# Appendix for "Topic Modeling Revisited: A Document Graph-based Neural Network Perspective"

## A Appendix

Here, we display additional material to support our content in the main manuscript, including:

- The mathematical notations in Table S1.
- The proof of Theorem 1 in Section A.1.
- The derivation of the objective function in Section A.2.
- The proof of Corollary 1 in Section A.3.
- The dataset statistics in Section A.4.
- Detailed experimental setup with the parameter analysis of the window size $s$ in Section A.5.
- The oblation study of representing topics with a distribution over word edges in Section A.6.
- More experimental results in Section A.7.
- More examples for learned topics in Section A.8.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

Table S1: Mathematical Notations

| Symbols | Descriptions |
|---|---|
| $w_{d,n}$ | The $n$-th word in the document $d$. |
| $z_{d,n}$ | The topic assignment for the word $w_{d,n}$. |
| $\theta_d$ | The topic proportion of the document $d$. |
| $N_d$ | The number of words in document $d$. |
| $V_d$ | The word set for the document $d$, i.e., $V_d = \{w_{d,n}\}_{n=1}^{N_d}$. |
| $E_d$ | The word edge set for the document $d$. |
| $Z_d$ | The topic assignment set for the document $d$ |
| $G_d$ | The graph for the document $d$, i.e., $G_d = (V_d, E_d)$. |
| $V_d^o$ | The placeholder set for the document $d$, i.e., $V_d^o = \{1, 2, ..., N_d\}$. |
| $E_d^o$ | The placeholder edge set for the document $d$. |
| $G_d^o$ | The graph structure for the document $d$, i.e., $G_d^o = (V_d^o, E_d^o)$. |
| $E$ | The word edge set on the whole document collection. |
| $V$ | The vocabulary on the whole document collection. |
| $M$ | The topic dependency matrix. |
| $m_{i,j}$ | The element of $M$ at $i$-th row and $j$-th column. |
| $\beta_k^v$ | The word distribution on $V$ for the $k$-th topic. |
| $\beta_k^e$ | The parameter to describe the distribution over word edges for $k$-th topic. |
| $\beta_k$ | The $k$-th topic, i.e., $\beta_k = (\beta_k^v, \beta_k^e)$. |
| $G_{d,k}$ | The graph among the nodes with the topic assignment $k$ in the document $d$. |
| $G_{d,k}^o$ | The graph structure among the nodes with the topic assignment $k$ in the document $d$. |
| $E_{d,k}$ | The word edge set among the nodes with the topic assignment $k$ in the document $d$. |
| $E_{d,k}^o$ | The placeholder edge set among the nodes with the topic assignment $k$ in the document $d$. |
| $\alpha$ | The parameter of the Dirichlet prior for $\theta_d$. |
| $\varphi_{d,n}$ | The paramter of the varaitionl multinomial distribution for $z_{d,n}$. |
| $\mu_d, \delta_d$ | The parameters of the varaitionl logistic normal distribution for $\theta_d$. |
| $r$ | The word vectors for computing topic set $\beta$. |
| $\hat{r}_{w_1,w_2}$ | The concatenation of word vectors, i.e., $\hat{r}_{w_1,w_2} = r_{w_1} \oplus r_{w_2}$. |
| $u^v$ | The topic-word vectors for computing topic set $\beta$. |
| $u^e$ | The topic-edge vectors for computing topic set $\beta$. |
| $W$ | The paramter matrix for computing topic dependency matrix $M$. |
| $a_k, b_k$ | The transitional vectors for the topic k to compute the possibility of directed edges among different topics. |
| $h_d$ | The node-level representation vector set for all word node on the document graph $G_d$. |
| $h_d^G$ | The graph-level representation vector for the whole document graph $G_d$. |
| $K$ | The number of topics. |
| $s$ | The window size for constructing the document graph. |
| $\mathrm{Dir}(\cdot)$ | The Dirichlet distribution. |
| $\mathcal{N}(\cdot)$ | The Gaussian distribution. |
| $\mathrm{Multi}(\cdot)$ | The multinomial distribution. |
| $f_*(\cdot)$ | The full connected layer. |
| $GNN(\cdot,\cdot)$ | The graph neural network. |
| $\sigma(\cdot)$ | The sigmod active function. |
| $\tanh(\cdot)$ | The $\tanh$ active function. |
| $\oplus$ | The vector concatenation operation. |
| $\cdot$ | The dot product operation. |
| $\odot$ | The Hadamard product operation. |

### A.1 Proof of Theorem 1

*Proof.* To prove Theorem 1, we introduce the following lemma:

**Lemma 1.** *Given the topic set $\beta$ defined in Equation 5 and the document graph structure $G_{d,k}^o$ under topic assignment $k$, the probability function $p(V_{d,k}|G_{d,k}^o;\beta_k)$ defined in Equation 6 is a legal probability measure on the vocabulary $V$.*

*Proof.* This proof is mainly inspired by [15]. Obsviouly, based on the definition in Equation 6, $p(V_{d,k}|G^o_{d,k};\beta_k) > 0$. Then we need to prove the summing of $p(V_{d,k}|G^o_{d,k};\beta_k)$ on all possible $V_{d,k}$ is equal to 1 for all possible graph structure $G^o_{d,k}$. Here, to provide a completed proof of the lemma, we cluster all situations into three categories based on the edge set $E^o_{d,k}$ of the graph structure, where 1) $E^o_{d,k} = \emptyset$, 2) $|E^o_{d,k}| = 1$, and 3) $|E^o_{d,k}| > 1$, respectively. In the following, we demonstrate that $p(V_{d,k}|G^o_{d,k};\beta_k)$ is a legal probability measure in each category.

**Category 1:** As the definition in equation 6, we set $\sum_{(w,w')\in E_{d,k}} \beta^e_{k,(w,w')}/|E_{d,k}| = 1$ if $E^o_{d,k} = \emptyset$. Then, $p(V_{d,k}|G^o_{d,k};\beta_k)$ will reduce to the following function:

$$p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \emptyset;\beta_k) = p(V_{d,k}|\beta_k) = \prod_{w\in V_{d,k}} \beta^v_{k,w}. \tag{S1}$$

The above function is a legal probability measure over all possible $V_{d,k}$, because we have:

$$\sum_{V_{d,k}} p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \emptyset;\beta_k) = \sum_{V_{d,k}} \prod_{w\in V_{d,k}} \beta^v_{k,w} = \prod_{n\in V^o_{d,k}} \sum_{w_n\in V} \beta^v_{k,w} = 1 \tag{S2}$$

**Category 2:** Here, we consider the graph structure $G^o_{d,k}$ with only one edge, i.e., $|E^o_{d,k}| = 1$. Without loss of generality, we let $E^o_{d,k} = \{(i,j)\}$ and $V^{\neg}_{d,k} = V_{d,k} - \{w_{d,i}, w_{d,j}\}$. Then, we have:

$$p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \{(i,j)\};\beta_k) = (\prod_{w\in V^{\neg}_{d,k}} \beta^v_{k,w}) \times (\beta^v_{k,w_{d,i}}\beta^v_{k,w_{d,j}}\beta^e_{k,(w_{d,i},w_{d,j})}) \tag{S3}$$

Summing of the above functions over all possible $V_{d,k}$, we have

$$\begin{aligned} \sum_{V_{d,k}} p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \{(i,j)\};\beta_k) = \\ (\sum_{V^{\neg}_{d,k}} \prod_{w\in V^{\neg}_{d,k}} \beta^v_{k,w}) \times (\sum_{w_i,w_j\in V} \beta^v_{k,w_{d,i}}\beta^v_{k,w_{d,j}}\beta^e_{k,(w_{d,i},w_{d,j})}) = (\prod_{w\in V^{o\neg}_{d,k}} \sum_{w\in V} \beta^v_{k,w}) \times 1 = 1 \end{aligned} \tag{S4}$$

**Category 3:** When $|E^o_{d,k}| > 1$, we first have the following derivation:

$$\begin{aligned} p(V_{d,k}|G^o_{d,k};\beta_k) = \frac{1}{|E_{d,k}|} \prod_{w\in V_{d,k}} \beta^v_{k,w} \sum_{(w',w'')\in E_{d,k}} \beta^e_{k,(w,w')} = \\ \frac{1}{|E^o_{d,k}|} \sum_{(i,j)\in E^o_{d,k}} (\prod_{w\in V_{d,k}} \beta^v_{k,w})\beta^e_{k,(w_i,w_j)} = \frac{1}{|E^o_{d,k}|} \sum_{(i,j)\in E^o_{d,k}} p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \{(i,j)\};\beta_k) \end{aligned} \tag{S5}$$

Then, summing the above function over all possible $V_{d,k}$, we have:

$$\begin{aligned} \sum_{V_{d,k}} p(V_{d,k}|G^o_{d,k};\beta_k) = \sum_{V_{d,k}} \frac{1}{|E^o_{d,k}|} \sum_{(i,j)\in E^o_{d,k}} p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \{(i,j)\};\beta_k) = \\ \frac{1}{|E^o_{d,k}|} \sum_{(i,j)\in E^o_{d,k}} \sum_{V_{d,k}} p(V_{d,k}|G^o_{d,k}, E^o_{d,k} = \{(i,j)\};\beta_k) = \frac{1}{|E^o_{d,k}|} \sum_{(i,j)\in E^o_{d,k}} 1 = 1. \end{aligned} \tag{S6}$$

where the above derivation is based on the discussion in situation 2. □

The above lemma tells us that each factor of $p(V_d|Z_d, G^o_d;\beta_k)$ in equation 6, i.e., $p(V_{d,k}|G^o_{d,k};\beta_k)$, is a legal possibility measure. Meanwhile, considering those factors are independent to each other when given topic assignment set $Z_d$, we have the conclusion that $p(V_{d,k}|G^o_{d,k};\beta_k)$ is also a legal possibility measure. □

### A.2 The Derivation of the Objective Function

Here we first provide the detailed derivation of Equation 9 with the guidance of variation inference algorithm [10]. Then, we further simplify each item among it.

**Detailed Derivation of Equation 9:** Given the joint model $p(G_d, \theta_d, Z_d; \alpha)$ of $G_d$ in Equation 3 and the variational family $p(\theta_d, Z_d|G_d)$ in Equation 7, the objective of inferring GNTM is minimize the KL divergence between approximate posterior $q(\theta_d, Z_d|G_d)$ and true posterior $p(\theta_d, Z_d|G_d)$ for all latent variables:

$$\begin{aligned} &arg \min_{Z_d,\theta_d} KL(q(\theta_d, Z_d|G_d)||p(\theta_d, Z_d|G_d)) \\ &= E_{q(\theta_d,Z_d|G_d)}[\log q(\theta_d, Z_d|G_d) - \log p(\theta_d, Z_d|G_d)] \\ &= E_{q(\theta_d,Z_d|G_d)}[\log q(\theta_d, Z_d|G_d) - \log p(G_d|\theta_d, Z_d) - \log p(\theta_d) - \log p(Z_d)] + \log p(G_d) \\ &= -\mathcal{L}_d + \log p(G_d). \end{aligned} \tag{S7}$$

Due to the $\log p(G_d)$ is constant, the above objective can be transformed to maximize the follow Evidence Lower Bound (ELBO) of $\log p(G_d)$:

$$\begin{aligned} \mathcal{L}_d &= E_{q(\theta_d,Z_d|G_d)}[\log p(G_d|\theta_d, Z_d)] - KL[q(\theta_d|G_d)||p(\theta)] - E_{q(\theta_d|G_d)}[KL(q(Z_d|G_d)||p(Z_d))] \\ &= E_{q(Z_d|G_d)}[\log p(G^o_d|Z_d; M)] + E_{q(Z_d|G_d)}[\log p(V_d|Z_d, G^o_d; \beta)] \\ &\quad -KL[q(\theta_d|G_d)|p(\theta_d)] - E_{q(\theta_d|G_d)}[\sum_{n=1}^{N_d} KL[q(z_{d,n}|G_d, w_{d,n})||p(z_{d,n}|\theta_d)]] \\ &= \mathcal{L}^1_d + \mathcal{L}^2_d - \mathcal{L}^3_d - \mathcal{L}^4_d, \end{aligned} \tag{S8}$$

where $\mathcal{L}_d$ can be split into four terms , denoted as $\mathcal{L}^1_d$, $\mathcal{L}^2_d$, $\mathcal{L}^3_d$, and $\mathcal{L}^4_d$, respectively.

