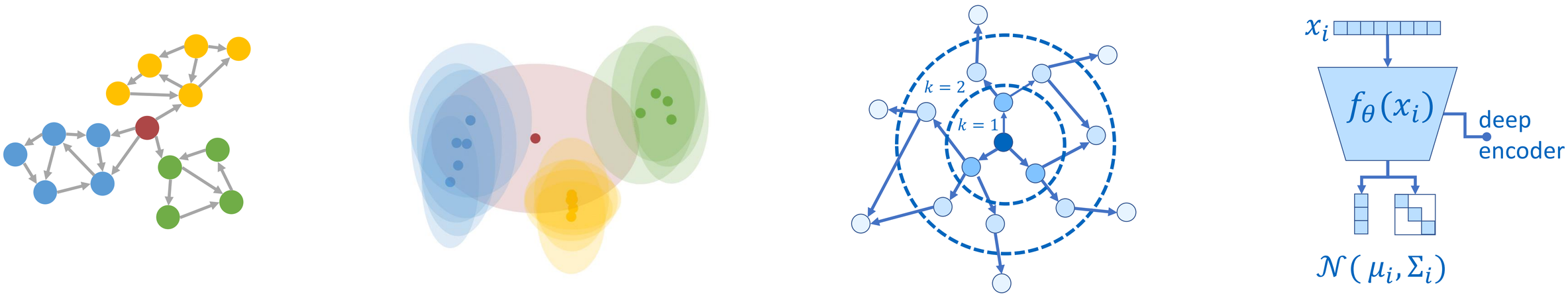
Professorship of Data Mining and Analytics
Department of Informatics
Technical University of Munich

Deep Gaussian Embedding of Graphs: Unsupervised Inductive Learning via Ranking

Aleksandar Bojchevski, Stephan Günnemann

Graph2Gauss - 3 key modeling ideas

1. Uncertainty
2. Personalized ranking
3. Inductiveness



1. Uncertainty: embedding nodes as Gaussian distributions captures uncertainty
2. Personalized ranking: for node i, nodes in k-hop neighborhood should be closer to i than nodes in $(k+1)$-hop neighborhood
3. Inductiveness: generalizes to unseen nodes by learning a mapping from node features to embeddings

Efficient optimization

❑ Personalized ranking implies the following constraints:

$$\mathrm{D}_{KL}(\mathcal{N}_j||\mathcal{N}_i) < \mathrm{D}_{KL}(\mathcal{N}_{j'}||\mathcal{N}_i) \quad \forall i, \forall j \in N_i^{(k)}, \forall j' \in N_i^{(k')}, \forall k < k'$$

D_{KL}: naturally handles directed graphs; $N_i^{(k)}$: set of nodes in the k-hop neighborhood of node i

❑ Learning via energy-based loss
- closer nodes should have lower energy
- naive optimization: $O(N^3)$ complexity

$$\mathcal{L} = \sum_{(i,j,j')} (E_{ij}^2 + \exp^{-E_{ij'}})$$

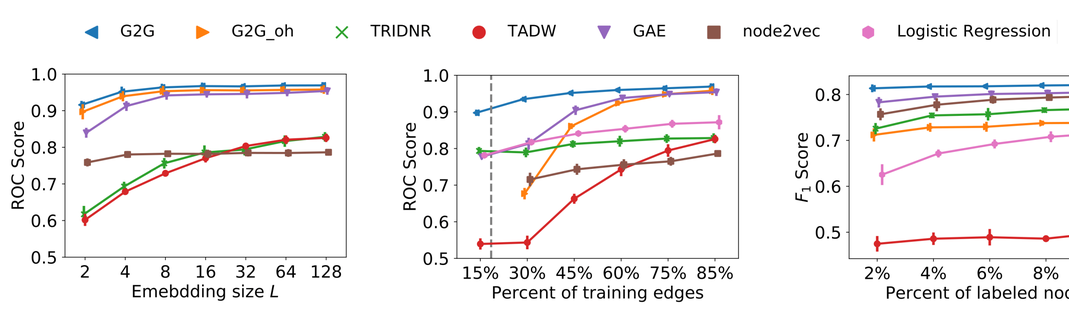
$$E_{ij} = \mathrm{D}_{KL}(N_j||N_i)$$

❑ Node-anchored sampling strategy

Sample $\forall i, (j_1, \dots, j_K) \sim (N_i^{(1)}, \dots, N_i^{(K)})$ and optimize over implied constraints

