The 3rd Workshop on Graph Learning Benchmarks
Long Beach, CA, USA 06-08-2023
KDD 2023 - Workshops

An Out-of-the-Box Application for Reproducible Graph Collaborative Filtering extending the Elliot Framework

Daniele Malitesta, Claudio Pomo, Vito Walter Anelli, Tommaso Di Noia, Antonio Ferrara

Politecnico di Bari

Bari, Italy

email: firstname.lastname@poliba.it

Politecnico di Bari

SISINF LAB

Disclaimer

This work has been presented as demonstration [*] at **UMAP 2023**.

preprint

[*] **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.

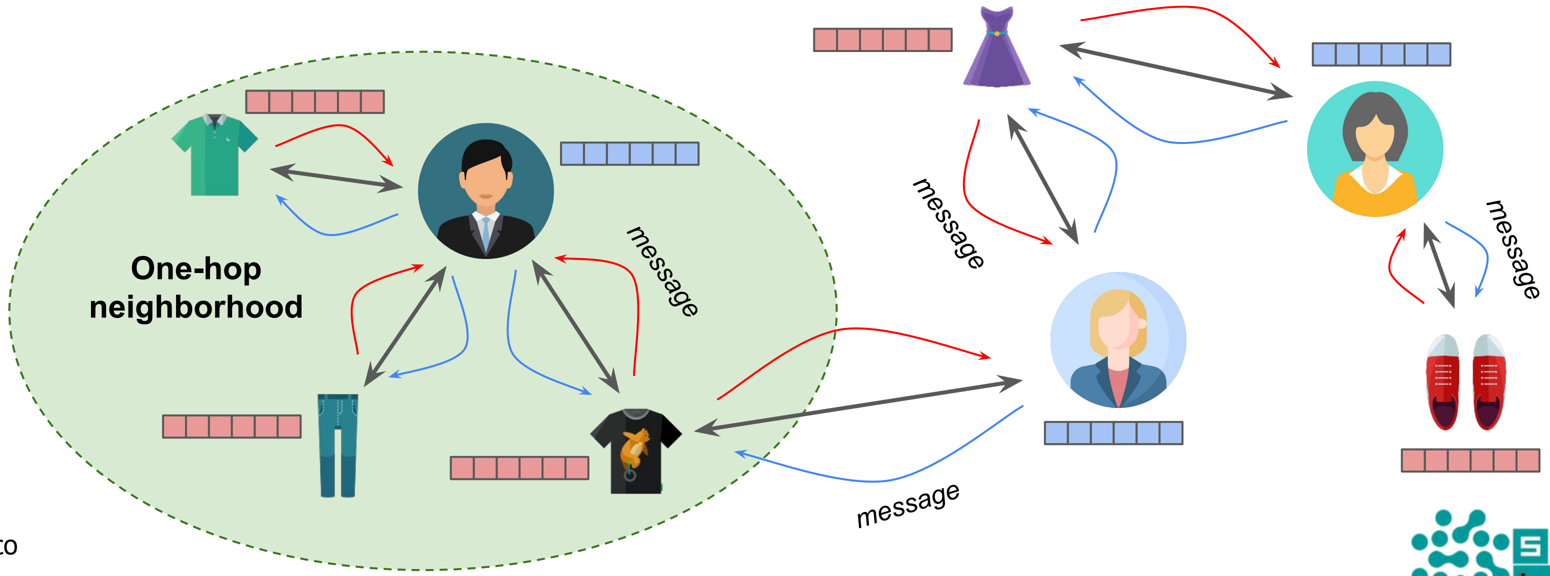
Politecnico di Bari

SISINF LAB

Introduction and Motivation

Graph collaborative filtering

In collaborative filtering (CF), graph convolutional networks (GCNs) have gained momentum thanks to their ability to **aggregate neighbor** nodes **information** into ego nodes at multiple hops (i.e., **message-passing**), thus effectively **distilling** the **collaborative signal**



Reproducibility and evaluation in graph CF

- Great effort has been put into building **frameworks** for **graph**-based **recommendation** (but only **recently**!)
- This is important to foster the **reproducibility** and the **fair evaluation** of graph collaborative filtering approaches in the literature
- Two noticeable mentions: **RecBole** [*] and **BARS** [**]
- **RecBole**: 8 graph-based models for general recommendation
- **BARS**: they add 5 more graph-based approaches

[*] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, Yingqian Min, Zhichao Feng, Xinyan Fan, Xu Chen, Pengfei Wang, Wendi Ji, Yaliang Li, Xiaoling Wang, Ji-Rong Wen: *RecBole: Towards a Unified, Comprehensive and Efficient Framework for Recommendation Algorithms*. CIKM 2021: 4653-4664.

[**] Jieming Zhu, Quanyu Dai, Liangcai Su, Rong Ma, Jinyang Liu, Guohao Cai, Xi Xiao, Rui Zhang: *BARS: Towards Open Benchmarking for Recommender Systems*. SIGIR 2022: 2912-2923.

So, why another framework for graph CF?

Our contributions

- We add **PyTorch Geometric** (PyG) to **Elliot** [*], our framework for the **rigorous reproducibility** of recommender systems; to date, **few** other frameworks have started **integrating PyG** into their **pipelines**.
- We **address** the **reproducibility issues** in PyG through **sparse** adjacency matrices.
- We propose a **twofold** models **categorization** [**][***]: graph-based approaches leveraging **explicit** and **implicit** message-passing.
- We provide a **Docker** image with the **CUDA** environment already set (**first** framework offering this).

