Graph-Theoretic Insights into Bayesian Personalized Ranking for Recommendation

Kai Zheng[1,2] Jianxin Wang[1] Jinhui Xu[3]

[1]School of Computer Science and Engineering, Central South University
[2]College of Computer Science and Technology, China University of Petroleum
[3]USTC, School of Information Science and Technology

October 18, 2025

Why revisit BPR in GSL?

Observation. In graph self-supervised learning (GSL), BPR is widely used but inherently local:

$$\hat{y}_{ui} = \boldsymbol{e}_u^\top \boldsymbol{e}_i = (EE^\top)_{ui}$$

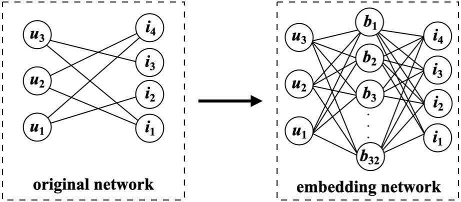
$\Rightarrow$ **counts 2-hop paths (Gram matrix).**

Limitations.

- **Locality**: ignores global connectivity and large-scale topology.
- **Coarse granularity**: similar topologies may map to the same score.
- **Norm bias**: scores inflate with larger embedding norms.

Goal. A *topology-aware*, *global* relation measure with theory and efficiency.

A network-geometry view of BPR



Embedding network with abstract nodes.

Treat E as an **embedding network**: users, items and abstract nodes; relation $\Rightarrow$ **path statistics**.

From maximum-entropy & latent hyperbolic geometry:

- 2-hop counts $\approx$ *energy distance* but are not precise.
- Need **global even-hop** information for finer topology fidelity.

Intuition

2-hop (common-neighbor) statistics are insufficient to recover latent distances; **weighted higher-order** structures give better topology-aware discrimination.

TopoLa distance & the BPR+ loss

Topology-encoded even-hop aggregation (TopoLa):

$$d_{\text{topo}}(u,i) = \frac{1}{\lambda}\left|\text{2-hop}\right| - \frac{1}{\lambda^2}\left|\text{4-hop}\right| + \frac{1}{\lambda^3}\left|\text{6-hop}\right| - \cdots$$

Plug into pairwise ranking (BPR+):

$$L_{\text{BPR+}} = -\sum_{u}\sum_{i\in\mathcal{N}_u}\sum_{j\notin\mathcal{N}_u} \ln\sigma\left(\frac{d_{\text{topo}}(u,i) - d_{\text{topo}}(u,j)}{\lambda}\right) + \tau\|\boldsymbol{E}^{(0)}\|^2.$$

Properties.

- Encodes **global connectivity** and **topological similarity**.
- **Finer discrimination**, reduced **norm bias** (Theorems 2 & 3).

Connection to graph convolution & layer fusion

LightGCN update:

$$\boldsymbol{E}^{(k)} = \alpha_k \left(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\right)^k \boldsymbol{E}^{(0)}.$$

Final embedding:

$$\boldsymbol{E} = \sum_{k=0}^{K} \alpha_k \widetilde{\boldsymbol{A}}^k \boldsymbol{E}^{(0)}, \quad \widetilde{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}.$$

Implications:

- Higher k reduces path-count variance $\Rightarrow$ over-smoothing intuition.
- Fusion captures degree/topology at *multiple hop scales*.
- Equivalent to adding *self-connections* in a path-count sense.

Efficient computation via SVD (practical BPR+)

Naïve series on $H = EE^\top$:

$$\boldsymbol{D}_{\text{topo}} = \frac{1}{\lambda}\boldsymbol{H} - \frac{1}{\lambda^2}\boldsymbol{H}^2 + \frac{1}{\lambda^3}\boldsymbol{H}^3 - \cdots \quad \Rightarrow \quad \text{costly (batch-cubic).}$$

SVD trick: $\boldsymbol{E} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$,

$$\boldsymbol{D}_{\text{topo}} = \boldsymbol{U}\left(\frac{1}{\lambda}\boldsymbol{\Sigma}^2 - \frac{1}{\lambda^2}\boldsymbol{\Sigma}^4 + \frac{1}{\lambda^3}\boldsymbol{\Sigma}^6 - \cdots\right)\boldsymbol{U}^\top.$$

Benefits.

- Complexity $\mathcal{O}(N_b N_e^2 + N_b^2 N_e + N_e^3)$, near BPR wall clock.
- In practice, **40-hop BPR+ (MF)** $\approx$ BPR runtime; $\ll$ naive H^k.

1. Introduction & Motivation
2. Method: TopoLa, BPR+ & Efficiency
3. **Results & Analyses**
4. Case Study
5. Conclusion

Five datasets, multiple backbones: consistent gains

- **Datasets**: Amazon, Gowalla, Yelp, LastFM, Beer.
- **Backbones**: LightGCN, SGL, NCL, LightGCL, AdaGCL.
- **Metrics**: Recall@N, NDCG@N (N=10,20).

