Prediction Is NOT Classification: On Formulation and Evaluation of Hyperedge Prediction

Anonymous Author(s)

A EXPERIMENTAL SETTINGS

A.1 Datasets

In this section, we provide the details of six real-world hypergraphs used in the paper.

- **Co-authorship datasets (Cora-A[1] [8])**: Each co-authorship dataset represents authors collaborating to write academic papers. Since a single author writes several papers, each paper becomes a node, and the papers of the same author form a hyperedge. The bag-of-words representation of the paper serves as the node features.
- **Co-citation datasets (Citeseer[2] [7], Cora[3] [8], and Pubmed[4] [5])**: Each co-citation dataset involves papers that reference each other through citations. Each node represents a paper, and each hyperedge corresponds to the collection of papers that have been cited by a paper. The node features are the same bag-of-words representation as the co-authorship dataset.
- **Metabolic reaction datasets (iAF1260b[5] and iJO1366 [6] [4])**: A metabolic reaction dataset expresses metabolic pathways, where nodes symbolize specific materials, and hyperedges represent reactions among these materials. The node features are the counts of the atoms (e.g. C, H, O) within each material.

We adopted the data preprocessing method used in [2, 9] to remove all duplicate hyperedges. Basic statistics can be found in Table 1.

A.2 Baselines

Heuristic methods We use four popular heuristic methods as our baselines:

- **Common Neighbors (CN).**

$$CN = |N(v_i) \cap N(v_j)| \tag{1}$$

- **Jaccard Index (JI).**

$$JI = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i)| + |N(v_j)|} \tag{2}$$

- **Adamic-Adar (AA).**

$$AA = \sum_{v \in N(v_i) \cap N(v_j)} \frac{1}{\log |N(z)|} \tag{3}$$

- **Katz Index.**

$$KZ = \sum_{l=1}^{\infty} \beta^{l} |path(v_i, v_j) = l| \tag{4}$$

[1] https://people.cs.umass.edu/~mccallum/data.html
[2] https://linqs.org/datasets/#citeseer-doc-classification
[3] https://linqs.org/datasets/#cora
[4] https://linqs.org/datasets/#pubmed-diabetes
[5] http://bigg.ucsd.edu/models/iAF1260b
[6] http://bigg.ucsd.edu/models/iJO1366

Table 1: Dataset statistics.

	Category	# Nodes	# Hedges*	# Features	Hedge* Size	
					Avg.	Max.
Cora-A	Co-authorship	2,388	1,072	1,433	4.28	43
Citeseer	Co-citation	1,458	1,079	3,703	3.20	26
Cora	Co-citation	1,434	1,579	1,433	3.03	5
Pubmed	Co-citation	3,840	7,962	500	4.35	171
iAF1260b	Metabolic reaction	1,668	2,084	26	4.30	67
iJO1366	Metabolic reaction	1,805	2,253	26	4.37	106

*Hedges indicate hyperedges.

For all, $N(v)$ stands for neighbors of v, and $path(v_i, v_j)$ stands for number of paths between v_i and v_j. Since the heuristic methods are originally defined in a graph manner, we transform the hypergraph into an ordinary graph using clique expansion [3]. Clique expansion is a common expansion method from graph to hypergraph, which connects all node pairs that exist within the same hyperedge with an edge.

Hyperparameter Settings. Heuristic methods (CN, JI, AA, Katz) do not have hyperparameters. For the deep-learning-based baselines, we conduct a search for the embedding dimension $d \in \{64, 128, 256\}$ and the learning rate $lr \in \{1e-3, 5e-4, 1e-4\}$. As AHP involves several other hyperparameters, we adopted the hyperparameters reported as the best in the original paper for our experiments. We give in Table 2 the hyperparameter search space of AHP and its variants.

A.3 MHP

We report the hyperparameter search space of MHP in Table 3.

B COMPLEXITY ANALYSIS

The time complexity of the beam search is $O(k|V|(s-|Q|))$. To achieve this, we maintain the node embeddings of size $O(|V| \cdot d)$ from HGNN and reuse them in the other steps. For beam search, we select the top-k nodes among $O(|V|)$ nodes in each of $s-|Q|$ steps; and thus, the running time of the beam search is $O(k|V|(s-|Q|))$.

C ADDITIONAL EXPERIMENTS

C.1 Full experiments

In this section, we provide the full experimental results of three scenarios in Table 5, Table 6, and Table 7, respectively.