- only few triplets seen ($< 4.2\%$) to match performance
- lower gradient variance compared to uniform random sampling
- our optimization: $O(N)$ complexity

Embedding quality

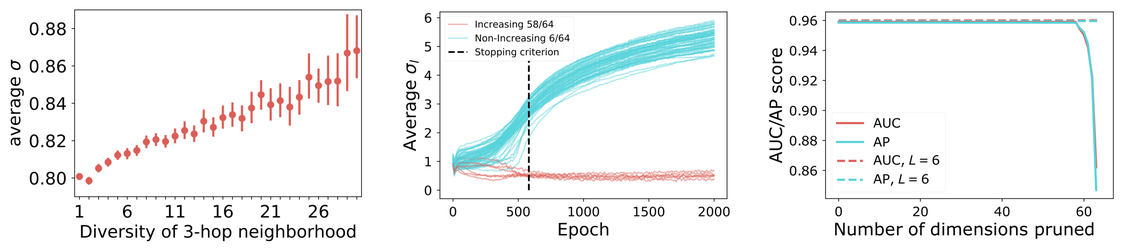


❑ Graph2Gauss shows strong performance for both link prediction and node classification tasks

❑ Strong performance even when using only the network structure

❑ Graph2Gauss is parameter and data efficient
- large performance gap for both small embedding size
- and in the sparse training setting (e.g. 15% edges)

Embedding uncertainty



❑ Uncertainty correlates with neighborhood diversity
- Diversity is number of distinct classes in a node's k-hop neighborhood

❑ Uncertainty reveals the intrinsic latent dimensionality of the graph
- Detected latent dimensions ≈ number ground-truth communities

❑ We can prune dimensions with high average uncertainty
- Without a decrease in link prediction performance

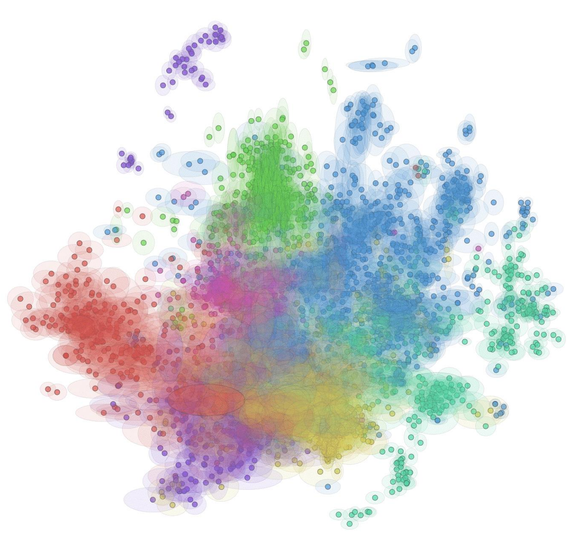
Inductiveness

❑ G2G is truly inductive
- After training only needs attributes
- Able to embed nodes w/o edges

❑ Maintains strong performance
- Even for ¼ of the nodes hidden

Dataset	Log. Reg. 10%	G2G 10%	G2G 25%
Cora-ML	75.95	90.93	87.83
Cora	78.53	94.18	92.96
Citeseer	73.09	88.58	87.30
DBLP	67.55	85.06	83.09
Pubmed	86.83	92.22	90.20

Visualization



2D Embedding of Cora-ML

ICLR 2018

April 30 – May 03, 2018
Vancouver, BC – Canada

github.com/abojchevski/graph2gauss