TaxoComplete: Self-Supervised Taxonomy Completion Leveraging Position-Enhanced Semantic Matching

Ines Arous
University of Fribourg
Fribourg, Switzerland
ines@exascale.info

Ljiljana Dolamic
armasuisse S+T
Thun, Switzerland
ljiljana.dolamic@armasuisse.ch

Philippe Cudré-Mauroux
University of Fribourg
Fribourg, Switzerland
pcm@unifr.ch

Code: https://github.com/eXascaleInfolab/TaxoComplete

2023. 6. 25 • ChongQing

—— WWW 2023

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Renhui Luo

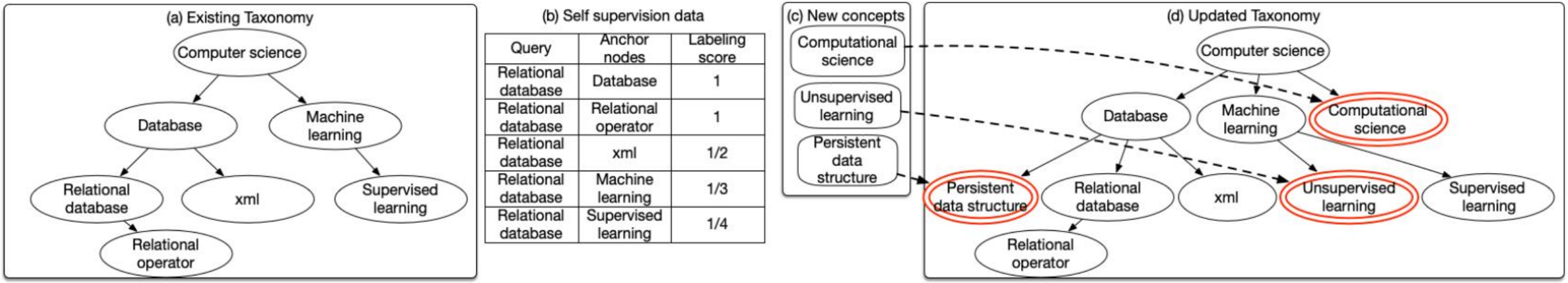
1.Introduction

2.Overview

3.Methods

4.Experiments

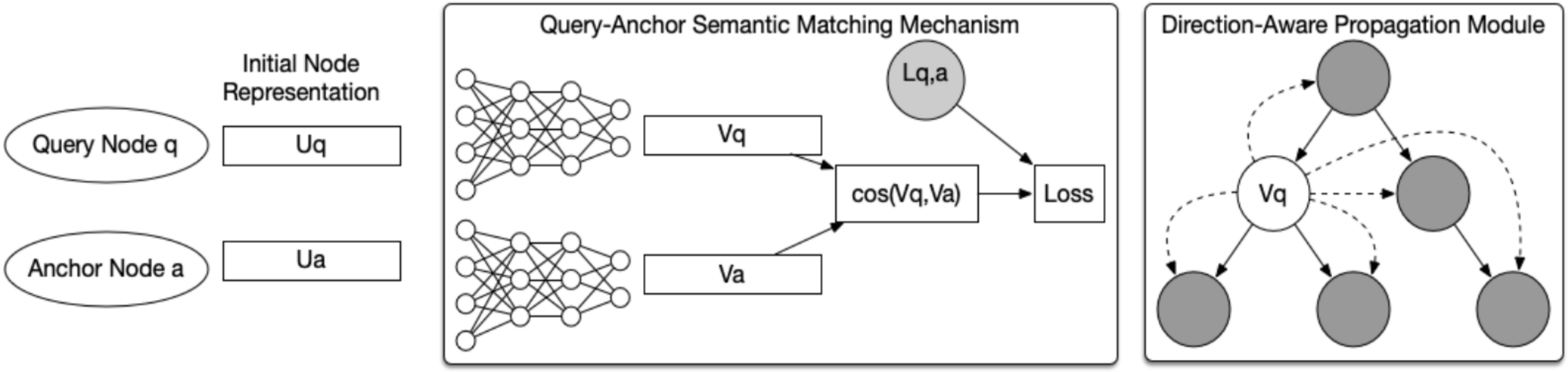
Introduction



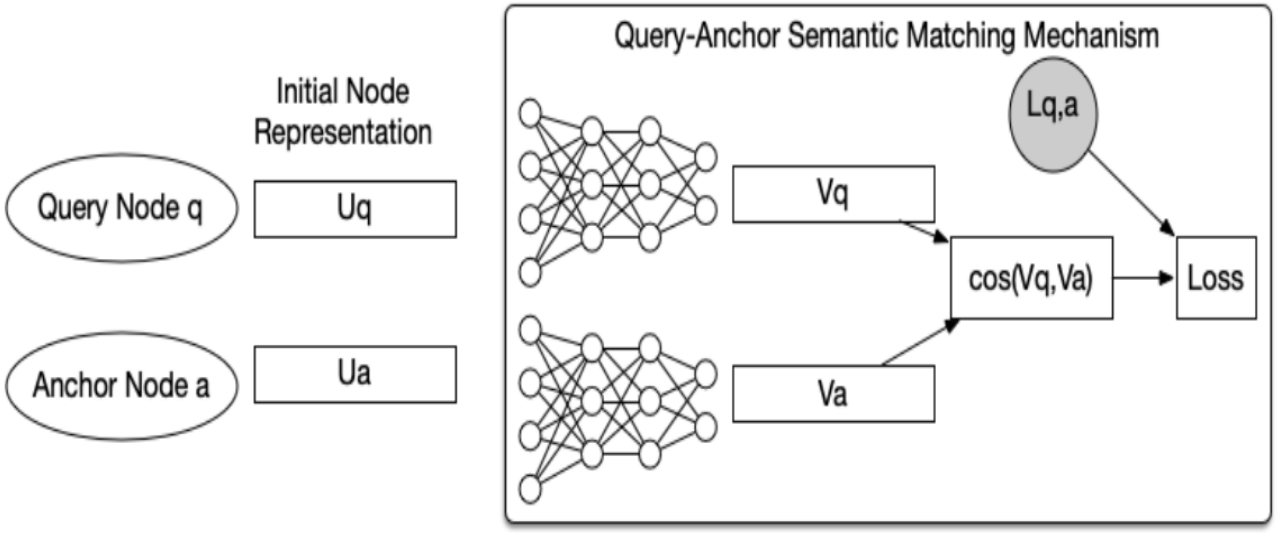
Query	Anchor nodes	Labeling score
Relational database	Database	1
Relational database	Relational operator	1
Relational database	xml	1/2
Relational database	Machine learning	1/3
Relational database	Supervised learning	1/4

Figure 1: An example of a completion task for a computer science taxonomy. The left figures illustrate: (a) the existing taxonomy of computer science; (b) the self-supervision data generated from the existing taxonomy by extracting a query node "relational database" and its anchor set composed of nodes from its close neighborhood such as "database" and from its distant neighborhood such as "supervised learning" and labeling each pair with the inverse of the graph distance separating them; the right figure illustrates: (c) new concepts that are added to (d) the updated taxonomy.

Overview


Initial Node Representation
Query Node q
Uq
Anchor Node a
Ua
Query-Anchor Semantic Matching Mechanism
Lq,a
Vq
cos(Vq,Va)
Loss
Va
Direction-Aware Propagation Module
Vq

