TUTORIALS

Graph Neural Networks for Recommendation: Reproducibility, Graph Topology, and Node Representation

LoG CONFERENCE

Part 1. Reproducibility

60 minutes

The 2nd Learning on Graphs Conference (LoG 2023)

TABLE OF CONTENTS

01

GNNs-BASED RECSYS IN ELLIOT

HANDS-ON

02

REPRODUCING GNNs-BASED RECSYS

03

GNNs-BASED *vs.* TRADITIONAL RECSYS

04

EXPLORING UNCOMMON DATASETS

05

INFORMATION FLOW IN GNNs-BASED RECSYS

06

TAKE-HOME MESSAGES

USEFUL RESOURCES

The content of the following slides is taken from:

- Vito Walter Anelli, Daniele Malitesta, Claudio Pomo, Alejandro Bellogín, Eugenio Di Sciascio, Tommaso Di Noia: **Challenging the Myth of Graph Collaborative Filtering: a Reasoned and Reproducibility-driven Analysis**. RecSys 2023: 350-361
- Daniele Malitesta, Claudio Pomo, Vito Walter Anelli, Tommaso Di Noia, Antonio Ferrara: **An Out-of-the-Box Application for Reproducible Graph Collaborative Filtering extending the Elliot Framework**. UMAP (Adjunct Publication) 2023: 12-15

Scan me!

01

HANDS-ON #1

GNNs-BASED RECSYS IN ELLIOT

SCAN ME AND GO TO GOOGLE COLAB!

or find me at:

https://sisinflab.github.io/tutorial-gnns-recsys-log2023/sections/reproducibility/

02

REPRODUCING GNNs-BASED RECSYS

SELECTED GNNs-BASED APPROACHES

Model	**Venue**	**Year**	**Strategy**
NGCF	SIGIR	2019	• Pioneer approach in graph CF • Inter-dependencies among *ego* and *neighbor* nodes
DGCF	SIGIR	2020	• Disentangles users' and items' into intents and weights their importance • Updates graph structure according to those learned intents
LightGCN	SIGIR	2020	• Lightens the graph convolutional layer • Removes feature transformation and non-linearities
SGL	SIGIR	2021	• Brings self-supervised and contrastive learning to recommendation • Learns multiple node views through node/edge dropout and random walk
UltraGCN	CIKM	2021	• Approximates infinite propagation layers through a constraint loss and negative sampling • Explores item-item connections
GFCF	CIKM	2021	• Questions graph convolution in recommendation through graph signal processing • Proposes a strong close-form algorithm

WHY REPLICATING GRAPH CF RESULTS?

Generally speaking

- **Several approaches** tend to **copy and paste previous results** from baselines
- Sometimes **they do not provide full details** about the **experimental settings**

Specifically

- Apart from **NGCF**, all **other approaches** take baselines' **results from previous papers**
- **Some authors are shared** across such works

WHAT WE HAVE DONE

1) Re-implement from scratch

all baselines by **carefully following the original works**

2) Train/evaluate them

within Elliot to provide a **fair and repeatable environment**

3) Use the same settings

as the **original papers and codes** in terms of **dataset splitting, hyper-parameters, and evaluation protocol**

4) Compare our results

to the **original ones** to assess the **numerical differences**

REPRODUCIBILITY RESULTS

Datasets	Models	Ours		Original		Performance Shift	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Gowalla	NGCF	0.1556	0.1320	0.1569	0.1327	$-1.3 \cdot 10^{-03}$	$-7 \cdot 10^{-04}$
	DGCF	0.1736	0.1477	0.1794	0.1521	$-5.8 \cdot 10^{-03}$	$-4.4 \cdot 10^{-03}$
	LightGCN	0.1826	0.1545	0.1830	0.1554	$-4 \cdot 10^{-04}$	$-9 \cdot 10^{-04}$
	SGL*	—	—	—	—	—	—
	UltraGCN	0.1863	0.1580	0.1862	0.1580	$+1 \cdot 10^{-04}$	0
	GFCF	0.1849	0.1518	0.1849	0.1518	0	0
Yelp 2018	NGCF	0.0556	0.0452	0.0579	0.0477	$-2.3 \cdot 10^{-03}$	$-2.5 \cdot 10^{-03}$
	DGCF	0.0621	0.0505	0.0640	0.0522	$-1.9 \cdot 10^{-03}$	$-1.7 \cdot 10^{-03}$
	LightGCN	0.0629	0.0516	0.0649	0.0530	$-2 \cdot 10^{-03}$	$-1.4 \cdot 10^{-03}$
	SGL	0.0669	0.0552	0.0675	0.0555	$-6 \cdot 10^{-04}$	$-3 \cdot 10^{-04}$
	UltraGCN	0.0672	0.0553	0.0683	0.0561	$-1.1 \cdot 10^{-03}$	$-8 \cdot 10^{-04}$
	GFCF	0.0697	0.0571	0.0697	0.0571	0	0
Amazon Book	NGCF	0.0319	0.0246	0.0337	0.0261	$-1.8 \cdot 10^{-03}$	$-1.5 \cdot 10^{-03}$
	DGCF	0.0384	0.0295	0.0399	0.0308	$-1.5 \cdot 10^{-03}$	$-1.3 \cdot 10^{-03}$
	LightGCN	0.0419	0.0323	0.0411	0.0315	$+8 \cdot 10^{-04}$	$+8 \cdot 10^{-04}$
	SGL	0.0474	0.0372	0.0478	0.0379	$-4 \cdot 10^{-04}$	$-7 \cdot 10^{-04}$
	UltraGCN	0.0688	0.0561	0.0681	0.0556	$+7 \cdot 10^{-04}$	$+5 \cdot 10^{-04}$
	GFCF	0.0710	0.0584	0.0710	0.0584	0	0

