How Do Higher-Order Interactions Persist in Real-World Hypergraphs? - Supplementary Document

A OBSERVATIONS

A.1 Global Analysis: Persistence vs. Frequency

The distributions of the persistence of HOIs in all 13 datasets are shown in Fig. 1. While the distributions from most datasets clearly obey power-laws, there exist anomalies that deviate from the fitted lines in the distributions from the Eu and Classes datasets. The anomalies from the Eu dataset indicate the surprising abundance of highly persistent HOIs.

A.2 Local Analysis (1): Group Features vs. Group Persistence

Observations. The mutual information (MI) and Pearson correlation coefficients (CC) between each structural group feature and the persistence in each dataset are shown in Table 1. Most features are positively correlated with persistence, and on average, the CC is strongest for #, (i.e., the number of hyperedges containing each HOI S), followed by $\mathcal{H}$ (i.e., the entropy in the sizes of hyperedges containing each HOI S), and then $\Sigma/\cap$. Notably, $\Sigma/\#$ (i.e., the average size of the hyperedges containing each HOI S) is the only feature that is negatively correlated with persistence. We show in Fig. 2 the distributions of # and $\Sigma/\#$ of HOIs with each level of persistence in all 13 datasets.

> Observation 1 (Group Features vs. Group Persistence). *In real-world hypergraphs, the persistence of each HOI S is positively correlated with (a) the number of hyperedges containing S and (b) the entropy in the sizes of hyperedges containing S.*

A.3 Local Analysis (2): Node Features vs. Group Persistence

Observations. The mutual information (MI) and Pearson correlation coefficients (CC) between each structural node feature, which is averaged over the nodes involved in each HOI, and the persistence in each dataset are shown in Table 1. On average, the MI is largest for $\bar{w}$ (i.e., the average weighted degree of neighbors), $\bar{d}$ (i.e., the average degree of neighbors), and r (i.e., PageRank). Notably, $\bar{w}$ and $\bar{d}$ are negatively correlated with persistence. In addition to r, w (i.e., weighted degree), and o (i.e., the number of occurrences) are positively correlated with persistence. The distributions of averaged w and $\bar{w}$ of HOIs with each level of persistence in all 13 datasets are shown in Fig. 3.

> Observation 2 (Node Features vs. Group Persistence). *In real-world hypergraphs, the persistence of each HOI is negatively correlated with the average (weighted) degree of neighbors of each node involved in the HOI.*

A.4 Local Analysis (3): Node Features vs. Node Persistence

Observations. We report in Table 1 the mutual information (MI) and Pearson correlation coefficients (CC) between each structural node feature and the k-node persistence in each dataset. Overall, the MIs are larger than those obtained in the previous subsections. On average, the MI is largest for r (i.e., PageRank), followed by $\bar{w}$ (i.e., the average weighted degree of neighbors), and then $\bar{d}$ (i.e., the average degree of neighbors). The correlation is strongest for o (i.e., the number of occurrences) and w (weighted node degree), which are positively correlated with k-node persistence. Among the features, only $\bar{w}$, $\bar{d}$, and l (i.e., the local clustering coefficient) are negatively correlated with k-node persistence. The distributions of w and $\bar{w}$ of nodes with each level of k-node persistence in all 13 datasets are shown in Fig. 4.

> Observation 3 (Node Features vs. Node Persistence). *In real-world hypergraphs, the weighted degree and number of occurrences of each node are positively correlated with the persistence of HOIs that the node is involved in.*

B LINEAR REGRESSION ANALYSIS

In Table 2, we report the average coefficient, standard error, and p-value of each structural feature obtained by linear regression analysis of each dataset. The results are summarized in Table 3.

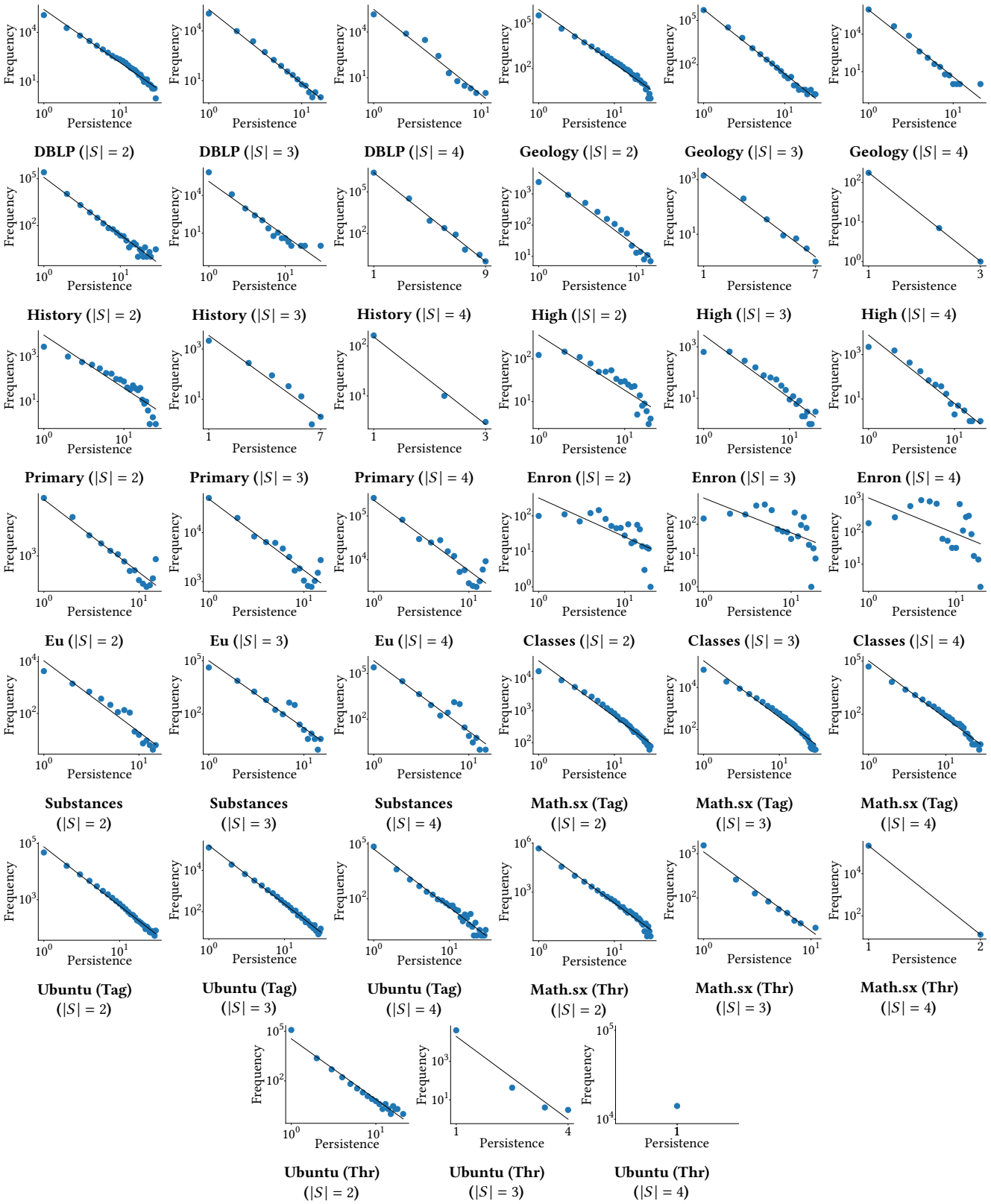


Figure 1: Power-Laws in the Persistence of HOIs. Note that there exist some anomalies in the Eu (Email) and Classes (NDC) datasets.

Table 1: Features vs. Persistence. Mutual information (MI) and correlation coefficients (CC) in all 13 datasets.