Method



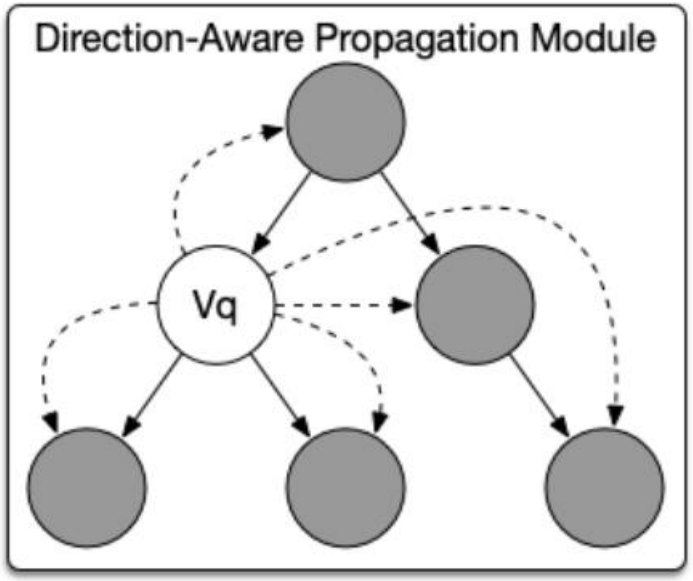
$$l_{q,a} = \frac{1}{f(d_{q,a})}, \tag{2}$$

where $d_{q,a}$ is the graph distance separating a query node q from an anchor node a in the close or the distant neighborhood and $f(.)$ is a linear function.

$$v_q = Pooling(E(d_q, \theta)), v_a = Pooling(E(d_a, \theta)) \tag{3}$$

$$\mathbf{L}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(cos(v_q, v_a) - l_{q,a})^2 \tag{4}$$

Method



$$\hat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-\frac{1}{2}} \tag{5}$$

$$\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_{n,n}$$

$$\widetilde{\mathbf{D}} = \sum_{j \in \mathcal{N}} \widetilde{\mathbf{A}}_{i,j}$$

$$w_q = \alpha(\mathbf{I}_{|\mathcal{N}|} - (1-\alpha)\hat{\mathbf{A}})^{-1} v_q \tag{6}$$

$$w_q^0 = v_q$$

$$w_q^k = (1-\alpha)\hat{\mathbf{A}} w_q^{(k-1)} + \alpha v_q \tag{7}$$

$$cos(w_q, w_a) - distance(q, a) \tag{1}$$

Experiments

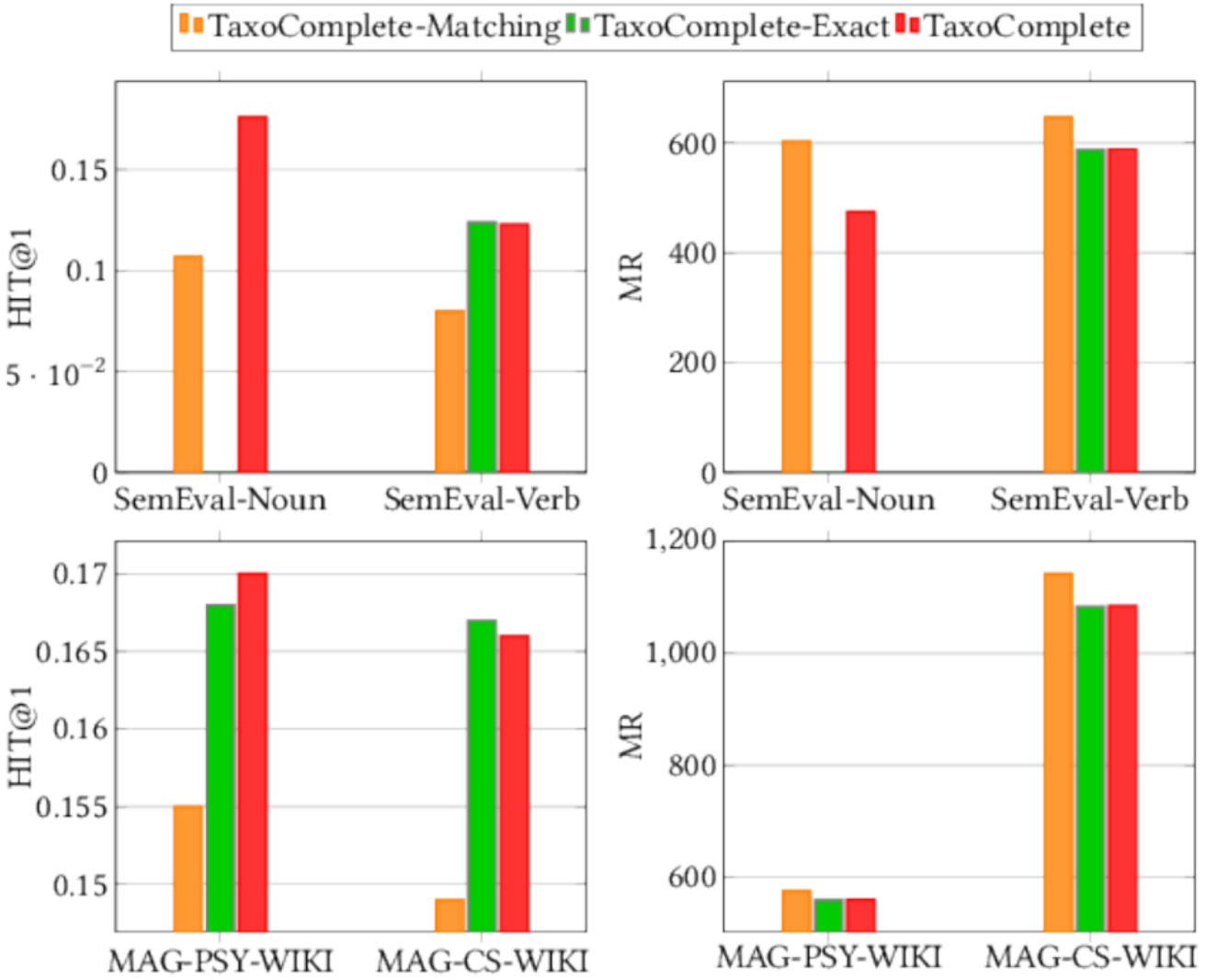
Table 1: Description of the taxonomy Datasets.