**Simplification for each item in Equation 9:** As for $\mathcal{L}^1_d$, we can derive:

$$\begin{aligned} \mathcal{L}^1_d &= E_{q(Z_d|G_d)}[\sum_{(n,n')\in E^o_d} \log m_{z_{d,n},z_{d,n'}} + \sum_{(n,n')\notin E^o_d} \log(1 - m_{z_{d,n},z_{d,n'}})] \\ &= \sum_{(n,n')\in E^o_d} E_{q(z_{d,n},z_{d,n'}|G_d)}[\log m_{z_{d,n},z_{d,n'}}] + \sum_{(n,n')\notin E^o_d} E_{q(z_{d,n},z_{d,n'}|G_d)}[\log(1 - m_{z_{d,n},z_{d,n'}})] \\ &= \sum_{(n,n')\in E^o_d} \varphi^T_{d,n} \cdot (\log M) \cdot \varphi_{d,n'} + \sum_{(n,n')\notin E^o_d} \varphi^T_{d,n} \cdot \log(1 - M) \cdot \varphi_{d,n'}. \end{aligned} \tag{S9}$$

Note that, in practice, to relax the unbalance between the two terms in above equation, we follow [12] and re-weight the second term with $|E^o_d|(\sum_{(n,n')\notin E^o_d} 1)^{-1}$. That is:

$$\mathcal{L}^1_d = \sum_{(n,n')\in E^o_d} \varphi^T_{d,n} \cdot (\log M) \cdot \varphi_{d,n'} + \frac{|E^o_d|}{\sum_{(n,n')\notin E^o_d} 1} \sum_{(n,n')\notin E^o_d} \varphi^T_{d,n} \cdot \log(1 - M) \cdot \varphi_{d,n'}. \tag{S10}$$

As for $\mathcal{L}^2_d$, we can derive:

$$\begin{aligned} \mathcal{L}^2_d &= E_{q(Z_d|G_d)}[\sum_{k=1}^{K}(\sum_{w\in V_{d,k}} \log \beta^v_{k,w} + \log \frac{\sum_{(w,w')\in E_{d,k}} \beta^e_{k,(w,w')}}{|E_{d,k}|})] \\ &= E_{q(Z_d|G_d)}[\sum_{n=1}^{N_d} \log \beta^v_{z_{d,n},w}] + E_{q(Z_d|G_d)}[\sum_{k=1}^{K} \log \frac{\sum_{(w,w')\in E_{d,k}} \beta^e_{k,(w,w')}}{|E_{d,k}|})] \\ &= \sum_{n=1}^{N_d} \varphi^T_{d,n} \cdot \log \beta^v_{\cdot,w_n} + \sum_{k=1}^{K} E_{q(Z_d|G_d)}[\log \frac{\sum_{(n,n')\in E^o_d} z^k_{d,n} z^k_{d,n'} \beta^e_{k,(w_{d,n},w_{d,n'})}}{\sum_{(n,n')\in E^o_d} z^k_{d,n} z^k_{d,n'}}]. \end{aligned} \tag{S11}$$

As for $\mathcal{L}_d^3$, we follow the idea in Laplace approximation [18] by rewriting $p(\theta_d) = Dir(\alpha)$ as the logistic normal distribution with mean $\mu^0$ and covariance matrix $\Sigma^0$. each dimension of them can be computed by:

$$\mu_{0,k} = \log \alpha_k - \frac{1}{K}\sum_i \alpha_i, \;\; \Sigma_{0,k,k} = \frac{1}{\alpha_k}(1-\frac{2}{K}) + \frac{1}{K^2}\sum_i \frac{1}{\alpha_i}. \tag{S12}$$

Then, the KL divergence $\mathcal{L}_d^3$ can be computed between two logistic normal distribution:

$$\mathcal{L}_d^3 = \frac{1}{2}\{Tr(\Sigma_0^{-1}\Sigma_d) + (\mu_d - \mu_0)^T\Sigma_0^{-1}(\mu_d - \mu_0) - K + \log\frac{|\Sigma_0|}{|\Sigma_d|}\}, \tag{S13}$$

where diagonal covariance $\Sigma_d = diag(\delta_d)$ returns a square diagonal matrix with the elements of vector $\delta_d$ on the main diagonal.

In term of $\mathcal{L}_d^4$, we have:

$$\mathcal{L}_d^4 = \sum_{n=1}^{N_d} \varphi_{d,n} \cdot (\log \varphi_{d,n} - E_{q(\theta_d|G_d)}[\log \theta_d]), \tag{S14}$$

where $E_{q(\theta_d|G_d)}[\log \theta_d])$ can also be computed analytically if $q(\theta_d|G_d)$ is Dirichlet distribution because $\log \theta_d$ is the sufficient statistics. However, we still use the sampling strategy to approximate its value, which is more general and can also be applied for more flexible situations without conjugate distributions, such as [1, 20].

**Algorithm 1:** The Generative Process of Graph Neural Topic Model

1. Draw the topic proportion of the document $d$, $\theta_d \sim Dir(\alpha)$.
2. Draw topic for $n$-th word $z_{d,n} \sim \text{Multi}(\theta_d)$, $n = 1, 2, ..., N$.
3. Draw the graph structure $G_d^o \sim p(G_d^o|Z_d)$ using Equation 4.
4. Draw the word set $p(V_d|G_d^o, Z_d, \beta)$ using Equation 6.

**Algorithm 2:** The Generative Process of Latent Dirichlet Allocation

1. Draw the topic proportion of the document $d$, $\theta_d \sim Dir(\alpha)$.
2. Draw topic for $n$-th word $z_{d,n} \sim \text{Multi}(\theta_d)$, $n = 1, 2, ..., N$.
3. Draw the word set $p(V_d|Z_d, \beta) = \prod_{n=1}^{N_d} p(w_{d,n}|z_{d,n}, \beta^v) = \prod_{n=1}^{N_d} \text{Multi}(\beta^v_{z_{d,n}})$.

### A.3 Proof of Corollary 1

*Proof.* Here, to provide one intuitive comparison, we first show the generative processes of GNTM and Latent Dirichlet Allocation (LDA) [2] in Algorithm 1 and 2, respectively. Then, for each document $d$, we refer to the variational family $q(\theta_d, Z_d|V_d)$ and objective function, i.e., ELBO, of LDA under our mathematical notation:

$$q(\theta_d, Z_d|V_d) = q(\theta_d; \gamma_d)\prod_{n=1}^{N_d} q(z_{d,n}; \varphi_{d,n})$$

$$\mathcal{L}_d^{LDA} = E_{q(Z_d)}[\log q(V_d|Z_d;\beta^v)] - KL[q(\theta_d)||p(\theta_d)] - E_{q(\theta)}[\sum_{n=1}^{N_d} KL[q(z_{d,n})||p(z_{d,n}|\theta_d)]], \tag{S15}$$

where $q(z_{d,n}; \varphi_{d,n})$ is a multinomial distribution with parameter $\varphi_{d,n}$ and $q(\theta_d; \gamma_d)$ is a Dirichlet distribution with parameter $\gamma_d$; The first term can be further derived as $\sum_{n=1}^{N_d} \varphi_{d,n} \cdot \log \beta^v_{\cdot, w_n}$.

By comparing the Equation S8 and S15, we can find that $\mathcal{L}_d^3$ and $\mathcal{L}_d^4$ is same as the last two terms in $\mathcal{L}_d^{LDA}$ except that $q(\theta_d|G_d)$ is the Laplace approximation of the Dirichlet distribution.

In addition, we can find that if $E_{d,k} = \emptyset$, both terms in Equation S10 will vanish to 0. In other words, $\mathcal{L}_d^1$ will reduce to 0 if $E_{d,k} = \emptyset$. Meanwhile, based on the discussion about Equation S1, we can find that $\mathcal{L}_d^2$ will reduce to the following function:

$$\begin{aligned} \mathcal{L}_d^2(E_d^o = \emptyset) &= E_{q(Z_d|G_d)}[\sum_{k=1}^{K} \log p(V_{d,k}|G_{d,k}^o, E_{d,k}^o = \emptyset; \beta_k)] \\ = E_{q(Z_d|G_d)}[\sum_{k=1}^{K} \sum_{w \in V_{d,k}} \log \beta_{k,w}^v] &= E_{q(Z_d|G_d)}[\sum_{n=1}^{N_d} \log \beta_{z_{d,n},w}^v] = \sum_{n=1}^{N_d} \varphi_{d,n} \cdot \log \beta_{\cdot,w}^v, \end{aligned} \tag{S16}$$

which indicates that $\mathcal{L}_d^2$ reduce to the first term in $\mathcal{L}_d^{LDA}$.

In sum, we can conclude that the objective of GNTM $\mathcal{L}_d$ in Equation 9 will be same as the ELBO $L_d^{LDA}$ of LDA if the edge set $E_d$ of the document graph $G_d$ is a empty set, except that we approach the Dirichlet distribution by the Laplace approximation. □

### A.4 Dataset Statistics

The statistics of the datasets in experiments are shown in Table S2. The links of the datasets can be found in the footnote: 20 News Group (**20NG**) [13][1], Tag My News (**TMN**) [19][2], the British National Corpus (**BNC**) [4][3], **Reuters** extracted from the Reuters-21578 dataset [4]. In particular, the stop word list is same as that in genism package[5]. The frequency threshold to filter out low-frequency words is set as 20 for 20NS, 5 for TMN, 120 for BNC, and 10 for Reuters. The frequency threshold to filter out low-frequency dependency edges is set as 10 for 20NG, 3 for TMN, 40 for BNC, and 15 for Reuters.

Table S2: The statistics of the datasets.

| | Docs | Data split (train/val/test) | Vocabulary | Word token | Edge set | Edge token | Label |
|---|---|---|---|---|---|---|---|
| 20NG | 17679 | 7105/3517/7057 | 9161 | 1265888 | 33758 | 665502 | 20 |
| TMN | 26171 | 18973/2428/4770 | 5572 | 138875 | 6799 | 44658 | 7 |
| BNC | 16963 | 14966/998/999 | 9401 | 6736890 | 28756 | 2089104 | N/A |
| Reuters | 10727 | 6757/965/3005 | 5907 | 675982 | 16668 | 857146 | N/A |

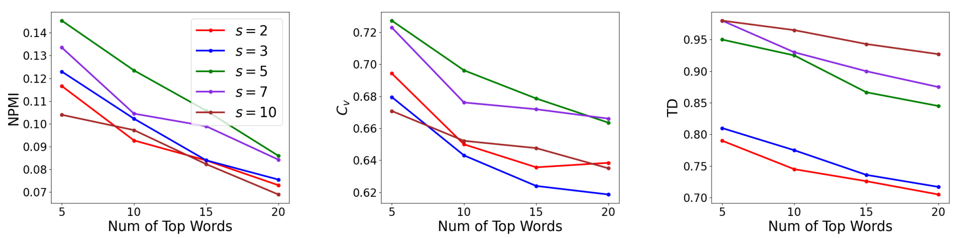
### A.5 Detailed Experimental Setup

We reproduced all baselines with the guidance of original papers and the codes provided by the original authors or other widely used sources. The code links of them can be found in the footnote: 1) **LDA** [2][6]; 2) **GSM** [16][7]; 3) **ProdLDA** [18][8]; 4) **ETM** [6][9]; 5) **GraphBTM** [21][10]; 6) **iDocNADE** and **iDocNADEe** [8][11]; 7) **TopicRNN** [7][12]; Actually, we also tried to reproduce the **GaussianLDA** [5] as a baseline, which also introduces word embedding into topic modeling, like ETM and GNTM. However, GaussianLDA is highly time-costing in practice if there is a large

[1] http://qwone.com/ jason/20Newsgroups/
[2] http://acube.di.unipi.it/tmn- dataset/
[3] https://www.sketchengine.eu/british-national-corpus/
[4] https://trec.nist.gov/data/reuters/reuters.html
[5] https://radimrehurek.com/gensim/
[6] https://radimrehurek.com/gensim/models/ldamodel.html
[7] https://github.com/linkstrife/NVDM-GSM
[8] https://github.com/akashgit/autoencoding_vi_for_topic_models
[9] https://github.com/adjidieng/ETM
[10] https://github.com/valdersoul/GraphBTM
[11] https://github.com/pgcool/iDocNADEe
[12] https://github.com/narratives-of-war/topic-rnn

vocabulary size, like that in our case. Without specific description, the parameters of baselines are the same as those recommended by the original paper or the official code. In particular, we set word embedding size in GSM and ETM as 300, same as that in GNTM, to provide a fair comparison.

As for GNTM, we set all main hyper-parameters as $s = 5$, $\alpha = 1.0$, $L = H = 300$, and $Y = 64$. In particular, Figure S1 shows the parameter analysis of the window size $s$, which is the most important parameter, based on 20NG dataset with the fixing number of topics $K = 20$. As the result shows, we can find that when $s = 5$, GNTM achieves best topic coherence both on NPMI and $C_v$ metrics with high topic diversity. Therefore, we set $s = 5$ in our experiments. The Glove word embeddings [17] used in GNTM and ETM can be downloaded in this link [13]. In the optimization, we followed [16] and alternately updated the decoder parameters with topic representation and the encoder parameters. Only one sample is used in neural variational inference for $\theta_d$ and $Z_d$ if needed. We use Adam [11] optimizer with the initial learning rate of 0.001 and decay it by a factor of 2 if the validation loss has not improved in 5 epochs and terminate training once the learning rate has decayed a total of 5 times. The batch size is set as 100. In particular, we pre-trained our model without the restructure of the graph structure in the first 15 epochs or 2000 iterations, which benefits for the robustness of our model empirically. Following [9], the temperature $\tau$ for the STGS estimator [9] is annealed from 1.0 to 0.3 using the schedule $\tau = max(0.3, exp(-\eta \text{iter}))$ of the global training step iter, where $\tau$ is updated every 1000 steps and $\eta = 0.00003$.




(a) NPMI (b) $C_v$ (c) TD

Figure S1: The parameter analysis of the window size $s$ on 20NG dataset with $K = 20$. We show both topic coherence measurements, i.e., NPMI and $C_v$, and topic diversity based on varying numbers of top words of each topic as the metrics.

### A.6 The oblation study

Here, we explore the impact of representing topics with a distribution over word dependency edges on the interpretability in our model. To be specific, we construct a variant of GNTM by removing the distribution over edges from topics, namely GNTM w/o $\beta^e$. Different from the description in Section 3.1.2, we assume that words are independent form each other given the topic assignments $Z_d$. In other words, different from Equation (6), we formulate $P(V_d|Z_d, G_d^o)$ as:

$$p(V_d|Z_d, G_d^o; \beta^v) = \prod_{n=1}^{N_d} \beta^v_{z_{d,n}, w_{d,n}}. \quad \text{(S17)}$$

Then, we can re-derive the loss function based on Equation (9). The experimental results on 20NG dataset are showed in Table S3. It demonstrates that removing the distribution over edges from topics causes the decrease of interpretability of topics.

### A.7 More Experimental Results

Here, we show more experimental results on topic quality with varying top words in Table S4-S15 and detailed results on document cluster with standard deviation in Table S16 and S17. Same as the description in Section 4.2, we can find that GNTM outperforms other baselines in terms of topic coherence with high topic diversity under all settings consistently. Note that, although iDcoNADe can outperform GNTM on TMN dataset based on $C_v$ metric with top 10, 15 or 20 words, I still think

[13] https://nlp.stanford.edu/projects/glove/

Table S3: The oblation study on the impact of representing topics with a distribution over edges.

| | Num. of topic | 20 | 30 | 50 | 70 | 100 |
|---|---|---|---|---|---|---|
| NPMI | GNTM | 0.1235 | 0.1074 | 0.0807 | 0.0710 | 0.0587 |
| | GNTM w/o $\beta^e$ | 0.0994 | 0.0953 | 0.0791 | 0.0701 | 0.0386 |
| $C_v$ | GNTM | 0.6962 | 0.6605 | 0.6328 | 0.6222 | 0.5939 |
| | GNTM w/o $\beta^e$ | 0.6479 | 0.6445 | 0.6055 | 0.5924 | 0.5532 |
| TD | GNTM | 0.9110 | 0.8759 | 0.8352 | 0.7231 | 0.5530 |
| | GNTM w/o $\beta^e$ | 0.9000 | 0.9000 | 0.7700 | 0.4900 | 0.3700 |

iDcoNADe is uncomparative due to the unignored gap compared with most of other baselines on NPMI metric, which has been demonstrated to be highly consistent with the human evaluated the quality of the topics [14], and topic diversity, where lower value indicates redundant topics.

