GFAR Ommitted Results

1 INTRODUCTION

This document contains the omitted results of our paper *Ensuring Fairness in Group Recommendations by Rank-Sensitive Balancing of Relevance.*

In our paper, due to space limitations, we only show the metric vs. group size plots for the zRecall metric; also we only show the results for the other metrics for group size $m = 8$.

2 MOVIELENS RESULTS

For the MovieLens dataset results see the following figures:

- For Recall metric, Figs. 1, 2 and 3 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
- For nDCG metric, Figs. 4, 5 and 6 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
- For DFH metric, Figs. 7, 8 and 9 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
- Tables 1, 2, 3, 4, 5, 6 shows all the results in table format for *Random*, *Divergent* and *Similar* groups, for group sizes $m = 2, 3, \ldots, 7$ respectively.

Note that,for comparison we also included an analysis of the MAX strategy which has been shown to align with human preferences [1]. For the MovieLens dataset, GFAR always outperforms MAX. Out of 210 cases, there is only one case where MAX outperforms the worst performing baseline AVG for the MovieLens dataset, for the nDCG(mean) metric of the *Divergent* groups of size $m = 2$. These results suggest that MAX is not a suitable strategy for the task of fairness as predicted.

3 KGREC-MUSIC RESULTS

For the KGRec-Music dataset results see the following figures:

- For Recall metric, Figs. 10, 11 and 12 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
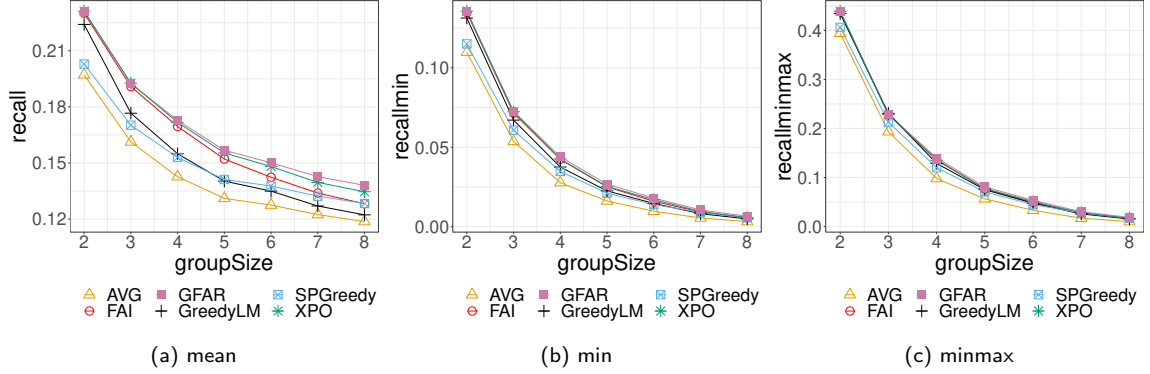
Author's address:

Fig. 1. MovieLens dataset, Recall@20 (mean, min, minmax) for varying group size, for *Random* groups.

Fig. 2. MovieLens dataset, Recall@20 (mean, min, minmax) for varying group size, for *Divergent* groups.

- For nDCG metric, Figs. 13, 14 and 15 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
- For DFH metric, Figs. 16, 17 and 18 show the mean, min, minmax results for *Random*, *Divergent* and *Similar* groups respectively.
- Tables 7, 8, 9, 10, 11, 12 shows all the results in table format for *Random*, *Divergent* and *Similar* groups, for group sizes $m = 2, 3, \ldots, 7$ respectively.

Similar to MoviLens dataset, for comparison we also include an analysis of MAX strategy. For the KGRec-Music dataset, generally, GFAR outperforms MAX. However, MAX outperforms GFAR for 6 out of 210 configurations: *Random* groups of size 2,3,4 for the nDCG(mean) metric; *Divergent* groups of size 2,3,4 for the nDCG(mean) metric. We note however the nDCG(mean) is slightly higher, MAX performs much worse for nDCG(min) and nDCG(minmax), suggesting it is not a suitable fairness strategy as predicted.

REFERENCES

[1] Allison JB Chaney, Mike Gartrell, Jake M Hofman, John Guiver, Noam Koenigstein, Pushmeet Kohli, and Ulrich Paquet. 2014. A large-scale exploration of group viewing patterns. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. 31–38.
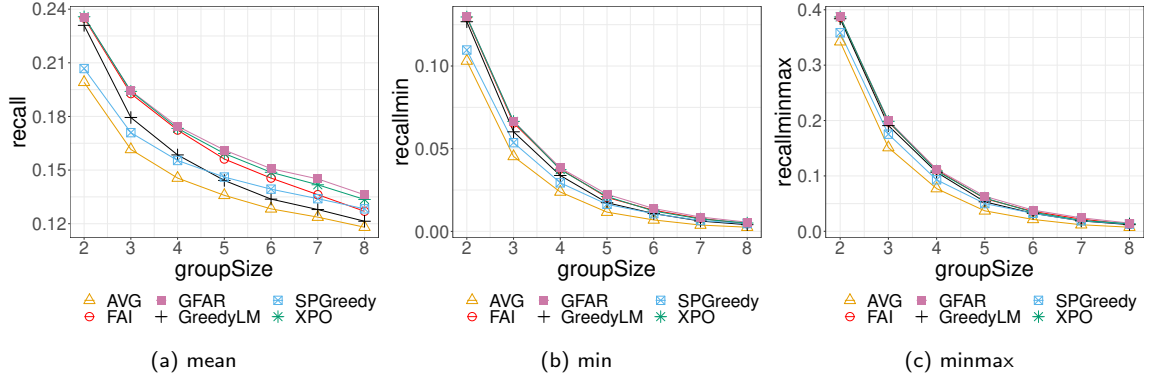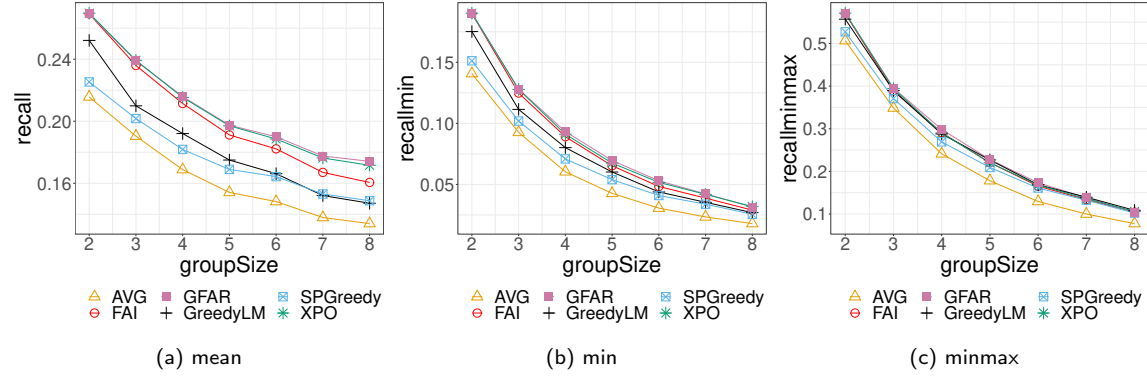
Fig. 3. MovieLens dataset, Recall@20 (mean, min, minmax) for varying group size, for *Similar* groups.
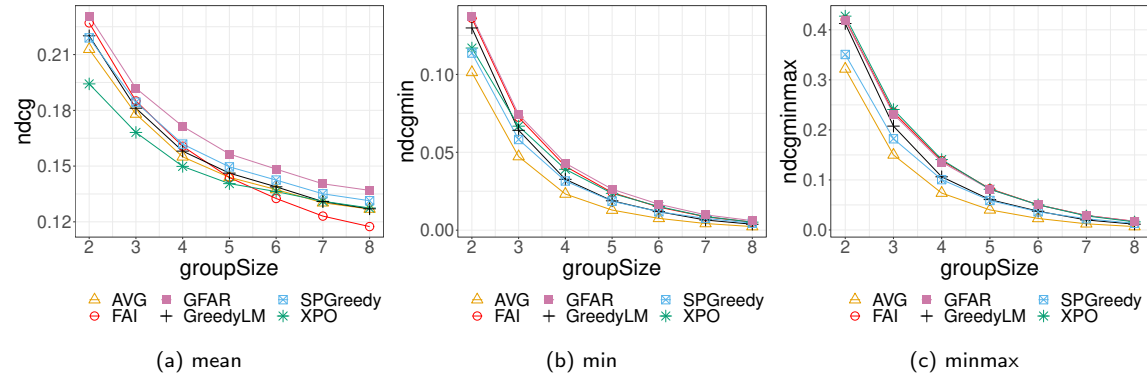
Fig. 4. MovieLens dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Random* groups.