DBLP

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	**0.12**	0.03	0.02	0.01	0.04	0.04	0.04	0.11	0.01	0.01	0.02	0.00	**0.07**	0.04	0.06	0.05	0.07	0.08	0.11	0.04	**0.21**	0.10	0.13	0.12
	3	0.07	0.01	0.01	0.00	0.01	0.02	0.01	**0.07**	0.00	0.00	0.00	0.00	**0.02**	0.02	0.02	0.02	0.05	0.06	0.07	0.03	**0.12**	0.07	0.09	0.09
	4	0.04	0.00	0.00	0.01	0.01	0.01	0.01	**0.05**	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.01**	0.02	0.03	0.03	0.02	**0.04**	0.03	0.04	0.04
	Avg.	0.08	0.02	0.01	0.00	0.02	0.02	0.02	**0.08**	0.00	0.00	0.01	0.00	**0.03**	0.02	0.03	0.03	0.05	0.06	0.07	0.03	**0.12**	0.07	0.09	0.09
CC	2	**0.53**	-0.03	-0.04	0.12	0.26	0.34	-0.09	0.38	0.11	0.16	**0.19**	0.02	0.00	0.02	-0.06	-0.16	0.15	0.25	**0.33**	0.07	0.00	0.06	-0.02	-0.09
	3	**0.39**	-0.02	-0.02	0.00	0.20	0.25	-0.06	0.25	0.05	**0.10**	0.10	-0.02	0.00	0.00	-0.05	-0.10	0.06	0.14	**0.15**	0.02	-0.01	0.03	-0.02	-0.09
	4	**0.28**	0.00	0.01	0.03	0.11	0.17	-0.02	0.21	0.02	**0.07**	0.04	0.00	0.00	0.01	-0.03	-0.04	0.03	**0.08**	0.07	0.01	0.00	0.02	-0.01	-0.05
	Avg.	**0.40**	-0.01	-0.02	0.05	0.19	0.26	-0.06	0.28	0.06	**0.11**	0.11	0.00	0.00	0.01	-0.05	-0.10	0.08	0.16	**0.19**	0.03	0.00	0.04	-0.02	-0.07

Geology

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.13	0.03	0.02	0.01	0.03	0.03	0.03	**0.13**	0.01	0.01	0.02	0.01	**0.06**	0.05	0.05	0.05	0.06	0.07	0.11	0.04	**0.16**	0.09	0.11	0.12
	3	0.08	0.01	0.01	0.00	0.01	0.02	0.01	**0.11**	0.00	0.00	0.01	0.00	0.02	0.02	0.02	**0.02**	0.05	0.06	0.08	0.03	0.09	0.07	0.08	**0.11**
	4	0.06	0.00	0.00	0.00	0.01	0.01	0.01	**0.09**	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.01**	0.02	0.03	0.03	0.02	0.03	0.03	0.03	**0.04**
	Avg.	0.09	0.01	0.01	0.00	0.02	0.02	0.02	**0.11**	0.00	0.01	0.01	0.00	**0.03**	0.02	0.03	0.03	0.05	0.05	0.07	0.03	**0.10**	0.06	0.08	0.09
CC	2	**0.50**	-0.09	-0.10	0.17	0.12	0.24	-0.04	0.44	0.19	0.21	**0.24**	0.10	-0.01	0.09	0.03	-0.19	0.21	0.27	**0.34**	0.13	0.00	0.13	0.06	-0.08
	3	**0.37**	-0.05	-0.06	0.04	0.10	0.17	-0.05	0.33	0.11	0.15	**0.15**	0.02	-0.01	0.05	0.00	-0.13	0.12	0.17	**0.19**	0.07	0.00	0.08	0.03	-0.12
	4	0.26	-0.03	-0.04	0.01	0.08	0.13	-0.04	**0.27**	0.08	**0.12**	0.11	0.00	-0.01	0.03	0.00	-0.09	0.06	**0.11**	0.10	0.04	0.00	0.05	0.02	-0.06
	Avg.	**0.38**	-0.06	-0.06	0.07	0.10	0.18	-0.04	0.35	0.13	0.16	**0.17**	0.04	-0.01	0.06	0.01	-0.14	0.13	0.19	**0.21**	0.08	0.00	0.09	0.04	-0.09

History

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	**0.07**	0.01	0.01	0.00	0.01	0.01	0.01	0.07	0.00	0.00	0.01	0.00	0.02	0.01	0.02	**0.02**	0.02	0.02	**0.06**	0.01	0.04	0.03	0.03	0.04
	3	0.05	0.00	0.01	0.00	0.01	0.01	0.00	**0.07**	0.00	0.00	0.00	0.00	**0.01**	0.01	0.01	0.01	0.01	0.02	0.03	0.01	0.02	0.02	0.02	**0.03**
	4	0.03	0.00	0.00	0.00	0.00	0.01	0.00	**0.06**	0.00	0.00	0.00	0.00	**0.01**	0.01	0.01	0.01	0.01	0.02	**0.02**	0.01	0.01	0.01	0.01	0.01
	Avg.	0.05	0.01	0.01	0.00	0.01	0.01	0.01	**0.07**	0.00	0.00	0.01	0.00	**0.01**	0.01	0.01	0.01	0.02	0.02	**0.04**	0.01	0.03	0.02	0.02	0.03
CC	2	0.12	-0.05	-0.05	0.02	0.07	0.07	-0.02	**0.22**	0.08	**0.13**	0.10	-0.01	0.01	0.02	0.00	-0.08	-0.01	0.05	**0.08**	-0.02	-0.01	0.01	0.00	-0.06
	3	0.14	-0.01	-0.02	0.04	0.04	0.07	0.01	**0.21**	0.08	**0.12**	0.07	0.02	-0.01	0.04	0.02	-0.05	0.02	0.08	**0.08**	0.01	0.00	0.03	0.02	-0.01
	4	0.10	0.02	0.01	0.03	0.02	0.04	0.01	**0.17**	0.05	**0.06**	0.03	0.02	-0.02	0.03	0.01	-0.01	0.01	**0.07**	0.06	0.01	0.00	0.03	0.03	0.00
	Avg.	0.12	-0.01	-0.02	0.03	0.04	0.06	0.00	**0.20**	0.07	**0.10**	0.07	0.01	0.00	0.03	0.01	-0.05	0.01	0.06	**0.07**	0.00	0.00	0.02	0.02	-0.02

High

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.02	0.03	**0.03**	0.01	0.02	0.03	0.03	0.03	0.01	0.02	0.02	0.01	**0.03**	0.03	0.03	0.03	0.43	0.61	0.60	0.30	**0.65**	0.63	**0.65**	0.61
	3	0.01	0.01	0.01	0.00	0.01	0.01	0.01	**0.01**	0.00	0.01	0.01	0.00	0.01	0.01	0.01	**0.01**	0.14	0.16	0.15	0.11	0.16	0.16	0.16	**0.16**
	Avg.	0.01	0.02	0.02	0.00	0.01	0.02	0.02	**0.02**	0.01	0.02	0.01	0.01	**0.02**	0.02	0.02	0.02	0.28	0.38	0.38	0.20	**0.41**	0.40	**0.41**	0.38
CC	2	0.08	0.06	0.06	0.05	0.08	**0.08**	-0.01	0.05	0.07	0.09	**0.10**	0.07	0.00	0.07	-0.01	0.05	0.03	-0.01	-0.01	0.03	0.02	0.06	**0.07**	0.02
	3	**0.06**	0.02	0.02	0.01	0.05	0.05	-0.02	0.02	-0.03	0.03	0.03	-0.01	**-0.04**	-0.01	-0.04	0.04	0.02	**0.24**	0.24	0.04	-0.03	0.08	-0.09	0.11
	Avg.	**0.07**	0.04	0.04	0.03	0.06	0.06	-0.02	0.04	0.02	0.06	**0.07**	0.03	-0.02	0.03	-0.02	0.05	0.02	**0.11**	0.11	0.04	0.00	0.07	-0.01	0.06

Primary

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.09	0.15	**0.18**	0.04	0.10	0.11	0.10	0.12	0.03	0.07	0.06	0.04	**0.20**	0.20	0.20	0.20	0.76	0.93	0.93	0.58	**0.99**	0.98	0.99	0.95
	3	0.03	0.04	**0.04**	0.01	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.01	0.05	0.05	**0.05**	0.05	0.62	0.77	0.76	0.38	**0.83**	0.82	**0.83**	0.82
	Avg.	0.06	0.09	**0.11**	0.02	0.04	0.07	0.04	0.08	0.02	0.05	0.05	0.03	0.13	0.12	**0.13**	0.12	0.69	0.85	0.85	0.48	**0.91**	0.90	0.91	0.88
CC	2	0.34	**0.40**	0.39	0.28	0.34	0.34	0.01	0.21	-0.11	-0.02	-0.01	-0.13	-0.14	-0.17	-0.16	**0.18**	0.31	**0.44**	0.43	0.31	0.17	0.27	-0.11	0.13
	3	**0.14**	0.13	0.13	0.04	0.12	0.12	-0.03	0.04	0.03	0.06	**0.06**	0.05	0.00	0.03	-0.02	0.00	0.05	0.08	0.07	-0.01	0.06	0.00	**-0.13**	0.01
	Avg.	0.24	**0.26**	0.26	0.16	0.23	0.23	-0.01	0.12	-0.04	0.02	0.02	-0.04	-0.07	-0.07	-0.09	**0.09**	0.18	**0.26**	0.25	0.15	0.11	0.13	-0.12	0.07

