Does Knowledge Graph Really Matter for Recommender Systems?

Anonymous Author(s)

ABSTRACT

Recommender systems (RSs) are designed to provide personalized recommendations to users. Recently, knowledge graphs (KGs) have been widely introduced in RSs, to improve the recommendation accuracy. In this study, however, we demonstrate that RSs do not necessarily perform worse even KG is downgraded to the user-item interaction graph only (or removed). We systematically evaluate how much a KG contributes to the recommendation accuracy of a KG-based RS, defining a metric KGER (*KG utilization efficiency in recommendation*) and proposing an evaluation framework KG4RecEval. We consider the scenarios where knowledge gets completely removed, randomly distorted and decreased, and also where recommendations are for cold-start users. Our extensive experiments on four commonly used datasets and a number of state-of-the-art KG-based RSs reveal that, to remove, randomly distort or decrease knowledge does not necessarily decrease recommendation accuracy, even for cold-start users. These findings inspire us to rethink how to better use knowledge from the existing KGs, whereby we discuss and provide insights about what characteristics of KGs and datasets may help improve their utilization efficiency. The code and supplementary material of this paper are available at: https://anonymous.4open.science/r/KGRecRole-CCC0/.

CCS CONCEPTS

• **Computing methodologies** → *Neural networks*; **Knowledge representation and reasoning**; • **Information systems** → **Recommender systems**.

KEYWORDS

Recommender System, Reproducibility, Knowledge Graph, Heterogeneous Information Network

ACM Reference Format:

Anonymous Author(s). 2024. Does Knowledge Graph Really Matter for Recommender Systems?. In *Proceedings of The 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR '24).* ACM, New York, NY, USA, 17 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION
2 BACKGROUND
3 EVALUATION FRAMEWORK DESIGN
3.1 What if there is no KG? (RQ1)
3.2 What if the knowledge is false? (RQ2)
3.3 What if the knowledge decreases? (RQ3)
3.4 What if KG is for cold-start users? (RQ4)
4 IMPLEMENTATION AND ANALYSIS
4.1 Implementation Settings
4.2 Investigation on No Knowledge
4.3 Investigation on False Knowledge
4.4 Investigation on Decreasing Knowledge
4.5 Investigation on Cold-start Setting
4.6 Comparison between Authenticity and Amount of Knowledge
5 DISCUSSION AND INSIGHTS
6 CONCLUSION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM SIGIR '24, July 14-18, 2024, Washington D.C., USA
© 2024 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

REFERENCES

ADDITIONAL EXPERIMENTAL RESULTS

A EVALUATION METRICS

- **MRR.** MRR measures on average how well a RS captures user preference, specifically by measuring whether the first preferred item is ranked top, defined as:

$$MRR = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{rank_u^*} \tag{1}$$

where $rank_u^*$ means the rank of the first preferred item of user u in his recommendation list.

- **Hit@10.** Hit ratio is a way of calculating how often you recommend at least 1 preferred item to users. If the recommendation list contains at least 1 item in the test set, we call it a hit.

$$Hit@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \delta\left(\hat{R}(u) \cap R(u) \neq \emptyset\right) \tag{2}$$

where δ is an indicator function. $\delta(x) = 1$ if x is true and 0 otherwise. $\hat{R}(u)$ is the recommendation list and $R(u)$ is the positive items in the test set for user u.

- **NDCG@10.** NDCG (Normalized Discounted Cumulative Gain) is a measure of ranking quality. It give higher scores to the hits at top ranks.

$$NDCG@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \left(\frac{1}{\sum_{i=1}^{min(|R(u)|,K)} \frac{1}{log_2(i+1)}} \sum_{i=1}^{K} \delta(\hat{R}_i(u) \in R(u)) \frac{1}{log_2(i+1)} \right) \tag{3}$$

where $\hat{R}_i(u)$ is the i-th item recommended to user u.

- **Precision@10.** Precision is the fraction of correctly recommended items out of all the recommended items for users.

$$Precision@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\hat{R}(u) \cap R(u)|}{|\hat{R}(u)|} \tag{4}$$

- **Recall@10.** Recall is the fraction of correctly recommended item out of all the preferred items for users.

$$Recall@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\hat{R}(u) \cap R(u)|}{|R(u)|} \tag{5}$$

B ADDITIONAL RESULTS OF FALSE KNOWLEDGE EXPERIMENT

Additional results of false knowledge experiments are presented in Figure B1. When using Hit, NDCG, Precision and Recall as metrics, the results are similar to those when using MRR as metric. The results also indicate that with more knowledge distorted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease, which is the same as the conclusion in Section 4.3.

