Profiling the Design Space for Graph Neural Networks based Collaborative Filtering

Zhenyi Wang[1] , Huan Zhao[2], Chuan Shi[1]

[1]Beijing University of Posts and Telecommunications, China
[2]4Paradigm Inc., China

The 15th ACM International Conference on Web Search and Data Mining
WSDM 2022, Phoenix, AZ, USA
February 21–25, 2022

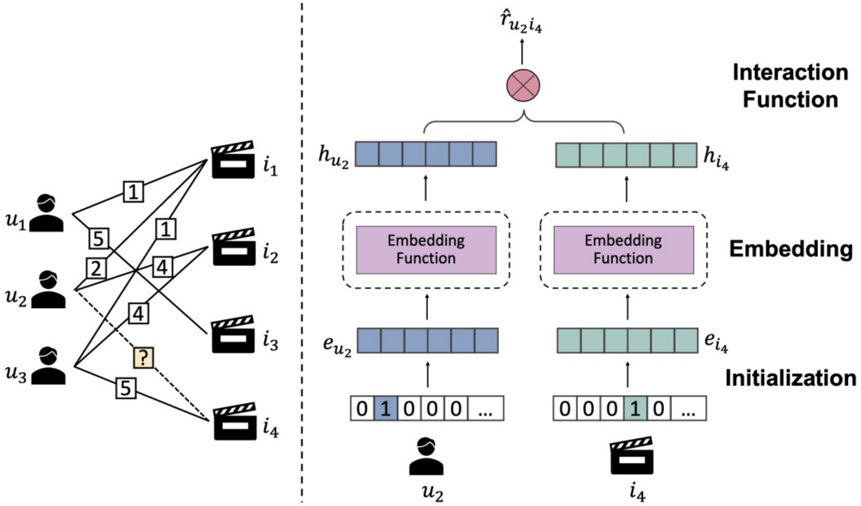
CONTENTS

CONTENTS

GNN-base CF

- **Collaborative Filtering (CF)** has been one of the most popular recommender system (RS) methods

- **Graph Neural Networks (GNNs)** have been incorporated for the CF tasks, since the user-item interactions can be naturally modeled as a bipartite graph



Motivation

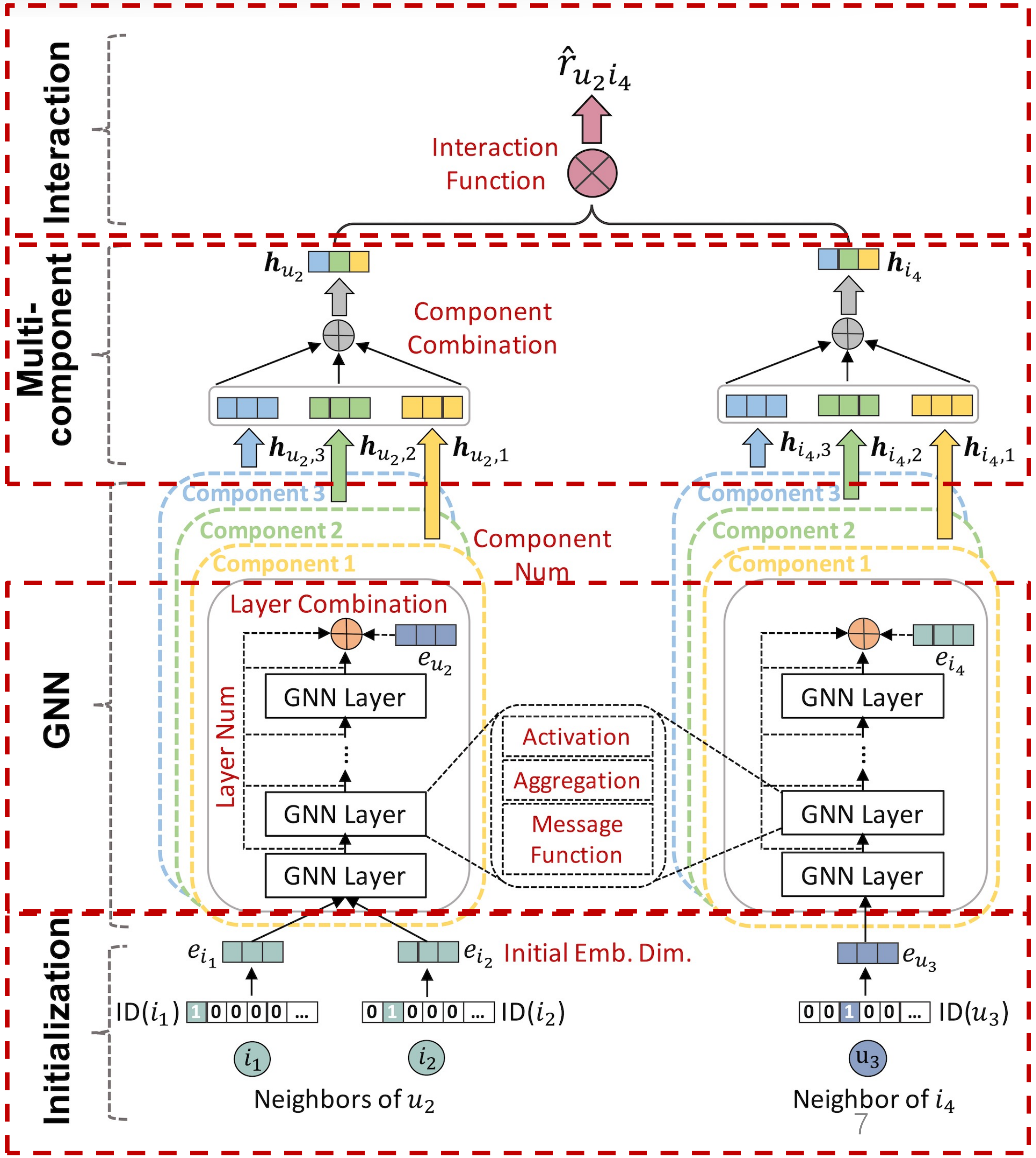
- Existing GNN-based methods are mainly limited to designing a single best architecture for a specific scenario
- Little has been done to systematically understand the influences of different design dimensions of GNN-based CF on recommendation performance