Enron

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.15	0.31	**0.38**	0.09	0.23	0.28	0.26	0.25	0.10	0.29	0.23	0.08	0.40	0.39	**0.40**	0.38	0.65	0.81	0.75	0.58	0.92	0.88	**0.92**	0.74
	3	0.10	0.14	0.18	0.10	0.16	**0.20**	0.16	0.15	0.06	0.12	0.09	0.09	0.20	0.20	**0.20**	0.20	0.64	0.79	0.75	0.55	0.86	0.84	**0.86**	0.79
	4	0.08	0.07	0.09	0.11	0.13	**0.15**	0.13	0.11	0.04	0.06	0.04	**0.11**	0.10	0.10	0.10	0.10	0.55	0.70	0.69	0.46	0.73	0.72	**0.74**	0.70
	Avg.	0.11	0.17	**0.22**	0.10	0.17	0.21	0.18	0.17	0.07	0.16	0.12	0.09	0.23	0.23	**0.24**	0.23	0.61	0.76	0.73	0.53	0.84	0.81	**0.84**	0.74
CC	2	0.36	0.32	0.24	-0.07	0.29	0.30	-0.33	**0.39**	-0.20	-0.06	0.00	-0.28	0.21	-0.33	**-0.34**	0.01	-0.12	0.24	**0.34**	-0.12	0.19	-0.25	-0.26	0.01
	3	0.34	0.24	0.20	-0.14	0.37	0.32	-0.35	**0.38**	-0.19	0.00	0.04	-0.28	0.24	-0.27	**-0.31**	0.07	-0.21	0.33	**0.37**	-0.18	0.19	-0.20	-0.28	0.12
	4	0.21	0.26	0.23	-0.20	0.28	0.21	-0.33	**0.34**	-0.23	-0.05	-0.05	-0.31	**0.34**	-0.31	-0.30	0.16	-0.22	0.11	0.13	-0.21	0.16	**-0.23**	-0.19	0.12
	Avg.	0.30	0.27	0.22	-0.14	0.31	0.28	-0.34	**0.37**	-0.21	-0.04	0.00	-0.29	0.27	-0.30	**-0.32**	0.08	-0.18	0.22	**0.28**	-0.17	0.18	-0.23	-0.24	0.08

Eu

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.18	0.19	**0.29**	0.05	0.15	0.21	0.18	0.23	0.02	0.12	0.07	0.02	0.31	0.31	**0.31**	0.31	0.68	0.83	0.78	0.53	0.89	0.88	**0.89**	0.85
	3	0.20	0.11	0.20	0.07	0.18	**0.24**	0.17	0.23	0.01	0.05	0.02	0.05	0.23	0.23	**0.23**	0.23	0.67	0.82	0.78	0.51	0.86	0.86	**0.86**	0.85
	4	0.22	0.08	0.15	0.08	0.22	**0.27**	0.19	0.24	0.01	0.02	0.01	0.07	0.18	0.18	**0.18**	0.18	0.61	0.74	0.70	0.44	**0.77**	0.77	**0.77**	0.76
	Avg.	0.20	0.13	0.21	0.07	0.18	**0.24**	0.18	0.23	0.02	0.06	0.03	0.05	0.24	0.24	**0.24**	0.24	0.65	0.80	0.75	0.49	0.84	0.83	**0.84**	0.82
CC	2	0.58	0.43	0.49	0.34	0.18	0.49	0.04	**0.59**	-0.02	0.19	0.07	0.02	-0.01	-0.15	**-0.32**	0.09	0.32	**0.52**	0.33	0.45	0.30	-0.10	-0.38	0.18
	3	**0.66**	0.49	0.58	0.15	0.49	0.64	-0.02	0.46	-0.16	0.16	-0.06	-0.14	-0.15	-0.20	**-0.32**	0.14	0.11	**0.40**	0.11	0.24	0.09	-0.15	-0.31	0.05
	4	**0.68**	0.55	0.61	0.07	0.58	0.66	-0.07	0.41	-0.18	0.16	-0.09	-0.19	-0.18	-0.21	**-0.32**	0.14	0.05	**0.36**	0.06	0.16	0.02	-0.11	-0.27	0.09
	Avg.	**0.64**	0.49	0.56	0.19	0.42	0.59	-0.02	0.49	-0.12	0.17	-0.02	-0.11	-0.11	-0.18	**-0.32**	0.12	0.16	**0.43**	0.16	0.28	0.14	-0.12	-0.32	0.11

Classes

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.31	0.36	0.41	0.21	0.38	**0.43**	0.34	0.31	0.25	0.41	0.35	0.24	0.45	0.41	**0.46**	0.32	0.31	0.51	0.46	0.26	**0.62**	0.47	0.55	0.26
	3	0.32	0.35	0.38	0.27	0.40	**0.42**	0.38	0.34	0.26	0.38	0.32	0.31	0.37	0.36	**0.41**	0.28	0.38	0.59	0.52	0.30	**0.70**	0.60	0.66	0.39
	4	0.32	0.31	0.33	0.33	0.40	**0.43**	0.41	0.33	0.23	0.31	0.25	**0.36**	0.30	0.30	0.33	0.24	0.43	0.62	0.56	0.36	**0.69**	0.63	0.66	0.47
	Avg.	0.32	0.34	0.37	0.27	0.39	**0.43**	0.37	0.32	0.25	0.37	0.31	0.30	0.37	0.36	**0.40**	0.28	0.37	0.57	0.51	0.31	**0.67**	0.56	0.62	0.37
CC	2	0.08	0.15	0.19	-0.08	**0.23**	0.12	-0.15	0.12	-0.19	-0.17	-0.15	-0.10	-0.18	**-0.27**	-0.20	-0.03	0.00	0.05	**0.26**	-0.04	-0.13	-0.16	-0.13	-0.06
	3	0.07	0.18	0.26	-0.24	0.13	0.09	**-0.31**	0.06	-0.36	-0.22	-0.17	-0.27	-0.09	**-0.44**	-0.29	0.10	-0.04	0.03	0.21	-0.11	-0.16	-0.23	-0.15	**-0.29**
	4	0.17	0.28	0.36	-0.40	0.19	0.19	**-0.49**	0.13	-0.47	-0.18	-0.15	-0.45	-0.01	**-0.58**	-0.37	0.17	0.06	0.05	0.21	-0.07	-0.14	-0.19	-0.14	**-0.41**
	Avg.	0.10	0.20	0.27	-0.24	0.18	0.13	**-0.32**	0.10	-0.34	-0.19	-0.15	-0.28	-0.09	**-0.43**	-0.29	0.08	0.01	0.05	0.23	-0.07	-0.14	-0.20	-0.14	**-0.25**

Substances

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.08	0.13	**0.20**	0.09	0.15	0.19	0.15	0.13	0.08	0.14	0.09	0.10	**0.26**	0.25	0.25	0.24	0.35	0.42	0.36	0.31	**0.53**	0.46	0.51	0.40
	3	0.08	0.08	0.11	0.07	0.10	**0.13**	0.09	0.10	0.06	0.08	0.05	0.09	**0.13**	0.13	0.13	0.12	0.30	0.37	0.31	0.28	**0.45**	0.41	0.44	0.34
	4	0.08	0.06	0.07	0.05	0.09	**0.10**	0.07	0.08	0.06	0.05	0.04	**0.08**	0.08	0.08	0.08	0.07	0.24	0.28	0.27	0.23	**0.34**	0.31	0.33	0.23
	Avg.	0.08	0.09	0.13	0.07	0.11	**0.14**	0.10	0.10	0.06	0.09	0.06	0.09	**0.16**	0.15	0.15	0.15	0.30	0.36	0.31	0.27	**0.44**	0.40	0.43	0.32
CC	2	0.08	0.05	0.03	-0.16	0.26	0.18	**-0.32**	0.01	-0.12	-0.17	-0.11	**-0.26**	-0.03	-0.24	-0.19	-0.25	-0.03	-0.01	**0.25**	-0.13	-0.05	-0.07	-0.07	-0.15
	3	0.17	**0.24**	0.23	-0.05	0.21	0.20	-0.22	0.11	-0.21	-0.15	-0.14	-0.24	-0.05	**-0.28**	-0.19	0.07	-0.11	-0.01	**0.22**	-0.16	-0.09	-0.17	-0.10	-0.10
	4	0.21	0.29	**0.30**	0.01	0.21	0.22	-0.15	0.16	-0.26	-0.17	-0.17	-0.25	-0.06	**-0.30**	-0.18	0.23	-0.13	0.02	**0.28**	-0.15	-0.11	-0.18	-0.09	-0.01
	Avg.	0.15	0.19	0.19	-0.06	0.23	0.20	**-0.23**	0.10	-0.20	-0.16	-0.14	-0.25	-0.05	**-0.27**	-0.19	0.02	-0.09	0.00	**0.25**	-0.15	-0.09	-0.14	-0.09	-0.09

Math.sx (Tags)

