

# Homogeneity-Based Transmissive Process To Model True and False News in Social Networks

Jooyeon Kim<sup>1</sup>, Dongkwan Kim<sup>1</sup>, and Alice Oh<sup>1</sup>

<sup>1</sup>KAIST, jooyeon.kim@kaist.ac.kr, dongkwan.kim@kaist.ac.kr, alice.oh@kaist.edu

## Abstract

An overwhelming number of true and false news stories are posted and shared in social networks, and users diffuse the stories based on multiple factors. Diffusion of news stories from one user to another depends not only on the stories' content and the genuineness but also on the alignment of the topical interests between the users. In this paper, we propose a novel Bayesian nonparametric model that incorporates *homogeneity* of news stories as the key component that regulates the topical similarity between the posting and sharing users' topical interests. Our model extends hierarchical Dirichlet process to model the topics of the news stories and incorporates Bayesian Gaussian process latent variable model to discover the homogeneity values. We train our model on a real-world social network dataset and find homogeneity values of news stories that strongly relate to their labels of genuineness and their contents. Finally, we show that the supervised version of our model predicts the labels of news stories better than the state-of-the-art neural network and Bayesian models.

## 1 Introduction

In recent years, users in social media and microblogging platforms have played major roles in disseminating online content and news stories [67, 22]. While some are based on factual information, a non-negligible portion of the news contains false information that often stir unnecessary disputes, sway people's opinions, give rise to political polarization, and even cause financial losses in the stock market [28, 5, 46, 6, 52].

In reaction to this, meaningful academic progress has been made to uncover the spreading pattern of true and false news in social networks [59, 71, 32], detect misinformation from content, network, and temporal information [31, 38, 39, 40, 73, 53], early-fact-check and curtail the spread of falsity [27, 19, 65, 4], build fact-checking systems [47, 66], and investigate the potential intervention of user biases in the process of fact-checking [3]. Also, recent work by Vosoughi *et al.* conducts an in-depth investigation into the differences between true and false information in the aspects of *topics*, *cascade patterns* and sizes, diffusion speeds, emotion, and sentiment [67]. Another recent paper by Del Vicario *et al.* suggests that *homogeneity* among users is the main driving factor of the dissemination of both true and false news [12].

Motivated by such progress, we propose Homogeneity-Based Transmissive Process (HBTP), a Bayesian nonparametric model that captures the complex interplay between the textual content, topics of true and false news, and the topical interests of users who share the news in social networks. Specifically, our model is operationalized as follows:

1. Each user is assigned *user interest* which is computed from the collection of content information of news stories that the user shares.
2. Each news story is assigned a *homogeneity index* which regulates the degree of homogeneity. We define this document-centric (rather than user-centric) homogeneity index to be the degree of uniformness of user interests among users who share the story.

Under our modeling assumption, a user’s topical interest *transmits* to another user through the process of them co-sharing a news story in a social network, and the degree of transmission between users depends on the homogeneity index of the news story that the users co-share. Thus, suppose there are two news stories, one with a high homogeneity index and the other with a low index. The news story with the high index will propagate through a highly biased subgroup within a social network, while propagation of the news story with low homogeneity index will look more like a random walk.

To formulate such modeling assumptions we design our HBTP model as follows:

- (a) Within the nonparametric topic modeling framework [60], we model the content of news stories from the topical interests of users who participate in diffusing the news story.
- (b) We formulate each user’s topical interest as a single probability measure in two-layer Dirichlet processes [18] and allow it to be drawn not only from the single, upper-level probability measure but also from multiple probability measures of the preceding users. Also, we adopt the Gamma process construction of HDP [49] with a modification to incorporate the homogeneity index and regulate the homogeneity among user interests.
- (c) We combine nonparametric topic models with Bayesian Gaussian process latent variable models (Bayesian GP-LVMs) [63] to infer the homogeneity index of a news story from the users’ and the stories’ topics in a highly nonlinear fashion.

Since both the content modeling and homogeneity discovery part of our modeling fall into the Bayesian nonparametric framework, the posterior inference of our model can be done using the variational Bayes methods [23]. We train our model on a real-world Twitter dataset [36, 37, 39] that consists of user posting and sharing with true and false news stories. Our main contributions and the findings through the experiments are as follows:

- I. We develop HBTP that jointly models contents of news stories and user sharing events. Our model discovers the latent topics, topical interests of users, and homogeneity indices of news stories; these mutually reshape one another in the joint modeling.
- II. We find that the homogeneity indices that our model discovered significantly varies according to the labels of genuineness of the stories.
- III. We develop a supervised extension of HBTP (sHBTP) that incorporates the genuineness of news stories in the training of the model, and conduct a classification task of predicting the genuineness of a news story in the test set. We compare our prediction results with state-of-the-art fake news detection models and with other recently proposed supervised topic models and show that our model outperforms the comparison models in most of the evaluation metrics.
- IV. We make our code and data (Twitter data augmented with the news articles that the tweets refer to) publicly available for future research (<https://github.com/todoaskit/HBTP>).

**Related work.** By incorporating the label information of the news stories, the supervised formulation of our model performs the classification task of predicting the labels of the news stories. There are feature-based methods that leverage text, social network, temporal traces and propagation models to classify true and false news in supervised fashion [39, 50, 20, 29, 72, 70, 31, 32]. Also, another line of research focuses on devising algorithms that mitigate false news and their diffusion in social networks [9, 64, 48, 27, 17].

From the modeling aspect, there are upstream models in parametric topic modeling literature that generate the topic indices in a document from the topic distribution of a set of entities [54, 45, 14, 43, 44]. HBTP falls into this category in a way that a set of users who share a news story participate in modeling its content. Also, to generate both homogeneity values and labels of news stories, we incorporate Bayesian GP-LVM [63, 11] in our model in the nonparametric topic modeling framework. Bayesian GP-LVM extends GP-LVM [33, 34] by incorporating additional priors and conducting posterior inference using variational methods [23, 68]. Kandemir *et al.* recently proposed GPSTM as a joint model between LDA and Bayesian GP-LVM and

demonstrate the performance gain of jointly modeling LDA and GPLVM over linear classifiers [24]. Lastly, while our model focus on the joint modeling of text and users' sharing patterns [26], textual data has also been aligned with users' temporal traces to model diffusion pattern and text [16, 21], to cluster the documents in both discrete [1] and continuous [15, 69, 41] time domains. These models can be complementary to our model and be co-trained to increase the predictive power in detecting false news.

## 2 Background on the Gamma process construction of HDP

Hierarchical Dirichlet process (HDP) for modeling document collections is a two-level Dirichlet process (DP) [18] that functions as a nonparametric Bayesian prior for mixed membership models [60]. Being nonparametric, we do not have to specify the number of topics a priori, but instead, it can be inferred from the data [60].

In the HDP topic model, the random probability measure drawn from the first-level Dirichlet process becomes the base measure for the second-level Dirichlet processes:

$$G_0 \sim \text{DP}(\alpha, H), \quad G_d \stackrel{iid}{\sim} \text{DP}(\beta, G_0),$$

where  $d$  is the document index,  $H$  is the base measure, and  $\alpha, \beta$  are the first and the second-level DP concentration parameters respectively.  $H$  is a Dirichlet distribution on the vocabulary simplex, and the atoms of  $G_0$  are an infinite set of word-topic distributions drawn from  $H$ .