Dataset	#nodes	#edges	Depth
SemEval-Noun	75,359	76,810	20
SemEval-Verb	13,715	13,407	13
MAG-WIKI-CS	25,170	40,314	6
MAG-WIKI-PSY	10,671	14,080	6

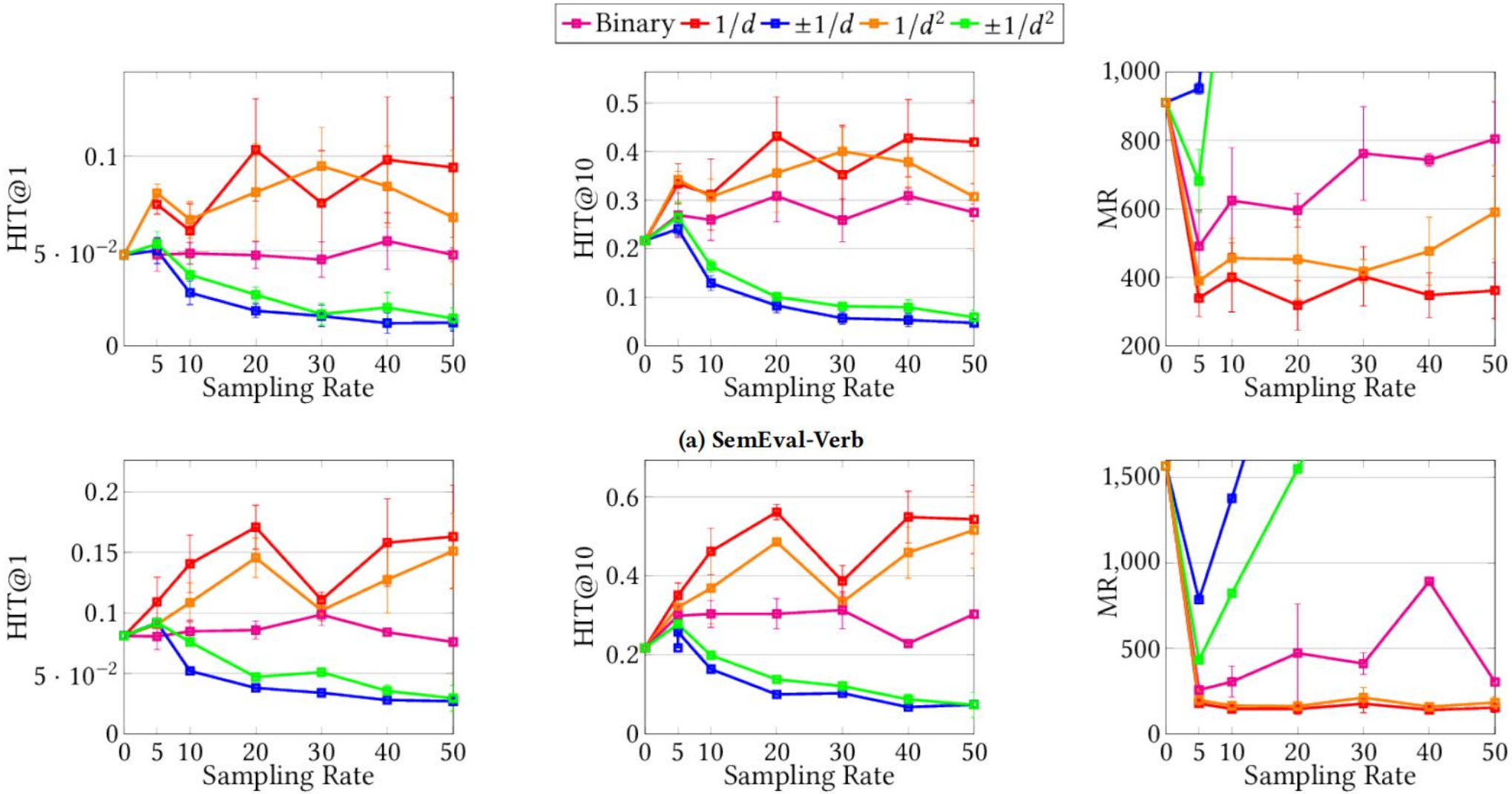
Experiments

Method	SemEval-Noun				SemEval-Verb			
	MR	HIT@1	HIT@5	HIT@10	MR	HIT@1	HIT@5	HIT@10
TaxoExpan	1236.4 ± 465*	0.069 ± 0.005	0.172 ± 0.023	0.248 ± 0.035	876.1 ± 123*	0.072 ± 0.010	0.186 ± 0.021	0.251 ± 0.021*
TMN	2237.4 ± 1087	0.036 ± 0.006	0.112 ± 0.009	0.174 ± 0.016	1931.9 ± 525	0.063 ± 0.007	0.160 ± 0.020	0.224 ± 0.026
Arborist	3993.1 ± 1295	0.020 ± 0.003	0.076 ± 0.009	0.122 ± 0.015	1878.8 ± 329	0.032 ± 0.005	0.100 ± 0.013	0.159 ± 0.018
TaxoEnrich	1703.5 ± 319	0.094 ± 0.015*	0.229 ± 0.033*	0.312 ± 0.038*	2762.0 ± 679	0.087 ± 0.027*	0.188 ± 0.046*	0.240 ± 0.063