We propose to profile the **design space** of existing GNN-based methods for CF

CONTENTS

The Unified Framework

- 4 modules of the unified framework:
  - **Initialization** performs embedding matrix ID-lookup
  - **GNN** refines the embeddings via the propagation of GNN layers
  - **Multi-component** models diverse user interests from different aspects
  - **Interaction** predicts the rating value of the user-item pair
- Covers extensive popular GNN-based CF models
- On top of the framework, 9 **design dimensions** can be extracted



- Popular choices in the 9 design dimensions:

Design Dimension	Choices
Initial Embedding Dimension d	64, 128, 256
Message Function $m(\cdot)$	Identity, Hadamard
Aggregation $f(\cdot)$	None, GCN, GAT, GIN, GraphSAGE
Activation $\sigma(\cdot)$	Identity, Sigmoid, Tanh, ReLU, PReLU, LeakyReLU
Layer Number L	1, 2, 3, 4
Layer Combination $g(\cdot)$	Stack, Concat, Sum, Mean
Component Number K	1, 2, 3, 4
Component Combination $c(\cdot)$	Concat, Mean, Att
Interaction Function $p(\cdot)$	Dot Product, Concat+MLP, Sum+MLP

$$\mathbf{h}_u^{(l+1)} = \sigma\Big(\ \ldots \ '(u)\big\}, \mathbf{h}_u^{(l)}\big]\Big)$$

$$\hat{r}_{ui} = \boxed{p}(\mathbf{h}_u, \mathbf{h}_i)$$

non-learnable interaction functions *Dot Product* and neural ones base on *MLPs*

4 GNN aggregators and *None* to enlarge the capacity of the design space to include those non-GNN-based models

- To explore the impacts of different design dimensions, we propose to profile the **design space**, defined as the Cartesian product of design dimensions[1]

[1] J. You et al. Design space for graph neural networks. NeurIPS 2020.

Relationship with Existing CF Methods

- Popular CF methods that can be instantiated from the proposed design space:

Category	Model	$m(\cdot)$	$f(\cdot)$	$\sigma(\cdot)$	$g(\cdot)$	$c(\cdot)$	$p(\cdot)$	Single-/Multi-component
Classic	MF [16, 24]	Identity	None	Identity	Stack	-	Dot Product	Single
	LLORMA [18, 46]	Identity	None	Identity	Stack	Att	Dot Product	Multiple
MLP-based	NCF [11]	Identity	None	ReLU	Stack	-	Concat+MLP	Single
GNN-based	NGCF [36]	Hadamard	GCN	LeakyReLU	Concat	-	Dot Product	Single
	LightGCN [10]	Identity	GCN	Identity	Mean	-	Dot Product	Single
	LR-GCCF [2]	Identity	GCN	Identity	Concat	-	Dot Product	Single
	SMOG-CF [43]	Hadamard	GCN	ReLU	Concat	-	Dot Product	Single
	PinSage [41]	Identity	GraphSAGE	ReLU	Stack	-	Dot Product	Single
	MCCF [38]	Identity	GAT	ReLU	Stack	Att	Concat+MLP	Multiple
	DGCF [37]	Identity	GCN	Tanh	Sum	Concat	Dot Product	Multiple

