Beyond Coherence: Improving Temporal Consistency and Interpretability in Dynamic Topic Models

Thanh Vinh Nguyen[1*], Ngo Van Dong[1*], Minh Chu Xuan[1*], Tung Nguyen[1†], Linh Van Ngo[1], Sang Dinh[1], Trung Le[2]
[1] Hanoi University of Science and Technology, Vietnam,
[2] University of Monash, Australia,

Abstract

Dynamic topic models aim to reveal how themes emerge, evolve, and dissolve in time-stamped corpora, but existing approaches still face three major challenges: (i) encoders capture bag-of-words statistics but fail to align with the rich semantic priors of large pre-trained language models, (ii) temporal linkages are often modeled as rigid one-to-one chains, limiting the ability to track non-linear evolution such as topic splits or merges, and (iii) interpretability remains shallow, relying on noisy top-word lists that obscure thematic clarity. We propose **L-DNTM** (*LLM-Augmented for **D**ynamic **N**eural **T**opic **M**odel*), a variational framework designed to capture more faithful temporal trajectories. Our model integrates three key components: multi-objective distillation to inject PLM-derived semantic knowledge into the encoder, entropy-regularized optimal transport to align entire topic constellations across time for smooth yet flexible evolution, and LLM-guided refinement to sharpen topicword distributions for improved interpretability. Extensive experiments on diverse corpora show that L-DNTM yields more coherent, temporally consistent, and interpretable topic dynamics, and further enhances downstream classification and clustering tasks.

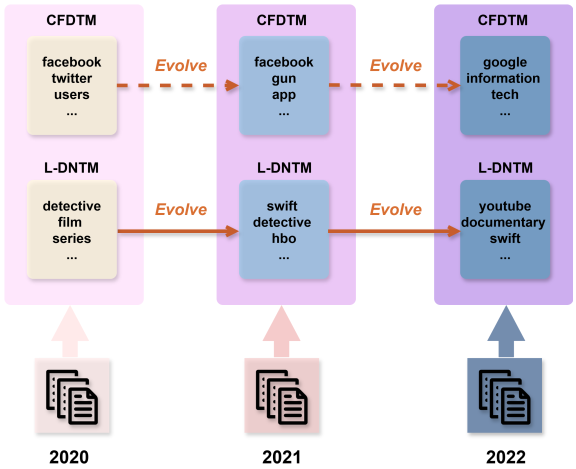

Figure 1: Illustration of dynamic topic modeling. Each time slice (year) contains latent topics, represented by their top words. Arrows indicate how topics evolve across consecutive time slices. While prior models such as CFDTM (Wu et al., 2024) can produce coherent topics, their temporal linkage remains weak, often leading to fragmented or drifting topic trajectories; in contrast, our proposed L-DNTM achieves smoother and more semantically consistent evolution across time.

1 Introduction

Topic modeling provides a compact representation of large text collections by uncovering latent themes (Blei et al., 2003; Blei, 2012). In temporal corporanews, scientific literature, or social mediathese themes are not static: they emerge, drift, split, and merge over time. Dynamic Topic Models (DTMs) therefore aim to characterize evolving semantics across discrete (or continuous) time (Blei and Lafferty, 2006; Wang et al., 2012). Neural approaches based on Variational Autoencoders (VAEs) (Kingma and Ba, 2014; Rezende et al., 2014) have advanced inference and representation learning (Miao et al., 2016; Srivastava and Sutton, 2017; Wu et al., 2020), culminating in dynamic neural topic models such as DETM (Dieng et al., 2019), attention-based designs (Miyamoto et al., 2023), and contrastive formulations (Wu et al., 2024).

Although prior dynamic topic models have shown promising performance in producing coherent topics with high topic quality, they remain limited in capturing temporal consistency and interpretability. **First**, their encoders primarily optimize for bag-of-words reconstruction, which produces locally coherent topics but misaligns with the richer semantic structures available in Pre-trained Language Models (PLMs). **Sec-**

*Equal contribution.
†Corresponding author: tungns@soict.hust.edu.vn

Findings of the Association for Computational Linguistics: EACL 2026, pages 3609–3629
March 24-29, 2026

ond, temporal alignment is often weak, as most models enforce near one-to-one correspondences between topics across adjacent time slices, limiting their ability to capture realistic dynamics such as merges, splits, or many-to-many transitions. **Third**, topic interpretability remains shallow, since it typically relies on short ranked word lists that are vulnerable to generic, noisy, or redundant terms, making it difficult for users to assign clear labels or track semantic shifts over time.

To overcome these limitations, we propose **L-DNTM** (LLM-Augmented for Dynamic Topic Model), a unified framework consisting of three key components: (i) *multi-objective distillation* from PLMs to inject semantic priors into the encoder, (ii) *optimal transport regularization* to capture flexible, many-to-many temporal alignments, and (iii) *LLM-guided refinement* to sharpen topic-word distributions and enhance interpretability.

L-DNTM is trained end-to-end on top of a standard VAE objective, with a warm-up schedule that introduces OT and LLM guidance after the model has learned a stable base of topics. Empirically, L-DNTM improves topic quality, temporal coherence, and labeling readiness over classic DTMs (Blei and Lafferty, 2006), neural baselines (Dieng et al., 2019), and contrastive/attention variants (Wu et al., 2024; Miyamoto et al., 2023). Figure 1 illustrates that while prior models such as CFDTM (Wu et al., 2024) can produce coherent topics, their temporal linkage remains weak, often leading to fragmented or drifting topic trajectories; in contrast, our proposed L-DNTM achieves smoother and more semantically consistent evolution across time. Our work makes the following key contributions:

- We propose a distillation framework that aligns topic proportions with PLM semantics at both the instance level and the latent geometry level, while maintaining accurate reconstruction.

- We design a novel many-to-many temporal evolution regularizer based on entropy-regularized optimal transport, complemented by ETC and UWE techniques to enhance topic diversity and reduce noise.

- We introduce a KLbased refinement strategy that leverages large language models (LLMs) to yield more interpretable topics while promoting the inclusion of emerging words.

- Through extensive experiments on multiple benchmark datasets, we demonstrate that L-DNTM consistently outperforms state-of-the-art dynamic topic models in terms of topic quality, temporal coherence, and performance on downstream tasks.

2 Related Work

Probabilistic Dynamic Topic Models Topic modeling seeks to uncover latent thematic structures in documents in an unsupervised manner, serving as a foundation for many downstream text analysis tasks (Pham et al., 2024b; Nguyen et al., 2025a; Vuong et al., 2025; Le et al., 2025; Vu et al., 2025c; Khanh et al., 2025). The earliest probabilistic formulation of dynamic topic modeling was introduced by Blei and Lafferty (2006), who extended Latent Dirichlet Allocation (LDA) (Blei et al., 2003) into a temporal setting. Their Dynamic Topic Model (DTM) employs a state-space framework in which the natural parameters of latent topics evolve over time through Gaussian noise, with inference approximated via Kalman filtering and wavelet regression. Building on this foundation, Wang et al. (2012) proposed a variant of DTM that operates in continuous time, while Caron and Doucet (2008) further generalized DTM into nonparametric formulations. Since then, numerous extensions have been explored (Wang and McCallum, 2006; Iwata et al., 2010; Bhadury et al., 2016; Jähnichen et al., 2018; Hida et al., 2018), employing either Variational Inference or Gibbs sampling to estimate the evolving topic distributions.

Neural Dynamic Topic Models With the recent advances in neural topic modeling (Miao et al., 2016; Srivastava and Sutton, 2017; Wu et al., 2020; Nguyen et al., 2025e; Vu et al., 2025a; Nguyen et al., 2025c,b,d), there has been a growing interest in extending these models to dynamic settings (Balepur et al., 2023). A seminal contribution in this direction is the Dynamic Embedded Topic Model (DETM) proposed by Dieng et al. (2019), which leverages the Variational Autoencoder (VAE) framework (Kingma and Ba, 2014; Rezende et al., 2014) to jointly model temporal evolution and topicword distributions. Building on this, Zhang and Lauw (2022) incorporated temporal document networks to capture the evolution of topics through document interactions. Cvejoski

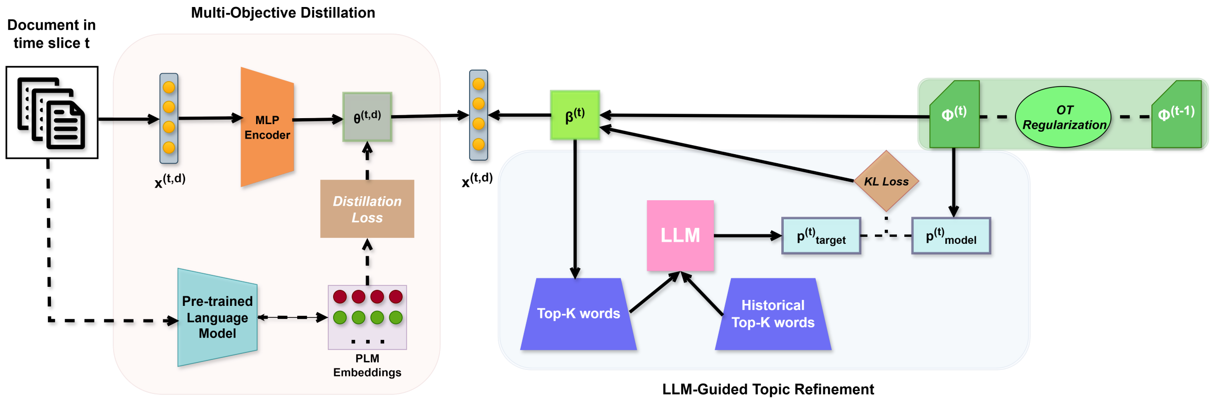

Figure 2: Overview of **L-DNTM**. The encoder leverages PLM-based distillation to enrich topic representations, the decoder applies entropy-regularized optimal transport (OT) to model smooth yet flexible temporal evolution, and an LLM-guided module refines topicword distributions for better interpretability. The framework is trained end-to-end with a composite objective.

et al. (2023) emphasized modeling topic activity patterns across time, while Rahimi et al. (2023) approached the problem by clustering documents to discover dynamic topics. Recent work has explored attention-based inter-topic dependencies (Miyamoto et al., 2023), and pairwise contrastive learning to track topic evolution (Wu et al., 2024).