TaxoComplete	**474.4 ± 57**	**0.176 ± 0.008**	**0.427 ± 0.009**	**0.541 ± 0.008**	**589.3 ± 132**	**0.123 ± 0.010**	**0.316 ± 0.016**	**0.421 ± 0.028**
Method	MAG-PSY-WIKI				MAG-CS-WIKI			
	MR	HIT@1	HIT@5	HIT@10	MR	HIT@1	HIT@5	HIT@10
TaxoExpan	2688.0 ± 1434	0.070 ± 0.021	0.187 ± 0.045	0.252 ± 0.062	7320.1 ± 3065	0.007 ± 0.003	0.026 ± 0.006	0.047 ± 0.012
TMN	3225.7 ± 1918	0.097 ± 0.022*	0.189 ± 0.043	0.226 ± 0.05	5271.9 ± 4154	0.040 ± 0.009	0.110 ± 0.022	0.150 ± 0.032
Arborist	3698.0 ± 2083	0.046 ± 0.023	0.134 ± 0.032	0.176 ± 0.04	5925.7 ± 4843	0.020 ± 0.007	0.062 ± 0.019	0.095 ± 0.029
TaxoEnrich	2664.9 ± 1473*	0.094 ± 0.023	0.215 ± 0.054*	0.272 ± 0.069*	4954.9 ± 3117*	0.049 ± 0.013*	0.131 ± 0.037*	0.183 ± 0.052*
TaxoComplete	**560.6 ± 23**	**0.170 ± 0.020**	**0.392 ± 0.025**	**0.488 ± 0.019**	**1085.9 ± 115**	**0.166 ± 0.019**	**0.346 ± 0.016**	**0.440 ± 0.018**

