1)

Hirschberg J, Litman D (1993) Empirical studies on the disambiguation of cue phrases. Computational Linguistics 19(3)

Hoffman MD, Blei DM, Cook PR (2009) Finding latent sources in recorded music with a shift-invariant hdp. In: Proceedings of the Conference on Digital Audio Effects

Horton S (2008) The Ifill factor. Harpers Magazine URL `http://harpers.org/archive/2008/10/hbc-90003659`

Hsueh Py, Moore JD, Renals S (2006) Automatic segmentation of multiparty dialogue. In: Proceedings of the European Chapter of the Association for Computational Linguistics

Hu Y, Boyd-Graber J, Satinoff B (2011) Interactive topic modeling. In: Proceedings of the Association for Computational Linguistics

Huffaker D (2010) Dimensions of leadership and social influence in online communities. Human Communication Research 36(4):593–617

Hung H, Huang Y, Friedland G, Gatica-Perez D (2011) Estimating dominance in multi-party meetings using speaker diarization. Transactions on Audio, Speech, and Language Processing 19(4)

Ireland ME, Slatcher RB, Eastwick PW, Scissors LE, Finkel EJ, Pennebaker JW (2011) Language style matching predicts relationship initiation and stability. Psychological Science 22(1)

Janin A, Baron D, Edwards J, Ellis D, Gelbart D, Morgan N, Peskin B, Pfau T, Shriberg E, Stolcke A, Wooters C (2003) The ICSI meeting corpus. In: IEEE International Conference on Acoustics, Speech, and Signal Processing

Jayagopi DB, Hung H, Yeo C, Gatica-Perez D (2009) Modeling dominance in group conversations using nonverbal activity cues. Transactions on Audio, Speech, and Language Processing 17(3)

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

Katz E, Lazarsfeld PF (1955) Personal influence: the part played by people in the flow of mass communications. Foundations of communications research, Free Press

Kellermann K (2004) Topical profiling: Emergent, co-occurring, and relationally defining topics in talk. Journal of Language and Social Psychology 23(3)

Marin A, Ostendorf M, Zhang B, Morgan JT, Oxley M, Zachry M, Bender EM (2010) Detecting authority bids in online discussions. In: SLT, pp 49–54

Mast MS (2002) Dominance as Expressed and Inferred Through Speaking Time. Human Communication Research 28(3):420–450

McCombs M, Reynolds A (2009) How the news shapes our civic agenda. In: Bryant J, Oliver MB (eds) Media effects: Advances in theory and research, Lawrence Erlbaum, pp 1–16

Mimno D, Wallach H, Naradowsky J, Smith D, McCallum A (2009) Polylingual topic models. In: Proceedings of Emperical Methods in Natural Language Processing

Mimno DM (2011) Reconstructing pompeian households. In: Proceedings of Uncertainty in Artificial Intelligence, pp 506–513

Morris J, Hirst G (1991) Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computational Linguistics 17

Müller P, Quintana FA (2004) Nonparametric Bayesian data analysis. Statistical Science 19(1)

Murray G, Renals S, Carletta J (2005) Extractive summarization of meeting recordings. In: European Conference on Speech Communication and Technology

Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics 9(2):249–265

Neal RM (2003) Slice sampling. Annals of Statistics 31:705–767

Ng SH, Bradac JJ (1993) Power in language: verbal communication and social influence. Language and language behaviors, Sage Publications

Ng SH, Bell D, Brooke M (1993) Gaining turns and achieving high influence ranking in small conversational groups. British Journal of Social Psychology 32(3):265–275

Nguyen VA, Boyd-Graber J, Resnik P (2012) SITS: A hierarchical nonparametric model using speaker identity for topic segmentation in multiparty conversations. In: Proceedings of the Association for Computational Linguistics

Olney A, Cai Z (2005) An orthonormal basis for topic segmentation in tutorial dialogue. In: Proceedings of the Human Language Technology Conference

Otsuka K, Yamato J, Takemae Y, Murase H (2006) Quantifying interpersonal influence in face-to-face conversations based on visual attention patterns. In: International Conference on Human Factors in Computing Systems

Palmer MT (1989) Controlling conversations: Turns, topics and interpersonal control. Communication Monographs 56(1):1–18

Pang B, Lee L (2008) Opinion Mining and Sentiment Analysis. Now Publishers Inc

Passonneau RJ, Litman DJ (1997) Discourse segmentation by human and automated means. Computational Linguistics 23(1)

Paul M, Girju R (2010) A two-dimensional topic-aspect model for discovering multi-faceted topics. In: Association for the Advancement of Artificial Intelligence

Paul R (2007) Political power and the rule of law. Texas Straight Talk URL
    http://www.lewrockwell.com/paul/paul366.html

Pele O, Werman M (2008) A linear time histogram metric for improved sift
    matching. In: ECCV, pp 495–508

Pele O, Werman M (2009) Fast and robust earth mover's distances. In: Inter-
    national Conference on Computer Vision

Pevzner L, Hearst MA (2002) A critique and improvement of an evaluation
    metric for text segmentation. Computational Linguistics 28

Planalp S, Tracy K (1980) Not to change the topic but ...: A cognitive approach
    to the managemen of conversations. In: Communication yearbook 4, New
    Brunswick, pp 237–258

