

Table 9: Recommendation performance (i.e., NDCG@10 and Recall@10) of all models across users with different numbers of interactions. Setting explored in this table is the same as what Table 2 has.

Method	NGCF	LightGCN	ENMF	+TAG-CF	Impr. (↑)	MF	+TAG-CF	Impr. (↑%)	UltraGCN	+TAG-CF	Impr. (↑%)
NDCG@10 – LOW-DEGREE USERS (LOWER PERCENTILE)											
Amazon-Book	3.33±0.09	5.05±0.11	3.32±0.02	3.54±0.03	6.8%	5.01±0.08	5.19±0.06	3.7%	3.49±0.17	3.80±0.24	8.7%
Anime	7.10±0.18	9.77±0.17	7.88±0.21	8.03±0.12	1.9%	8.48±0.07	9.68±0.03	14.2%	9.98±0.16	10.69±0.24	7.0%
Gowalla	5.54±0.07	6.59±0.12	2.54±0.17	2.68±0.09	5.4%	6.54±0.07	6.73±0.04	2.9%	5.38±0.10	5.70±0.11	6.0%
Yelp-2018	2.80±0.07	3.51±0.10	1.80±0.08	1.89±0.04	5.2%	3.52±0.09	3.57±0.05	1.6%	2.82±0.10	3.16±0.14	12.1%
MovieLens-1M	15.14±0.30	17.81±0.32	12.55±0.19	15.55±0.22	23.9%	14.44±0.11	20.11±0.17	39.2%	16.39±0.17	19.54±0.23	19.2%
NDCG@10 – OVERALL											
Amazon-Book	4.35±0.13	5.03±0.10	3.81±0.11	4.09±0.08	7.4%	4.98±0.03	5.08±0.03	1.9%	3.60±0.25	3.83±0.23	6.6%
Anime	7.99±0.27	9.92±0.24	10.63±0.10	10.97±0.11	3.1%	8.52±0.05	9.73±0.03	14.2%	10.66±0.10	10.99±0.09	3.1%
Gowalla	5.65±0.11	6.54±0.11	3.42±0.05	3.49±0.05	2.1%	6.41±0.07	6.53±0.04	1.9%	5.59±0.14	5.93±0.14	6.1%
Yelp-2018	3.19±0.05	3.64±0.05	2.18±0.11	2.26±0.05	3.6%	3.59±0.07	3.67±0.03	2.3%	2.88±0.11	3.21±0.10	11.5%
MovieLens-1M	15.94±0.20	18.24±0.27	14.12±0.17	15.80±0.21	11.9%	15.47±0.11	20.51±0.18	32.6%	18.23±0.13	20.43±0.25	12.1%
RECALL@10 – LOW-DEGREE USERS (LOWER PERCENTILE)											
Amazon-Book	3.75±0.12	4.64±0.19	3.65±0.14	3.93±0.13	7.6%	4.57±0.10	4.74±0.10	3.8%	2.77±0.15	2.93±0.12	5.8%
Anime	10.10±0.28	12.84±0.24	14.64±0.64	15.15±0.57	3.5%	11.41±0.08	12.61±0.06	10.5%	13.34±0.25	14.52±0.30	8.8%
Gowalla	7.22±0.28	7.88±0.17	3.59±0.06	3.74±0.07	4.2%	7.82±0.21	7.96±0.13	1.8%	6.40±0.15	6.66±0.17	4.2%
Yelp-2018	3.45±0.12	3.64±0.13	2.44±0.06	2.57±0.10	5.5%	3.60±0.12	3.73±0.16	3.6%	2.62±0.16	2.92±0.19	11.5%
MovieLens-1M	6.60±0.15	7.54±0.18	5.73±0.11	7.07±0.14	23.4%	6.90±0.18	8.22±0.22	19.2%	7.66±0.22	8.54±0.23	11.5%
RECALL@10 – OVERALL											
Amazon-Book	3.62±0.22	4.46±0.17	3.81±0.17	3.99±0.08	4.8%	4.44±0.06	4.56±0.07	2.6%	2.81±0.24	3.00±0.26	6.7%
Anime	11.12±0.18	12.86±0.22	13.45±0.26	13.71±0.26	1.9%	11.43±0.08	12.71±0.04	11.2%	14.19±0.46	14.64±0.39	3.2%
Gowalla	7.43±0.06	7.70±0.12	4.01±0.07	4.04±0.08	0.8%	7.57±0.18	7.72±0.09	2.0%	6.56±0.17	6.77±0.21	3.4%
Yelp-2018	3.39±0.06	3.73±0.09	2.32±0.09	2.41±0.03	3.7%	3.67±0.12	3.80±0.11	3.6%	2.83±0.17	3.36±0.24	18.7%
MovieLens-1M	7.00±0.13	7.66±0.20	6.20±0.23	7.02±0.25	13.2%	7.68±0.12	8.33±0.16	8.4%	7.93±0.20	8.75±0.22	10.3%

Table 10: Recommendation performance (i.e., NDCG@10/20 and Recall@10/20) of efficient matrix factorization methods and performance improvement (%) of TAG-CF over them. Setting explored in this table is the same as what Table 2 has. For methods that directly outputs ranking predictions per user, we use an user-derived propagation matrix to conduct test-time augmentation.