**Results are not provided since SGL was not originally trained and tested on Gowalla.*

The most **significant performance shift** is in the **order of 10e-3**

REPRODUCIBILITY RESULTS

Datasets	Models	Ours		Original		Performance Shift	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Gowalla	NGCF	0.1556	0.1320	0.1569	0.1327	$-1.3 \cdot 10^{-03}$	$-7 \cdot 10^{-04}$
	DGCF	0.1736	0.1477	0.1794	0.1521	$-5.8 \cdot 10^{-03}$	$-4.4 \cdot 10^{-03}$
	LightGCN	0.1826	0.1545	0.1830	0.1554	$-4 \cdot 10^{-04}$	$-9 \cdot 10^{-04}$
	SGL*	—	—	—	—	—	—
	UltraGCN	0.1863	0.1580	0.1862	0.1580	$+1 \cdot 10^{-04}$	0
	GFCF	0.1849	0.1518	0.1849	0.1518	0	0
Yelp 2018	NGCF	0.0556	0.0452	0.0579	0.0477	$-2.3 \cdot 10^{-03}$	$-2.5 \cdot 10^{-03}$
	DGCF	0.0621	0.0505	0.0640	0.0522	$-1.9 \cdot 10^{-03}$	$-1.7 \cdot 10^{-03}$
	LightGCN	0.0629	0.0516	0.0649	0.0530	$-2 \cdot 10^{-03}$	$-1.4 \cdot 10^{-03}$
	SGL	0.0669	0.0552	0.0675	0.0555	$-6 \cdot 10^{-04}$	$-3 \cdot 10^{-04}$
	UltraGCN	0.0672	0.0553	0.0683	0.0561	$-1.1 \cdot 10^{-03}$	$-8 \cdot 10^{-04}$
	GFCF	0.0697	0.0571	0.0697	0.0571	0	0
Amazon Book	NGCF	0.0319	0.0246	0.0337	0.0261	$-1.8 \cdot 10^{-03}$	$-1.5 \cdot 10^{-03}$
	DGCF	0.0384	0.0295	0.0399	0.0308	$-1.5 \cdot 10^{-03}$	$-1.3 \cdot 10^{-03}$
	LightGCN	0.0419	0.0323	0.0411	0.0315	$+8 \cdot 10^{-04}$	$+8 \cdot 10^{-04}$
	SGL	0.0474	0.0372	0.0478	0.0379	$-4 \cdot 10^{-04}$	$-7 \cdot 10^{-04}$
	UltraGCN	0.0688	0.0561	0.0681	0.0556	$+7 \cdot 10^{-04}$	$+5 \cdot 10^{-04}$
	GFCF	0.0710	0.0584	0.0710	0.0584	0	0

**Results are not provided since SGL was not originally trained and tested on Gowalla.*

GFCF is the **best replicated model** as it does not implement **any random initialization** of the weights

REPRODUCIBILITY RESULTS

Datasets	Models	Ours		Original		Performance Shift	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Gowalla	NGCF	0.1556	0.1320	0.1569	0.1327	$-1.3 \cdot 10^{-03}$	$-7 \cdot 10^{-04}$
	DGCF	0.1736	0.1477	0.1794	0.1521	$-5.8 \cdot 10^{-03}$	$-4.4 \cdot 10^{-03}$
	LightGCN	0.1826	0.1545	0.1830	0.1554	$-4 \cdot 10^{-04}$	$-9 \cdot 10^{-04}$
	SGL*	—	—	—	—	—	—
	UltraGCN	0.1863	0.1580	0.1862	0.1580	$+1 \cdot 10^{-04}$	0
	GFCF	0.1849	0.1518	0.1849	0.1518	0	0
Yelp 2018	NGCF	0.0556	0.0452	0.0579	0.0477	$-2.3 \cdot 10^{-03}$	$-2.5 \cdot 10^{-03}$
	DGCF	0.0621	0.0505	0.0640	0.0522	$-1.9 \cdot 10^{-03}$	$-1.7 \cdot 10^{-03}$
	LightGCN	0.0629	0.0516	0.0649	0.0530	$-2 \cdot 10^{-03}$	$-1.4 \cdot 10^{-03}$
	SGL	0.0669	0.0552	0.0675	0.0555	$-6 \cdot 10^{-04}$	$-3 \cdot 10^{-04}$
	UltraGCN	0.0672	0.0553	0.0683	0.0561	$-1.1 \cdot 10^{-03}$	$-8 \cdot 10^{-04}$
	GFCF	0.0697	0.0571	0.0697	0.0571	0	0
Amazon Book	NGCF	0.0319	0.0246	0.0337	0.0261	$-1.8 \cdot 10^{-03}$	$-1.5 \cdot 10^{-03}$
	DGCF	0.0384	0.0295	0.0399	0.0308	$-1.5 \cdot 10^{-03}$	$-1.3 \cdot 10^{-03}$
	LightGCN	0.0419	0.0323	0.0411	0.0315	$+8 \cdot 10^{-04}$	$+8 \cdot 10^{-04}$
	SGL	0.0474	0.0372	0.0478	0.0379	$-4 \cdot 10^{-04}$	$-7 \cdot 10^{-04}$
	UltraGCN	0.0688	0.0561	0.0681	0.0556	$+7 \cdot 10^{-04}$	$+5 \cdot 10^{-04}$
	GFCF	0.0710	0.0584	0.0710	0.0584	0	0