The first- and the second-level Dirichlet processes yield random discrete probability measures  $G_0 = \sum_{k=1}^{\infty} p_k \delta_{\phi_k}$  and  $G_d = \sum_{k=1}^{\infty} p_k^{(d)} \delta_{\phi_k}$ . Here, the weights  $\{p_k\}_{k=1}^{\infty}$  and  $\{p_k^{(d)}\}_{k=1}^{\infty}$  depend on the corresponding concentration parameters  $\alpha$  and  $\beta$  and atoms  $\{\phi_k\}_{k=1}^{\infty}$  are drawn from the base probability measure:  $\phi_k \stackrel{iid}{\sim} H$ . Finally, each second-level  $G_d$  generates its  $n^{th}$  associated topic-indicator variable  $z_n^{(d)}$  and each word  $x_n^{(d)}$  is drawn from  $\phi_{z_n^{(d)}}$  for  $N_d$  times as follows:

$$z_n^{(d)} | G_d \stackrel{iid}{\sim} G_d, \quad x_n^{(d)} | z_n^{(d)} \stackrel{iid}{\sim} \text{Multinomial}(\phi_{z_n^{(d)}}).$$

Following the work of Paisley *et al.* [49], we constructively define the first and the second-level Dirichlet processes using the stick-breaking process [57] and normalized Gamma processes [49]. For the first-level  $G_0$ , we represent the weights  $\{p_k\}_{k=1}^{\infty}$  as

$$p_k = V_k \prod_{j=1}^{j < k} (1 - V_j), \quad V_k \stackrel{iid}{\sim} \text{Beta}(1, \alpha).$$

The second-level probability measures  $G_d$  are conditionally distributed on the first-level random probability measure  $G_0$ , and the corresponding weights  $\{p_k^{(d)}\}_{k=1}^{\infty}$  can be represented as

$$p_k^{(d)} = \frac{\pi_k^{(d)}}{\sum_{j=1}^{\infty} \pi_j^{(d)}}, \quad \pi_k^{(d)} | G_0 \stackrel{iid}{\sim} \text{Gamma}(\beta p_k, 1).$$

## 3 Homogeneity-Based Transmissive Process

In this section, we introduce the modeling scheme of the homogeneity based transmissive process. We start by specifying different kinds of data structure the model works on. Next, we describe the generative process of the first- and second-level Dirichlet processes, explain the transmissive property when generating the second-level probability measures and explain how the homogeneity indices of news stories function in regulating those probability measures. We then explain how we generate the content of a news story as a mixture of users' probability measures and illustrate how we draw homogeneity indices of news stories. Lastly, we discuss the parametric and supervised formulations of our model.

### 3.1 Event representation

In social networks, users introduce news stories and these stories are propagated throughout the network by means of sharing (*e.g.*, tweets and retweets). We first represent such tweet and retweet events in a social network as a triplet

$$e := \begin{matrix} & \text{preceding user} \\ & \downarrow \\ (u, & a, & s) \\ \uparrow & & \uparrow \\ \text{user} & & \text{story} \end{matrix}.$$

Here, we denote a set of events as  $\mathbb{E} \ni e$ . The first item  $u$  is a user who created the tweet/retweet, and the second item  $a$  is a user who created the preceding tweet if the event is a retweet. We assume that there are  $M$  users and denote a set of users as  $\mathbb{U} \ni u, a$ . The last item  $s$  is a news story. We represent each news story as a triplet

$$s := \begin{matrix} & \text{homogeneity index} \\ & \downarrow \\ (d, & h, & l), \\ \uparrow & & \uparrow \\ \text{content} & & \text{label} \end{matrix},$$

where  $d$  is a bag-of-words representation of content of the story,  $h$  is a homogeneity index, and  $l$  is the label of the story, *e.g.*, true or false. We remark that the homogeneity index  $h$  is hidden and will be uncovered using our model. There are  $L$  stories and the set of stories is denoted as  $\mathbb{S} \ni s$ .

### 3.2 Modeling the first and the second-level DPs

Our model inherits the two-level DP construction of the HDP, and the first-level DP formulates identically to that of the HDP. We draw a probability measure  $G_0$  from DP with the base measure  $H$  and the concentration parameter  $\alpha$ . Using the stick-breaking construction [57], we represent the first-level probability measure as  $G_0 = \sum_{k=1}^{\infty} V_k \prod_{j < k}^{j=1} (1 - V_j) \delta_{\phi_k}$ ,  $V_k \stackrel{iid}{\sim} \text{Beta}(1, \alpha)$  and  $\phi_k \stackrel{iid}{\sim} H$ .

**Transmissive modeling of user DPs.** For the second-level, we endow each user in a social network, instead of a news story, with a single DP. The key idea is that each user's probability measure is transmitted to the users who retweeted her (re)tweet, and the degree of transmission, *i.e.*, the similarity between the two users' probability measure, depends on the homogeneity index of the news story the users are sharing. To formalize this, for each user  $u$ , we first construct  $\mathbb{E}_u$  which is a set of events  $e$  whose first element is  $u$ . From  $\mathbb{E}_u$  we retrieve  $\mathbb{A}_u \ni v$ , a set of users preceding  $u$  and denote the set of news stories that are shared by  $u$  and  $v$  as  $s^{(u,v)}$ .

When a user  $u$  has no predecessor, *i.e.*,  $\mathbb{A}_u = \emptyset$ , the user's probability measure  $G_u$  is drawn from

$$G_u = \sum_{k=1}^{\infty} \frac{\pi_k^{(u)}}{\sum_{j=1}^{\infty} \pi_j^{(u)}} \delta_{\phi_k} \sim \text{DP}(\beta, G_0), \quad \pi_k^{(u)} \sim \text{Gamma}(\beta p_k, 1).$$

Recall that  $p_k = V_k \prod_{j=1}^{k-1} (1 - V_j)$  is the weight of the  $k^{th}$  atom of the first-level probability measure.

If the user  $u$  has preceding users, for each  $v \in \mathbb{A}$  we draw  $G_u^{(v)}$ , which is user  $u$ 's probability measure *transmitted* by the preceding user  $v$ . Denoting the homogeneity index of story  $s^{(u,v)}$  as  $h^{(u,v)}$ ,  $G_u^{(v)}$  is drawn from the DP with the base measure  $G_v$  and the concentration parameter  $\beta \exp(h^{(u,v)})$  as follows:

$$\begin{aligned} G_u^{(v)} &= \sum_{k=1}^{\infty} \frac{\pi_k^{(u,v)}}{\sum_{j=1}^{\infty} \pi_j^{(u,v)}} \delta_{\phi_k} \sim \text{DP}(\beta e^{h^{(u,v)}}, G_v), \\ \pi_k^{(u,v)} &\sim \text{Gamma}(\beta e^{h^{(u,v)}} p_k^{(v)}, 1), \end{aligned} \tag{1}$$

where  $p_k^{(v)} = \sum_{j=1}^{\infty} \frac{\pi_j^{(v)}}{\sum_{j=1}^{\infty} \pi_j^{(v)}}$  is the weight of the  $k^{th}$  atom of the preceding user  $v$ 's probability measure.

Here, we posit that there is no circular retweet among users. Finally, denoting the weight of the  $k^{th}$  atom of  $G_u^{(v)}$  as  $p_k^{(u,v)}$ , we constructively define user  $u$ 's probability measure as

$$G_u := \sum_{k=1}^{\infty} \frac{\sum_{v \in \mathbb{A}_u} p_k^{(u,v)}}{|\mathbb{A}_u|} \delta_{\phi_k}.$$

Note that  $G_u$  derived from multiple  $G_u^{(v)}$ s is still a probability measure as it meets the countable additivity property and the weights of the atoms lie in the unit interval [55].