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	**0.15**	0.08	0.12	0.05	0.08	0.09	0.07	0.14	0.01	0.06	0.04	0.01	0.20	0.20	**0.20**	0.20	0.50	0.56	0.48	0.46	**0.93**	0.90	0.93	0.61
	3	**0.12**	0.04	0.07	0.03	0.05	0.06	0.05	0.11	0.01	0.03	0.02	0.00	0.11	0.11	**0.11**	0.11	0.51	0.58	0.51	0.47	**0.93**	0.90	0.92	0.64
	4	**0.10**	0.03	0.05	0.03	0.04	0.06	0.04	0.08	0.01	0.04	0.03	0.00	0.06	0.06	**0.06**	0.06	0.51	0.58	0.53	0.45	**0.81**	0.79	0.81	0.65
	Avg.	**0.12**	0.05	0.08	0.04	0.06	0.07	0.05	0.11	0.01	0.04	0.03	0.00	0.12	0.12	**0.12**	0.12	0.51	0.58	0.50	0.46	**0.89**	0.86	0.89	0.63
CC	2	**0.57**	0.06	0.06	0.42	0.36	0.51	-0.11	0.52	0.08	0.07	0.08	-0.06	**0.34**	-0.26	-0.23	-0.18	0.32	0.33	0.31	**0.34**	0.20	0.13	0.03	-0.05
	3	**0.51**	0.03	0.03	0.21	0.26	0.34	-0.07	0.40	0.06	0.08	0.09	-0.02	0.16	-0.15	**-0.17**	-0.10	0.14	**0.19**	0.17	0.18	0.05	0.10	0.01	0.00
	4	**0.36**	0.02	0.02	0.12	0.22	0.25	-0.04	0.29	0.06	0.10	0.11	0.01	0.10	-0.10	**-0.14**	-0.08	0.12	**0.20**	0.18	0.13	0.07	0.02	-0.04	-0.02
	Avg.	**0.48**	0.04	0.03	0.25	0.28	0.37	-0.08	0.40	0.07	0.09	0.09	-0.02	**0.20**	-0.17	-0.18	-0.12	0.19	**0.24**	0.22	0.21	0.11	0.08	0.00	-0.03

Ubuntu (Tags)

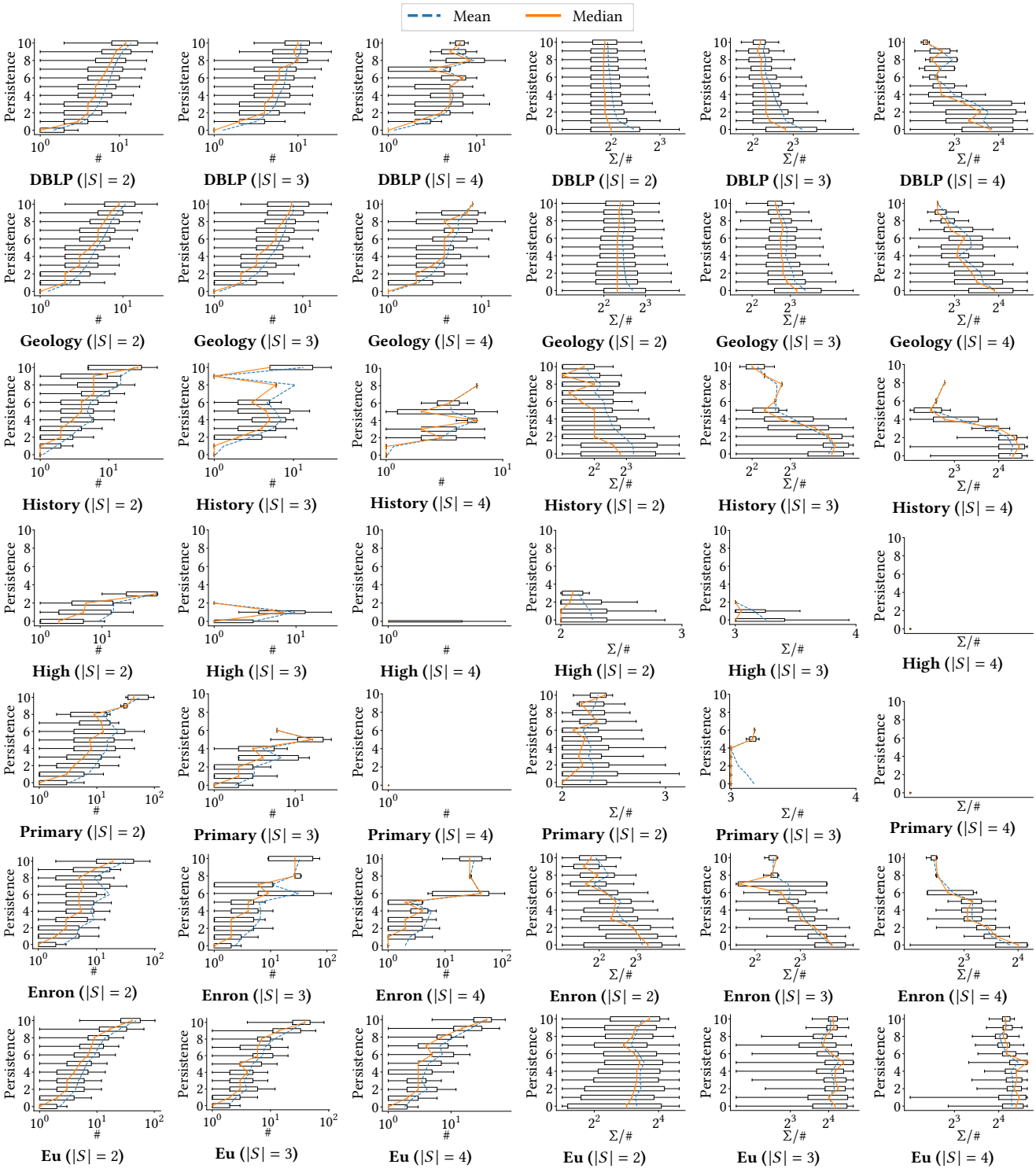
		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	**0.17**	0.07	0.09	0.08	0.09	0.10	0.08	0.17	0.03	0.05	0.05	0.01	**0.15**	0.15	0.15	0.15	0.43	0.47	0.41	0.38	**0.83**	0.80	0.82	0.59
	3	**0.16**	0.03	0.05	0.06	0.06	0.07	0.05	0.14	0.01	0.03	0.02	0.01	**0.07**	0.07	0.07	0.07	0.46	0.50	0.45	0.41	**0.74**	0.72	0.73	0.62
	4	**0.18**	0.02	0.03	0.05	0.05	0.08	0.04	0.11	0.01	0.02	0.02	0.01	0.04	0.04	0.04	**0.04**	0.42	0.45	0.42	0.35	0.51	0.50	0.51	**0.52**
	Avg.	**0.17**	0.04	0.06	0.07	0.07	0.08	0.06	0.14	0.02	0.04	0.03	0.01	**0.09**	0.09	0.09	0.09	0.44	0.47	0.43	0.38	**0.69**	0.67	0.69	0.58
CC	2	0.52	-0.03	-0.03	0.55	0.27	0.52	-0.04	**0.56**	0.35	0.36	**0.36**	0.19	0.32	-0.24	-0.21	-0.26	0.35	0.31	0.31	**0.42**	0.05	-0.05	-0.07	-0.11
	3	**0.57**	-0.02	-0.02	0.44	0.18	0.32	-0.01	0.42	0.21	**0.25**	0.25	0.15	0.19	-0.15	-0.17	-0.16	0.20	0.19	0.19	**0.26**	0.07	-0.02	-0.06	-0.12
	4	**0.58**	0.00	0.00	0.26	0.20	0.28	0.00	0.29	0.11	**0.16**	0.15	0.12	0.09	-0.09	-0.13	-0.11	0.14	0.15	0.14	**0.15**	0.12	0.00	-0.02	-0.08
	Avg.	**0.56**	-0.02	-0.02	0.42	0.22	0.37	-0.02	0.42	0.23	**0.25**	0.25	0.15	0.20	-0.16	-0.17	-0.18	0.23	0.22	0.21	**0.28**	0.08	-0.02	-0.05	-0.10

Math.sx (Threads)

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	**0.11**	0.02	0.02	0.02	0.02	0.03	0.02	0.11	0.01	0.01	0.01	0.01	0.03	0.03	0.03	**0.03**	0.15	0.16	0.19	0.13	0.09	0.10	0.10	**0.21**
	3	**0.09**	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.03	0.03	**0.04**	0.02	0.01	0.01	0.01	0.03
	4	0.05	0.00	0.00	0.00	0.00	0.00	0.00	**0.07**	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00
	Avg.	0.08	0.01	0.01	0.01	0.01	0.01	0.01	**0.09**	0.00	0.00	0.00	0.00	0.01	0.01	0.01	**0.01**	0.06	0.06	0.08	0.05	0.03	0.04	0.04	**0.08**
CC	2	**0.46**	-0.04	-0.04	0.33	0.00	0.08	0.04	0.35	0.23	0.22	0.21	0.19	**0.25**	-0.12	-0.12	-0.12	0.25	0.25	0.25	**0.26**	0.22	-0.01	-0.02	-0.03
	3	**0.24**	-0.01	-0.01	0.02	-0.01	0.00	0.00	0.19	0.08	0.09	0.08	0.04	**0.10**	-0.06	-0.06	-0.04	0.07	0.08	0.07	0.05	**0.09**	-0.01	-0.01	-0.03
	4	**0.18**	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.01	**0.01**	0.01	0.00	0.01	0.00	-0.01	0.00	0.01	**0.01**	0.01	0.01	0.01	0.00	0.00	0.00
	Avg.	**0.29**	-0.02	-0.02	0.11	0.00	0.03	0.01	0.24	0.11	0.11	0.10	0.08	**0.12**	-0.06	-0.06	-0.06	0.11	**0.11**	0.11	0.11	0.10	-0.01	-0.01	-0.02