Experimental settings. Evaluation protocols are the same as in the main paper, while we add two additional measures to evaluate the prediction performance of MHP. To evaluate a variety of aspects in hyperedge prediction (HP) performance, we used Hits@k, Mean Reciprocal Rank (MRR@k), and our defined metric Coverage@k. *Coverage*, is defined as $|\widehat{E} \cap E_T| / |E_T|$, where $\widehat{E}$ is the union of the answers for all the query node sets. Note that larger coverage indicates better performance. We fixed $k = 10$.

Table 2: Hyperparameter search space of AHP and its variants

Hyperparameter	Selection pool
Optimizer for the discriminator	Adam
Optimizer for the generator	Adam
Maximum epoch*	400
Learning rate for the discriminator	5e-03, 5e-04, 5e-05, 5e-06
Learning rate for the generator	1e-04, 1e-05, 1e-06, 1e-07
Normalization factor (α, β)**	(0,0), (1,1)
Memory size	0, 32, 128

(a) Search space for AHP

Hyperparameter	Selection pool
Optimizer	Adam
Learning rate	5e-02, 5e-03, 5e-04, 5e-05, 5e-06
Normalization factor (α, β)	(0,0), (1,1)
Maximum epoch*	200

*Early stopped when the validation AUROC was maximized.

**α and β are degree normalization factors of hyperedges and nodes [1].

(b) Search space for the variants of AHP

Table 3: Hyperpameter search space of MHP

Hyperparameter	Search Space
Maxmimum Epoch	500
Number of HGNN layers	$\{1, 2\}$
Dimension of hidden embedding	$\{64, 128\}$
Number of samplings for each hedge	$\{2, 5, 8, 10, 12, 15\}$
Optimizer	Adam
Learning rate	$\{1e-3, 5e-4, 1e-4\}$
Batch size	$\{128, 256, 512\}$

Experimental results. In all datasets, measures, and scenarios, MHP significantly outperforms the baseline models and achieves the best rank overall (Tables 5-7).

C.2 Ablation studies

We show that MHP performs well regardless of the selection of k. We conducted single-node removal scenario experiments on the iJO1366 dataset for MHP and our baselines with various $k \in \{1, 5, 10, 20, 50\}$. As in Figure 1, MHP outperformed all baselines on all k values. Similar results can be made across the other datasets and measures.

In addition, we present the ablation study of the MHP on the iJO1366 dataset in Table 4, focusing on a single-node removal scenario. Five distinct models were considered in this study. The first four models involved employing the classification formulation of MHP, wherein the model was trained to distinguish between positive and negative hyperedges. To accomplish this, we employed three different negative sampling strategies, SNS, MNS, and CNS, defined as follows:

- **Sized negative sampling (SNS)**: fill each hyperedge with nodes drawn uniformly at random,

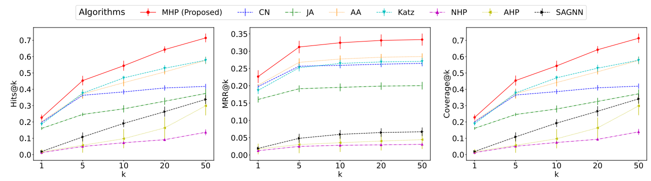

Figure 1: Single-node removal scenario on iJO1366 dataset. MHP performs best regardless of k, with respect to Hits@k, MRR@k, and Coverage@k.

Table 4: Ablation study on iJO1366 dataset. Single-node removal scenario.

Method	Hits@10	MRR@10
MHP + Classification Form. (SNS)	0.1730 ± 0.0229	0.0803 ± 0.0121
MHP + Classification Form. (MNS)	0.1410 ± 0.0119	0.0451± 0.0125
MHP + Classification Form. (CNS)	0.0044 ± 0.0027	0.0004 ± 0.0004
MHP + Classification Form. (ALL)	0.0820 ± 0.0211	0.0301 ± 0.0074
MHP - Structure Embedding	0.4741 ± 0.0353	0.2750 ± 0.0263
MHP	**0.5415 ± 0.0369**	**0.3217 ± 0.0214**

- **Motif negative sampling (MNS)**: grow each hyperedge by repeatedly adding adjacent nodes,
- **Clique negative sampling (CNS)**: pick a hyperedge at random and replace a constituent node (chosen at random) with a node that is adjacent to all other constituent nodes (chosen at random).

ALL stands for using both SNS, MNS, and CNS together (refer to [6] for their details). While both structure embedding and our novel formulation exhibited utility, our new formulation played a more crucial role in enhancing performance. Similar results can be made across the other datasets.