Table S4: The performance of topic ocherence based on NPMI metric with top 5 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0856±0.0207 | 0.0617±0.0217 | 0.0541±0.0156 | 0.0503±0.0142 | 0.0305±0.0119 | -0.0590±0.0090 | -0.1045±0.0223 | -0.1359±0.0087 | -0.1719±0.0097 | -0.1036±0.0132 |
| GSM | 0.1049±0.0257 | 0.1032±0.0208 | 0.1046±0.0129 | 0.0846±0.0114 | 0.0778±0.0063 | 0.0217±0.0346 | 0.0299±0.0152 | 0.0291±0.0178 | -0.0246±0.0108 | -0.1776±0.0182 |
| ProdLDA | 0.0544±0.0544 | 0.0450±0.0349 | 0.0167±0.0350 | 0.0071±0.0220 | -0.0108±0.0283 | -0.1292±0.0446 | -0.1802±0.0196 | -0.2220±0.0226 | -0.2482±0.0248 | -0.2656±0.0164 |
| ETM | 0.0912±0.0135 | 0.0866±0.0092 | 0.0730±0.0087 | 0.0591±0.0067 | 0.0555±0.0062 | 0.0016±0.0091 | 0.0099±0.0031 | -0.0062±0.0121 | -0.0255±0.0107 | -0.0369±0.0088 |
| GraphBTM | 0.0383±0.0030 | 0.0448±0.0188 | 0.0646±0.0029 | 0.0366±0.0321 | -0.0062±0.0578 | -0.1534±0.0174 | -0.1552±0.0073 | -0.2438±0.0131 | -0.2831±0.0148 | -0.2922±0.0065 |
| iDocNADE | -0.0608±0.0397 | -0.0801±0.0410 | -0.1141±0.0393 | -0.1290±0.0347 | -0.1550±0.0195 | -0.5393±0.0372 | -0.5180±0.0343 | -0.4840±0.0255 | -0.4588±0.0241 | -0.4445±0.0199 |
| iDocNADEe | - | - | -0.0236±0.0578 | - | -0.0853±0.0533 | - | - | -0.4712±0.0144 | - | -0.4569±0.0118 |
| TopicRNN | -0.1403±0.0124 | -0.1956±0.0236 | -0.2795±0.0088 | -0.3036±0.0092 | -0.3159±0.0143 | -0.4229±0.0354 | -0.4100±0.0151 | -0.4063±0.0270 | -0.4066±0.0252 | -0.4119±0.0096 |
| GNTM | 0.1453±0.0128 | 0.1284±0.0073 | 0.1084±0.0190 | 0.1049±0.0026 | 0.0888±0.0070 | -0.0192±0.0143 | 0.0005±0.0151 | 0.0284±0.0074 | 0.0428±0.0071 | 0.0607±0.0178 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0929±0.0083 | 0.0886±0.0115 | 0.1077±0.0160 | 0.1087±0.0201 | 0.1025±0.0186 | 0.0916±0.0110 | 0.0994±0.0176 | 0.0866±0.0127 | 0.0769±0.0110 | 0.0518±0.0177 |
| GSM | 0.0989±0.0031 | 0.1138±0.0021 | 0.1179±0.0061 | 0.0935±0.0191 | 0.0750±0.0037 | 0.1284±0.0080 | 0.1305±0.0079 | 0.0968±0.0178 | 0.0637±0.0165 | 0.0561±0.0080 |
| ProdLDA | 0.0950±0.0108 | 0.0902±0.0184 | 0.0807±0.0064 | 0.0471±0.0128 | 0.0373±0.0070 | 0.0842±0.0230 | 0.0926±0.0194 | 0.0983±0.0126 | 0.0665±0.0045 | 0.0693±0.0214 |
| ETM | 0.1055±0.0071 | 0.0934±0.0030 | 0.0874±0.0030 | 0.0821±0.0067 | 0.0741±0.0046 | 0.1040±0.0127 | 0.0878±0.0125 | 0.0852±0.0098 | 0.0719±0.0043 | 0.0686±0.0048 |
| GraphBTM | 0.0173±0.0088 | 0.0077±0.0101 | 0.0291±0.0121 | 0.0401±0.0173 | 0.0057±0.0014 | 0.0762±0.0027 | 0.0711±0.0118 | 0.0706±0.0068 | 0.0748±0.0164 | 0.0552±0.0216 |
| iDocNADE | -0.0146±0.0487 | -0.0321±0.0378 | -0.0921±0.0350 | -0.1257±0.0159 | -0.1370±0.0142 | -0.0517±0.0420 | -0.1184±0.0574 | -0.1652±0.0445 | -0.1913±0.0214 | -0.2176±0.0113 |
| iDocNADEe | - | - | -0.0687±0.0145 | - | -0.1506±0.0115 | - | - | -0.1369±0.0235 | - | -0.2299±0.0039 |
| TopicRNN | -0.0599±0.0326 | -0.1439±0.0304 | -0.1934±0.0150 | -0.2318±0.0124 | -0.2607±0.0124 | -0.2290±0.0462 | -0.2637±0.0169 | -0.3218±0.0180 | -0.3367±0.0186 | -0.3553±0.0111 |
| GNTM | 0.1268±0.0126 | 0.1179±0.0065 | 0.1107±0.0059 | 0.1122±0.0058 | 0.1048±0.0040 | 0.1392±0.0160 | 0.1445±0.0173 | 0.1395±0.0061 | 0.1171±0.0075 | 0.0972±0.0088 |

Table S5: The performance of topic coherence based on $C_v$ metric with top 5 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.6825±0.0251 | 0.6486±0.0289 | 0.6437±0.0128 | 0.6251±0.0159 | 0.6087±0.0119 | 0.4047±0.0149 | 0.3841±0.0185 | 0.3592±0.0102 | 0.3419±0.0093 | 0.4148±0.0113 |
| GSM | 0.6795±0.0270 | 0.6634±0.0134 | 0.6526±0.0163 | 0.6276±0.0088 | 0.6373±0.0065 | 0.5152±0.0339 | 0.5274±0.0133 | 0.5060±0.0239 | 0.4056±0.0164 | 0.1770±0.0209 |
| ProdLDA | 0.6212±0.0611 | 0.6136±0.0489 | 0.5823±0.0313 | 0.5820±0.0261 | 0.5596±0.0180 | 0.3601±0.0330 | 0.3231±0.0126 | 0.2887±0.0098 | 0.2883±0.0108 | 0.2841±0.0055 |
| ETM | 0.6779±0.0229 | 0.6690±0.0165 | 0.6604±0.0065 | 0.6420±0.0086 | 0.6340±0.0092 | 0.5101±0.0244 | 0.5223±0.0062 | 0.5036±0.0157 | 0.4946±0.0181 | 0.4797±0.0081 |
| GraphBTM | 0.6010±0.0042 | 0.6073±0.0252 | 0.6369±0.0083 | 0.5909±0.0477 | 0.5467±0.0664 | 0.2237±0.0102 | 0.2584±0.0110 | 0.2840±0.0072 | 0.2742±0.0045 | 0.2762±0.0074 |
| iDocNADE | 0.4769±0.0288 | 0.4526±0.0290 | 0.4283±0.0286 | 0.4119±0.0163 | 0.3901±0.0074 | 0.3779±0.0211 | 0.3613±0.0164 | 0.3543±0.0097 | 0.3471±0.0078 | 0.3427±0.0078 |
| iDocNADEe | - | - | 0.4716±0.0464 | - | 0.4195±0.0378 | - | - | 0.3393±0.0078 | - | 0.3284±0.0096 |
| TopicRNN | 0.4546±0.0313 | 0.3939±0.0251 | 0.3231±0.0073 | 0.3040±0.0079 | 0.3061±0.0130 | 0.3467±0.0251 | 0.3326±0.0146 | 0.3181±0.0153 | 0.3122±0.0070 | 0.3097±0.0096 |
| GNTM | 0.7272±0.0118 | 0.6951±0.0054 | 0.6749±0.0158 | 0.6752±0.0081 | 0.6584±0.0066 | 0.4819±0.0069 | 0.5208±0.0168 | 0.5610±0.0121 | 0.5784±0.0090 | 0.5938±0.0173 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.6376±0.0086 | 0.6394±0.0120 | 0.6639±0.0063 | 0.6669±0.0078 | 0.6638±0.0037 | 0.5862±0.0080 | 0.5985±0.0203 | 0.5803±0.0066 | 0.5703±0.0094 | 0.5494±0.0150 |
| GSM | 0.6375±0.0046 | 0.6540±0.0042 | 0.6573±0.0097 | 0.6300±0.0136 | 0.5966±0.0040 | 0.6489±0.0085 | 0.6461±0.0082 | 0.6083±0.0181 | 0.5807±0.0196 | 0.5898±0.0111 |
| ProdLDA | 0.6449±0.0145 | 0.6293±0.0071 | 0.6193±0.0155 | 0.5858±0.0130 | 0.5738±0.0083 | 0.6272±0.0343 | 0.6179±0.0213 | 0.6157±0.0153 | 0.5788±0.0077 | 0.5681±0.0142 |
| ETM | 0.6653±0.0083 | 0.6436±0.0040 | 0.6337±0.0059 | 0.6217±0.0064 | 0.6115±0.0091 | 0.6090±0.0162 | 0.6047±0.0095 | 0.6051±0.0105 | 0.5916±0.0059 | 0.5905±0.0054 |
| GraphBTM | 0.5218±0.0133 | 0.5061±0.0158 | 0.5317±0.0184 | 0.5569±0.0298 | 0.4954±0.0022 | 0.6033±0.0060 | 0.5925±0.0121 | 0.5975±0.0061 | 0.6015±0.0235 | 0.5747±0.0303 |
| iDocNADE | 0.5157±0.0451 | 0.5008±0.0276 | 0.4539±0.0359 | 0.4220±0.0121 | 0.4133±0.0137 | 0.5407±0.0334 | 0.5061±0.0457 | 0.4563±0.0351 | 0.4217±0.0043 | 0.4061±0.0023 |
| iDocNADEe | - | - | 0.4563±0.0164 | - | 0.3775±0.0130 | - | - | 0.4413±0.0256 | - | 0.3748±0.0081 |
| TopicRNN | 0.5164±0.0464 | 0.4442±0.0417 | 0.3804±0.0111 | 0.3443±0.0206 | 0.3186±0.0090 | 0.3445±0.0272 | 0.3316±0.0204 | 0.2995±0.0186 | 0.2957±0.0064 | 0.2803±0.0076 |
| GNTM | 0.6868±0.0204 | 0.6780±0.0114 | 0.6664±0.0056 | 0.6677±0.0074 | 0.6583±0.0038 | 0.6696±0.0117 | 0.6612±0.0175 | 0.6455±0.0084 | 0.6298±0.0116 | 0.6151±0.0065 |

Table S6: The performance of topic diversity with top 5 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.8380±0.0117 | 0.8450±0.0222 | 0.8464±0.0246 | 0.8640±0.0138 | 0.8816±0.0129 | 0.9900±0.0089 | 0.9958±0.0056 | 0.9944±0.0020 | 0.9940±0.0019 | 0.8816±0.0242 |
| GSM | 0.7380±0.0453 | 0.7128±0.0439 | 0.6920±0.0188 | 0.5682±0.0185 | 0.3628±0.0254 | 0.9720±0.0117 | 0.9586±0.0245 | 0.6472±0.0416 | 0.2874±0.0141 | 0.0608±0.0143 |
| ProdLDA | 0.9420±0.0147 | 0.8810±0.0271 | 0.7944±0.0213 | 0.7534±0.0272 | 0.7104±0.0146 | 0.8600±0.0374 | 0.6742±0.0363 | 0.7016±0.0531 | 0.7124±0.0641 | 0.6956±0.0397 |
| ETM | 0.8520±0.0319 | 0.7876±0.0090 | 0.6880±0.0152 | 0.6474±0.0100 | 0.5788±0.0086 | 0.9640±0.0224 | 0.9158±0.0209 | 0.7648±0.0174 | 0.6888±0.0132 | 0.5804±0.0160 |
| GraphBTM | 0.3080±0.0214 | 0.2385±0.0770 | 0.2776±0.0362 | 0.2682±0.0768 | 0.4380±0.1570 | 0.4120±0.0954 | 0.4785±0.0114 | 0.7712±0.0265 | 0.8185±0.0290 | 0.7856±0.0194 |
| iDocNADE | 0.7280±0.0376 | 0.6662±0.0311 | 0.6504±0.0280 | 0.6380±0.0217 | 0.6120±0.0100 | 0.5360±0.0242 | 0.4850±0.0155 | 0.3936±0.0169 | 0.3368±0.0201 | 0.2800±0.0066 |
| iDocNADEe | - | - | 0.5776±0.0635 | - | 0.5628±0.0659 | - | - | 0.4280±0.0173 | - | 0.3728±0.0109 |
| TopicRNN | 0.9800±0.0063 | 0.9676±0.0108 | 0.9768±0.0047 | 0.9752±0.0052 | 0.9536±0.0062 | 0.7160±0.0595 | 0.7316±0.0369 | 0.6392±0.0266 | 0.6206±0.0202 | 0.6304±0.0141 |
| GNTM | 0.9540±0.0102 | 0.9251±0.0179 | 0.9016±0.0190 | 0.8037±0.0324 | 0.6528±0.0217 | 1.0000±0.0000 | 1.0000±0.0000 | 0.9984±0.0032 | 0.9940±0.0019 | 0.9368±0.0372 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.8080±0.0421 | 0.8234±0.0483 | 0.8368±0.0250 | 0.8452±0.0069 | 0.8516±0.0062 | 0.6500±0.0141 | 0.6690±0.0311 | 0.6776±0.0130 | 0.6882±0.0204 | 0.6964±0.0102 |
| GSM | 0.8200±0.0253 | 0.7744±0.0184 | 0.7176±0.0161 | 0.6018±0.0440 | 0.4140±0.0133 | 0.6060±0.0488 | 0.5717±0.0225 | 0.5240±0.0378 | 0.3916±0.0439 | 0.2376±0.0165 |
| ProdLDA | 0.9140±0.0287 | 0.8636±0.0125 | 0.7904±0.0169 | 0.7368±0.0206 | 0.6340±0.0252 | 0.9580±0.0117 | 0.9104±0.0150 | 0.8360±0.0143 | 0.7712±0.0130 | 0.6748±0.0118 |
| ETM | 0.9380±0.0133 | 0.8943±0.0108 | 0.8112±0.0117 | 0.7564±0.0148 | 0.6896±0.0246 | 0.8940±0.0427 | 0.8462±0.0168 | 0.7560±0.0181 | 0.6734±0.0122 | 0.5636±0.0189 |
| GraphBTM | 0.1700±0.0603 | 0.1211±0.0670 | 0.1776±0.0360 | 0.2008±0.0549 | 0.1536±0.0267 | 0.1940±0.0185 | 0.1665±0.0084 | 0.1776±0.0419 | 0.1997±0.0075 | 0.2444±0.0345 |
| iDocNADE | 0.5980±0.0279 | 0.6022±0.0274 | 0.6392±0.0452 | 0.6492±0.0629 | 0.6276±0.0414 | 0.5320±0.0337 | 0.5264±0.0250 | 0.4816±0.0130 | 0.4628±0.0103 | 0.4460±0.0132 |
| iDocNADEe | - | - | 0.6904±0.0255 | - | 0.6296±0.0051 | - | - | 0.5528±0.0223 | - | 0.4672±0.0129 |
| TopicRNN | 0.9880±0.0117 | 0.9810±0.0077 | 0.9832±0.0064 | 0.9746±0.0040 | 0.9676±0.0066 | 0.9640±0.0150 | 0.9612±0.0065 | 0.9624±0.0149 | 0.9606±0.0103 | 0.9508±0.0116 |
| GNTM | 0.9820±0.0117 | 0.9598±0.0164 | 0.9024±0.0203 | 0.7990±0.0762 | 0.7056±0.0589 | 0.9060±0.0150 | 0.8582±0.0265 | 0.8024±0.0235 | 0.7874±0.0167 | 0.7440±0.0102 |

