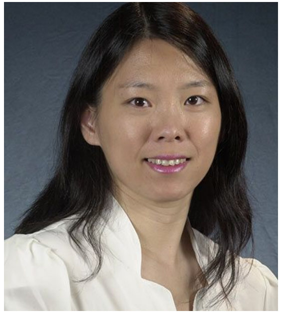
NEURAL INFORMATION PROCESSING SYSTEMS

UCLA Samueli Computer Science

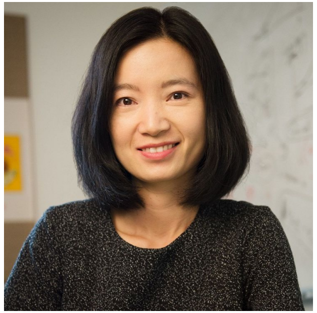
Non-Euclidean Mixture Model for Social Network Embedding

Roshni G. Iyer

Yewen Wang



Wei Wang



Yizhou Sun

Introduction and Preliminaries

- **Social networks** are omnipresent because they model interactions on social platforms
- Social network analysis is key to community detection, user connectivity etc.
- Widely agreed that social network links are formed from **homophily** or **social influence**
- **Homophily:** associated nodes imply feature similarity (form cycles)
- **Social Influence:** popular nodes have direct influence in forming links (form hierarchies)

Social Network Embedding Models

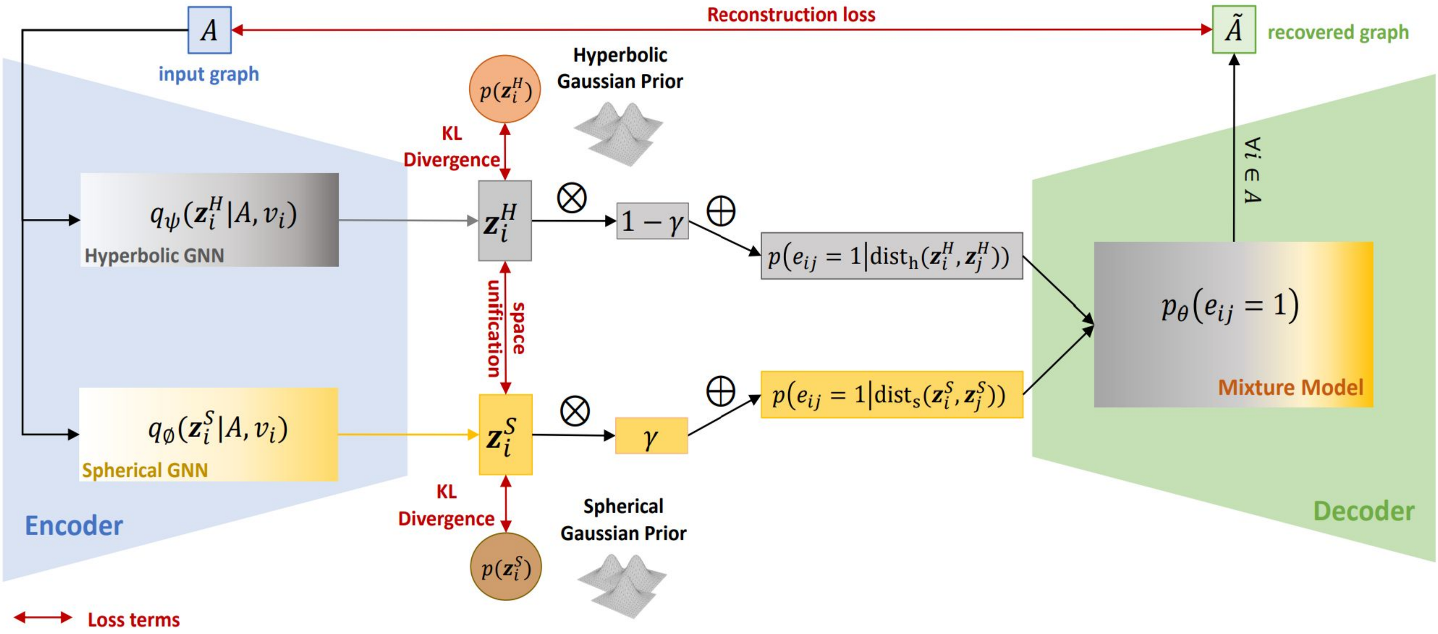
Observations:

- Shallow embedding models (structural embedding) do not effectively learn graph structure (limited to attributes)
- Models do not capture all social network factors e.g., social influence
- All network structures are modeled in same space e.g., flat Euclidean space

Category	Description
Structural Embedding Models	**GraRep** [Cao *et al.*, 2015], shallow embedding integrating global structural information **RolX** [Henderson *et al.*, 2012], unsupervised learning approach using structural role based similarity **GraphWave** [Donnat *et al.*, 2018], shallow embedding model using spectral graph wavelet diffusion patterns
GNN Embedding Models (Euclidean space)	**GraphSAGE** [Hamilton *et al.*, 2017], inductive framework using node features and neighbor aggregation **GCN** [Kipf and Welling, 2017], semi-supervised learning model via graph convolution on local neighborhoods **GAT** [Veličković *et al.*, 2018], graph attention model using mask self-attention layers on local neighborhoods
Homophily-based Embedding Models	**GELTOR** [Hamedani *et al.*, 2023], embedding method using learning-to-rank with AdaSim* similarity metric **NRP** [Yang *et al.*, 2020], embedding model using pairwise personalized PageRank on the global graph
GNN Embedding Models (non-Euclidean space)	**HGCN** [Chami *et al.*, 2019], hyperbolic GCN model utilizing Riemannian geometry and hyperboloid model **κ-GCN** [Bachmann *et al.*, 2020], GCN model using product space e.g., product of constant curvature spaces
Mixture Models (homophily and social influence)	**RaRE** [Gu *et al.*, 2018], Bayesian probabilistic model for node proximity/popularity via posterior estimation **NMM**, our non-Euclidean mixture model (see Eqn. 9), without use of GraphVAE framework **NMM-GNN** , our non-Euclidean mixture model with non-Euclidean GraphVAE framework

SOTA baseline models

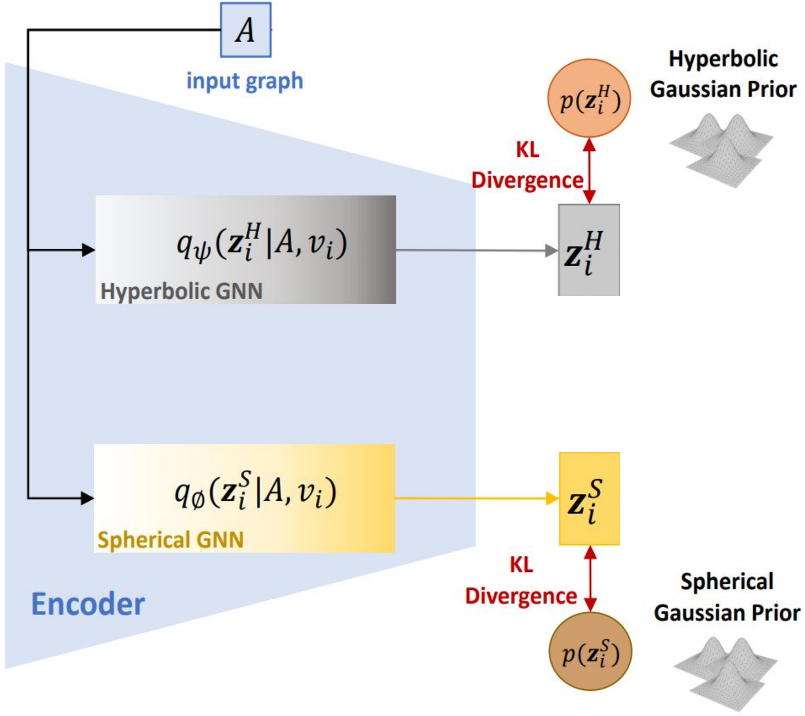
Motivation and Problem Definition



- A social network $G = (V, A)$ consists of vertices $V = \{v_i\}^N_{i=1}$, and adjacency matrix $e_{ij} \in A$
- We aim to design a model to jointly learn both node homophily and social influence representation, denoted $\mathbf{z}_i^S$ and $\mathbf{z}_i^H$ respectively, that can best **explain** the social network for link reconstruction.