REFERENCES

[1] Yihe Dong, Will Sawin, and Yoshua Bengio. 2020. Hnhn: Hypergraph networks with hyperedge neurons. *arXiv preprint arXiv:2006.12278* (2020).

[2] Hyunjin Hwang, Seungwoo Lee, Chanyoung Park, and Kijung Shin. 2022. Ahp: Learning to negative sample for hyperedge prediction. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2237–2242.

[3] Sunwoo Kim, Soo Yong Lee, Yue Gao, Alessia Antelmi, Mirko Polato, and Kijung Shin. 2024. A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide. *arXiv preprint arXiv:2404.01039* (2024).

[4] Zachary A King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research* 44, D1 (2016), D515–D522.

[5] Galileo Namata, Ben London, Lise Getoor, Bert Huang, and U Edu. 2012. Query-driven active surveying for collective classification. In *MLG*.

[6] Prasanna Patil, Govind Sharma, and M Narasimha Murty. 2020. Negative sampling for hyperlink prediction in networks. In *Advances in Knowledge Discovery and Data Mining*. Springer, 607–619.

[7] Ryan Rossi and Nesreen Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *AAAI*.

[8] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine* 29, 3 (2008), 93–93.

[9] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. 2019. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In *Advances in neural information processing systems*.

Table 5: Hyperedge prediction performance when a single node is masked.

Datasets	Metric	CN	JA	AA	Katz	SAGNN	NHP	AHP	MHP
Cora-A	Hits@10	0.7070 ± 0.0093	0.7377 ± 0.0071	0.7274 ± 0.0107	0.6791 ± 0.0154	0.0140 ± 0.0104	0.1544 ± 0.0260	0.4958 ± 0.0863	**0.7702 ± 0.0252**
	MRR@10	0.3654 ± 0.0107	0.4078 ± 0.0069	0.3907 ± 0.0156	0.3453 ± 0.0147	0.0070 ± 0.0080	0.0698 ± 0.0174	0.3931 ± 0.1014	**0.5033 ± 0.0206**
	Coverage@10	0.7033 ± 0.0111	0.7321 ± 0.0062	0.7237 ± 0.0134	0.6753 ± 0.0159	0.0158 ± 0.0138	0.1544 ± 0.0295	0.4949 ± 0.0855	**0.7647 ± 0.0252**
Citeseer	Hits@10	0.6463 ± 0.0367	0.6231 ± 0.0395	0.6889 ± 0.0408	0.6926 ± 0.0292	0.1194 ± 0.0331	0.4407 ± 0.0318	0.4861 ± 0.0386	**0.7111 ± 0.0370**