Fig. 5. MovieLens dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Divergent* groups.

Fig. 6. MovieLens dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Similar* groups.

Fig. 7. MovieLens dataset, DFH@20 (mean, min, minmax) for varying group size, for *Random* groups.

Fig. 8. MovieLens dataset, DFH@20 (mean, min, minmax) for varying group size, for *Divergent* groups.

Fig. 9. MovieLens dataset, DFH@20 (mean, min, minmax) for varying group size, for *Similar* groups.

Fig. 10. KGRec-Music dataset, Recall@20 (mean, min, minmax) for varying group size, for *Random* groups.
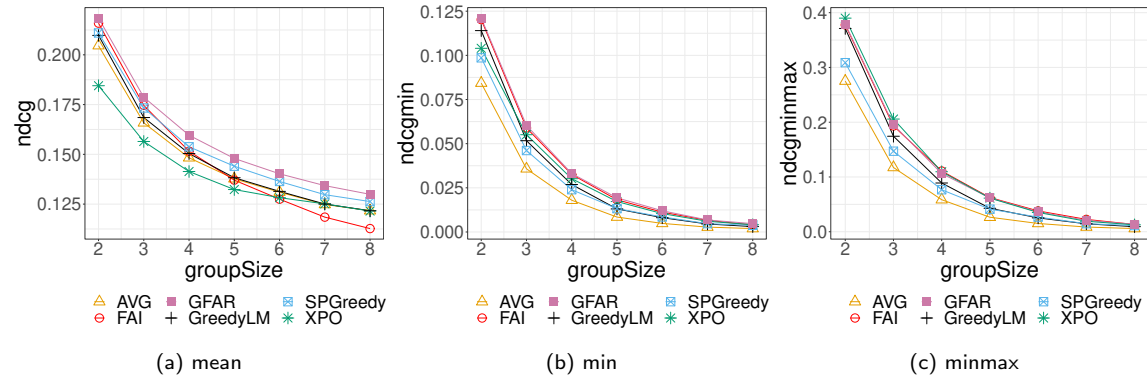

Fig. 11. KGRec-Music dataset, Recall@20 (mean, min, minmax) for varying group size, for *Divergent* groups.

(a) mean   (b) min   (c) minmax

Fig. 12. KGRec-Music dataset, Recall@20 (mean, min, minmax) for varying group size, for *Similar* groups.

(a) mean   (b) min   (c) minmax

Fig. 13. KGRec-Music dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Random* groups.