Ubuntu (Threads)

		Group Features vs. Group Persistence								Node Features vs. Group Persistence								Node Features vs. Node Persistence							
	Size of HOIs	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
MI	2	0.11	0.01	0.01	0.01	0.01	0.02	0.01	**0.11**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.01**	0.06	0.07	**0.09**	0.05	0.02	0.03	0.03	0.08
	3	0.15	0.00	0.00	0.00	0.00	0.00	0.00	**0.17**	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	**0.01**	0.01	0.00	0.00	0.00	0.01
	4	**0.12**	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00	0.00	0.00	0.00	0.00
	Avg.	0.12	0.00	0.00	0.00	0.01	0.01	0.01	**0.13**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.01**	0.03	0.03	**0.03**	0.02	0.01	0.01	0.01	0.03
CC	2	**0.43**	-0.04	-0.04	0.25	-0.01	0.05	0.04	0.29	0.15	**0.16**	0.14	0.15	0.14	-0.02	-0.02	-0.05	0.13	**0.13**	0.13	0.13	0.03	0.02	0.02	0.00
	3	**0.40**	-0.01	-0.01	0.09	-0.02	-0.01	0.03	0.35	0.09	**0.10**	0.09	0.06	0.08	-0.01	-0.01	-0.05	0.11	**0.12**	0.12	0.08	0.09	0.01	0.01	-0.03
	4	0.25	0.00	-0.01	0.02	-0.01	0.00	0.00	**0.25**	0.06	**0.07**	0.06	0.03	0.05	-0.01	-0.01	-0.02	0.18	**0.21**	0.19	0.07	0.13	-0.01	-0.01	-0.03
	Avg.	**0.36**	-0.02	-0.02	0.12	-0.01	0.01	0.02	0.30	0.10	**0.11**	0.10	0.08	0.09	-0.01	-0.02	-0.04	0.14	**0.15**	0.14	0.09	0.08	0.01	0.01	-0.02


Mean
Median
Persistence
#
Σ/#
DBLP (|S| = 2)
DBLP (|S| = 3)
DBLP (|S| = 4)
Geology (|S| = 2)
Geology (|S| = 3)
Geology (|S| = 4)
History (|S| = 2)
History (|S| = 3)
History (|S| = 4)
High (|S| = 2)
High (|S| = 3)
High (|S| = 4)
Primary (|S| = 2)
Primary (|S| = 3)
Primary (|S| = 4)
Enron (|S| = 2)
Enron (|S| = 3)
Enron (|S| = 4)
Eu (|S| = 2)
Eu (|S| = 3)
Eu (|S| = 4)

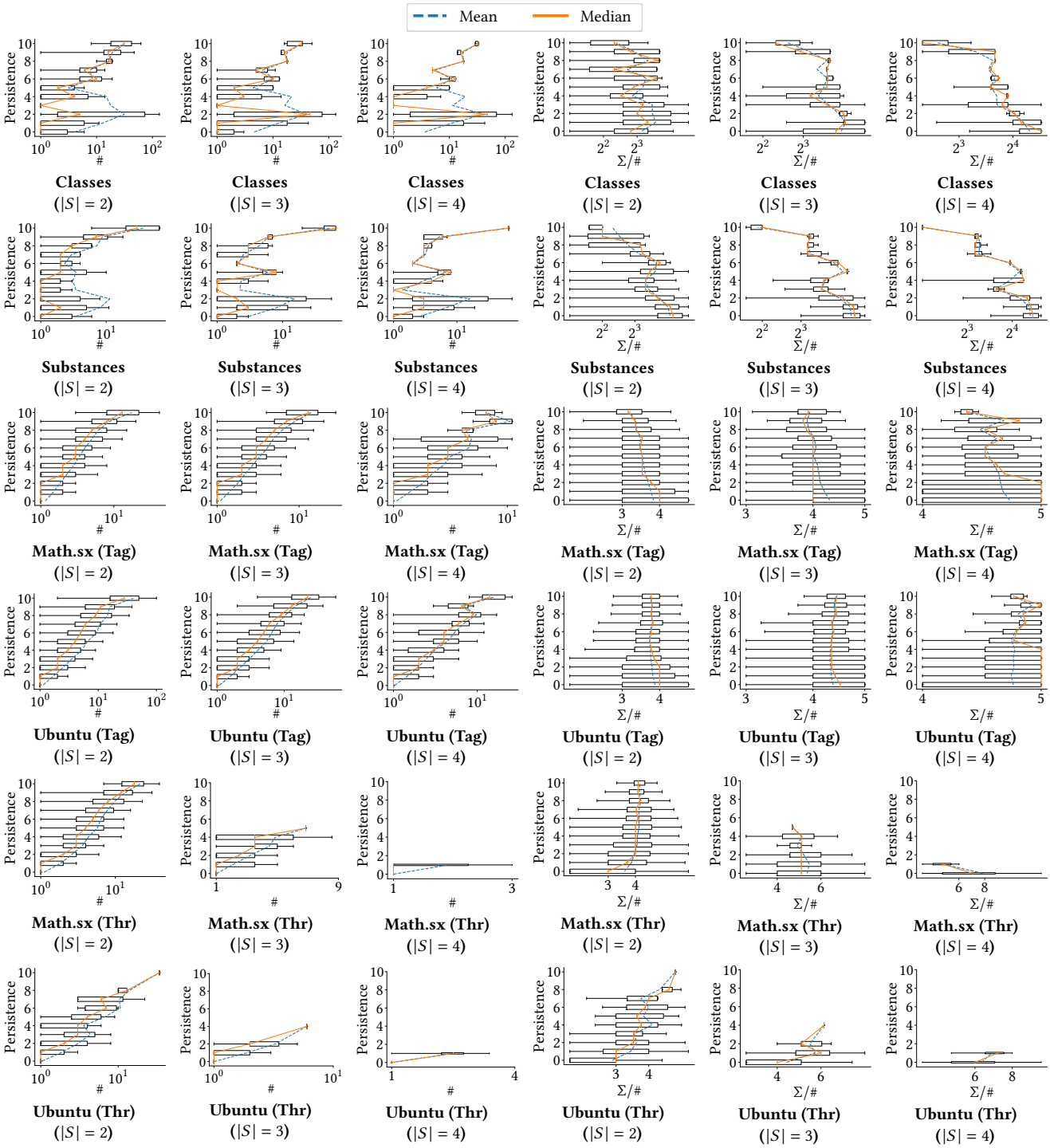
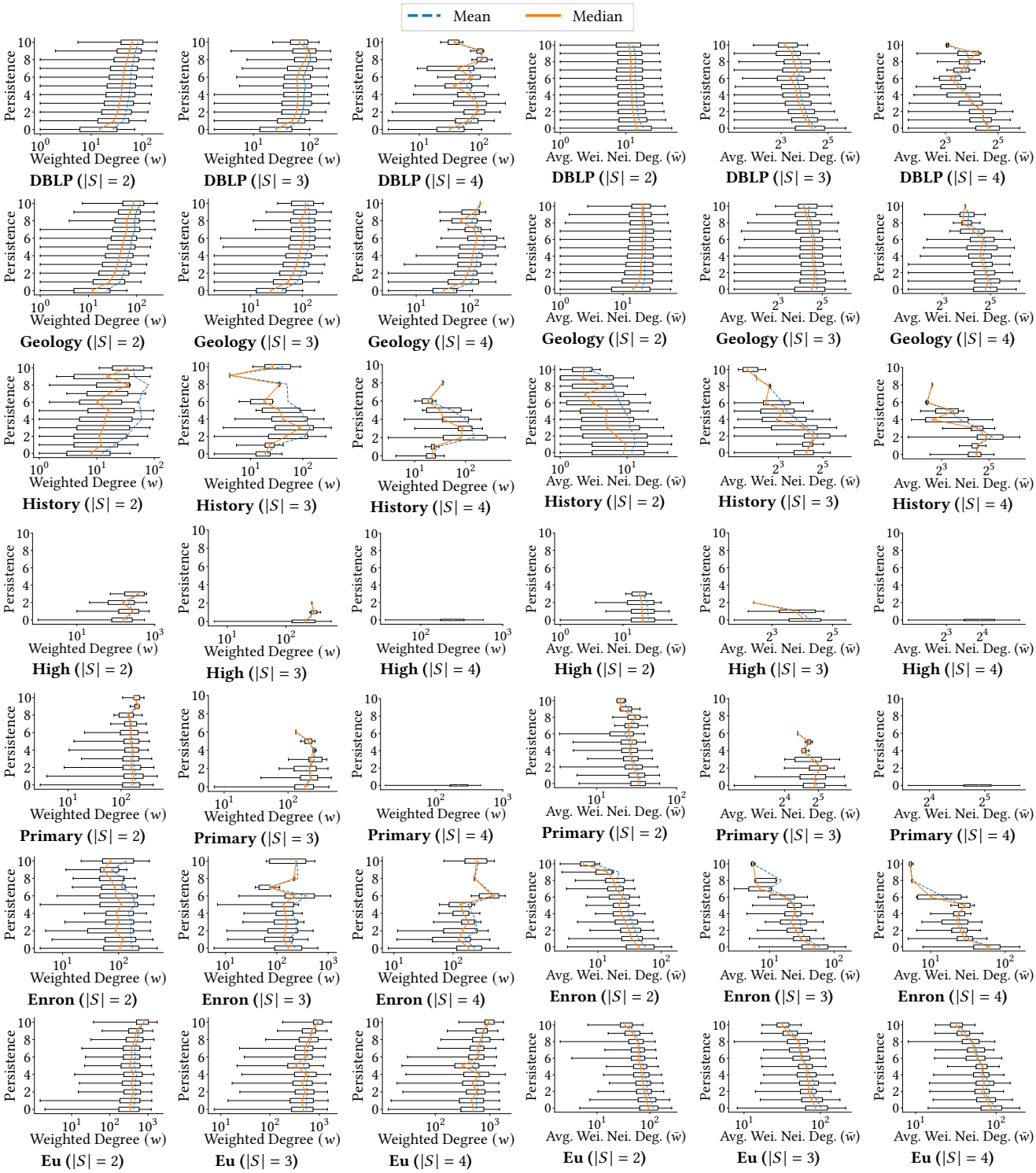


