Table 2: Experimental results of author node classification on the AMiner dataset.

Macro-F1	10%	30%	50%	70%	90%
DeepWalk	0.8534	0.8584	0.8587	0.8585	0.8588
Node2vec	0.8665	0.8714	0.8733	0.8738	0.8748
LINE	0.8886	0.8921	0.8929	0.8936	0.8934
PTE	0.8940	0.8982	0.8990	0.8999	0.9005
Metapath2vec	0.9262	0.9303	0.9314	0.9316	0.9320
PMSW-metapath	0.9390	0.9462	0.9472	0.9473	0.9485
PMSW-metaschema	**0.9429**	**0.9575**	**0.9598**	**0.9597**	**0.9604**
Micro-F1	10%	30%	50%	70%	90%
DeepWalk	0.8684	0.8727	0.8733	0.8733	0.8739
Node2vec	0.8731	0.8773	0.8787	0.8792	0.8799
LINE	0.8969	0.9002	0.9010	0.9016	0.9017
PTE	0.9023	0.9061	0.9068	0.9077	0.9082
Metapath2vec	0.9319	0.9356	0.9365	0.9365	0.9369
PMSW-metapath	0.9440	0.9503	0.9513	0.9516	0.9519
PMSW-metaschema	**0.9479**	**0.9617**	**0.9637**	**0.9640**	**0.9638**

Table 3: Experimental results of author node classification on the DBIS dataset.

Macro-F1	10%	30%	50%	70%	90%
DeepWalk	0.3195	0.3823	0.3898	0.3992	0.3945
Node2vec	0.3245	0.3873	0.3938	0.4042	0.3995
LINE	0.3096	0.3765	0.3828	0.3934	0.3885
PTE	0.3080	0.3748	0.3810	0.3922	0.3872
Metapath2vec	0.4306	0.5116	0.5342	0.5476	0.5448
PMSW-metapath	0.5067	0.5561	0.5751	0.5775	0.5839
PMSW-metaschema	**0.5294**	**0.5798**	**0.5909**	**0.5919**	**0.5956**
Micro-F1	10%	30%	50%	70%	90%
DeepWalk	0.3488	0.3953	0.4011	0.4069	0.4067
Node2vec	0.3538	0.4003	0.4051	0.4119	0.4107
LINE	0.3430	0.3905	0.3951	0.3969	0.4009
PTE	0.3418	0.3894	0.3937	0.3950	0.3989
Metapath2vec	0.4866	0.5308	0.5459	0.5545	0.5535
PMSW-metapath	0.5331	0.5619	0.5808	0.5838	0.5910
PMSW-metaschema	**0.5538**	**0.5865**	**0.6043**	**0.6063**	**0.6096**

(since there are too few venues to be classified in the DBIS dataset, we only test author classification on this dataset).

By varying the train-test split ratio, we can observe that the proposed *PMSW-metapath* and *PMSW-metaschema* consistently outperform all baselines in terms of both metrics. For the author node classification, *PMSW-metapath* and *PMSW-metaschema* improve the classification performance in different train-test ratios by 6%-23% over DeepWalk, LINE, and Node2vec, and by 1%-23% over PTE and Metapath2vec, respectively. For the venue node classification, the improvement is consistent, on average by 1%-28% over baseline models. Especially, we can see that *PMSW-metapath* and *PMSW-metaschema* are quite robust and can achieve big gains when given a small size of training data. In addition, we find that *PMSW-metaschema* can achieve on average 1% improvement compared

Table 4: Experimental results of venue node classification on the AMiner dataset.

Macro-F1	10%	30%	50%	70%	90%
DeepWalk	0.3796	0.6695	0.7812	0.8674	0.8357
Node2vec	0.4486	0.7767	0.8394	0.8935	0.9177
LINE	0.4629	0.8473	0.9203	0.9466	0.9466
PTE	0.3388	0.8304	0.9210	0.9505	0.9489
Metapath2vec	0.5247	0.8971	0.9532	0.9701	0.9670
PMSW-metapath	0.5844	0.9403	0.9772	0.9786	0.9711
PMSW-metaschema	**0.6056**	**0.9591**	**0.9774**	**0.9800**	**0.9789**
Micro-F1	10%	30%	50%	70%	90%
DeepWalk	0.4042	0.7166	0.7990	0.8877	0.9186
Node2vec	0.4981	0.7957	0.8586	0.9145	0.9451
LINE	0.5167	0.8457	0.9209	0.9500	0.9571
PTE	0.4267	0.8372	0.9239	0.9550	0.9571
Metapath2vec	0.5975	0.9011	0.9522	0.9725	0.9857
PMSW-metapath	0.6309	0.9675	0.9704	0.9724	0.9772
PMSW-metaschema	**0.6721**	**0.9723**	**0.9806**	**0.9838**	**0.9890**

Table 5: P-value for significance test for author node classification (50% training data) and clustering on AMiner dataset.

Classification	Metapath2vec	PMSW-metapath	PMSW-metaschema
Metapath2vec	–	7.010e-12	2.588e-05
PMSW-metapath	–	–	1.047e-02
PMSW-metaschema	–	–	–
Clustering	**Metapath2vec**	**PMSW-metapath**	**PMSW-metaschema**
Metapath2vec	–	7.184e-03	1.641e-04
PMSW-metapath	–	–	3.768e-03
PMSW-metaschema	–	–	–

with *PMSW-metapath*, which indicates that an ensemble of multiple meta-paths is more effective than a single meta-path, which may capture an ensemble of semantic and structural correlations. Due to the lack of space, we sample one datasets (50% training for author node classification) and present the p-values in t-test results in Table 5. The results prove the effectiveness of the proposed two frameworks.

4.3.2 Node Clustering. We also use the eight-category author and venue nodes in AMiner dataset to evaluate the performance of the representations learned by embedding methods on the node clustering task. We use the k-means algorithm to cluster the data and evaluate the clustering results with metric normalized mutual information (NMI). Table 6 shows the experimental results.

Overall, the results of node clustering are consistent with the results of node classification. And we can reach a similar conclusion as analyzed in node classification experiments. We can see that *PMSW-metapath* and *PMSW-metaschema* outperform all the comparative methods. When clustering for authors, *PMSW-metapath* and *PMSW-metaschema* achieve 12%-28% improvements over DeepWalk, LINE, and Node2vec, and 2%-13% gains over PTE and Metapath2vec.