A Multiscale Geometric Method for Capturing Relational Topic Alignment

Conrad D. Hougen
University of Michigan
Ann Arbor, MI, USA
chougen@umich.edu

Karl T. Pazdernik
Pacific Northwest National Laboratory
Richland, WA, USA
karl.pazdernik@pnnl.gov

Alfred O. Hero
University of Michigan
Ann Arbor, MI, USA
hero@eecs.umich.edu

***Abstract*—Interpretable topic modeling is essential for tracking how research interests evolve within co-author communities. In scientific corpora, where novelty is prized, identifying underrepresented niche topics is particularly important. However, contemporary models built from dense transformer embeddings tend to miss rare topics and therefore also fail to capture smooth temporal alignment. We propose a geometric method that integrates multimodal text and co-author network data, using Hellinger distances and Ward's linkage to construct a hierarchical topic dendrogram. This approach captures both local and global structure, supporting multiscale learning across semantic and temporal dimensions. Our method effectively identifies rare-topic structure and visualizes smooth topic drift over time. Experiments highlight the strength of interpretable bag-of-words models when paired with principled geometric alignment.**

***Index Terms*—relational topic models, topic alignment, LDA, multiscale learning, transformer embeddings**

I. Introduction

Topic modeling enables automated discovery of hidden semantic structure within documents. Traditional topic models operate on bag-of-words representations, using linear algebraic or probabilistic methods such as in latent semantic analysis (LSA), non-negative matrix factorization (NMF), probabilistic LSA (pLSA), and latent Dirichlet allocation (LDA) [1]–[4]. These models use raw word counts or term-frequency inverse document frequency (TF-IDF) inputs [5]. Word2Vec [6] motivated the usage of neural networks to translate words, phrases, and sentences into dense vector embeddings. A similar approach is now used by large language models (LLMs) including BERT [7], RoBERTa [8], and GPT [9], with transformer architectures. Derivative topic models, like the contextualized topic model (CTM) [10], BERTopic [11], GPTopic [12], and TopicGPT [13], all use transformer embeddings.

While the above methods are useful, these models operate on text alone, overlooking relational information within document networks. Relational topic models (RTMs) help bridge that gap [14]–[18]. Citation networks, which are relatively dense and stable over time, are often used to relate documents [19]. In contrast, co-author network links tend to be highly variable, with co-author networks being globally sparse but locally clique-structured. This complicates the integration of co-authorship information into RTMs. Prior literature validates RTMs using link prediction benchmarks, but this is a difficult task in co-author networks.

Incorporating the temporal dimension into dynamic document networks introduces additional complexity. This is particularly true when the goal is to simultaneously align topics across time and across document networks. Few existing approaches tackle this joint challenge. Also, existing methods do not capture the time-evolving nature of author interests [20]. Author-based topic drift varies with the granularity at which topics are defined, motivating the need for a multiscale learning approach.

To address these challenges, we introduce Multiscale Topic Manifold Learning (MSTML), an ensemble model that combines temporal alignment [21] with hierarchical learning [22] to capture multiscale structure in topics and co-author networks. MSTML integrates text and network data through an interpretable probabilistic model, aligning topics using Hellinger distances [23] and Ward's linkage [24], and visualizing them with Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) embeddings [25]. Unlike prior methods [26], [27], MSTML relies only on texts and author lists, without external metadata. Applied to arXiv data, it reveals topic trends and co-authorship dynamics. Sections II–V present the concepts and results. Additional method details and experiments can be found in [28].

II. Background

A. Topic Modeling

Topic modeling aims to uncover thematic structures in document corpora, but the objectives of topic models vary. Model evaluation metrics assess the quality and distinctiveness of discovered topics. Topic coherence metrics, such as C_{V}, C_{UCI}, and C_{NPMI} [29], [30], are among the most popular. Topic coherence quantifies semantic similarity among the top words within each topic, based on co-occurrence statistics and external resources. In this work we promote assessments of topic alignment as well, which refers to topic consistency and smooth topic evolution across time. This is crucial for analyzing longitudinal corpora.