Purver M (2011) Topic segmentation. In: Spoken Language Understanding:
    Systems for Extracting Semantic Information from Speech

Purver M, Körding K, Griffiths TL, Tenenbaum J (2006) Unsupervised topic
    modelling for multi-party spoken discourse. In: Proceedings of the Association
    for Computational Linguistics

Regula R, Julian W (1973) The impact of quality and frequency of task
    contributions on perceived ability. Journal of Social Psychology 89(1):115–
    122

Reid SA, Ng SH (2000) Conversation as a resource for influence: evidence for
    prototypical arguments and social identification processes. European Journal
    of Social Psychology 30(1):83–100

Ren L, Dunson DB, Carin L (2008) The dynamic hierarchical Dirichlet process.
    In: Proceedings of the International Conference of Machine Learning

Resnik P, Hardisty E (2010) Gibbs sampling for the uniniti-
    ated. Tech. Rep. UMIACS-TR-2010-04, University of Maryland,
    http://drum.lib.umd.edu//handle/1903/10058

Rienks R, Heylen D (2005) Dominance detection in meetings using easily ob-
    tainable features. In: Proceedings of the 2nd Joint Workshop on Multimodal
    Interaction and Related Machine Learning Algorithms

Rienks R, Zhang D, Gatica-Perez D, Post W (2006) Detection and applica-
    tion of influence rankings in small group meetings. In: Proceedings of the
    International Conference on Multimodal Interfaces, ICMI '06

Rubner Y, Tomasi C, Guibas LJ (2000) The earth mover's distance as a metric
    for image retrieval. International Journal of Computer Vision 40(2)

Sayeed AB, Boyd-Graber J, Rusk B, Weinberg A (2012) Grammatical struc-
    tures for word-level sentiment detection. In: North American Association of
    Computational Linguistics

Scheer LK, Stern LW (1992) The effect of influence type and performance
    outcomes on attitude toward the influencer. Journal of Marketing Research
    29(1):pp. 128–142

Schlenker BR, Nacci P, Helm B, Tedeschi JT (1976) Reactions to coercive and
    reward power: The effects of switching influence modes on target compliance.
    Sociometry 39(4):316–323

Sorrentino RM, Boutiller RG (1972) The effect of quantity and quality of verbal
    interaction on ratings of leadership ability. Journal of Experimental Social

Psychology (5):403–411

Stang DJ (1973) Effect of Interaction Rate on Ratings of Leadership and Liking. Journal of Personality and Social Psychology 27(3):405–408

Teh YW (2006) A hierarchical Bayesian language model based on Pitman-Yor processes. In: Proceedings of the Association for Computational Linguistics

Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. Journal of the American Statistical Association 101(476):1566–1581

Thomas M, Pang B, Lee L (2006) Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In: Proceedings of Emperical Methods in Natural Language Processing

Trammell KD, Keshelashvili A (2005) Examining the new influencers: A self-presentation study of a-list blogs. Journalism & Mass Communication Quarterly 82(4):968–982

Tur G, Stolcke A, Voss L, Peters S, Hakkani-Tür D, Dowding J, Favre B, Fernández R, Frampton M, Frandsen M, Frederickson C, Graciarena M, Kintzing D, Leveque K, Mason S, Niekrasz J, Purver M, Riedhammer K, Shriberg E, Tien J, Vergyri D, Yang F (2010) The CALO meeting assistant system. Transactions on Audio, Speech, and Language Processing 18(6)

Wallach HM (2006) Topic modeling: Beyond bag-of-words. In: Proceedings of the International Conference of Machine Learning

Wallach HM (2008) Structured topic models for language. PhD thesis, University of Cambridge

Wang C, Blei DM, Heckerman D (2008) Continuous time dynamic topic models. In: Proceedings of Uncertainty in Artificial Intelligence

Weimann G (1994) The Influentials: People Who Influence People. Suny Series in Human Communication Processes, State University of New York Press

Zhai K, Boyd-Graber J, Asadi N, Alkhouja M (2012) Mr. LDA: A flexible large scale topic modeling package using variational inference in mapreduce. In: Proceedings of World Wide Web Conference

Zhang D, Gatica-Perez D, Bengio S, Roy D (2005) Learning influence among interacting Markov chains. In: Proceedings of Advances in Neural Information Processing Systems

A Derivation of Sampling Equations

A.1 Nonparametric SITS

In this section, we describe the general Gibbs sampler for our nonparametric model, without using the minimal or maximal path assumption (Cowans, 2006; Wallach, 2008). The state space of our chain consists of the topic indices assigned to all tokens $z = \{z_{c,t,n}\}$ and topic shift indicators assigned to all turns $l = \{l_{c,t}\}$. In order to obtain $z_{c,t,n}$ we need to know the *path* assigned for token $w_{c,t,n}$ through the hierarchy which includes $k^{\mathcal{T}}_{c,t,n}$, $k^{\mathcal{S}}_{c,s,j}$ and $k^{\mathcal{C}}_{c,i}$. For ease of reference, the meaning of these symbols (and others used in this Appendix) are listed in Table 10. Figure 7b shows the relationship among the latent variables in our model. Once we know the three seating assignments $k^{\mathcal{T}}_{c,t,n}$, $k^{\mathcal{S}}_{c,s,j}$ and $k^{\mathcal{C}}_{c,i}$, $z_{c,t,n}$ can obtained by

$$z_{c,t,n} \equiv k^{\mathcal{C}}_{c,k^{\mathcal{S}}_{c,s_t,k^{\mathcal{T}}_{c,t,n}}} \tag{14}$$

To perform inference, we marginalize over all other latent variables and alternate between sampling paths z and the sampling topic shift indicators l.