CONTENTS

Evaluation of the Design Space

Datasets

Dataset	# of Users	# of Items	# of Interactions	Rating Scale	Density
Yelp[1]	58,069	31,721	1,160,605	[1,5]	0.063%
Amazon-CDs [9]	31,296	24,379	622,163	[1,5]	0.082%
Amazon-Movies [9]	44,439	25,047	1,070,860	[1,5]	0.096%
YahooMusic [5, 25]	1,357	1,363	5,335	[1,100]	0.28%
Amazon-Beauty [9]	7,068	3,570	79,506	[1,5]	0.32%
Flixster[13, 25]	2,341	2,956	26,173	[0.5,5]	0.38%
Douban [23, 25]	2,999	3,000	136,891	[1,5]	1.52%
MovieLens-1M[2]	6,040	3,706	1,000,209	[1,5]	4.47%
MovieLens-100K[3]	943	1,682	100,000	[1,5]	6.31%

[1] https://www.yelp.com/dataset/
[2] https://grouplens.org/datasets/movielens/100k/
[3] https://grouplens.org/datasets/movielens/1m/

Loss: *MSE*

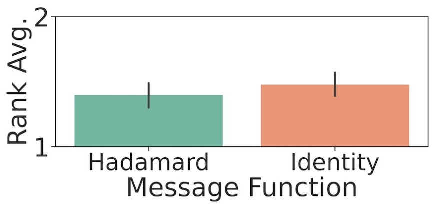
$$L = \frac{\sum_{(u,i)\in O_t} (\hat{r}_{ui} - r_{ui})^2}{|O_t|} + \lambda \|\Theta\|^2$$

Evaluation Metric: *RMSE*

$$RMSE = \sqrt{\frac{\sum_{(u,i)\in O_e} (\hat{r}_{ui} - r_{ui})^2}{|O_e|}}$$

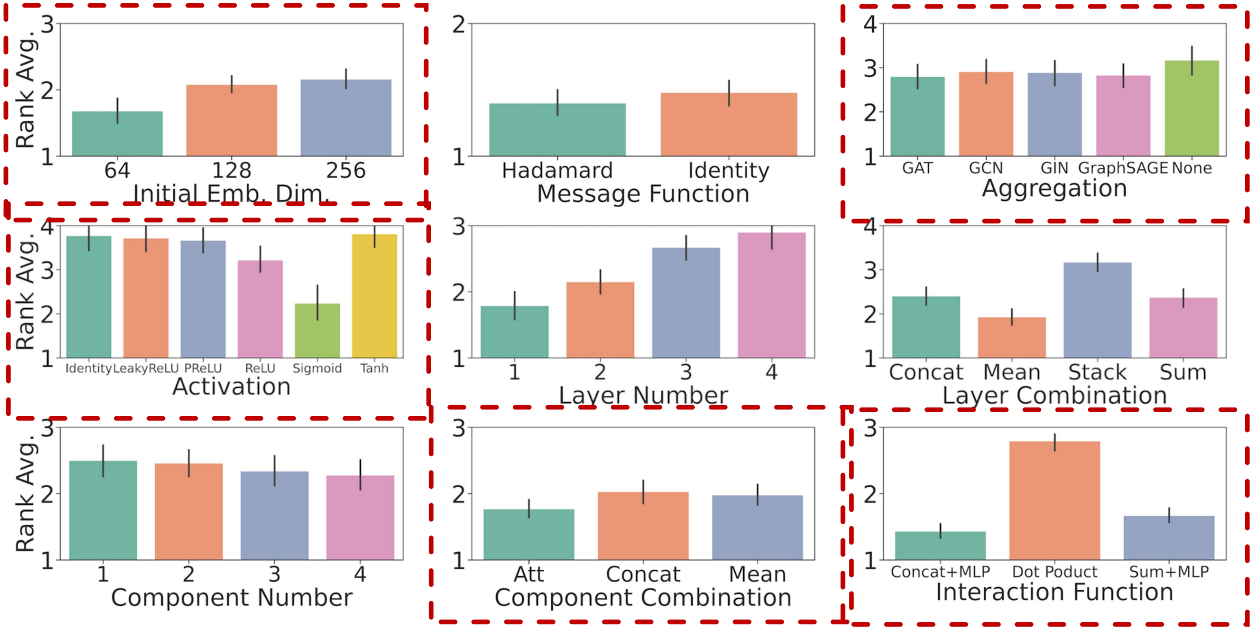
Evaluation technique: *controlled random search*[1]