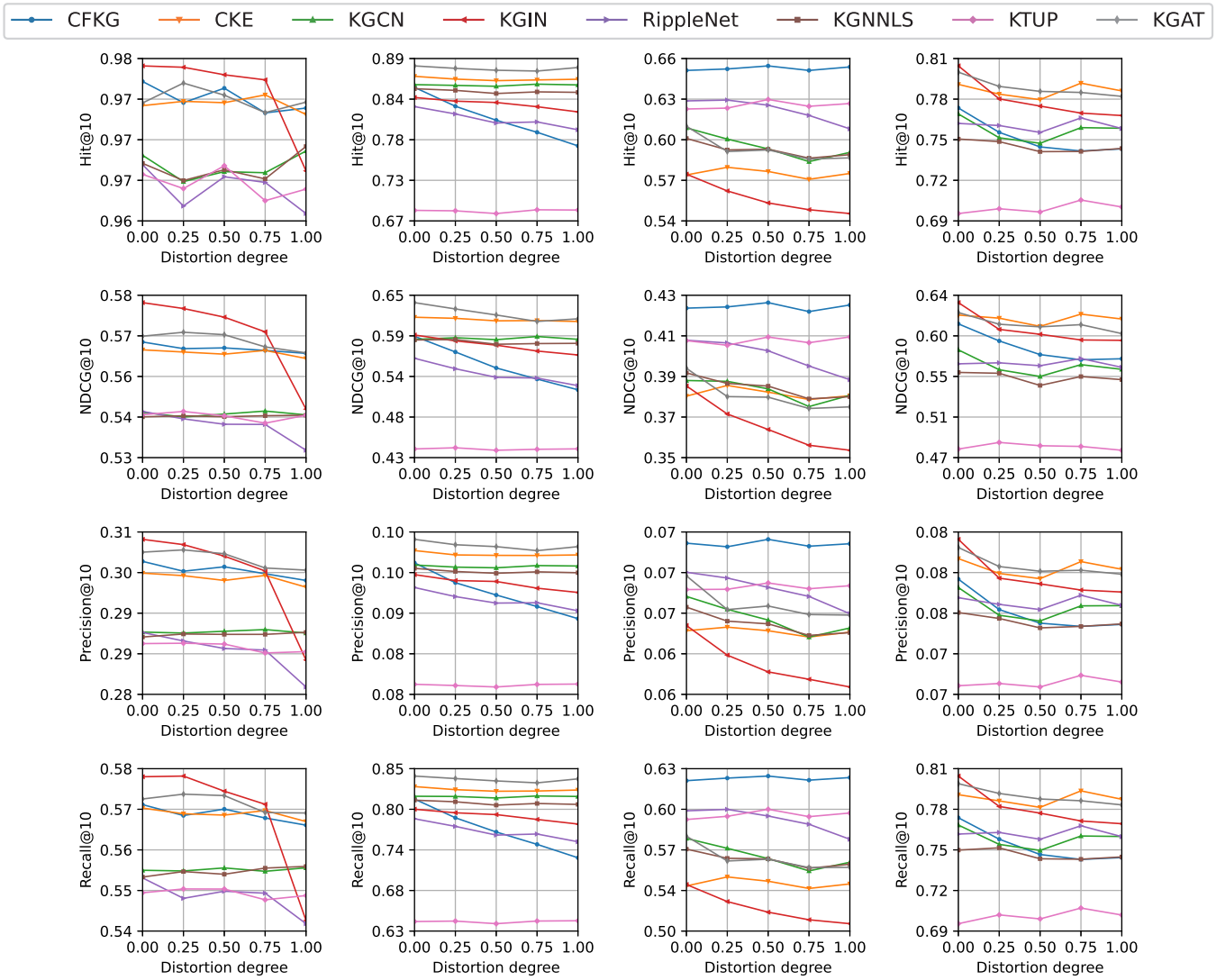
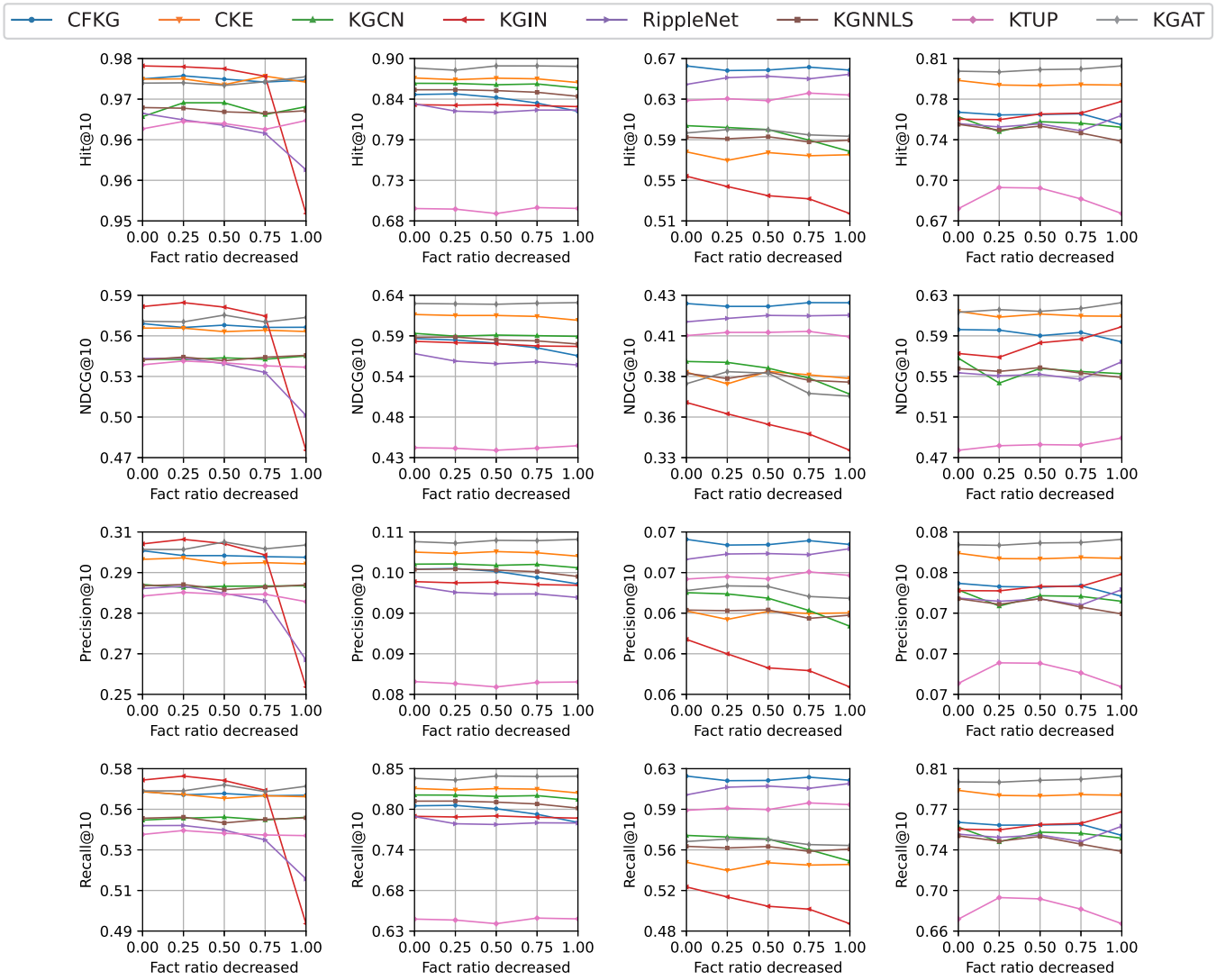


Figure B1: Additional results of randomly distorted KG. The horizontal axis represents the distortion degree. The results are similar to those when using MRR as metric, i.e., with more knowledge distorted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease.

C ADDITIONAL RESULTS OF DECREASING KNOWLEDGE EXPERIMENT

Additional results of decreasing knowledge experiments are presented in Figure C2-C4. When using Hit, NDCG, Precision and Recall as metrics, the results are similar to those when using MRR as metric. Decreasing entities and relations generally have similar effects as decreasing facts. The results indicate that with more knowledge deleted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease, which is the same as the conclusion in Section 4.4.



(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure C2: Additional results of decreasing the facts. The horizontal axis denotes the ratio that decreases. The results are similar to those when using MRR as metric, i.e., with more facts deleted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease.