**Results are not provided since SGL was not originally trained and tested on Gowalla.*

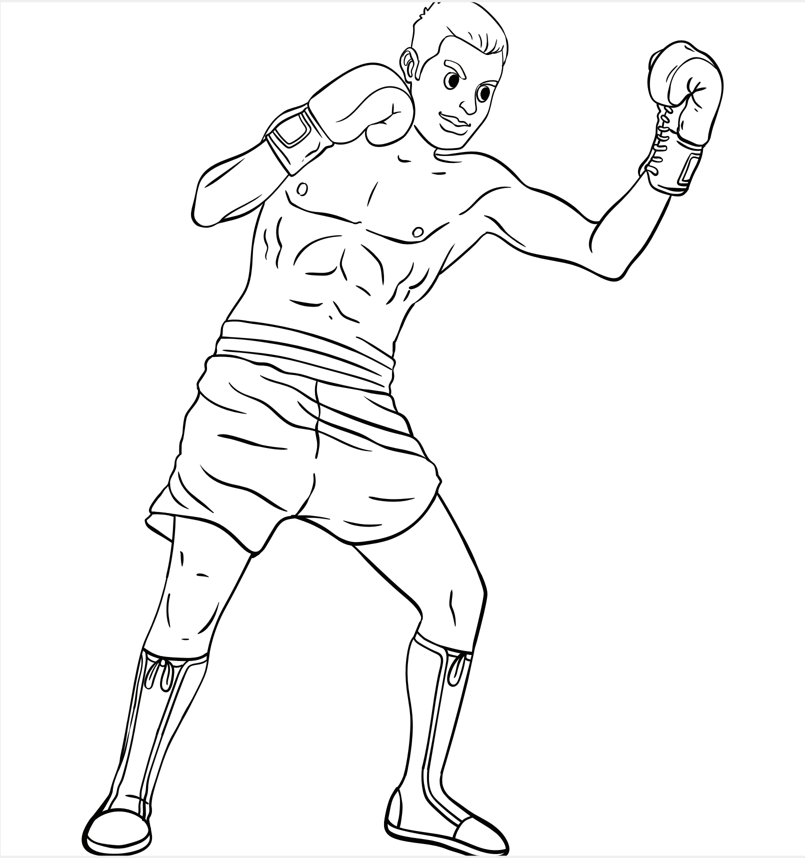
NGCF and DGCF rarely achieve 10e-4 because of **the random initializations** and **stochastic learning processes** involved

REPRODUCIBILITY RESULTS

Datasets	Models	Ours		Original		Performance Shift	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Gowalla	NGCF	0.1556	0.1320	0.1569	0.1327	$-1.3 \cdot 10^{-03}$	$-7 \cdot 10^{-04}$
	DGCF	0.1736	0.1477	0.1794	0.1521	$-5.8 \cdot 10^{-03}$	$-4.4 \cdot 10^{-03}$
	LightGCN	0.1826	0.1545	0.1830	0.1554	$-4 \cdot 10^{-04}$	$-9 \cdot 10^{-04}$
	SGL*	—	—	—	—	—	—
	UltraGCN	0.1863	0.1580	0.1862	0.1580	$+1 \cdot 10^{-04}$	0
	GFCF	0.1849	0.1518	0.1849	0.1518	0	0
Yelp 2018	NGCF	0.0556	0.0452	0.0579	0.0477	$-2.3 \cdot 10^{-03}$	$-2.5 \cdot 10^{-03}$
	DGCF	0.0621	0.0505	0.0640	0.0522	$-1.9 \cdot 10^{-03}$	$-1.7 \cdot 10^{-03}$
	LightGCN	0.0629	0.0516	0.0649	0.0530	$-2 \cdot 10^{-03}$	$-1.4 \cdot 10^{-03}$
	SGL	0.0669	0.0552	0.0675	0.0555	$-6 \cdot 10^{-04}$	$-3 \cdot 10^{-04}$
	UltraGCN	0.0672	0.0553	0.0683	0.0561	$-1.1 \cdot 10^{-03}$	$-8 \cdot 10^{-04}$
	GFCF	0.0697	0.0571	0.0697	0.0571	0	0
Amazon Book	NGCF	0.0319	0.0246	0.0337	0.0261	$-1.8 \cdot 10^{-03}$	$-1.5 \cdot 10^{-03}$
	DGCF	0.0384	0.0295	0.0399	0.0308	$-1.5 \cdot 10^{-03}$	$-1.3 \cdot 10^{-03}$
	LightGCN	0.0419	0.0323	0.0411	0.0315	$+8 \cdot 10^{-04}$	$+8 \cdot 10^{-04}$
	SGL	0.0474	0.0372	0.0478	0.0379	$-4 \cdot 10^{-04}$	$-7 \cdot 10^{-04}$
	UltraGCN	0.0688	0.0561	0.0681	0.0556	$+7 \cdot 10^{-04}$	$+5 \cdot 10^{-04}$
	GFCF	0.0710	0.0584	0.0710	0.0584	0	0

**Results are not provided since SGL was not originally trained and tested on Gowalla.*

Replicability is ensured!



03

GNNs-BASED *vs.* TRADITIONAL RECSYS

WHY COMPARING TRADITIONAL RECSYS?

Families	Baselines	Models					
		NGCF [71]	DGCF [73]	LightGCN [28]	SGL [78]	UltraGCN [47]	GFCF [59]
		Used as graph CF baseline in (2021 — present)					
		[10, 13, 32, 62, 77, 84]	[19, 39, 46, 74, 75, 92]	[40, 54, 78, 82, 88, 89]	[22, 46, 77, 82, 85, 93]	[17, 24, 42, 48, 95, 96]	[4, 5, 41, 50, 80, 96]
Classic CF	MF-BPR [55]	✓	✓			✓	
	NeuMF [29]	✓					
	CMN [18]	✓					
	MacridVAE [44]		✓				
	Mult-VAE [38]			✓	✓		✓
	DNN+SSL [86]				✓		
	ENMF [11]					✓	
	CML [30]					✓	
	DeepWalk [52]					✓	
	LINE [66]					✓	
	Node2Vec [25]					✓	
	NBPO [91]					✓	

Most of the approaches are compared **against a small subset of classical CF** solutions. However, the **recent literature** has **raised concerns** about **usually-untested strong CF baselines!**

WHY COMPARING TRADITIONAL RECSYS?