Table S7: The performance of topic coherence based on NPMI metric with top 10 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0552±0.0133 | 0.0425±0.0172 | 0.0209±0.0098 | 0.0069±0.0106 | -0.0156±0.0047 | -0.1222±0.0100 | -0.1655±0.0096 | -0.2152±0.0088 | -0.2447±0.0040 | -0.1714±0.0087 |
| GSM | 0.0902±0.0181 | 0.0872±0.0023 | 0.0767±0.0117 | 0.0560±0.0063 | 0.0543±0.0052 | -0.0249±0.0048 | -0.0289±0.0076 | -0.0040±0.0113 | -0.0335±0.0058 | -0.1607±0.0093 |
| ProdLDA | 0.0059±0.0265 | -0.0017±0.0286 | -0.0296±0.0195 | -0.0410±0.0190 | -0.0511±0.0237 | -0.1900±0.0324 | -0.2019±0.0154 | -0.2348±0.0176 | -0.2603±0.0220 | -0.2842±0.0132 |
| ETM | 0.0750±0.0097 | 0.0749±0.0088 | 0.0549±0.0051 | 0.0423±0.0043 | 0.0328±0.0017 | -0.0391±0.0216 | -0.0505±0.0122 | -0.0594±0.0140 | -0.0677±0.0080 | -0.0829±0.0081 |
| GraphBTM | 0.0309±0.0017 | 0.0372±0.0163 | 0.0509±0.0019 | 0.0271±0.0295 | -0.0271±0.0546 | -0.1460±0.0113 | -0.1732±0.0098 | -0.2541±0.0021 | -0.2883±0.0048 | -0.3020±0.0041 |
| iDocNADE | -0.1679±0.0179 | -0.1960±0.0157 | -0.2180±0.0154 | -0.2208±0.0122 | -0.2370±0.0147 | -0.5261±0.0202 | -0.5054±0.0105 | -0.4744±0.0132 | -0.4555±0.0091 | -0.4409±0.0118 |
| iDocNADEe | - | - | -0.1197±0.0592 | - | -0.1680±0.0583 | - | - | -0.4601±0.0145 | - | -0.4492±0.0055 |
| TopicRNN | -0.1959±0.0061 | -0.2398±0.0217 | -0.2950±0.0052 | -0.3154±0.0041 | -0.3332±0.0089 | -0.4500±0.0197 | -0.4289±0.0093 | -0.4327±0.0137 | -0.4334±0.0076 | -0.4311±0.0037 |
| GNTM | 0.1235±0.0132 | 0.1074±0.0111 | 0.0807±0.0105 | 0.0710±0.0022 | 0.0587±0.0066 | -0.0327±0.0137 | -0.0273±0.0101 | -0.0235±0.0041 | -0.0253±0.0018 | -0.0245±0.0179 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0749±0.0052 | 0.0712±0.0131 | 0.0820±0.0130 | 0.0743±0.0142 | 0.0669±0.0197 | 0.0753±0.0085 | 0.0651±0.0157 | 0.0474±0.0106 | 0.0297±0.0134 | 0.0019±0.0092 |
| GSM | 0.0798±0.0023 | 0.0895±0.0022 | 0.0893±0.0029 | 0.0792±0.0084 | 0.0578±0.0015 | 0.1018±0.0072 | 0.0965±0.0055 | 0.0640±0.0096 | 0.0269±0.0119 | 0.0220±0.0032 |
| ProdLDA | 0.0694±0.0129 | 0.0628±0.0137 | 0.0417±0.0043 | 0.0218±0.0101 | 0.0141±0.0103 | 0.0232±0.0146 | 0.0329±0.0106 | 0.0289±0.0050 | 0.0099±0.0051 | 0.0224±0.0106 |
| ETM | 0.0883±0.0066 | 0.0794±0.0062 | 0.0720±0.0027 | 0.0680±0.0046 | 0.0634±0.0034 | 0.0609±0.0079 | 0.0572±0.0057 | 0.0476±0.0098 | 0.0369±0.0031 | 0.0317±0.0046 |
| GraphBTM | 0.0195±0.0051 | 0.0076±0.0117 | 0.0231±0.0106 | 0.0364±0.0165 | 0.0024±0.0008 | 0.0427±0.0062 | 0.0441±0.0085 | 0.0467±0.0059 | 0.0402±0.0103 | 0.0175±0.0226 |
| iDocNADE | -0.0426±0.0288 | -0.0660±0.0309 | -0.1228±0.0214 | -0.1637±0.0246 | -0.1792±0.0068 | -0.1433±0.0197 | -0.2049±0.0245 | -0.2464±0.0162 | -0.2623±0.0149 | -0.2832±0.0067 |
| iDocNADEe | - | - | -0.1002±0.0066 | - | -0.1673±0.0069 | - | - | -0.2160±0.0064 | - | -0.2773±0.0049 |
| TopicRNN | -0.1022±0.0159 | -0.1668±0.0240 | -0.2077±0.0141 | -0.2440±0.0134 | -0.2770±0.0103 | -0.2565±0.0278 | -0.2971±0.0141 | -0.3311±0.0078 | -0.3475±0.0023 | -0.3570±0.0054 |
| GNTM | 0.0967±0.0054 | 0.0971±0.0048 | 0.0884±0.0021 | 0.0893±0.0042 | 0.0836±0.0033 | 0.0965±0.0086 | 0.0945±0.0080 | 0.0835±0.0050 | 0.0543±0.0081 | 0.0417±0.0051 |

Table S8: The performance of topic coherence based on $C_v$ metric with top 10 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.6182±0.0153 | 0.5953±0.0313 | 0.5807±0.0136 | 0.5521±0.0110 | 0.5259±0.0041 | 0.2907±0.0072 | 0.2969±0.0078 | 0.3064±0.0076 | 0.3260±0.0056 | 0.3342±0.0060 |
| GSM | 0.6393±0.0256 | 0.6189±0.0151 | 0.5991±0.0183 | 0.5656±0.0066 | 0.5652±0.0078 | 0.4332±0.0166 | 0.4203±0.0092 | 0.4002±0.0103 | 0.3024±0.0121 | 0.1955±0.0117 |
| ProdLDA | 0.5799±0.0372 | 0.5605±0.0425 | 0.5121±0.0218 | 0.5131±0.0254 | 0.4909±0.0145 | 0.3206±0.0167 | 0.3059±0.0109 | 0.3129±0.0132 | 0.3343±0.0196 | 0.3599±0.0171 |
| ETM | 0.6421±0.0156 | 0.6365±0.0148 | 0.6077±0.0072 | 0.5916±0.0079 | 0.5745±0.0026 | 0.4320±0.0238 | 0.4259±0.0194 | 0.4199±0.0124 | 0.4211±0.0124 | 0.4033±0.0024 |
| GraphBTM | 0.5268±0.0058 | 0.5452±0.0333 | 0.5797±0.0072 | 0.5253±0.0650 | 0.4480±0.0778 | 0.2128±0.0147 | 0.2570±0.0074 | 0.3400±0.0034 | 0.3639±0.0092 | 0.3799±0.0018 |
| iDocNADE | 0.4422±0.0282 | 0.4319±0.0161 | 0.4418±0.0123 | 0.4309±0.0060 | 0.4224±0.0085 | 0.5825±0.0190 | 0.5656±0.0114 | 0.5410±0.0109 | 0.5249±0.0085 | 0.5147±0.0099 |
| iDocNADEe | - | - | 0.3894±0.0143 | - | 0.3942±0.0106 | - | - | 0.5292±0.0109 | - | 0.5196±0.0057 |
| TopicRNN | 0.4029±0.0249 | 0.3947±0.0170 | 0.3779±0.0090 | 0.3943±0.0124 | 0.4089±0.0131 | 0.5169±0.0203 | 0.5066±0.0085 | 0.5083±0.0097 | 0.5091±0.0073 | 0.5086±0.0041 |
| GNTM | 0.6962±0.0126 | 0.6605±0.0094 | 0.6328±0.0108 | 0.6222±0.0055 | 0.5939±0.0091 | 0.3977±0.0117 | 0.4237±0.0155 | 0.4489±0.0114 | 0.4585±0.0071 | 0.4521±0.0163 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.5705±0.0077 | 0.5815±0.0107 | 0.5969±0.0051 | 0.5983±0.0062 | 0.5938±0.0136 | 0.5327±0.0085 | 0.5273±0.0261 | 0.5102±0.0086 | 0.4951±0.0141 | 0.4622±0.0135 |
| GSM | 0.5755±0.0062 | 0.5912±0.0070 | 0.5863±0.0051 | 0.5573±0.0125 | 0.5122±0.0035 | 0.5868±0.0167 | 0.5732±0.0037 | 0.5252±0.0112 | 0.4817±0.0215 | 0.4678±0.0067 |
| ProdLDA | 0.5793±0.0220 | 0.5620±0.0098 | 0.5321±0.0129 | 0.5047±0.0096 | 0.4907±0.0098 | 0.5337±0.0118 | 0.5290±0.0217 | 0.5249±0.0107 | 0.4954±0.0121 | 0.4978±0.0091 |
| ETM | 0.6071±0.0044 | 0.5918±0.0082 | 0.5740±0.0073 | 0.5591±0.0078 | 0.5465±0.0081 | 0.5494±0.0100 | 0.5362±0.0039 | 0.5312±0.0085 | 0.5138±0.0046 | 0.5073±0.0081 |
| GraphBTM | 0.4368±0.0156 | 0.4008±0.0326 | 0.4413±0.0306 | 0.4793±0.0459 | 0.3825±0.0026 | 0.5017±0.0105 | 0.4986±0.0154 | 0.5032±0.0086 | 0.4966±0.0227 | 0.4619±0.0433 |
| iDocNADE | 0.4349±0.0201 | 0.4255±0.0237 | 0.3963±0.0239 | 0.3868±0.0103 | 0.3823±0.0145 | 0.5116±0.0142 | 0.4859±0.0158 | 0.4651±0.0140 | 0.4569±0.0073 | 0.4632±0.0041 |
| iDocNADEe | - | - | 0.3350±0.0099 | - | 0.3091±0.0082 | - | - | 0.3885±0.0108 | - | 0.4154±0.0082 |
| TopicRNN | 0.4139±0.0318 | 0.3598±0.0261 | 0.3176±0.0065 | 0.3015±0.0041 | 0.2943±0.0063 | 0.3776±0.0155 | 0.3900±0.0098 | 0.4120±0.0053 | 0.4345±0.0083 | 0.4524±0.0108 |
| GNTM | 0.6218±0.0137 | 0.6253±0.0104 | 0.6067±0.0037 | 0.6024±0.0070 | 0.5930±0.0032 | 0.5917±0.0048 | 0.5781±0.0150 | 0.5623±0.0116 | 0.5276±0.0133 | 0.5154±0.0086 |