Dynamic Topic Modeling vs. Streaming Topic Modeling The concept of dynamic topic modeling is often conflated with streaming topic modeling (Nguyen et al., 2021; Linh et al., 2022; Nguyen et al., 2022; Bach et al., 2023; Nguyen et al., 2025f), despite the fact that the two paradigms address fundamentally different problem settings. This confusion mainly arises from their shared emphasis on temporality. However, streaming topic models are designed for online or incremental learning scenarios, where data arrive sequentially and the model is continuously updated at each time step, accumulating knowledge from past observations to refine a single global topic representation. As a result, streaming approaches are typically evaluated using predictive likelihood or held-out perplexity, reflecting their focus on real-time adaptability and forward prediction. In contrast, dynamic topic models assume access to the entire corpus in advance and are trained in an offline manner. Rather than incrementally updating a global model, they explicitly partition the corpus into discrete time slices and learn a sequence of time-dependent topic distributions, where each timestamp maintains its own locally defined topics. Temporal dependencies are then modeled across adjacent time slices to capture smooth topic evolution.

Representation Alignment and Knowledge Transfer. Several studies have explored integrating knowledge from pretrained language models (Wang et al., 2023; Bianchi et al., 2021) such as BERT (Devlin et al., 2019) and GPT (Radford and Narasimhan, 2018). Trained on massive text corpora, these models capture rich linguistic patterns and contextual semantics, providing informative representations that can be fed into the encoder (Zeng et al., 2023; Wang et al., 2023) to enhance a topic models ability to produce coherent and meaningful topics. Building on this idea, Pham et al. (2024b) leveraged pretrained knowledge without incurring costly inference overhead, while effectively modeling semantic interrelationships at the topic level.

Optimal Transport. Optimal Transport (OT) has emerged as a powerful mathematical framework for comparing probability distributions and has found extensive applications across machine learning and related fields (Cuturi, 2013a; Frogner et al., 2015; Seguy et al., 2018; Peyré and Cuturi, 2019). In topic modeling, OT enables more accurate alignment of word and topic distributions, enabling more accurate measurement of semantic distances and relationships (Xu et al., 2023; Zhang and Lauw, 2024; Granese et al., 2025; Vu et al., 2025b). Various efficient solvers and regularization techniques have been proposed to make OT scalable to large datasets and high-dimensional spaces (Flamary et al., 2021), further broadening its applicability in modern machine learning systems.

LLMs in Topic Modeling While Large Language Models (LLMs) (OpenAI, 2023; Touvron et al., 2023) have revolutionized natural language processing, topic models remain highly valuable for large-scale text analysis. LLMs excel at providing fine-grained, context-rich understanding of individual documents; however, when dealing with extensive domain-specific corpora, topic models offer a more interpretable global perspective of thematic structures at a fraction of the computational cost. This complementary relationship has sparked growing interest in leveraging LLMs to enhance topic modeling (Yang et al., 2024; Wang et al., 2023; Doi et al., 2024; Pham et al., 2024a; Mu et al., 2024; Chang et al., 2024).

3 Background

Problem Setup and Notation We observe documents partitioned into T discrete time slices. At time $t \in \{1, \dots, T\}$ the collection contains N_t documents with bag-of-words (BoW) vectors $\mathbf{x}^{(t,d)} \in N^{|V|}$ over a global vocabulary V. The model maintains a shared D-dimensional word embedding matrix $\mathbf{W} \in R^{|V| \times D}$ and time-varying topic embeddings $\Phi^{(t)} \in R^{K \times D}$ with rows $\phi_k^{(t)}$. Given embeddings, each topic k at time t induces a topic-word distribution via a distance-softmax over the vocabulary:

$$\beta_{k,i}^{(t)} = \frac{\exp\Big(-\|\phi_k^{(t)} - \mathbf{w}_i\|^2/\tau_\beta\Big)}{\sum_{j=1}^{|V|} \exp\Big(-\|\phi_k^{(t)} - \mathbf{w}_j\|^2/\tau_\beta\Big)}, \quad (1)$$

where $\mathbf{w}_i$ is the embedding of word i and $\tau_\beta > 0$ is a temperature. A documents topic proportions are $\theta^{(t,d)} \in \Delta^{K-1}$.

Generative story. For each (t, d), we draw a logistic-normal latent $\mathbf{r}^{(t,d)} \sim \mathcal{N}(\mu_0, \Sigma_0)$, map to proportions $\theta^{(t,d)} = \mathrm{softmax}(\mathbf{r}^{(t,d)})$, and generate words with multinomial likelihood $\mathbf{x}^{(t,d)} \sim \mathrm{Mult}\big(n^{(t,d)}, \theta^{(t,d)}\beta^{(t)}\big)$. An amortized encoder $q_\Theta(\mathbf{r}^{(t,d)} \mid \mathbf{x}^{(t,d)})$ produces the variational Gaussian.

Based on the generative process, the core of our objective function is the negative Evidence Lower Bound (ELBO) for the VAE, which we aim to minimize. The VAE loss, summed over all documents and time slices, is composed of a reconstruction term and a KL divergence term that regularizes the latent space:

$$\mathcal{L}_{\mathrm{TM}} = \sum_{t=1}^{T} \sum_{d=1}^{N_t} \Big(- E_{q_\Theta}[\log p(\mathbf{x}^{(t,d)} \mid \theta^{(t,d)}, \beta^{(t)})] + \mathrm{KL}(q_\Theta(\mathbf{r}^{(t,d)} \mid \mathbf{x}^{(t,d)}) \,\|\, p(\mathbf{r}))\Big). \quad (2)$$

The reconstruction term is equivalent to the Negative Log-Likelihood (NLL), while the KL term encourages the learned posterior distributions to stay close to the prior.

Evolution-Tracking Contrastive Learning (ETC) and Unassociated Word Exclusion (UWE). Wu et al. (2024) replaced Markov-chain linkage with a contrastive framework to better capture topic evolution and reduce redundancy. At each time slice t, the embedding of topic k, $\phi_k^{(t)}$, is pulled toward its previous counterpart $\phi_k^{(t-1)}$ via a positive loss weighted by $\lambda(t)$, while a negative loss scaled by γ pushes apart topics within the same slice:

$$\mathcal{L}_{\mathrm{ETC}} = \mathcal{L}_{\mathrm{pos}} + \mathcal{L}_{\mathrm{neg}}.$$

To address "unassociated topics," UWE removes words irrelevant to the current slice. It defines the unassociated set $V_{\mathrm{UW}}^{(t)} = V_{\mathrm{top}}^{(t)} \setminus V^{(t)}$, where $V_{\mathrm{top}}^{(t)}$ are the top-N_{top} words and $V^{(t)}$ the actual slice vocabulary. Topic embeddings $\phi_k^{(t)}$ are then pushed away from these words using

$$\mathcal{L}_{\mathrm{UWE}} = \sum_{t=1}^{T} \sum_{k=1}^{K} \log \sum_{x \in V_{\mathrm{UW}}^{(t)}} \exp\big(g(\phi_k^{(t)}, w_{\mathrm{id}(x)})\big),$$

which sharpens semantics, removes irrelevant words, and improves robustness to varying evolution intensity $\lambda(t)$.

Evolutionary Dynamic Topic Modeling Loss. Combining the above components, our Evolutionary Dynamic Topic Modeling Loss is

$$\mathcal{L}_{\mathrm{EvoDTM}} = \mathcal{L}_{\mathrm{TM}} + \lambda_{\mathrm{ETC}}\mathcal{L}_{\mathrm{ETC}} + \lambda_{\mathrm{UWE}}\mathcal{L}_{\mathrm{UWE}}, \quad (3)$$

where λ_{ETC} and λ_{UWE} control the relative importance of temporal contrastive tracking and unassociated word exclusion.

4 Methodology

We tackle dynamic topic modeling (DTM) with **L-DNTM** (LLM-Augmented for Dynamic Neural

Topic Model), a variational neural topic model that (i) distills semantic structure from a strong pre-trained language model (PLM) into the encoder, (ii) regularizes temporal evolution with entropic optimal transport (OT) to allow splits/merges, and (iii) refines topic interpretability with LLM-guided listwise supervision. L-DNTM trains end-to-end with a single composite objective and a warm-up schedule for stability. Figure 2 shows the overview of our framework.

4.1 Distillation via Multi-Objective Optimization from Pretrained Language Models

To enrich the encoder of our dynamic topic model with semantic priors from Pre-trained Language Models (PLMs) while preserving efficiency, we propose a multi-objective distillation framework, which is motivated by recent work on neural topic model distillation (Bianchi et al., 2021; Pham et al., 2024b), where PLMs serve as teachers guiding the temporal latent space. At each time slice t, the encoder outputs a topic proportion $\theta^{(t,d)}$ that is jointly optimized for (i) reconstructing slice-specific word distributions to capture local dynamics and (ii) aligning with PLM embeddings to maintain global semantic consistency. This integration ensures topic trajectories that evolve smoothly while preserving rich semantic structure, effectively bridging temporal modeling with neural semantics.

