Table 2: Test performance of different methods. Best results are in bold face. (*Improve.* denotes the improvement of TransCF over the best competitor, which is CML.)

Datasets	Metrics	WMF	BPR	CDAE	NeuMF	CML	TransCF^{alt}	TransCF	*Improve.*
Delicious	HR@10	0.1503±0.0047	0.1981±0.0065	0.1059±0.0035	0.1164±0.0159	0.2470±0.0021	0.2174±0.0037	**0.2557**±0.0049	3.52%
	HR@20	0.2531±0.0039	0.3177±0.0111	0.1747±0.0067	0.2171±0.0184	0.3649±0.0082	0.3084±0.0060	**0.3823**±0.0042	4.77%
	NDCG@10	0.0826±0.0017	0.1122±0.0032	0.0531±0.0029	0.0558±0.0075	0.1389±0.0014	0.1281±0.0009	**0.1444**±0.0034	3.96%
	NDCG@20	0.1078±0.0005	0.1418±0.0028	0.0707±0.0040	0.0789±0.0085	0.1678±0.0018	0.1494±0.0014	**0.1762**±0.0013	5.01%
	MRR@10	0.0624±0.0010	0.0875±0.0025	0.0379±0.0038	0.0383±0.0056	0.1066±0.0012	0.1025±0.0007	**0.1117**±0.0026	4.78%
	MRR@20	0.0689±0.0010	0.0953±0.0023	0.0427±0.0040	0.0444±0.0059	0.1143±0.0011	0.1082±0.0006	**0.1205**±0.0023	5.42%
Ciao	HR@10	0.1493±0.0005	0.2027±0.0027	0.1612±0.0034	0.1535±0.0023	0.2085±0.0019	0.1991±0.0013	**0.2292**±0.0020	9.93%
	HR@20	0.2611±0.0017	0.3564±0.0024	0.2856±0.0024	0.2788±0.0036	0.3337±0.0012	0.3270±0.0033	**0.3740**±0.0022	12.08%
	NDCG@10	0.0739±0.0003	0.0973±0.0013	0.0781±0.0009	0.0741±0.0020	0.1053±0.0006	0.0989±0.0007	**0.1167**±0.0009	10.83%
	NDCG@20	0.1015±0.0003	0.1356±0.0010	0.1090±0.0007	0.1040±0.0015	0.1358±0.0009	0.1309±0.0010	**0.1525**±0.0010	12.30%
	MRR@10	0.0516±0.0004	0.0668±0.0011	0.0536±0.0005	0.0511±0.0023	0.0747±0.0003	0.0698±0.0009	**0.0833**±0.0010	11.51%
	MRR@20	0.0589±0.0004	0.0769±0.0010	0.0618±0.0005	0.0590±0.0020	0.0828±0.0005	0.0784±0.0009	**0.0929**±0.0010	12.20%
Bookcrossing	HR@10	0.1875±0.0011	0.2792±0.0008	0.2207±0.0015	0.2286±0.0018	0.2885±0.0006	0.2828±0.0014	**0.3180**±0.0073	10.23%
	HR@20	0.3127±0.0009	0.4357±0.0020	0.3553±0.0019	0.3747±0.0091	0.4053±0.0007	0.4069±0.0011	**0.4636**±0.0053	14.38%
	NDCG@10	0.0968±0.0004	0.1457±0.0011	0.1126±0.0007	0.1158±0.0006	0.1663±0.0003	0.1578±0.0008	**0.1755**±0.0050	5.53%
	NDCG@20	0.1279±0.0004	0.1850±0.0011	0.1461±0.0010	0.1482±0.0007	0.1956±0.0003	0.1890±0.0007	**0.2121**±0.0045	8.44%
	MRR@10	0.0695±0.0003	0.1059±0.0010	0.0805±0.0007	0.0840±0.0006	0.1294±0.0004	0.1201±0.0006	**0.1328**±0.0043	2.63%
	MRR@20	0.0779±0.0003	0.1166±0.0010	0.0894±0.0008	0.0925±0.0005	0.1374±0.0004	0.1286±0.0005	**0.1428**±0.0041	3.93%
Amazon C&A	HR@10	0.2161±0.0008	0.2968±0.0019	0.1993±0.0012	0.1317±0.0015	0.3011±0.0007	0.3184±0.0006	**0.3436**±0.0040	14.11%
	HR@20	0.3480±0.0008	0.4515±0.0023	0.2906±0.0006	0.2390±0.0025	0.4123±0.0009	0.4509±0.0011	**0.4658**±0.0044	12.98%
	NDCG@10	0.1064±0.0005	0.1563±0.0014	0.1131±0.0007	0.0613±0.0007	0.1752±0.0002	0.1766±0.0008	**0.2019**±0.0038	15.24%
	NDCG@20	0.1397±0.0004	0.1951±0.0016	0.1358±0.0004	0.0880±0.0010	0.2031±0.0004	0.2094±0.0008	**0.2323**±0.0037	14.38%
	MRR@10	0.0739±0.0003	0.1141±0.0014	0.0872±0.0006	0.0405±0.0005	0.1368±0.0003	0.1339±0.0010	**0.1588**±0.0037	16.08%
	MRR@20	0.0830±0.0002	0.1246±0.0014	0.0933±0.0005	0.0477±0.0005	0.1443±0.0003	0.1428±0.0009	**0.1671**±0.0036	15.80%

