FREEDOM

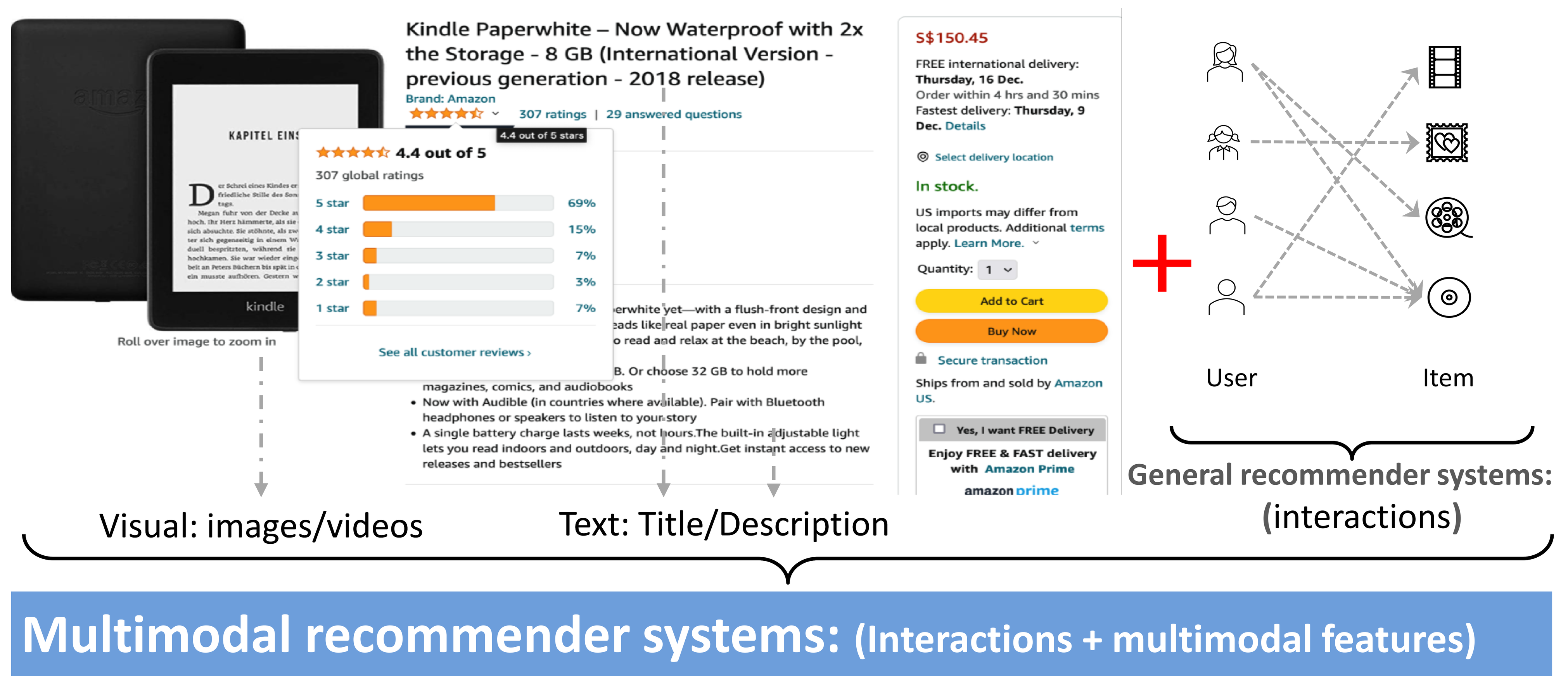
(A Tale of Two Graphs: Freezing and Denoising Graph Structures for Multimodal Recommendation)

Xin Zhou (xin.zhou@ntu.edu.sg), Zhiqi Shen (zqshen@ntu.edu.sg)

Alibaba-NTU Singapore Joint Research Institute

Introduction

- **General recommender systems** utilize user-item interactions for recommendation.
- **Multimodal recommender systems** can further exploit item multimodal information (e.g., images and textual descriptions) to improve the recommendation accuracy.