In addition to choosing proper metrics, topic models must contend with the inherent complexity of high-dimensional and noisy datasets. Large vocabulary sizes are computationally challenging [31]. To mitigate this, connective stop words like

The work in this paper was partially supported by grants from DOE NNSA Grant DE-NA0003921 and NSF grant CCF-2246213. PNNL-SA-213388.

conjunctions and articles are usually removed prior to model training. Variations of the same word can also be joined into a common root form, which is called stemming. Lemmatization further combines semantically-similar terms, like ”better” and ”good.” These operations reduce the vocabulary size.

Techniques for filtering the vocabulary also often utilize term-frequency thresholds. Prior research has revealed a balance between noise and topic granularity [32], [33]. Terms that appear only once in the corpus can be safely removed without effect, but terms that appear more than once, and high-frequency terms, add to document discriminative capacity [32].

LDA is a generative probabilistic topic model and one of the most commonly-cited techniques [4]. In LDA, each word in a document is an outcome of a generative process which randomly selects a topic z, then a word w. Topic z is first sampled according to multinomial distribution $\boldsymbol{\theta}^{(\boldsymbol{j})}$. Second, w is sampled from the vocabulary, according to the multinomial distribution $\boldsymbol{\phi}^{(\boldsymbol{z})}$. This repeats for all $N_W^{(j)}$ words in document j, and for all documents $j \in \{1, \ldots, N_D\}$. LDA uses Bayesian inference to learn the vectors of interest, $\{\boldsymbol{\theta}^{(\boldsymbol{j})}\}_{j=1}^{N_D}$ and $\{\boldsymbol{\phi}^{(\boldsymbol{k})}\}_{k=1}^{K}$.

LDAvis [34] is a visualization tool based on LDA that uses a linear embedding and the concept of term relevancy, defined in (1). $P(w \mid k)$ is the probablity of term w given the topic k, and $P(w)$ is the marginal probability of the term w across the entire corpus. $\lambda \in (0, 1)$ is a weight hyperparameter. Term relevancy balances frequency within a topic against exclusivity to that topic, providing a basis for vocabulary filtering by ranking terms using relevancy scores from an auxiliary model.

$$r(w, k \mid \lambda) = \lambda \log P(w \mid k) + (1 - \lambda) \log \left(\frac{P(w \mid k)}{P(w)} \right) \quad (1)$$

B. Manifold Learning and Information Geometry

Manifold learning describes a subset of nonlinear methods for dimension reduction of high-dimensional data. Unlike linear methods like principal component analysis (PCA) [35], manifolds are assumed to be globally non-linear but locally linear. Well known algorithms include locally linear embedding (LLE), t-distributed stochastic neighbor embedding (t-SNE) and multi-dimensional scaling (MDS) [36]–[38]. In this work, we take advantage of potential of heat-diffusion for affinity-based trajectory embedding (PHATE) [25], which is advantageous for time-evolving data due to a density-adaptive diffusion process which preserves local and global distances.

Preserving distances across multiple scales is also central to information geometry, which applies differential geometric tools to probability distributions. Unlike standard Euclidean vectors, the L_2 distance is ineffective for comparing multinomial document-topic and topic-word distributions, such as $\boldsymbol{\theta}$ or $\boldsymbol{\phi}$ from LDA [21], [39], [40]. Instead, information geometry exploits the Fisher information metric, which defines a Riemannian structure on statistical manifolds like the probability simplex. In practice, the Hellinger metric may be used as a computationally-simple approximation of the Fisher information. For multinomial vectors, $\boldsymbol{p}$ and $\boldsymbol{q}$, the Hellinger metric is defined according to (2). The Hellinger distance exhibits useful properties of being bounded between 0 and 1 and symmetric between $\boldsymbol{p}$ and $\boldsymbol{q}$.

$$H(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} \left(\sqrt{\boldsymbol{p}_i} - \sqrt{\boldsymbol{q}_i}\right)^2} \quad (2)$$

C. Co-Author Networks

A co-author network is defined by observable author lists associated with documents. When multiple authors share approximately the same name, author disambiguation is required, using fuzzy matching or additional meta-data [26]. Co-author networks are then constructed as described in Definition 1. The network evolves over time, with each time window producing a snapshot of collaboration structure.