(a) mean   (b) min   (c) minmax
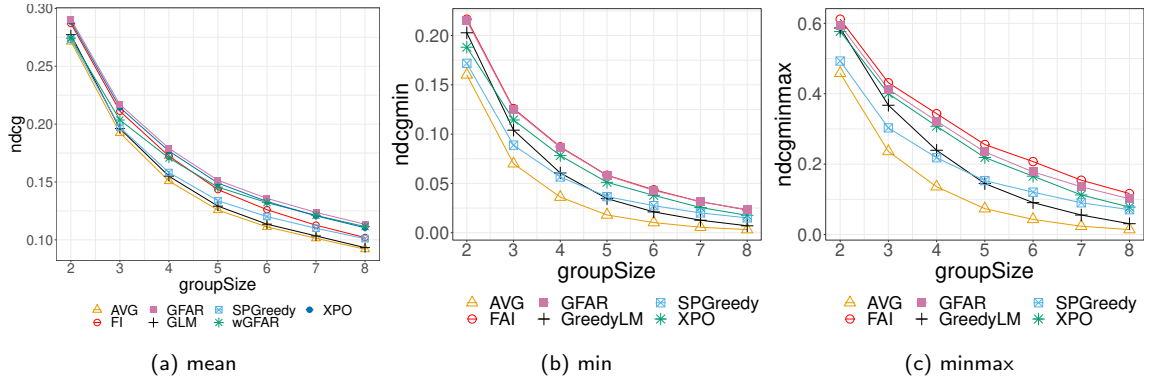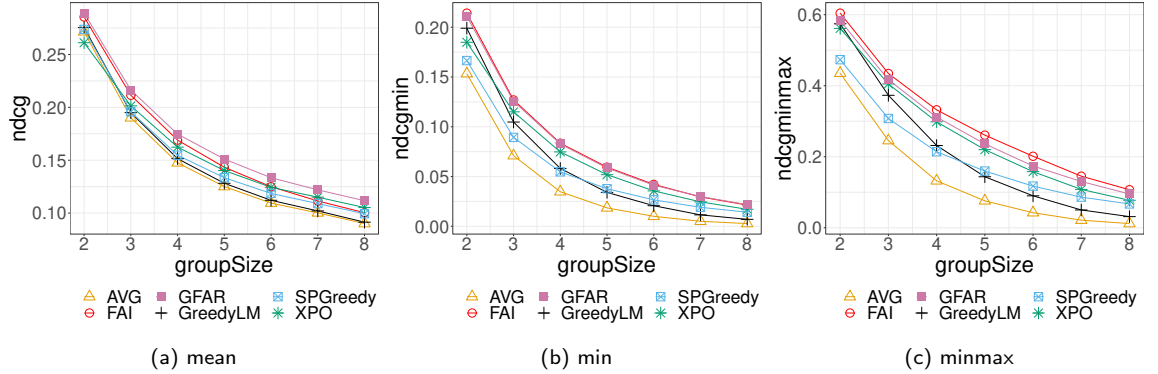
Fig. 14. KGRec-Music dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Divergent* groups.
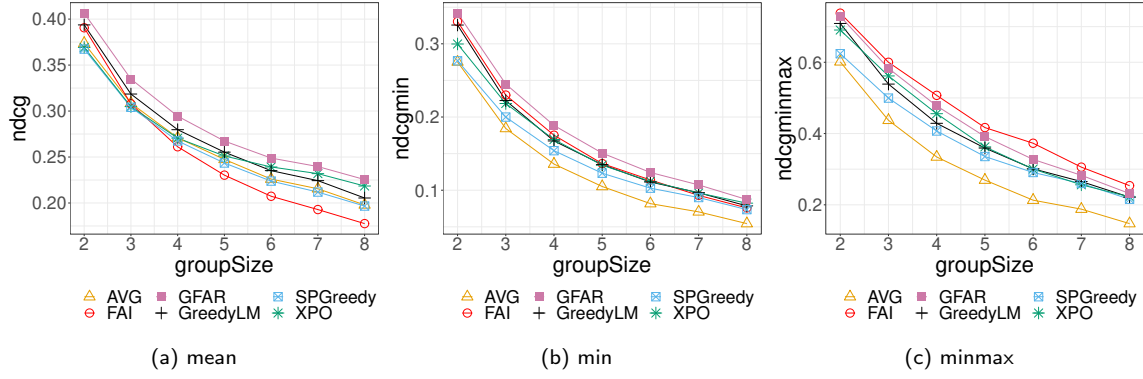
Fig. 15. KGRec-Music dataset, nDCG@20 (mean, min, minmax) for varying group size, for *Similar* groups.
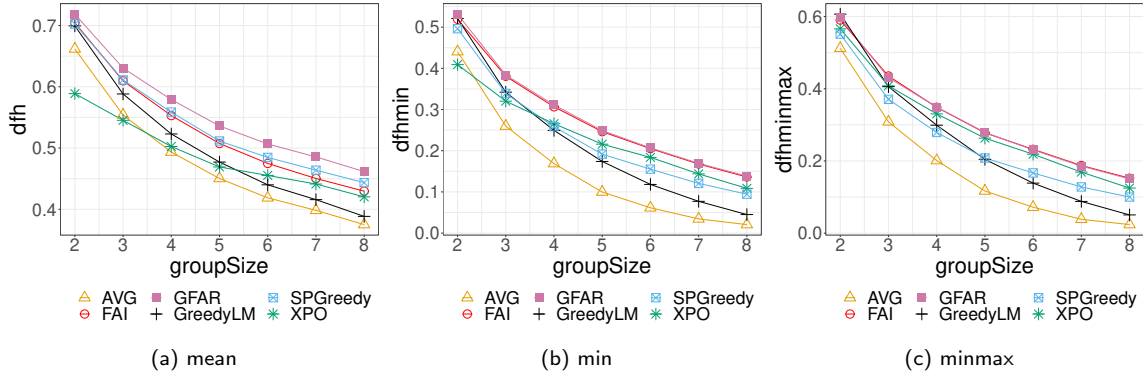
Fig. 16. KGRec-Music dataset, DFH@20 (mean, min, minmax) for varying group size, for *Random* groups.
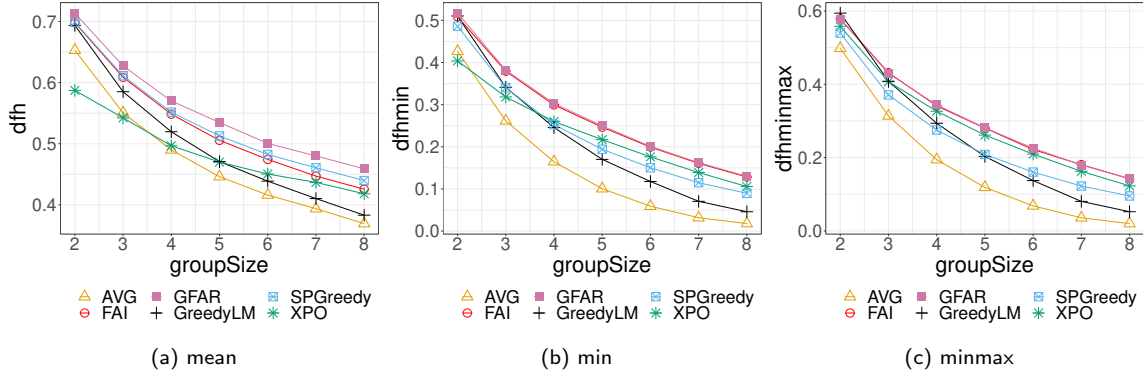
Fig. 17. KGRec-Music dataset, DFH@20 (mean, min, minmax) for varying group size, for *Divergent* groups.
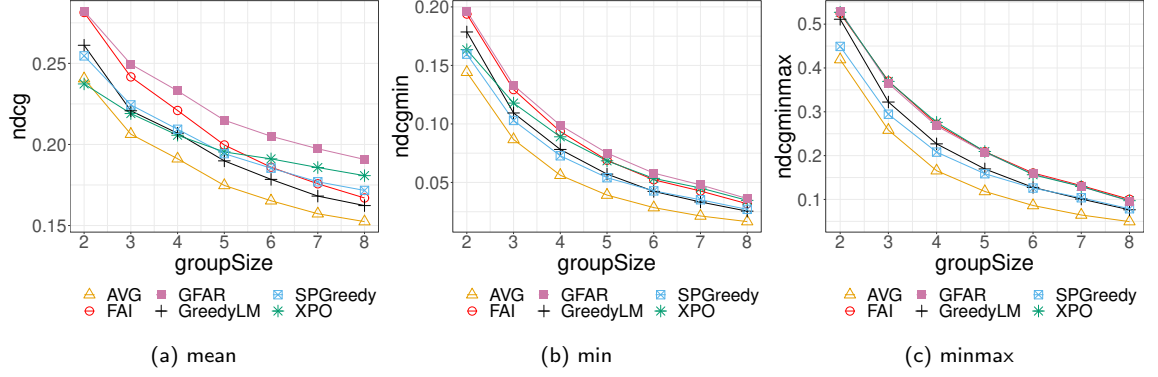
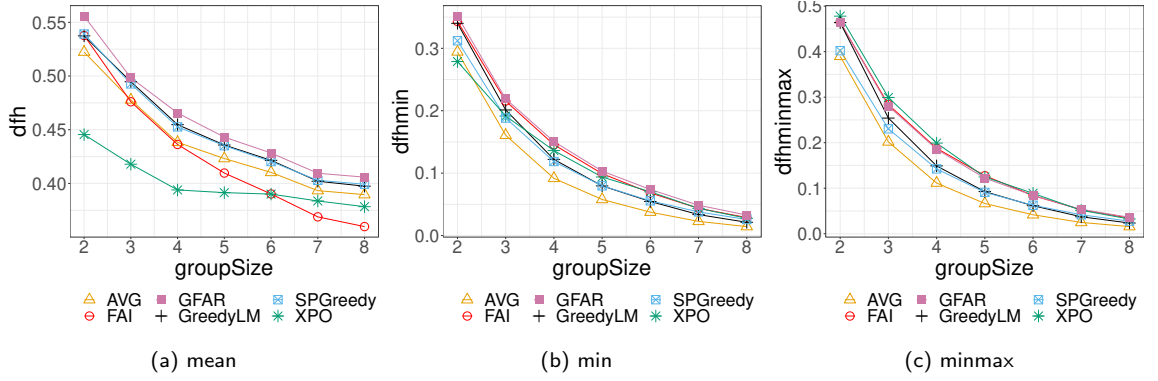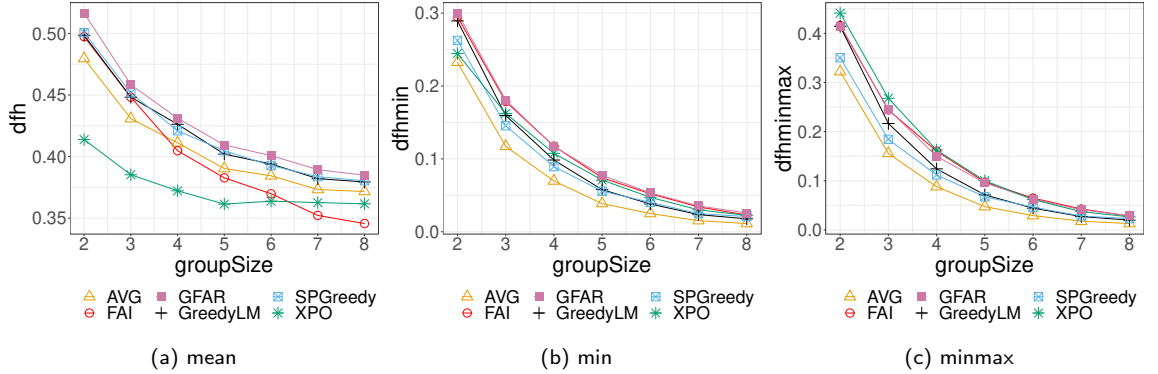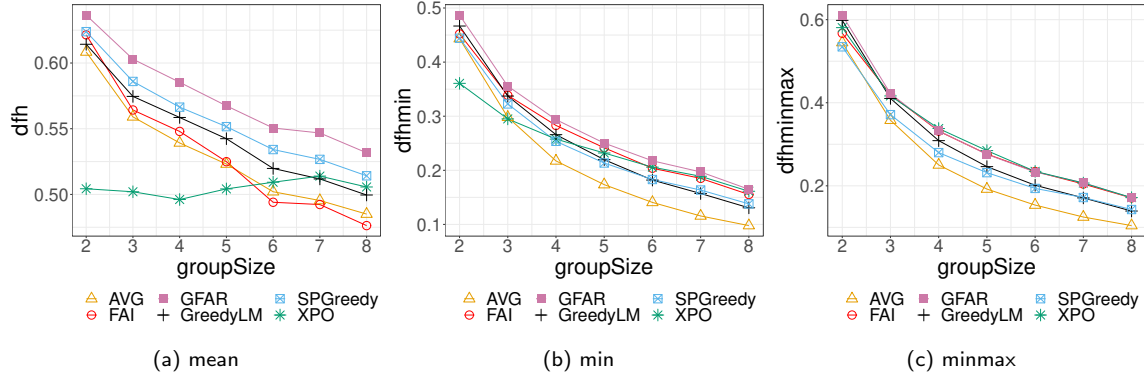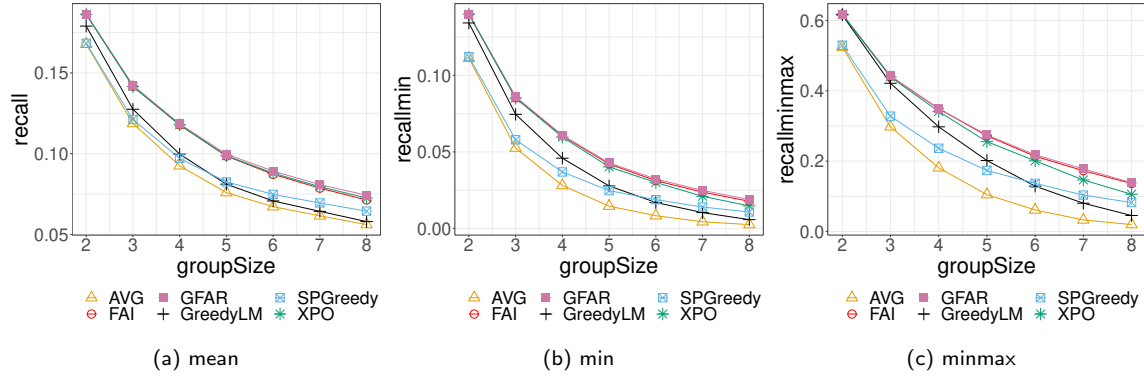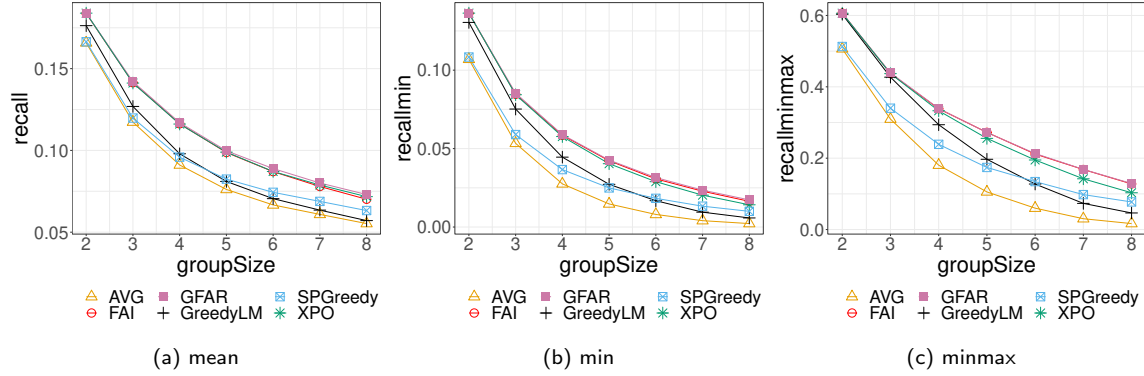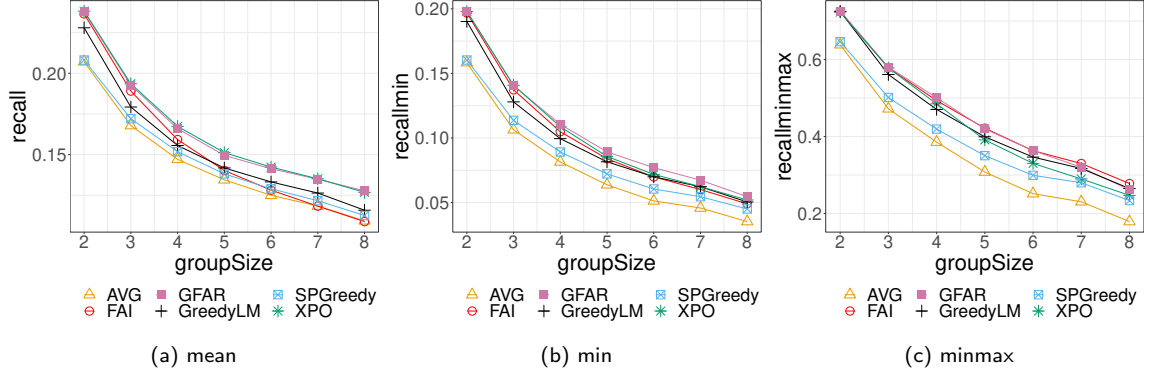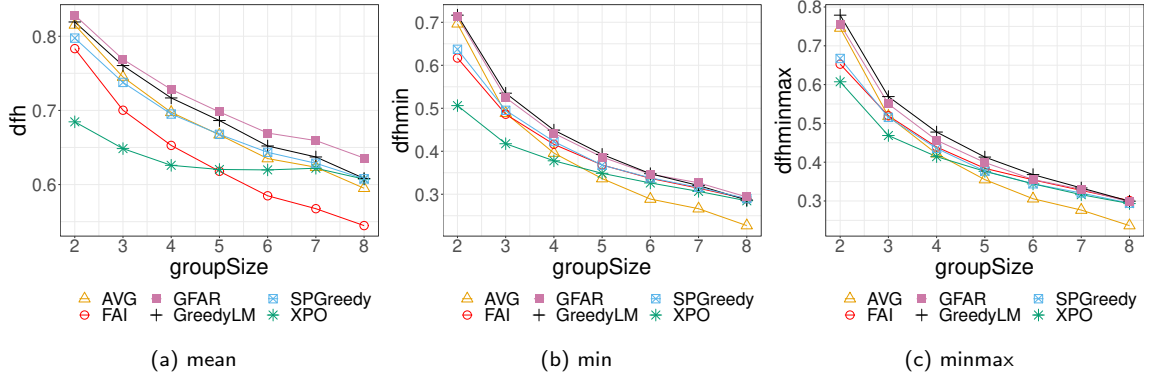Fig. 18. KGRec-Music dataset, DFH@20 (mean, min, minmax) for varying group size, for *Similar* groups.