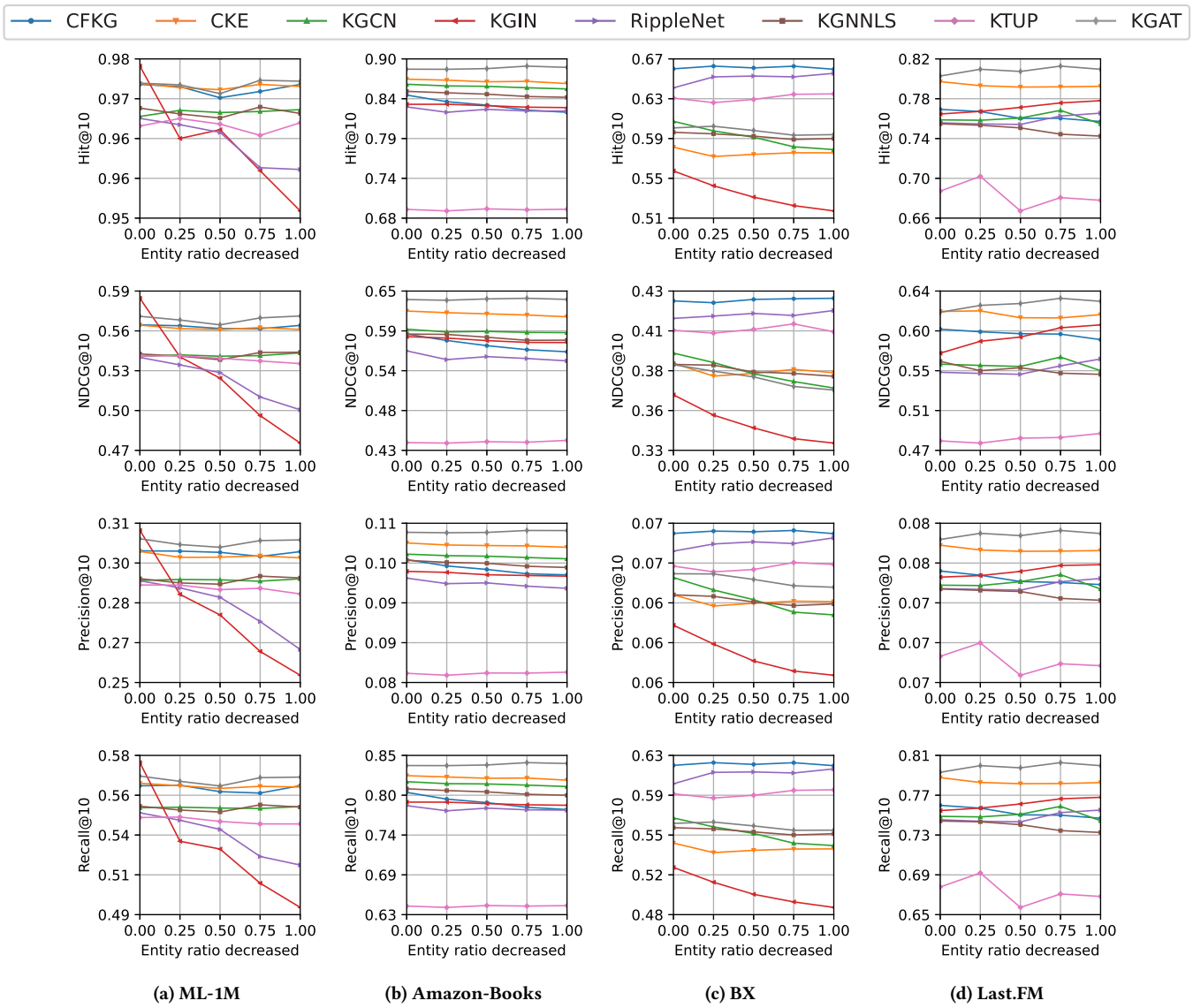
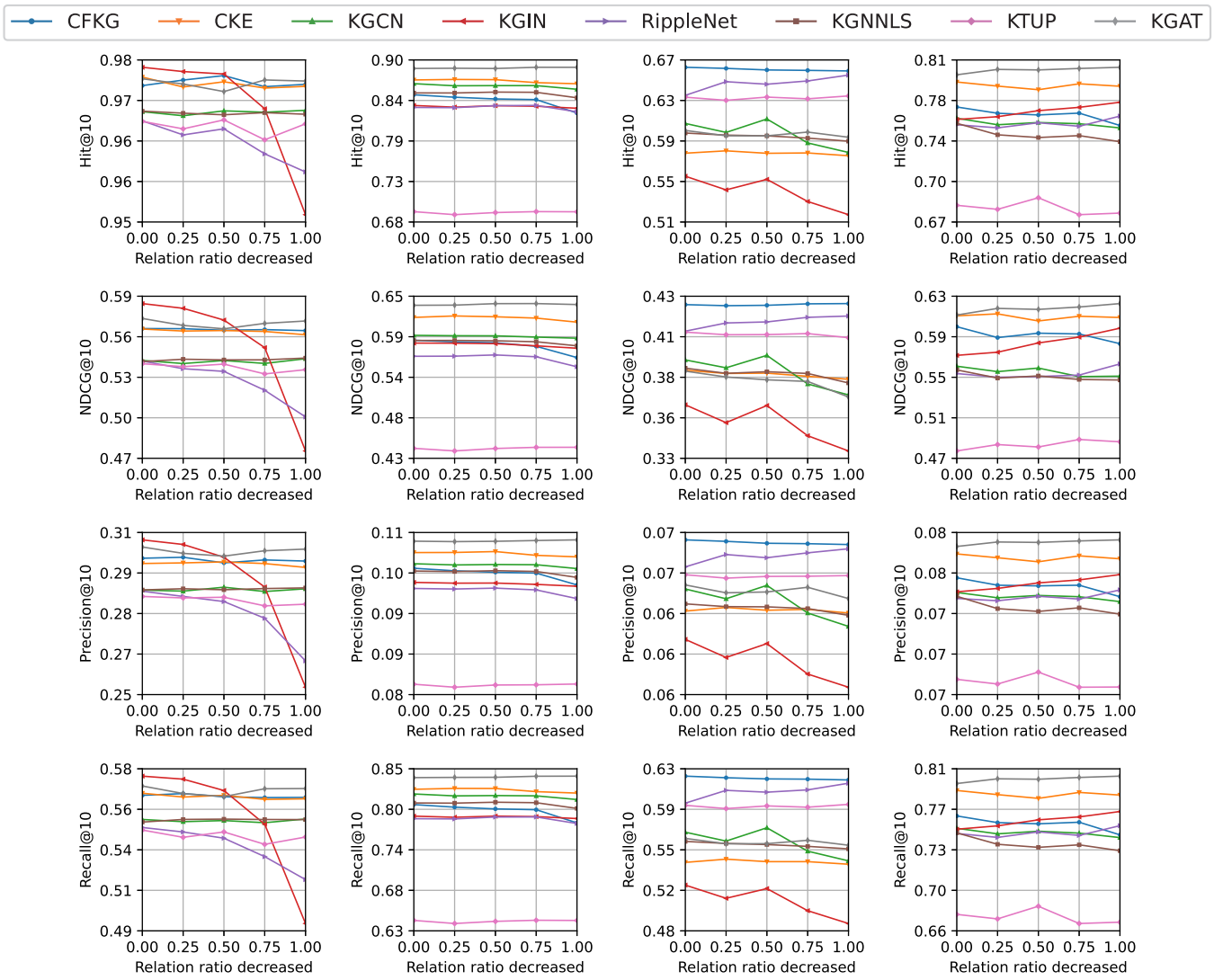


Figure C3: Additional results of decreasing the entities (nodes). The horizontal axis denotes the ratio that decreases. The results are similar to those when decreasing facts, i.e., with more entities deleted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease.