**Remark 1.** We modify the shape parameter of  $\text{Gamma}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$  in equation 1 to incorporate the effect of homogeneity of news stories on the transmission of the probability measure. This approach differs from past approaches which modify the rate parameter  $b$  to achieve topic-wise re-scaling [49, 25]. In our model, changing the rate parameter would be ineffective because the expected impact of the homogeneity index is identical for all topics, and scaling the rate parameter using the same multiplier for all topics yields identical outcome without the multiplier in the Gamma process after the normalization step. For this particular reason, the rate parameters in the Gamma processes disappear when used to represent DP [13]. Conversely, by scaling the shape parameter by, say,  $\tau$ , the mean of the normalized Gamma process remains unchanged while the variance is reduced by  $1/\tau$ . Therefore, when a probability measure of a user is *transmitted* by a news story with a high homogeneity index, the probability measure of the user will be more similar to a preceding user’s probability measure. Finally, note that since the rate parameter remains unchanged, our formulation of the second-level probability measures can be represented using the stick-breaking process. We use the normalized Gamma process representation to enjoy the merit of generating  $\pi_k$  independently, which will be useful for approximate posterior inference.

### 3.3 Modeling news stories using mixtures of user DPs

We model the contents of a news story  $s$  using probability measures of users who participated in propagating the story. There are multiple previous parametric topic models that adopt this *upstream* approach to model a document using probability measures from multiple sources [54, 45, 14, 43, 44]. Here, using  $\mathbb{E}$ , we define  $\mathbb{U}_s$  as the set of users who propagated the story  $s$ . Then, following the work of Rosen-Zvi *et al.* [54], for  $n^{th}$  word, we sample a user  $w$  uniformly from  $\mathbb{U}_s$  and draw the topic indicator variable  $z_n^{(s)}$  and word  $x_n^{(s)}$  from  $G_w$  as follows:

$$w \sim \text{Uniform}(\mathbb{U}_s), \quad z_n^{(s)} | G_w \sim G_w, \quad x_n^{(s)} | z_n^{(s)} \sim \text{Mult.}(\phi_{z_n^{(s)}}).$$

**Generating homogeneity indices of stories.** In equation 1, we assume that a user’s probability measure is generated from her preceding user’s probability measure and the similarity of the two probability measures is calibrated by the homogeneity index of a news story that bridges the two users. Here, the homogeneity index of a news story is generated by  $\bar{z}^{(s)} = \frac{1}{N_s} \sum_{n=1}^{N_s} z_n^{(s)}$ , which is a normalization of indicator one-hot vectors in the topic space. Following the Bayesian Gaussian process latent variable model (Bayesian GP-LVM) [63, 11], we first generate a hidden input vector  $c^{(s)}$  from the Gaussian distribution centered at  $\bar{z}^{(s)}$ , and draw latent function values  $\mathbf{f} \in \mathbb{R}^L$  from the Gaussian distribution centered at zero and the covariance matrix constructed from the pairings of news stories’  $c^{(s)}$ s. Finally, the homogeneity index for a news story  $h^{(s)}$  is a noisy observation of  $f^{(s)}$  with Gaussian noise. The formal constructions are as follows:

$$c^{(s)} | \bar{z}^{(s)} \sim \mathcal{N}(\bar{z}^{(s)}, \zeta^{-1}I), \quad \mathbf{f} | \mathbf{c} \sim \mathcal{N}(0, K_{LL}), \quad \mathbf{h} | \mathbf{f} \sim \mathcal{N}(\mathbf{f}, \kappa^{-1}I). \quad (2)$$

We use squared exponential kernel  $k(l, l') = \sigma^2 \exp(-\frac{1}{2}\|c - c'\|^2)$  for the covariance matrix  $K_{LL}$ .

**Supervised HBTP.** HBTP can be used for supervised learning, *e.g.*, to predict the type of genuineness of news stories. Similar to drawing the homogeneity indices, we draw labels of the stories from the same hidden input vector. The difference is that the homogeneity indices are hidden in the modeling state and have to be inferred from the model, while the labels are given in the training data.

**Parametric counterpart of HBTP.** There is a much overlap in the modeling scheme when we develop a parametric version of HBTP. The major differences are that in the parametric model, (1) we predefine the number of topics and (2) we use the logistic normal prior [2] for the user-topic distributions used to capture the transmissive property among users. With the parametric HBTP, we sacrifice the model flexibility, but we gain an algorithmic benefit of inferring the homogeneity variables for news stories with closed-form updates by solving the Lambert W function. Refer to the Appendix for the generative process of parametric HBTP.

**Remark 2.** In our model, we jointly model Bayesian GP-LVM with a nonparametric topic model which provides both computational and modeling benefits. First, incorporating GP-LVM with hidden inputs reduces

the time complexity of drawing function values  $f^{(s)}$  from  $\mathcal{O}(L^3)$  to  $\mathcal{O}(LP^2)$ , where  $L$  is the number of stories and  $P$  is the number of auxiliary inducing variables [10, 58, 62, 56]. In the modeling aspect, we can model multi-dimensional or categorical labels from the topic indicator variable  $z^{(s)}$  without directly conditioning on the document-topic simplex, which has been pointed out as one of the weaknesses of the supervised topic model [42] compared to the labeled LDA [51]. Finally, we remark that when computing vector-wise norms of  $c^{(s)}$ , we resort to a truncated approximation method [7, 30, 61] to predefine the upper-bound of the number of topics used.

## 4 Posterior Inference of HBTP

We derive a joint variational inference algorithm to approximate the posterior of HBTP. Since both the text modeling and the homogeneity generation of HBTP fall into the Bayesian nonparametrics, we refer to the well-established literature for variational inference in nonparametric topic models [7, 30, 61] and Bayesian GP-LVMs [11]. We derive the log joint distribution of HBTP as:

$$\begin{aligned} \ln p(\mathbf{x}, \mathbf{z}, \mathbf{c}, \mathbf{h}, \Pi, V, \Phi) = & \sum_{s=1}^L \sum_{n=1}^{N_s} \left\{ \ln p(x_n^{(s)} | z_n^{(s)}, \Phi) + \ln \left( \frac{1}{|\mathbb{U}_s|} \sum_{w \in \mathbb{U}_s} p(z_n^{(s)} | \pi^{(w)}) \right) \right\} \\ & + \ln p(\mathbf{h} | \mathbf{c}) + \sum_{s=1}^L \ln p(c^{(s)} | z^{(s)}) + \sum_{k=1}^{\infty} \left\{ \ln p(V_k | \alpha) + \ln p(\phi_k | \alpha_0) \right\} \\ & + \sum_{u=1}^M \sum_{k=1}^{\infty} \ln \left( \frac{1}{|\mathbb{A}_u|} \sum_{v \in \mathbb{A}_u} p(\pi_k^{(u,v)} | V_k, \Pi_k, h^{(u,v)}, \beta) \right). \end{aligned}$$

The log variational distribution of HBTP can be factorized into

$$\begin{aligned} \ln q(\mathbf{z}, \mathbf{c}, \mathbf{f}, \Pi, V, \Phi) = & \sum_{s=1}^L \sum_{n=1}^{N_s} \ln q(z_n^{(s)}) + \ln q(\mathbf{f}) + \sum_{s=1}^L \ln q(c^{(s)}) \\ & + \sum_{u=1}^M \sum_{k=1}^T \ln \left( \frac{1}{|\mathbb{A}_u|} \sum_{v \in \mathbb{A}_u} q(\pi_k^{(u,v)}) \right) + \sum_{k=1}^T \left\{ \ln q(V_k) + \ln q(\phi_k) \right\}, \end{aligned}$$

with the truncation value  $T$  with  $V_T = 1$ . We specify the variational distributions for the topic-related latent variables with their variational parameters as

$$\begin{aligned} q(z_n^{(s)}) &= \text{Multinomial}(z_n^{(s)} | \gamma_n^{(s)}), \\ q(\pi_k^{(u,v)}) &= \text{Gamma}(\pi_k^{(u,v)} | a_k^{(u,v)}, b_k^{(u,v)}). \end{aligned}$$