FREEDOM

Freezing the latent item-item graph

Constructing Frozen Item-Item Graph. We uses kNN to construct an initial modality-aware item-item graph S^m using raw features x_i^m from each modality m.

Graph sparsification. We further employ kNN sparsification and convert the weighted S^m into an **unweighted matrix**.

$$S_{ij}^m = \frac{(x_i^m)^\top x_j^m}{\|x_i^m\| \|x_j^m\|}, \quad \widehat{S}_{ij}^m = \begin{cases} 1, & S_{ij}^m \in \text{top}-k(S_i^m), \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

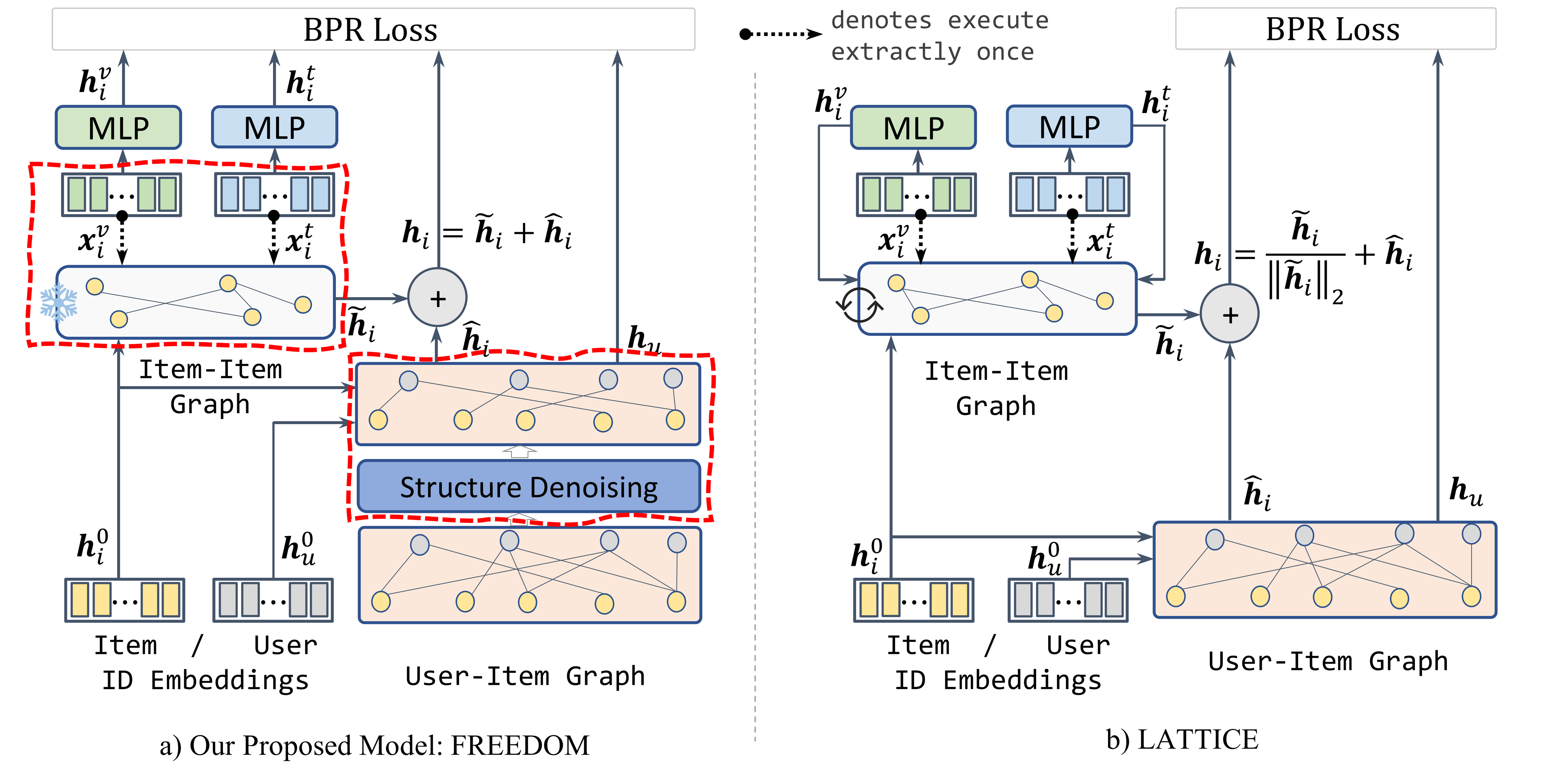
Freezing. Finally, we construct the latent item-item graph by aggregating the structures from each modality and freeze it for recommendation.