Notation. Let the entire collection of documents across all timestamps be:

$$D = \bigcup_{t=1}^{T} D^{(t)} = \{d_1, d_2, \ldots, d_{|D|}\},$$

where $\mathcal{D}^{(t)}$ denotes the set of documents at timestamp t and $|D| = \sum_{t=1}^{T} |\mathcal{D}^{(t)}|$. For each document $d \in D$, we write its topic proportion as $\theta^{(t,d)}$. The set of all topic proportions is:

$$\Theta = \{\theta^{(t,d)} : d \in D\}.$$

Mutual-Information Alignment. The first objective maximizes the mutual information $I(\mathbf{X}_{\text{PLM}}; \Theta)$ between the distribution of topic proportions Θ and the distribution of PLM embeddings $\mathbf{X}_{\text{PLM}}$. To obtain a tractable optimization, we maximize the InfoNCE lower bound (van den Oord et al., 2019):

$$I(\mathbf{X}_{\text{PLM}}; \Theta) \geq \log B - \mathcal{L}_{\text{InfoNCE}}, \tag{4}$$

where B denotes a set of topic proportions drawn from documents within the same temporal batch. Since B is constant, maximizing the mutual information is equivalent to *minimizing* the InfoNCE loss:

$$\mathcal{L} = \frac{-1}{D} \sum_{d=1}^{D} \sum_{t=1}^{T} \log \frac{e^{f(\theta^{(t,d)}, \mathbf{x}_{i,\text{PLM}})}}{\sum_{\theta'^{(t,d)} \in B} e^{f(\theta'^{(t,d)}, \mathbf{x}_{d,\text{PLM}})}}, \tag{5}$$

where D is the number of documents at time t. The similarity function is defined as

$$f(a, b) = \frac{\langle \phi_\theta(a), b \rangle}{\|\phi_\theta(a)\| \, \|b\|},$$

where ϕ_θ is a learnable linear projection.

Temporal Structural Alignment. While InfoNCE enforces instance-level semantic alignment within each time slice, dynamic topic models also require global structural preservation across time. To capture the pairwise document relationships within each temporal batch, we employ the Centered Kernel Alignment (CKA) loss (Cárdenas-Peña et al., 2016):

$$\mathcal{L}_{\text{CKA}} = 1 - \text{CKA}(\Theta, \mathbf{E}_{\text{PLM}}), \tag{6}$$

where $\mathbf{E}_{\text{PLM}}$ denotes the embeddings from a Pre-trained Language Model.

Distillation Objective. The overall distillation loss is a weighted combination of the instance-level and structure-level objectives:

$$\mathcal{L}_{\text{distill}} = \lambda_{\text{InfoNCE}} \mathcal{L}_{\text{InfoNCE}} + \lambda_{\text{CKA}} \mathcal{L}_{\text{CKA}}. \tag{7}$$

4.2 Temporal Regularization via Entropic Optimal Transport

Local smoothness alone cannot model non-isomorphic dynamics (splits/merges). We therefore align the *sets* of topics across adjacent slices with entropic OT (Peyré and Cuturi, 2020), which shifts the paradigm from matching individual topics to aligning the entire geometric constellation of the topic space over time. We assume that each topic contains an equal amount of information and conduct a process where information is transferred between topics based on their relationships, ensuring the total amount remains unchanged. Let $\Phi^{(t-1)} = \{\phi_i^{(t-1)}\}_{i=1}^{K}$ and $\Phi^{(t)} = \{\phi_j^{(t)}\}_{j=1}^{K}$ with uniform masses $\mathbf{a} = \mathbf{b} = \frac{1}{K}\mathbf{1}$, and the cost matrix $\mathbf{C}_{ij}^{(t)} = \|\phi_i^{(t-1)} - \phi_j^{(t)}\|_2^2$. The entropic OT loss for step $t-1 \to t$ is:

$$\mathcal{L}_{\text{OT}}^{(t)} = \min_{\mathbf{P}\in\mathcal{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P}, \mathbf{C}^{(t)} \rangle + \varepsilon \sum_{ij} \mathbf{P}_{ij} \big(\log \mathbf{P}_{ij} - 1 \big). \tag{8}$$

This optimization problem is solved with exceptional efficiency using the Sinkhorn-Knopp algorithm (Cuturi, 2013b). Averaging across time, the complete loss function is:

$$\mathcal{L}_{\text{OT}} = \frac{1}{T-1} \sum_{t=2}^{T} \mathcal{L}_{\text{OT}}^{(t)}. \tag{9}$$

4.3 LLM-Guided Topic Refinement

During training, dynamic topic models represent evolving themes as word distributions over time. However, the highest-probability words at each slice often suffer from poor interpretability, being overly generic (e.g., model, data), ambiguous, or noisy, which obscures the trajectory of topic evolution. To better capture temporal dynamics, inspired by Yang et al. (2024), we propose an LLM-driven refinement framework that continuously tracks and adjusts topic vocabularies across timestamps. At each time t, we prompt an LLM with a structured query that merges the historical top words of topic k with its current candidate set. The LLM then (i) discards terms that are nonsensical or irrelevant to the evolving context, (ii) assigns novelty scores in $[0.0, 1.0]$ to highlight emerging words relative to topic history, and (iii) returns a temporally-aware ranking of the refined keyword set $\mathcal{W}_k^{(t)}$. This process ensures that topics evolve smoothly while incorporating meaningful innovations over time.

Let $q^{(t,k)}(w)$ be the LLM scores temperature-softmaxed to a target listwise distribution, we define the distribution represents the LLMs "ideal" ranking:

$$p_{\text{target}}^{(t,k)}(w) = \frac{\exp\left(q^{(t,k)}(w)/\tau'\right)}{\sum_{u\in\mathcal{W}_k^{(t)}} \exp\left(q^{(t,k)}(u)/\tau'\right)}, \tag{10}$$

and the models belief induced from cosine similarities between $\phi_k^{(t)}$ and $\mathbf{w}$:

$$p_{\text{model}}^{(t,k)}(w) = \frac{\exp\big(\cos(\phi_k^{(t)}, \mathbf{w})/\tau\big)}{\sum_{u\in\mathcal{W}_k^{(t)}} \exp\big(\cos(\phi_k^{(t)}, \mathbf{w}_u)/\tau\big)}. \tag{11}$$

We apply a KL loss:

$$\mathcal{L}_{\text{LLM}} = E_{(t,k)}\left[\text{KL}\Big(p_{\text{target}}^{(t,k)} \,\Big\|\, p_{\text{model}}^{(t,k)}\Big)\right]. \tag{12}$$

By minimizing this objective, we apply a targeted gradient that refines the topic embeddings to better reflect the LLMs sophisticated, context-aware ranking, leading to more coherent and interpretable topics. This nudges topics toward LLM-adjudicated, temporally consistent semantics without overriding the probabilistic decoder.

4.4 Overall Training Objective

We train with a short curriculum: diversity losses are active from the start; both OT and LLM refinement are initiated after warm-up epochs (E_{OT} and E_{LLM}):

$$\begin{aligned}\mathcal{L}_{\text{L-DNTM}} = {} & \mathcal{L}_{\text{EvoDTM}} + \lambda_{\text{distill}}\mathcal{L}_{\text{distill}} \\ & + I(e > E_{\text{OT}})\lambda_{\text{OT}}\mathcal{L}_{\text{OT}} \\ & + I(e > E_{\text{LLM}})\lambda_{\text{LLM}}\mathcal{L}_{\text{LLM}}.\end{aligned} \tag{13}$$

5 Experimental Setup

5.1 Datasets

We conduct experiments on several benchmark corpora: (i) **NeurIPS** contains papers published between 1987 and 2017 at the NeurIPS conference; (ii) **ACL** (Bird et al., 2008), a collection of articles from the ACL Anthology spanning 1973–2006; (iii) **UN** (Baturo et al., 2017), which provides statement transcripts from United Nations meetings between 1970 and 2015; (iv) **NYT**, a New York Times news dataset covering 2012–2022 with 12 categories such as “Arts”, “Business”, and “Health”; and (v) **WHO** (Li et al., 2020), containing weekly articles on non-pharmacological interventions released by the World Health Organization from January to May 2020. All datasets are preprocessed using the `TopMost` toolkit (Wu et al., 2023b). Note that NYT and WHO consist of relatively shorter documents compared with the others.

5.2 Baselines

We evaluate our approach against a range of representative dynamic topic models: (i) **DTM** (Blei and Lafferty, 2006), a classical probabilistic dynamic topic model; (ii) **NDTM** (Cvejoski et al., 2023), which extends DTM with neural variational inference; (iii) **NDTM-b**, a variant of NDTM that leverages our Eq.(1) to parameterize the topic–word distributions, and is introduced here to enable a fair comparison; (iv) **DETM** (Dieng et al., 2019), a neural dynamic topic model incorporating pre-trained word embeddings; (v)

Model	NeurIPS			ACL			UN			NYT			WHO		
	TQ	TTQ	DTQ	TQ	TTQ	DTQ	TQ	TTQ	DTQ	TQ	TTQ	DTQ	TQ	TTQ	DTQ
DTM	0.385†	0.090†	0.237†	0.415†	0.109†	0.262†	0.199†	0.080†	0.139†	0.263†	0.139†	0.201†	0.350†	0.035†	0.192†
NDTM	0.420†	0.095†	0.258†	0.440†	0.115†	0.278†	0.235†	0.085†	0.160†	0.310†	0.145†	0.228†	0.385†	0.042†	0.214†
NDTM-b	0.432†	0.098†	0.265†	0.452†	0.118†	0.285†	0.250†	0.088†	0.169†	0.325†	0.148†	0.237†	0.395†	0.045†	0.220†
DETM	0.392†	0.093†	0.242†	0.423†	0.112†	0.268†	0.208†	0.082†	0.145†	0.271†	0.141†	0.207†	0.358†	0.037†	0.196†
BERTopic	0.536†	0.090†	0.317†	0.512†	0.121†	0.321†	0.323†	0.071†	0.197†	0.538†	0.109†	0.324†	**0.483**†	**0.076**†	**0.275**†
DSNTM	0.465†	0.102†	0.284†	0.475†	**0.125**†	0.300†	0.280†	0.092†	0.186†	0.385†	0.155†	0.270†	0.445†	-0.050†	0.248†
CFDTM	0.476†	0.062†	0.269†	0.431†	0.072†	0.252†	0.390†	0.059†	0.231†	0.580†	0.118†	0.349†	0.460†	-0.031†	0.223†
L-DNTM	**0.611**†	**0.106**†	**0.359**†	**0.572**†	0.120†	**0.346**†	**0.586**†	**0.106**†	**0.346**†	**0.654**†	**0.194**†	**0.422**†	0.465†	0.054†	0.260†

Table 1: Comprehensive quality assessment with Topic Quality (TQ), Temporal Topic Quality (TTQ), and Dynamic Topic Quality (DTQ). The best are in **bold**. † indicates statistical significance ($p<0.05$).