Figure 2: Group Features vs. Group Persistence. The distribution of # (i.e., the number of hyperedges containing each HOI) and Σ/# (i.e., the average size of the hyperedges containing each HOI) of HOIs with each level of persistence in two datasets.


Mean
Median
Persistence
Weighted Degree (w)
Avg. Wei. Nei. Deg. ($\bar{w}$)
DBLP ($|S| = 2$)
DBLP ($|S| = 3$)
DBLP ($|S| = 4$)
DBLP ($|S| = 2$)
DBLP ($|S| = 3$)
DBLP ($|S| = 4$)
Geology ($|S| = 2$)
Geology ($|S| = 3$)
Geology ($|S| = 4$)
Geology ($|S| = 2$)
Geology ($|S| = 3$)
Geology ($|S| = 4$)
History ($|S| = 2$)
History ($|S| = 3$)
History ($|S| = 4$)
History ($|S| = 2$)
History ($|S| = 3$)
History ($|S| = 4$)
High ($|S| = 2$)
High ($|S| = 3$)
High ($|S| = 4$)
High ($|S| = 2$)
High ($|S| = 3$)
High ($|S| = 4$)
Primary ($|S| = 2$)
Primary ($|S| = 3$)
Primary ($|S| = 4$)
Primary ($|S| = 2$)
Primary ($|S| = 3$)
Primary ($|S| = 4$)
Enron ($|S| = 2$)
Enron ($|S| = 3$)
Enron ($|S| = 4$)
Enron ($|S| = 2$)
Enron ($|S| = 3$)
Enron ($|S| = 4$)
Eu ($|S| = 2$)
Eu ($|S| = 3$)
Eu ($|S| = 4$)
Eu ($|S| = 2$)
Eu ($|S| = 3$)
Eu ($|S| = 4$)

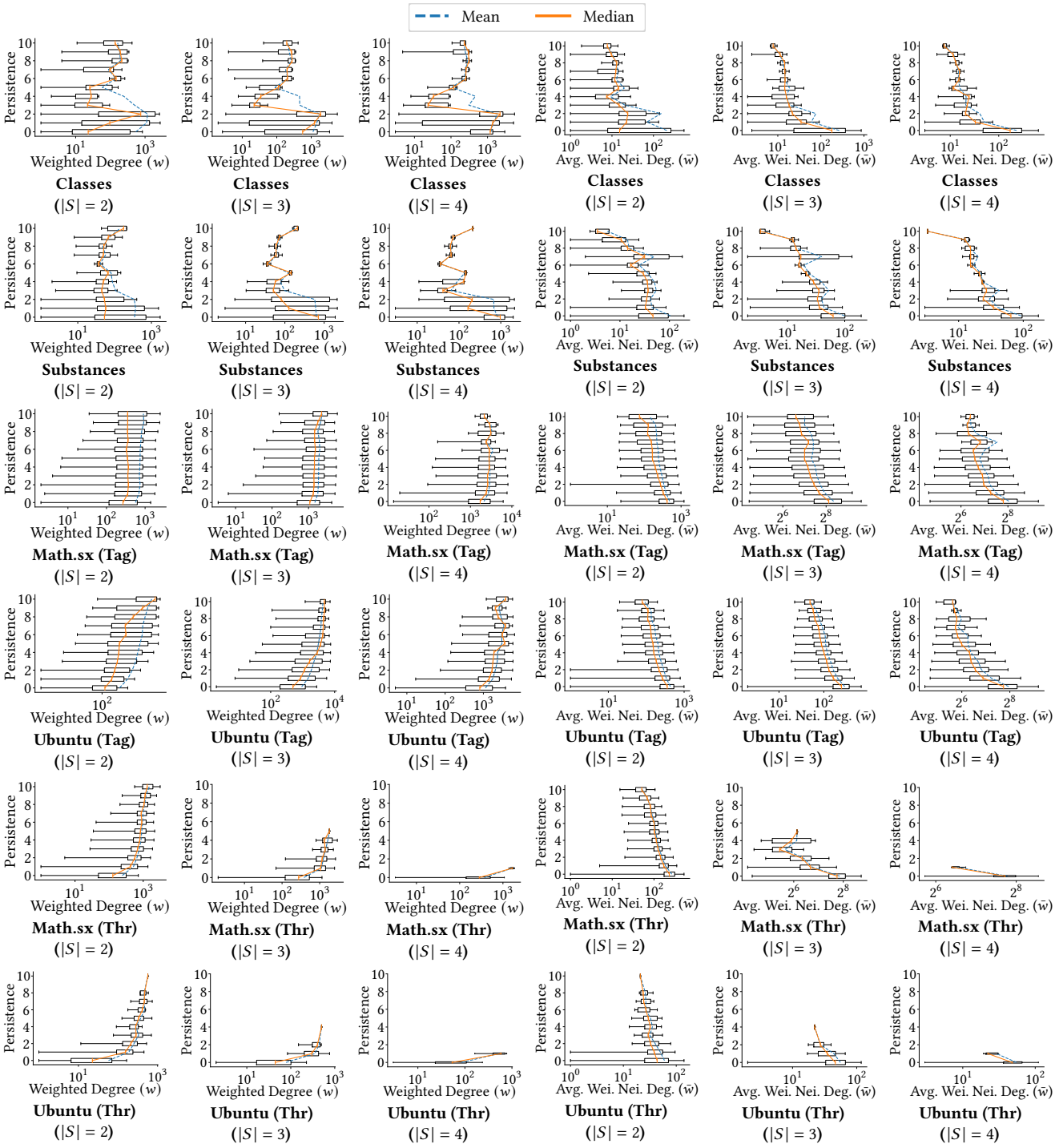
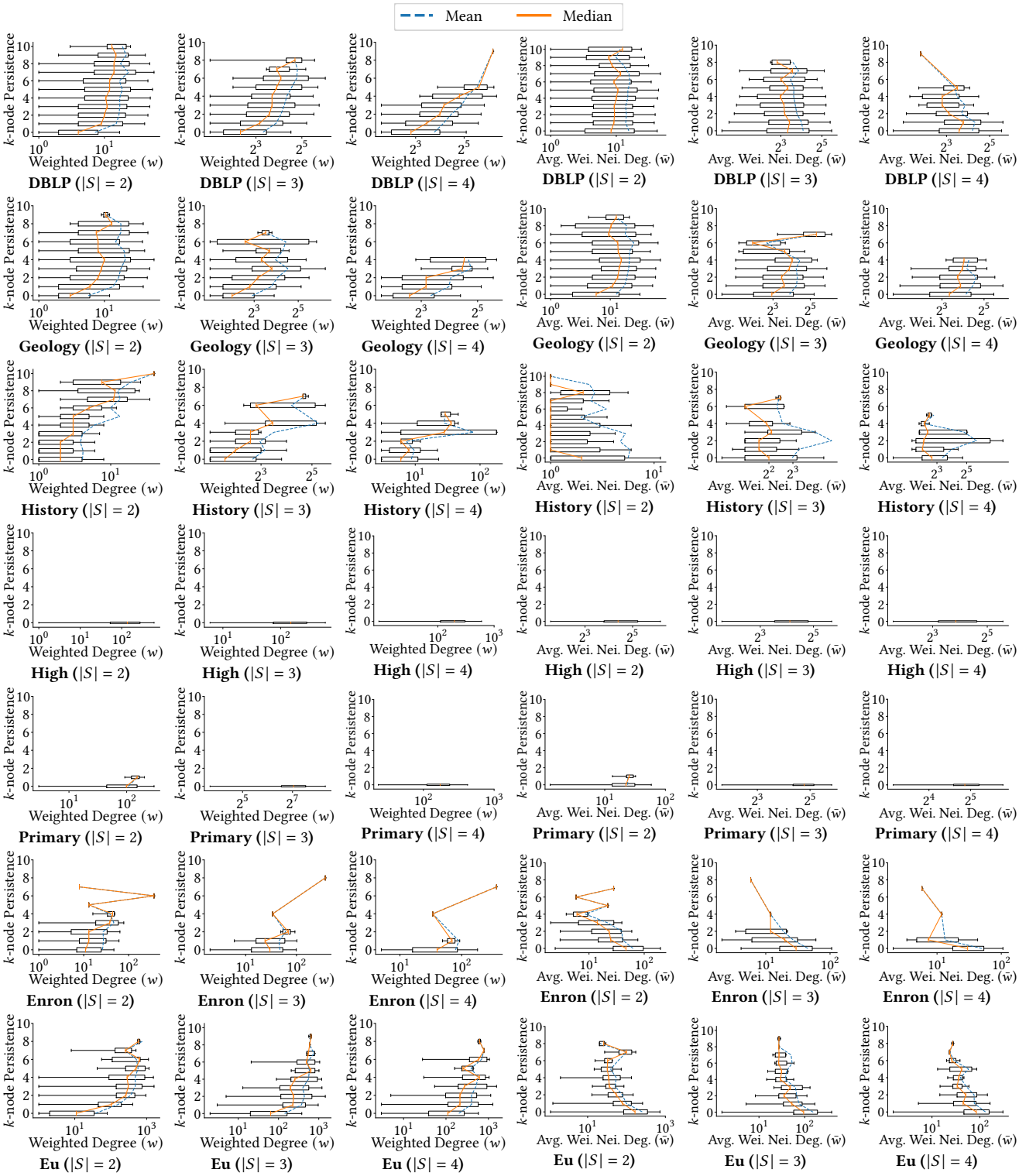