Denoising the user-item graph

Degree-sensitive edge pruning. Given a specific edge e_k which connects node i and j, we calculate its probability as

$$p_k = \frac{1}{\sqrt{\omega_i}\sqrt{\omega_j}}, \tag{2}$$

where ω_i and ω_j are the degrees of nodes i and j in the user-item graph, respectively.



a) Our Proposed Model: FREEDOM

b) LATTICE

Performance Comparison

FREEDOM improves LATTICE[1] by an average of ***19.07***% across all datasets.

Dataset	Metric	BPR	LightGCN	VBPR	MMGCN	GRCN	DualGNN	SLMRec	LATTICE	FREEDOM	*improv.*
Baby	R@10	0.0357	0.0479	0.0423	0.0421	0.0532	0.0513	0.0521	<u>0.0547</u>	**0.0627**	14.63%
	R@20	0.0575	0.0754	0.0663	0.0660	0.0824	0.0803	0.0772	<u>0.0850</u>	**0.0992**	16.71%
	N@10	0.0192	0.0257	0.0223	0.0220	0.0282	0.0278	0.0289	<u>0.0292</u>	**0.0330**	13.01%
	N@20	0.0249	0.0328	0.0284	0.0282	0.0358	0.0352	0.0354	<u>0.0370</u>	**0.0424**	14.59%
Sports	R@10	0.0432	0.0569	0.0558	0.0401	0.0599	0.0588	<u>0.0663</u>	0.0620	**0.0717**	15.65%
	R@20	0.0653	0.0864	0.0856	0.0636	0.0919	0.0899	<u>0.0990</u>	0.0953	**0.1089**	14.27%
	N@10	0.0241	0.0311	0.0307	0.0209	0.0330	0.0324	<u>0.0365</u>	0.0335	**0.0385**	14.93%
	N@20	0.0298	0.0387	0.0384	0.0270	0.0413	0.0404	<u>0.0450</u>	0.0421	**0.0481**	14.25%
Clothing	R@10	0.0206	0.0361	0.0281	0.0227	0.0421	0.0452	0.0442	<u>0.0492</u>	**0.0629**	27.85%
	R@20	0.0303	0.0544	0.0415	0.0361	0.0657	0.0675	0.0659	<u>0.0733</u>	**0.0941**	28.38%
	N@10	0.0114	0.0197	0.0158	0.0120	0.0224	0.0242	0.0241	<u>0.0268</u>	**0.0341**	27.24%
	N@20	0.0138	0.0243	0.0192	0.0154	0.0284	0.0298	0.0296	<u>0.0330</u>	**0.0420**	27.27%

Motivation

LATTICE[1] model demonstrates state-of-the-art performance in multimodal recommendation due to two key factors:

1. **Latent Item-Item Structures**: LATTICE **learns the latent item-item graph structures** that are inherent in the multimodal contents of items.
2. **High-Order Interaction Semantics**: LATTICE exploits the high-order interaction semantics from the user-item graph.

