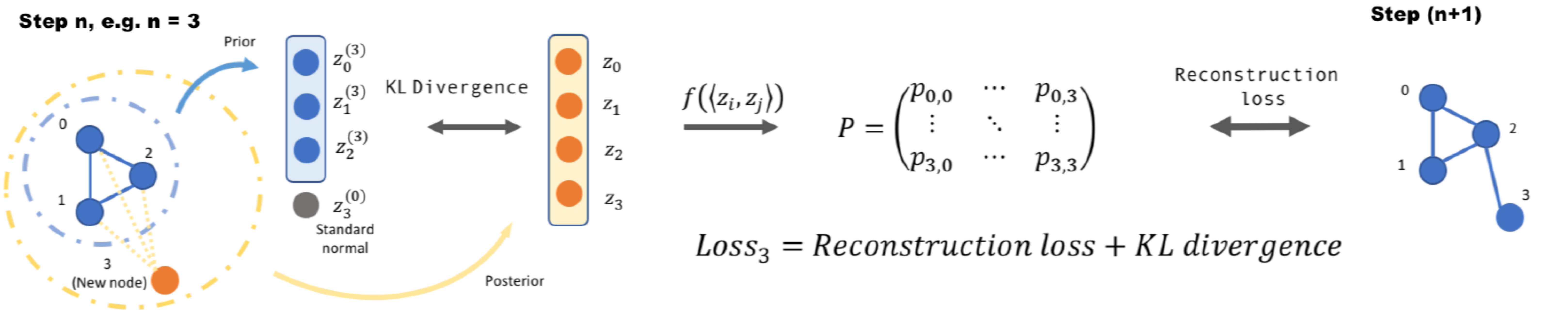
Generative Graph Convolutional Network for Growing Graphs

Da Xu, Chuanwei Ruan, Kamiya Motwani, Evren Korpeoglu, Sushant Kumar, Kannan Achan

Walmart Labs, Sunnyvale, California, USA

Contribution Highlights

- Proposed a unified generative graph convolutional network to model the growing graphs.
- The framework models the underlying graph generation sequence considering not only the topological information, but also the node attributes information.
- Learns the vector representations for both observed nodes in the graph and new isolated nodes that might potentially link to the graph, where-as most existing methods rely on fixed graphs.
- The proposed model shows superior results on several benchmark data sets on link prediction tasks.



Background and Applications

Representation on graphs

Learning the vector representations for nodes in the graph which makes the topological information contained in the graph accessible by down stream tasks.

- Use cases includes but not limited to **recommender systems**, outlier detection and community detection.
- Personalizing recommendations for a new user based on other users seen so far.
- Applications in identifying optimal audience for targeting based on the social graph.

The rise of growing graphs

Existing methods are great, but they only work for fixed graphs. In many applications, the graphs are not fixed and keep growing, example :

- New items / New users / New campaigns

References

- Thomas N Kipf and Max Welling,"Variational graph auto-encoders", *arXiv preprint arXiv:1611.107308*, 2016.
- Jainxuan You, Rex Ying, Xiang Ren, William L Hamilton and Jure Leskovec,"Graphrnn: A deep generative model for graphs", *arXiv preprint arXiv:1802.03480*, 2018.
- Diederik P Kinema and Max Welling,"Auto-encoding variation bayes", *arXiv preprint arXiv:1312.6114*, 2013.

Proposed Method

- Addresses the challenge of generating graph structure for growing graphs with new nodes that are unconnected to the previous observed graph.
- Major assumption : underlying generating mechanism is stationary during growth.
- We learn the sequential generation of graph structures, for cases where both node attributes and topological information exist as well as for cases where only node attributes are available.

Proposed Approach

- When a new node is added, we treat it as connected to all of previous nodes with the same probability $\tilde{p}$, where $\tilde{p}$ may reflect the overall sparsity of the graph.
- When formulating encoding distribution, due to the efficiency of *GCN* in node classification and linkage prediction, we adopt their convolutional layers.
- We use what the model has informed us till the i^{th} step in an adaptive way by treating $\boldsymbol{z}^{i+1} \in \mathbb{R}^{(i+1)\times d_2}$ as $[\boldsymbol{z}^{i+1}_{1:i}, \boldsymbol{z}^{i+1}_{i+1}]$
- Hidden factors for previous nodes use the encoding distribution where the candidate adjacency matrix $\tilde{\boldsymbol{A}}_{\leq i}$ passes information from previous steps.

Adaptive Evidence Lower Bound

$$-\sum_{i=1}^{n-1} E_{q_\phi^i(\boldsymbol{z}^{i+1}|\tilde{\boldsymbol{A}}_{\leq i+1}, \boldsymbol{X}_{\leq i+1})}[\log p_\theta(\boldsymbol{A}_{\leq i}|\boldsymbol{z}^i)]$$
$$+\beta \sum_{i=1}^{n-1} KL(q_\phi^i(\boldsymbol{z}^{i+1}|\tilde{\boldsymbol{A}}_{\leq i+1}, \boldsymbol{X}_{\leq i+1}) \| p_0^i(\boldsymbol{z}^{i+1}|\boldsymbol{A}_{\leq i}, \boldsymbol{X}_{\leq i})).$$

- First term : Reconstruction loss in each generation step.
- Second term : Adaptive regularizer that enforces the posterior of latent factors for observed nodes to remain close to their priors which contain information from previous steps.

Experiment Results

Method	Cora		Citeseer		Pubmed	
	AUC	AP	AUC	AP	AUC	AP
Isolated new nodes						
GCN-VAE	75.1	76.3	79.4	82.1	85.5	85.4
MLP-VAE	75.6	75.6	81.8	83.7	77.1	77.2
G-GCN	**83.3**	**85.0**	**89.5**	**91.3**	**87.5**	**87.2**
Nodes in observed graph						
GCN-VAE	93.1	94.4	93.3	94.4	96.7	96.9
MLP-VAE	86.5	87.2	87.1	89.3	79.4	79.5
G-GCN	**94.1**	**95.2**	**94.6**	**95.9**	**96.9**	**97.3**

Conclusion

- Generative graph convolution model for growing graphs that incorporates graph representation learning and graph convolutional network into a sequential generative model.
- Our approach outperforms both baselines in new node link prediction :
  - Comparison to *MLP-VAE* shows our advantage of learning with topological information.
  - Better performance over *GCN-VAE* indicates the importance of modeling the sequential generating process when making predictions on new nodes.
- Our approach has comparable or even slightly better results than *GCN-VAE* on link prediction task in observed graph : our superior performance on isolated new nodes is not at the cost of the performance on nodes in observed graph.