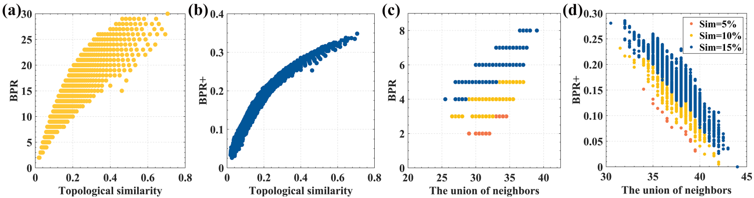
Representative results on **Amazon**

- AdaGCL ⇒ **AdaGCL+**: **Recall@10 +11.9%, NDCG@10 +13.4%.**
- Similar gains across models/datasets; p-values confirm significance.

Dataset	R@10	N@10
Amazon	**+11.9**	**+13.4**
Gowalla	+0.8	+0.9
Yelp	+1.5	+1.9
LastFM	+4.4	+2.0
Beer	+1.3	+2.0

R@10 = Recall@10, N@10 = NDCG@10 (relative %).

Finer granularity & less norm bias



Score vs. topology similarity.

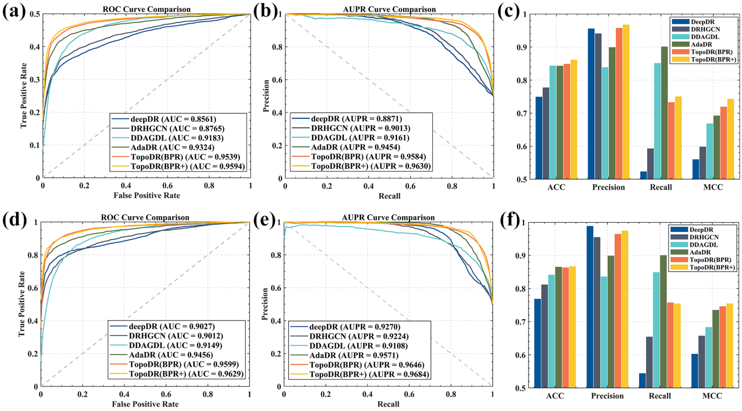
Findings (1)

- BPR+ scores vary **monotonically** with topology similarity.
- **Higher resolution** than BPR for close similarities.

Findings (2)

- Less sensitive to union-of-neighbors effect (embedding norm).
- More topology-aware discrimination overall.

Application: drug repositioning with TopoDR (I)



ROC/PR on Fdataset & Cdataset.

Application: drug repositioning with TopoDR (II)

Pipeline

LightGCN+ embeddings + multimodal drug/disease features ⇒ **TopoDR**.

- Graph SSL embedding with *TopoLa*/BPR+.
- Feature fusion: chemistry, ATC, side effects, DDI, targets; disease ontology & phenotypes.
- Classifier: Random Forest for link prediction.

Results

- On *Fdataset* & *Cdataset*: **AUC/AUPR/MCC** surpass strong baselines.
- Produces plausible oncology candidates (CRC, BC, GC, Leukemia).
- Gains stem from **global topology** captured by BPR+.

Takeaways & Outlook

Takeaways.

- BPR's 2-hop locality limits global topology fidelity.
- **BPR+** uses **TopoLa** to encode *all even-hop paths* with alternating weights.
- SVD-based computation makes BPR+ **practically efficient**.

Outlook.

- Inspire **new GNN modules** from a topology/geometry lens.
- Further **acceleration** and **larger-scale** validations.

References

[1] Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian Personalized Ranking from Implicit Feedback.

[2] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. (2020). LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation.

[3] Wu, J., Wang, X., Feng, F., He, X., et al. (2021). Self-supervised Graph Learning for Recommendation (SGL).

[4] Cai, X., Wang, B., et al. (2023). LightGCL: Simple yet effective graph contrastive learning for recommendation.

[5] Wang, Z., et al. (2022). AdaGCL: Adaptive graph contrastive learning for recommendation.

Thanks!

Formulas & Theorems (for Q&A)

BPR (pairwise):

$$L_{\text{BPR}} = -\sum_{u}\sum_{i\in\mathcal{N}_u}\sum_{j\notin\mathcal{N}_u} \ln\sigma(\hat{y}_{ui}-\hat{y}_{uj}) + \tau\|\boldsymbol{E}^{(0)}\|^2, \quad \hat{y}_{ui} = \boldsymbol{e}_u^\top \boldsymbol{e}_i.$$

TopoLa & BPR+: see Method slide.

Thm 1 (intuition). 2-hop insufficient for precise latent energy distance.

Thm 2. TopoLa yields finer topology-aware discrimination bounds.

Thm 3. d_{topo} is proportional to topology similarity.

Datasets & Metrics (for Q&A)

Dataset	#Users	#Items	#Interactions	Density
Amazon	76,469	83,761	966,680	1.5×10^{-4}
Gowalla	25,557	19,747	294,983	5.8×10^{-4}
Yelp	42,712	26,822	182,357	1.6×10^{-4}
LastFM	1,892	17,632	92,834	2.8×10^{-3}
Beer	10,456	13,845	1,381,094	9.5×10^{-3}

Metrics. Recall@N, NDCG@N ($N \in \{10, 20\}$).