Table S9: The performance of topic diversity with top 10 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.7900_{\pm0.0130}$ | $0.8158_{\pm0.0222}$ | $0.8176_{\pm0.0125}$ | $0.8256_{\pm0.0087}$ | $0.8274_{\pm0.0109}$ | $0.9840_{\pm0.0086}$ | $0.9964_{\pm0.0031}$ | $0.9936_{\pm0.0020}$ | $0.9912_{\pm0.0035}$ | $0.8066_{\pm0.0156}$ |
| GSM | $0.7040_{\pm0.0271}$ | $0.6924_{\pm0.0349}$ | $0.6368_{\pm0.0131}$ | $0.5156_{\pm0.0234}$ | $0.3246_{\pm0.0188}$ | $0.9530_{\pm0.0125}$ | $0.9176_{\pm0.0210}$ | $0.6100_{\pm0.0363}$ | $0.2632_{\pm0.0155}$ | $0.0600_{\pm0.0091}$ |
| ProdLDA | $0.9300_{\pm0.0077}$ | $0.8582_{\pm0.0152}$ | $0.7628_{\pm0.0139}$ | $0.7038_{\pm0.0162}$ | $0.6356_{\pm0.0064}$ | $0.7890_{\pm0.0278}$ | $0.6302_{\pm0.0339}$ | $0.5992_{\pm0.0647}$ | $0.6108_{\pm0.0634}$ | $0.6056_{\pm0.0415}$ |
| ETM | $0.8550_{\pm0.0327}$ | $0.7796_{\pm0.0119}$ | $0.6636_{\pm0.0054}$ | $0.6092_{\pm0.0019}$ | $0.5412_{\pm0.0084}$ | $0.9540_{\pm0.0222}$ | $0.8976_{\pm0.0120}$ | $0.7380_{\pm0.0145}$ | $0.6470_{\pm0.0091}$ | $0.5502_{\pm0.0138}$ |
| GraphBTM | $0.2780_{\pm0.0121}$ | $0.2211_{\pm0.0565}$ | $0.2556_{\pm0.0211}$ | $0.2404_{\pm0.0605}$ | $0.3936_{\pm0.1320}$ | $0.4450_{\pm0.0875}$ | $0.5091_{\pm0.0216}$ | $0.6960_{\pm0.0125}$ | $0.7306_{\pm0.0171}$ | $0.6902_{\pm0.0135}$ |
| iDocNADE | $0.7160_{\pm0.0337}$ | $0.6636_{\pm0.0181}$ | $0.6428_{\pm0.0200}$ | $0.6166_{\pm0.0134}$ | $0.5892_{\pm0.0121}$ | $0.5850_{\pm0.0235}$ | $0.4990_{\pm0.0242}$ | $0.3940_{\pm0.0109}$ | $0.3246_{\pm0.0081}$ | $0.2714_{\pm0.0059}$ |
| iDocNADEe | - | - | $0.5448_{\pm0.0416}$ | - | $0.5338_{\pm0.0538}$ | - | - | $0.3904_{\pm0.0096}$ | - | $0.3238_{\pm0.0115}$ |
| TopicRNN | $0.9590_{\pm0.0097}$ | $0.9482_{\pm0.0141}$ | $0.9504_{\pm0.0076}$ | $0.9422_{\pm0.0065}$ | $0.9262_{\pm0.0105}$ | $0.7190_{\pm0.0493}$ | $0.6944_{\pm0.0390}$ | $0.6080_{\pm0.0258}$ | $0.5834_{\pm0.0188}$ | $0.5812_{\pm0.0151}$ |
| GNTM | $0.9110_{\pm0.0086}$ | $0.8759_{\pm0.0145}$ | $0.8352_{\pm0.0115}$ | $0.7231_{\pm0.0319}$ | $0.5530_{\pm0.0139}$ | $0.9990_{\pm0.0020}$ | $0.9970_{\pm0.0027}$ | $0.9932_{\pm0.0048}$ | $0.9870_{\pm0.0045}$ | $0.9014_{\pm0.0541}$ |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.7770_{\pm0.0466}$ | $0.7776_{\pm0.0364}$ | $0.7692_{\pm0.0213}$ | $0.7840_{\pm0.0140}$ | $0.7890_{\pm0.0105}$ | $0.6640_{\pm0.0206}$ | $0.6742_{\pm0.0207}$ | $0.6744_{\pm0.0143}$ | $0.6730_{\pm0.0164}$ | $0.6752_{\pm0.0042}$ |
| GSM | $0.7630_{\pm0.0144}$ | $0.7082_{\pm0.0120}$ | $0.6280_{\pm0.0150}$ | $0.5348_{\pm0.0386}$ | $0.3586_{\pm0.0119}$ | $0.6020_{\pm0.0186}$ | $0.5530_{\pm0.0163}$ | $0.5004_{\pm0.0207}$ | $0.3646_{\pm0.0330}$ | $0.2274_{\pm0.0201}$ |
| ProdLDA | $0.8890_{\pm0.0285}$ | $0.8056_{\pm0.0091}$ | $0.7280_{\pm0.0125}$ | $0.6544_{\pm0.0156}$ | $0.5544_{\pm0.0126}$ | $0.9180_{\pm0.0223}$ | $0.8690_{\pm0.0224}$ | $0.7764_{\pm0.0130}$ | $0.6974_{\pm0.0080}$ | $0.5898_{\pm0.0105}$ |
| ETM | $0.9180_{\pm0.0103}$ | $0.8671_{\pm0.0141}$ | $0.7824_{\pm0.0098}$ | $0.7151_{\pm0.0090}$ | $0.6426_{\pm0.0143}$ | $0.8990_{\pm0.0150}$ | $0.8166_{\pm0.0055}$ | $0.7224_{\pm0.0085}$ | $0.6274_{\pm0.0080}$ | $0.5122_{\pm0.0130}$ |
| GraphBTM | $0.1810_{\pm0.0466}$ | $0.1383_{\pm0.0664}$ | $0.1760_{\pm0.0270}$ | $0.1804_{\pm0.0415}$ | $0.1468_{\pm0.0220}$ | $0.2020_{\pm0.0227}$ | $0.1591_{\pm0.0063}$ | $0.1700_{\pm0.0300}$ | $0.1846_{\pm0.0134}$ | $0.2356_{\pm0.0193}$ |
| iDocNADE | $0.6030_{\pm0.0474}$ | $0.5914_{\pm0.0316}$ | $0.6076_{\pm0.0336}$ | $0.6172_{\pm0.0498}$ | $0.5966_{\pm0.0393}$ | $0.5140_{\pm0.0224}$ | $0.4998_{\pm0.0155}$ | $0.4372_{\pm0.0180}$ | $0.4208_{\pm0.0107}$ | $0.3996_{\pm0.0056}$ |
| iDocNADEe | - | - | $0.6368_{\pm0.0259}$ | - | $0.6110_{\pm0.0149}$ | - | - | $0.4832_{\pm0.0077}$ | - | $0.4252_{\pm0.0035}$ |
| TopicRNN | $0.9710_{\pm0.0097}$ | $0.9602_{\pm0.0098}$ | $0.9592_{\pm0.0095}$ | $0.9430_{\pm0.0065}$ | $0.9414_{\pm0.0071}$ | $0.9560_{\pm0.0146}$ | $0.9508_{\pm0.0075}$ | $0.9360_{\pm0.0157}$ | $0.9256_{\pm0.0098}$ | $0.9092_{\pm0.0047}$ |
| GNTM | $0.9230_{\pm0.0298}$ | $0.9294_{\pm0.0142}$ | $0.8552_{\pm0.0220}$ | $0.7290_{\pm0.0752}$ | $0.6432_{\pm0.0575}$ | $0.8470_{\pm0.0163}$ | $0.7862_{\pm0.0225}$ | $0.7060_{\pm0.0178}$ | $0.6716_{\pm0.0116}$ | $0.6236_{\pm0.0112}$ |

Table S10: The performance of topic coherence based on NPMI metric with top 15 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.0357_{\pm0.0134}$ | $0.0241_{\pm0.0127}$ | $-0.0065_{\pm0.0091}$ | $-0.0195_{\pm0.0086}$ | $-0.0471_{\pm0.0052}$ | $-0.1700_{\pm0.0087}$ | $-0.2086_{\pm0.0040}$ | $-0.2637_{\pm0.0080}$ | $-0.2908_{\pm0.0034}$ | $-0.2186_{\pm0.0066}$ |
| GSM | $0.0724_{\pm0.0128}$ | $0.0634_{\pm0.0084}$ | $0.0496_{\pm0.0105}$ | $0.0312_{\pm0.0046}$ | $0.0376_{\pm0.0050}$ | $-0.0560_{\pm0.0123}$ | $-0.0667_{\pm0.0091}$ | $-0.0358_{\pm0.0106}$ | $-0.0442_{\pm0.0056}$ | $-0.1585_{\pm0.0095}$ |
| ProdLDA | $-0.0301_{\pm0.0151}$ | $-0.0367_{\pm0.0281}$ | $-0.0650_{\pm0.0128}$ | $-0.0754_{\pm0.0174}$ | $-0.0818_{\pm0.0183}$ | $-0.2062_{\pm0.0180}$ | $-0.2153_{\pm0.0124}$ | $-0.2488_{\pm0.0196}$ | $-0.2712_{\pm0.0224}$ | $-0.2973_{\pm0.0146}$ |
| ETM | $0.0611_{\pm0.0083}$ | $0.0648_{\pm0.0082}$ | $0.0425_{\pm0.0051}$ | $0.0297_{\pm0.0041}$ | $0.0173_{\pm0.0022}$ | $-0.0717_{\pm0.0174}$ | $-0.0816_{\pm0.0064}$ | $-0.0873_{\pm0.0076}$ | $-0.1002_{\pm0.0104}$ | $-0.1140_{\pm0.0065}$ |
| GraphBTM | $0.0273_{\pm0.0027}$ | $0.0287_{\pm0.0134}$ | $0.0403_{\pm0.0034}$ | $0.0196_{\pm0.0253}$ | $-0.0384_{\pm0.0523}$ | $-0.1573_{\pm0.0148}$ | $-0.1863_{\pm0.0091}$ | $-0.2663_{\pm0.0019}$ | $-0.2997_{\pm0.0034}$ | $-0.3135_{\pm0.0037}$ |
| iDocNADE | $-0.2158_{\pm0.0172}$ | $-0.2466_{\pm0.0114}$ | $-0.2605_{\pm0.0097}$ | $-0.2677_{\pm0.0084}$ | $-0.2795_{\pm0.0079}$ | $-0.5166_{\pm0.0077}$ | $-0.4966_{\pm0.0054}$ | $-0.4702_{\pm0.0077}$ | $-0.4545_{\pm0.0094}$ | $-0.4378_{\pm0.0089}$ |
| iDocNADEe | - | - | $-0.1779_{\pm0.0438}$ | - | $-0.2180_{\pm0.0477}$ | - | - | $-0.4561_{\pm0.0109}$ | - | $-0.4500_{\pm0.0009}$ |
| TopicRNN | $-0.2212_{\pm0.0051}$ | $-0.2638_{\pm0.0180}$ | $-0.3037_{\pm0.0057}$ | $-0.3264_{\pm0.0030}$ | $-0.3384_{\pm0.0069}$ | $-0.4532_{\pm0.0122}$ | $-0.4377_{\pm0.0089}$ | $-0.4421_{\pm0.0079}$ | $-0.4397_{\pm0.0026}$ | $-0.4401_{\pm0.0035}$ |
| GNTM | $0.1058_{\pm0.0096}$ | $0.0855_{\pm0.0082}$ | $0.0542_{\pm0.0070}$ | $0.0454_{\pm0.0035}$ | $0.0342_{\pm0.0039}$ | $-0.0539_{\pm0.0113}$ | $-0.0564_{\pm0.0078}$ | $-0.0700_{\pm0.0039}$ | $-0.0791_{\pm0.0056}$ | $-0.0747_{\pm0.0203}$ |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.0652_{\pm0.0045}$ | $0.0632_{\pm0.0111}$ | $0.0674_{\pm0.0115}$ | $0.0586_{\pm0.0121}$ | $0.0479_{\pm0.0195}$ | $0.0605_{\pm0.0058}$ | $0.0396_{\pm0.0158}$ | $0.0180_{\pm0.0102}$ | $-0.0046_{\pm0.0125}$ | $-0.0315_{\pm0.0062}$ |
| GSM | $0.0671_{\pm0.0025}$ | $0.0757_{\pm0.0018}$ | $0.0762_{\pm0.0034}$ | $0.0675_{\pm0.0052}$ | $0.0503_{\pm0.0014}$ | $0.0852_{\pm0.0078}$ | $0.0711_{\pm0.0070}$ | $0.0409_{\pm0.0091}$ | $0.0115_{\pm0.0126}$ | $0.0081_{\pm0.0035}$ |
| ProdLDA | $0.0552_{\pm0.0129}$ | $0.0490_{\pm0.0115}$ | $0.0244_{\pm0.0042}$ | $0.0080_{\pm0.0085}$ | $0.0029_{\pm0.0102}$ | $-0.0124_{\pm0.0083}$ | $-0.0039_{\pm0.0047}$ | $-0.0041_{\pm0.0092}$ | $-0.0236_{\pm0.0082}$ | $-0.0183_{\pm0.0094}$ |
| ETM | $0.0776_{\pm0.0068}$ | $0.0700_{\pm0.0056}$ | $0.0648_{\pm0.0032}$ | $0.0600_{\pm0.0034}$ | $0.0546_{\pm0.0029}$ | $0.0332_{\pm0.0110}$ | $0.0287_{\pm0.0048}$ | $0.0155_{\pm0.0095}$ | $0.0060_{\pm0.0037}$ | $0.0010_{\pm0.0038}$ |
| GraphBTM | $0.0177_{\pm0.0030}$ | $0.0067_{\pm0.0095}$ | $0.0219_{\pm0.0115}$ | $0.0336_{\pm0.0166}$ | $-0.0012_{\pm0.0003}$ | $0.0186_{\pm0.0075}$ | $0.0235_{\pm0.0083}$ | $0.0283_{\pm0.0071}$ | $0.0210_{\pm0.0086}$ | $-0.0039_{\pm0.0241}$ |
| iDocNADE | $-0.0697_{\pm0.0308}$ | $-0.0917_{\pm0.0294}$ | $-0.1470_{\pm0.0196}$ | $-0.1895_{\pm0.0176}$ | $-0.1999_{\pm0.0077}$ | $-0.2190_{\pm0.0149}$ | $-0.2576_{\pm0.0176}$ | $-0.2843_{\pm0.0076}$ | $-0.3039_{\pm0.0101}$ | $-0.3196_{\pm0.0056}$ |
| iDocNADEe | - | - | $-0.1199_{\pm0.0069}$ | - | $-0.1773_{\pm0.0051}$ | - | - | $-0.2520_{\pm0.0064}$ | - | $-0.3019_{\pm0.0039}$ |
| TopicRNN | $-0.1291_{\pm0.0189}$ | $-0.1867_{\pm0.0247}$ | $-0.2243_{\pm0.0116}$ | $-0.2557_{\pm0.0113}$ | $-0.2832_{\pm0.0089}$ | $-0.2827_{\pm0.0104}$ | $-0.3108_{\pm0.0090}$ | $-0.3409_{\pm0.0057}$ | $-0.3543_{\pm0.0031}$ | $-0.3616_{\pm0.0031}$ |
| GNTM | $0.0832_{\pm0.0041}$ | $0.0854_{\pm0.0029}$ | $0.0760_{\pm0.0033}$ | $0.0766_{\pm0.0037}$ | $0.0705_{\pm0.0032}$ | $0.0743_{\pm0.0073}$ | $0.0610_{\pm0.0097}$ | $0.0522_{\pm0.0069}$ | $0.0205_{\pm0.0056}$ | $0.0121_{\pm0.0082}$ |