	MRR@10	0.3949 ± 0.0278	0.3427 ± 0.0313	0.4269 ± 0.0259	0.4213 ± 0.0308	0.0522 ± 0.0262	0.2159 ± 0.0136	0.3385 ± 0.0452	**0.4692 ± 0.0237**
	Coverage@10	0.5991 ± 0.0371	0.5713 ± 0.0361	0.6343 ± 0.0387	0.6444 ± 0.0277	0.1019 ± 0.0282	0.3870 ± 0.0335	0.4435 ± 0.0341	**0.6509 ± 0.0273**
Pubmed	Hits@10	0.2247 ± 0.0151	0.2393 ± 0.0158	0.3063 ± 0.0138	0.2939 ± 0.0104	0.0290 ± 0.0063	0.0418 ± 0.0010	0.0286 ± 0.0222	**0.4134 ± 0.0345**
	MRR@10	0.1087 ± 0.0080	0.1056 ± 0.0085	0.1474 ± 0.0040	0.1339 ± 0.0037	0.0082 ± 0.0024	0.0120 ± 0.0001	0.0083 ± 0.0085	**0.2239 ± 0.0183**
	Coverage@10	0.2068 ± 0.0112	0.2117 ± 0.0132	0.2754 ± 0.0135	0.2758 ± 0.0071	0.0261 ± 0.0057	0.0360 ± 0.0035	0.0243 ± 0.0176	**0.3704 ± 0.0266**
Cora	Hits@10	0.6000 ± 0.0162	0.5342 ± 0.0255	0.6000 ± 0.0198	**0.6063 ± 0.0180**	0.0741 ± 0.0377	0.3709 ± 0.0374	0.3209 ± 0.0722	0.5918 ± 0.0190
	MRR@10	0.3405 ± 0.0207	0.2607 ± 0.0154	0.3390 ± 0.0159	**0.3578 ± 0.0215**	0.0295 ± 0.0207	0.1687 ± 0.0306	0.1778 ± 0.0429	0.3430 ± 0.0222
	Coverage@10	0.5335 ± 0.0248	0.4734 ± 0.0322	0.5310 ± 0.0192	**0.5386 ± 0.0302**	0.0551 ± 0.0248	0.3076 ± 0.0178	0.2658 ± 0.0626	0.5297 ± 0.0231
iAF1260b	Hits@10	0.4125 ± 0.0277	0.3453 ± 0.0334	0.4489 ± 0.0222	0.4878 ± 0.0371	0.1947 ± 0.0350	0.0647 ± 0.0193	0.0676 ± 0.0745	**0.5376 ± 0.0172**
	MRR@10	0.2648 ± 0.0176	0.2311 ± 0.0209	0.2848 ± 0.0158	0.2718 ± 0.0128	0.0652 ± 0.0143	0.0250 ± 0.0097	0.0204 ± 0.0248	**0.3171 ± 0.0186**
	Coverage@10	0.4153 ± 0.0257	0.3448 ± 0.0330	0.4518 ± 0.0210	0.4882 ± 0.0364	0.1971 ± 0.0346	0.0652 ± 0.0194	0.0681 ± 0.0741	**0.5396 ± 0.0143**
iJO1366	Hits@10	0.3840 ± 0.0168	0.2803 ± 0.0248	0.4404 ± 0.0249	0.4696 ± 0.0146	0.1920 ± 0.0229	0.0687 ± 0.0150	0.0976 ± 0.0654	**0.5415 ± 0.0369**
	MRR@10	0.2591 ± 0.0093	0.1954 ± 0.0126	0.2774 ± 0.0112	0.2646 ± 0.0113	0.0525 ± 0.0102	0.0276 ± 0.0055	0.0355 ± 0.0299	**0.3217 ± 0.0214**
	Coverage@10	0.3858 ± 0.0169	0.2794 ± 0.0240	0.4417 ± 0.0268	0.4705 ± 0.0156	0.1933 ± 0.0237	0.0701 ± 0.0160	0.0971 ± 0.0658	**0.5410 ± 0.0380**
Average Ranking		3.917	4.389	2.806	2.667	7.222	7.000	6.611	**1.389**