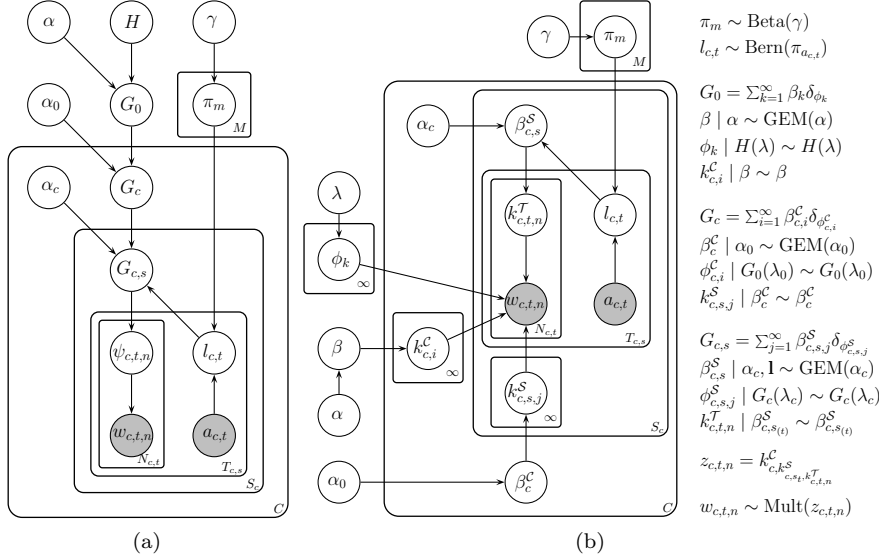(a)   (b)

Fig. 7: Plate diagram representations of our nonparametric model: (a) representation using the notion of *segment*. We use S_c to denote the number of segments in conversation c and $T_{c,t}$ to denote the number of turns in segment s of conversation c; (b) representation where explicit path assignments are shown.

A.1.1 Sampling topic assignments

Before deriving the sampling equations, let us use $f_k^{-c,t,n}(w_{c,t,n})$ to denote the conditional density of token $w_{c,t,n}$ under topic k given all other items except $w_{c,t,n}$

$$f_k^{-c,t,n}(w_{c,t,n}) = \frac{\int_{\phi_k} P(\boldsymbol{w} \mid \phi_k) P(\phi_k \mid \lambda) \, \mathrm{d}\phi_k}{\int_{\phi_k} P(\boldsymbol{w}_{-c,t,n} \mid \phi_k) P(\phi_k \mid \lambda) \, \mathrm{d}\phi_k} = \begin{cases} \dfrac{M_{k,w_{ctn}} + \lambda}{M_{k,\cdot} + V\lambda}, & \text{if } k \text{ exists;} \\ \dfrac{1}{V}, & \text{if } k \text{ is new.} \end{cases} \quad (15)$$

where $M_{k,w}$ is the number of times word type w assigned to topic k, $M_{k,\cdot}$ represents marginal count, and superscript $_{-c,t,n}$ denotes the same count excluding $w_{c,t,n}$.

To sample the path for a token (i.e. a customer), we take a similar approach to the first sampling method described in Teh et al (2006). We first sample a segment-level table $k_{c,t,n}^{\mathcal{T}}$ for customer $w_{c,t,n}$. The customer can sit at an existing table j or create a new one j^{new}. If a new segment-level table is created, a table in conversation-level restaurant c will be sampled. Again, this j^{new} can be assigned to an existing conversation-level table i or a new one i^{new}. If i^{new} is sampled, it can be assigned to an existing dish k or a new one k^{new} in the corpus-level restaurant. The details of each step is as follows:

Notation	Descriptions
H	Base distribution over topics (a symmetric Dirichlet distribution)
G_0	Distribution over topics drawn from H and is shared across all conversations
G_c	Distribution over topics drawn from G_0 for each conversation c and is shared across all turns in c
$G_{c,t}$	Distribution over topics drawn from G_c for turn t in conversation c
$G_{c,s}$	Distribution over topics drawn from G_c for segment s in conversation c
ϕ_k	k^{th} multinomial distribution over words (i.e. the kth topic)
$\psi_{c,t,n}$	Multinomial distribution over words generating token $w_{c,t,n}$ ($\psi_{c,t,n} \equiv \phi_{z_{c,t,n}}$)
$l_{c,t}$	Topic shift indicator assigned to turn t of conversation c
π_m	Topic shift tendency of speaker m
$w_{c,t,n}$	n^{th} token of turn t of conversation c
$a_{c,t}$	Speaker of turn t of conversation c
s_t	Index of the segment that contains turn t
$N_{c,t}$	Number of tokens in turn t of conversations c
$T_{c,s}$	Number of turns in segment s of conversation c
T_c	Number of turns in conversation c
S_c	Number of segments in conversation c
C	Number of conversations
$N^{\mathcal{S}}_{c,s,j}$	Number of customers sitting at table j in segment-level restaurant c,s
$N^{\mathcal{C}}_{c,i}$	Number of segment-level tables in restaurant c assigned to conversation-level table i
N_k	Number of conversation-level tables assigned to corpus-level table k
$k^{\mathcal{T}}_{c,t,n}$	Index of table assigned to token $w_{c,t,n}$
$k^{\mathcal{S}}_{c,s,j}$	Index of conversation-level table assigned to table j in segment-level restaurant c,s
$k^{\mathcal{C}}_{c,i}$	Index of corpus-level table assigned to table i in conversation-level restaurant c
$J^{\mathcal{S}}_{c,s}$	Number of tables in segment-level restaurant c,s
$I^{\mathcal{C}}_c$	Number of tables in conversation-level restaurant c
K^+	Number of tables in corpus-level restaurant
$*$	The set of all hyperparameters
$\cdot$	This notation is used to denote the marginal count

Table 10: List of notations used.

Sampling $k^{\mathcal{T}}_{c,t,n}$: The probability of assigning a segment-level table j to token $w_{c,t,n}$ is

$$P(k^{\mathcal{T}}_{c,t,n} = j \mid \boldsymbol{k}^{\mathcal{T}}_{-c,t,n}, \boldsymbol{k}^{\mathcal{S}}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, \boldsymbol{w}, *)$$

$$\propto P(k^{\mathcal{T}}_{c,t,n} = j \mid \boldsymbol{k}^{\mathcal{T}}_{-c,t,n}) P(w_{c,t,n} \mid k^{\mathcal{T}}_{c,t,n}, \boldsymbol{k}^{\mathcal{T}}_{-c,t,n}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{S}}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *)$$

$$= \begin{cases} \dfrac{N^{\mathcal{S}}_{c,s_t,j}}{N^{\mathcal{S}}_{c,s_t,\cdot} + \alpha_c} f^{-c,t,n}_{k^{\mathcal{C}}_{c,k^{\mathcal{S}}_{c,s_t,j}}}(w_{c,t,n}), & \text{if } j \text{ exists;} \\[3ex] \dfrac{\alpha_c}{N^{\mathcal{S}}_{c,s_t,\cdot} + \alpha_c} P(w_{c,t,n} \mid k^{\mathcal{T}}_{c,t,n} = j^{new}, \boldsymbol{k}^{\mathcal{T}}_{-c,t,n}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{S}}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *), & \text{if } j \text{ is new.} \end{cases} \quad (16)$$

Marginalizing out all assignments of to $k^{\mathcal{S}}_{c,s_t,j^{new}}$ (i.e. all possible tables i's including the new table i^{new} of conversation-level restaurant c), we have:

$$P(w_{c,t,n} \mid k^{\mathcal{T}}_{c,t,n} = j^{new}, \boldsymbol{k}^{\mathcal{T}}_{-c,t,n}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{S}}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *)$$

$$= \sum_{i=1}^{I^{\mathcal{C}}_c} \frac{N^{\mathcal{C}}_{c,i}}{N^{\mathcal{C}}_{c,\cdot} + \alpha_0} f^{-c,t,n}_{k^{\mathcal{C}}_{c,i}}(w_{c,t,n})$$

$$+ \frac{\alpha_0}{N^{\mathcal{C}}_{c,\cdot} + \alpha_0} P(w_{c,t,n} \mid k^{\mathcal{S}}_{c,s_t,j^{new}} = i^{new}, \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *) \qquad (17)$$

Again, marginalizing out all possible assignments of $k^{\mathcal{C}}_{c,i^{new}}$ (i.e. all possible global dishes k's including a new k^{new}), we have:

$$P(w_{c,t,n} \mid k^{\mathcal{S}}_{c,s_t,j^{new}} = i^{new}, \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *)$$

$$= \sum_{k=1}^{K} \frac{N_k}{N_{\cdot} + \alpha} f^{-c,t,n}_k(w_{c,t,n}) + \frac{\alpha}{N_{\cdot} + \alpha} f^{-c,t,n}_{k^{new}}(w_{c,t,n}) \qquad (18)$$

Sampling $k^{\mathcal{S}}_{c,s,j}$: When a new segment-level table j^{new} is created after sampling $k^{\mathcal{T}}_{c,t,n}$, we need to assign it to a table in the conversation-level restaurant. The probability of assigning it to conversation-level table i is

$$P(k^{\mathcal{S}}_{c,s_t,j^{new}} = i \mid \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{w}, \boldsymbol{l}, *)$$

$$\propto P(k^{\mathcal{S}}_{c,s_t,j^{new}} = i \mid \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}) P(w_{c,t,n} \mid k^{\mathcal{S}}_{c,s_t,j^{new}} = i^{new}, \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *)$$

$$= \begin{cases} \dfrac{N^{\mathcal{C}}_{c,i}}{N^{\mathcal{C}}_{c,\cdot} + \alpha_0} f^{-c,t,n}_{k^{\mathcal{C}}_{c,i}}(w_{c,t,n}), & \text{if } i \text{ exists;} \\[2ex] \dfrac{\alpha_0}{N^{\mathcal{C}}_{c,\cdot} + \alpha_0} P(w_{c,t,n} \mid k^{\mathcal{S}}_{c,s_t,j^{new}} = i^{new}, \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *), & \text{if } i \text{ is new.} \end{cases} \qquad (19)$$

The value of $P(w_{c,t,n} \mid k^{\mathcal{S}}_{c,s_t,j^{new}} = i^{new}, \boldsymbol{k}^{\mathcal{S}}_{-c,s_t,j^{new}}, \boldsymbol{w}_{-c,t,n}, \boldsymbol{k}^{\mathcal{C}}, \boldsymbol{l}, *)$ can be obtained in Equation 18.