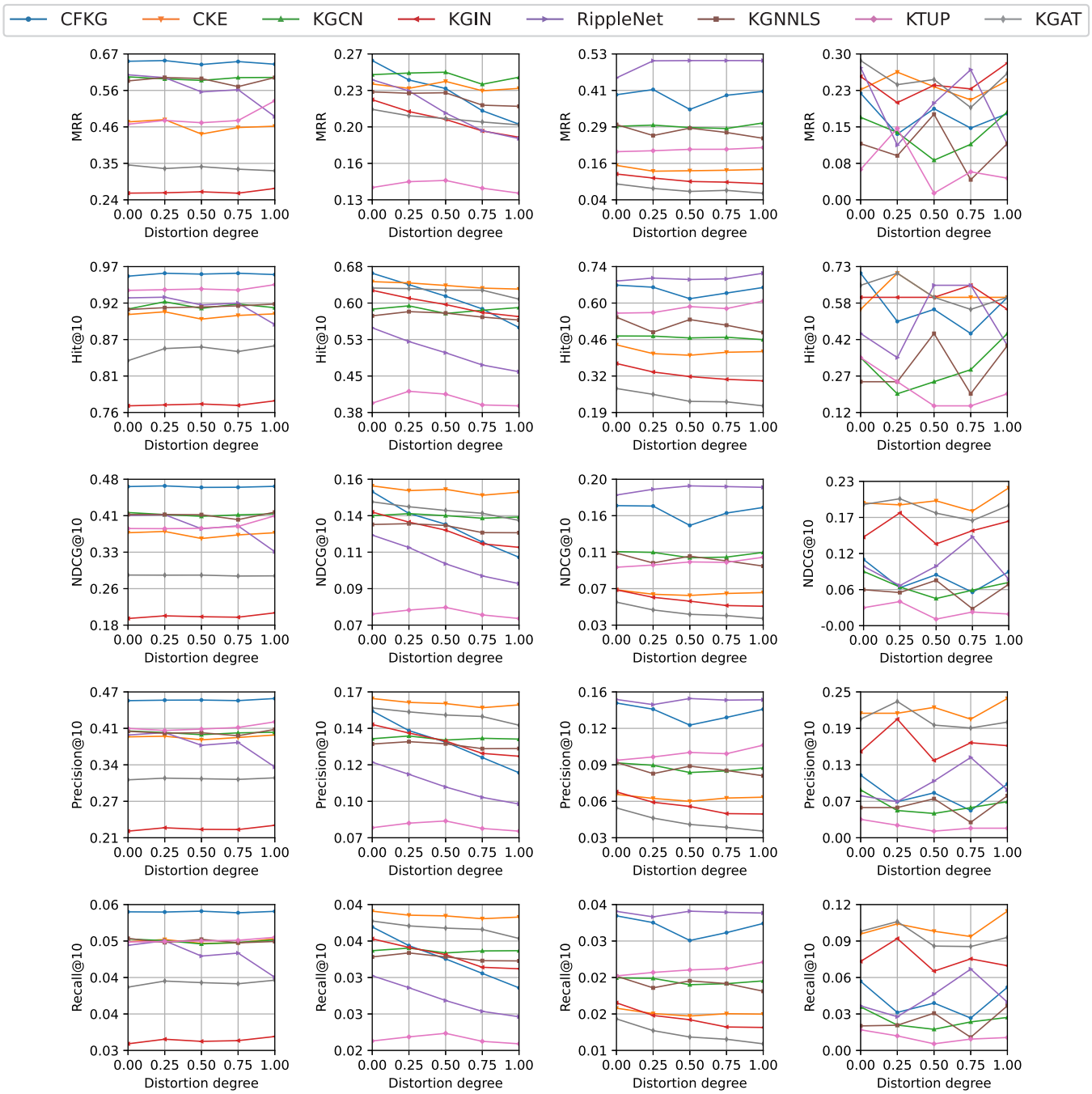


(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure C4: Additional results of decreasing the relations. The horizontal axis denotes the ratio that decreases. The results are similar to those when decreasing facts, i.e., with more relations deleted in a KG, the recommendation accuracy of a KG-based RS does not necessarily decrease. And the change degree is generally smaller than that when decreasing facts.

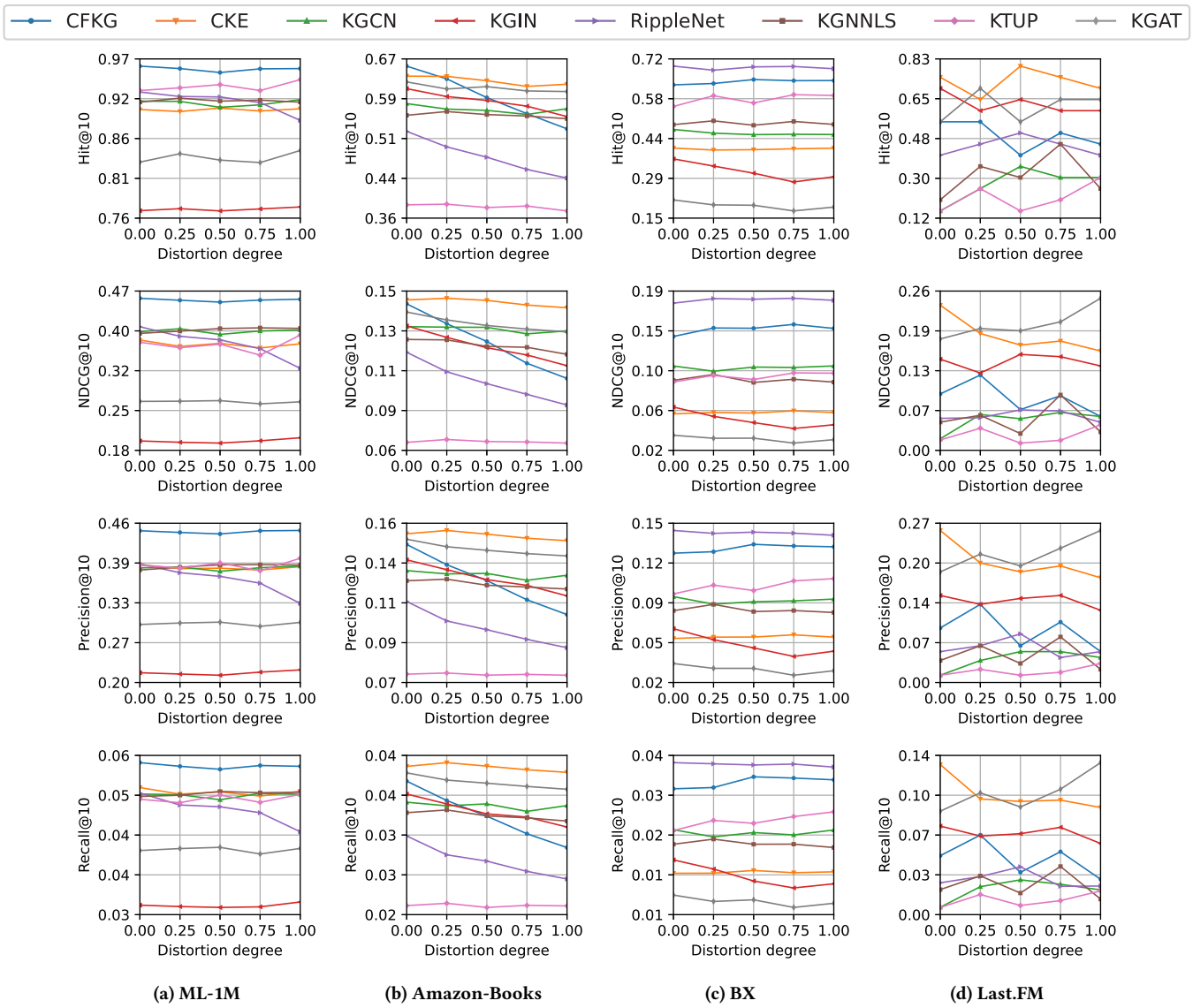
D ADDITIONAL RESULTS OF COLD-START EXPERIMENT

Additional results of cold-start experiments are presented in Figure D5-D10. When using different metrics and T, the results are generally similar to those when using MRR as metric and T=3. Decreasing entities and relations generally have similar effects as decreasing facts. The results indicate that with more knowledge randomly distorted or deleted, recommendation accuracy for cold-start users does not necessarily decrease, and how it gets influenced depends on both the model and the dataset being used., which is the same as the conclusion in Section 4.5.