Definition 1. *A* ***co-author network*** *is a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ is the set of authors and $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ contains an edge (u, v) if authors u and v co-authored at least one document in the* ***active*** *document set. The graph is typically undirected, so $(u, v) \in \mathscr{E}$ implies $(v, u) \in \mathscr{E}$. The degree $\deg(v)$ of a vertex is the number of co-authors of author v.*

III. Methods

A. General Algorithm and Framework Diagram

MSTML integrates topic manifold learning [21] with hierarchical network models [22]. The core method fits a dendrogram, parameterized by internal node probabilities, $\{\mathscr{D}; \{p_m\}\}$, to a co-author network $\mathscr{G}$. The authors (nodes) in $\mathscr{G}$ are represented by embeddings derived from an LDA topic model ensemble. LDA-derived topic vectors $\{\boldsymbol{\phi}^{(\boldsymbol{k})}\}$ are mapped to dendrogram leaf nodes. These topic vectors reside in the probability simplex, $\Delta^{\nu-1} \subset \mathbb{R}^{\nu}$, where ν is the vocabulary size. The MCMC process of the HRG model, which learns the dendrogram topology, is replaced by agglomeratve clustering, using the Hellinger distance (2). Figure 1 illustrates an example topic dendrogram model.

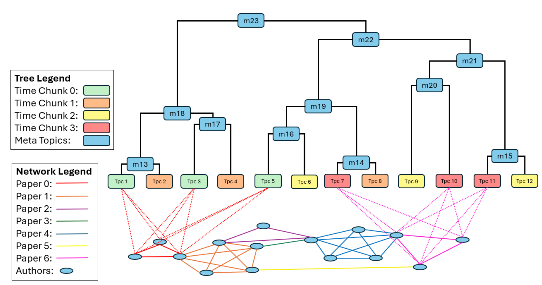


Fig. 1: The topic space dendrogram (hierarchical tree, top) links chunk topics (multi-colored rectangles), meta-topics (m13-m23), and the co-author network (network, below). Several example links (red, pink) between the dendrogram and co-author network are included. These links represent author distributions over multiple chunk topic leaf nodes.

Instead of one LDA model, MSTML employs an ensemble learning approach. The corpus is split into uniform time chunks, with LDA applied to each sub-corpus $C_1, C_2, \ldots, C_T$

independently. Temporal smoothing is applied to improve continuity in the topic geometry [21], [28]. Each LDA model uses uniform Dirichlet priors, $\boldsymbol{\alpha} = \mathbf{1}, \boldsymbol{\beta} = \mathbf{1}$. A key challenge in LDA is selecting the number of topics, K, which is based on: avoiding topic duplication, high isolation between topics, and repeatability [41]. We have chosen to scale K as an affine function of the number of documents per time chunk to balance topic continuity against discovery of emerging topics [28]. Chunk topics are finally clustered into meta-topics using agglomerative clustering with Ward's linkage [24]. The resultant, smoothed manifold allows for interpretable traversal and alignment with co-author network trends.

The ensemble approach mitigates overfitting, improving the odds of capturing niche topics that could be diluted by a single, global topic model. However, a global LDA model is uniquely used in MSTML for vocabulary filtering. We found that using global term relevancy (1) to rank and filter terms improves the interpretable quality of topic word clouds, with λ set around 0.4, adaptive based on the data [28]. Additionally, the learned topic manifold appears to be represented more smoothly, with better topic alignment and clustering of similar topics.

B. Dendrogram Learning

The MSTML topic dendrogram provides a multi-scale representation of topic relationships and also links together the observed text and co-author network data. $\{\mathscr{D}; \{p_m\}\}$ is constructed in two stages. In the first, tree topology is determined using agglomerative hierarchical clustering of topic vectors $\{\boldsymbol{\phi}^{(\boldsymbol{k})}\}$. Second, probabilities $\{p_m\}$ are assigned to the internal nodes.