Sampling $k^{\mathcal{C}}_{c,i}$: When a new conversation-level table i^{new} is created, a table in the corpus-level restaurant is sampled using the following probabilities

$$P(k^{\mathcal{C}}_{c,i^{new}} \mid \boldsymbol{k}^{\mathcal{C}}_{-c,i^{new}}, \boldsymbol{w}, \boldsymbol{l}, *) \propto \begin{cases} \dfrac{N_k}{N_{\cdot} + \alpha} f^{-c,t,n}_k(w_{c,t,n}), & \text{if } k \text{ exists;} \\[2ex] \dfrac{\alpha}{N_{\cdot} + \alpha} f^{-c,t,n}_{k^{new}}(w_{c,t,n}), & \text{if } k \text{ is new.} \end{cases} \qquad (20)$$

By having all the seating assignments, we obtain the topic assignment $z_{c,t,n}$ for every token $w_{c,t,n}$ by using Equation 14.

A.1.2 Sampling topic shift indicators

Given all the path assignments for all token $w_{c,t,n}$, we will sample the topic shift $l_{c,t}$ for every turn t of conversation c. This probability is given as

$$P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{k}^{\mathcal{T}}, \boldsymbol{w}, \boldsymbol{a}, *) \propto P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{a}, *) \cdot P(\boldsymbol{k}^{\mathcal{T}}_{c,t} \mid \boldsymbol{k}^{\mathcal{T}}_{-c,t}, l_{c,t}, \boldsymbol{l}_{-c,t}, *) \qquad (21)$$

Computing $P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{a}, *)$: Recall that the topic shifts $\boldsymbol{l} = \{l_{c,t}\}$ are drawn from a Bernoulli distribution parameterized by the topic shift tendency π, which is drawn from a conjugate prior Beta(γ). Marginalizing out π we have

$$P(\mathbf{l}) = \int_0^1 P(\mathbf{l} \mid \pi) P(\pi; \gamma)\, \mathrm{d}\pi = \prod_{m=1}^{M} \frac{\Gamma(2\gamma)}{\Gamma(\gamma)^2} \frac{\Gamma(S_{m,1} + \gamma)\Gamma(S_{m,0} + \gamma)}{\Gamma(S_{m,\cdot} + 2\gamma)} \tag{22}$$

When $l_{c,t} = 0$, the counts of number of times being assigned topic shift of value 1 for all speakers will remain unchanged. Similarly, for the case of $l_{c,t} = 1$. Thus, we have $P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{a}, *)$ for the two cases following Resnik and Hardisty (2010) as follows

$$P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{a}, *) = \frac{P(\mathbf{l} \mid \boldsymbol{a}, *)}{P(\boldsymbol{l}_{-c,t} \mid \boldsymbol{a}, *)} \propto \begin{cases} \dfrac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma}, \text{ if } l_{c,t} = 0 \\[2ex] \dfrac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma}, \text{ if } l_{c,t} = 1 \end{cases} \tag{23}$$

where $S_{a,x}^{-c,t}$ denotes the number of times speaker a is assigned topic shift of value $x \in \{0,1\}$ excluding $l_{c,t}$.

Computing $P(\boldsymbol{k}_{c,t}^{\mathcal{T}} \mid \boldsymbol{k}_{-c,t}^{\mathcal{T}}, l_{c,t}, \boldsymbol{l}_{-c,t}, *)$:

$$P(\boldsymbol{k}_{c,t}^{\mathcal{T}} \mid \boldsymbol{k}_{-c,t}^{\mathcal{T}}, l_{c,t}, \boldsymbol{l}_{-c,t}, *) \propto \frac{P(\boldsymbol{k}_{c,t}^{\mathcal{T}}, \boldsymbol{k}_{-c,t}^{\mathcal{T}} \mid l_{c,t}, \boldsymbol{l}_{-c,t}, *)}{P(\boldsymbol{k}_{-c,t}^{\mathcal{T}} \mid l_{c,t}, \boldsymbol{l}_{-c,t}, *)} = \frac{P(\boldsymbol{k}^{\mathcal{T}} \mid l_{c,t}, \boldsymbol{l}_{-c,t}, *)}{P(\boldsymbol{k}_{-c,t}^{\mathcal{T}} \mid l_{c,t}, \boldsymbol{l}_{-c,t}, *)} \tag{24}$$