Families	Baselines	Models					
		NGCF [71]	DGCF [73]	LightGCN [28]	SGL [78]	UltraGCN [47]	GFCF [59]
		Used as graph CF baseline in (2021 — present)					
		[10, 13, 32, 62, 77, 84]	[19, 39, 46, 74, 75, 92]	[40, 54, 78, 82, 88, 89]	[22, 46, 77, 82, 85, 93]	[17, 24, 42, 48, 95, 96]	[4, 5, 41, 50, 80, 96]
Graph CF	HOP-Rec [83]	✓					
	GC-MC [68]	✓	✓				
	PinSage [87]	✓					
	NGCF [71]		✓	✓	✓	✓	✓
	DisenGCN [43]		✓				
	GRMF [53]			✓			✓
	GRMF-Norm [28]			✓			✓
	NIA-GCN [64]					✓	
	LightGCN [28]				✓	✓	✓
	DGCF [73]					✓	
	LR-GCCF [14]					✓	
	SCF [94]					✓	
	BGCF [63]					✓	
	LCFN [90]					✓	

Orange ticks indicate that **no extensive comparison among the selected graph CF** baselines is performed (**mainly for chronological reasons**)

WHAT WE HAVE DONE

1) Expand the investigation

to **four classical CF** recommender systems:
UserkNN [Resnick et al.], ItemkNN [Sarwar et al.], $RP^3\beta$ [Paudel et al.], $EASE^R$ [Harald Steck]

2) Fine-tune classic CF models

by retaining the **10% of the training** set as validation for **fair comparison** to graph CF

3) Use the TPE algorithm

which is a **strong strategy** for **hyper-parameter search**

4) Compare to unpersonalized

recommendation approaches such as **MostPop and Random**

COMPARISON RESULTS

Families	**Models**	**Gowalla**		**Yelp 2018**		**Amazon Book**	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Reference	MostPop	0.0416	0.0316	0.0125	0.0101	0.0051	0.0044
	Random	0.0005	0.0003	0.0005	0.0004	0.0002	0.0002
Classic CF	UserkNN	0.1685	0.1370	0.0630	0.0528	0.0582	0.0477
	ItemkNN	0.1409	0.1165	0.0610	0.0507	0.0634	0.0524
	$RP^3\beta$	0.1829	0.1520	0.0671	<u>0.0559</u>	0.0683	0.0565
	$EASE^R$*	0.1661	0.1384	0.0655	0.0552	**0.0710**	<u>0.0567</u>
Graph CF	NGCF	0.1556	0.1320	0.0556	0.0452	0.0319	0.0246
	DGCF	0.1736	0.1477	0.0621	0.0505	0.0384	0.0295
	LightGCN	0.1826	<u>0.1545</u>	0.0629	0.0516	0.0419	0.0323
	SGL	—	—	0.0669	0.0552	0.0474	0.0372
	UltraGCN	**0.1863**	**0.1580**	<u>0.0672</u>	0.0553	<u>0.0688</u>	0.0561
	GFCF	<u>0.1849</u>	0.1518	**0.0697**	**0.0571**	**0.0710**	**0.0584**

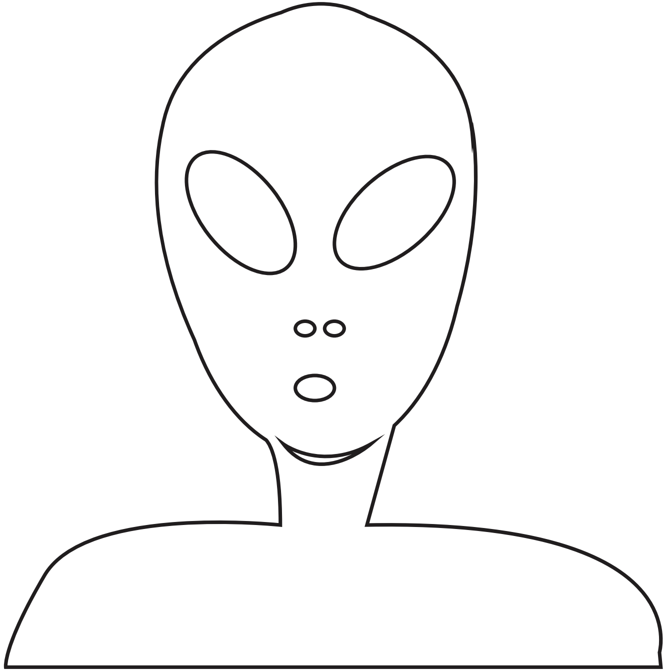
**Results for $EASE^R$ on Amazon Book are taken from <u>BARS Benchmark</u>.*

COMPARISON RESULTS

Families	Models	Gowalla		Yelp 2018		Amazon Book	
		Recall	nDCG	Recall	nDCG	Recall	nDCG
Reference	MostPop	0.0416	0.0316	0.0125	0.0101	0.0051	0.0044
	Random	0.0005	0.0003	0.0005	0.0004	0.0002	0.0002
Classic CF	UserkNN	0.1685	0.1370	0.0630	0.0528	0.0582	0.0477
	ItemkNN	0.1409	0.1165	0.0610	0.0507	0.0634	0.0524
	$RP^3\beta$	0.1829	0.1520	0.0671	0.0559	0.0683	0.0565
	$EASE^{R}*$	0.1661	0.1384	0.0655	0.0552	**0.0710**	0.0567
Graph CF	NGCF	0.1556	0.1320	0.0556	0.0452	0.0319	0.0246
	DGCF	0.1736	0.1477	0.0621	0.0505	0.0384	0.0295
	LightGCN	0.1826	0.1545	0.0629	0.0516	0.0419	0.0323
	SGL	—	—	0.0669	0.0552	0.0474	0.0372
	UltraGCN	**0.1863**	**0.1580**	0.0672	0.0553	0.0688	0.0561
	GFCF	0.1849	0.1518	**0.0697**	**0.0571**	**0.0710**	**0.0584**

**Results for $EASE^{R}$ on Amazon Book are taken from BARS Benchmark.*

On **Yelp-2018** and **Amazon Book**, **classic CF** approaches are **best or second-to-best approaches**



04

EXPLORING UNCOMMON DATASETS

WHY CONSIDERING OTHER DATASETS?