For the topology of $\mathscr{D}$, LDA-derived topic vectors $\{\boldsymbol{\phi}^{(\boldsymbol{k})}\}$ are treated as multinomial distributions over the vocabulary $\mathcal{V}$. A k-nearest neighbors graph, $\mathcal{X}$, is constructed to capture meso-scale structures among topics to guide agglomerative clustering in alignment with diffusion-based PHATE embeddings. For large topic sets, FAISS is used to approximate the k-nearest neighbors graph [42]. Clustering proceeds by iteratively merging topic clusters, where each merge produces an internal node m with a height h_m, representing inter-cluster dissimilarity. These heights h_m increase up the tree and may exceed the Hellinger distance bounds of $[0, 1]$ when using Ward's linkage. For consistent interpretations with various linkages and distances, heights are re-normalized to the $[0, 1]$ interval.

To compute the internal node probabilities $\{p_m\}$, MSTML treats $\mathscr{E}_m$, L_m, and R_m as random variables based on author-topic distributions $\{\boldsymbol{\psi}^{(\boldsymbol{u})}\}$, which are derived as weighted averages over their associated document-topic vectors. Each document-topic vector is weighted inversely by the number of authors on the document. The author-topic distributions map authors to chunk topics, which then map directly to dendrogram leaf nodes. Following [22], the MLE estimator for internal node probabilities is defined as $\hat{p}_m = \frac{\mathscr{E}_m}{L_m R_m}$, approximated here as $\hat{p}_m \approx \frac{\mathbb{E}[\mathscr{E}_m]}{\mathbb{E}[L_m]\mathbb{E}[R_m]}$ under the assumption of weak correlation between $\mathscr{E}_m$, L_m, and R_m.

Critically, this approach avoids the computational complexity of MCMC sampling by leveraging topic manifold clustering to learn the dendrogram topology only once. The dendrogram topology captures temporal topic trends and multi-scale author-topic relationships. The topology itself need only be "close" to optimal in order to estimate the internal node probabilities [22]. Probabilities $\{p_m\}$ are computed as the ratio of the number of expected edges to possible edges, conditioned on the LDA-learned topic distributions (3). $\hat{L}_m$, $\hat{R}_m$, and $\hat{\mathscr{E}}_m$ are computed as the expected numbers of authors in the left and right subtrees, and the number of edges between these subtrees, respectively [28].

$$p_m \triangleq \frac{\hat{\mathscr{E}}_m}{\hat{L}_m \hat{R}_m}, \forall m \in \mathscr{D} \tag{3}$$

IV. Results

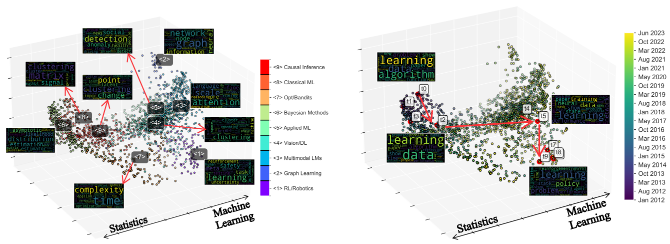
This section shows experimental results using an arXiv corpus, collated by Cornell University and hosted on Kaggle (https://www.kaggle.com/datasets/Cornell-University/arxiv). The arxiv-stat-ml corpus is an extraction of 194,035 self-labeled document abstracts from statistics (stat) and machine learning (cs.LG) categories. These abstracts are supported on a vocabulary of 136,113 terms which is eventually filtered down to 8,465 terms. Code was written in Python and Cython.

A. Visualizations and Multimodal Analysis

MSTML visualizes the topic manifold, discovering smooth topic alignment across time. The topic dendrogram can be cut at various heights. Using default MSTML parameters, a cut height of $h = 0.55$ results in 9 meta topic clusters. These meta topic clusters can be identified by word clouds or by ranking top contributing documents. The 9 meta topics were mapped back to the top 30 contributing documents, using inferred probability mass. Document titles were then passed to ChatGPT in order to produce automated topic labeling. This prompt produced 9 labels that largely matched human labels: {1: RL/Robotics, 2: Graph Learning, 3: Multimodal LMs, 4: Vision/DL, 5: Applied ML, 6: Bayesian Methods, 7: Opt/Bandits, 8: Classical ML, 9: Causal Inference}.