To specify the variational distributions for homogeneity-related variables, we introduce  $G$  auxiliary inducing variables  $y \in \mathbb{R}^G$  with corresponding inducing input locations  $Y \in \mathbb{R}^{G \times T}$  with  $p(y|Y) = \mathcal{N}(y|0, K_{GG})$ , and express the variational distributions for the hidden input vectors  $c^{(s)}$  and noise-free GP latent functions as

$$\begin{aligned} q(c_k^{(s)}) &= \mathcal{N}(c_k^{(s)} | \lambda_k^{(s)}, \xi_{s,k}^{-1} I), \\ q(f^{(s)}) &= \mathcal{N}(f^{(s)} | K_{LG}^\top K_{GG}^{-1} y, \quad K_{LL} - K_{LG} K_{GG}^{-1} K_{GL}), \end{aligned}$$

where we set  $q(y) = \mathcal{N}(y | \mu^{(y)}, \Sigma^{(s)})$  and assume independence among topics for  $c^{(s)}$ . Here we set  $G$  to be drastically smaller than  $L$  to increase the inference speed. Finally, we specify the variational distributions for topic variables as follows:

$$q(V_k) = \delta_{V_k}, \quad q(\phi_k) = \text{Dirichlet}(\phi_k | \eta_k),$$

where we use the delta function for both simplicity and tractability in inference steps as demonstrated in the work of Liang et al. [35]. Given the evidence, the task of approximating the variational distribution with

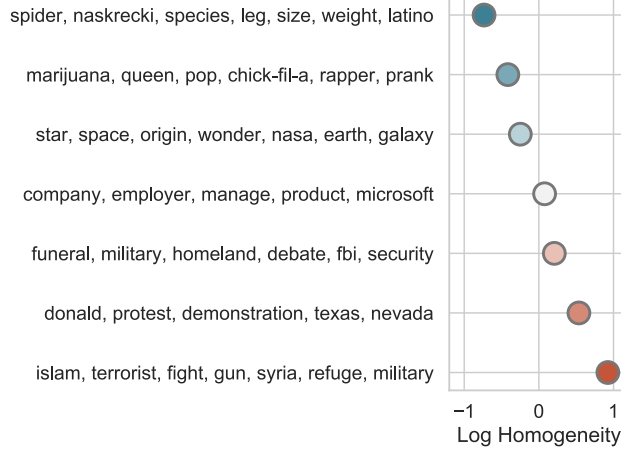


Figure 1: Topics retrieved by HBTP and their corresponding homogeneity indices.

respect to the original posterior is equivalent to minimizing the KL divergence between the joint distribution and the factorized variational distribution or maximizing the evidence lower bound. The main derivations come from the expectations of the log-probability of the latent variables with respect to the variational distribution  $q$ , and the optimization is done using the coordinate ascent algorithm. For the remainder of the section, we report updating formulas for the latent variables that are unique for HTBP compared to previous nonparametric topic models and GP-LVMs. Refer to the Appendix for the full updates.

**Story-level updates.** For the story-level latent variables,  $z^{(s)}$  is at the intersection of the topic modeling part and the GP part, and the update for  $\gamma_{nk}^{(s)}$  is

$$\gamma_{n,k}^{(s)} \propto \exp \left\{ \mathbb{E}_q[\ln \eta_{k,x_n^{(s)}}] + \mathbb{E}_q[\ln \pi_k^{(s)}] + \frac{\zeta}{N_s} \mathbb{E}_q[c^{(s)}] - \frac{\zeta}{2N_s^2} \left( \sum_{n' \neq n} \gamma_{n',k}^{(s)} + 1 \right) \right\}.$$

Also, the update for the homogeneity variable  $h^{(s)}$  is

$$\frac{\partial \mathcal{L}}{\partial h^{(s)}} = \sum_{k=1}^T \sum_{u,v:s^{(u,v)}=s} \beta e^{h^{(u,v)}} p_k^{(v)} \left\{ \mathbb{E}_q[\ln \pi_k^{(u,v)}] - \psi \left( \beta e^{h^{(u,v)}} p_k^{(v)} \right) \right\} + \sum_{s'=1}^L h^{(s')} (W_{s,s'} + W_{s',s}),$$

where the full expression for the  $L \times L$  matrix  $W$  is specified in the Appendix.

**User-level updates.** Updating  $\pi_k^{(u,v)}$  is done as follows:

$$a_k^{(u,v)} = \beta \exp(h^{(u,v)}) p_k^{(v)} + \sum_{n=1}^{N_s} \gamma_{n,k}^s, \quad b_k^{(u,v)} = 1 + \frac{N_s}{\chi_s},$$

where  $\chi_s = \sum_{k=1}^T \mathbb{E}_q[\pi_k^{(u,v)}]$  and we simplify  $s^{(u,v)}$  to  $s$ .

**Computational complexity.** For text modeling, HBTP has the time complexity  $\mathcal{O}(|\mathbb{E}|NT + VT)$  with  $|\mathbb{E}|$  being the number of events,  $N$  the average length of a story (in words) and  $T$  the truncation size. The time complexity for the homogeneity variables is  $\mathcal{O}(LP^2)$ .

## 5 Experiments

We first investigate the interplay among news stories' content, homogeneity indices and users' topical interests modeled in nonparametric HBTP and examine how the homogeneity indices of news stories regulate the alignment of topical interests of users who tweet/retweet the news stories. Through the process, we find

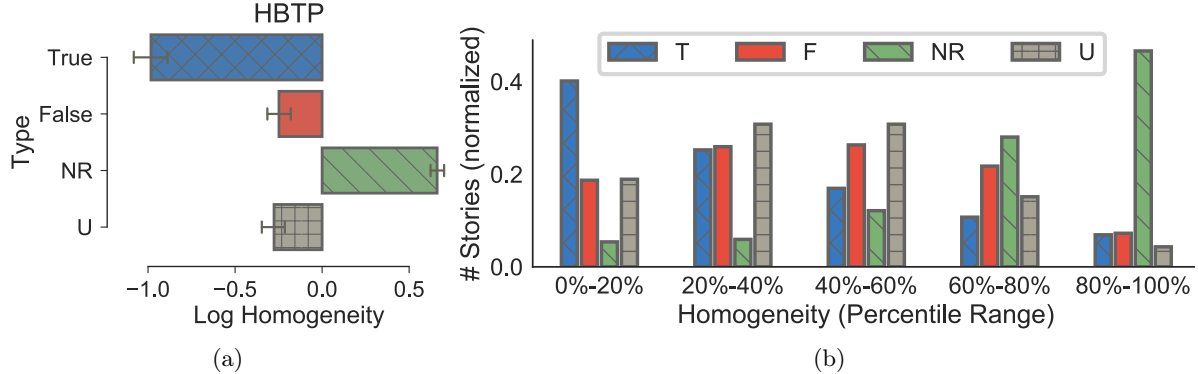


Figure 2: Relationship between the homogeneity indices by HBTP and the labels (genuineness) of news stories when the labels have not been used during the training. (a) Mean homogeneity indices for stories with different labels retrieved using *unsupervised* HBTP. Homogeneity indices of news stories vary significantly according to their labels except for "False" and "Unverified" stories. (b) Distribution of news stories with different labels over the homogeneity indices.

interesting relationship between the homogeneity indices and the labels (genuineness) of news stories when the labels are not observed during training. Motivated by the finding, we demonstrate how the supervised version of our model (sHBTP) is able to detect the labels of news stories better than the state-of-the-art comparison models.