Models	Gowalla	Yelp 2018	Amazon Book	Alibaba-iFashion	Movielens 1M	Amazon Electronics	Amazon CDs
NGCF	✓	✓	✓				
DGCF	✓	✓	✓				
LightGCN	✓	✓	✓				
SGL		✓	✓	✓			
UltraGCN	✓	✓	✓		✓	✓	✓
GFCF	✓	✓	✓				

A **limited set of shared datasets**, so we decide to **consider un-explored and novel** datasets with **specific topological properties**

TWO NEW DATASETS

Statistics	**Gowalla**	**Yelp 2018**	**Amazon Book**	**Allrecipes**	**BookCrossing**
Users	29,858	31,668	52,643	10,084	6,754
Items	40,981	38,048	91,599	8,407	13,670
Edges	810,128	1,237,259	2,380,730	80,540	234,762
Density	0.0007	0.0010	0.0005	0.0010	0.0025
Avg. Deg. (U)	27.1327	39.0697	45.2241	7.9869	34.7590
Avg. Deg. (I)	19.7684	32.5184	25.9908	9.5801	17.1735

Two new datasets (**Allrecipes [Gao et al.] and BookCrossing [Ziegler et al.]**) which **have discordant characteristics** to **Gowalla, Yelp 2018, and Amazon Book**

TWO NEW DATASETS

Statistics	Gowalla	Yelp 2018	Amazon Book	**Allrecipes**	BookCrossing
Users	29,858	31,668	52,643	10,084	6,754
Items	40,981	38,048	91,599	8,407	13,670
Edges	810,128	1,237,259	2,380,730	80,540	234,762
Density	0.0007	0.0010	0.0005	0.0010	0.0025
Avg. Deg. (U)	27.1327	39.0697	45.2241	7.9869	34.7590
Avg. Deg. (I)	19.7684	32.5184	25.9908	9.5801	17.1735

Users are **more numerous** than **items**; there is a **much lower average** of users and items **node degree**

TWO NEW DATASETS

Statistics	Gowalla	Yelp 2018	Amazon Book	Allrecipes	BookCrossing
Users	29,858	31,668	52,643	10,084	6,754
Items	40,981	38,048	91,599	8,407	13,670
Edges	810,128	1,237,259	2,380,730	80,540	234,762
Density	0.0007	0.0010	0.0005	0.0010	0.0025
Avg. Deg. (U)	27.1327	39.0697	45.2241	7.9869	34.7590
Avg. Deg. (I)	19.7684	32.5184	25.9908	9.5801	17.1735

Lowest ratio between **users and items**; much **higher density** than the other datasets

COMPARISON RESULTS

Families	Models	Allrecipes		BookCrossing	
		Recall	nDCG	Recall	nDCG
Reference	MostPop	0.0472	0.0242	0.0352	0.0319
	Random	0.0024	0.0010	0.0013	0.0011
Classic CF	UserkNN	0.0339	0.0188	0.0871	0.0769
	ItemkNN	0.0326	0.0180	0.0779	0.0739
	$RP^3\beta$	0.0170	0.0089	**0.0941**	0.0821
	$EASE^R$	0.0351	0.0192	0.0925	**0.0847**
Graph CF	NGCF	0.0291	0.0144	0.0670	0.0546
	DGCF	0.0448	0.0234	0.0643	0.0543
	LightGCN	0.0459	0.0236	0.0803	0.0660
	SGL	0.0365	0.0192	0.0716	0.0600
	UltraGCN	**0.0475**	**0.0248**	0.0800	0.0651
	GFCF	0.0101	0.0051	0.0819	0.0712

COMPARISON RESULTS

Families	Models	Allrecipes		BookCrossing	
		Recall	nDCG	Recall	nDCG
Reference	MostPop	<u>0.0472</u>	<u>0.0242</u>	0.0352	0.0319
	Random	0.0024	0.0010	0.0013	0.0011
Classic CF	UserkNN	0.0339	0.0188	0.0871	0.0769
	ItemkNN	0.0326	0.0180	0.0779	0.0739
	$RP^3\beta$	0.0170	0.0089	**0.0941**	<u>0.0821</u>
	$EASE^R$	0.0351	0.0192	<u>0.0925</u>	**0.0847**
Graph CF	NGCF	0.0291	0.0144	0.0670	0.0546
	DGCF	0.0448	0.0234	0.0643	0.0543
	LightGCN	0.0459	0.0236	0.0803	0.0660
	SGL	0.0365	0.0192	0.0716	0.0600
	UltraGCN	**0.0475**	**0.0248**	0.0800	0.0651
	GFCF	0.0101	0.0051	0.0819	0.0712

Classic CF approaches are **very competitive**, especially **on BookCrossing**!

COMPARISON RESULTS

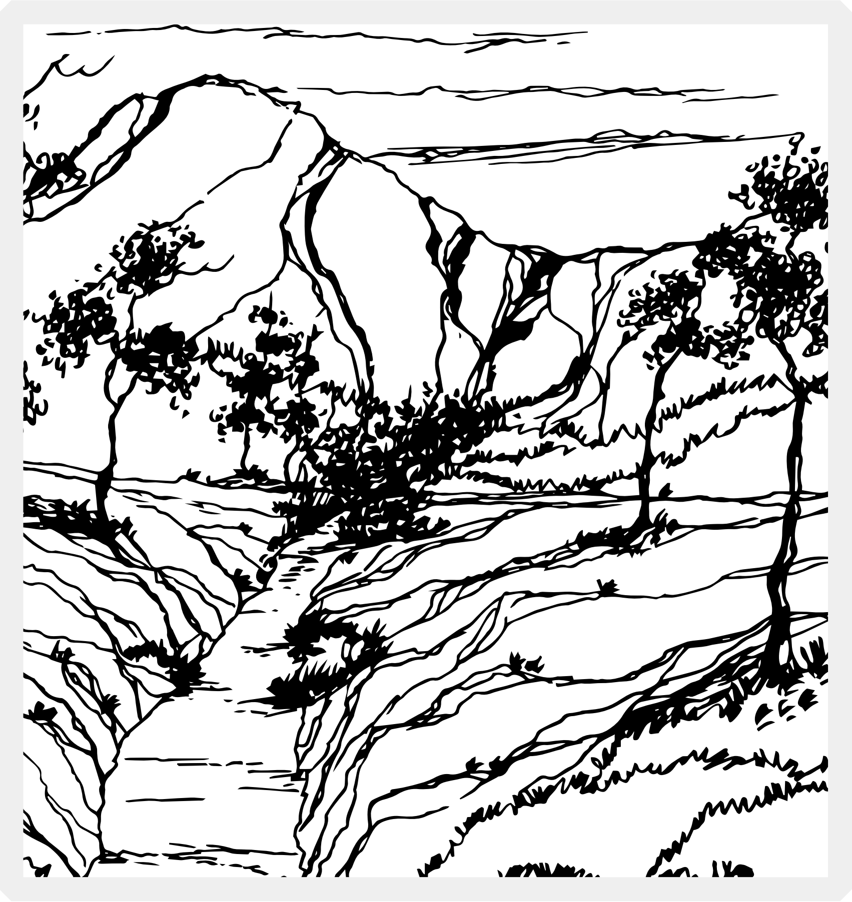
Families	Models	Allrecipes		BookCrossing	
		Recall	nDCG	Recall	nDCG
Reference	MostPop	0.0472	0.0242	0.0352	0.0319
	Random	0.0024	0.0010	0.0013	0.0011
Classic CF	UserkNN	0.0339	0.0188	0.0871	0.0769
	ItemkNN	0.0326	0.0180	0.0779	0.0739
	$RP^3\beta$	0.0170	0.0089	**0.0941**	0.0821
	$EASE^R$	0.0351	0.0192	0.0925	**0.0847**
Graph CF	NGCF	0.0291	0.0144	0.0670	0.0546
	DGCF	0.0448	0.0234	0.0643	0.0543
	LightGCN	0.0459	0.0236	0.0803	0.0660
	SGL	0.0365	0.0192	0.0716	0.0600
	UltraGCN	**0.0475**	**0.0248**	0.0800	0.0651
	GFCF	0.0101	0.0051	0.0819	0.0712