Given all the customers assigned to all tables, the joint probability of all tables (Gershman and Blei, 2012) is

$$P(\boldsymbol{k}^{\mathcal{T}} \mid \boldsymbol{l}) = \prod_{c=1}^{C} \prod_{s=1}^{S_c} \frac{\alpha_c^{J_{c,s}^{\mathcal{S}}} \prod_{j=1}^{J_{c,s}^{\mathcal{S}}} (N_{c,s,j}^{\mathcal{S}} - 1)!}{\prod_{x=1}^{N_{c,s,\cdot}^{\mathcal{S}}} (x - 1 + \alpha_c)} \tag{25}$$

where S_c denotes the number of segments in restaurant c. Substituting Equation 25 to Equation 24, we have:
$$P(\boldsymbol{k}_{c,t}^{\mathcal{T}} \mid l_{c,t}, \boldsymbol{l}_{-c,t}, \boldsymbol{k}_{-c,t}^{\mathcal{T}})$$

$$\propto \begin{cases} \dfrac{\alpha_c^{J_{c,s_t}^{\mathcal{S},0}} \prod_{j=1}^{J_{c,s_t}^{\mathcal{S},0}} (N_{c,s_t,j}^{\mathcal{S},0} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{\mathcal{S},0}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 0 \\[3ex] \dfrac{\alpha_c^{J_{c,s_t-1}^{\mathcal{S},1}} \prod_{j=1}^{J_{c,s_t-1}^{\mathcal{S},1}} (N_{c,s_t-1,j}^{\mathcal{S},1} - 1)!}{\prod_{x=1}^{N_{c,s_t-1,\cdot}^{\mathcal{S},1}} (x - 1 + \alpha_c)} \dfrac{\alpha_c^{J_{c,s_t}^{\mathcal{S},1}} \prod_{j=1}^{J_{c,s_t}^{\mathcal{S},1}} (N_{c,s_t,j}^{\mathcal{S},1} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{\mathcal{S},1}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 1 \end{cases} \tag{26}$$

where

- $J_{c,s_t}^{\mathcal{S},x}$ denotes the number of tables during segment s of restaurant c if $l_{c,t} = x$.
- $N_{c,s_t,j}^{\mathcal{S},x}$ denotes the number of customers sitting at table j during segment s of restaurant c if $l_{c,t} = x$. The marginal count $N_{c,s_t,\cdot}^{\mathcal{S},x}$ denotes the total number of customers during segment s of restaurant c.

Combining Equations 23 and 26, we have the sampling equation for topic shift indicator as follow

$$P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{k}^{\mathcal{T}}, \boldsymbol{w}, \boldsymbol{a}, *)$$

$$
\propto
\begin{cases}
\dfrac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \dfrac{\alpha_c^{J_{c,s_t}^{\mathcal{S},0}} \prod_{j=1}^{J_{c,s_t}^{\mathcal{S},0}} (N_{c,s_t,j}^{\mathcal{S},0} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{\mathcal{S},0}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 0 \\[3ex]
\dfrac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \dfrac{\alpha_c^{J_{c,s_t}^{\mathcal{S},1}-1} \prod_{j=1}^{J_{c,s_t}^{\mathcal{S},1}-1} (N_{c,s_t-1,j}^{\mathcal{S},1} - 1)!}{\prod_{x=1}^{N_{c,s_t-1,\cdot}^{\mathcal{S},1}} (x - 1 + \alpha_c)} \dfrac{\alpha_c^{J_{c,s_t}^{\mathcal{S},1}} \prod_{j=1}^{J_{c,s_t}^{\mathcal{S},1}} (N_{c,s_t,j}^{\mathcal{S},1} - 1)!}{\prod_{x=1}^{N_{c,s_t,\cdot}^{\mathcal{S},1}} (x - 1 + \alpha_c)}, & \text{if } l_{c,t} = 1
\end{cases}
$$

$$(27)$$

Thus, using Equations 16, 19 and 20 to obtain the topic assignment $z_{c,t,n}$ as in Equation 14 and using Equation 27 to obtain the topic shift indicator $l_{c,t}$, we complete the derivation of sampling equations for our Gibbs samplers.

A.2 Parametric SITS

Similar to the nonparametric version, the state space of the Markov chain includes the topic indices assigned to all tokens $\boldsymbol{z} = \{z_{c,t,n}\}$ and topic shifts assigned to all turns $\boldsymbol{l} = \{l_{c,t}\}$. Here, we present the sampling equations for both variables.

A.2.1 Sampling topic assignments

$$P(z_{c,t,n} = y \mid \boldsymbol{z}_{-c,t,n}, \boldsymbol{l}, \boldsymbol{w}, *)$$

$$
\begin{aligned}
&= \frac{P(z_{c,t,n} = y, \boldsymbol{z}_{-c,t,n}, \boldsymbol{l}, \boldsymbol{w})}{P(\boldsymbol{z}_{-c,t,n}, \boldsymbol{l}, \boldsymbol{w})} \\
&= \frac{P(\boldsymbol{w} \mid \boldsymbol{z})P(\boldsymbol{z} \mid \boldsymbol{l})P(\boldsymbol{l})}{P(\boldsymbol{w} \mid \boldsymbol{z}_{-c,t,n})P(\boldsymbol{z}_{-c,t,n} \mid \boldsymbol{l})P(\boldsymbol{l})} = \frac{P(\boldsymbol{w} \mid \boldsymbol{z})}{P(\boldsymbol{w} \mid \boldsymbol{z}_{-c,t,n})} \frac{P(\boldsymbol{z} \mid \boldsymbol{l})}{P(\boldsymbol{z}_{-c,t,n} \mid \boldsymbol{l})} \\
&\propto \frac{M_{y,w_{c,t,n}}^{-c,t,n} + \beta}{M_{y,\cdot}^{-c,t,n} + V\beta} \cdot \frac{N_{c,s_t,y}^{-c,t,n} + \alpha}{N_{c,s_t,\cdot}^{-c,t,n} + K\alpha}
\end{aligned}
$$