$\{0.0005, 0.001, 0.005, 0.01, 0.05, 0.1\}$, $\gamma \in \{0.0, 0.1, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$, $\lambda, \lambda_{\text{nbr}}, \lambda_{\text{dist}} \in \{0.0, 0.001, 0.01, 0.1\}$, where λ is the regularization coefficient for the competing methods. All results reported are the test performance measured using the hyperparameter values that give the best HR@10 with the validation set. We fix the number of samples in a mini-batch to 1000 for mini-batch SGD.

5.1 Performance Analysis

5.1.1 ***Recommendation performance (RQ 1)***. Table 2 shows the performance in terms of various ranking metrics averaged over five runs that used different random seeds for the model initializations. We have the following observations from Table 2. First, we observe that CML outperforms the MF-based competitors (WMF, BPR and NeuMF). This is consistent with the previous work [12], indicating that metric learning approaches overcome the inherent limitation of MF by learning a metric space wherein the triangle inequality is satisfied. Second, we observe that TransCF considerably outperforms the state-of-the-art competitor, namely CML, by up to 16.08% (achieved for MRR@10 on Amazon C&A dataset). This verifies the benefit of the translation vectors that translate each user toward items according to the user's relationships with those items. Third, TransCF$^{\text{alt}}$ generally performs worse than CML, which implies that the translation vectors should be carefully designed, or else the performance will rather deteriorate. Fourth, the superior performance of TransCF over TransCF$^{\text{alt}}$ confirms that incorporating the neighborhood information is indeed crucial in collaborative filtering [15, 17, 27]. Lastly, it is noteworthy that the performance improvements on Delicious dataset are smaller than those on the rest of the datasets. We attribute this to the fact that Delicious dataset is a small dataset containing relatively fewer items than the other datasets. The user–item interactions pertaining to CF are less complex in such a dataset, and thus projecting a user to a single point, as is done in CML, yields comparable results.

Table 3: Effect of the regularization coefficients.

Ciao (HR@10)		λ_{nbr} 0.0	0.001	0.01	0.1
λ_{dist}	0.0	0.2069	0.2098	0.2193	0.2181
	0.001	0.2096	0.2119	0.2206	0.2183
	0.01	0.2194	0.2206	0.2240	0.2180
	0.1	**0.2320**	0.2310	0.2289	0.2196

Delicious (HR@10)		λ_{nbr} 0.0	0.001	0.01	0.1
λ_{dist}	0.0	0.2388	0.2388	0.2401	0.2454
	0.001	0.2375	0.2401	0.2401	0.2454
	0.01	0.2401	0.2401	0.2427	0.2467
	0.1	0.2520	0.252	0.2533	**0.2586**

5.1.2 ***Benefit of regularizers (RQ 2)***. Table 3 shows the effect of the regularization coefficients on the performance of TransCF on Ciao and Delicious datasets, where λ_{nbr} and λ_{dist} denote the strengths of the neighborhood regularizer (reg_{nbr}) and of the distance regularizer (reg_{dist}), respectively. Larger values imply a stronger contribution of the regularizer to the model, and $\lambda_* = 0.0$ indicates no regularization. We have the following observations: 1) The regularizers are indeed beneficial for the model performance, and their impact varies across different datasets. 2) Although reg_{nbr} is beneficial, its impact on the model performance decreases as the reg_{dist} dominates the model; i.e., as λ_{dist} increases. 3) reg_{dist} is especially helpful for the model performance. This confirms that explicitly pulling each translated user toward all of the positive items rather than merely pushing negative items away from each user helps to model the complex user–item interactions.

5.2 Qualitative Evaluations

5.2.1 ***Benefit of translation vectors (RQ 3)***. In this section, we conduct experiments to verify whether the translation vectors learned by TransCF indeed translate each user closer to the observed (positive) items as in the **Toy example** shown in Figure 1. To this