Metrics	GF-CF	+TAG-CF	Impr. (↑)	SVD-GCN-S	+TAG-CF	Impr. (↑%)	SVD-AE	+TAG-CF	Impr. (↑%)	Turbo-CF	+TAG-CF	Impr. (↑%)
AMAZON-BOOK												
Recall@10	7.31	7.93	8.5%	6.72	6.88	2.5%	6.74	7.23	7.3%	7.37	7.70	4.4%
NDCG@10	4.37	4.75	8.6%	3.99	4.10	2.6%	3.97	4.24	6.8%	4.40	4.58	4.0%
Recall@20	12.72	13.87	9.1%	11.66	11.92	2.2%	11.58	12.39	7.0%	12.73	13.30	4.5%
NDCG@20	7.89	8.59	8.9%	7.24	7.45	3.0%	7.24	7.77	7.3%	7.93	8.19	3.4%
ANIME												
Recall@10	21.08	22.99	9.1%	16.44	17.65	7.4%	20.59	21.80	5.9%	20.98	23.60	12.5%
NDCG@10	17.19	18.85	9.7%	13.62	14.61	7.3%	16.83	17.60	4.6%	17.12	19.31	12.8%
Recall@20	33.54	36.67	9.3%	26.45	28.59	8.1%	32.79	34.97	6.7%	33.40	37.70	12.9%
NDCG@20	28.87	31.58	9.4%	22.38	24.25	8.4%	28.03	29.64	5.7%	28.62	32.42	13.3%
GOWALLA												
Recall@10	9.56	9.97	4.3%	8.32	9.04	8.7%	9.84	10.38	5.5%	9.31	9.39	0.8%
NDCG@10	5.61	5.87	4.7%	4.87	5.25	7.8%	5.81	6.12	5.3%	5.46	5.53	1.3%
Recall@20	19.29	20.22	4.8%	16.77	18.23	8.7%	20.06	21.26	6.0%	18.61	18.69	0.4%
NDCG@20	10.03	10.47	4.4%	8.67	9.45	8.9%	10.42	11.03	5.9%	9.71	9.83	1.2%
YELP-2018												
Recall@10	6.81	7.61	11.8%	6.23	6.74	8.1%	7.12	7.79	9.5%	6.60	7.25	9.8%
NDCG@10	3.62	4.04	11.7%	3.33	3.60	8.1%	3.79	4.14	9.3%	3.51	3.82	8.9%
Recall@20	10.35	11.57	11.8%	9.36	10.21	9.1%	10.65	11.69	9.7%	9.96	10.95	9.9%
NDCG@20	6.17	6.86	11.2%	5.67	6.13	8.1%	6.46	7.08	9.6%	5.92	6.54	10.6%
MOVIELENS-1M												
Recall@10	18.10	19.82	9.5%	15.56	17.33	11.4%	18.01	19.41	7.8%	17.96	19.60	9.2%
NDCG@10	20.15	22.14	9.9%	17.48	19.37	10.8%	20.00	21.81	9.1%	19.89	21.72	9.2%
Recall@20	25.72	28.15	9.4%	22.40	25.02	11.7%	25.89	27.77	7.3%	25.56	27.81	8.8%
NDCG@20	26.84	29.42	9.6%	23.41	26.13	11.6%	26.83	28.96	8.0%	26.69	29.30	9.8%

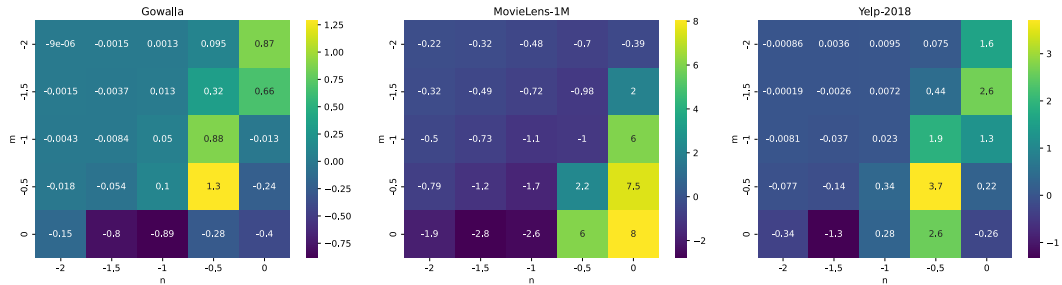


Figure 4: The sensitivity of TAG-CF to m and n in Equation (7). Numbers reported in these plots are performance improvement (%) brought by TAG-CF to MF trained by DirectAU [46] on Recall@20.