Figure 3: Node Features vs. Group Persistence. The distribution of averaged w (i.e., weighted degree) and $\bar{w}$ (i.e., the average weighted degree of neighbors) of HOIs with each level of persistence in two datasets.


Mean
Median
k-node Persistence
Weighted Degree (w)
Avg. Wei. Nei. Deg. (w̄)
DBLP (|S| = 2)
DBLP (|S| = 3)
DBLP (|S| = 4)
Geology (|S| = 2)
Geology (|S| = 3)
Geology (|S| = 4)
History (|S| = 2)
History (|S| = 3)
History (|S| = 4)
High (|S| = 2)
High (|S| = 3)
High (|S| = 4)
Primary (|S| = 2)
Primary (|S| = 3)
Primary (|S| = 4)
Enron (|S| = 2)
Enron (|S| = 3)
Enron (|S| = 4)
Eu (|S| = 2)
Eu (|S| = 3)
Eu (|S| = 4)

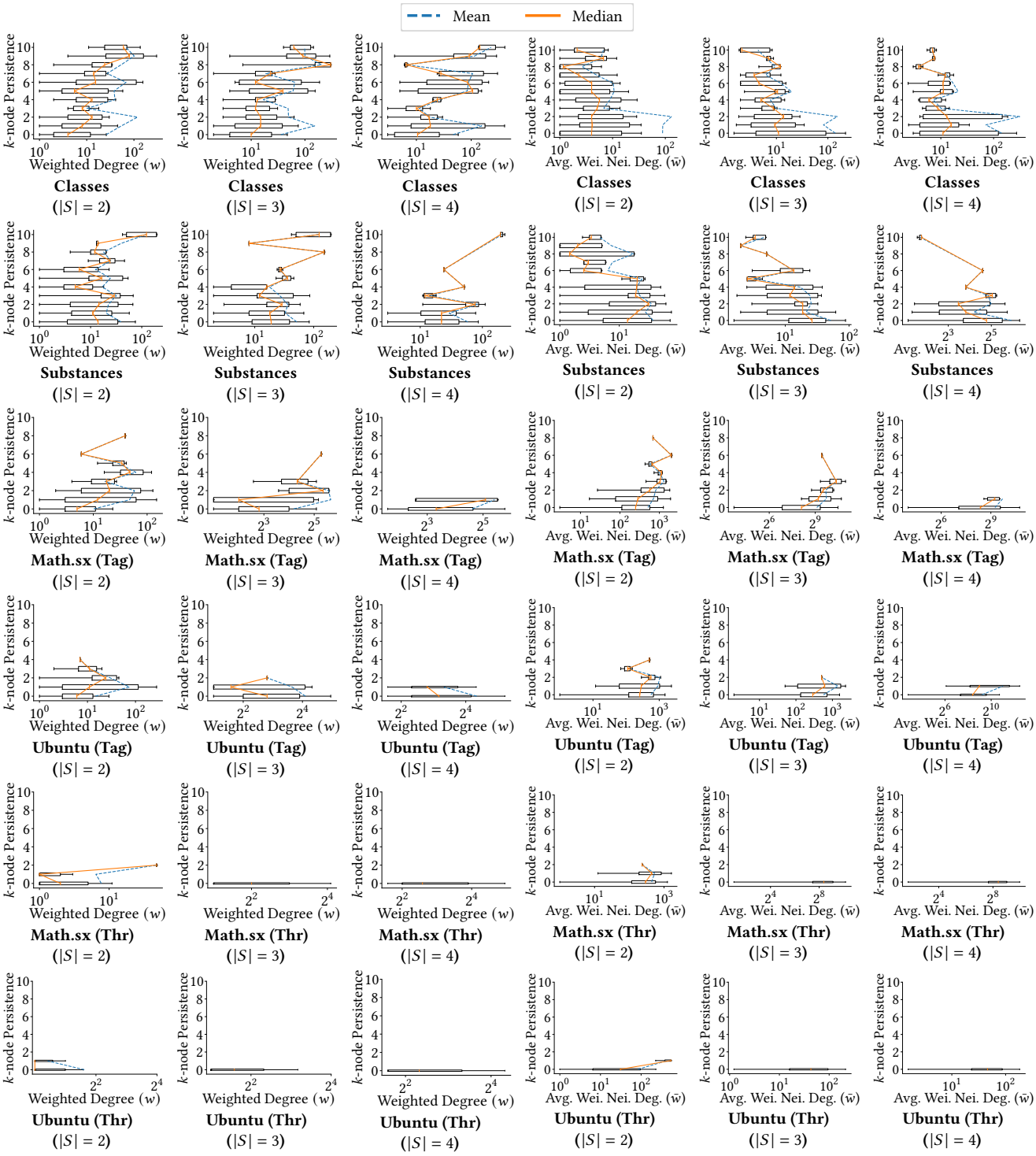


Figure 4: Node Features vs. Node Persistence. The distribution of w (i.e., weighted degree) and $\bar{w}$ (i.e., the average weighted degree of neighbors) of nodes with each level of k-node persistence in two datasets.

Table 2: The average coefficient, standard error, and p-value of each structural feature obtained by linear regression analysis of each dataset.

DBLP

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	**Coef.**	0.25	0.09	-0.13	0.01	0.31	-0.17	-0.01	0.17	0.00	0.00	0.01	0.00	-3.62	-0.01	0.00	-0.06
	Std. Err.	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.39	0.00	0.00	0.00
	p-value	**0**	**1.2e-10**	**3.2e-18**	**4.7e-37**	**0**	**0**	**1.2e-05**	**0**	**0**	**0**	**0**	0.57	0.30	**0**	**4.0e-96**	**2.4e-67**
3	**Coef.**	0.17	0.01	0.01	0.00	0.36	-0.15	0.00	-0.03	0.00	0.00	0.00	0.00	-40.00	0.00	0.00	-0.06
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.26	0.00	0.00	0.00
	p-value	**0**	**0.01**	**1.8e-04**	**3.7e-42**	**0**	**0**	**1.0e-07**	**9.8e-88**	0.23	**9.2e-25**	**7.2e-21**	0.45	**6.3e-04**	**1.9e-64**	**4.3e-32**	**0**
4	**Coef.**	0.09	0.00	0.00	0.01	0.27	-0.08	-0.01	-0.04	0.00	0.00	0.00	0.00	26.99	0.00	0.00	0.02
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.03	0.00	0.00	0.00
	p-value	**0**	0.61	**0.02**	**0**	**0**	**0**	**0**	**0**	**1.6e-29**	**0**	**0**	**7.4e-19**	**6.2e-11**	0.26	**8.3e-05**	**0**