Table S11: The performance of topic coherence based on $C_v$ metric with top 15 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.6047_{\pm0.0148}$ | $0.5797_{\pm0.0283}$ | $0.5521_{\pm0.0175}$ | $0.5287_{\pm0.0095}$ | $0.4992_{\pm0.0051}$ | $0.3070_{\pm0.0079}$ | $0.3403_{\pm0.0101}$ | $0.3990_{\pm0.0083}$ | $0.4302_{\pm0.0050}$ | $0.3579_{\pm0.0099}$ |
| GSM | $0.6184_{\pm0.0216}$ | $0.5925_{\pm0.0246}$ | $0.5672_{\pm0.0201}$ | $0.5346_{\pm0.0067}$ | $0.5388_{\pm0.0099}$ | $0.4035_{\pm0.0188}$ | $0.3941_{\pm0.0151}$ | $0.3450_{\pm0.0115}$ | $0.2613_{\pm0.0104}$ | $0.2488_{\pm0.0165}$ |
| ProdLDA | $0.5613_{\pm0.0176}$ | $0.5328_{\pm0.0407}$ | $0.4870_{\pm0.0114}$ | $0.4902_{\pm0.0215}$ | $0.4680_{\pm0.0129}$ | $0.3554_{\pm0.0175}$ | $0.3517_{\pm0.0100}$ | $0.3858_{\pm0.0237}$ | $0.4124_{\pm0.0297}$ | $0.4472_{\pm0.0220}$ |
| ETM | $0.6299_{\pm0.0164}$ | $0.6287_{\pm0.0149}$ | $0.5958_{\pm0.0074}$ | $0.5743_{\pm0.0086}$ | $0.5546_{\pm0.0043}$ | $0.4175_{\pm0.0089}$ | $0.4075_{\pm0.0091}$ | $0.4011_{\pm0.0101}$ | $0.3995_{\pm0.0100}$ | $0.4017_{\pm0.0043}$ |
| GraphBTM | $0.5048_{\pm0.0115}$ | $0.5135_{\pm0.0401}$ | $0.5617_{\pm0.0050}$ | $0.5038_{\pm0.0663}$ | $0.4171_{\pm0.0803}$ | $0.2642_{\pm0.0293}$ | $0.3124_{\pm0.0165}$ | $0.4162_{\pm0.0015}$ | $0.4560_{\pm0.0060}$ | $0.4698_{\pm0.0062}$ |
| iDocNADE | $0.4555_{\pm0.0181}$ | $0.4756_{\pm0.0130}$ | $0.4824_{\pm0.0121}$ | $0.4807_{\pm0.0068}$ | $0.4790_{\pm0.0052}$ | $0.6703_{\pm0.0073}$ | $0.6547_{\pm0.0059}$ | $0.6298_{\pm0.0094}$ | $0.6152_{\pm0.0097}$ | $0.6010_{\pm0.0082}$ |
| iDocNADEe | - | - | $0.3753_{\pm0.0271}$ | - | $0.4119_{\pm0.0385}$ | - | - | $0.6154_{\pm0.0115}$ | - | $0.6140_{\pm0.0019}$ |
| TopicRNN | $0.4283_{\pm0.0168}$ | $0.4342_{\pm0.0126}$ | $0.4426_{\pm0.0093}$ | $0.4749_{\pm0.0066}$ | $0.4925_{\pm0.0129}$ | $0.6121_{\pm0.0147}$ | $0.6032_{\pm0.0092}$ | $0.6096_{\pm0.0088}$ | $0.6108_{\pm0.0032}$ | $0.6150_{\pm0.0052}$ |
| GNTM | $0.6786_{\pm0.0121}$ | $0.6411_{\pm0.0091}$ | $0.6028_{\pm0.0073}$ | $0.5905_{\pm0.0039}$ | $0.5594_{\pm0.0108}$ | $0.3745_{\pm0.0109}$ | $0.3915_{\pm0.0163}$ | $0.4082_{\pm0.0077}$ | $0.4105_{\pm0.0111}$ | $0.4054_{\pm0.0042}$ |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | $0.5458_{\pm0.0060}$ | $0.5629_{\pm0.0101}$ | $0.5739_{\pm0.0039}$ | $0.5717_{\pm0.0075}$ | $0.5714_{\pm0.0135}$ | $0.5183_{\pm0.0095}$ | $0.4991_{\pm0.0284}$ | $0.4819_{\pm0.0112}$ | $0.4605_{\pm0.0106}$ | $0.4407_{\pm0.0091}$ |
| GSM | $0.5445_{\pm0.0077}$ | $0.5650_{\pm0.0068}$ | $0.5607_{\pm0.0090}$ | $0.5304_{\pm0.0106}$ | $0.4814_{\pm0.0039}$ | $0.5583_{\pm0.0162}$ | $0.5310_{\pm0.0101}$ | $0.4840_{\pm0.0093}$ | $0.4382_{\pm0.0252}$ | $0.4145_{\pm0.0082}$ |
| ProdLDA | $0.5510_{\pm0.0332}$ | $0.5331_{\pm0.0121}$ | $0.5007_{\pm0.0168}$ | $0.4767_{\pm0.0112}$ | $0.4651_{\pm0.0086}$ | $0.4982_{\pm0.0061}$ | $0.4927_{\pm0.0127}$ | $0.4893_{\pm0.0131}$ | $0.4653_{\pm0.0150}$ | $0.4613_{\pm0.0085}$ |
| ETM | $0.5839_{\pm0.0078}$ | $0.5745_{\pm0.0092}$ | $0.5592_{\pm0.0077}$ | $0.5403_{\pm0.0076}$ | $0.5228_{\pm0.0068}$ | $0.5250_{\pm0.0129}$ | $0.5008_{\pm0.0084}$ | $0.4931_{\pm0.0116}$ | $0.4780_{\pm0.0052}$ | $0.4741_{\pm0.0048}$ |
| GraphBTM | $0.4061_{\pm0.0113}$ | $0.3572_{\pm0.0378}$ | $0.4113_{\pm0.0424}$ | $0.4533_{\pm0.0580}$ | $0.3305_{\pm0.0021}$ | $0.4445_{\pm0.0124}$ | $0.4473_{\pm0.0163}$ | $0.4596_{\pm0.0133}$ | $0.4502_{\pm0.0197}$ | $0.4111_{\pm0.0498}$ |
| iDocNADE | $0.3986_{\pm0.0208}$ | $0.3962_{\pm0.0213}$ | $0.3868_{\pm0.0166}$ | $0.3936_{\pm0.0111}$ | $0.3942_{\pm0.0158}$ | $0.4895_{\pm0.0045}$ | $0.4965_{\pm0.0169}$ | $0.4898_{\pm0.0179}$ | $0.5023_{\pm0.0057}$ | $0.5126_{\pm0.0031}$ |
| iDocNADEe | - | - | $0.2964_{\pm0.0099}$ | - | $0.3068_{\pm0.0040}$ | - | - | $0.4121_{\pm0.0151}$ | - | $0.4672_{\pm0.0059}$ |
| TopicRNN | $0.3809_{\pm0.0232}$ | $0.3433_{\pm0.0208}$ | $0.3250_{\pm0.0065}$ | $0.3260_{\pm0.0052}$ | $0.3300_{\pm0.0030}$ | $0.4354_{\pm0.0096}$ | $0.4696_{\pm0.0093}$ | $0.5038_{\pm0.0060}$ | $0.5296_{\pm0.0057}$ | $0.5514_{\pm0.0104}$ |
| GNTM | $0.6015_{\pm0.0120}$ | $0.6041_{\pm0.0069}$ | $0.5856_{\pm0.0069}$ | $0.5807_{\pm0.0079}$ | $0.5664_{\pm0.0040}$ | $0.5526_{\pm0.0046}$ | $0.5226_{\pm0.0185}$ | $0.5127_{\pm0.0136}$ | $0.4825_{\pm0.0101}$ | $0.4755_{\pm0.0098}$ |

Table S12: The performance of topic diversity with top 15 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.7770±0.0147 | 0.7950±0.0226 | 0.7952±0.0196 | 0.7974±0.0076 | 0.7934±0.0094 | 0.9808±0.0053 | 0.9886±0.0033 | 0.9890±0.0026 | 0.9882±0.0013 | 0.6820±0.0141 |
| GSM | 0.6890±0.0161 | 0.6628±0.0330 | 0.6070±0.0149 | 0.4872±0.0227 | 0.3028±0.0155 | 0.9228±0.0105 | 0.8828±0.0238 | 0.5732±0.0333 | 0.2452±0.0172 | 0.0552±0.0103 |
| ProdLDA | 0.9114±0.0170 | 0.8330±0.0137 | 0.7290±0.0095 | 0.6672±0.0178 | 0.5926±0.0079 | 0.7430±0.0338 | 0.5772±0.0419 | 0.5376±0.0674 | 0.5374±0.0609 | 0.5436±0.0487 |
| ETM | 0.8378±0.0331 | 0.7556±0.0156 | 0.6454±0.0097 | 0.5802±0.0032 | 0.5116±0.0087 | 0.9390±0.0213 | 0.8910±0.0124 | 0.7164±0.0136 | 0.6276±0.0124 | 0.5264±0.0096 |
| GraphBTM | 0.2745±0.0213 | 0.2195±0.0424 | 0.2419±0.0175 | 0.2260±0.0581 | 0.3626±0.1139 | 0.4423±0.0776 | 0.5228±0.0146 | 0.6503±0.0077 | 0.6672±0.0093 | 0.6203±0.0133 |
| iDocNADE | 0.7146±0.0284 | 0.6746±0.0210 | 0.6448±0.0122 | 0.6072±0.0199 | 0.5786±0.0119 | 0.5890±0.0133 | 0.5058±0.0155 | 0.3956±0.0143 | 0.3274±0.0088 | 0.2698±0.0077 |
| iDocNADEe | - | - | 0.5285±0.0440 | - | 0.5225±0.0497 | - | - | 0.3668±0.0122 | - | 0.3044±0.0068 |
| TopicRNN | 0.9478±0.0075 | 0.9372±0.0105 | 0.9288±0.0073 | 0.9188±0.0024 | 0.8924±0.0117 | 0.7216±0.0428 | 0.6844±0.0345 | 0.5974±0.0205 | 0.5742±0.0177 | 0.5460±0.0196 |
| GNTM | 0.8691±0.0153 | 0.8392±0.0190 | 0.7842±0.0193 | 0.6622±0.0319 | 0.5019±0.0133 | 0.9992±0.0016 | 0.9942±0.0010 | 0.9882±0.0077 | 0.9710±0.0099 | 0.8648±0.0649 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.7470±0.0341 | 0.7422±0.0319 | 0.7312±0.0302 | 0.7384±0.0147 | 0.7408±0.0113 | 0.6824±0.0162 | 0.6778±0.0161 | 0.6562±0.0128 | 0.6510±0.0155 | 0.6518±0.0060 |
| GSM | 0.7284±0.0072 | 0.6570±0.0112 | 0.5838±0.0115 | 0.4938±0.0313 | 0.3306±0.0101 | 0.5958±0.0192 | 0.5415±0.0089 | 0.4775±0.0193 | 0.3439±0.0296 | 0.2172±0.0169 |
| ProdLDA | 0.8590±0.0304 | 0.7698±0.0171 | 0.6794±0.0180 | 0.6070±0.0112 | 0.5010±0.0115 | 0.9016±0.0140 | 0.8482±0.0118 | 0.7314±0.0135 | 0.6448±0.0078 | 0.5364±0.0087 |
| ETM | 0.8785±0.0062 | 0.8312±0.0080 | 0.7549±0.0093 | 0.6834±0.0082 | 0.6091±0.0132 | 0.8758±0.0128 | 0.7930±0.0070 | 0.6880±0.0072 | 0.5934±0.0050 | 0.4828±0.0121 |
| GraphBTM | 0.1965±0.0437 | 0.1445±0.0534 | 0.1747±0.0202 | 0.1696±0.0367 | 0.1383±0.0210 | 0.2150±0.0156 | 0.1726±0.0052 | 0.1749±0.0276 | 0.1852±0.0128 | 0.2300±0.0222 |
| iDocNADE | 0.5890±0.0467 | 0.5710±0.0411 | 0.5854±0.0252 | 0.6012±0.0417 | 0.5788±0.0333 | 0.5106±0.0249 | 0.4790±0.0177 | 0.4266±0.0123 | 0.4038±0.0109 | 0.3860±0.0083 |
| iDocNADEe | - | - | 0.6240±0.0232 | - | 0.5884±0.0129 | - | - | 0.4494±0.0104 | - | 0.3980±0.0062 |
| TopicRNN | 0.9562±0.0022 | 0.9440±0.0133 | 0.9354±0.0070 | 0.9246±0.0075 | 0.9148±0.0079 | 0.9416±0.0123 | 0.9348±0.0126 | 0.9188±0.0165 | 0.9022±0.0075 | 0.8822±0.0022 |
| GNTM | 0.8890±0.0367 | 0.9034±0.0170 | 0.8224±0.0264 | 0.6834±0.0732 | 0.5944±0.0544 | 0.7890±0.0253 | 0.7128±0.0236 | 0.6384±0.0131 | 0.5976±0.0108 | 0.5438±0.0077 |