Table 1. MovieLens dataset: Results when group size $m = 2$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
					Random groups					
AVG	0.1758	0.1972	0.1098	0.3943	0.2129	0.1013	0.3218	0.5223	0.2943	0.3897
FAI	0.1121	0.2301	0.1348	0.4381	0.2271$\star$	0.136$\star$	0.4207	0.5377	0.343$\star$	0.4645
XPO	0.1118$\star$	**0.2312**	**0.1357**	0.4389$\star$	0.1942	0.1169	**0.4274**	0.4455	0.2788	**0.478**
GreedyLM	0.1204	0.2241	0.1313	0.4343	0.2201	0.1298	0.4126	0.5374	0.3401	0.4639
SPGreedy	0.164	0.2029	0.1151	0.4061	0.219	0.1137	0.3507	0.5391$\star$	0.3122	0.4021
GFAR	**0.1116**	0.2309$\star$	0.1355$\star$	**0.4393**	**0.2307**$^\iota$	**0.1374**	0.4192$\star$	**0.5557**$^\iota$	**0.3515**$^\iota$	0.4648$\star$
					Similar groups					
AVG	0.1089	0.2157	0.1407	0.5066	0.2406	0.1442	0.4192	0.6084	0.4434	0.5448
FAI	0.0518	0.2692	0.1899	0.5702	0.2814$\star$	0.1939$\star$	0.5248	0.6214	0.4521	0.5672
XPO	**0.0522**	**0.2697**	**0.1902**	**0.5711**	0.2373	0.1635	0.527$\star$	0.5043	0.3607	0.5808
GreedyLM	0.0672	0.2521	0.1752	0.557	0.2612	0.1786	0.5112	0.6143	0.4667$\star$	0.5983$\star$
SPGreedy	0.0945	0.2253	0.1512	0.5273	0.2546	0.1596	0.4491	0.6239$\star$	0.4442	0.5347
GFAR	0.0523$\star$	0.2693$\star$	0.19$\star$	**0.5711**	**0.2822**	**0.1964**	**0.5296**	**0.6364**$^\iota$	**0.4868**$^\iota$	**0.6108**$^\iota$
					Divergent groups					
AVG	0.2155	0.1991	0.103	0.3423	0.2046	0.0843	0.2751	0.4796	0.2325	0.3221
FAI	0.1445	0.2351	0.1293	0.3863	0.2159$\star$	0.1202$\star$	0.3777	0.4974	0.2949$\star$	0.416
XPO	**0.1442**	**0.2357**	0.1297$\star$	0.3871$\star$	0.1845	0.104	**0.39**$^\iota$	0.4138	0.2445	**0.4411**$^\iota$
GreedyLM	0.1492	0.2309	0.1268	0.3838	0.2098	0.114	0.3712	0.4985	0.2892	0.4144
SPGreedy	0.1984	0.2068	0.1097	0.3585	0.211	0.0986	0.3087	0.5007$\star$	0.2624	0.3504
GFAR	0.1443$\star$	0.2356$\star$	**0.1298**	**0.3873**	**0.2184**$^\iota$	**0.1211**	0.3792$\star$	**0.5165**$^\iota$	**0.2999**$^\iota$	0.4136$\star$