The topic space forms interpretable geometric regions that can be smoothly traversed (Figure 2a) and easily annotated with word clouds. This is advantageous compared to dense vector embeddings which often lack such intuitive structure. By mapping meta topic clusters back to the co-author network, MSTML can also visualize community topic drift by choosing topic points with high cumulative probability mass (Figure 2b).

Figure 3 shows temporal snapshots of both the co-author network and topic space for the RL/Robotics community. Node sizes and edges in the network reflect the active document set, while node colors indicate each author's dominant topic during a given snapshot. In the corresponding PHATE diagrams, points are colored and sized according to the topic probability mass across the RL/Robotics community. Star-shaped markers highlight the topic distribution learned for Dorsa Sadigh, a selected representative author from the RL/Robotics subnetwork. These visualizations illustrate how MSTML's smooth manifold captures temporal alignment and topic drift in an intuitive way. In Sadigh's case, the trajectory shows a shift from earlier research interests into a central role within the RL/Robotics space.



(a) 9 meta topic clusters (b) RL/Robotics community drift

Fig. 2: (a) PHATE visualization of the topic manifold, using a particular dendrogram cut height that reveals 9 meta topics. (b) The RL/Robotics community originally published in topics related to statistics before drifting toward RL topics.

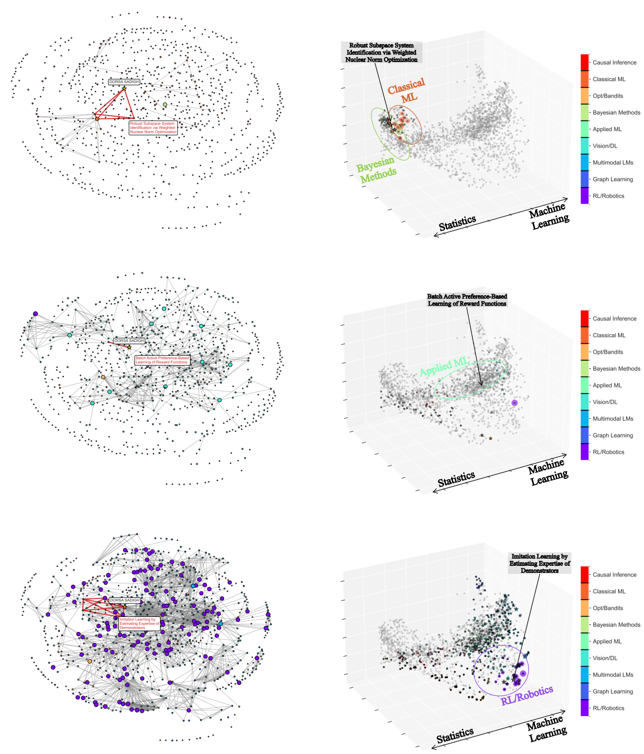


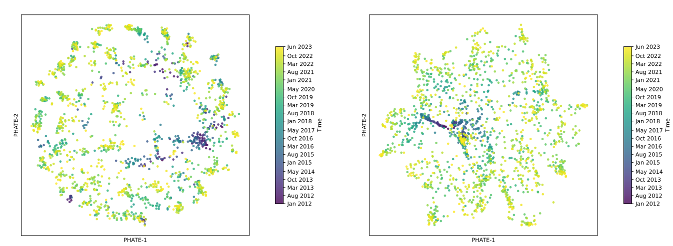
Fig. 3: The RL/Robotics community and sub-network is highlighted across time snapshots (t1, t5, t9 shown). In the left panels, red links indicate articles authored by Sadigh, with authors (nodes) colored by dominant topic contributions. In the right panels, community-wide topic distributions are represented as colored points. Star-shaped points represent Sadigh, specifically. Point sizes indicate probability mass.