Table S13: The performance of topic coherence based on NPMI metric with top 20 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0214±0.0150 | 0.0075±0.0121 | -0.0257±0.0030 | -0.0437±0.0053 | -0.0691±0.0028 | -0.1973±0.0115 | -0.2329±0.0061 | -0.2955±0.0035 | -0.3221±0.0017 | -0.2529±0.0049 |
| GSM | 0.0590±0.0108 | 0.0487±0.0085 | 0.0313±0.0071 | 0.0167±0.0046 | 0.0266±0.0033 | -0.0784±0.0101 | -0.0928±0.0109 | -0.0554±0.0092 | -0.0531±0.0058 | -0.1534±0.0061 |
| ProdLDA | -0.0531±0.0151 | -0.0700±0.0212 | -0.0900±0.0131 | -0.0960±0.0132 | -0.1010±0.0149 | -0.2195±0.0198 | -0.2283±0.0121 | -0.2581±0.0219 | -0.2804±0.0234 | -0.3075±0.0152 |
| ETM | 0.0482±0.0081 | 0.0489±0.0093 | 0.0319±0.0052 | 0.0182±0.0024 | 0.0041±0.0031 | -0.0966±0.0126 | -0.1026±0.0028 | -0.1103±0.0065 | -0.1247±0.0085 | -0.1368±0.0058 |
| GraphBTM | 0.0244±0.0022 | 0.0266±0.0132 | 0.0327±0.0025 | 0.0137±0.0242 | -0.0463±0.0508 | -0.1661±0.0201 | -0.2056±0.0065 | -0.2757±0.0042 | -0.3064±0.0034 | -0.3200±0.0039 |
| iDocNADE | -0.2435±0.0155 | -0.2698±0.0086 | -0.2826±0.0065 | -0.2913±0.0069 | -0.3002±0.0045 | -0.5105±0.0058 | -0.4895±0.0045 | -0.4659±0.0075 | -0.4536±0.0071 | -0.4399±0.0054 |
| iDocNADEe | - | - | -0.2072±0.0450 | - | -0.2440±0.0419 | - | - | -0.4530±0.0091 | - | -0.4513±0.0023 |
| TopicRNN | -0.2394±0.0085 | -0.2783±0.0112 | -0.3125±0.0054 | -0.3303±0.0023 | -0.3423±0.0063 | -0.4584±0.0092 | -0.4443±0.0072 | -0.4426±0.0060 | -0.4412±0.0015 | -0.4442±0.0028 |
| GNTM | 0.0860±0.0064 | 0.0643±0.0072 | 0.0356±0.0064 | 0.0263±0.0046 | 0.0172±0.0039 | -0.0739±0.0114 | -0.0862±0.0108 | -0.1051±0.0038 | -0.1158±0.0072 | -0.1139±0.0228 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.0572±0.0050 | 0.0571±0.0098 | 0.0580±0.0111 | 0.0468±0.0105 | 0.0344±0.0165 | 0.0442±0.0069 | 0.0184±0.0141 | -0.0058±0.0125 | -0.0284±0.0093 | -0.0572±0.0045 |
| GSM | 0.0612±0.0032 | 0.0678±0.0024 | 0.0679±0.0027 | 0.0589±0.0048 | 0.0431±0.0017 | 0.0684±0.0044 | 0.0525±0.0057 | 0.0222±0.0100 | 0.0006±0.0124 | -0.0010±0.0038 |
| ProdLDA | 0.0478±0.0116 | 0.0359±0.0125 | 0.0160±0.0054 | -0.0009±0.0067 | -0.0065±0.0087 | -0.0382±0.0064 | -0.0402±0.0075 | -0.0374±0.0076 | -0.0528±0.0079 | -0.0489±0.0075 |
| ETM | 0.0687±0.0072 | 0.0634±0.0062 | 0.0591±0.0024 | 0.0537±0.0029 | 0.0472±0.0029 | 0.0133±0.0096 | 0.0054±0.0036 | -0.0110±0.0081 | -0.0178±0.0038 | -0.0232±0.0033 |
| GraphBTM | 0.0170±0.0039 | 0.0050±0.0101 | 0.0204±0.0120 | 0.0316±0.0166 | -0.0035±0.0001 | 0.0069±0.0098 | 0.0130±0.0100 | 0.0165±0.0066 | 0.0101±0.0097 | -0.0196±0.0222 |
| iDocNADE | -0.0889±0.0412 | -0.1130±0.0316 | -0.1648±0.0184 | -0.2056±0.0132 | -0.2161±0.0048 | -0.2554±0.0119 | -0.2863±0.0128 | -0.3089±0.0058 | -0.3251±0.0070 | -0.3358±0.0036 |
| iDocNADEe | - | - | -0.1323±0.0073 | - | -0.1847±0.0055 | - | - | -0.2742±0.0053 | - | -0.3176±0.0008 |
| TopicRNN | -0.1469±0.0178 | -0.1981±0.0220 | -0.2341±0.0087 | -0.2627±0.0091 | -0.2870±0.0087 | -0.2974±0.0102 | -0.3219±0.0080 | -0.3466±0.0060 | -0.3566±0.0033 | -0.3632±0.0022 |
| GNTM | 0.0752±0.0025 | 0.0754±0.0024 | 0.0671±0.0022 | 0.0676±0.0031 | 0.0609±0.0024 | 0.0542±0.0103 | 0.0396±0.0082 | 0.0296±0.0042 | -0.0012±0.0040 | -0.0084±0.0058 |

Table S14: The performance of topic coherence based on $C_v$ metric with top 20 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.5959±0.0139 | 0.5662±0.0266 | 0.5347±0.0130 | 0.5106±0.0055 | 0.4856±0.0055 | 0.3638±0.0103 | 0.3983±0.0080 | 0.4821±0.0042 | 0.5200±0.0028 | 0.4334±0.0061 |
| GSM | 0.6065±0.0202 | 0.5789±0.0224 | 0.5522±0.0191 | 0.5208±0.0068 | 0.5260±0.0119 | 0.3898±0.0271 | 0.3890±0.0198 | 0.3228±0.0125 | 0.2461±0.0063 | 0.2805±0.0116 |
| ProdLDA | 0.5535±0.0149 | 0.5124±0.0280 | 0.4729±0.0143 | 0.4770±0.0144 | 0.4573±0.0115 | 0.4000±0.0226 | 0.4042±0.0150 | 0.4394±0.0286 | 0.4692±0.0316 | 0.5083±0.0245 |
| ETM | 0.6258±0.0173 | 0.6199±0.0157 | 0.5934±0.0085 | 0.5670±0.0084 | 0.5471±0.0052 | 0.4077±0.0075 | 0.4011±0.0031 | 0.4021±0.0104 | 0.4024±0.0094 | 0.4061±0.0057 |
| GraphBTM | 0.4994±0.0153 | 0.5121±0.0491 | 0.5469±0.0078 | 0.4891±0.0692 | 0.4003±0.0773 | 0.3120±0.0380 | 0.3774±0.0140 | 0.4749±0.0050 | 0.5119±0.0053 | 0.5254±0.0062 |
| iDocNADE | 0.4917±0.0278 | 0.5188±0.0096 | 0.5267±0.0079 | 0.5267±0.0074 | 0.5279±0.0043 | 0.7202±0.0059 | 0.7032±0.0058 | 0.6807±0.0091 | 0.6687±0.0072 | 0.6558±0.0055 |
| iDocNADEe | - | - | 0.3998±0.0410 | - | 0.4480±0.0515 | - | - | 0.6642±0.0101 | - | 0.6702±0.0030 |
| TopicRNN | 0.4508±0.0207 | 0.4708±0.0141 | 0.4914±0.0055 | 0.5248±0.0044 | 0.5448±0.0122 | 0.6723±0.0136 | 0.6641±0.0097 | 0.6659±0.0079 | 0.6697±0.0029 | 0.6768±0.0053 |
| GNTM | 0.6635±0.0139 | 0.6253±0.0049 | 0.5870±0.0061 | 0.5693±0.0086 | 0.5386±0.0101 | 0.3606±0.0155 | 0.3765±0.0085 | 0.4009±0.0066 | 0.4046±0.0114 | 0.3972±0.0105 |
| | BNC | | | | | Reuters | | | | |
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.5305±0.0080 | 0.5557±0.0071 | 0.5639±0.0040 | 0.5581±0.0062 | 0.5543±0.0146 | 0.4979±0.0099 | 0.4809±0.0219 | 0.4635±0.0159 | 0.4466±0.0068 | 0.4327±0.0092 |
| GSM | 0.5351±0.0096 | 0.5525±0.0076 | 0.5487±0.0080 | 0.5135±0.0125 | 0.4611±0.0049 | 0.5363±0.0111 | 0.5031±0.0073 | 0.4547±0.0080 | 0.4133±0.0234 | 0.3822±0.0081 |
| ProdLDA | 0.5403±0.0365 | 0.5163±0.0096 | 0.4891±0.0185 | 0.4632±0.0115 | 0.4504±0.0084 | 0.4744±0.0069 | 0.4594±0.0085 | 0.4628±0.0094 | 0.4424±0.0130 | 0.4397±0.0102 |
| ETM | 0.5707±0.0101 | 0.5669±0.0108 | 0.5535±0.0081 | 0.5314±0.0083 | 0.5103±0.0073 | 0.5073±0.0107 | 0.4749±0.0061 | 0.4678±0.0066 | 0.4548±0.0066 | 0.4538±0.0038 |
| GraphBTM | 0.3955±0.0163 | 0.3314±0.0462 | 0.3985±0.0521 | 0.4428±0.0678 | 0.2995±0.0021 | 0.4096±0.0209 | 0.4160±0.0244 | 0.4290±0.0143 | 0.4270±0.0237 | 0.3786±0.0509 |
| iDocNADE | 0.3866±0.0171 | 0.3819±0.0124 | 0.3822±0.0093 | 0.4054±0.0172 | 0.4092±0.0181 | 0.5001±0.0087 | 0.5136±0.0141 | 0.5222±0.0119 | 0.5368±0.0068 | 0.5510±0.0060 |
| iDocNADEe | - | - | 0.2846±0.0068 | - | 0.3162±0.0043 | - | - | 0.4480±0.0078 | - | 0.5090±0.0052 |
| TopicRNN | 0.3727±0.0190 | 0.3510±0.0174 | 0.3485±0.0133 | 0.3546±0.0094 | 0.3715±0.0083 | 0.4826±0.0090 | 0.5239±0.0070 | 0.5609±0.0072 | 0.5898±0.0070 | 0.6113±0.0092 |
| GNTM | 0.5889±0.0087 | 0.5917±0.0070 | 0.5724±0.0038 | 0.5683±0.0074 | 0.5517±0.0045 | 0.5220±0.0140 | 0.4920±0.0188 | 0.4825±0.0104 | 0.4569±0.0073 | 0.4513±0.0129 |

Table S15: The performance of topic diversity with top 20 words.

| | 20NG | | | | | TMN | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.7688±0.0152 | 0.7810±0.0226 | 0.7736±0.0099 | 0.7698±0.0051 | 0.7648±0.0092 | 0.9784±0.0045 | 0.9784±0.0059 | 0.9842±0.0013 | 0.9866±0.0010 | 0.5504±0.0150 |
| GSM | 0.6696±0.0170 | 0.6372±0.0282 | 0.5770±0.0110 | 0.4586±0.0214 | 0.2862±0.0131 | 0.9036±0.0087 | 0.8568±0.0168 | 0.5454±0.0301 | 0.2374±0.0128 | 0.0520±0.0089 |
| ProdLDA | 0.8988±0.0118 | 0.8054±0.0119 | 0.7070±0.0083 | 0.6388±0.0160 | 0.5578±0.0050 | 0.7146±0.0284 | 0.5320±0.0373 | 0.5022±0.0675 | 0.4988±0.0598 | 0.4970±0.0459 |
| ETM | 0.8142±0.0321 | 0.7416±0.0113 | 0.6316±0.0111 | 0.5622±0.0098 | 0.4966±0.0072 | 0.9332±0.0253 | 0.8742±0.0112 | 0.7020±0.0123 | 0.6124±0.0096 | 0.5090±0.0155 |
| GraphBTM | 0.2723±0.0166 | 0.2118±0.0408 | 0.2312±0.0171 | 0.2143±0.0545 | 0.3404±0.1057 | 0.4523±0.0797 | 0.5217±0.0156 | 0.6180±0.0064 | 0.6235±0.0075 | 0.5622±0.0075 |
| iDocNADE | 0.7272±0.0306 | 0.6706±0.0122 | 0.6376±0.0121 | 0.5982±0.0154 | 0.5638±0.0117 | 0.5842±0.0149 | 0.5060±0.0129 | 0.3966±0.0096 | 0.3264±0.0029 | 0.2668±0.0061 |
| iDocNADEe | - | - | 0.5274±0.0417 | - | 0.5098±0.0443 | - | - | 0.3552±0.0100 | - | 0.2922±0.0039 |
| TopicRNN | 0.9342±0.0117 | 0.9228±0.0063 | 0.9182±0.0079 | 0.9042±0.0016 | 0.8686±0.0069 | 0.7248±0.0399 | 0.6748±0.0330 | 0.5914±0.0174 | 0.5586±0.0153 | 0.5264±0.0199 |
| GNTM | 0.8468±0.0099 | 0.7991±0.0262 | 0.7394±0.0199 | 0.6169±0.0307 | 0.4632±0.0125 | 0.9962±0.0026 | 0.9926±0.0024 | 0.9814±0.0099 | 0.9512±0.0122 | 0.8278±0.0720 |

| | BNC | | | | | Reuters | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.7222±0.0314 | 0.7234±0.0324 | 0.6932±0.0284 | 0.7028±0.0183 | 0.7034±0.0126 | 0.6748±0.0143 | 0.6666±0.0217 | 0.6394±0.0060 | 0.6286±0.0123 | 0.6284±0.0068 |
| GSM | 0.7076±0.0056 | 0.6220±0.0122 | 0.5506±0.0048 | 0.4620±0.0266 | 0.3114±0.0118 | 0.5867±0.0152 | 0.5253±0.0101 | 0.4564±0.0130 | 0.3243±0.0256 | 0.2027±0.0136 |
| ProdLDA | 0.8458±0.0362 | 0.7530±0.0108 | 0.6506±0.0205 | 0.5736±0.0107 | 0.4618±0.0136 | 0.8888±0.0085 | 0.8268±0.0113 | 0.7058±0.0078 | 0.6076±0.0137 | 0.5062±0.0120 |
| ETM | 0.8497±0.0074 | 0.8093±0.0088 | 0.7264±0.0056 | 0.6526±0.0067 | 0.5799±0.0123 | 0.8560±0.0080 | 0.7686±0.0117 | 0.6634±0.0077 | 0.5696±0.0057 | 0.4640±0.0092 |
| GraphBTM | 0.2019±0.0346 | 0.1431±0.0514 | 0.1664±0.0171 | 0.1616±0.0326 | 0.1301±0.0200 | 0.2223±0.0197 | 0.1797±0.0081 | 0.1716±0.0275 | 0.1779±0.0146 | 0.2210±0.0248 |
| iDocNADE | 0.5764±0.0570 | 0.5698±0.0460 | 0.5824±0.0240 | 0.5910±0.0359 | 0.5660±0.0322 | 0.5118±0.0204 | 0.4772±0.0128 | 0.4144±0.0149 | 0.3950±0.0070 | 0.3764±0.0039 |
| iDocNADEe | - | - | 0.6106±0.0172 | - | 0.5720±0.0128 | - | - | 0.4320±0.0099 | - | 0.3850±0.0088 |
| TopicRNN | 0.9452±0.0103 | 0.9270±0.0129 | 0.9152±0.0043 | 0.9054±0.0059 | 0.8910±0.0066 | 0.9278±0.0108 | 0.9198±0.0083 | 0.9052±0.0138 | 0.8786±0.0079 | 0.8532±0.0059 |
| GNTM | 0.8724±0.0411 | 0.8782±0.0173 | 0.7874±0.0252 | 0.6450±0.0717 | 0.5526±0.0496 | 0.7402±0.0240 | 0.6676±0.0203 | 0.5892±0.0079 | 0.5442±0.0137 | 0.4918±0.0110 |