$$(28)$$

where

- $M_{k,w}$ denotes the number of times that topic k is assigned to token w in the vocabulary.
- $N_{c,s,k}$ denotes the number of times topic k is assigned to segment s of conversation c.
- V denotes the size of the vocabulary.
- K denotes the number of predefined topics.

A.2.2 Sampling topic shifts

$$P(l_{c,t} \mid \boldsymbol{l}_{-c,t}, \boldsymbol{z}, \boldsymbol{w}, \boldsymbol{a}, *)$$

$$
\propto
\begin{cases}
\dfrac{S_{a_{c,t},0}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \cdot \dfrac{\prod_{k=1}^{K} \Gamma(N_{c,s_t,k}^{0} + \alpha)}{\Gamma(N_{c,s_t,\cdot}^{0} + K\alpha)}, & \text{if } l_{c,t} = 0 \\[3ex]
\dfrac{S_{a_{c,t},1}^{-c,t} + \gamma}{S_{a_{c,t},\cdot}^{-c,t} + 2\gamma} \cdot \dfrac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \dfrac{\prod_{k=1}^{K} \Gamma(N_{c,s_t-1,k}^{1} + \alpha)}{\Gamma(N_{c,s_t-1,\cdot}^{1} + K\alpha)} \dfrac{\prod_{k=1}^{K} \Gamma(N_{c,s_t,k}^{1} + \alpha)}{\Gamma(N_{c,s_t,\cdot}^{1} + K\alpha)}, & \text{if } l_{c,t} = 1
\end{cases}
$$

$$(29)$$

where

- $S_{m,x}$ denotes the number of times that topic shift of value x is assigned to speaker m.
- $N_{c,t,k}^{x}$ denotes the number of times that topic k is assigned to segment s of conversation c if $l_{c,t} = x$.

B Crossfire Annotation

Identifying Influencers: A discussant was considered an influencer if he or she initiated a topic shift that steered the conversation in a different direction, convinced others to agree to a certain viewpoint, or used an authoritative voice that caused others to defer to or reference that person's expertise. A discussant was not identified as an influencer if he or she merely initiated a topic at the start of a conversation, did not garner any support from others for the points he or she made, or was not recognized by others as speaking with authority.

Hosts: The two hosts on Crossfire are tasked with changing the course of the discussion, as defined by their role on the program. Therefore, the hosts were considered annotated as influencers when they shifted the direction of the conversation. As a result, the hosts were generally included as influencers. To address this issue, up to four influencers could be identified within an interaction; then once all influencers were identified based on the interaction, the hosts were tagged within the influencers identified. In previous annotations, the number of influencers was limited to two per interaction. Allowing for up to four influencers and then tagging the hosts after all the influencers were identified allowed for guests to be tagged as influencers; otherwise they may have been left off because of being dominated by the roles of the hosts.

Polite Agreement: Crossfire is a television show in which debate is the format and goal of the interaction, with the two hosts guiding the conversation. Certain patterns emerge from this type of debate, including some normative politeness that involves acknowledgement of another's point-such as polite agreement-but then making a shift in a way that counters that agreement. For example, when a discussant responded to another person by initially conceding a small point only to argue a larger point, that agreement was considered an argumentation tactic and not an expression of actual agreement; therefore, the initial agreement was not tagged as an indicator of an influencer.

- PRESS: Better marketing. Frank Gaffney, let's start by looking at those two polls back to back. The first one was a Gallup poll, which was December/January, released just about 10 days ago that surveyed residents of nine Muslim countries asking the question, do you think favorably about the U.S. or unfavorably, showed 22 percent only favorable opinion of the United States. 53 percent unfavorable. And then just a couple of days ago, CNN conducted its own poll here in the United States of Americans. Our opinion of Muslim nations, favorable opinion, 24 percent, unfavorable 41 percent. They're the mirror image of each other. So before you and I maybe tango about what we do about it, can we agree that this is a serious problem and we ought to be doing something about the spin war while we're fighting the military war?
- GAFFNEY: I think we can agree to that. It is a serious problem. I think the genesis of this problem, however, was not September 11. It certainly wasn't what we've been doing since September 11. I think it has its roots in fundamentally what most of the people of these countries have been receiving in the way of propaganda from government-controlled media, al-Jazeera and similar kinds of outlets that frankly, given the vitriol against America that is pumped out by these media sources almost 24/7, it's surprising the poll numbers aren't worse for the United States in most of those countries.

Influencer: NONE.

Reason: GAFFNEY agrees on a small point in preparation to make a larger argument against what PRESS is claiming. GAFFNEY agrees that there is a problem with US-Arab relations, but disagrees with PRESS that the genesis of the problem was 9/11.