UCLA Samueli Computer Science

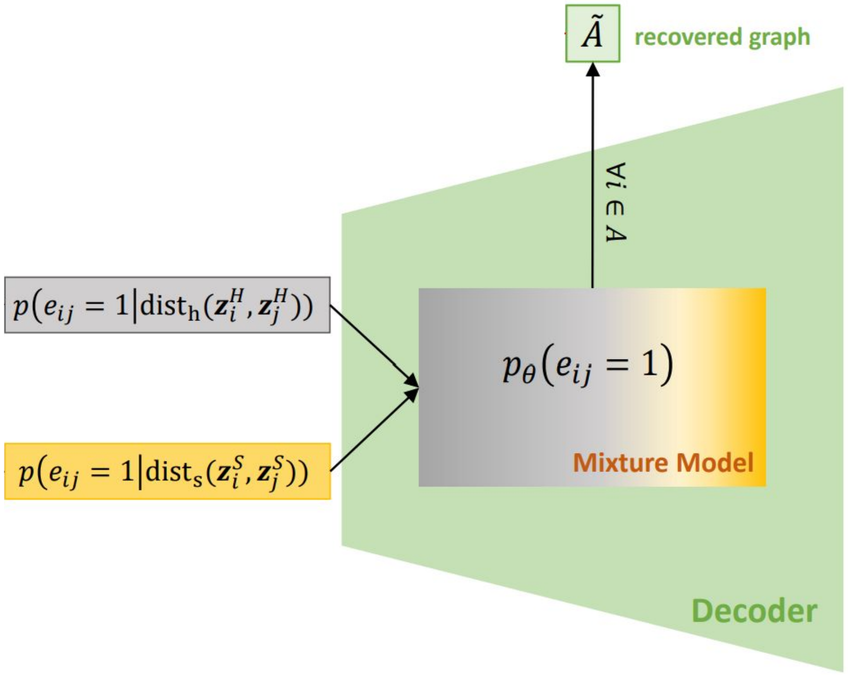
Encoder and Prior Distributions



- The encoder maps nodes into $\mathbf{z}^S$ (homophily) and $\mathbf{z}^H$ (social influence), which follow non-Euclidean prior distributions.

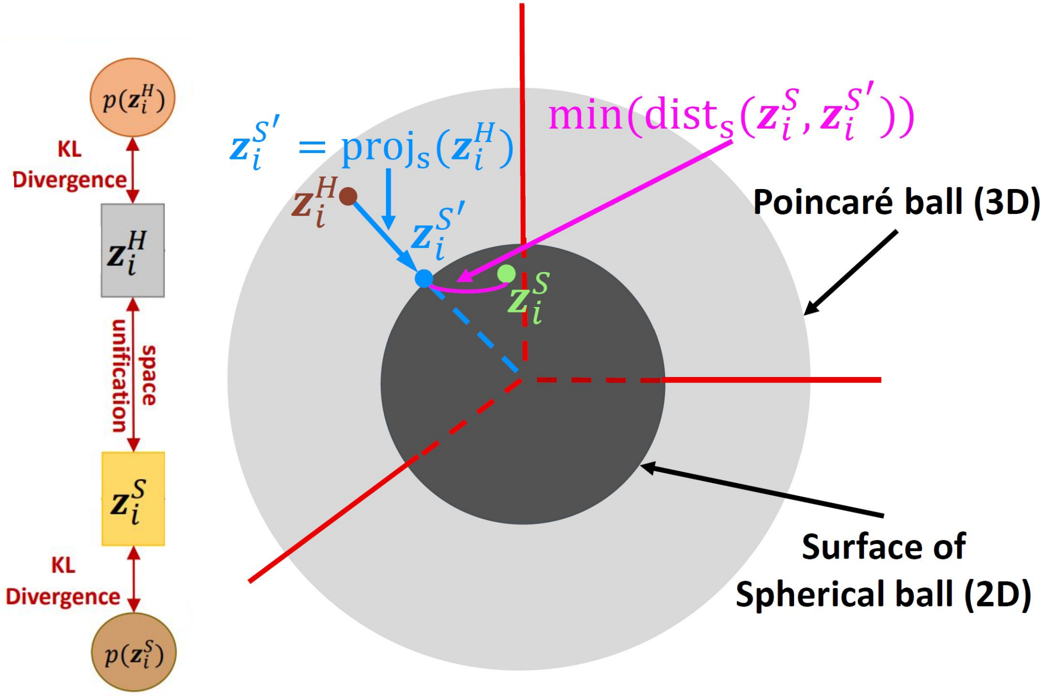
Decoder: Non-Euclidean Mixture Model

- Embeddings are passed into our mixture model decoder (homophily + social influence).
- Objective: maximize likelihood to observe links (= minimize link reconstruction loss).