Is graph learning necessary? No

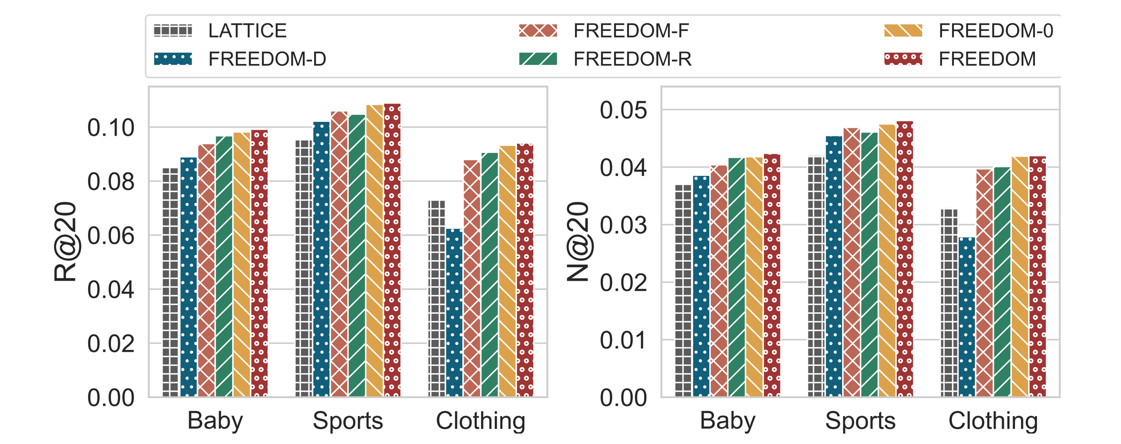
We compare the recommendation performance between LATTICE and the variant of LATTICE, i.e., LATTICE-Frozen, which freezes the item-item graph structure.

Dataset	Metric	LATTICE	LATTICE-Frozen
Baby	R@10	0.0547	0.0551
	R@20	0.0850	0.0873
	N@10	0.0292	0.0291
	N@20	0.0370	0.0373
Sports	R@10	0.0620	0.0626
	R@20	0.0953	0.0964
	N@10	0.0335	0.0336
	N@20	0.0421	0.0423
Clothing	R@10	0.0492	0.0434
	R@20	0.0733	0.0635
	N@10	0.0268	0.0227
	N@20	0.0330	0.0279

Although LATTICE-Frozen outperforms its original version in **Baby** and **Sports**, its frozen item-item graph, which uses edge weights to represent item affinities, can be noisy. $\Longrightarrow$ FREEDOM.

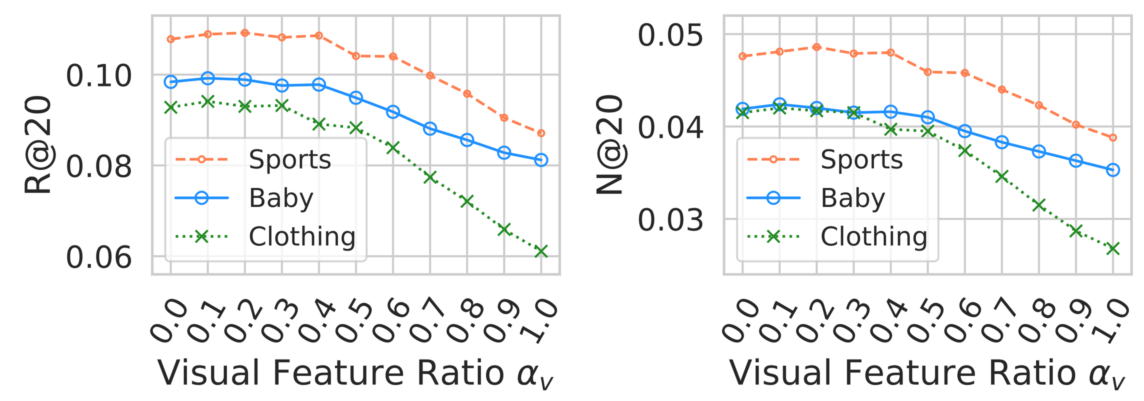
FREEDOM Ablation

We compare different components of FREEDOM, and the following figure reveals that **Freezing** contributes the most to our model.



Modality Ablation

Imapct of FREEDOM with multimodal features: **textual features** play a more important role than visual features.



Reference

[1]. Zhang, Jinghao, et al. "Mining latent structures for multimedia recommendation." Proceedings of the 29th ACM International Conference on Multimedia. 2021.

MMRec Framework

- 10+ multimodal models
- Including all baselines
- https://github.com/enoche/mmrec