Video Segments: When discussants in the transcripts appeared in video clips, they were not considered as potential influencers; instead, video clips were treated as supporting material used, usually by the host, to make a point. Therefore, if a host showed a video that supported an argument and someone else agreed with the argument made in the video, then the influencer would be the host who introduced it.

- NOVAK: Congressman Menendez again, my new best friend, the former President Clinton, was asked on "LARRY KING" last night, all this stuff about 16 words, 10 words. He was asked whether biological and chemical weapons in Iraq when he was president, just a short time ago. Let's listen to what he said.
- CLINTON: When I left office there was a substantial amount of biological and chemical material unaccounted for.
- NOVAK: That was from "LARRY KING LIVE" last night. I mean, he is saying that there was material then. That's what the issue is. It isn't the issue of this parsing words and clipping off like Joe McCarthy does, the front side of the...
- PRYCE: ...that's exactly.

Influencer: NOVAK.

Reason: NOVAK is the influencer because he presents a recording of CLINTON to make his point, and PRYCE supports the argument in the video.

However, below is an example where the guest expressed agreement with what was said in the recording but disagreed with the argument the host was trying to make with the video. In this case, the guest provided a different interpretation of the video clip than the host had proposed.

- NOVAK: Senator Murray, welcome. When John Ashcroft, a former senator, was appointed – nominated by President- elect Bush, my Senate sources said if the vote were taken right then, there would be more than 10 votes against him. How did it get up to 42 votes against him? I think the code was broken today by the Senate – the chairman of the Senate Judiciary Committee. And let's listen to Orrin Hatch and how he explains it.
- SEN. ORRIN HATCH (R-UT), CHAIRMAN, JUDICIARY COMMITTEE: This is one of the worst episodes in the onward trashing soldiers routine that I've ever seen. Outside groups basically cowing United States senators, telling them that they all have primaries and that they'll do everything they can to defeat them unless they send this message to President Bush.
- NOVAK: Isn't that right? You and your companions buckled under to the vast left-wing conspiracy. ***
- DREIER: I don't think it means that all, Bill. I mean, if you look at what has taken place here, it's very obvious to me that, as Orrin Hatch said, there are a wide range of groups that led the American people to say the things that they did to my colleague Patty Murray, to me, and a wide range of others. And I just happen to believe that it's a mischaracterization of the record of John Ashcroft. John Ashcroft is someone who, as we all know, was the chairman of the National Governor's Association. He was the chairman of the Attorney General's Association, selected by his peers to that position. They tried to paint him as some kind of extremist, and I don't believe that George Bush will be appointing nominees to the United States Supreme Court or to any other spots who are deserving of the kind of criticism that was leveled at John Ashcroft. So, it's true, Bill, that these groups have gotten out there and successfully convinced many that John Ashcroft is something other than what he is. And the statement that Bob just put up on to the screen, that was made during the hearings and I listened to the hearings, I mean, it's very clear that John Ashcroft offered very thoughtful answers to the questions and I believe was very deserving of broader support than he got.

Influencer: NOVAK.

Reason: NOVAK is the influencer because he presents a recording of SEN. ORRIN HATCH, and DREIER supports the argument in the video about why Ashcroft faced great criticism.

Shift In Topic: At times, guests were able to shift the topic effectively, but not for too long because the hosts would jump in and shift the topic back or in a different direction. Nonetheless, this effort to shift the topic was viewed as a topic shift, as in the following example.

- NOVAK: Well, senator, you have been a very successful electoral politician, and you've never found the necessity to move to the center. But don't you think very possibly that Al Gore, assuming he's going to be nominated, is going to have to get away from this extremist politics, and try to get – at least pose as a centrist? Don't you feel that...

– BOXER: Bobby, Bob, thank you for saying I'm a successful politician. But I have to say to you this. As much as you would like to make these campaigns about left and right, and actually you do that on this show all the time, so I hate to undercut you...
– NOVAK: Yes, we do.
– BOXER: ... it's really about what people want, Bob. It's about opportunity, education, a good economy, a decent minimum wage. When George Bush gets up and defends – and I'd like to see this, both Paul and I – a $3.35 minimum wage, I want to see what people think. Is this compassionate? So these are the issues we have to deal with.
– NOVAK: Well, let me – let me – let me try – let me try...
– BOXER: It's not...
– WELLSTONE: Do you think, Bob, in a cafe in Minnesota, anybody ever comes up to me and says, Paul, are you left, are you right, or are you center?
– NOVAK: Well, they know where you are, they know where you are, Paul.
– WELLSTONE: They – no, no, no, no, no. They talk very concrete issues, about jobs, about their children and education...
– NOVAK: Minnesota's a funny place.
– WELLSTONE: ... about health care. So do people in the country.
– NOVAK: Senator Boxer, let me try something else.
– BOXER: Try it again, Bob.
– WELLSTONE: Yes, try it, we're ready, we're ready. Go ahead and try it, we're ready.

Influencer: BOXER.

Reason: BOXER is an influencer because she changes the topic and challenges the premise of NOVAK's question about right versus left, and WELLSTONE joins her in challenging NOVAK on this point.