Table S16: The performance of the document cluster on 20NG.

| | Purity | | | | | NMI | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.2980±0.0131 | 0.3340±0.0106 | 0.3375±0.0086 | 0.3510±0.0098 | 0.3740±0.0138 | 0.2908±0.0097 | 0.3013±0.0064 | 0.2852±0.0033 | 0.2878±0.0085 | 0.2858±0.0092 |
| GSM | 0.4133±0.0268 | 0.4379±0.0133 | 0.4629±0.0180 | 0.4429±0.0137 | 0.4210±0.0290 | 0.4394±0.0083 | 0.4369±0.0158 | 0.4433±0.0059 | 0.4449±0.0064 | 0.4412±0.0202 |
| ProdLDA | 0.3306±0.0168 | 0.3450±0.0221 | 0.3641±0.0108 | 0.3638±0.0210 | 0.3807±0.0244 | 0.3405±0.0178 | 0.3345±0.0133 | 0.3350±0.0035 | 0.3298±0.0090 | 0.3343±0.0170 |
| ETM | 0.3496±0.0173 | 0.4154±0.0134 | 0.4380±0.0136 | 0.4510±0.0174 | 0.4616±0.0079 | 0.3842±0.0097 | 0.4227±0.0063 | 0.4296±0.0017 | 0.4297±0.0090 | 0.4356±0.0033 |
| GraphBTM | 0.1448±0.0084 | 0.1210±0.0454 | 0.1630±0.0125 | 0.1068±0.0609 | 0.0992±0.0497 | 0.1552±0.0185 | 0.1108±0.0825 | 0.1807±0.0135 | 0.0816±0.0956 | 0.0707±0.0848 |
| iDocNADE | 0.2175±0.1087 | 0.2844±0.0796 | 0.3064±0.0758 | 0.3187±0.0758 | 0.3371±0.0756 | 0.1128±0.0444 | 0.1723±0.0771 | 0.1802±0.0728 | 0.1901±0.0737 | 0.1990±0.0684 |
| iDocNADEe | - | - | 0.1300±0.0005 | - | 0.1507±0.0035 | - | - | 0.1185±0.0016 | - | 0.1228±0.0017 |
| TopicRNN | 0.0728±0.0022 | 0.0761±0.0020 | 0.0865±0.0022 | 0.0920±0.0010 | 0.1001±0.0002 | 0.0109±0.0007 | 0.0141±0.0006 | 0.0229±0.0006 | 0.0303±0.0011 | 0.0400±0.0004 |
| GNTM | 0.4500±0.0222 | 0.4882±0.0328 | 0.5089±0.0136 | 0.5090±0.0156 | 0.5021±0.0241 | 0.4436±0.0084 | 0.4419±0.0102 | 0.4416±0.0060 | 0.4362±0.0084 | 0.4371±0.0044 |

Table S17: The performance of the document cluster on TMN.

| | Purity | | | | | NMI | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 30 | 50 | 70 | 100 | 20 | 30 | 50 | 70 | 100 |
| LDA | 0.3509±0.0093 | 0.3692±0.0026 | 0.3725±0.0072 | 0.4031±0.0033 | 0.4228±0.0108 | 0.0622±0.0055 | 0.0665±0.0016 | 0.0754±0.0040 | 0.0901±0.0016 | 0.1064±0.0069 |
| GSM | 0.5933±0.0170 | 0.6054±0.0182 | 0.6184±0.0096 | 0.5934±0.0233 | 0.2632±0.0205 | 0.2848±0.0127 | 0.2787±0.0138 | 0.2996±0.0099 | 0.3246±0.0084 | 0.0204±0.0149 |
| ProdLDA | 0.3141±0.0159 | 0.2808±0.0208 | 0.2505±0.0065 | 0.2535±0.0194 | 0.2438±0.0000 | 0.0508±0.0198 | 0.0334±0.0108 | 0.0056±0.0047 | 0.0053±0.0107 | 0.0000±0.0000 |
| ETM | 0.5841±0.0160 | 0.5967±0.0109 | 0.6347±0.0141 | 0.6358±0.0065 | 0.6420±0.0086 | 0.2764±0.0095 | 0.2705±0.0059 | 0.2829±0.0088 | 0.2784±0.0052 | 0.2767±0.0066 |
| GraphBTM | 0.2438±0.0000 | 0.2438±0.0000 | 0.2438±0.0000 | 0.2438±0.0000 | 0.2438±0.0000 | 0.0000±0.0000 | 0.0000±0.0000 | 0.0000±0.0000 | 0.0000±0.0000 | 0.0000±0.0000 |
| iDocNADE | 0.2712±0.0962 | 0.3116±0.0766 | 0.3314±0.0766 | 0.3512±0.0751 | 0.3730±0.0727 | 0.1074±0.0643 | 0.1436±0.0728 | 0.1488±0.0681 | 0.1585±0.0654 | 0.1701±0.0647 |
| iDocNADEe | - | - | 0.2623±0.0094 | - | 0.4171±0.0067 | - | - | 0.0332±0.0054 | - | 0.1728±0.0022 |
| TopicRNN | 0.2493±0.0014 | 0.2537±0.0007 | 0.2586±0.0017 | 0.2684±0.0024 | 0.2775±0.0053 | 0.0048±0.0006 | 0.0071±0.0005 | 0.0108±0.0014 | 0.0154±0.0004 | 0.0221±0.0014 |
| GNTM | 0.6150±0.0161 | 0.6252±0.0140 | 0.6472±0.0040 | 0.6656±0.0054 | 0.6773±0.0069 | 0.2851±0.0067 | 0.2801±0.0099 | 0.2789±0.0039 | 0.2821±0.0037 | 0.2870±0.0043 |

## A.8 More Examples for Learned Topics

Figure S2-S4 show the two randomly selected learned topics on TMN, BNC, and Reuters datasets, respectively, with $K = 20$. We report the top 8 words, and top 8 edges and corresponding topic graph representation in the top 30 words, where the sizes of nodes and edges are positively correlated to $\beta_k^v$ and $A_k = (\beta_k^v \cdot (\beta_k^v)^T) \odot \beta_k^e$, respectively. Nodes are colored by their cluster labels detected by fast unfloding algorithm [3], where isolated nodes are coloured in gray.

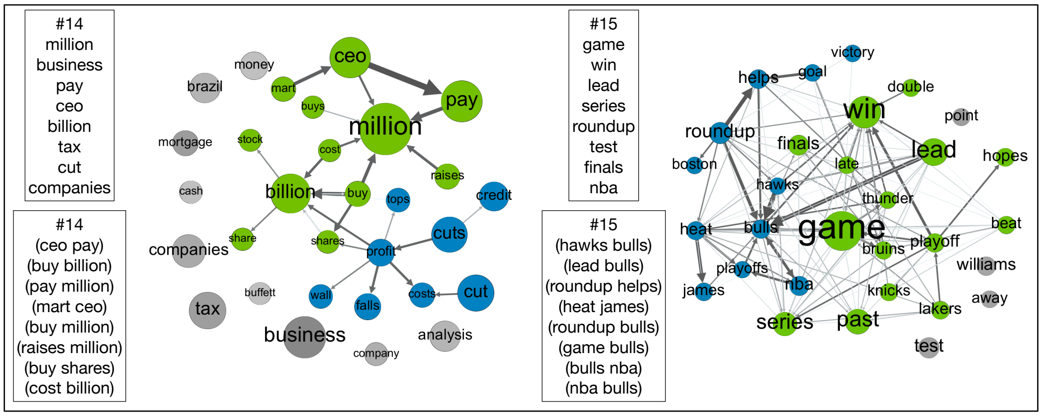



Figure S2: The illustration of learned topics on TMN dataset.

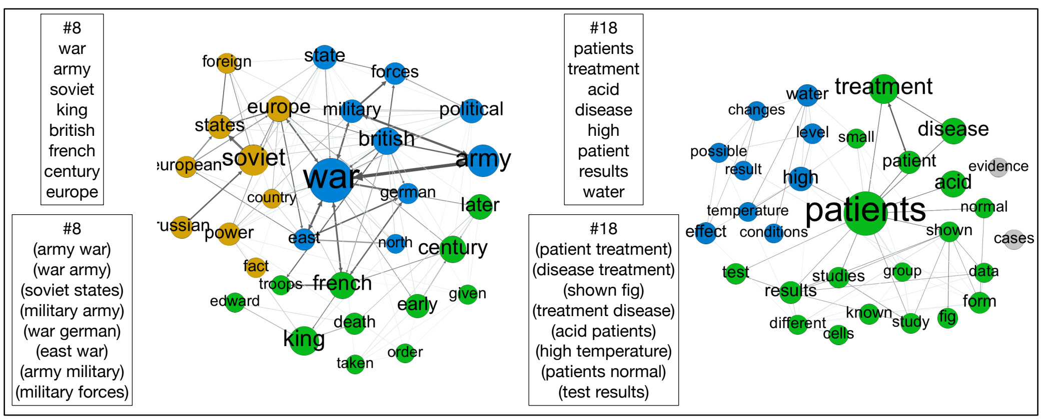



Figure S3: The illustration of learned topics on BNC dataset.

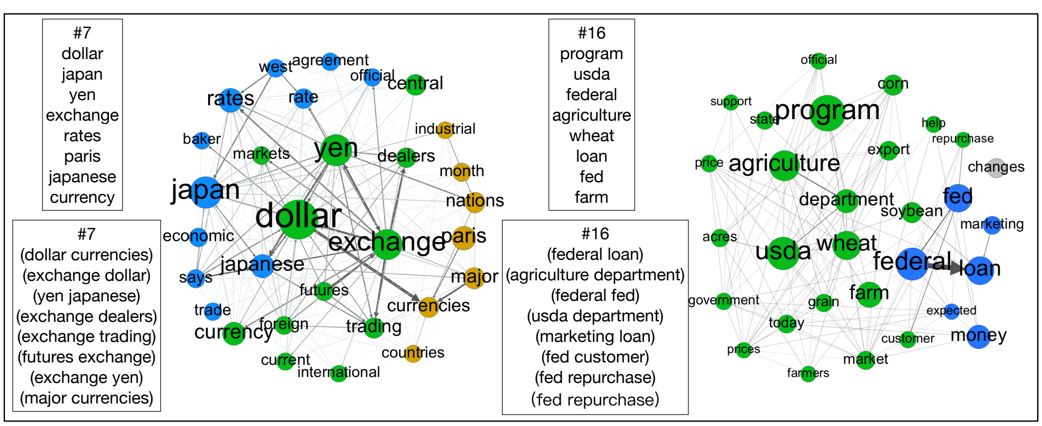



Figure S4: The illustration of learned topics on Reuters dataset.

## References


[1] David M Blei, John D Lafferty, et al. A correlated topic model of science. *The annals of applied statistics*, 1(1):17–35, 2007.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[3] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[4] BNC Consortium et al. British national corpus. *Oxford Text Archive Core Collection*, 2007.

[5] Rajarshi Das, Manzil Zaheer, and Chris Dyer. Gaussian lda for topic models with word embeddings. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 795–804, 2015.

[6] Adji B Dieng, Francisco JR Ruiz, and David M Blei. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453, 2020.

[7] Adji B Dieng, Chong Wang, Jianfeng Gao, and John Paisley. Topicrnn: A recurrent neural network with long-range semantic dependency. *arXiv preprint arXiv:1611.01702*, 2016.

[8] Pankaj Gupta, Yatin Chaudhary, Florian Buettner, and Hinrich Schütze. Document informed neural autoregressive topic models with distributional prior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6505–6512, 2019.

[9] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

[10] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

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

[12] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[13] Ken Lang. Newsweeder: Learning to filter netnews. In *Machine Learning Proceedings 1995*, pages 331–339. Elsevier, 1995.

[14] Jey Han Lau, David Newman, and Timothy Baldwin. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 530–539, 2014.

[15] Zhixing Li, Siqiang Wen, Juanzi Li, Peng Zhang, and Jie Tang. On modelling non-linear topical dependencies. In *International Conference on Machine Learning*, pages 458–466. PMLR, 2014.

[16] Yishu Miao, Edward Grefenstette, and Phil Blunsom. Discovering discrete latent topics with neural variational inference. In *International Conference on Machine Learning*, pages 2410–2419. PMLR, 2017.

[17] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[18] Akash Srivastava and Charles Sutton. Autoencoding variational inference for topic models. *arXiv preprint arXiv:1703.01488*, 2017.

[19] Daniele Vitale, Paolo Ferragina, and Ugo Scaiella. Classification of short texts by deploying topical annotations. In *European Conference on Information Retrieval*, pages 376–387. Springer, 2012.

[20] Chong Wang and David M Blei. Variational inference in nonconjugate models. *Journal of Machine Learning Research*, 14(4), 2013.

[21] Qile Zhu, Zheng Feng, and Xiaolin Li. Graphbtm: Graph enhanced autoencoded variational inference for biterm topic model. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 4663–4672, 2018.