BERTopic (Grootendorst, 2022), a clustering-based topic discovery method built upon document embeddings; (vi) **DSNTM** (Miyamoto et al., 2023), a recent neural model employing an attention mechanism to capture dependencies among evolving topics; and (vii) **CFDTM** (Wu et al., 2024), a novel chain-free neural dynamic topic model employing evolution-tracking contrastive learning and unassociated word exclusion to address repetitive and unassociated topic issues.

5.3 Metrics

We adopt a comprehensive evaluation framework that considers both the static quality of topics at each time slice and their temporal dynamics. For static quality, we follow Dieng et al. (2019) and compute two metrics for every time period. **Topic Coherence (TC)** evaluates the semantic interpretability of the top words in a topic using the C_V score (Röder et al., 2015), while **Topic Diversity (TD)** measures the distinctiveness of topics within the same slice (Dieng et al., 2020). Their product defines the **Topic Quality (TQ)** for that slice, i.e., $\mathrm{TQ} = \mathrm{TC} \times \mathrm{TD}$.

To capture temporal evolution, we incorporate two measures from James et al. (2024). **Temporal Topic Coherence (TTC)** quantifies the semantic consistency of a topic across consecutive slices, and **Temporal Topic Smoothness (TTS)** reflects the gradualness of vocabulary change over time. The average product of these two metrics yields the **Temporal Topic Quality (TTQ)** for each topic.

Finally, we define the **Dynamic Topic Quality (DTQ)** score by combining static and temporal aspects, computed as the mean of TQ across time slices and TTQ across topics. All experiments use $K = 50$ topics, and results are averaged over five independent runs.

6 Results

6.1 Dynamic Topic Quality

Table 1 summarizes results on five benchmarks. L-DNTM consistently achieves the best Dynamic Topic Quality (DTQ) and strong or state-of-the-art scores in Topic Quality (TQ) and Temporal Topic Quality (TTQ). It outperforms all baselines on NeurIPS, UN, and NYT, and remains the overall strongest model on ACL.

These gains stem from three contributions: (i) multi-objective distillation transfers semantic structure from PLMs, boosting TQ; (ii) optimal transport regularization ensures smooth yet flexible temporal evolution, improving TTQ and DTQ; and (iii) LLM-guided refinement sharpens topic-word distributions for greater interpretability.

On WHO, BERTopic performs best across all metrics, likely due to the repetitive and domain-specific nature of health reports with limited temporal variation. In this setting, clustering-based methods are more effective, though L-DNTM still delivers competitive results.

6.2 Temporal Evaluation

Table 2 reports the temporal component analysis. L-DNTM achieves the highest Temporal Topic Coherence (TTC) on all datasets, showing stronger semantic consistency over time. For Temporal Topic Smoothness (TTS), baselines often produce overly high values that risk topic stagnation, while L-DNTM yields more moderate scores (0.5550.629), striking a better balance between stability and flexibility. Overall, L-DNTM improves coherence and avoids over-smoothing, resulting in more adaptive topic trajectories.

6.3 Static topic quality

Table 3 presents the static topic quality results in terms of Topic Coherence (TC) and Topic Di-

Model	NeurIPS		ACL		UN	
	TTC	TTS	TTC	TTS	TTC	TTS
DTM	0.118†	0.754	0.159†	0.646	0.089†	0.888
NDTM	0.125†	0.782	0.165†	0.672	0.095†	0.905
NDTM-b	0.132†	0.798	0.168†	0.685	0.098†	0.912
DETM	0.159†	0.856	0.146†	0.857	0.089†	0.888
BERTopic	0.142†	0.609	0.162†	0.676	0.090†	0.879
DSNTM	0.159†	0.910	0.141†	0.497	0.085†	0.690
CFDTM	0.106†	0.532	0.132†	0.497	0.100†	0.590
L-DNTM	**0.183**†	0.555	**0.195**†	0.594	**0.160**†	0.629

Table 2: Temporal component analysis showing Temporal Topic Coherence (TTC) and Temporal Topic Smoothness (TTS) across datasets. The best results for **TTC** are in **bold**. For **TTS**, moderate values are desirable as they indicate a balance between stability and evolution, whereas excessively high values can signal topic stagnation. † indicates statistical significance ($p<0.05$).

Model	NeurIPS		ACL		NYT	
	TC	TD	TC	TD	TC	TD
DTM	0.440†	0.256†	0.468†	0.334†	0.464†	0.632†
NDTM	0.445†	0.612†	0.459†	0.525†	0.473†	0.446†
NDTM-b	0.448†	0.618†	0.467†	0.603†	0.424†	0.479†
DETM	0.428†	0.256†	0.447†	0.334†	0.372†	0.368†
BERTopic	0.502†	0.494†	0.392†	0.264†	0.412†	0.597†
DSNTM	0.427†	0.685†	0.370†	0.609†	0.374†	0.414†
CFDTM	0.492†	0.830†	0.441†	0.839†	0.555†	0.692†
L-DNTM	**0.594**†	**0.841**†	**0.554**†	**0.853**†	**0.629**†	**0.862**†

Table 3: Topic coherence and stability analysis showing Topic Coherence (TC) and Topic Diversity (TD) across datasets. The best results are in **bold**. † indicates statistical significance ($p<0.05$).

versity (TD). L-DNTM consistently achieves the best performance across all datasets, surpassing all baselines, including CFDTM, on both metrics.

In particular, L-DNTM obtains the highest TC on NeurIPS (0.594), ACL (0.554), and NYT (0.629), while also achieving the highest TD scores of 0.841, 0.853, and 0.862, respectively. These results indicate that L-DNTM learns topics that are both semantically coherent and highly diverse.

Compared to CFDTM, which improves diversity via contrastive learning, L-DNTM further enhances this trade-off by simultaneously improving coherence and diversity, leading to more informative and interpretable topic representations.

6.4 Ablations

Table 4 reports the ablation study on the NYT and NeurIPS datasets. We observe that removing either distillation or optimal transport leads to clear drops in both Topic Quality (TQ) and Temporal

Model Variant	NeurIPS			NYT		
	TQ	TTC	TTS	TQ	TTC	TTS
w/o LLM	0.608†	**0.212**†	*0.941*	**0.677**†	**0.299**†	*0.933*
w/o Distill	0.583†	0.154†	0.545	0.606†	0.202†	0.517
w/o OT	0.605†	0.174†	0.520	0.618†	0.233†	0.568
w/o Distill + OT	0.578†	0.146†	0.508	0.583†	0.180†	0.440
L-DNTM	**0.611**†	0.183†	0.555	0.654†	0.270†	0.685

Table 4: Ablation study on L-DNTM with Topic Quality (TQ), Temporal Topic Coherence (TTC), and Temporal Topic Smoothness (TTS) on NeurIPS and NYT datasets. The best results for TQ and TTC are in **bold**. The highest TTS scores are *underlined and italicized* to highlight potential topic stagnation, as excessively high values are not desirable. † indicates statistical significance ($p<0.05$).

Topic Coherence (TTC), confirming their complementary roles in semantic transfer and smooth evolution. The full L-DNTM achieves consistently strong performance across all three metrics, balancing topic interpretability with adaptive temporal dynamics.

Interestingly, the w/o LLM variant shows slightly higher TQ and TTC compared to the full model, suggesting that pretrained knowledge is not strictly necessary for local coherence. However, this comes at the cost of excessively high TTS values (0.933 on NYT and 0.941 on NeurIPS), indicating topic stagnation where topics fail to evolve meaningfully over time. By contrast, L-DNTM achieves more moderate TTS values (0.685 and 0.555), striking a better balance between stability and evolution.

These findings highlight that LLM integration is essential not only for semantic enrichment but also for avoiding rigid, stagnant topic trajectories, enabling L-DNTM to maintain both high quality and dynamic adaptability.

6.5 Text Classification and Clustering

In addition to evaluating dynamic topic quality, we assess the learned documenttopic distributions on downstream tasks using the NYT dataset, following Wu et al. (2024). For classification, SVMs are trained on documenttopic vectors to predict categories, while clustering assigns each document to its most probable topic and is evaluated with Purity and NMI (Wu et al., 2023a). Our focus is on evaluating the quality of document representations, not on pushing state-of-the-art results.

Figure 3 visualizes the performance of all models on text classification (Accuracy and F1) and

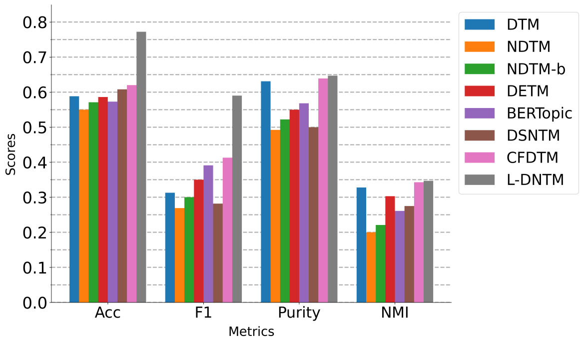

Figure 3: Performance metrics for text classification (Accuracy and F1) and clustering (Purity and NMI)

clustering (Purity and NMI). Our proposed **L-DNTM** consistently outperforms the baselines, achieving the highest scores across all four metrics. In particular, it yields substantial improvements over CFDTM (Wu et al., 2024), highlighting that the documenttopic representations learned by L-DNTM are more discriminative and semantically consistent, which translates into stronger performance on downstream tasks.

7 Conclusion

We proposed L-DNTM, a dynamic neural topic model that integrates PLM-based distillation, OT-based temporal regularization, and LLM-guided refinement. Experiments on multiple benchmarks show that L-DNTM consistently outperforms strong baselines in topic quality, temporal coherence, and downstream tasks. It effectively balances coherence and diversity, producing robust and interpretable topic trajectories.