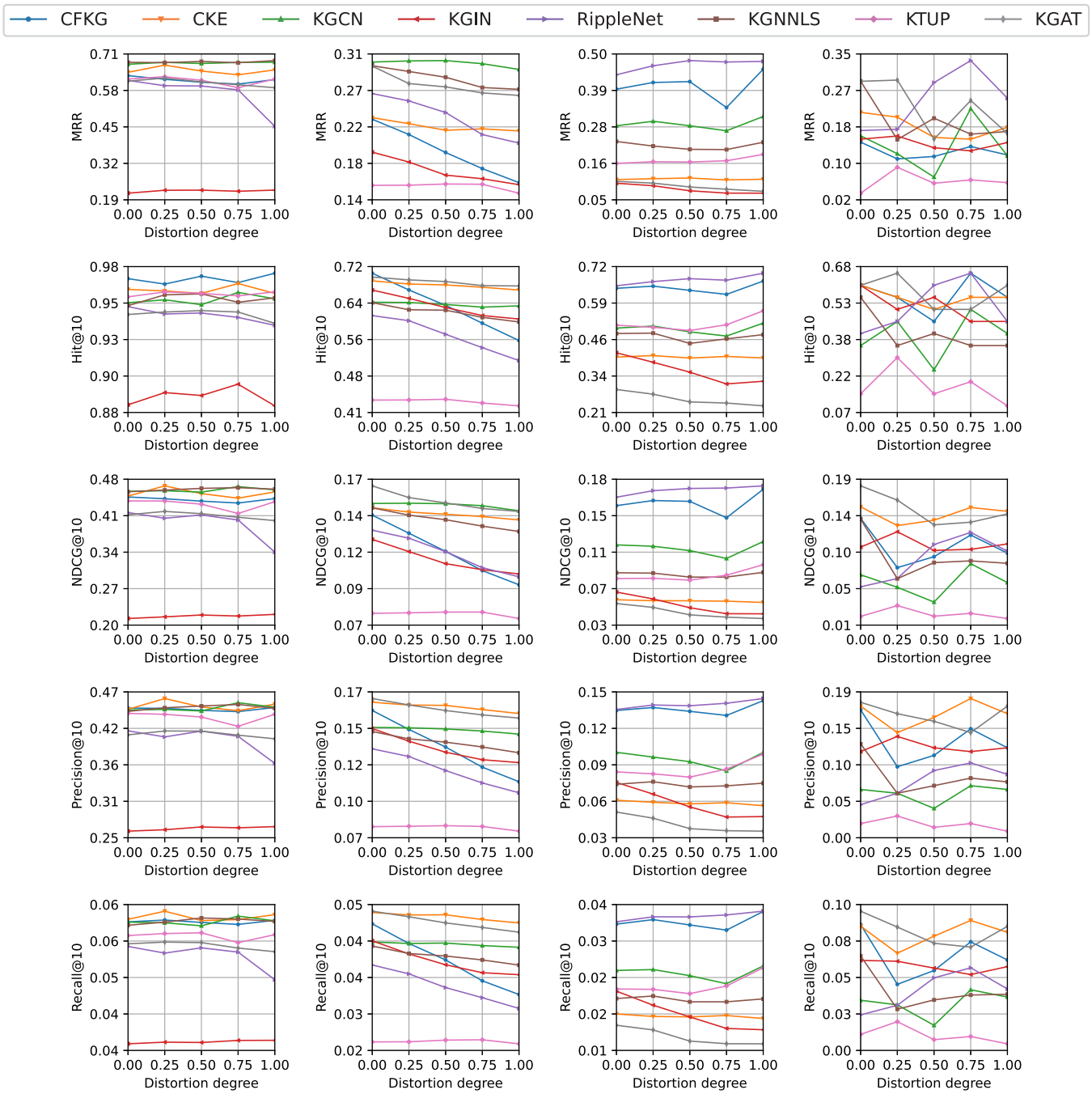
(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D5: Additional result of cold-start experiment with false knowledge, T=1. Horizontal axis denotes distortion degree.The results are similar to those when T=3, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.



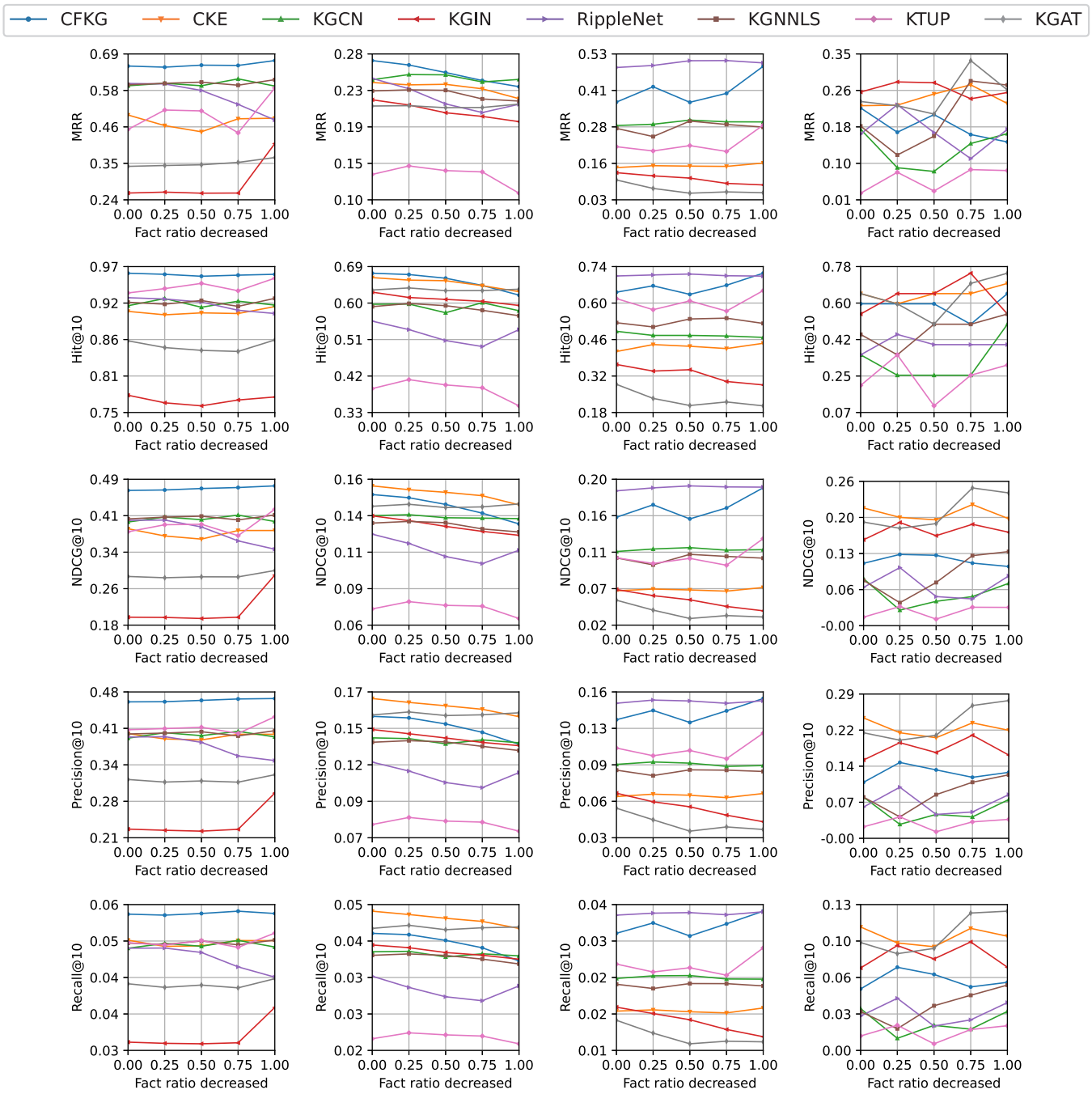
(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D6: Additional result of cold-start experiment with false knowledge, T=3. Horizontal axis denotes distortion degree.The results are similar to those using MRR as metric, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.