**Dataset.** We use the extension of Twitter rumor dataset that has been previously used in rumor detection research [36, 37, 39, 40]. The original dataset contains, for each tweet, (1) the tweet ID, (2) URL of the news story (3) users' tweet and retweet logs, and (4) a label, which is either true/T, false/F, non-rumor/NR, or unverified/U. The "true", "false", and "unverified" stories are potential "rumors" that need to be fact-checked by the debunking websites such as [snopes.com](http://snopes.com) to inspect their genuineness. "True" stories are the ones that turned out to be genuine after the fact-checking, while "false" stories are the ones that contain misinformation and "unverified" stories are the ones that cannot be conclusively judged even after the human evaluation. Finally, "non-rumor" stories are the ones that do not need the fact-checking because *e.g.*, they came from reliable sources.

For the pre-processing of the original dataset, we exclude the news stories whose content information cannot be retrieved with the Twitter API. Then, we remove "leaf" users, the ones who shared a story once and are at the leaf of the diffusion cascades, since they do not affect the calculation of the homogeneity indices and thus, the outcome of our modeling. Also, we confirm that there is no "circular" retweets among users regarding news stories so that our graphical model remains a directed acyclic graph (DAG). After the pre-processing, there are 79,416 users and 175,389 tweets and retweets. For the news stories, there are 1,107 stories in total that contain 12,515 unique tokens. The news stories are divided into 289, 262, 371, and 185 true, false, non-rumor and unverified stories. Finally, for the prediction task, we conduct 5-fold cross validation on the Twitter dataset.

**Parameter settings.** For HBTP and sHBTP, we set the topic Dirichlet prior  $\alpha_0$  to 0.1. For GP-LVM in the model, we set 50 inducing points for the homogeneity variable and the labels, noise precision parameters  $\zeta$ ,  $\kappa$  to 10, and the variational precision parameter  $\xi$  to 0.1. For the comparison methods, we set the values to be consistent with ours for overlapping parameters. Otherwise, we follow the parameter settings disclosed by the original paper or optimize the values through exhaustive search.

**Analyses on the homogeneity of news stories and their topic and labels using unsupervised HBTP.** We analyze the interplay between the homogeneity indices of news stories uncovered by HBTP and the alignment of topical interests of users linked to the story and examine how news stories with different topics and labels have different homogeneity indices.

First, figure 1 lists 6 topics chosen from HBTP and their corresponding homogeneity indices. We observe



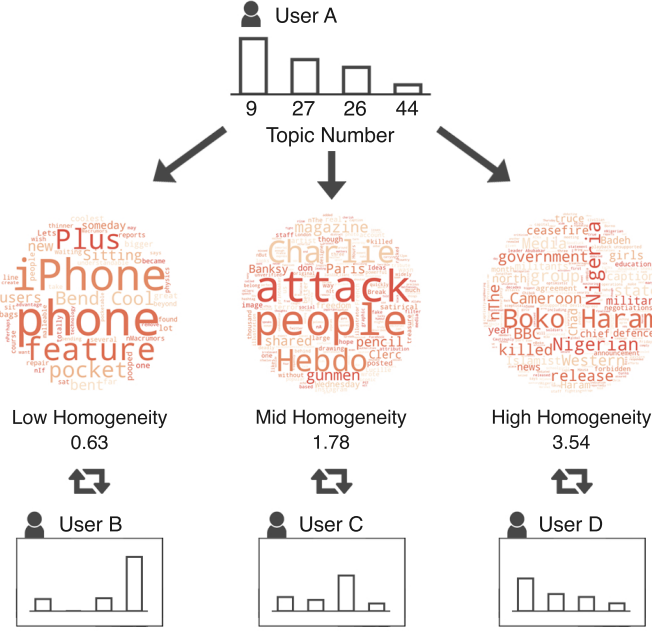


Figure 3: Three example false news stories with different homogeneity indices. All three news stories are originally tweeted by user A, but retweeted by different users B, C, and D respectively. Although all three stories have same labels, their contents are different and have different homogeneity indices, which affect the topical similarities between the tweeting user A and retweeting users B, C, and D.

that topics related to nature, urban culture, and science in general have low homogeneity indices, suggesting that users who share news stories with these topics are in general less topically aligned than those who share other news stories. On the other hand, the homogeneity indices of economic and industry-related topics are in the middle-range, and topics regarding domestic and international politics have high homogeneity indices. Note that our topic-word probability measures are attached with homogeneity indices in a similar way to that of supervised topic models [42], since both HBTP and supervised topic models leverage the labels conditioned on topics, unlike other models that incorporate label information but the conditioning is the other way around [51]. However, the modeling assumption of HBTP diverges from that of supervised topic models in that in supervised topic models, the labels are "given", whereas in our model, both topics and homogeneity indices are "latent", and the homogeneity indices need to be inferred from users' tweet/retweeting patterns and both topics and homogeneity indices mutually reshape one another during the joint-modeling.

Second, from figure 2a, we observe that the homogeneity indices of stories vary significantly according to their labels. We note that HBTP computes the homogeneity indices of news stories in unsupervised fashion, without observing the labels of the stories while training. Interestingly, "non-rumors" have the highest homogeneity indices while the other types of news stories that went through fact-checking process have much lower values. Among the fact-checked stories, the stories that are verified to be "true" have significantly lower homogeneity indices than "false" and "unverified" stories. Figure 2b shows the distribution of news stories with different labels over homogeneity indices. From the figure, we confirm that majority of "non-rumors" have high homogeneity indices (top 40%), majority of "true" stories have low homogeneity indices (bottom 40%), while "false" and "unverified" stories are distributed in the middle range. Overall, the difference in homogeneity indices of news stories with respect to different labels suggest that we can leverage such homogeneity indices and topics to classify the news labels in prediction tasks.

Lastly, we highlight the *transmissive* nature of our model and inspect how the homogeneity of a news story affects the degree of transmission between the topical interests of user pairs who tweet and retweet the story. In figure 3, we visualize three news stories tweeted by a user (user A). All three stories are labeled

Table 1: Results on label classification task using supervised HBTP and other methods. The plus-minus sign indicates one standard error.

Method	Accuracy	True $F_1$	False $F_1$	Non-Rumor $F_1$	Unverified $F_1$
HDP+SVM: Linear	$0.484 \pm 0.019$	0.593	0.490	0.514	0.357
HDP+SVM: RBF	$0.605 \pm 0.024$	0.644	0.522	0.629	0.514
HBTP+SVM: Linear	$0.533 \pm 0.014$	0.621	0.464	0.641	0.593
HBTP+SVM: RBF	$0.717 \pm 0.006$	0.855	0.550	0.682	0.653
GPSTM (Kandemir <i>et al.</i> , 2018)	$0.664 \pm 0.013$	0.662	0.686	0.687	0.435
BU-RvNN (Ma <i>et al.</i> , 2018)	$0.622 \pm 0.010$	0.616	0.571	0.687	0.584
TD-RvNN (Ma <i>et al.</i> , 2018)	$0.698 \pm 0.004$	0.664	0.668	0.723	<b>0.693</b>
Supervised HBTP	<b><math>0.781 \pm 0.012</math></b>	<b>0.891</b>	<b>0.740</b>	<b>0.812</b>	0.622

as "false", but the content of the stories differ, leading to different homogeneity indices. Here, the topic proportion of user A is transmitted to three other users, B, C, and D, each of whom retweeted different tweets of user A with different stories. Among users B, C, and D, since user D retweeted a tweet that contains a high homogeneity-valued story, her topic distribution is more similar to user A than, say, that of users B and C. By generalizing this observation to multiple tweeting and retweeting users, we can presume that the topical interests of users that share a high homogeneity-valued story will be more uniform than that of the users who share news stories with low homogeneity indices.

