First of all, we thank all the reviewers for their valuable comments and suggestions.

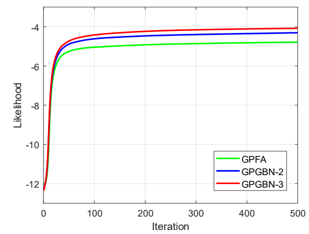

Figure 1: Graph likelihood of GPGBNs as a function of iteration with various network depths.

To Reviewer 1: $\boldsymbol{i})$ For SIG-VAE, the memory cost is theoretically $(K+J)$ times larger than the most basic VGAE, where K and J represents the sampling numbers of SIVI [32] in each iteration. With the official release, SIG-VAE takes nearly 0.7G ($K=1, J=1$), 4.4G ($K=5, J=10$), and 10.6G ($K=15, J=20$) RAM cost on Cora (the smallest dataset with 2708 nodes). Following this trend, SIG-VAE is estimated to take at least 600G RAM on Cora with their original setting ($K=150, J=2000$), which is a normally unaffordable memory. By contrast, a 3-layer WGCAE, which takes only 1.3G RAM, has achieved a comparable link prediction performance and outperformed SIG-VAE on both node clustering and classification tasks, showing the proposed WGCAE is more efficient than SIG-VAE. $\boldsymbol{ii})$ Thanks for pointing out these references, we will carefully investigate them and comprehensively discuss their relations to our work. We will fix the listed typos in our revision. $\boldsymbol{iii})$ For an intuitive quantitative comparison of network modeling, we estimate three GPGBNs with different depths $T \in \{1,2,3\}$ on 20news dataset and exhibit the likelihood of adjacency matrix as a function of iterations. As shown in Fig. 1 here, increasing the network depth in general improves the quality of adjacency matrix fitting, showing the benefit of capturing document relations with a hierarchical structure. Moreover, we will provide more visualized GPGBNs with different depths in our revision.

To Reviewers 2&4 (Novelty): As claimed in our contributions, we propose the first hierarchical relational topic model (RTM) named GPGBN, and successfully illustrate the connections at different semantic levels. Moreover, our work provides a novel solution to combine the RTM and graph autoencoders, firstly adopting the GCN to estimate the posteriors of the latent representations of RTMs (note related theoretical proofs [44] have only recently been proposed). Moving beyond deterministic projecting, the uncertainty and sparseness provided by Weibull reparameterization effectively alleviate overfitting of GCN and further improve the performance in a hierarchical fashion.

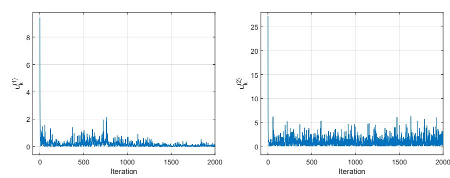

Figure 2: Gibbs sampling of variables selected from $\boldsymbol{u}^{(1)}$ and $\boldsymbol{u}^{(2)}$ as a function of iteration.

To Reviewer 2: $\boldsymbol{i})$ Gibbs sampling is applicable when there exist local conjugacies for latent variables, whose conditional distributions will then become tractable and simple to sample from, even though the posterior of the joint distribution of these variables is often intractable. Gibbs sampling uses a Markov chain to sample the latent variables in turn to iteratively approach the true posteriors. In Fig. 2, we show the trace plot of a random dimension of $\boldsymbol{u}^{(1)}$ and that of $\boldsymbol{u}^{(2)}$ from a 3-layer GPGBN, suggesting the Markov chain under the proposed Gibbs sampler converges faster and mixes well. $\boldsymbol{ii})$ That is a nice idea. We note a full matrix for U could provide extra flexibility to model stochastic equivalence/disassortativity (e.g., in protein-protein interaction network), while a diagonal one is more suitable to model an assortative relational network exhibiting homophily (e.g., co-author network) but not necessarily stochastic equivalence. Similar conclusion can be found in [34] and we will add a discussion in Appendix. $\boldsymbol{iii})$ Actually, the Weibull inference network only approximates the posterior of latent document representation $\boldsymbol{\theta}$ and can't directly improve the performance. However, moving beyond treating the importance of document content and relations equally like GPGBN, WGCAE is a VAE-like model that can be trained via optimizing the loss function, where we can introduce a trade-off parameter β to control the focus of the model. By adjusting β, WGCAE can provide more expressive latent representations for down-stream tasks and we have discussed this phenomenon in Section 5.3.

Table 1: Topic-coherence comparisons on 20news.

Topic layers	hardware	christian	guns	space	graphics
LDA [11]	0.530	0.561	0.491	0.538	0.564
PFA [33]	0.494	0.560	0.483	0.520	0.555
AVITM [45]	0.434	0.495	0.422	0.451	0.483
DPFA [19]	0.581	0.604	0.535	0.562	0.575
PGBN [20]	0.607	0.615	0.550	0.578	0.583
GPGBN	**0.638**	**0.641**	**0.602**	**0.623**	**0.613**

To Reviewer 3: We clarify that the proposed WGCAE is a basic VGAE-like model, which has a significant improvement compared to the original VGAE on various graph analytics tasks. Other relevant improvement techniques, such as SIVI [32] and GAT [27], can potentially be incorporated into our models to further improve the model performance; we leave these further extensions for future study.

To Reviewer 4: $\boldsymbol{i})$ We'd like to emphasize that we have compared with many GCN-based methods in our experiments, including node clustering (2nd block of Table 1), link prediction (Table 2), and node classification (Table 4 in Appendix). As far as we know, VGAE could be the most popular GCN-based method for network modeling and other variants like SIG-VAE, S-VGAE, and NF-VGAE have all been included in our comparison. We are also glad to compare with other VGAEs (if any) for network modeling. $\boldsymbol{ii})$ Thanks for your suggestions, we have included topic-coherence comparison between hierarchical topics learned by PGBN and GPGBN in Appendix, showing that the words among the topics learned by GPGBN are more relevant (or co-occurrence) than those learned by PGBN. Moreover, we also add additional topic-model baselines including LDA, PFA, AVITM [45], and DPFA for topic-coherence comparisons as shown in Table. 1, indicating the benefit of introducing hierarchical graph regularization. We will put these results in our revision.

[44] Zhao L, Akoglu L. Connecting Graph Convolutional Networks and Graph-Regularized PCA. In ICML, 2020.
[45] Srivastava, A. and Sutton, C. Autoencoding variational inference for topic models. In ICLR, 2017.