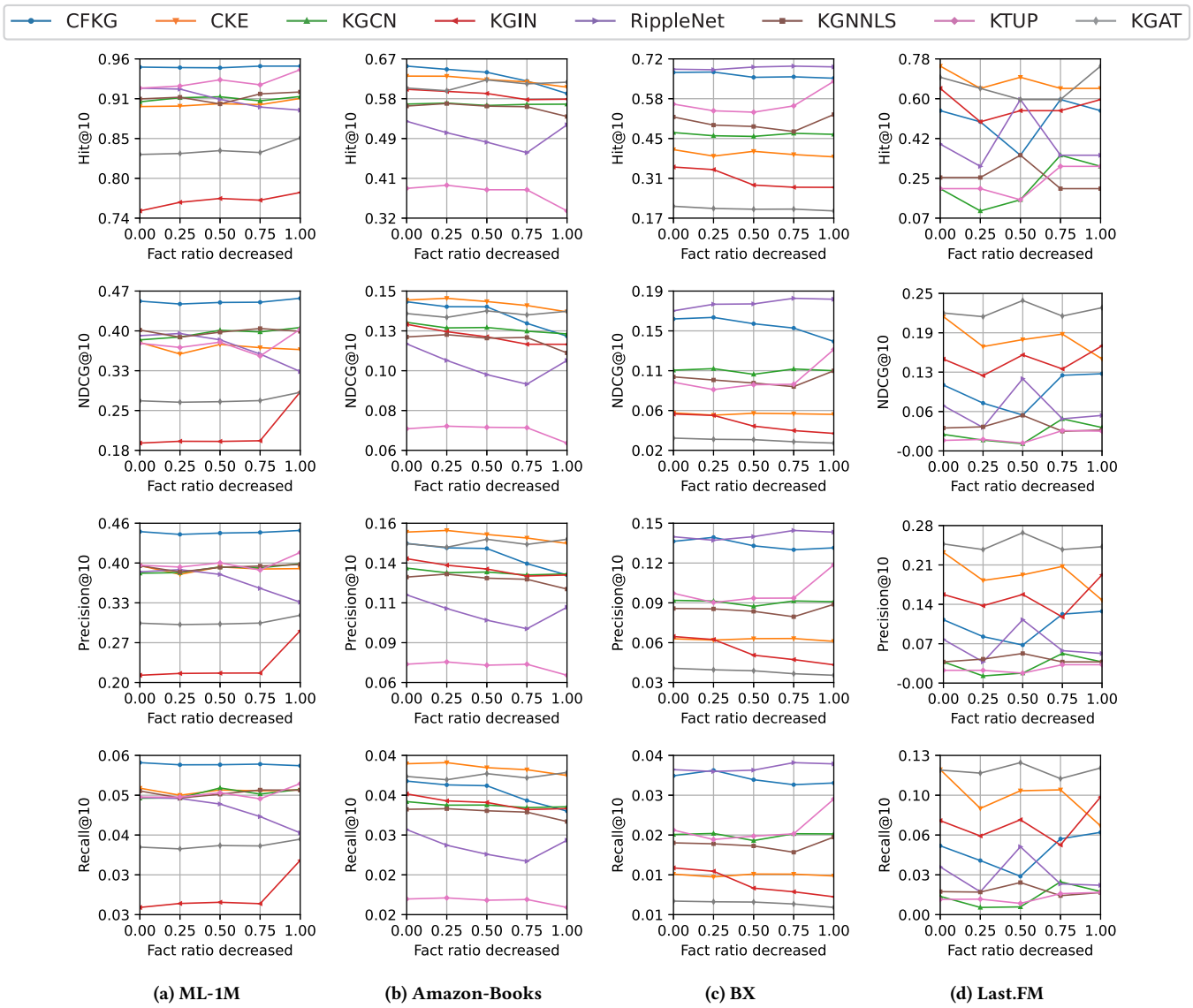
(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D7: Additional result of cold-start experiment with false knowledge, T=5. Horizontal axis denotes distortion degree.The results are similar to those when T=3, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.