The **performance** of **graph CF** significantly **drops**

COMPARISON RESULTS

Families	**Models**	**Allrecipes**		**BookCrossing**	
		Recall	nDCG	Recall	nDCG
Reference	MostPop	0.0472	0.0242	0.0352	0.0319
	Random	0.0024	0.0010	0.0013	0.0011
Classic CF	UserkNN	0.0339	0.0188	0.0871	0.0769
	ItemkNN	0.0326	0.0180	0.0779	0.0739
	$RP^3\beta$	0.0170	0.0089	**0.0941**	0.0821
	$EASE^R$	0.0351	0.0192	0.0925	**0.0847**
Graph CF	NGCF	0.0291	0.0144	0.0670	0.0546
	DGCF	0.0448	0.0234	0.0643	0.0543
	LightGCN	0.0459	0.0236	0.0803	0.0660
	SGL	0.0365	0.0192	0.0716	0.0600
	UltraGCN	**0.0475**	**0.0248**	0.0800	0.0651
	GFCF	0.0101	0.0051	0.0819	0.0712

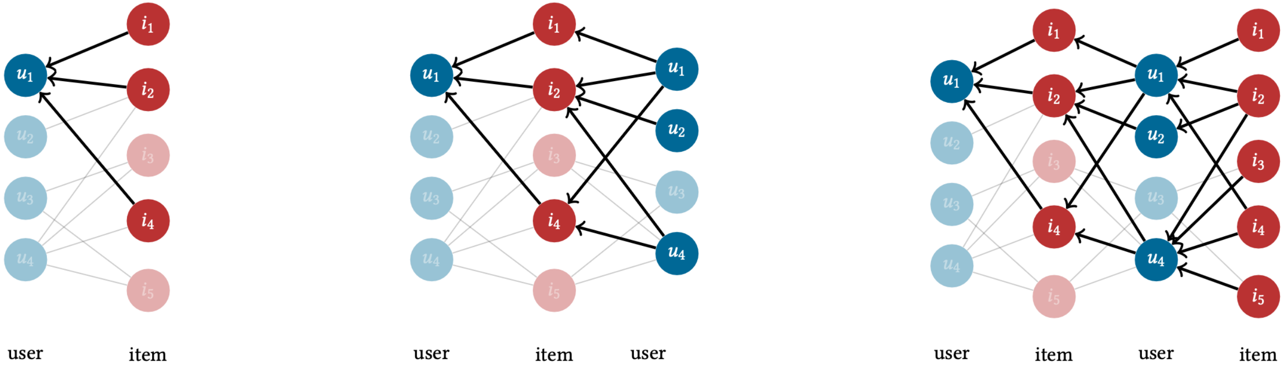
Only **LightGCN** and **UltraGCN** **keep their performance** up!



05

INFORMATION FLOW IN GNNs-BASED RECSYS

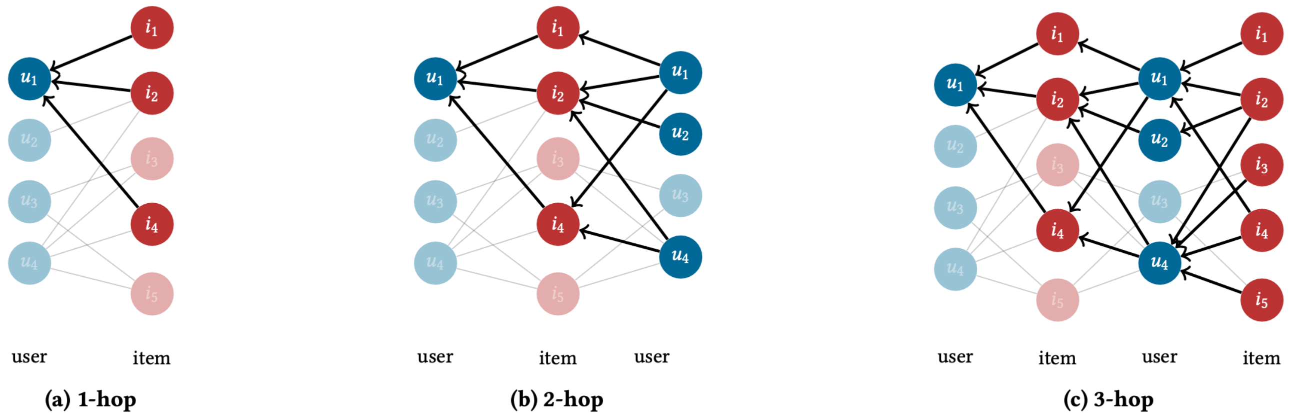
NODE DEGREE AS INFORMATION FLOW



(a) 1-hop (b) 2-hop (c) 3-hop

Node degree as information flow from the neighborhood nodes to the ego node after multiple hops