Space Unification

- We design **space unification** component to align distinct geometric spaces
  - Ensures two embeddings of same node correspond to each other
  - $\boldsymbol{z}_i^H$ in the Poincaré ball is projected on the surface of the sphere and its distance to $\boldsymbol{z}_i^S$ is minimized



Space unification architecture

Dataset Statistics

Table 1: Dataset statistics for evaluation datasets.

Dataset	# Vertices	# Edges	Type	# Classes
BlogCatalog	10.3K	334.0K	undirected	39
LiveJournal	4.8M	69.0M	directed	10
Friendster	65.6M	1.8B	undirected	–

- **Experiment Evaluation:** 90% of links randomly sampled as training. Do not perform cross-validation as it may cause overfitting of learnable parameters: z_S, z_H, J, B, C, D, γ, W_l.

Evaluation: Classification & Link Prediction

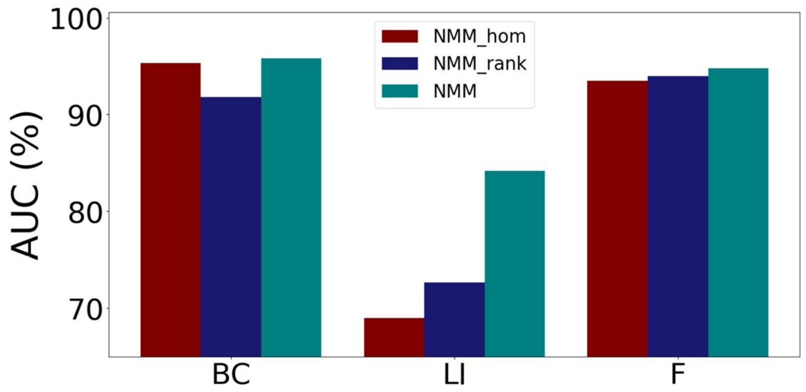
Table 3: Results of social network classification and link prediction for **Jaccard Index (%)**, **Hamming Loss (%)**, **F1 Score (%)**, **AUC (%)** using embedding dimension 64. Our **NMM** and its variants are in gray shading. For each group of models, the best results are bold-faced. The overall best results on each dataset are underscored. †Ablation study variant models using distinct non-Euclidean geometric spaces for **NMM** (homophily/social influence) where $\mathbb{E}$, $\mathbb{S}$, and $\mathbb{H}$ denote Euclidean, Spherical, and Hyperbolic spaces.

Datasets	BlogCatalog				LiveJournal				Friendster			
Metrics	Jaccard Index	Hamming Loss	F1 Score	AUC	Jaccard Index	Hamming Loss	F1 Score	AUC	Jaccard Index	Hamming Loss	F1 Score	AUC
GraRep	36.0	28.2	45.6	87.9	40.1	41.1	35.2	56.7	53.6	34.2	40.6	89.8
RolX	37.2	25.4	48.7	90.4	40.9	38.0	35.6	**60.1**	58.8	33.9	40.9	90.3
GraphWave	**39.5**	**22.8**	**48.9**	**92.3**	**42.2**	**37.6**	**35.9**	**60.1**	**59.0**	**31.5**	**41.1**	**90.5**
GraphSAGE	45.4	20.1	49.3	**92.0**	45.5	34.7	34.1	59.0	64.1	28.7	43.4	90.5
GCN	47.3	19.5	**55.1**	91.6	46.7	31.2	47.8	62.6	**66.5**	**28.0**	**47.2**	91.9
GAT	**47.9**	**19.3**	54.5	91.4	**47.4**	**28.5**	**49.0**	**65.3**	66.3	**28.0**	46.8	**92.0**
GELTOR	47.4	**19.3**	54.9	92.0	51.0	28.9	48.6	65.3	66.7	27.9	47.5	91.7
NRP	**61.6**	20.4	**65.2**	**95.5**	**69.7**	**24.5**	**64.0**	**78.7**	**72.2**	**22.6**	**52.8**	**92.2**
HGCN	56.7	**19.2**	60.9	92.7	58.8	**27.1**	**57.7**	68.5	**69.9**	24.3	49.9	**93.3**
κ-GCN	**61.6**	20.7	**65.4**	**95.3**	**63.6**	27.3	57.2	**69.1**	69.4	**24.1**	**50.3**	93.1
RaRE	61.4	20.6	65.6	95.1	74.2	23.8	65.1	79.9	75.7	22.5	55.0	94.4
NMM($\mathbb{H}^d/\mathbb{S}^d$)†	56.6	19.8	62.3	95.1	74.0	28.4	55.5	68.8	74.6	26.9	50.6	93.0
NMM($\mathbb{S}^d/\mathbb{S}^d$)†	57.1	19.6	65.9	94.0	74.7	27.6	57.1	69.0	75.3	26.2	52.5	93.4
NMM($\mathbb{E}^d/\mathbb{E}^d$)†	57.9	19.5	66.3	95.4	75.1	25.0	58.4	71.2	77.0	24.7	52.8	94.5
NMM($\mathbb{S}^d/\mathbb{E}^d$)†	59.2	19.2	67.1	95.5	75.3	24.4	59.3	74.5	77.5	23.3	54.3	94.5
NMM($\mathbb{H}^d/\mathbb{H}^d$)†	58.4	19.0	66.7	95.3	75.6	24.6	61.9	76.0	78.8	23.3	55.0	94.7
NMM($\mathbb{E}^d/\mathbb{H}^d$)†	60.3	19.1	67.8	95.7	76.2	23.2	64.4	79.2	79.1	22.6	55.4	94.5
NMM (ours)	**62.7**	19.0	70.9	95.8	76.5	22.7	**67.3**	84.2	79.8	22.1	56.3	94.8
NMM-GNN (ours)	62.6	**17.3**	**78.8**	**96.9**	**78.6**	**20.4**	**67.3**	**86.8**	**83.3**	**21.8**	**57.7**	**94.9**