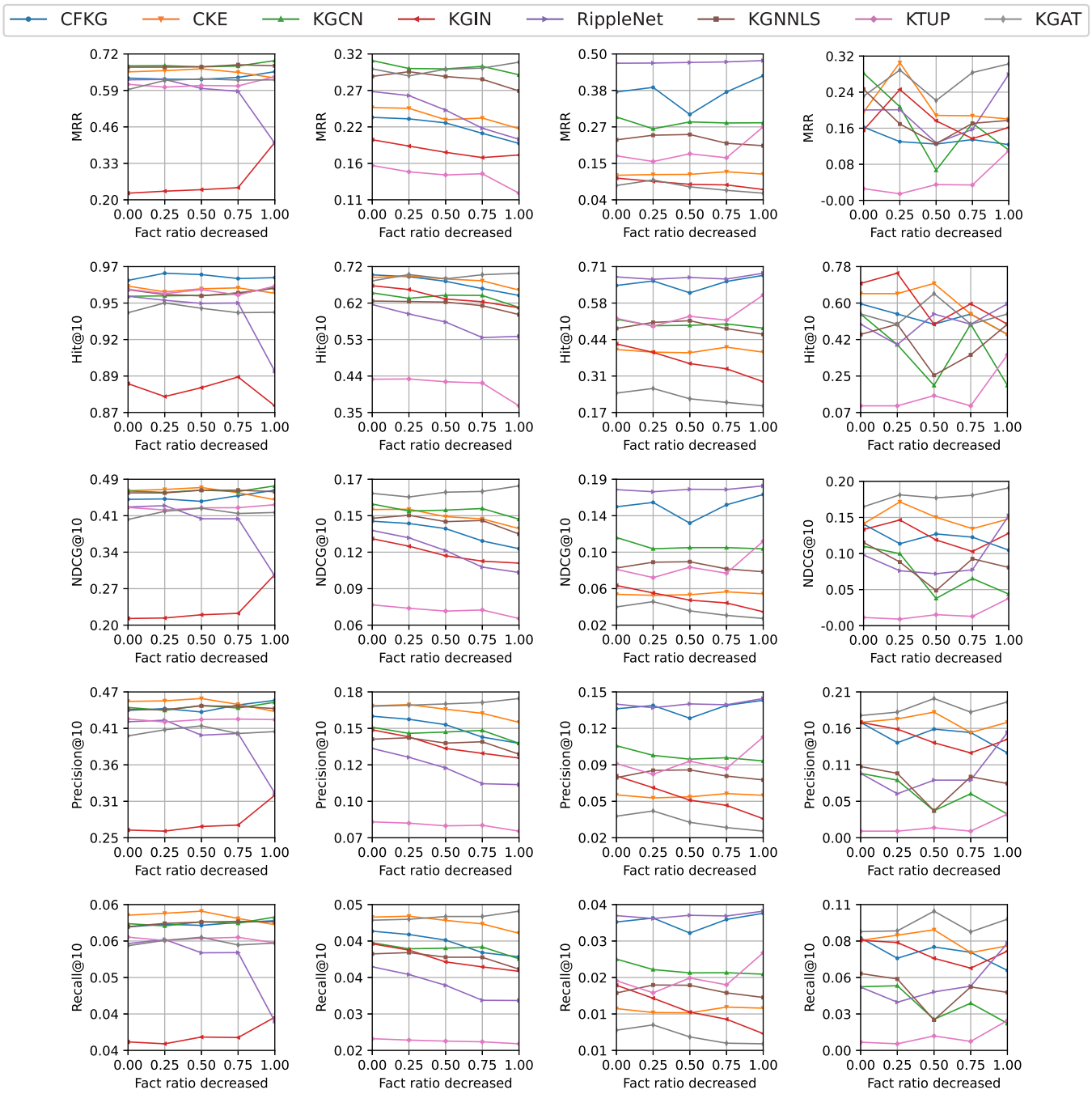
(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D8: Additional result of cold-start experiment with decreased facts, T=1. Horizontal axis denotes distortion degree.The results are similar to those when T=3, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.



(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D9: Additional result of cold-start experiment with decreased facts, T=3. Horizontal axis denotes distortion degree.The results are similar to those when using MRR as metric, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.



(a) ML-1M (b) Amazon-Books (c) BX (d) Last.FM

Figure D10: Additional result of cold-start experiment with decreased facts, T=5. Horizontal axis denotes distortion degree.The results are similar to those when T=3, i.e., how cold-start performance of RSs is influenced depends on both the model and the dataset being used.

E ADDITIONAL RESULTS OF KGER

Additional experimental results of KGER are presented in Table E1-E14.

Table E1: Under the normal settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.25.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	-0.002	0.007	0.149	0.021	-0.007	0.009	0.136	0.002
CKE	0.001	-0.002	0.005	0.005	-0.055	0.089	0.027	0.044
KGCN	0.014	0.004	-0.020	0.038	-0.028	0.003	0.175	0.227
KGIN	0.021	-0.010	0.068	0.016	0.181	0.081	0.209	0.037
RippleNet	0.000	0.000	0.122	0.077	0.025	-0.004	0.004	0.028
KGNNLS	-0.001	-0.017	-0.010	-0.001	0.065	0.045	0.021	0.024
KTUP	0.003	-0.015	-0.012	0.003	0.049	-0.019	-0.059	0.004
KGAT	-0.011	0.005	0.074	-0.001	0.157	-0.115	0.104	-0.024

Table E2: Under the normal settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.75.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.002	0.001	0.140	0.031	0.006	-0.004	0.093	0.008
CKE	-0.003	0.004	0.009	0.007	0.009	0.006	0.002	0.012
KGCN	-0.001	-0.002	-0.013	0.010	0.038	0.036	0.045	0.042
KGIN	0.018	0.010	0.060	0.021	0.126	0.082	0.094	-0.045
RippleNet	0.007	0.021	0.077	0.030	0.055	-0.010	-0.010	0.019
KGNNLS	0.004	-0.005	0.012	0.012	0.051	0.015	0.011	0.013
KTUP	0.011	0.005	0.004	0.003	0.011	-0.004	0.002	-0.012
KGAT	0.004	-0.001	0.071	0.001	0.078	0.034	0.031	-0.010

Table E3: Under the normal settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 1.0.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.002	0.000	0.130	0.042	-0.005	-0.005	0.068	0.023
CKE	0.002	0.004	0.011	0.014	0.002	0.014	0.009	0.009
KGCN	0.001	-0.006	-0.000	0.006	0.010	0.057	0.045	0.035
KGIN	0.033	0.129	0.054	0.017	0.102	0.088	0.070	-0.060
RippleNet	0.016	0.063	0.077	0.034	0.056	-0.005	0.010	-0.026
KGNNLS	0.001	-0.006	0.007	0.016	0.038	0.019	0.017	0.016
KTUP	0.002	-0.000	0.001	-0.009	-0.002	0.009	0.010	-0.042
KGAT	0.009	-0.001	0.048	-0.001	0.054	0.030	0.043	-0.020

Table E4: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.25, T=1.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	-0.013	0.021	0.282	0.079	-0.174	-0.568	1.503	0.995
CKE	-0.057	0.262	0.069	0.047	0.486	-0.178	-0.629	-0.022
KGCN	0.040	-0.052	-0.025	-0.100	-0.058	-0.058	0.715	2.016
KGIN	-0.015	-0.043	0.200	0.112	0.414	0.338	0.834	-0.355
RippleNet	0.055	0.010	0.174	0.201	-0.505	-0.057	2.308	-1.637
KGNNLS	-0.068	-0.031	0.022	-0.023	0.495	0.402	0.828	1.454
KTUP	-0.097	-0.525	-0.157	-0.305	-0.061	0.269	-5.074	-6.717
KGAT	0.125	-0.031	0.113	-0.008	0.617	1.123	0.682	0.161

Table E5: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.5, T=1.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.030	-0.009	0.203	0.106	0.249	0.004	0.286	0.136
CKE	0.153	0.208	-0.021	0.018	0.224	-0.069	-0.054	-0.235
KGCN	0.035	-0.000	-0.019	-0.047	0.028	-0.127	1.014	1.105
KGIN	-0.031	0.004	0.166	0.140	0.383	0.289	0.139	-0.163
RippleNet	0.166	0.072	0.259	0.247	-0.256	-0.095	0.529	-0.041
KGNNLS	-0.024	-0.028	0.007	-0.007	0.080	-0.179	-1.012	0.253
KTUP	-0.019	-0.251	-0.095	-0.069	-0.077	-0.035	1.465	-0.348
KGAT	0.030	-0.028	0.080	0.018	0.514	0.879	0.268	0.241

Table E6: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.75, T=1.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.002	-0.004	0.243	0.118	0.006	-0.108	0.428	0.363
CKE	0.047	0.031	0.036	0.042	0.135	-0.037	0.121	-0.278
KGCN	0.005	-0.048	0.049	0.014	0.033	-0.058	0.425	0.249
KGIN	0.001	0.000	0.176	0.119	0.278	0.382	0.132	0.079
RippleNet	0.097	0.144	0.265	0.220	-0.171	-0.065	0.020	0.459
KGNNLS	0.038	0.003	0.072	0.055	0.119	-0.064	0.827	-0.756
KTUP	-0.031	0.031	0.006	-0.030	-0.052	0.099	0.098	-2.526
KGAT	0.049	-0.048	0.074	0.011	0.294	0.529	0.443	-0.524