8 Limitations

Our proposed method has several limitations. On the encoder side, its reliance on PLM-based distillation introduces potential bias and computational overhead; future work could explore lighter distillation methods or reduce dependence on large PLMs. For the decoder, optimal transport remains costly when scaling to many topics or long horizons, suggesting the need for more efficient or approximate OT solvers. The LLM-guided refinement requires repeated prompting, which may limit scalability and propagate errors; adaptive or budgeted refinement strategies could mitigate this issue. Finally, our experiments are restricted to English corpora, leaving multilingual and noisy domains unexploredan important direction for broader evaluation.

Acknowledgments

This project is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2025.16.

References

Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Slav Petrov, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy P. Lillicrap, and 33 others. 2023. Gemini: A family of highly capable multimodal models. *CoRR*, abs/2312.11805.

Jason Ansel, Edward Z. Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, and 30 others. 2024. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2, ASPLOS 2024, La Jolla, CA, USA, 27 April 2024- 1 May 2024*, pages 929–947. ACM.

Tran Xuan Bach, Nguyen Duc Anh, Ngo Van Linh, and Khoat Than. 2023. Dynamic transformation of prior knowledge into bayesian models for data streams. *IEEE Trans. Knowl. Data Eng.*, 35(4):3742–3750.

Nishant Balepur, Shivam Agarwal, Karthik Venkat Ramanan, Susik Yoon, Diyi Yang, and Jiawei Han. 2023. Dynamite: Discovering explosive topic evolutions with user guidance. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 194–217. Association for Computational Linguistics.

Alexander Baturo, Niheer Dasandi, and Slava J. Mikhaylov. 2017. Understanding state preferences with text as data: Introducing the un general debate corpus. *Research Politics*, 4(2).

Arnab Bhadury, Jianfei Chen, Jun Zhu, and Shixia Liu. 2016. Scaling up dynamic topic models. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 381–390. ACM.

Federico Bianchi, Silvia Terragni, Dirk Hovy, Debora Nozza, and Elisabetta Fersini. 2021. Cross-lingual contextualized topic models with zero-shot learning. In *Proceedings of the 16th Conference of the*

European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, pages 1676–1683. Association for Computational Linguistics.

Steven Bird, Robert Dale, Bonnie J. Dorr, Bryan R. Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R. Radev, and Yee Fan Tan. 2008. The ACL anthology reference corpus: A reference dataset for bibliographic research in computational linguistics. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.

David M. Blei. 2012. Probabilistic topic models. *Commun. ACM*, 55(4):77–84.

David M. Blei and John D. Lafferty. 2006. Dynamic topic models. In *Machine Learning, Proceedings of the Twenty-Third International Conference (ICML 2006), Pittsburgh, Pennsylvania, USA, June 25-29, 2006*, volume 148 of *ACM International Conference Proceeding Series*, pages 113–120. ACM.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

David Cárdenas-Peña, Diego Fabian Collazos-Huertas, and Germán Castellanos-Domínguez. 2016. Centered kernel alignment enhancing neural network pretraining for mri-based dementia diagnosis. *Comput. Math. Methods Medicine*, 2016:9523849:1–9523849:10.

Francois Caron and Arnaud Doucet. 2008. Sparse bayesian nonparametric regression. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008*, volume 307 of *ACM International Conference Proceeding Series*, pages 88–95. ACM.

Shuyu Chang, Rui Wang, Peng Ren, and Haiping Huang. 2024. Enhanced short text modeling: Leveraging large language models for topic refinement. *CoRR*, abs/2403.17706.

Marco Cuturi. 2013a. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2292–2300.

Marco Cuturi. 2013b. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Kostadin Cvejoski, Ramsés J. Sánchez, and César Ojeda. 2023. Neural dynamic focused topic model. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 12719–12727. AAAI Press.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Adji B. Dieng, Francisco J. R. Ruiz, and David M. Blei. 2019. The dynamic embedded topic model. *CoRR*, abs/1907.05545.

Adji Bousso Dieng, Francisco J. R. Ruiz, and David M. Blei. 2020. Topic modeling in embedding spaces. *Trans. Assoc. Comput. Linguistics*, 8:439–453.

Tomoki Doi, Masaru Isonuma, and Hitomi Yanaka. 2024. Topic modeling for short texts with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, ACL 2024 - Student Research Workshop, Bangkok, Thailand, August 11-16, 2024*, pages 21–33. Association for Computational Linguistics.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T. H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, and 3 others. 2021. POT: python optimal transport. *J. Mach. Learn. Res.*, 22:78:1–78:8.

Charlie Frogner, Chiyuan Zhang, Hossein Mobahi, Mauricio Araya-Polo, and Tomaso A. Poggio. 2015. Learning with a wasserstein loss. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 2053–2061.

Federica Granese, Benjamin Navet, Serena Villata, and Charles Bouveyron. 2025. Merging embedded topics with optimal transport for online topic modeling on data streams. *CoRR*, abs/2504.07711.

Arthur Gretton, Olivier Bousquet, Alexander J. Smola, and Bernhard Schölkopf. 2005. Measuring statistical dependence with hilbert-schmidt norms. In *Algorithmic Learning Theory, 16th International Conference, ALT 2005, Singapore, October 8-11, 2005, Proceedings*, volume 3734 of *Lecture Notes in Computer Science*, pages 63–77. Springer.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based TF-IDF procedure. *CoRR*, abs/2203.05794.

Philipp Hennig, David H. Stern, Ralf Herbrich, and Thore Graepel. 2012. Kernel topic models. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012, La Palma, Canary Islands, Spain, April 21-23, 2012*, volume 22 of *JMLR Proceedings*, pages 511–519. JMLR.org.

Rem Hida, Naoya Takeishi, Takehisa Yairi, and Koichi Hori. 2018. Dynamic and static topic model for analyzing time-series document collections. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 516–520. Association for Computational Linguistics.

Tomoharu Iwata, Takeshi Yamada, Yasushi Sakurai, and Naonori Ueda. 2010. Online multiscale dynamic topic models. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 663–672. ACM.

Patrick Jähnichen, Florian Wenzel, Marius Kloft, and Stephan Mandt. 2018. Scalable generalized dynamic topic models. In *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, volume 84 of *Proceedings of Machine Learning Research*, pages 1427–1435. PMLR.

Charu James, Mayank Nagda, Nooshin Haji Ghassemi, Marius Kloft, and Sophie Fellenz. 2024. Evaluating dynamic topic models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 160–176. Association for Computational Linguistics.

Chi Pham Khanh, Hoang Tran Vuong, Sang Dinh Viet, and Linh Ngo Van. 2025. A framework for neural topic modeling using hierarchical clustering and contrastive learning with optimal transport. *Neurocomputing*, page 132530.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Tue Le, Hoang Tran Vuong, Tung Nguyen, Linh Ngo Van, Dinh Viet Sang, Trung Le, and Thien Huu Nguyen. 2025. Multi-surrogate-objective optimization for neural topic models. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 135–151, Suzhou, China. Association for Computational Linguistics.

Yue Li, Pratheeksha Nair, Zhi Wen, Imane Chafi, Anya Okhmatovskaia, Guido Powell, Yannan Shen, and David L. Buckeridge. 2020. Global surveillance of COVID-19 by mining news media using a multi-source dynamic embedded topic model. In *BCB '20: 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Virtual Event, USA, September 21-24, 2020*, pages 34:1–34:14. ACM.

Ngo Van Linh, Tran Xuan Bach, and Khoat Than. 2022. A graph convolutional topic model for short and noisy text streams. *Neurocomputing*, 468:345–359.

Yishu Miao, Lei Yu, and Phil Blunsom. 2016. Neural variational inference for text processing. In *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1727–1736. JMLR.org.

Nozomu Miyamoto, Masaru Isonuma, Sho Takase, Junichiro Mori, and Ichiro Sakata. 2023. Dynamic structured neural topic model with self-attention mechanism. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 5916–5930. Association for Computational Linguistics.

Yida Mu, Chun Dong, Kalina Bontcheva, and Xingyi Song. 2024. Large language models offer an alternative to the traditional approach of topic modelling. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024, 20-25 May, 2024, Torino, Italy*, pages 10160–10171. ELRA and ICCL.

Duc Anh Nguyen, Ngo Van Linh, Nguyen Kim Anh, Canh Hao Nguyen, and Khoat Than. 2021. Boosting prior knowledge in streaming variational bayes. *Neurocomputing*, 424:143–159.

Ha Nguyen, Hoang Pham, Son Nguyen, Ngo Van Linh, and Khoat Than. 2022. Adaptive infinite dropout for noisy and sparse data streams. *Mach. Learn.*, 111(8):3025–3060.

Quang Duc Nguyen, Tung Nguyen, Duc Anh Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen. 2025a. Glocom: A short text neural topic model via global clustering context. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2025 - Volume 1: Long Papers, Albuquerque, New Mexico, USA, April 29 - May 4, 2025*, pages 1109–1124. Association for Computational Linguistics.

Tien-Phat Nguyen, Vu Minh Ngo, Tung Nguyen, Linh Van Ngo, Duc Anh Nguyen, Sang Dinh, and Trung Le. 2025b. XTRA: cross-lingual topic modeling with topic and representation alignments. *CoRR*, abs/2510.02788.

Tung Nguyen, Duy-Tung Doan, Huu-Thanh Luong, and Ngo Van Linh. 2025c. Sub-document neural topic model. *Knowl. Based Syst.*, 331:114762.

Tung Nguyen, Linh Van Ngo, Duc Anh Nguyen, and Sang Dinh Viet. 2025d. A framework for neural topic modeling with mutual information and group regularization. *Neurocomputing*, 645:130420.

Tung Nguyen, Duy-Tung Pham, Quang Duc Nguyen, Ngo Van Linh, Duc Anh Nguyen, and Sang Dinh Viet. 2025e. Topicot: Neural topic model aligning with pre-trained clustering and optimal transport. *Neurocomputing*, 654:131268.

Tung Nguyen, Tung Pham, Ngo Van Linh, Ha-Bang Ban, and Khoat Than. 2025f. Out-of-vocabulary handling and topic quality control strategies in streaming topic models. *Neurocomputing*, 614:128757.

OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543. ACL.

Gabriel Peyré and Marco Cuturi. 2019. Computational optimal transport. *Found. Trends Mach. Learn.*, 11(5-6):355–607.