**Story label prediction using supervised HBTP.** We validate our modeling scheme by showing the predictive power of our model compared with the other competitive models. For this classification task, we use sHBTP that observes the labels of the news stories in the training set. Our comparison models are as follows:

- **HDP + SVM:** This method uses document-topic probability measures for news stories using HDP, and uses support vector machine (SVM) as a classifier. For SVM, we use linear and radial basis function (RBF) kernels.
- **HBTP + SVM:** Unlike HDP, document-topic probability measures are drawn per users, and the topics of news stories are represented by aggregating users' probability measures who shared the story. Also, the homogeneity of news stories are incorporated in the model which affect users' probability measures.
- **GPSTM:** Gaussian process supervised topic models [24] extend parametric supervised topic models [42] by using Bayesian GP-LVMs as classifiers instead of linear classifiers and the labels are conditioned on news stories' topics.
- **BU-RvNN and TD-RvNN:** Bottom-up and Top-down recursive neural networks [40] are the state-of-the-art models designed for classifying true and false news. The models are operationalized on bottom-up and top-down propagation trees where each node contains tweet contents. Thus, our model and the other topic-based comparison models exploit the content information of news sources, whereas BU-RvNN and TD-RvNN use content information of tweets.

Note that there are trade-offs between using the content information of news stories and tweets. From one standpoint, contents in news stories have richer description about the actual event from which the models judge the genuineness. From another standpoint, as stated in the work of Ma *et al.* [39], the tweets often

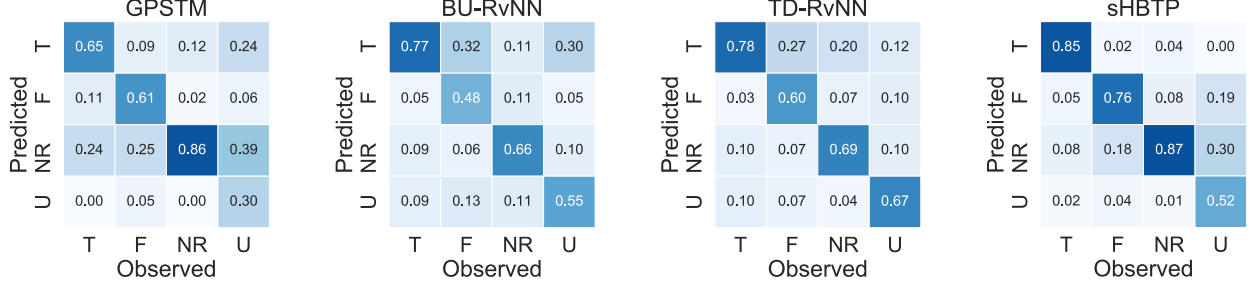


Figure 4: Confusion matrices for classification results of GPSTM, BU-RvNN, TD-RvNN, and supervised HBTP.

contain user judgments or debates regarding the news stories’ genuineness that can be critical hints the models exploit.

In table 1, we report the overall classification accuracy and the  $F_1$  scores for the four labels. We confirm that our model outperforms the comparison methods in the overall accuracy and in the  $F_1$  scores for all labels except "unverified". Specifically, in figure 4, we confirm that our model excels at differentiating "true" and "false" news, two controversial groups of news stories that needed to be fact-checked by human experts for the classification. On the flip side, our model is relatively poor at detecting "unverified" stories. One possible reason for the performance edge of our model is the inclusion of homogeneity variable in our model; as observed in figure 2, the log homogeneity indices of stories with four labels differ significantly, which help the model to classify confusing label pairs, *e.g.*, high-valued non-diagonal entries in the confusion matrices in figure 4. Also, the performance edge over GPSTM highlights the efficacy of our model’s nonparametric and joint modeling-nature among user interests, news stories’ contents and the homogeneity variables that encode information of users’ sharing patterns.

## 6 Conclusion and Future Work

We developed HBTP, a novel Bayesian nonparametric model that jointly models topics and homogeneity indices of news stories and user interests. HBTP for modeling the content of news stories extends hierarchical Dirichlet process in a way that instead of a single upper-level probability measure affecting the lower-level probability measures, we allowed the probability measures of the bottom-layer to be transmitted from one another. To discover homogeneity values of the news stories, we incorporated Bayesian GP-LVM in the model. By conducting quantitative and qualitative analyses, we found interesting relationships between the homogeneity index and the genuineness of news stories, and between the homogeneity index and the topics extracted. Finally, we showed how HBTP can be extended easily to predict genuineness of news stories better than the state-of-the-art rumor detection methods.

Our model can be extended to leverage different types of data to increase the flexibility of the model and to boost the predictive power. One direction is to explicitly incorporate users’ diffusion cascades and the contents of the user tweets in the model. For example, our model and RvNN based model of Ma *et al.* [40] can work in a complementary fashion because the input spaces these models operate on are almost orthogonal to each other.

Also, while the RvNN based models require the users’ diffusion cascades and thus cannot predict completely new news stories, we can explore the predictive power of our model on newly created stories with no user sharing by directly inferring the topic probability measures of these held-out stories. In this setting, we can only leverage the homogeneity indices of the training stories.

## Appendix A Parametric HBTP

For parametric HBTP, we sample a user's topic distribution  $\theta$  using a multivariate normal distribution with a mean being the preceding user's topic distribution. For the variance, we incorporate the homogeneity value to regulate the similarity between the two users' topic distributions. Then, we put a sigmoid function over  $\theta$  to turn it into the simplex, and draw topic labels for each words for news stories. Note that the logistic normal distribution has been previously used in correlated topic models [8], but with a different purpose: to capture correlations among topic. Finally, we illustrate the generative process of the parametric counterpart of HBTP as follows:

1. For each topic  $k$ , draw word-topic distribution  $\phi_k \sim \text{Dir}(\alpha)$ .
2. For user  $u$ , draw  $\theta_u$  with  $p(\theta_u) = \frac{1}{|\mathbb{A}_u|} \sum_{v \in \mathbb{A}_u} \mathcal{N}(\theta_u | \theta_v, \exp(-h^{(u,v)})\Sigma)$  if  $u$  has preceding users  $v$ . If not, draw  $\theta_u \sim \mathcal{N}(\mu, \Sigma)$ .
3. For news story  $s$ :
  - For each word index  $n$ :
    - Draw user  $w$  who propagated the story  $s$ .
    - Draw topic indicator variable  $z_n^{(s)} \sim \text{Sigmoid}(\theta_w)$  and draw word  $x_n^{(s)} \sim \phi_{z_n^{(s)}}$ .
  - Draw homogeneity value  $h^{(s)}$  using equation (9).