- Modeling for both social network factors jointly achieves superior performance
- Homophily is best modeled in spherical space and social influence is best modeled in hyperbolic space

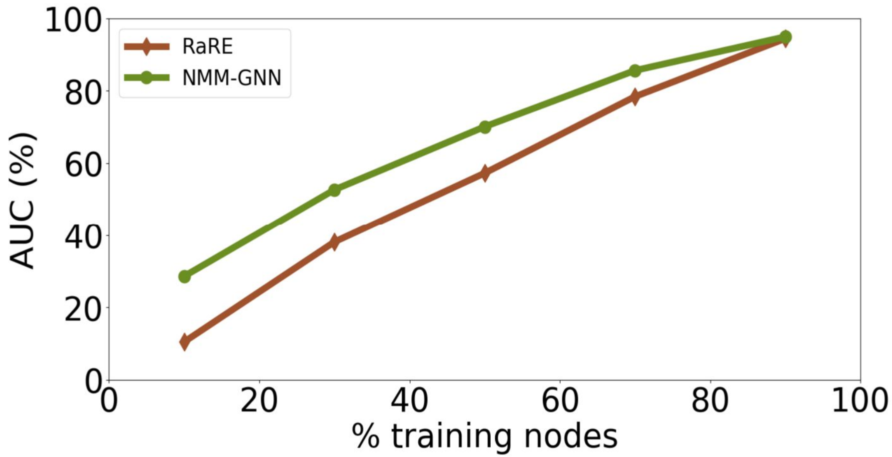
UCLA Samueli Computer Science

Additional Ablation Studies



- **NMM$_{hom}$**: $p_\theta(e_{ij} = 1) = \alpha \cdot p_{hom}(e_{ij} = 1)$
- **NMM$_{rank}$**: $p_\theta(e_{ij} = 1) = \beta \cdot p_{rank}(e_{ij} = 1)$
- **NMM**: our mixed distribution model

On embedding dimension 64 for AUC score



- **NMM-GNN** and **RaRE** on *LiveJournal*
- As less training nodes are observed, NMM-GNN outperforms RaRE by larger margins (e.g., 10% vs. 70% training nodes)
  - better generalization to unseen graphs

UCLA Samueli Computer Science

NMM-GNN Contributions

(1) We propose Graph-based Non-Euclidean Mixture Model (NMM) to explain social network generation. NMM represents nodes via joint influence by homophily (spherical space) and social influence (hyperbolic space), with space alignment component.

(2) The first to couple NMM with graph-based VAE learning framework, NMM-GNN.

  (a) We introduce a novel non-Euclidean VAE framework where node embeddings are learned with a powerful encoder of GNNs using spherical and hyperbolic spaces, non-Euclidean Gaussian priors, and unified non-Euclidean optimization.