Gabriel Peyré and Marco Cuturi. 2020. Computational optimal transport. *Preprint*, arXiv:1803.00567.

Chau Pham, Alexander Miserlis Hoyle, Simeng Sun, Philip Resnik, and Mohit Iyyer. 2024a. Topicgpt: A prompt-based topic modeling framework. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, pages 2956–2984. Association for Computational Linguistics.

Duy-Tung Pham, Thien Trang Nguyen Vu, Tung Nguyen, Linh Ngo Van, Duc Anh Nguyen, and Thien Huu Nguyen. 2024b. Neuromax: Enhancing neural topic modeling via maximizing mutual information and group topic regularization. *Preprint*, arXiv:2409.19749.

Alec Radford and Karthik Narasimhan. 2018. Improving language understanding by generative pre-training.

Hamed Rahimi, Hubert Naacke, Camélia Constantin, and Bernd Amann. 2023. ANTM: an aligned neural topic model for exploring evolving topics. *CoRR*, abs/2302.01501.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. 2014. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1278–1286. JMLR.org.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, Shanghai, China, February 2-6, 2015*, pages 399–408. ACM.

Vivien Seguy, Bharath Bhushan Damodaran, Rémi Flamary, Nicolas Courty, Antoine Rolet, and Mathieu Blondel. 2018. Large scale optimal transport and mapping estimation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. Mpnet: Masked and permuted pre-training for language understanding. *CoRR*, abs/2004.09297.

Akash Srivastava and Charles Sutton. 2017. Autoencoding variational inference for topic models. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *CoRR*, abs/2302.13971.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2019. Representation learning with contrastive predictive coding. *Preprint*, arXiv:1807.03748.

Tu Vu, Manh Do, Tung Nguyen, Ngo Van Linh, Sang Dinh, and Thien Huu Nguyen. 2025a. Topic modeling for short texts via optimal transport-based clustering. In *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pages 7666–7680. Association for Computational Linguistics.

Tu Vu, Manh Do, Tung Nguyen, Ngo Van Linh, Sang Dinh, and Thien Huu Nguyen. 2025b. Topic modeling for short texts via optimal transport-based clustering. In *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pages 7666–7680. Association for Computational Linguistics.

Tu Vu, Manh Do, Tung Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen. 2025c. Topic modeling for short texts via optimal transport-based clustering. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 7666–7680, Vienna, Austria. Association for Computational Linguistics.

Hoang Tran Vuong, Tue Le, Tu Vu, Tung Nguyen, Linh Ngo Van, Sang Dinh, and Thien Huu Nguyen. 2025. Hicot: Improving neural topic models via optimal transport and contrastive learning. In *Findings of the Association for Computational Linguistics, ACL 2025, Vienna, Austria, July 27 - August 1, 2025*, pages 13894–13920. Association for Computational Linguistics.

Chong Wang, David M. Blei, and David Heckerman. 2012. Continuous time dynamic topic models. *CoRR*, abs/1206.3298.

Han Wang, Nirmalendu Prakash, Nguyen-Khoi Hoang, Ming Shan Hee, Usman Naseem, and Roy Ka-Wei Lee. 2023. Prompting large language models for topic modeling. In *IEEE International Conference on Big Data, BigData 2023, Sorrento, Italy, December 15-18, 2023*, pages 1236–1241. IEEE.

Xuerui Wang and Andrew McCallum. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, August 20-23, 2006*, pages 424–433. ACM.

Xiaobao Wu, Xinshuai Dong, Thong Thanh Nguyen, and Anh Tuan Luu. 2023a. Effective neural topic modeling with embedding clustering regularization. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 37335–37357. PMLR.

Xiaobao Wu, Xinshuai Dong, Liangming Pan, Thong Nguyen, and Anh Tuan Luu. 2024. Modeling dynamic topics in chain-free fashion by evolution-tracking contrastive learning and unassociated word exclusion. In *Findings of the Association for Computational Linguistics, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024*, pages 3088–3105. Association for Computational Linguistics.

Xiaobao Wu, Chunping Li, Yan Zhu, and Yishu Miao. 2020. Learning multilingual topics with neural variational inference. In *Natural Language Processing and Chinese Computing - 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14-18, 2020, Proceedings, Part I*, volume 12430 of *Lecture Notes in Computer Science*, pages 840–851. Springer.

Xiaobao Wu, Fengjun Pan, and Anh Tuan Luu. 2023b. Towards the topmost: A topic modeling system toolkit. *CoRR*, abs/2309.06908.

Weijie Xu, Xiaoyu Jiang, Srinivasan Sengamedu Hanumantha Rao, Francis Iannacci, and Jinjin Zhao. 2023. vontss: vmf based semi-supervised neural topic modeling with optimal transport. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 4433–4457. Association for Computational Linguistics.

Xiaohao Yang, He Zhao, Weijie Xu, Yuanyuan Qi, Jueqing Lu, Dinh Phung, and Lan Du. 2024. Neural topic modeling with large language models in the loop. *arXiv preprint arXiv:2411.08534*.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting? In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 11121–11128. AAAI Press.

Delvin Ce Zhang and Hady W. Lauw. 2022. Dynamic topic models for temporal document networks. In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 26281–26292. PMLR.

Delvin Ce Zhang and Hady W. Lauw. 2024. Topic modeling on document networks with dirichlet optimal transport barycenter. *IEEE Trans. Knowl. Data Eng.*, 36(3):1328–1340.

Dataset	#doc	Average Length	Vocab Size	#time slices
NeurIPS	7,237	2,085.9	10,000	31
ACL	10,560	2,023.0	10,000	31
UN	7,507	1,421.6	10,000	46
NYT	9,172	175.4	10,000	11
WHO	12,145	41.3	10,000	15

Table 5: Statistics of datasets.

A Preprocessing

To pre-process the datasets, we follow the procedure outlined in Wu et al. (2023b): (i) tokenize documents and convert all text to lowercase, (ii) remove punctuation marks, (iii) discard tokens containing numerical characters, (iv) filter out tokens shorter than three characters, and (v) eliminate stopwords. The resulting dataset statistics after applying these preprocessing steps are reported in Table 5.

B Implementation Details

Our model is implemented using PyTorch Ansel et al. (2024) and the publicly available topic modeling toolkit TopMost (Wu et al., 2023b). We employ the *"all-mpnet-base-v2"* (Song et al., 2020) as our pre-trained language model. Glove (Pennington et al., 2014) serves as the initial word embedding. Following Wu et al. (2024), the prior distribution is approximated using the Laplace method (Hennig et al., 2012), which serves as a surrogate for the symmetric Dirichlet prior with $\mu_{0,k} = 0$ and $\Sigma_{0,kk} = (K-1)/K$. The encoder network f_Θ is designed as a two-layer MLP with softplus activation, followed by two separate linear layers that output the mean and covariance matrices. For Section 4.3, we implement with *"gemini-2.0-flash-lite"* (Anil et al., 2023) as our LLM.

Hyperparameters Selection We set τ_β in Eq. 1 to 0.7. The parameters τ and τ', appearing in Eq. 11 and Eq. 10, respectively, are both fixed at 0.1. Our models were trained on a single NVIDIA P100 GPU (Kaggle). We employ Adam (Kingma and Ba, 2014) to optimize the model parameters and train our model for 300 epochs with a learning rate of 0.002, In practice, the first 125–150 epochs are allocated to stabilize the topic embeddings before applying the OT regularization, while the LLM guidance is only introduced during the final 5–10 epochs.

See our code for more implementation details.

C Mutual Information Maximization and Temporal Structural Alignment

C.1 Mutual Information Maximization

Let X and Y denote two random variables. The mutual information between them, which measures the degree of statistical dependence, is given by:

$$I(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy. \quad (14)$$

Directly maximizing this objective is generally intractable. To address this, we adopt a tractable lower bound following van den Oord et al. (2019):

$$\begin{aligned} I(X;Y) &\geq \mathcal{L}_{\text{InfoNCE}} \\ &= \log N + E_{p(x,y)} \left[\log \frac{f(x,y)}{\sum_{y' \in \mathcal{B}} f(x,y')} \right]. \end{aligned} \quad (15)$$

Here, $f(x,y)$ denotes a similarity function between x and y, while $\mathcal{B}$ represents the set containing one positive and $N-1$ negative samples. Maximizing this bound encourages each instance to achieve high similarity with its corresponding positive pair, while maintaining low similarity with negative ones.

C.2 Temporal Structure Alignment with Central Kernel Alignment (CKA) loss

Let $G_S^{(l)}$ and $G_T^{(l)}$ denote the Gram matrices of the student and teacher representations at layer l, respectively. To measure their similarity, we make use of the Hilbert-Schmidt Independence Criterion (HSIC) (Gretton et al., 2005), which is defined as:

$$\begin{aligned} \text{HSIC}(G_S, G_T) &= \frac{\text{tr}(G_S H G_T H)}{(n-1)^2} \\ &= \frac{\langle \text{vec}(G_S H), \text{vec}(G_T H) \rangle}{(n-1)^2}, \end{aligned} \quad (16)$$

where $H = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is the centering matrix.

While HSIC accounts for centering, it is not invariant to isotropic scaling. To address this, the Centered Kernel Alignment (CKA) (Cárdenas-Peña et al., 2016) normalizes HSIC as:

$$\text{CKA}(G_S, G_T) = \frac{\text{HSIC}(G_S, G_T)}{\sqrt{\prod_{i \in \{S,T\}} \text{HSIC}(G_i, G_i)}}. \quad (17)$$

Intuitively, CKA focuses on the shape of the distribution rather than the absolute values of the Gram matrices. This normalization makes CKA a more robust similarity measure compared to distance-based metrics such as ℓ_2 or Huber losses. Furthermore, if the centering step in Eq. (16) is omitted, CKA reduces to cosine similarity:

$$\mathrm{CKA}_{\mathrm{WC}}(G_S, G_T) = \cos(G_S, G_T). \tag{18}$$

By leveraging cosine similarity on centered and normalized Gram matrices, CKA yields values in $[0, 1]$ that robustly capture the alignment between teacher and student representations, while allowing for differences in scale.