Table 2. MovieLens dataset: Results when group size $m = 3$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
Random groups										
AVG	0.2089	0.1613	0.0537	0.1934	0.178	0.0473	0.15	0.4777	0.1607	0.2017
FAI	0.1438	0.1909	0.0716	0.2281	0.1851*	0.0725*	0.2352	0.476	0.2155*	0.2836
XPO	0.1423*	**0.1927**	0.0724*	0.2287	0.1681	0.0668	**0.2408**$^\iota$	0.4179	0.1915	**0.2993**$^\iota$
GreedyLM	0.1661	0.1766	0.0671	**0.2308**	0.1811	0.0641	0.2076	0.4947*	0.2012	0.2541
SPGreedy	0.1849	0.1702	0.0607	0.213	0.1844	0.0582	0.1823	0.4927	0.1883	0.2305
GFAR	**0.1419**	0.1924*	**0.0725**	0.2291*	**0.1919**$^\iota$	**0.0742**$^\iota$	0.2314*	**0.4984**	**0.2191**	0.2804*
Similar groups										
AVG	0.1104	0.1904	0.0926	0.3487	0.2063	0.0867	0.2582	0.559	0.2981	0.3579
FAI	0.0589	0.236	0.1249	0.389	0.2417*	0.1294*	**0.3702**	0.5643	0.3386*	0.4198*
XPO	0.0579*	**0.2393**	0.1278*	0.3942*	0.2191	0.1178	0.3694	0.502	0.295	0.4164
GreedyLM	0.0837	0.2099	0.1114	0.3895	0.2208	0.1094	0.3222	0.5746	0.3372	0.4104
SPGreedy	0.0926	0.2017	0.1019	0.3708	0.2244	0.103	0.2944	0.586*	0.3227	0.3714
GFAR	**0.0577**	0.2389*	**0.128**	**0.3954**	**0.2495**$^\iota$	**0.1331**$^\iota$	0.3648*	**0.6029**$^\iota$	**0.3557**$^\iota$	**0.4219**
Divergent groups										
AVG	0.2547	0.1615	0.0453	0.1512	0.1658	0.0358	0.1173	0.4308	0.1172	0.1556
FAI	0.1738*	0.1928	0.0656	0.1974	0.1746*	0.0589*	0.1931	0.4483	0.1787*	0.2442
XPO	0.1747	0.194*	0.0665*	0.1986*	0.1565	0.0551	**0.2058**$^\iota$	0.3853	0.1624	**0.2679**$^\iota$
GreedyLM	0.2002	0.1794	0.0602	0.1908	0.1685	0.0517	0.1744	0.4482	0.1595	0.2165
SPGreedy	0.228	0.1709	0.0536	0.1753	0.1733	0.0461	0.1471	0.4517*	0.1455	0.1843
GFAR	**0.1731**	**0.1943**	**0.0666**	**0.1998**	**0.1787**$^\iota$	**0.0603**	0.1951*	**0.4584**	**0.1807**	0.2455*

Table 3. MovieLens dataset: Results when group size $m = 4$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
Random groups										
AVG	0.2425	0.1427	0.0277	0.098	0.155	0.0232	0.0737	0.4383	0.0916	0.1114
FAI	0.1659$\star$	0.1695	0.0424	0.1364$\star$	0.1607	0.041$\star$	0.1386	0.4363	0.1456$\star$	0.1877
XPO	0.166	0.1717$\star$	0.0426$\star$	0.1348	0.1498	0.0391	**0.1407**	0.3939	0.1361	**0.1991**
GreedyLM	0.1996	0.1549	0.0375	0.1287	0.1581	0.0327	0.1063	0.4547$\star$	0.122	0.149
SPGreedy	0.21	0.153	0.0348	0.1201	0.1618$\star$	0.0316	0.1006	0.4526	0.1186	0.1424
GFAR	**0.1627**	**0.1728**	**0.044**$^{\iota}$	**0.1397**	**0.1714**$^{\iota}$	**0.043**$^{\iota}$	0.1356$\star$	**0.4655**$^{\iota}$	**0.1505**$^{\iota}$	0.1855$\star$
Similar groups										
AVG	0.1264	0.1688	0.0604	0.2413	0.1912	0.0562	0.1656	0.5392	0.2173	0.2503
FAI	0.0587$\star$	0.2115	0.0891	0.2909$\star$	0.2209$\star$	0.0936$\star$	0.2717	0.5479	0.283$\star$	0.3316
XPO	0.0593	0.2154$\star$	0.0911$\star$	0.2902	0.2057	0.089	**0.2759**$^{\iota}$	0.4961	0.2577	**0.3379**
GreedyLM	0.0866	0.1921	0.0801	0.2878	0.2069	0.0782	0.2268	0.5586	0.2659	0.3088
SPGreedy	0.1013	0.1818	0.0709	0.2693	0.2094	0.0728	0.208	0.5664$\star$	0.2536	0.28
GFAR	**0.0563**	**0.2159**	**0.0933**$^{\iota}$	**0.2985**$^{\iota}$	**0.233**$^{\iota}$	**0.0988**$^{\iota}$	0.2676$\star$	**0.5853**$^{\iota}$	**0.2937**$^{\iota}$	0.3324$\star$
Divergent groups										
AVG	0.2678	0.1455	0.0238	0.0771	0.1483	0.018	0.0585	0.4112	0.0697	0.088
FAI	0.1931$\star$	0.1724	0.0378$\star$	0.1108$\star$	0.1516	0.0326$\star$	**0.1108**	0.4048	**0.1175**	0.1603
XPO	0.1958	0.1733$\star$	0.0374	0.109	0.1414	0.0302	0.1095	0.3722	0.1077	**0.1626**$^{\iota}$
GreedyLM	0.2225	0.1585	0.0338	0.1067	0.1504	0.0269	0.089	0.4263$\star$	0.0984	0.1242
SPGreedy	0.2402	0.1554	0.0294	0.093	0.1538$\star$	0.0239	0.0767	0.421	0.0895	0.1119
GFAR	**0.1923**	**0.1745**	**0.0384**	**0.112**	**0.1596**$^{\iota}$	**0.0331**	0.1063$\star$	**0.4312**	0.1171$\star$	0.1497$\star$

Table 4. MovieLens dataset: Results when group size $m = 5$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
Random groups										
AVG	0.2528	0.1311	0.0162	0.0564	0.144	0.0129	0.0401	0.4232	0.0578	0.0663
FAI	0.1831$\star$	0.1521	0.0251	0.077	0.1438	0.0242$\star$	0.0824	0.4097	0.0992$\star$	0.1275
XPO	0.1856	0.1554$\star$	0.0253$\star$	0.0774$\star$	0.1406	0.0236	0.0814	0.3914	0.0938	0.1246
GreedyLM	0.2152	0.1403	0.0225	0.075	0.1462	0.0189	0.0607	0.4358$\star$	0.0796	0.0925
SPGreedy	0.2169	0.1412	0.021	0.0706	0.1496$\star$	0.0186	0.0589	0.4351	0.0797	0.0908
GFAR	**0.1788**$^\iota$	**0.1568**$^\iota$	**0.0267**$^\iota$	**0.0809**	**0.1565**$^\iota$	**0.0263**$^\iota$	0.0812$\star$	**0.4431**$^\iota$	**0.1032**$^\iota$	0.1222$\star$
Similar groups										
AVG	0.1282	0.1541	0.0427	0.1781	0.1747	0.0392	0.1182	0.5231	0.1735	0.1922
FAI	0.0647	0.1912	0.0653	0.2205	0.1997$\star$	0.069$\star$	0.209	0.5249	0.2416$\star$	0.2765
XPO	0.0642$\star$	0.197$\star$	0.0675$\star$	0.2203	0.1954	0.0685	**0.2094**	0.5042	0.232	**0.2848**$^\iota$
GreedyLM	0.0888	0.175	0.0602	0.2269$\star$	0.1899	0.0571	0.1696	0.5424	0.2197	0.2462
SPGreedy	0.0968	0.169	0.0539	0.2092	0.1943	0.0541	0.1587	0.5516$\star$	0.2135	0.2317
GFAR	**0.062**	**0.1975**	**0.0696**$^\iota$	**0.2279**	**0.2149**$^\iota$	**0.0751**$^\iota$	0.2076$\star$	**0.5674**$^\iota$	**0.2503**$^\iota$	0.2757$\star$
Divergent groups										
AVG	0.2874	0.1359	0.0116	0.0369	0.1377	0.0084	0.0265	0.3903	0.0387	0.0474
FAI	0.2199	0.1562	0.0205	0.059	0.137	0.0182$\star$	**0.0624**	0.3827	0.0741$\star$	**0.0974**
XPO	0.2186$\star$	0.1593$\star$	0.0211$\star$	0.06$\star$	0.1323	0.0172	0.0614	0.3614	0.071	0.1001
GreedyLM	0.25	0.1442	0.0173	0.0532	0.1382	0.0131	0.0428	0.402	0.0578	0.0713
SPGreedy	0.2542	0.1461	0.0164	0.05	0.144$\star$	0.0132	0.0412	0.4048$\star$	0.0557	0.0674
GFAR	**0.212**$^\iota$	**0.1612**$^\iota$	**0.0224**$^\iota$	**0.0632**	**0.1479**$^\iota$	**0.0193**	0.0616$\star$	**0.4094**	**0.0775**$^\iota$	0.0956$\star$