## Appendix B Posterior Inference

**Expectations w.r.t.  $q$ .** By taking expectations of the latent variables in HBTP with respect to the variational distribution  $q$ , we get the following results:

$$\mathbb{E}_q[\ln p(x_n^{(s)} | z_n^{(s)})] = \sum_{k=1}^T \gamma_{nk}^{(s)} \mathbb{E}_q[\ln p(x_n^{(s)} | \phi_k)] = \sum_{k=1}^T \gamma_{nk}^{(s)} \left\{ \psi(\eta_{kx_n^{(s)}}) - \psi\left(\sum_{i=1}^{|V|} \eta_{ki}\right) \right\}.$$

For stories and user variables,

$$\begin{aligned} \mathbb{E}_q[\ln p(z_n^{(s)} | \pi^{(w)})] &= \sum_{k=1}^T \gamma_{nk}^{(s)} \mathbb{E}_q[\ln p(z_n^{(s)} = k | \pi^{(w)})] \\ &= \sum_{k=1}^T \gamma_{nk}^{(s)} \left\{ \mathbb{E}_q[\ln \pi_k^{(w)}] - \mathbb{E}_q[\ln \sum_{k'=1}^T \pi_{k'}^{(w)}] \right\} \\ &\geq \sum_{k=1}^T \gamma_{nk}^{(s)} \left\{ \mathbb{E}_q[\ln \pi_k^{(w)}] - \ln \chi - \frac{\sum_{k=1}^T \mathbb{E}_q[\pi_k^{(w)}] - \chi}{\chi} \right\} \\ &\geq \sum_{k=1}^T \gamma_{nk}^{(s)} \left\{ \mathbb{E}_q[\psi(a_k^{(w)})] - \ln b_k^{(w)} - \ln \chi - \frac{\sum_{k=1}^T \mathbb{E}_q[\pi_k^{(w)}] - \chi}{\chi} \right\}, \end{aligned}$$

$$\mathbb{E}_q[\ln p(\pi_k^{(u,v)} | V_k, \Pi_k, h^{(u,v)}, \beta)] = \left( \beta e^{h^{(u,v)}} p_k^{(v)} - 1 \right) \left\{ \psi(a_k^{(u,v)}) - \ln b_k^{(u,v)} \right\} - \frac{a_k^{(u,v)}}{b_k^{(u,v)}} - \ln \Gamma(\beta e^{h^{(u,v)}} p_k^{(v)}).$$

For the corpus-level,

$$\mathbb{E}_q[\ln p(V_k | \alpha)] = \sum_{k=1}^T \ln \Gamma(\alpha + 1) - \ln \Gamma(\alpha) + (\alpha - 1) \ln(1 - V_k).$$

Finally, for GPLVM latent variables,

$$\mathbb{E}_q[\ln p(c^{(s)}|z^{(s)})] = \frac{T}{2} \ln 2\pi\zeta - \frac{1}{2}\zeta \left\{ \sum_{k=1}^T \xi_{s,k}^{-1} + (\lambda^{(s)} - \frac{1}{N_s} \sum_{n=1}^{N_s} \gamma_n^{(s)})^\top (\lambda^{(s)} - \frac{1}{N_s} \sum_{n=1}^{N_s} \gamma_n^{(s)}) \right\},$$

$$\mathbb{E}_q[\ln p(\mathbf{h}|\mathbf{c})] \geq \ln \left[ \frac{\kappa^{\frac{L}{2}} |K_{GG}|^{\frac{1}{2}}}{(2\pi)^{\frac{L}{2}} |\kappa\psi_2 + K_{GG}|^{\frac{1}{2}}} \right] - \frac{1}{2} \mathbf{h}^\top W \mathbf{h} - \frac{\kappa}{2} \psi_0 + \frac{\kappa}{2} \text{tr}(K_{GG}^{-1} \psi_2).$$

**Corpus-level updates.** Updating  $\eta_k$  can be done in a closed-form:

$$\eta_{k,i} = \alpha_0 + \sum_{s=1}^L \sum_{n=1}^{N_s} \gamma_{n,k}^{(s)} \mathbb{1}(x_n^{(s)} = i).$$

To update  $V_k$ , we use the steepest ascent algorithm:

$$\frac{\partial \mathcal{L}}{\partial V_k} = -\frac{\alpha-1}{1-V_k} + \beta \left[ \sum_{s=1}^L \left( \mathbb{E}_q[\ln \pi_k^{(s)}] - \mathbb{E}_q[\pi_k^{(s)}] - \psi(\beta p_k) \right) \right] \left[ \frac{p_k}{V_k} - \sum_{k' > k} \frac{p_{k'}}{1-V_k} \right].$$

**GPLVM updates.** Updating formula for  $W$ , with  $\psi_0 = L\sigma^2$  and  $\psi_2 = \sum_{s=1}^L \psi_2^{(s)}$ , is:

$$W = \kappa I - \kappa^2 \psi_1 (\kappa \psi_2 + K_{GG})^{-1} \psi_1^\top,$$

where

$$(\psi_1)_{s,g} = \sigma^2 \prod_{k=1}^T \exp \left\{ -\frac{1}{2} (\lambda_k^{(s)} - Y_k^{(g)})^2 (\xi_{s,k}^{-1} + 1)^{-1} \right\} (\xi_{s,k}^{-1} + 1)^{-\frac{1}{2}},$$

$$(\psi_2^{(s)})_{g,g'} = \sigma^4 \prod_{k=1}^T \exp \left\{ -\frac{1}{4} (Y_k^{(g)} - Y_k^{(g')})^2 - \frac{(\lambda_k^{(s)} - \bar{Y}_k)^2}{(2\xi_{s,k}^{-1} + 1)} \right\} (2\xi_{s,k}^{-1} + 1)^{-\frac{1}{2}}.$$

To update  $\mu^{(y)}$  and  $\Sigma^{(y)}$ , we calculate

$$\mu^{(y)} = \kappa \Sigma^{(y)} K_{GG}^{-1} \mathbb{E}_q[K_{GL}] \mathbf{h},$$

$$\Sigma^{(y)} = \left( K_{GG}^{-1} + \kappa K_{GG}^{-1} \mathbb{E}_q[K_{GL} K_{GL}^\top] K_{GG}^{-1} \right)^{-1}.$$

Finally, for  $\lambda^{(s)}$ , we have:

$$\frac{\partial \mathcal{L}}{\partial \lambda^{(s)}} = -\zeta \lambda^{(s)} + \frac{\zeta}{N_s} \sum_{n=1}^{N_s} \gamma_n^{(s)} + \kappa \sum_{s=1}^L h^{(s)} \frac{\partial E_q[k_{Gc^{(s)}}]}{\partial \lambda^{(s)}} K_{GG}^{-1} \mu^{(y)}$$

$$- \frac{\kappa}{2} \sum_{s=1}^L \text{tr} \left( \frac{\partial \mathbb{E}_q[k_{Gc^{(s)}} k_{Gc^{(s)}}^\top] + I}{\partial \lambda^{(s)}} K_{GG}^{-1} (\mu^{(y)} \mu^{(y)\top} + \Sigma^{(y)}) K_{GG}^{-1} \right).$$