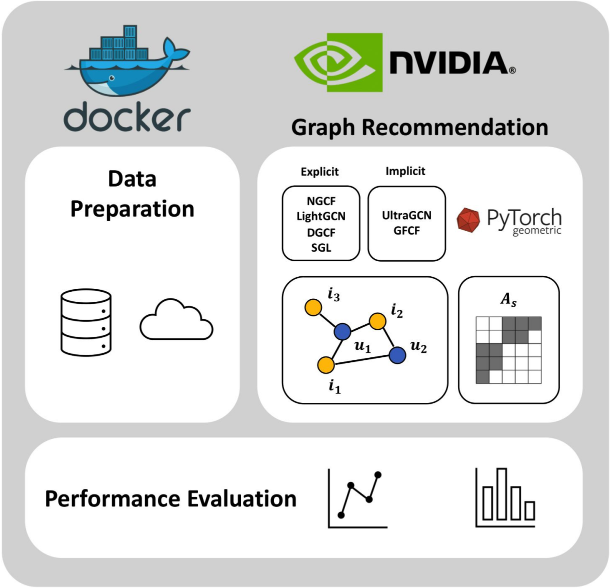
[*] **Vito Walter Anelli**, Alejandro Bellogín, Antonio Ferrara, **Daniele Malitesta**, Felice Antonio Merra, **Claudio Pomo**, Francesco Maria Donini, Tommaso Di Noia: *Elliot: A Comprehensive and Rigorous Framework for Reproducible Recommender Systems Evaluation*. SIGIR 2021: 2405-2414.

[**] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, Eugenio Di Sciascio, Antonio Ferrara, **Daniele Malitesta**, **Claudio Pomo**: *How Neighborhood Exploration influences Novelty and Diversity in Graph Collaborative Filtering*. MORS@RecSys 2022.

[***] Vito Walter Anelli, Yashar Deldjoo, Tommaso Di Noia, **Daniele Malitesta**, Vincenzo Paparella, **Claudio Pomo**: *Auditing Consumer- and Producer-Fairness in Graph Collaborative Filtering*. ECIR (1) 2023: 33-48.

Graph collaborative filtering in Elliot

Overall architecture and reproducibility issues



- **Explicit** (i.e., NGCF, LightGCN, DGCF, SGL) and **implicit** (i.e., UltraGCN, GFCF) message-passing.
- The **`MessagePassing`** class allows to implement any (**explicit**) message-passing formulation.
- Reproducibility issues in the **`scatter`** [*] method (called by **`aggregate`**).
- We solve [**] with the sparse adjacency matrix and the matrix formulation of the message-passing (this calls **`message_and_aggregate`** which does not use **`scatter`**).

$$\mathbf{E}^{(l)} = \mathbf{A}_s \mathbf{E}^{(l-1)}$$

user/item embedding

sparse adjacency matrix (e.g., COO format)

[*] https://github.com/pyg-team/pytorch_geometric/issues/2788.

[**] https://pytorch-geometric.readthedocs.io/en/latest/notes/sparse_tensor.html.

Politecnico di Bari

Dockerization

We provide a **ready-to-use** environment equipped with **CUDA**, **cuDNN**, **Python**, and all the required packages to run our framework.

No need to install everything **from scratch** and find the **one** configuration where all libraries and packages **work smoothly** 😎 🚀

Operating System	Ubuntu 20.04
CUDA	11.6.2
cuDNN	8.0
Python	3.8
Volume mounted at	./results/

scan me!

Politecnico di Bari

An experimental workflow

Run our framework locally…

```
$ git clone https://github.com/sisinflab/Graph-Demo.git
$ cd Graph-Demo/
$ sudo docker compose run demo-graph
```

… or on Google Colab

scan me!

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.

Politecnico di Bari

KDD2023
LONG BEACH, CA
AUGUST 6 - 10

Find us on GitHub

GitHub

Demo

Elliot

Politecnico
di Bari

SISINF
LAB

Wait, is that everything??

Of course not 😎

Come and meet us at RecSys 2023!

Challenging the Myth of Graph Collaborative Filtering: a Reasoned and Reproducibility-driven Analysis

VITO WALTER ANELLI*, Politecnico di Bari, Italy
DANIELE MALITESTA*, Politecnico di Bari, Italy
CLAUDIO POMO*, Politecnico di Bari, Italy, Italy
ALEJANDRO BELLOGÍN, Universidad Autónoma de Madrid, Spain
EUGENIO DI SCIASCIO, Politecnico di Bari, Italy
TOMMASO DI NOIA, Politecnico di Bari, Italy

The success of graph neural network-based models (GNNs) has significantly advanced recommender systems by effectively modeling users and items as a bipartite, undirected graph. However, many original graph-based works often adopt results from baseline papers without verifying their validity for the specific configuration under analysis. Our work addresses this issue by focusing on the replicability of results. We present a code that successfully replicates results from six popular and recent graph recommendation models (NGCF, DGCF, LightGCN, SGL, UltraGCN, and GFCF) on three common benchmark datasets (Gowalla, Yelp 2018, and Amazon Book). Additionally, we compare these graph models with traditional collaborative filtering models that historically performed well in

our brand-new
reproducibility paper
<u>accepted at RecSys 2023</u>
is out on arXiv!

RecSys
acm.org

Singapore, 18-22 September 2023

GitHub

Preprint

Politecnico di Bari

SISINF LAB

Thank you! Any questions?