D Entropy Regularized Optimal Transport

Let u and v be two discrete probability measures supported on $\{x_1, x_2, \ldots, x_n\} \subset R^d$ and $\{y_1, y_2, \ldots, y_m\} \subset R^d$, respectively, with associated weights $(u_1, u_2, \ldots, u_n)$ and $(v_1, v_2, \ldots, v_m)$ such that $\sum_i u_i = \sum_j v_j$. Given a cost matrix $C \in R^{n\times m}$, the optimal transport plan is obtained by solving the following optimization problem (Peyré and Cuturi, 2020):

$$\min_{P\in R^{n\times m}} \langle P, C\rangle \quad \text{subject to } P\mathbf{1} = u,\ P^\top\mathbf{1} = v. \tag{19}$$

The minimum value of this objective defines a distance between the two distributions. To make the computation more efficient, an entropic regularization term can be added to (19), and the resulting problem can be solved using iterative methods such as the Sinkhorn algorithm (Cuturi, 2013a).

E Prompts

For the task of dynamic topic modeling, we constructed a carefully designed prompt to guide the model in refining topic keywords across different time slices. The goal of this prompt is to ensure that each topic at time t is analyzed in light of its historical trajectory, while also capturing new emerging signals. The structure of the prompt can be broken down into four main components:

Semantic Analysis and Summarization. The prompt first asks the model to compare the current set of candidate keywords for a topic with its historical keywords. Based on this comparison, the model must produce a one-sentence summary that captures the core semantic meaning of the topic at time t, together with a description of how it has evolved. This ensures that every subsequent filtering and scoring decision is grounded in a clear semantic interpretation of topic change.

Filtering of Irrelevant Candidates. Next, the model is instructed to remove any candidate words that are typographical errors, meaningless tokens, or unrelated to the identified semantic focus. This step is crucial to avoid noise propagation and to keep the final keyword list semantically coherent.

Scoring with a Differentiated Novelty Scale. The prompt then requires the model to assign a *novelty score* to each remaining candidate keyword, using the full range from 0.0 to 1.0. The scoring scheme is designed to encourage differentiation:

- A score of 1.0 is given to words that are highly relevant but completely new compared to the topics history, representing truly emerging signals.
- Scores between 0.5 and 0.9 are used for words that already appeared in the history but whose importance has significantly increased or shifted in meaning.
- Scores between 0.1 and 0.4 correspond to stable words that remain consistently relevant over time.
- A score of 0.0 is assigned to words that, while still listed as candidates, no longer reflect the topics current focus.

This design explicitly prevents the model from assigning uniform scores, thereby forcing nuanced judgments about word novelty and stability.

Enforcing Structured Output. Finally, the prompt specifies that the output must be a single valid JSON object, without additional explanations or formatting. Each JSON entry contains both the reasoning summary for the topic and the ranked list of candidate keywords with their novelty scores. By enforcing this strict output format, we guarantee that the results are machine-readable and easily integrated into downstream evaluation pipelines.

F Details of temporal metrics.

James et al. (2024) provided the measures for evaluating temporal dynamics in topic models. These measures extend classical topic coherence by incorporating temporal consistency and smoothness across consecutive timestamps.

Temporal Topic Coherence (TTC) Temporal topic coherence (TTC) extends topic coherence (TC) by considering word co-occurrences across consecutive timestamps of the same topic. Formally, for topic k at timestamp t with window size L, we define:

$$\mathrm{TTC}_{k,t} = \sum_{j=1}^{N}\sum_{i=1}^{N} \log \frac{P\big(w_i^{(k,t)}, w_j^{(k,t+L)}\big) + \epsilon}{P\big(w_i^{(k,t)}\big)\, P\big(w_j^{(k,t+L)}\big)} - \log\Big(P\big(w_i^{(k,t)}, w_j^{(k,t+L)}\big) + \epsilon\Big). \tag{20}$$

where $w_i^{(k,t)}$ denotes the i-th word in topic k at timestamp t, N is the number of top words considered, and ϵ is a smoothing constant. A high TTC indicates temporal semantic stability of the topic.

Temporal Topic Smoothness (TTS) Temporal topic smoothness (TTS) measures how gradually the semantic content of a topic evolves across consecutive timestamps. While temporal topic coherence (TTC) captures the degree of semantic consistency by relying on an external reference corpus, TTS is an intrinsic measure that evaluates the smoothness of topic transitions directly from the learned topic representations.

To compute TTS, we adapt the diversity measure $r_{k,\tilde{C}}$ to the temporal axis. Intuitively, if the word distribution vectors of a topic remain very similar over consecutive timestamps, the topic evolves smoothly, leading to a high TTS value. Conversely, if the topic changes abruptly, the diversity across time steps will be large and the TTS value will be low.

Formally, for topic k at timestamp t and a temporal window of size L, the temporal topic smoothness is defined as:

$$\mathrm{TTS}_{k,t} = r_{k,\tilde{C}}, \tag{21}$$

where $\tilde{C} = \big(v^{(k,t)}, v^{(k,t+1)}, \ldots, v^{(k,t+L-1)}\big)$ denotes the sequence of word distribution vectors of topic k across the window. Smooth transitions therefore correspond to low diversity in $\tilde{C}$, and hence to higher values of $\mathrm{TTS}_{k,t}$.

Temporal Topic Quality (TTQ) While TTC relies on a reference corpus, TTS only depends on the topic word distributions. They capture complementary aspects of temporal quality. Combining them yields the temporal topic quality (TTQ), which balances coherence and smoothness:

$$\mathrm{TTQ}_k = \frac{1}{T-L+1} \sum_{t=1}^{T-L+1} \mathrm{TTC}_{k,t} \cdot \mathrm{TTS}_{k,t}, \tag{22}$$

where T is the number of timestamps. The window size L controls the resolution: smaller L detects rapid changes, while larger L highlights long-term transitions.

Dynamic Topic Quality (DTQ) To evaluate both temporal consistency and per-timestamp coherence of topics, we propose the dynamic topic quality (DTQ). This aggregated measure combines topic quality (TQ, vertical coherence/diversity within each timestamp) and TTQ (horizontal temporal consistency across time):

$$\mathrm{DTQ} = \frac{1}{2}\left(\frac{1}{T}\sum_{t=1}^{T}\mathrm{TQ}_t + \frac{1}{K}\sum_{k=1}^{K}\mathrm{TTQ}_k\right), \tag{23}$$

where K is the number of topics. A high DTQ indicates both coherent topics at each time slice and smooth, semantically.

G Scalability and Robustness to Varying Numbers of Topics

Beyond the main setting with $K = 50$, we evaluate the robustness of L-DNTM on the NYT dataset under varying numbers of topics ($K = 20, 70, 100$). Table 6 compares L-DNTM with two strong baselines, CFDTM and DETM.

Overall, L-DNTM consistently delivers the best performance across almost all metrics and topic settings. As K increases, L-DNTM maintains strong topic quality (TQ), stable dynamic modeling (TTQ, DTQ, TTC), and high topic diversity (TD), demonstrating robustness to increasingly fine-grained topic partitions. In contrast, the baselines degrade notably as K grows. DETM completely collapses at $K = 100$, producing NaN values across all metrics, while CFDTM exhibits unstable temporal modeling, with TTQ and TTC dropping to negative values.

Importantly, L-DNTM achieves the highest TD across all K values (0.927 at $K = 20$, 0.894 at $K = 70$, and 0.859 at $K = 100$), while simultaneously preserving strong TC and downstream clustering performance. These results confirm that L-

DNTM effectively scales with the number of topics and remains reliable even when the optimal K is unknown.

K = 20

Model	TQ	TTQ	DTQ	TTC	TTS	TC	TD	Purity	NMI	Acc	F1
DETM	0.439†	0.091†	0.265†	0.150†	0.571†	0.459†	0.512†	0.515†	0.296†	0.562†	0.279†
CFDTM	0.601†	0.133†	0.367†	0.190†	0.574†	**0.625**†	0.855†	0.572†	0.305†	0.692†	0.487†
L-DNTM	**0.616**†	**0.139**†	**0.377**†	**0.233**†	0.562†	0.568†	**0.927**†	**0.599**†	**0.316**†	**0.771**†	**0.596**†

K = 70

Model	TQ	TTQ	DTQ	TTC	TTS	TC	TD	Purity	NMI	Acc	F1
DETM	0.249†	0.118†	0.183†	0.133†	0.884†	0.350†	0.132†	0.577†	0.325†	0.598†	0.345†
CFDTM	0.542†	0.020†	0.281†	0.024†	0.565†	0.501†	0.678†	0.633†	0.346†	0.580†	0.345†
L-DNTM	**0.590**†	**0.160**†	**0.375**†	**0.234**†	0.660†	**0.583**†	**0.894**†	**0.674**†	**0.361**†	**0.760**†	**0.575**†

K = 100

Model	TQ	TTQ	DTQ	TTC	TTS	TC	TD	Purity	NMI	Acc	F1
DETM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
CFDTM	0.506†	-0.051†	0.238†	-0.058†	0.577†	0.533†	0.575†	0.573†	0.299†	0.510†	0.280†
L-DNTM	**0.538**†	**0.134**†	**0.336**†	**0.182**†	0.691†	**0.548**†	**0.859**†	**0.665**†	**0.353**†	**0.764**†	**0.578**†

Table 6: Robustness analysis on the NYT dataset with varying topic numbers (K). Metrics include Topic Quality (TQ), Temporal Topic Quality (TTQ), Dynamic Topic Quality (DTQ), Temporal Topic Coherence (TTC), Temporal Topic Smoothness (TTS), Topic Coherence (TC), Topic Diversity (TD), Purity, NMI, Accuracy, and F1. The best results (excluding TTS) are in **bold**. At $K = 100$, DETM collapses, yielding NaN scores across all metrics. † indicates statistical significance (p<0.05).