Group No.	Configuration Space					Experimental Results	
	Message Function	Activation	...	Interaction Function	Dataset	Performance	Ranking
1	Identity	ReLU	...	Dot Product	Yelp	0.8996	2
	Hadamard					0.8665	1
2	Identity	Sigmoid	...	Concat+MLP	Amazon-CDs	0.7636	1
	Hadamard					0.7812	2
...							
S	Identity	Tanh	...	Sum+MLP	MovieLens-1M	0.8231	1(tie)
	Hadamard					0.8231	1(tie)



[1] J. You et al. Design space for graph neural networks. NeurIPS 2020.

Evaluation Results



- The experimental findings provide valuable insights for effectively designing top-performing models
  - *GAT* and *GraphSAGE* slightly outperform, and *None* is competitive
  - *Sigmoid* stands out
  - *Att* is the preferable multi-component combination
  - Neural interaction is superior to *Dot Product*
  - …

- There exists some **redundancy** in the design space
- The vanilla design space can be further **pruned**
  - e.g., the initial embedding dimension can be fixed as 64
- Boost the searching efficiency of top-performing models

[1] J. You et al. Design space for graph neural networks. NeurIPS 2020.

The Pruned Design Space

- By remaining favorable design choices, the pruned design space contains **a higher concentration** of **top-performing models**
- Choices in the design dimensions of the pruned design space:

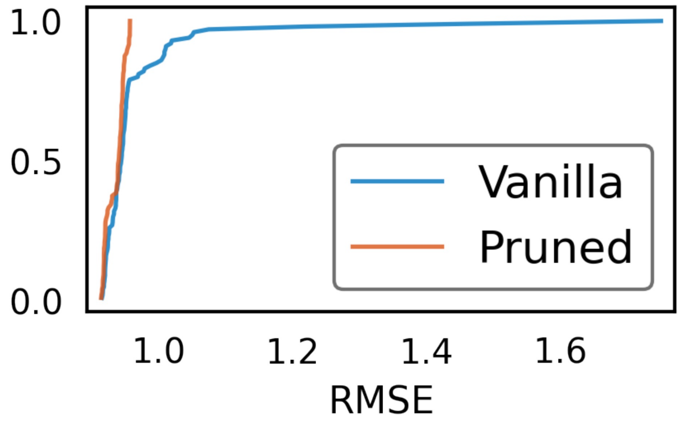
Design Dimension	Choices
Initial Embedding Dimension d	`64`
Message Function $m(\cdot)$	`Identity`, `Hadamard`
Aggregation $f(\cdot)$	`None`, `GraphSAGE`
Activation $\sigma(\cdot)$	`Identity`, `Sigmoid`, `ReLU`
Layer Number L	`1`, `2`
Layer Combination $g(\cdot)$	`Mean`
Component Number K	`1`, `4`
Component Combination $c(\cdot)$	`Att`
Interaction Function $h(\cdot)$	`Concat+MLP`, `Sum+MLP`

- 96 v.s. 103,680, reduction of three orders of magnitude (1,080x)

Evaluation of the Pruned Design Space

- Evaluation technique
  - *RMSE empirical distribution function (EDF)*[1]
  - Suppose $\mathbf{1}$ as the indicator function, n as the number of sampled models, each with RMSE x_i. The RMSE EDF is given by:

$$F(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left[x_i < x\right]$$



  - $F(x)$ gives the fraction of models with RMSE less than x
  - Comparing distributions can help to obtain more robust and informative conclusions

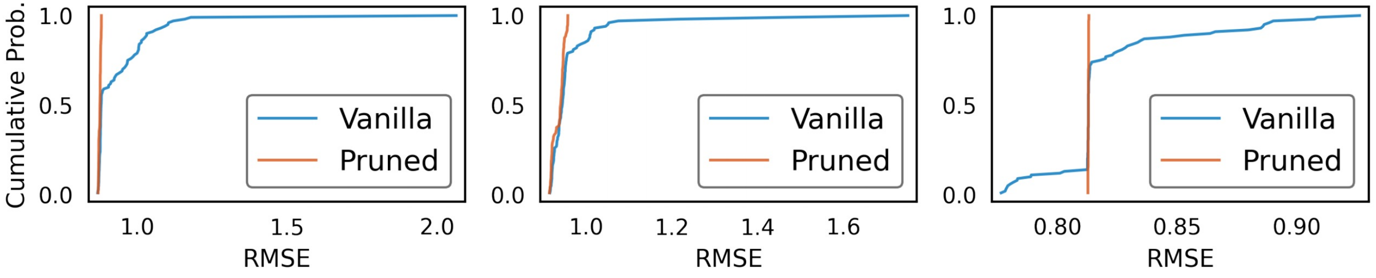
[1] I. Radosavovic et al. On network design spaces for visual recognition. ICCV 2019.