Table 5. MovieLensdataset: Results when group size $m = 6$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
					Random groups					
AVG	0.2596	0.1275	0.0097	0.0331	0.1374	0.0076	0.023	0.4103	0.0375	0.0416
FAI	0.2006	0.1424	0.016	0.0486	0.1326	0.0148	**0.0503**	0.3902	0.0683	0.0839
XPO	0.1975$\star$	0.1481$\star$	0.0171$\star$	0.0508$\star$	0.1363	0.0152$\star$	**0.0503**	0.3901	0.07$\star$	**0.0888**
GreedyLM	0.2264	0.1349	0.0146	0.0476	0.1388	0.0119	0.0375	0.4215$\star$	0.0546	0.0615
SPGreedy	0.2238	0.1377	0.0139	0.0449	0.1425$\star$	0.0118	0.0364	0.4204	0.0556	0.0626
GFAR	**0.1904**$^\iota$	**0.1501**$^\iota$	**0.018**	**0.0532**	**0.1484**$^\iota$	**0.0167**$^\iota$	0.0501$\star$	**0.4283**$^\iota$	**0.0738**$^\iota$	0.0842$\star$
					Similar groups					
AVG	0.1366	0.1482	0.0306	0.1294	0.1653	0.0285	0.0862	0.502	0.1406	0.1537
FAI	0.0747	0.1822	0.0483	0.1654	0.186	0.0523	**0.1604**	0.494	0.2038	0.2345
XPO	0.0725$\star$	0.1886$\star$	0.0517$\star$	0.1679	0.1911$\star$	0.0535$\star$	0.1564	0.5093	0.2058$\star$	**0.2348**
GreedyLM	0.1003	0.1662	0.0439	0.1687$\star$	0.1785	0.0424	0.1275	0.5198	0.1819	0.2014
SPGreedy	0.101	0.1643	0.0409	0.1609	0.1854	0.0428	0.126	0.5341$\star$	0.183	0.1938
GFAR	**0.0687**$^\iota$	**0.19**$^\iota$	**0.053**	**0.1732**	**0.2051**$^\iota$	**0.058**$^\iota$	0.1583$\star$	**0.5505**$^\iota$	**0.2176**$^\iota$	0.2333$\star$
					Divergent groups					
AVG	0.2921	0.1283	0.0069	0.0214	0.131	0.0049	0.0151	0.3844	0.0251	0.0291
FAI	0.2287	0.1455	0.0128$\star$	0.0359$\star$	0.1275	0.0112$\star$	**0.038**	0.3698	0.0521$\star$	**0.0648**
XPO	0.2284$\star$	0.1488$\star$	0.0123	0.0345	0.1282	0.0104	0.0352	0.3638	0.0472	0.0618
GreedyLM	0.2617	0.1338	0.0108	0.0324	0.1314	0.0081	0.0253	0.394$\star$	0.0381	0.0443
SPGreedy	0.2542	0.1393	0.0107	0.0325	0.1364$\star$	0.0085	0.0268	0.3927	0.0404	0.0463
GFAR	**0.2204**$^\iota$	**0.1508**$^\iota$	**0.0138**	**0.0382**	**0.1402**$^\iota$	**0.012**	0.0365$\star$	**0.4008**$^\iota$	**0.0534**	0.0628$\star$

Table 6. MovieLens dataset: Results when group size $m = 7$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
					Random groups					
AVG	0.2753	0.1224	0.0055	0.017	0.1303	0.0043	0.0123	0.3933	0.0227	0.0248
FAI	0.2198	0.1339	0.0092	0.0273	0.1231	0.0086	**0.0291**	0.3689	0.0436$\star$	0.0528$\star$
XPO	0.219$\star$	0.1396$\star$	0.0096$\star$	0.0279$\star$	0.1311	0.0088$\star$	0.0281	0.3837	0.0435	0.0514
GreedyLM	0.2481	0.127	0.0082	0.0257	0.1309	0.0065	0.02	0.402	0.0336	0.0372
SPGreedy	0.2352	0.1322	0.009	0.0276	0.1351$\star$	0.0076	0.0224	0.4028$\star$	0.0379	0.0416
GFAR	**0.2103**$^\iota$	**0.1427**$^\iota$	**0.0105**$^\iota$	**0.0301**	**0.1405**$^\iota$	**0.0098**$^\iota$	0.029$\star$	**0.4095**$^\iota$	**0.0486**$^\iota$	**0.0533**
					Similar groups					
AVG	0.1372	0.1379	0.0234	0.1	0.1573	0.0214	0.0645	0.4954	0.1156	0.1249
FAI	0.0737	0.167	0.0388	0.1338	0.1757	0.0427	**0.1316**	0.4923	0.1851	0.2044
XPO	0.0711$\star$	0.1763$\star$	0.0419$\star$	0.1359	0.1858$\star$	0.0453$\star$	0.1296	0.5139	0.189$\star$	0.2068$\star$
GreedyLM	0.1006	0.1521	0.0354	**0.1394**	0.1682	0.0333	0.1009	0.5118	0.1571	0.1709
SPGreedy	0.099	0.1531	0.0338	0.1324	0.177	0.0352	0.1038	0.5267$\star$	0.1635	0.1718
GFAR	**0.0694**	**0.1776**$^\iota$	**0.0423**	0.1382$\star$	**0.1973**$^\iota$	**0.0479**$^\iota$	0.1296$\star$	**0.5468**$^\iota$	**0.197**$^\iota$	**0.2073**
					Divergent groups					
AVG	0.299	0.1237	0.0038	0.0118	0.1248	0.0028	0.0086	0.3733	0.0153	0.0178
FAI	0.2414$\star$	0.1363	0.0079$\star$	0.022$\star$	0.1185	0.0066$\star$	**0.0226**	0.3522	0.0341$\star$	**0.0426**
XPO	0.2425	0.1418$\star$	0.0075	0.0197	0.1251	0.0061	0.0191	0.3626	0.0301	0.0371
GreedyLM	0.2706	0.1279	0.0062	0.0187	0.125	0.0045	0.0145	0.3821	0.0235	0.0272
SPGreedy	0.2607	0.134	0.0066	0.0189	0.1297$\star$	0.0047	0.0147	0.3837$\star$	0.0249	0.0284
GFAR	**0.2304**$^\iota$	**0.1451**$^\iota$	**0.0087**	**0.0244**	**0.1343**$^\iota$	**0.0068**	0.0208$\star$	**0.3894**$^\iota$	**0.0355**	0.041$\star$

Table 7. KGRec-Music dataset: Results when group size $m = 2$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
					Random groups					
AVG	0.037	0.1678	0.1111	0.5242	0.2714	0.1602	0.4581	0.6618	0.4404	0.5121
FAI	0.0125	0.1863*	0.1398	0.6175	0.287*	**0.2167**$^\iota$	**0.6118**$^\iota$	0.7044*	0.5193	0.5902
XPO	**0.0123**	0.1863*	**0.14**	0.6177*	0.2612	0.1882	0.5769	0.5889	0.409	0.566
GreedyLM	0.0128	0.179	0.1342	0.6157	0.2773	0.2029	0.5866	0.6994	0.5212*	**0.6066**$^\iota$
SPGreedy	0.0337	0.1683	0.1123	0.5295	0.2732	0.1719	0.4928	0.7024	0.4963	0.5512
GFAR	0.0124*	**0.1864**	**0.14**	**0.6178**	**0.2906**$^\iota$	0.2149*	0.595*	**0.7186**$^\iota$	**0.5315**$^\iota$	0.5962*
					Similar groups					
AVG	0.0039	0.2071	0.1583	0.6384	0.3737	0.2752	0.6013	0.8152	0.6965	0.7459
FAI	0.0008	0.2365	0.1968	0.725*	0.3906	0.3305*	**0.7378**	0.7833	0.6169	0.6527
XPO	0.0008	**0.2382**	0.1979*	0.7231	0.3692	0.2998	0.6905	0.6846	0.5062	0.6073
GreedyLM	**0.0003**	0.228	0.1902	0.7253	0.3937*	0.3256	0.7089	0.8193*	**0.7166**	**0.779**$^\iota$
SPGreedy	0.0031	0.2081	0.1603	0.6465	0.3674	0.2768	0.624	0.7975	0.6372	0.6674
GFAR	0.0008*	0.2379*	**0.1982**	**0.7256**	**0.4065**$^\iota$	**0.3419**$^\iota$	0.7289*	**0.8284**$^\iota$	0.7124*	0.7555*
					Divergent groups					
AVG	0.0422	0.1658	0.1069	0.5065	0.2715	0.1533	0.4352	0.6529	0.4265	0.4981
FAI	**0.0166**	0.1838	0.1362	**0.6065**	0.2855*	**0.2142**$^\iota$	**0.6045**$^\iota$	0.6979	0.5085	0.5784
XPO	0.0167	**0.1839**	**0.1364**	0.6063	0.2612	0.1849	0.5616	0.5873	0.4037	0.5582
GreedyLM	0.0171	0.1763	0.1304	0.6034	0.2755	0.199	0.5744	0.6936	0.5105*	**0.594**
SPGreedy	0.0371	0.1664	0.1084	0.513	0.2737	0.1665	0.4734	0.6988*	0.4861	0.5396
GFAR	0.0169*	**0.1839**	0.1363*	0.6063*	**0.2891**$^\iota$	0.211*	0.5834*	**0.7136**$^\iota$	**0.5172**$^\iota$	0.5777*