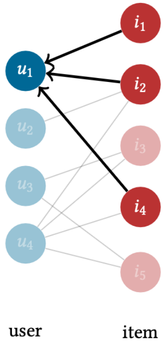
NODE DEGREE AS INFORMATION FLOW



Node degree as information flow from the neighborhood nodes to the ego node after multiple hops

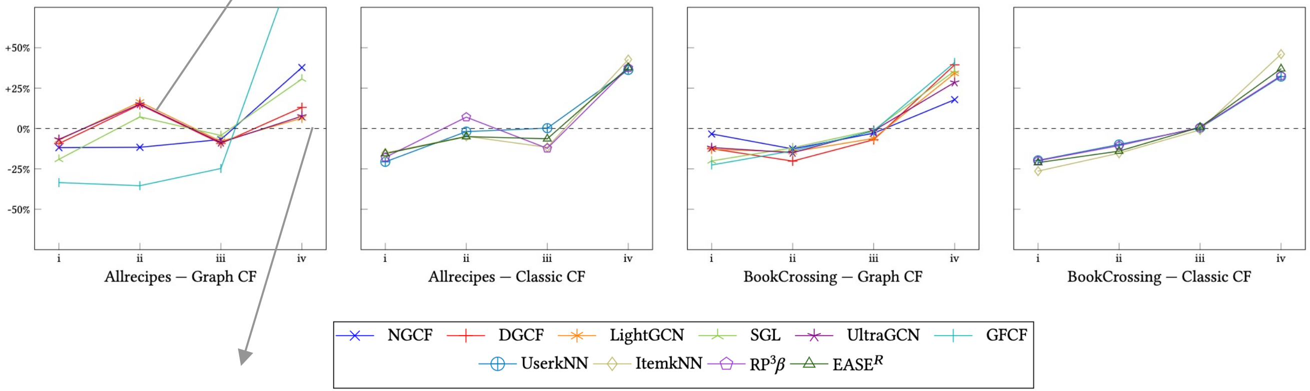
$$\Upsilon_{\mathcal{U}}^{(1)} = \mathbf{R}\mathbf{1}_{\mathcal{I}}, \qquad \Upsilon_{\mathcal{U}}^{(2)} = (\mathbf{R} \odot (\mathbf{1}_{\mathcal{U}}\mathbf{R}))\mathbf{1}_{\mathcal{I}}, \qquad \Upsilon_{\mathcal{U}}^{(3)} = (\mathbf{R}\mathbf{R}^\top \odot \mathbf{R}\mathbf{1}_{\mathcal{I}})\mathbf{1}_{\mathcal{I}},$$

FIRST HOP



(a) 1-hop

Improvement

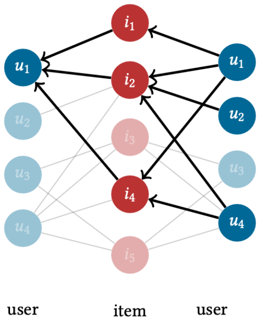


Average performance

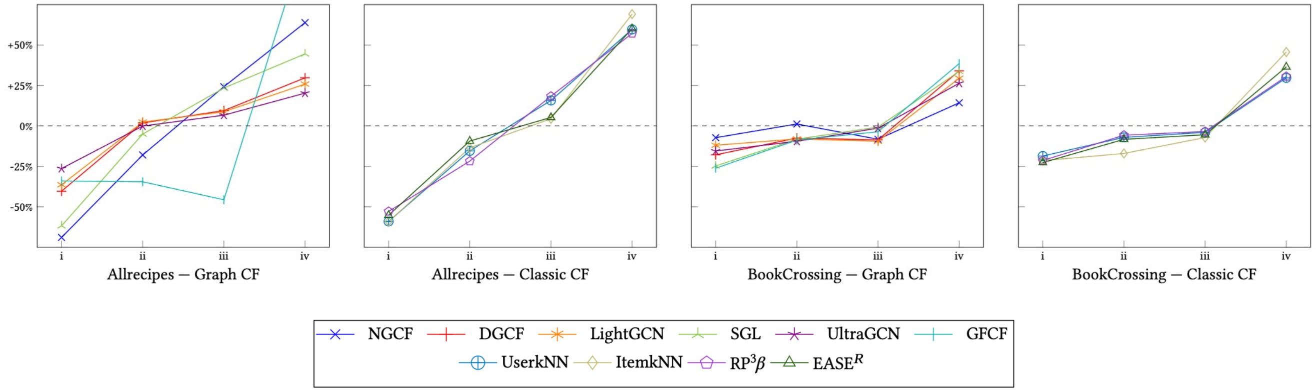
Indication of user's activeness on the platform

- The **fourth quartile** is **favoured** over the other ones
- The **trend is even more** evident on **GFCF**

SECOND HOP



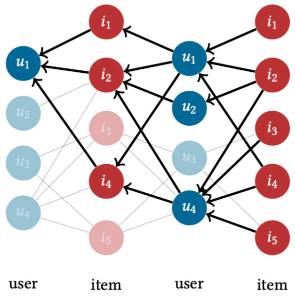
(b) 2-hop



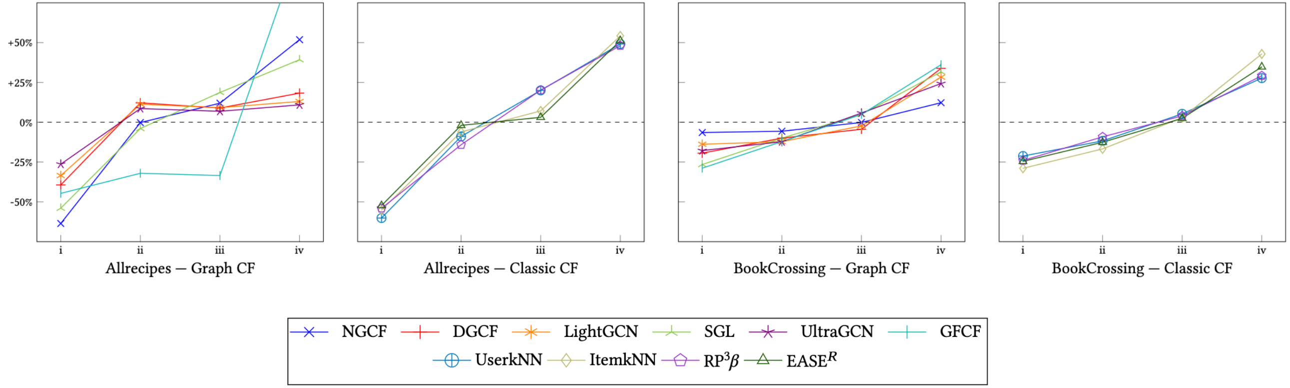
Indication of the influence of items' popularity on users