Evaluation of the Pruned Design Space

The 15th ACM International Conference on Web Search and Data Mining
wsdm Phoenix, AZ, USA 2022
February 21–25, 2022

Generalization evaluation

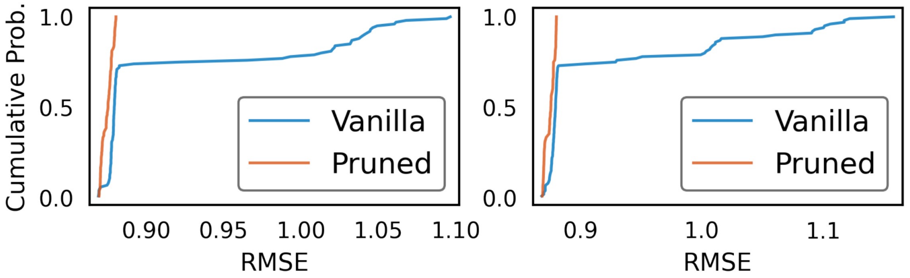
Different levels of density



Yelp　　Amazon-Beauty　　MovieLens-1M

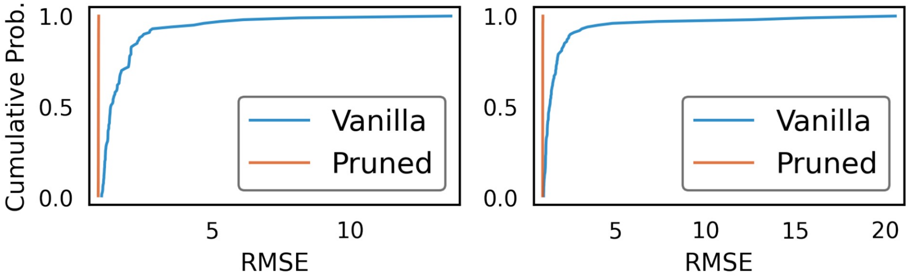
Dataset	Density
Yelp[1]	0.063%
Amazon-CDs [9]	0.082%
Amazon-Movies [9]	0.096%
YahooMusic [5, 25]	0.28%
Amazon-Beauty [9]	0.32%
Flixster[13, 25]	0.38%
Douban [23, 25]	1.52%
MovieLens-1M[2]	4.47%
MovieLens-100K[3]	6.31%

Different model complexity



Low model comlexity　　High model comlexity

Generalization to new datasets



Epinions　　Amazon-Sports

- The pruned design space consistently holds **better quality**, which demonstrates its **strong generalization** to various new settings.

Case Study: Random Search

- We perform random search on the pruned design space
  - The searched model architectures on new datasets :

Dataset	$m(\cdot)$	$f(\cdot)$	$\sigma(\cdot)$	L	K	$p(\cdot)$
Epinions	Hadamard	None	ReLU	1	1	Sum+MLP
Amazon-Sports	Identity	GraphSAGE	Sigmoid	1	4	Sum+MLP

  - Performance comparisons with baselines:

Model	Epinions	Amazon-Sports
MF [16]	0.9945 ± 0.0000	0.9882 ± 0.0007
NCF [11]	1.0070 ± 0.0055	<u>0.9342 ± 0.0008</u>
NGCF [36]	1.1437 ± 0.0240	1.0668 ± 0.0038
LightGCN [10]	<u>0.9926 ± 0.0001</u>	0.9705 ± 0.0003
DGCF [37]	1.6800 ± 0.2272	0.9894 ± 0.0000
RS-10	**0.8729 ± 0.0014**	**0.9327 ± 0.0006**

- The randomly searched models outperform all the baselines

- The pruned design space helps to **efficiently** design top-performing models for **new recommendation scenarios**

CONTENTS

Problem

- The **first attempt** to profile the design space of GNN-based CF

Evaluation of the Design Space

- Propose **a unified framework** covering popular GNN-based CF models
- Develop a design space and **evaluate** it by extensive experiments

Evaluation of the Pruned Design Space

- Prune the design space for **a higher concentration** of top-performing models
- Empirical studies demonstrate its **high quality** and **strong generalization ability**

Thank you!

Paper

Code

Contact: Zhenyi Wang, zy_wang@bupt.edu.cn