B. Comparisons with Transformer Models

Table I compares BERTopic against CTM [10], Embed-Cluster (BERT with k-means clustering), and LDA ensemble models. The LDA ensembles are the backbone of MSTML, but they skip term relevancy filtering and meso-scale k-NN graph construction to avoid test bias. BERTopic achieves the highest scores on standard topic coherence metrics (C_{V}, C_{UCI}, C_{NPMI}), consistent with its use of semantically clustered sentence-transformer embeddings. However, this advantage comes at the cost of interpretability and niche topic identification. When considering topic alignment, the LDA ensembles reveal a continuous and temporally coherent geometry, in contrast to the fragmented clusters from the BERTopic ensemble (Figure 4). This smoothness enables MSTML to trace topic drift and identify emerging themes. These capabilities are obscured in models optimized for semantic similarity alone.

TABLE I: Topic Coherence Comparison

Model Name	C_{V}	C_{UCI}	C_{NPMI}
BERTopic	**0.6081**	**0.5236**	**0.1198**
CTM	0.5922	0.4673	0.0823
EmbedCluster	0.4680	0.2059	0.0316
LDA Ensemble $\lambda = 0.00$	0.4054	-5.5958	-0.1959
LDA Ensemble $\lambda = 0.25$	0.3810	-4.4006	-0.1282
LDA Ensemble $\lambda = 0.50$	0.5291	-0.8150	0.0325
LDA Ensemble $\lambda = 0.75$	0.5391	0.0381	0.0559
LDA Ensemble $\lambda = 1.00$	0.5033	0.1565	0.0471



(a) BERTopic (Cosine) (b) LDA ensemble (Hellinger)

Fig. 4: PHATE embeddings for time alignment comparison. BERTopic (left) shows tight, discrete clusters, while LDA ensemble (right) is more smooth, with clear temporal clustering.

V. Conclusion

We developed MSTML for temporal and multiscale relational topic modeling, in longitudinal corpora. MSTML combines probabilistic models with information geometry to build aligned and interpretable topic manifolds. Smooth topic alignment captures temporal topic drift within author and document networks. Compared to transformer-based models, which prioritize topic coherence, MSTML better preserves niche topics critical for understanding scientific novelty. Future work should explore additional ensemble-based topic models and also develop quantitative metrics for niche topic representation. Topic diversity offers a partial solution but does not fully capture the distinctiveness of niche content [28]. GPT-based models like GPTopic [43] have also been compared to MSTML but are computationally expensive [28] due to reliance on costly API calls using large models. Regardless, future work should address comparisons with alternative models based on GPT embeddings. Finally, we would also like to thoroughly analyze the effects of term relevancy filtering for ensemble model representations.

REFERENCES

[1] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, vol. 41, no. 6, pp. 391–407, 1990.

[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[3] T. Hofmann, "Probabilistic latent semantic analysis." in *UAI*, vol. 99, 1999, pp. 289–296.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[5] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of ICLR*, 2013.

[7] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," vol. 1, no. 2, 2019.

[8] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[9] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI Blog*, 2018.

[10] F. Bianchi, S. Terragni, and D. Hovy, "Pre-training is a hot topic: Contextualized document embeddings improve topic coherence," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Association for Computational Linguistics, 2021, pp. 759–766.

[11] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[12] F. Bianchi, S. Terragni, A. Hogenboom, M. Palmonari, and D. Nozza, "GPTopic: Generative pretrained transformers for topic modeling," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023.

[13] M. Grootendorst, "Topicgpt: Zero-shot topic modeling with gpt," https://github.com/MaartenGr/TopicGPT, 2023.

[14] C. Wang, H. Zhang, B. Chen, D. Wang, Z. Wang, and M. Zhou, "Deep relational topic modeling via graph poisson gamma belief network," in *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[15] Z. Wang, C. Wang, H. Zhang, Z. Duan, M. Zhou, and B. Chen, "Learning dynamic hierarchical topic graph with graph convolutional network for document classification," in *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 108. PMLR, 2020, pp. 1397–1407.