Table 8. KGRec-Music dataset: Results when group size $m = 3$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
					Random groups					
AVG	0.0859	0.1187	0.0525	0.2976	0.1924	0.07	0.2361	0.554	0.2598	0.3082
FAI	**0.0237**	0.1417$\star$	0.0853$\star$	0.4409$\star$	0.2112$\star$	**0.126**	**0.4312**$^{\iota}$	0.6097	0.3808$\star$	**0.4353**
XPO	0.0244	0.1415	0.0851	0.4392	0.2016	0.1144	0.4004	0.5452	0.3203	0.408
GreedyLM	0.0363	0.1275	0.0745	0.4212	0.1961	0.104	0.3673	0.5883	0.3421	0.4057
SPGreedy	0.0636	0.1212	0.0579	0.3277	0.1972	0.0888	0.303	0.6111$\star$	0.3389	0.3706
GFAR	0.0247$\star$	**0.1424**$^{\iota}$	**0.0861**$^{\iota}$	**0.4426**	**0.2168**$^{\iota}$	0.1252$\star$	0.4128$\star$	**0.6305**$^{\iota}$	**0.3838**$^{\iota}$	0.4307$\star$
					Similar groups					
AVG	0.0086	0.1678	0.1064	0.472	0.3085	0.1843	0.4372	0.745	0.4883	0.5183
FAI	**0.0015**	0.189	0.1373	0.5785$\star$	0.3078	0.2299$\star$	**0.6**$^{\iota}$	0.7002	0.4861	0.5196
XPO	0.0017	**0.1935**$^{\iota}$	**0.1408**	0.578	0.3048	0.2185	0.5617	0.6485	0.4175	0.4687
GreedyLM	0.0026	0.1793	0.128	0.5611	0.3185$\star$	0.2225	0.5388	0.7605$\star$	**0.5351**	**0.5691**$^{\iota}$
SPGreedy	0.0051	0.1721	0.1136	0.5015	0.304	0.1999	0.4995	0.7377	0.4952	0.5155
GFAR	**0.0015**	0.1927$\star$	0.1405$\star$	**0.5803**	**0.3348**$^{\iota}$	**0.245**$^{\iota}$	0.5829$\star$	**0.7688**	0.5247$\star$	0.5501$\star$
					Divergent groups					
AVG	0.0823	0.1173	0.0534	0.3095	0.1901	0.0709	0.2455	0.5506	0.2616	0.3138
FAI	**0.0245**	0.1414$\star$	0.0846$\star$	0.4373$\star$	0.2114$\star$	**0.1269**	**0.4345**$^{\iota}$	0.6087	0.3794$\star$	**0.4316**
XPO	0.0251	0.1412	0.0843	0.4369	0.2014	0.115	0.4051	0.5424	0.3183	0.4069
GreedyLM	0.0346	0.1269	0.0752	0.4271	0.195	0.1047	0.3731	0.5851	0.3412	0.4084
SPGreedy	0.0602	0.1198	0.0589	0.3401	0.1957	0.0893	0.3077	0.6113$\star$	0.3405	0.3711
GFAR	0.0249$\star$	**0.142**$^{\iota}$	**0.0853**$^{\iota}$	**0.439**	**0.2164**$^{\iota}$	0.1256$\star$	0.4169$\star$	**0.6273**$^{\iota}$	**0.382**	0.4292$\star$

Table 9. KGRec-Music dataset: Results when group size $m = 4$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
Random groups										
AVG	0.125	0.0925	0.0281	0.1809	0.1508	0.0361	0.1358	0.4934	0.1688	0.2009
FAI	**0.0335**	0.1178	0.0605$\star$	**0.349**	0.1724$\star$	**0.0872**	0.3434$^{\iota}$	0.553	0.3066$\star$	0.3487$\star$
XPO	0.0339	0.1179$\star$	0.0597	0.3411	0.1664	0.0782	0.3076	0.5025	0.2656	0.3303
GreedyLM	0.0617	0.0997	0.0459	0.2975	0.155	0.0606	0.2394	0.523	0.2493	0.299
SPGreedy	0.0774	0.0971	0.0369	0.2365	0.1582	0.0566	0.2182	0.5586$\star$	0.2584	0.2788
GFAR	**0.0335**	**0.1185$^{\iota}$**	**0.0609**	0.3489$\star$	**0.1788$^{\iota}$**	0.0866$\star$	0.3221$\star$	**0.5792$^{\iota}$**	**0.3111$^{\iota}$**	**0.3492**
Similar groups										
AVG	0.0095	0.1471	0.0815	0.3855	0.271	0.1357	0.3343	0.6974	0.3958	0.4217
FAI	0.0013	0.1592	0.1048	0.4933$\star$	0.2612	0.1752$\star$	**0.5072$^{\iota}$**	0.6528	0.4161	0.4391
XPO	0.0012$\star$	**0.1674$^{\iota}$**	0.1087$\star$	0.485	0.2706	0.1692	0.4557	0.626	0.3778	0.4143
GreedyLM	0.0034	0.1556	0.0995	0.4704	0.2797$\star$	0.1673	0.4287	0.7168$\star$	**0.4492**	**0.4777$^{\iota}$**
SPGreedy	0.0032	0.1518	0.0891	0.4193	0.2675	0.154	0.4071	0.6951	0.4226	0.4357
GFAR	**0.001**	0.166$\star$	**0.1106$^{\iota}$**	**0.5009**	**0.2944$^{\iota}$**	**0.1885$^{\iota}$**	0.4776$\star$	**0.7283$^{\iota}$**	0.4418$\star$	0.4564$\star$
Divergent groups										
AVG	0.1278	0.0911	0.0275	0.1804	0.1475	0.0347	0.1322	0.4899	0.1646	0.1943
FAI	0.0381$\star$	0.1163$\star$	0.0585$\star$	0.3385$\star$	0.169$\star$	**0.0835**	0.3321$^{\iota}$	0.5484	0.2994$\star$	0.3411$\star$
XPO	0.0389	0.1161	0.0577	0.3329	0.1624	0.0747	0.2983	0.4968	0.2599	0.3266
GreedyLM	0.0632	0.0979	0.0446	0.2938	0.1515	0.0581	0.2317	0.5196	0.2458	0.2937
SPGreedy	0.0808	0.0959	0.0366	0.2388	0.1547	0.0547	0.2141	0.5516$\star$	0.2524	0.2752
GFAR	**0.0379**	**0.117$^{\iota}$**	**0.0589**	0.3393	**0.1747$^{\iota}$**	0.0827$\star$	0.3117$\star$	**0.5701$^{\iota}$**	**0.3029$^{\iota}$**	**0.3439**