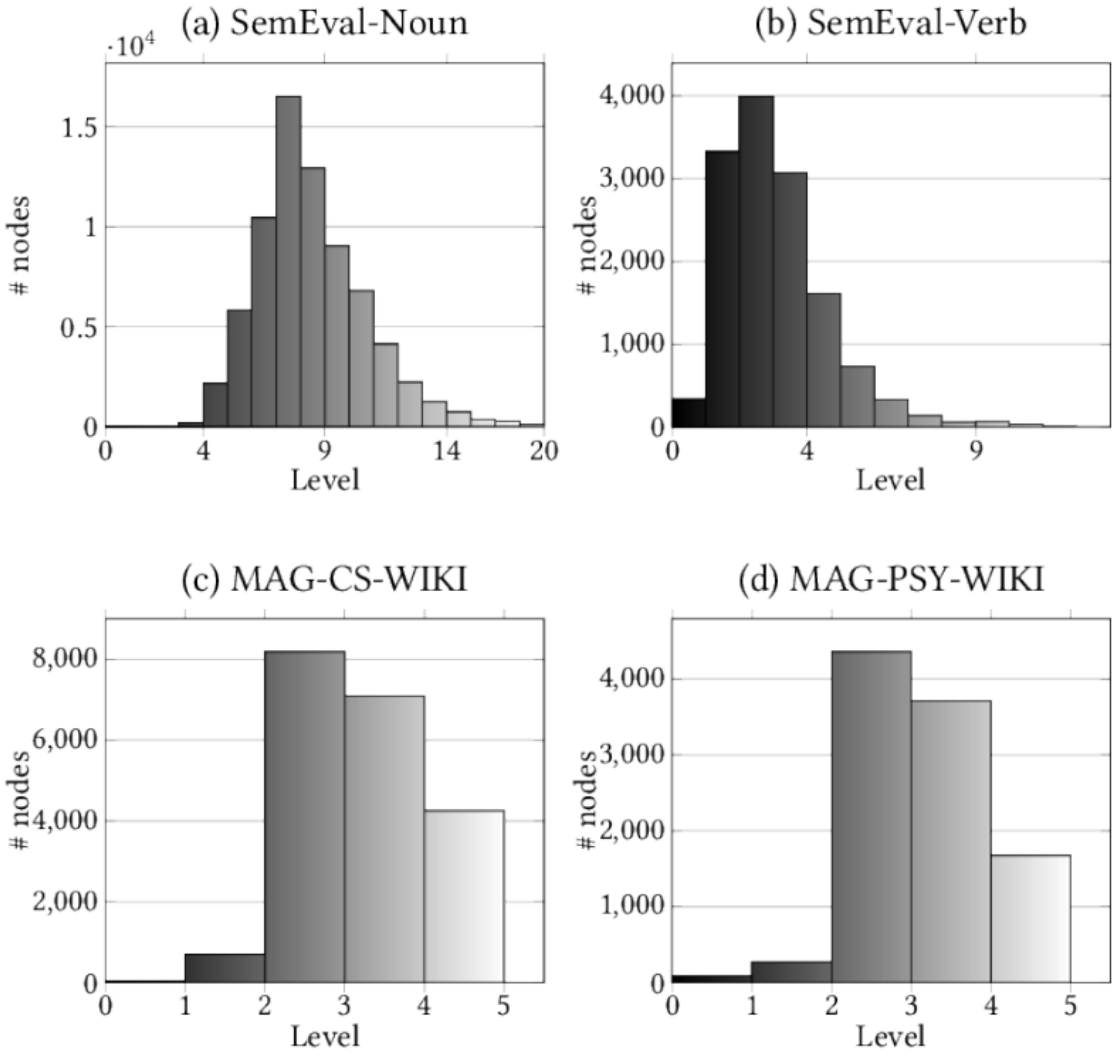
Experiments