Table E7: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 1.0, T=1.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.013	-0.026	0.231	0.116	-0.028	-0.332	0.188	0.348
CKE	0.027	0.020	0.017	0.080	0.084	-0.107	-0.082	-0.022
KGCN	0.003	0.001	0.010	-0.001	-0.040	-0.042	-0.064	0.057
KGIN	-0.054	-0.589	0.159	0.118	0.247	0.328	-0.107	0.005
RippleNet	0.203	0.189	0.229	0.125	-0.128	-0.032	0.572	-0.066
KGNNLS	-0.017	-0.027	0.059	0.052	0.155	-0.015	-0.004	-0.515
KTUP	-0.150	-0.280	0.038	0.169	-0.067	-0.341	0.269	-1.820
KGAT	0.050	-0.080	0.068	-0.011	0.317	0.423	0.090	-0.111

Table E8: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.25, T=3.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.040	0.034	0.159	0.053	-0.527	0.047	0.062	0.472
CKE	0.232	0.371	0.040	-0.032	-0.083	0.312	0.568	0.890
KGCN	-0.106	-0.173	-0.055	0.074	0.063	-0.086	-7.153	-4.312
KGIN	0.037	-0.087	0.231	0.105	0.646	-0.012	1.340	1.647
RippleNet	0.171	-0.054	0.222	0.312	-0.315	-0.463	0.961	0.905
KGNNLS	0.005	0.139	0.080	-0.010	-0.185	0.321	0.353	0.016
KTUP	0.251	0.191	-0.178	-0.089	-0.300	0.333	-5.927	-0.882
KGAT	0.058	0.025	0.096	0.069	0.314	0.245	0.425	-0.337

Table E9: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.75, T=3.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.036	0.002	0.262	0.098	-0.172	0.084	0.404	-0.333
CKE	0.085	0.061	0.035	0.024	-0.054	0.086	0.254	0.277
KGCN	-0.009	-0.088	0.008	0.052	-0.024	-0.035	-1.553	-3.671
KGIN	0.006	-0.044	0.157	0.121	0.436	0.347	-0.007	-0.349
RippleNet	0.141	0.122	0.225	0.268	-0.102	-0.163	-1.377	0.465
KGNNLS	-0.029	-0.039	0.037	-0.030	-0.038	0.191	-1.463	0.557
KTUP	0.197	0.208	-0.013	-0.034	-0.107	-0.005	0.402	-2.325
KGAT	0.048	-0.017	0.086	0.010	0.274	0.131	0.150	0.132

Table E10: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 1.0, T=3.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.022	-0.021	0.246	0.133	-0.078	0.273	0.310	-0.156
CKE	0.045	0.076	0.034	0.056	-0.040	0.032	0.219	0.055
KGCN	-0.003	-0.086	0.019	0.082	-0.022	0.003	-1.497	-1.872
KGIN	-0.053	-0.693	0.161	0.101	0.278	0.249	0.058	0.214
RippleNet	0.186	0.200	0.224	0.072	-0.073	-0.122	0.285	-0.037
KGNNLS	-0.011	0.027	0.098	0.099	0.018	-0.082	0.081	0.305
KTUP	-0.033	-0.103	0.007	0.199	-0.039	-0.504	-1.470	-1.633
KGAT	0.041	-0.099	0.075	0.011	0.095	0.166	-0.085	-0.128

Table E11: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.25, T=5.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.082	0.024	0.315	0.038	-0.209	-0.151	1.037	0.802
CKE	-0.158	-0.028	0.123	0.017	-0.092	-0.072	0.205	-2.265
KGCN	-0.040	-0.005	-0.018	0.148	-0.195	0.497	1.008	1.063
KGIN	-0.188	-0.130	0.246	0.189	0.289	0.376	-0.187	-2.336
RippleNet	0.121	-0.008	0.135	0.088	-0.252	-0.005	-0.066	-0.007
KGNNLS	0.007	-0.000	0.093	-0.095	0.240	-0.257	1.859	1.260
KTUP	-0.054	0.068	-0.005	0.231	-0.122	0.403	-7.636	1.866
KGAT	-0.072	-0.221	0.283	0.135	0.197	-0.826	-0.039	-1.011

Table E12: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.5, T=5.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.074	0.014	0.344	0.070	-0.118	0.375	0.443	0.467
CKE	-0.014	-0.034	0.129	0.145	-0.082	-0.048	0.531	0.073
KGCN	-0.008	0.010	-0.012	0.078	-0.000	0.104	1.163	1.534
KGIN	-0.098	-0.115	0.286	0.187	0.437	0.363	0.254	-0.268
RippleNet	0.064	0.099	0.173	0.203	-0.199	-0.011	-1.257	0.734
KGNNLS	-0.009	-0.003	0.093	0.001	0.206	-0.148	0.589	0.986
KTUP	0.014	0.014	-0.021	0.173	-0.050	-0.073	-1.467	-0.733
KGAT	0.008	-0.130	0.168	0.003	0.308	0.103	0.921	0.084

Table E13: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 0.75, T=5.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.063	-0.005	0.340	0.135	0.190	0.000	0.090	0.230
CKE	0.018	0.004	0.077	0.084	0.012	-0.126	0.380	0.055
KGCN	-0.013	0.001	0.009	0.036	0.072	0.084	-0.511	0.529
KGIN	-0.042	-0.120	0.221	0.177	0.382	0.266	0.230	0.162
RippleNet	0.072	0.088	0.249	0.267	-0.119	-0.012	-1.224	0.290
KGNNLS	0.003	-0.017	0.118	0.019	0.141	0.063	0.563	0.412
KTUP	0.065	0.013	-0.011	0.101	-0.064	0.046	-1.304	-0.444
KGAT	0.027	-0.077	0.145	-0.005	0.289	0.241	0.204	-0.304

Table E14: Under the cold-start settings, comparison of KGER under false knowledge and decreased knowledge, distortion (F) / decreased (D) ratio is 1.0, T=5.

	ML-1M		Amazon-Books		BX		Last.FM	
	F	D	F	D	F	D	F	D
CFKG	0.022	-0.036	0.327	0.166	-0.156	-0.135	0.198	0.241
CKE	-0.013	0.033	0.068	0.126	-0.009	-0.031	0.159	0.078
KGCN	-0.011	-0.028	0.029	0.066	-0.101	0.061	0.283	0.607
KGIN	-0.050	-0.830	0.203	0.113	0.288	0.334	0.049	-0.038
RippleNet	0.263	0.363	0.226	0.261	-0.094	-0.018	-0.424	-0.396
KGNNLS	-0.009	-0.008	0.096	0.073	0.010	0.081	0.401	0.283
KTUP	0.000	-0.045	0.062	0.254	-0.165	-0.515	-0.778	-3.435
KGAT	0.038	-0.059	0.119	-0.033	0.273	0.289	0.417	-0.312