Table 6: Hyperedge prediction performance when two nodes are masked.

Datasets	Metric	CN	JA	AA	Katz	SAGNN	NHP	AHP	MHP
Cora-A	Hits@10	0.5414 ± 0.0156	0.5544 ± 0.0156	0.5395 ± 0.0216	0.4930 ± 0.0211	0.0074 ± 0.0096	0.1377 ± 0.0239	0.4084 ± 0.0797	**0.5935 ± 0.0156**
	MRR@10	0.2896 ± 0.0126	0.3258 ± 0.0112	0.3114 ± 0.0197	0.2676 ± 0.0151	0.0050 ± 0.0081	0.0648 ± 0.0151	0.3335 ± 0.0900	**0.4156 ± 0.0136**
	Coverage@10	0.5377 ± 0.0121	0.5507 ± 0.0149	0.5358 ± 0.0193	0.4902 ± 0.0196	0.0093 ± 0.0136	0.1358 ± 0.0274	0.4047 ± 0.0744	**0.5870 ± 0.0162**
Citeseer	Hits@10	0.5046 ± 0.0217	0.4250 ± 0.0267	0.5380 ± 0.0323	0.5333 ± 0.0288	0.0611 ± 0.0237	0.3176 ± 0.0357	0.3333 ± 0.0463	**0.5620 ± 0.0306**
	MRR@10	0.3026 ± 0.0162	0.2386 ± 0.0203	0.3231 ± 0.0199	0.3230 ± 0.0199	0.0294 ± 0.0191	0.1581 ± 0.0164	0.2437 ± 0.0523	**0.3720 ± 0.0096**
	Coverage@10	0.4602 ± 0.0195	0.3806 ± 0.0346	0.4787 ± 0.0290	0.4861 ± 0.0274	0.0491 ± 0.0137	0.2620 ± 0.0265	0.2963 ± 0.038	**0.4935 ± 0.0193**
Pubmed	Hits@10	0.1450 ± 0.0126	0.1380 ± 0.0113	0.1765 ± 0.0096	0.1557 ± 0.0106	0.0178 ± 0.0053	0.0257 ± 0.0023	0.0151 ± 0.0125	**0.2586 ± 0.0257**
	MRR@10	0.0717 ± 0.0072	0.0605 ± 0.0070	0.0895 ± 0.0042	0.0758 ± 0.0048	0.0053 ± 0.0013	0.0071 ± 0.0004	0.0043 ± 0.0045	**0.1448 ± 0.0118**
	Coverage@10	0.1265 ± 0.0071	0.1098 ± 0.0084	0.1486 ± 0.0081	0.1390 ± 0.0068	0.0137 ± 0.0041	0.0203 ± 0.0030	0.0093 ± 0.0075	**0.2117 ± 0.0181**
Cora	Hits@10	0.4582 ± 0.0135	0.3500 ± 0.0196	0.4494 ± 0.0081	**0.4658 ± 0.0088**	0.0614 ± 0.0331	0.2639 ± 0.0238	0.2228 ± 0.0542	0.4519 ± 0.0209
	MRR@10	0.2563 ± 0.0171	0.1738 ± 0.0185	0.2500 ± 0.0135	**0.2689 ± 0.0100**	0.0269 ± 0.0219	0.1245 ± 0.0288	0.1232 ± 0.0301	0.2535 ± 0.0233
	Coverage@10	0.3677 ± 0.0157	0.2835 ± 0.0139	0.3589 ± 0.0113	**0.3778 ± 0.0156**	0.0418 ± 0.0241	0.1956 ± 0.0186	0.1525 ± 0.0430	0.3633 ± 0.0189
iAF1260b	Hits@10	0.1434 ± 0.0147	0.0940 ± 0.0064	0.1400 ± 0.0162	0.1362 ± 0.0151	0.0120 ± 0.0051	0.0168 ± 0.0051	0.0043 ± 0.0049	**0.1578 ± 0.0316**
	MRR@10	0.0867 ± 0.0089	0.0674 ± 0.0073	0.0847 ± 0.0103	0.0824 ± 0.0107	0.0041 ± 0.0032	0.0082 ± 0.0020	0.0019 ± 0.0025	**0.0915 ± 0.0168**
	Coverage@10	0.1400 ± 0.0145	0.0854 ± 0.0075	0.1329 ± 0.0203	0.1276 ± 0.0168	0.0120 ± 0.0034	0.0168 ± 0.0051	0.0043 ± 0.0049	**0.1482 ± 0.0294**
iJO1366	Hits@10	0.1326 ± 0.0174	0.0736 ± 0.0073	0.1290 ± 0.0197	0.1268 ± 0.0273	0.0191 ± 0.0112	0.0226 ± 0.0082	0.0044 ± 0.0054	**0.1512 ± 0.0200**
	MRR@10	0.0818 ± 0.0133	0.0518 ± 0.0102	0.0825 ± 0.0143	0.0755 ± 0.0136	0.0047 ± 0.0012	0.0116 ± 0.0056	0.0018 ± 0.0031	**0.0932 ± 0.0153**
	Coverage@10	0.1277 ± 0.0167	0.0692 ± 0.0088	0.1224 ± 0.0206	0.1215 ± 0.0288	0.0142 ± 0.0040	0.0226 ± 0.0082	0.0044 ± 0.0054	**0.1446 ± 0.0224**
Average Ranking		3.000	4.611	3.000	3.333	7.500	6.333	6.889	**1.333**

Table 7: Hyperedge prediction performance when half of nodes are masked.