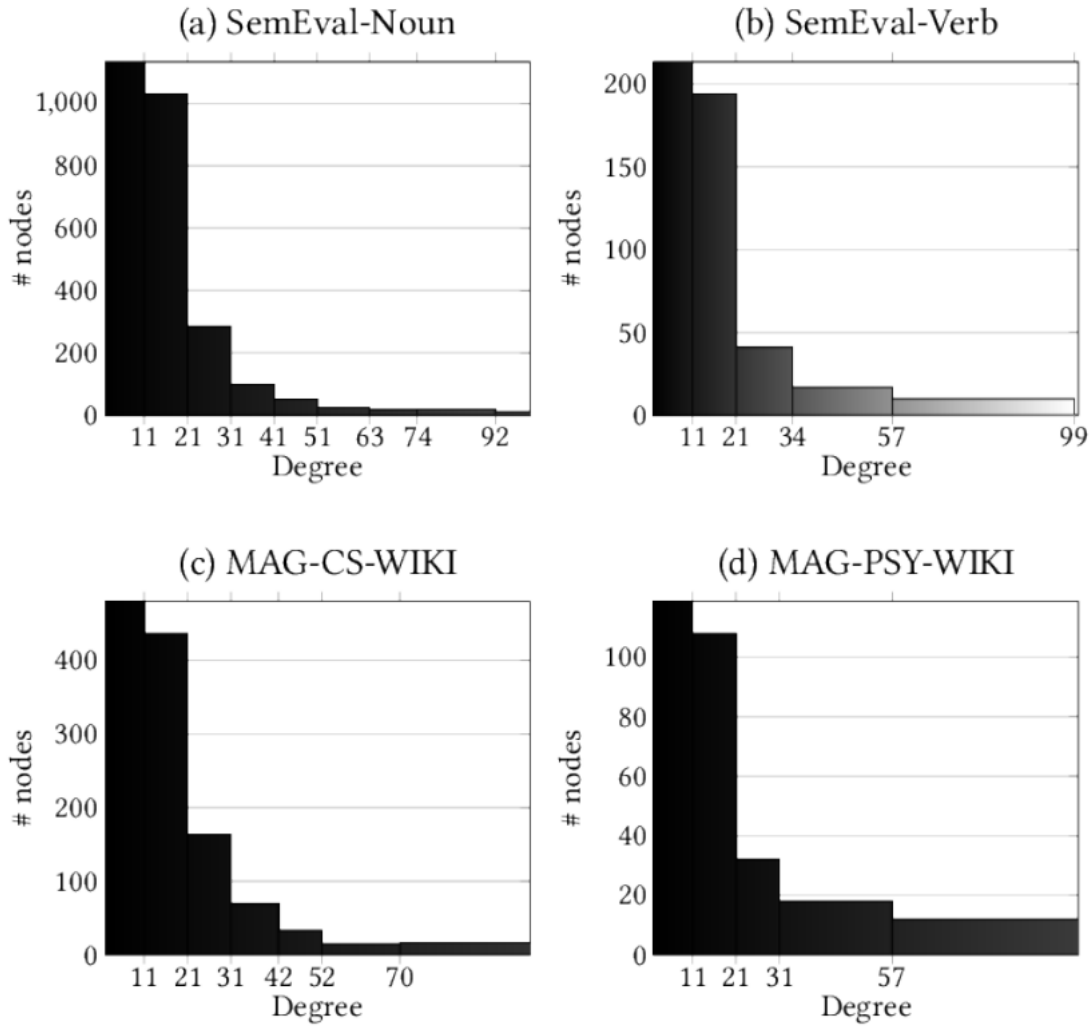
Experiments



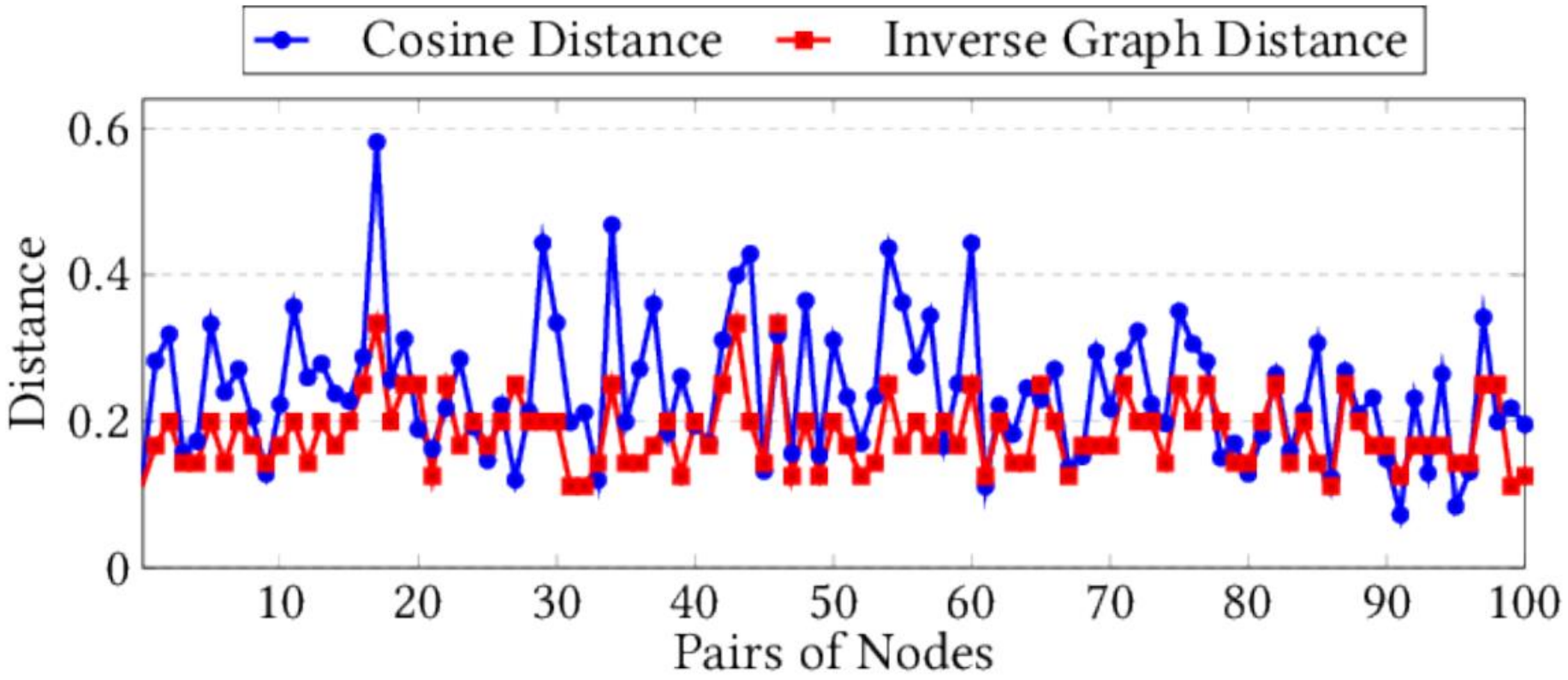
Experiments



Experiments



Experiments



Thanks!