[16] Y. Wang, X. Li, and J. Ouyang, "Layer-assisted neural topic modeling over document networks," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conferences on Artificial Intelligence Organization, 2021, pp. 3148–3154.

[17] C. Wang, X. Liu, D. Wang, H. Zhang, B. Chen, and M. Zhou, "Scalable weibull graph attention autoencoder for modeling document networks," Oct. 2024, unpublished.

[18] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

[19] C. L. Giles, K. D. Bollacker, and S. Lawrence, "Citeseer: An automatic citation indexing system," in *Proceedings of the Third ACM Conference on Digital Libraries*. New York, NY, USA: ACM, 1998, pp. 89–98.

[20] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, ser. UAI '04, 2004, p. 487–494.

[21] Y. Wang, C. Hougen, B. Oselio, W. Dempsey, and A. Hero, "A geometry-driven longitudinal topic model," *Harvard Data Science Review*, vol. 3, 2021.

[22] A. Clauset, C. Moore, and M. E. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, no. 7191, pp. 98–101, 2008.

[23] E. Hellinger, *Neue Begründung der Theorie der quadratischen Formen von unendlich vielen Veränderlichen*, 1909, vol. 136.

[24] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[25] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, S. Burkhardt, W. S. Chen, X. Yim, A. van den Elzen, K. Hirn, R. Wolf, D. Krishnaswamy, G. Wolf, and S. Krishnaswamy, "Visualizing structure and transitions in high-dimensional biological data," *Nature Biotechnology*, vol. 37, pp. 1482–1492, 2019.

[26] S.-Y. Hwang, C.-P. Wei, C.-H. Lee, and Y.-S. Chen, "Coauthorship network-based literature recommendation with topic model," *Online Information Review*, vol. 41, no. 3, pp. 318–336, 2017.

[27] Y. Zhang, M. Wu, G. Zhang, and J. Lu, "Stepping beyond your comfort zone: Diffusion-based network analytics for knowledge trajectory recommendation," *Journal of the Association for Information Science and Technology*, vol. 74, no. 7, pp. 775–790, 2023.

[28] C. D. Hougen, K. T. Pazdernik, and A. O. Hero, "Multi-scale topic manifold learning for understanding interdisciplinary collaborations in co-author networks," unpublished.

[29] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the eighth ACM international conference on Web search and data mining*. ACM, 2015, pp. 399–408.

[30] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," in *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. Association for Computational Linguistics, 2010, pp. 100–108.

[31] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[32] K. Lu, X. Cai, I. Ajiferuke, and D. Wolfram, "Vocabulary size and its effect on topic representation," *Information Processing & Management*, vol. 53, no. 3, pp. 653–665, 2017.

[33] D. Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler *et al.*, "Applying lda topic modeling in communication research: Toward a valid and reliable methodology," pp. 13–38, 2021.

[34] C. Sievert and K. Shirley, "LDAvis: A method for visualizing and interpreting topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*. Baltimore, Maryland, USA: Association for Computational Linguistics, 2014, pp. 63–70.

[35] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 2, no. 11, pp. 559–572, 1901.

[36] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[37] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[38] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, pp. 1–27, 1964.

[39] F. Nielsen, "An elementary introduction to information geometry," *Entropy*, vol. 22, no. 10, p. 1100, 2020.

[40] K. Sun and S. Marchand-Maillet, "An information geometry of statistical manifold learning," pp. 1–9, 2014.

[41] J. Gan and Y. Qi, "Selection of the optimal number of topics for lda topic model—taking patent policy analysis as an example," *Entropy*, vol. 23, no. 10, p. 1301, 2021.

[42] J. Johnson, M. Douze, and H. Jegou, "Billion-scale similarity search with gpus," in *IEEE Transactions on Big Data*, 2019, pp. 535–547.

[43] A. Reuter *et al.*, "Gptopic: Dynamic and interactive topic representations," https://arxiv.org/abs/2403.03628, 2024, arXiv preprint, March 2024.