## References

- [1] A. Ahmed and E. Xing. Timeline: A dynamic hierarchical Dirichlet process model for recovering birth/death and evolution of topics in text stream. In *UAI*, 2010.
- [2] J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B*, 44(2):139–177, 1982.
- [3] M. Babaei, A. Chakraborty, J. Kulshrestha, E. Redmiles, M. Cha, and K. Gummadi. Analysing biases in perception of truth in news stories and their implications for fact checking. 2018.
- [4] O. Balmau, R. Guerraoui, A.-M. Kermarrec, A. Maurer, M. Pavlovic, and W. Zwaenepoel. Limiting the spread of fake news on social media platforms by evaluating users’ trustworthiness. *arXiv preprint arXiv:1808.09922*, 2018.
- [5] A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi. Science vs conspiracy: Collective narratives in the age of misinformation. *PloS one*, 10(2):e0118093, 2015.
- [6] A. Bessi, A. Scala, L. Rossi, Q. Zhang, and W. Quattrociocchi. The economy of attention in the age of (mis) information. *Journal of Trust Management*, 1(1):12, 2014.
- [7] D. Blei and M. Jordan. Variational inference for Dirichlet process mixtures. *Bayesian analysis*, 1(1):121–143, 2006.
- [8] D. Blei and J. Lafferty. Correlated topic models. In *NIPS*, 2005.
- [9] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, 2011.
- [10] L. Csató and M. Oppor. Sparse on-line Gaussian processes. *Neural computation*, 14(3):641–668, 2002.
- [11] A. Damianou, M. Titsias, and N. Lawrence. Variational inference for latent variables and uncertain inputs in Gaussian processes. *JMLR*, 17(1):1425–1486, 2016.
- [12] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, E. Stanley, and W. Quattrociocchi. The spreading of misinformation online. *PNAS*, 113(3):554–559, 2016.
- [13] L. Devroye. Sample-based non-uniform random variate generation. In *Winter Simulation Conference*, 1986.
- [14] L. Dietz, S. Bickel, and T. Scheffer. Unsupervised prediction of citation influences. In *ICML*, 2007.
- [15] N. Du, M. Farajtabar, A. Ahmed, A. Smola, and L. Song. Dirichlet-Hawkes processes with applications to clustering continuous-time document streams. In *KDD*, 2015.
- [16] N. Du, L. Song, H. Woo, and H. Zha. Uncover topic-sensitive information diffusion networks. In *AISTATS*, 2013.
- [17] M. Farajtabar, J. Yang, X. Ye, H. Xu, R. Trivedi, E. Khalil, S. Li, L. Song, and H. Zha. Fake news mitigation via point process based intervention. In *ICML*, 2017.
- [18] T. Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.
- [19] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng. Rumor cascades. In *ICWSM*, 2014.
- [20] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *WWW*, 2013.

- [21] X. He, T. Rekatsinas, J. Foulds, L. Getoor, and Y. Liu. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades. In *ICML*, 2015.
- [22] J. Holcomb, J. Gottfried, A. Mitchell, and J. Schillinger. News use across social media platforms. *Pew Research Journalism Project*, 2013.
- [23] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.
- [24] M. Kandemir, T. Kekeç, and R. Yeniterzi. Supervising topic models with Gaussian processes. *Pattern Recognition*, 77:226–236, 2018.
- [25] D. Kim and A. Oh. Hierarchical Dirichlet scaling process. *Machine Learning*, 3(106):387–418, 2017.
- [26] J. Kim, D. Kim, and A. Oh. Joint modeling of topics, citations, and topical authority in academic corpora. *TACL*, 5:191–204, 2017.
- [27] J. Kim, B. Tabibian, A. Oh, B. Schölkopf, and M. Gomez Rodriguez. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *WSDM*, 2018.
- [28] S. Kumar and N. Shah. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*, 2018.
- [29] S. Kumar, R. West, and J. Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *WWW*, 2016.
- [30] K. Kurihara, M. Welling, and Y. W. Teh. Collapsed variational Dirichlet process mixture models. In *IJCAI*, 2007.
- [31] S. Kwon, M. Cha, and K. Jung. Rumor detection over varying time windows. *PLoS one*, 12(1):e0168344, 2017.
- [32] S. Kwon, M. Cha, K. Jung, W. Chen, et al. Prominent features of rumor propagation in online social media. In *ICDM*, 2013.
- [33] N. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *NIPS*, 2004.
- [34] N. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *JMLR*, 6(Nov):1783–1816, 2005.
- [35] P. Liang, S. Petrov, M. Jordan, and D. Klein. The infinite PCFG using hierarchical Dirichlet processes. In *EMNLP-CoNLL*, 2007.
- [36] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah. Real-time rumor debunking on twitter. In *CIKM*, 2015.
- [37] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI*, 2016.
- [38] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong. Detect rumors using time series of social context information on microblogging websites. In *CIKM*, 2015.
- [39] J. Ma, W. Gao, and K.-F. Wong. Detect rumors in microblog posts using propagation structure via kernel learning. In *ACL*, 2017.
- [40] J. Ma, W. Gao, and K.-F. Wong. Rumor detection on twitter with tree-structured recursive neural networks. In *ACL*, 2018.

- [41] C. Mavroforakis, I. Valera, and M. Gomez Rodriguez. Modeling the dynamics of online learning activity. In *WWW*, 2017.
- [42] J. Mcauliffe and D. Blei. Supervised topic models. In *NIPS*, 2008.
- [43] A. McCallum, A. Corrada-Emmanuel, and X. Wang. Topic and role discovery in social networks. In *IJCAI*, 2005.
- [44] D. Mimno and A. McCallum. Expertise modeling for matching papers with reviewers. In *KDD*, 2007.
- [45] D. Mimno and A. McCallum. Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. *UAI*, 2008.
- [46] D. Mocanu, L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi. Collective attention in the age of (mis) information. *Computers in Human Behavior*, 51:1198–1204, 2015.
- [47] A. Nguyen, A. Kharosekar, S. Krishnan, S. Krishnan, E. Tate, B. Wallace, and M. Lease. Believe it or not: Designing a human-AI partnership for mixed-initiative fact-checking. In *UIST*, 2018.
- [48] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz. Containment of misinformation spread in online social networks. In *WebSci*, 2012.
- [49] J. Paisley, C. Wang, and D. Blei. The discrete infinite logistic normal distribution. *Bayesian Analysis*, 7(4):997–1034, 2012.
- [50] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *EMNLP*, 2011.
- [51] D. Ramage, D. Hall, R. Nallapati, and C. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *EMNLP*, 2009.
- [52] K. Rapoza. Can ‘fake news’ impact the stock market, Feb 2017.
- [53] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *EMNLP*, 2017.
- [54] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, 2004.
- [55] G. Roussas. *An introduction to measure-theoretic probability*. Academic Press, 2014.
- [56] M. Seeger, C. Williams, and N. Lawrence. Fast forward selection to speed up sparse Gaussian process regression. In *AISTATS*, 2003.
- [57] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.
- [58] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In *NIPS*, 2006.
- [59] M. Tambuscio, G. Ruffo, A. Flammini, and F. Menczer. Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In *WWW*, 2015.
- [60] Y. W. Teh, M. Jordan, M. Beal, and D. Blei. Sharing clusters among related groups: Hierarchical Dirichlet processes. In *NIPS*, 2005.
- [61] Y. W. Teh, K. Kurihara, and M. Welling. Collapsed variational inference for HDP. In *NIPS*, 2008.
- [62] M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *AISTATS*, 2009.
- [63] M. Titsias and N. Lawrence. Bayesian Gaussian process latent variable model. In *AISTATS*, 2010.



- [64] R. M. Tripathy, A. Bagchi, and S. Mehta. A study of rumor control strategies on social networks. In *CIKM*, 2010.
- [65] S. Tschitschek, A. Singla, M. Gomez Rodriguez, A. Merchant, and A. Krause. Fake news detection in social networks via crowd signals. In *WWW*, 2018.
- [66] N. Vo and K. Lee. The rise of guardians: Fact-checking URL recommendation to combat fake news. In *SIGIR*, 2018.
- [67] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.
- [68] M. Wainwright and M. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.
- [69] S. Wang, X. Hu, P. Yu, and Z. Li. MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades. In *KDD*, 2014.
- [70] K. Wu, S. Yang, and K. Q. Zhu. False rumors detection on sina weibo by propagation structures. In *ICDE*, 2015.
- [71] L. Wu and H. Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *WSDM*, 2018.
- [72] F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on sina weibo. In *KDD*, 2012.
- [73] Z. Zhao, P. Resnick, and Q. Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *WWW*, 2015.