H Full Topic Lists

Here are the discovered topics of different models in 2022 of NYT (the latest time slice).
DETM

```
Topic#1: city people said many home years year country family work
Topic#2: american many dream one phrase like political university people leaders
Topic#3: like just really speaker get right now school know people
Topic#4: said new people group public last one york members chief
Topic#5: said new city people york years many last year one
Topic#6: said city office people new year many years last one
Topic#7: said one new year last national president group week office
Topic#8: said people new york news one public media last social
Topic#9: said new people office many year one now members part
Topic#10: said people home year one work many working new children
Topic#11: said new city people many week last public group york
Topic#12: said new city home people many work year children family
Topic#13: said people new public year last many news york month
Topic#14: said new people year many last one week group years
Topic#15: said one many year people now president new country day
Topic#16: said new people city one year many york work days
Topic#17: said year people new news one city week last day
Topic#18: said new people york news public office made week get
Topic#19: said city people many year new home work country cities
Topic#20: said people city recently many public country last work time
Topic#21: company said musk twitter companies billion chief executive deal percent
Topic#22: said new city people many york year public office home
Topic#23: information media sources anonymous questions know facebook times twitter online
Topic#24: said court law new state federal investigation department report case
Topic#25: workers said abortion tax percent debt financial bank market companies
Topic#26: said people new city now one year many long take
Topic#27: police said one officers capitol killed school two video officer
Topic#28: said new people city york one many home public month
Topic#29: said new city york people mayor one many office two
Topic#30: drug opioid states fentanyl plan overdose company one addiction treatment
Topic#31: said city home day office last one first now many
Topic#32: said news last people days week media members made month
Topic#33: said new people city work many office day public now
Topic#34: said new news public last chief monday week president office
Topic#35: said new city york people many now country work public
Topic#36: trump house former election republican president republicans democrats abortion said
Topic#37: said city new one year york people home years many
Topic#38: said year new now people work many home like last
Topic#39: said people new news public social members one group director
Topic#40: said people new many now city public home country told
Topic#41: russian ukrainian ukraine said war soldiers russia forces kyiv times
Topic#42: first one like new time just team also two best
Topic#43: said new people city many old country years public office
Topic#44: said new people city york year many one last week
Topic#45: russia russian ukraine gas war said putin president energy united
Topic#46: inflation percent prices economy year said pandemic rates price covid
Topic#47: said new people many one get even work now made
Topic#48: said city new people work year home time many one
Topic#49: said new one people city last news many year years
Topic#50: said new city people last york week month days told
```

CFDTM

Topic#1: murder father mother wine mothers husband mexico toronto drug blaze

Topic#2: union opponent coup rio democracy brazil criticism cabinet fury challenges

Topic#3: inflation rates prices fed economy percent interest investors market year

Topic#4: like just really michael know get right now going people

Topic#5: scene movies film queue viewers episodes listening directors sequence motives

Topic#6: democrats legislation senate bill republicans vote votes measure senator house

Topic#7: team season players ball game play win toe nba coach

Topic#8: stadium space soccer thailand stampede ballistic condemns launch mourn philippines

Topic#9: democratic voters candidate presidential candidates poll race sanders campaign win

Topic#10: queen royal elizabeth prince palace andrew walker london throne monarch

Topic#11: putin sanctions moscow vladimir russian kremlin soviet russia western opposition

Topic#12: league sports art museum team game games sport formula ancient

Topic#13: border migrants immigration asylum homeland immigrants humanitarian order undocumented thousands

Topic#14: book shame really mother writes writing character hard horror essays

Topic#15: city york new mayor trash correction station housing adams manhattan

Topic#16: taliban afghan afghanistan kabul forces american airstrikes air afghans insurgents

Topic#17: gas prices oil russia companies price energy sanctions russian economy

Topic#18: prime minister johnson boris downing parties london parliament british gatherings

Topic#19: trial prosecutors charges case manhattan judge lawyer guilty weinstein accused

Topic#20: getting worried doctor got felt oxygen doctors keep tennis theory

Topic#21: russian forces ukraine soldiers war said ukrainian military russia york

Topic#22: ukrainian israeli kyiv plant brittney palestinian zaporizhzhia mourners pilots attacks

Topic#23: hate black asian mass white church francis catholic violence descent

Topic#24: tax irs treasury financial revenue banks yellen income overhaul taxes

Topic#25: parliament israeli coalition minister prime israel netanyahu shiite government jerusalem

Topic#26: twitter musk company tesla stock board elon buy investors shares

Topic#27: best opera won series comedy show actress actor awards night

Topic#28: book novel painting books mccarthy prize published science art fiction

Topic#29: airlines merger tiktok operations chinese customers cigarettes airline technology food

Topic#30: health cases vaccine disease response virus covid infections monkeypox coronavirus

Topic#31: vaccine doses study patients disease researchers treatment drug dose symptoms

Topic#32: police officers capitol said now two killed chief school rioters

Topic#33: sales quarter profit apple tesla cars revenue vehicles tech electric

Topic#34: court law abortion texas supreme transgender ruling abortions rights abuse

Topic#35: school students schools university high sexual abuse teachers campus classes

Topic#36: rights human arrested detained criminal egypt saudi released charges arrest

Topic#37: variant cases infections tested restrictions isolation quarantine coronavirus indoor omicron

Topic#38: gas energy natural germany europe supplies nord pipeline stream pipelines

Topic#39: google information tech app facebook gun content internet users privacy

Topic#40: workers income schools teachers labor job work gig students jobs

Topic#41: storm hurricane residents water florida winds storms homes fort myers

Topic#42: china chinese taiwan beijing jinping hong kong communist mao island

Topic#43: nato putin ukraine finland vladimir war sweden turkey russia alliance

Topic#44: gap executive discovery chief warner media bros partnership space disney

Topic#45: prosecutors officer capitol officers guilty charges trial police floyd jan

Topic#46: growth economy prices oil recession global economic dollar jobs price

Topic#47: republican republicans democrats abortion voters said senate democratic vote senator

Topic#48: protests iranian officers protesters arrested hong students tehran kong crackdown

Topic#49: documents trump justice former investigation fbi committee house classified search

Topic#50: republican election republicans fauci senator wisconsin fraud victory donald primary

L-DNTM

Topic#1: twitter elon musk bid offer stake tweet filing merger board

Topic#2: stone roger bannon journalists campaign reporters fox cnn politics wikileaks

Topic#3: mourners bomber bombings explosion attackers blast khan pleads carnage wounding

Topic#4: successor chief executive retire leadership stepping hired appointment ceo retirement

Topic#5: harvey podcast consent misconduct rape allegations complaint workplace catholic sexual

Topic#6: omicron surge contagious hospitalizations variant infected restrictions outbreak hospitalized wave

Topic#7: crackdown detained democracy increasingly beijing jinping taiwan reform protesters taiwanese

Topic#8: equity bitcoin assets fund financial funds hedge investment finance investments

Topic#9: instagram privacy platforms facebook app developers apple users google apps

Topic#10: puerto florida rescue flooding water damage rubble collapse destruction emergency

Topic#11: election senate vote voters senator democratic candidates representative voting support

Topic#12: symptoms fentanyl overdose opioid addiction medical treatment risk deaths overdoses

Topic#13: energy renewable gas climate emissions change wind environmental carbon electricity

Topic#14: abuses torture coup sisi egypt saudi ouster navalny graft arrests

Topic#15: infrastructure irs legislation overhaul funding bipartisan pass act subsidies medicaid

Topic#16: shame memoir horror essays fiction dark essay book characters novel

Topic#17: hate nazi anti migrants nationalist muslim immigration extremist refugees immigrants

Topic#18: jan capitol archives classified riot invoke panel privilege committee administration

Topic#19: speech question trump biden presidency meeting words white donald recording

Topic#20: capitol rioters gunman mob crowd attack video riot inside footage

Topic#21: warhead explosion strike explosives explosions killed killing drone attack strikes

Topic#22: death lethal injection murder executions sentence execution guilty trial convicted

Topic#23: frontier spirit electric tesla battery cars manufacturing traffic recall vehicles

Topic#24: overrun taliban afghanistan afghan fighting military forces troops army soldiers

Topic#25: launch launched missile ballistic missiles diplomatic intercontinental dialogue nuclear arms

Topic#26: rates interest fed rising rate inflation economy recession higher prices

Topic#27: rabbi clashes palestinians gaza jerusalem israel israeli palestinian hamas annex

Topic#28: racism systemic statue racist confederate injustice symbols remove slavery history

Topic#29: investigating investigators special prosecutors investigation criminal lawyers lawyer charges attorney

Topic#30: fraud georgia votes ballots results ballot mail runoff voting electoral

Topic#31: warnock ocasio cortez walker donors debate campaigns race campaign candidates

Topic#32: throne monarch london queen royal british prince meghan britain harry

Topic#33: attempted japanese combination flip landed anna olympics olympic skating chen

Topic#34: youtube documentary swift episodes scene sequence comedy bros detective writer

Topic#35: mandate mandates mask staffing classrooms classroom learning teachers school schools

Topic#36: nba nfl finals championship basketball soccer football baseball team game

Topic#37: monkeypox variant response covid coronavirus virus omicron vaccine vaccines pandemic

Topic#38: adams kathy hochul eric cuomo mayor city nyc governor andrew

Topic#39: benjamin netanyahu labour modi parliament party opposition leader coalition parties

Topic#40: roe wade abortion ruling ban rights amendment constitutional unconstitutional arguments

Topic#41: artillery zaporizhzhia soldiers donetsk invasion war sanctions russian ukraine russia

Topic#42: rhetoric phrase change powerful events aggressive challenges climate social attention

Topic#43: cybersecurity cyberattacks hackers security attacks intelligence spying information target sources

Topic#44: said one people now just like time also two new

Topic#45: agreement nuclear trade negotiations accord deal talks enrichment oil uranium

Topic#46: gap families factors hispanic age poor women education child parents

Topic#47: moratorium homelessness eviction homeless housing rent tenants landlords apartments rents

Topic#48: streaming broadway opera met performances tony musical awards show theater

Topic#49: chancellor macron emmanuel european germany europe france union britain euros

Topic#50: married dies life famous celebrity prize sarah career jennifer patricia