Table 10. KGRec-Music dataset: Results when group size $m = 5$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
Random groups										
AVG	0.1712	0.0759	0.0146	0.1047	0.1258	0.018	0.0734	0.4503	0.0998	0.1167
FAI	0.0487$\star$	0.0985	0.0421$\star$	0.272$\star$	0.1434$\star$	**0.0584**	**0.2556$^\iota$**	0.5075	0.2462$\star$	0.2779$\star$
XPO	0.0532	0.0987$\star$	0.0402	0.2554	0.1398	0.0512	0.2188	0.4691	0.2159	0.2632
GreedyLM	0.098	0.0811	0.0277	0.2017	0.1291	0.0346	0.1447	0.4771	0.174	0.2044
SPGreedy	0.0989	0.0825	0.0247	0.1734	0.1338	0.0366	0.1531	0.5115$\star$	0.1916	0.2081
GFAR	**0.0481**	**0.0997$^\iota$**	**0.0428$^\iota$**	**0.2743**	**0.1512$^\iota$**	0.0581$\star$	0.2346$\star$	**0.5361$^\iota$**	**0.2489$^\iota$**	**0.2789**
Similar groups										
AVG	0.019	0.1345	0.0638	0.3077	0.2474	0.1051	0.2694	0.6672	0.3364	0.355
FAI	**0.0026**	0.1402	0.0839	**0.4218**	0.2303	0.1364$\star$	**0.4168$^\iota$**	0.6179	0.3681	0.3841
XPO	0.0035	**0.1512$^\iota$**	0.0857$\star$	0.3913	0.2512	0.1342	0.3627	0.6204	0.3487	0.3765
GreedyLM	0.0068	0.1418	0.0816	0.3994	0.2553$\star$	0.135	0.3593	0.6863$\star$	**0.3927**	**0.4132$^\iota$**
SPGreedy	0.0093	0.1382	0.0722	0.3504	0.2435	0.1232	0.3355	0.6677	0.3677	0.3771
GFAR	**0.0026**	0.1495$\star$	**0.0892$^\iota$**	0.4197$\star$	**0.2674$^\iota$**	**0.1507$^\iota$**	0.3915$\star$	**0.6979$^\iota$**	0.3864$\star$	0.3993$\star$
Divergent groups										
AVG	0.1699	0.0761	0.0147	0.1054	0.1251	0.0185	0.0758	0.4462	0.1006	0.1188
FAI	0.0495$\star$	0.0987	0.0423$\star$	**0.2723**	0.1433$\star$	**0.0594**	**0.2609$^\iota$**	0.5056	0.2466$\star$	0.2804$\star$
XPO	0.0531	0.0989$\star$	0.0404	0.2556	0.1402	0.0518	0.22	0.47	0.2169	0.2614
GreedyLM	0.1015	0.081	0.0272	0.1969	0.1282	0.0341	0.1437	0.4708	0.1702	0.2031
SPGreedy	0.0988	0.0824	0.0249	0.174	0.1335	0.0377	0.1597	0.5129$\star$	0.1942	0.2086
GFAR	**0.0488**	**0.0998$^\iota$**	**0.0427**	0.2719$\star$	**0.1511$^\iota$**	0.0586$\star$	0.2362$\star$	**0.5348$^\iota$**	**0.2495$^\iota$**	**0.2812**

Table 11. KGRec-Music dataset: Results when group size $m = 6$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
				Random groups						
AVG	0.2042	0.0672	0.0083	0.0611	0.1112	0.0103	0.043	0.4187	0.0615	0.0717
FAI	0.0602$\star$	0.0872	0.0312$\star$	0.2138$\star$	0.126	**0.0435**	**0.2072**$^\iota$	0.4749	0.2052$\star$	**0.2321**$^\iota$
XPO	0.0638	0.088$\star$	0.03	0.2008	0.1269$\star$	0.0379	0.1657	0.4553	0.1839	0.2186
GreedyLM	0.1363	0.0707	0.017	0.1288	0.1135	0.0211	0.0906	0.4399	0.1182	0.1385
SPGreedy	0.106	0.0748	0.0188	0.1369	0.1201	0.0274	0.1203	0.4847$\star$	0.1553	0.1668
GFAR	**0.0587**	**0.0892**$^\iota$	**0.0323**$^\iota$	**0.218**	**0.1357**$^\iota$	0.0429$\star$	0.1779$\star$	**0.5066**$^\iota$	**0.2068**	0.2304$\star$
				Similar groups						
AVG	0.024	0.125	0.0513	0.2516	0.226	0.0817	0.213	0.6351	0.2888	0.3058
FAI	0.0036$\star$	0.1281	0.0698	0.3632$\star$	0.2073	0.114$\star$	**0.3728**$^\iota$	0.5849	0.337	0.3545
XPO	0.0052	**0.1423**$^\iota$	0.0718$\star$	0.3303	0.2392$\star$	0.112	0.3004	0.6198	0.3262	0.3442
GreedyLM	0.0079	0.1332	0.07	0.3466	0.2352	0.1112	0.301	0.652$\star$	**0.3478**	**0.368**$^\iota$
SPGreedy	0.0091	0.129	0.0605	0.2992	0.2238	0.1028	0.2907	0.6435	0.3382	0.3444
GFAR	**0.0034**	0.1413$\star$	**0.077**$^\iota$	**0.3633**	**0.2489**$^\iota$	**0.1242**$^\iota$	0.3275$\star$	**0.6695**$^\iota$	0.3469$\star$	0.3553$\star$
				Divergent groups						
AVG	0.2053	0.0667	0.008	0.06	0.1094	0.01	0.0424	0.4161	0.0591	0.0683
FAI	0.0628$\star$	0.087	0.0307$\star$	0.2117$\star$	0.125$\star$	**0.0421**	**0.2014**$^\iota$	0.4744	0.1992$\star$	0.2219$\star$
XPO	0.069	0.0871$\star$	0.0289	0.1949	0.1244	0.0356	0.1581	0.4505	0.1761	0.2096
GreedyLM	0.1333	0.0705	0.0168	0.1276	0.1121	0.0207	0.0899	0.4389	0.1177	0.1373
SPGreedy	0.109	0.0744	0.0183	0.1344	0.1186	0.0266	0.1172	0.4825$\star$	0.1502	0.1604
GFAR	**0.0625**	**0.0889**$^\iota$	**0.0315**$^\iota$	**0.2132**	**0.1334**$^\iota$	0.0413$\star$	0.1739$\star$	**0.5005**$^\iota$	**0.201**	**0.2256**

Table 12. KGRec-Music dataset: Results when group size $m = 7$. For each group type, the best result for each metric is highlighted in **bold**. If the best result is achieved by GFAR, the second best is marked with $\star$; otherwise GFAR is marked with $\star$. The metrics for the **best result** are also marked with ι if the results are statistically significant with respect to the approach marked with $\star$. (Paired t-test with $p < 0.05$, with Bonferroni correction)

	zRecall	Recall			nDCG			DFH		
		AVG	min	minmax	AVG	min	minmax	AVG	min	minmax
				Random groups						
AVG	0.228	0.0615	0.0043	0.0328	0.1014	0.0055	0.0235	0.3983	0.0345	0.0383
FAI	0.0711*	0.0786	0.0237*	0.1728*	0.1126	**0.0315**	**0.1543**$^\iota$	0.4503	0.1677*	**0.1872**
XPO	0.0802	0.0795*	0.021	0.1464	0.1162*	0.0255	0.1124	0.4412	0.1438	0.1691
GreedyLM	0.1675	0.0643	0.0103	0.0803	0.1034	0.0127	0.0554	0.4157	0.0771	0.0878
SPGreedy	0.1148	0.0695	0.014	0.1034	0.1101	0.0201	0.0902	0.4638*	0.1207	0.1279
GFAR	**0.069**	**0.0808**$^\iota$	**0.0248**$^\iota$	**0.1779**$^\iota$	**0.1233**$^\iota$	**0.0315**	0.1357*	**0.4855**$^\iota$	**0.1694**	0.1856*
				Similar groups						
AVG	0.0242	0.1184	0.0458	0.2306	0.2155	0.0704	0.1879	0.6234	0.2663	0.2766
FAI	0.0047*	0.1183	0.0601	**0.3299**	0.1929	0.0929	**0.3057**$^\iota$	0.5675	0.314	0.3272
XPO	0.0077	**0.1351**	0.0623*	0.2898	0.2319*	0.0963	0.2562	0.622	0.3067	0.3162
GreedyLM	0.0104	0.1264	0.0623*	0.317	0.2243	0.0966*	0.2653	0.6372*	0.3202*	**0.3336**
SPGreedy	0.0085	0.1217	0.0546	0.2801	0.2119	0.09	0.26	0.6287	0.316	0.3205
GFAR	**0.0037**	0.1349*	**0.0673**$^\iota$	0.3202*	**0.2396**$^\iota$	**0.1072**$^\iota$	0.2828*	**0.6593**$^\iota$	**0.3266**	0.3302*
				Divergent groups						
AVG	0.2338	0.0609	0.0041	0.0303	0.1002	0.005	0.0214	0.3938	0.0317	0.0358
FAI	0.0743*	0.0779	0.0228*	0.1682*	0.1115	0.0295*	**0.1453**$^\iota$	0.4471	0.1607*	**0.1808**
XPO	0.0848	0.0788*	0.0203	0.1422	0.1151*	0.0245	0.108	0.4368	0.1388	0.1626
GreedyLM	0.1768	0.0635	0.0095	0.0733	0.1021	0.0115	0.0502	0.4104	0.0706	0.0806
SPGreedy	0.1207	0.0688	0.0133	0.0979	0.109	0.0191	0.0862	0.4607*	0.1149	0.1224
GFAR	**0.0736**	**0.0801**$^\iota$	**0.0235**$^\iota$	**0.1686**	**0.122**$^\iota$	**0.0299**	0.1303*	**0.4801**$^\iota$	**0.1623**	0.1798*