Datasets	Metric	CN	JA	AA	Katz	SAGNN	AHP	NHP	MHP
Cora-A	Hits@10	0.6140 ± 0.0093	0.6335 ± 0.0129	0.6121 ± 0.0107	0.5702 ± 0.0153	0.0112 ± 0.0071	0.1656 ± 0.0341	0.4437 ± 0.0776	**0.6595 ± 0.0187**
	MRR@10	0.3278 ± 0.0065	0.3725 ± 0.0142	0.3474 ± 0.0190	0.3018 ± 0.0130	0.0057 ± 0.0074	0.0746 ± 0.0171	0.3649 ± 0.0963	**0.4583 ± 0.0246**
	Coverage@10	0.6102 ± 0.0120	0.6316 ± 0.0116	0.6084 ± 0.0137	0.5674 ± 0.0164	0.0121 ± 0.0096	0.1647 ± 0.0370	0.4419 ± 0.0745	**0.6530 ± 0.0176**
Citeseer	Hits@10	0.5781 ± 0.0392	0.5270 ± 0.0200	0.6190 ± 0.0404	0.6069 ± 0.0449	0.0966 ± 0.0375	0.3913 ± 0.0326	0.4275 ± 0.0329	**0.6376 ± 0.0471**
	MRR@10	0.3537 ± 0.0209	0.3020 ± 0.0178	0.3856 ± 0.0181	0.3739 ± 0.0263	0.0367 ± 0.0200	0.1964 ± 0.0125	0.3114 ± 0.0383	**0.4211 ± 0.0152**
	Coverage@10	0.5389 ± 0.0447	0.4815 ± 0.0278	0.5713 ± 0.0434	0.5657 ± 0.0461	0.0722 ± 0.0273	0.3352 ± 0.0367	0.3889 ± 0.0386	**0.5852 ± 0.0475**
Pubmed	Hits@10	0.1814 ± 0.0154	0.1736 ± 0.0163	0.2300 ± 0.0130	0.2057 ± 0.0149	0.0270 ± 0.0034	0.0366 ± 0.0040	0.0188 ± 0.0170	**0.3204 ± 0.0290**
	MRR@10	0.0894 ± 0.0065	0.0764 ± 0.0066	0.1152 ± 0.0050	0.0959 ± 0.0066	0.0077 ± 0.0020	0.0113 ± 0.0026	0.0055 ± 0.0057	**0.1758 ± 0.0133**
	Coverage@10	0.1552 ± 0.0111	0.1414 ± 0.0117	0.1896 ± 0.0124	0.1794 ± 0.0118	0.0217 ± 0.0025	0.0278 ± 0.0042	0.0130 ± 0.0108	**0.2655 ± 0.0248**
Cora	Hits@10	0.5519 ± 0.0240	0.4570 ± 0.0130	0.5468 ± 0.0188	**0.5690 ± 0.0152**	0.0772 ± 0.0380	0.3184 ± 0.0428	0.2703 ± 0.0601	0.5519 ± 0.0248
	MRR@10	0.3050 ± 0.0270	0.2250 ± 0.0176	0.3092 ± 0.0248	**0.3306 ± 0.0293**	0.0301 ± 0.0225	0.1494 ± 0.0305	0.1466 ± 0.0389	0.3065 ± 0.0219
	Coverage@10	0.4810 ± 0.0235	0.4006 ± 0.0057	0.4753 ± 0.0214	**0.4930 ± 0.0190**	0.0557 ± 0.0309	0.2551 ± 0.0202	0.2203 ± 0.0567	0.4899 ± 0.0140
iAF1260b	Hits@10	0.1334 ± 0.0074	0.0806 ± 0.0088	0.1382 ± 0.0115	0.1329 ± 0.0146	0.0163 ± 0.0102	0.0178 ± 0.0055	0.0014 ± 0.0032	**0.1454 ± 0.0213**
	MRR@10	0.0794 ± 0.0086	0.0515 ± 0.0081	0.0793 ± 0.0082	0.0771 ± 0.0090	0.0076 ± 0.0087	0.0083 ± 0.0035	0.0007 ± 0.0015	**0.0795 ± 0.0087**
	Coverage@10	0.1309 ± 0.0087	0.0734 ± 0.0047	0.1290 ± 0.0098	0.1242 ± 0.0128	0.0129 ± 0.0060	0.0177 ± 0.0055	0.0014 ± 0.0032	**0.1376 ± 0.0180**
iJO1366	Hits@10	0.1195 ± 0.0114	0.0662 ± 0.0029	0.1226 ± 0.0147	0.1062 ± 0.0087	0.0204 ± 0.0055	0.0249 ± 0.0043	0.0040 ± 0.0040	**0.1297 ± 0.0187**
	MRR@10	0.0702 ± 0.0057	0.0440 ± 0.0039	0.0708 ± 0.0077	0.0552 ± 0.0040	0.0058 ± 0.0015	0.0124 ± 0.0029	0.0016 ± 0.0018	**0.0715 ± 0.0137**
	Coverage@10	0.1166 ± 0.0146	0.0594 ± 0.0045	0.1180 ± 0.0163	0.1024 ± 0.0079	0.0186 ± 0.0062	0.0248 ± 0.0043	0.0040 ± 0.0040	**0.1206 ± 0.0187**
Average Ranking		3.361	4.556	2.667	3.389	7.500	6.333	6.944	**1.250**