- Models **favour** the **warm users who enjoyed popular items**
- On **Allrecipes**, **UltraGCN, DGCF, and LightGCN** show a **less discriminatory behaviour** across quartiles
- On **BookCrossing**, the trend is **almost aligned** across models

THIRD HOP

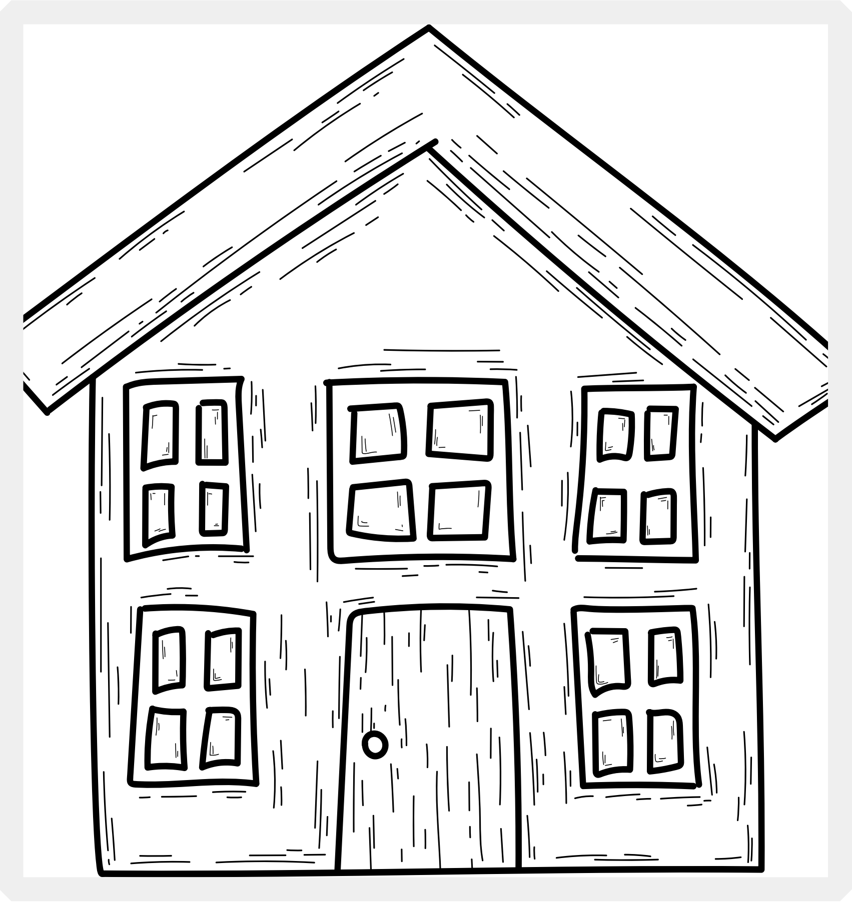


(c) 3-hop



Indication of the influence of co-interacting users' activeness on users

- On **Allrecipes**, **UltraGCN, DGCF, and LightGCN** exhibit **more consistency** across quartiles
- On **BookCrossing**, the 3-hop is **not providing more information** than the **2-hop**



06

TAKE-HOME MESSAGES

WHAT WE HAVE LEARNED

01

GNNs-BASED RECSYS IN ELLIOT

WHAT WE HAVE LEARNED

01

GNNs-BASED RECSYS IN ELLIOT

02

REPRODUCIBILITY OF GNNs-BASED RECSYS

WHAT WE HAVE LEARNED

01

GNNs-BASED RECSYS IN ELLIOT

02

REPRODUCIBILITY OF GNNs-BASED RECSYS

03

COMPARISON TO TRADITIONAL RECSYS

WHAT WE HAVE LEARNED

01

GNNs-BASED RECSYS IN ELLIOT

02

REPRODUCIBILITY OF GNNs-BASED RECSYS

03

COMPARISON TO TRADITIONAL RECSYS

04

COMPARISON ON NEW DATASETS

WHAT WE HAVE LEARNED

01

GNNs-BASED RECSYS IN ELLIOT

02

REPRODUCIBILITY OF GNNs-BASED RECSYS

03

COMPARISON TO TRADITIONAL RECSYS

04

COMPARISON ON NEW DATASETS

05

NODE DEGREE AS INFORMATION FLOW

REFERENCES

[Anelli et al.] Challenging the Myth of Graph Collaborative Filtering: a Reasoned and Reproducibility-driven Analysis. RecSys 2023: 350-361

[Malitesta et al.] An Out-of-the-Box Application for Reproducible Graph Collaborative Filtering extending the Elliot Framework. UMAP (Adjunct Publication) 2023: 12-15

[Resnick et al.] GroupLens: An Open Architecture for Collaborative Filtering of Netnews. CSCW 1994: 175-186

[Sarwar et al.] Item-based collaborative filtering recommendation algorithms. WWW 2001: 285-295

[Paudel et al.] Updatable, Accurate, Diverse, and Scalable Recommendations for Interactive Applications. ACM Trans. Interact. Intell. Syst. 7(1): 1:1-1:34 (2017)

[Harald Steck] Embarrassingly Shallow Autoencoders for Sparse Data. WWW 2019: 3251-3257

[Gao et al.] Hierarchical Attention Network for Visually-Aware Food Recommendation. IEEE Trans. Multim. 22(6): 1647-1659 (2020)

[Ziegler et al.] Improving recommendation lists through topic diversification. WWW 2005: 22-32

THANKS!

Scan me!