Geology

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.34	0.04	-0.06	0.00	0.46	-0.34	-0.01	0.28	0.01	0.00	0.01	0.01	-1.15	0.00	0.00	-0.01
	Std. Err.	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.71	0.00	0.00	0.00
	p-value	**0**	**3.2e-05**	**1.1e-11**	**2.1e-06**	**0**	**0**	**1.6e-56**	**0**	**0**	**0**	**0**	**4.1e-46**	0.11	0.73	**6.0e-08**	0.12
3	Coeff.	0.28	0.00	0.01	-0.01	0.63	-0.29	0.01	0.08	0.00	0.00	0.00	0.00	-0.10	0.00	0.00	-0.02
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00
	p-value	**0**	0.35	**0.04**	**0**	**0**	**0**	**0**	**0**	**2.6e-88**	**4.1e-08**	**3.0e-11**	**1.3e-14**	0.62	**4.1e-10**	**4.8e-54**	**1.7e-69**
4	Coeff.	0.19	0.00	0.01	-0.01	0.64	-0.20	0.01	0.07	0.00	0.00	0.00	0.00	0.09	0.00	0.00	-0.02
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00
	p-value	**0**	0.07	**2.0e-14**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0.04**	**5.4e-58**	**0**	**0**

History

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.02	0.01	0.03	0.01	0.00	-0.01	-0.01	0.24	0.00	0.00	0.00	0.00	6.07	0.00	0.00	-0.03
	Std. Err.	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00
	p-value	**0**	0.20	**2.4e-08**	**0**	0.22	**1.7e-22**	**9.7e-69**	**0**	**1.2e-59**	**0.03**	**7.3e-15**	**2.0e-13**	**0.01**	**3.7e-57**	**1.3e-57**	**8.2e-27**
3	Coeff.	0.03	0.02	-0.01	0.01	0.05	-0.02	-0.01	0.15	0.00	0.00	0.00	0.00	-3.75	0.00	0.00	0.01
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.00	0.00	0.00
	p-value	**0**	**4.3e-36**	**7.2e-04**	**0**	**7.0e-67**	**5.1e-40**	**0**	**0**	**3.8e-50**	**1.7e-06**	**9.6e-21**	**1.5e-57**	**7.0e-08**	**0**	**0**	**0.00**
4	Coeff.	0.09	0.02	-0.01	0.00	0.31	-0.10	0.00	0.14	0.00	0.00	0.00	0.00	-8.51	0.00	0.00	0.01
	Std. Err.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00
	p-value	**0**	**0**	**3.4e-21**	**1.2e-06**	**0**	**0**	**1.3e-11**	**0**	**0**	**1.8e-04**	**2.5e-49**	**6.5e-15**	**0**	**0**	**0**	0.26

High

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.00	0.13	0.00	0.01	-0.01	0.00	-0.03	0.01	0.00	0.00	0.00	0.00	1.71	0.00	0.00	0.09
	Std. Err.	0.00	0.96	0.96	0.01	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	3.06	0.00	0.00	0.03
	p-value	0.34	0.25	0.30	0.33	0.21	0.22	**0.00**	0.55	0.42	**8.3e-05**	**6.3e-06**	0.54	0.51	0.67	0.37	**0.01**
3	Coeff.	0.00	-1.10	0.67	-0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-1.76	0.00	0.00	0.03
	Std. Err.	0.00	2.07	1.46	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	2.41	0.00	0.00	0.02
	p-value	0.28	0.57	0.61	0.63	0.25	0.26	0.62	0.32	0.69	0.47	0.42	0.66	0.46	0.47	0.75	0.31

Primary

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.00	12.80	-5.61	0.18	0.48	-0.22	-0.04	0.37	0.04	0.00	0.01	-0.04	-111.39	-0.03	0.00	1.90
	Std. Err.	0.01	9.64	9.38	0.03	0.10	0.05	0.05	0.09	0.00	0.00	0.00	0.01	16.15	0.01	0.00	0.16
	p-value	0.49	0.29	0.53	**1.1e-06**	**1.8e-05**	**5.4e-05**	0.40	**1.9e-04**	**4.2e-15**	0.09	**0.01**	**0.00**	**1.5e-11**	**1.2e-06**	0.14	**1.1e-28**
3	Coeff.	0.01	-35.04	30.43	0.07	0.61	-0.20	-0.04	-0.02	0.00	0.00	0.00	0.01	-21.58	0.00	0.00	-0.02
	Std. Err.	0.01	17.24	11.83	0.03	0.12	0.04	0.03	0.05	0.00	0.00	0.00	0.00	6.18	0.00	0.00	0.06
	p-value	0.27	0.32	0.21	0.13	**3.1e-05**	**9.0e-05**	0.15	0.41	0.12	0.34	0.16	0.22	**0.01**	0.40	0.78	0.71

Enron

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.02	5.38	-2.56	0.01	0.08	0.02	-0.04	1.58	0.04	-0.01	0.01	-0.18	40.24	0.01	0.00	2.82
	Std. Err.	0.01	1.47	1.13	0.03	0.07	0.03	0.03	0.20	0.02	0.00	0.00	0.06	6.39	0.03	0.00	0.52
	p-value	**0.04**	**0.00**	**0.05**	0.45	0.19	0.32	0.23	**1.3e-12**	**0.01**	**8.1e-06**	**3.4e-05**	**0.01**	**3.8e-09**	0.76	0.40	**5.0e-06**
3	Coeff.	0.04	3.63	-1.52	-0.02	0.41	-0.05	0.01	0.95	0.00	0.00	0.01	-0.07	21.52	0.02	0.00	0.86
	Std. Err.	0.01	0.99	0.52	0.01	0.05	0.01	0.01	0.07	0.00	0.00	0.00	0.02	1.93	0.01	0.00	0.19
	p-value	**2.0e-05**	**0.00**	**0.02**	0.12	**8.7e-09**	0.11	0.34	**6.8e-40**	0.52	**3.1e-08**	**6.7e-09**	**1.6e-04**	**4.3e-27**	**0.02**	**0.02**	**1.1e-05**
4	Coeff.	0.03	3.43	-1.30	-0.03	0.40	-0.05	0.03	0.75	-0.01	0.00	0.00	-0.03	17.28	0.02	0.00	0.23
	Std. Err.	0.00	0.96	0.35	0.00	0.03	0.01	0.01	0.03	0.00	0.00	0.00	0.01	0.73	0.00	0.00	0.08
	p-value	**4.3e-16**	**0.00**	**3.6e-04**	**2.3e-08**	**1.1e-29**	**1.5e-16**	**2.0e-05**	**0**	**7.2e-05**	**1.2e-07**	**5.3e-07**	**5.1e-08**	**0**	**1.1e-10**	**5.2e-08**	**0.01**

Eu

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.03	-0.89	3.01	0.09	-0.13	0.11	-0.11	1.19	-0.06	0.00	0.00	0.03	1586.84	0.00	0.00	0.63
	Std. Err.	0.00	0.58	0.39	0.00	0.01	0.01	0.01	0.05	0.00	0.00	0.00	0.01	104.74	0.00	0.00	0.15
	p-value	**1.8e-22**	0.24	**6.9e-12**	**9.4e-64**	**5.1e-09**	**1.5e-65**	**3.7e-75**	**0**	**7.6e-46**	0.33	0.84	**1.4e-06**	**9.4e-51**	0.44	**4.7e-19**	**2.7e-05**
3	Coeff.	0.06	-8.02	7.77	0.05	0.00	0.07	-0.06	0.48	-0.03	0.00	0.00	0.02	830.95	0.00	0.00	0.35
	Std. Err.	0.00	0.46	0.24	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00	40.26	0.00	0.00	0.07
	p-value	**3.1e-76**	**1.4e-58**	**0**	**1.0e-57**	0.25	**1.4e-42**	**7.3e-74**	**0**	**4.5e-73**	**0**	**0**	**1.1e-05**	**5.5e-91**	**0.02**	**0**	**1.1e-05**
4	Coeff.	0.15	-13.15	10.26	0.00	-0.19	0.00	-0.02	0.38	-0.02	0.00	0.00	0.01	570.27	0.00	0.00	0.15
	Std. Err.	0.00	0.38	0.16	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	17.81	0.00	0.00	0.04
	p-value	**0**	**0**	**0**	**0.01**	**3.5e-05**	**2.7e-06**	**4.5e-09**	**0**	**0**	**0**	**0**	**4.2e-14**	**0**	**6.5e-10**	**0**	**9.4e-05**

Classes

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.00	-3.39	4.46	-0.25	0.39	-0.04	-0.01	2.79	0.18	0.00	0.02	0.21	-25.32	-0.13	0.00	2.02
	Std. Err.	0.01	0.81	0.86	0.08	0.05	0.02	0.06	0.29	0.02	0.00	0.00	0.07	4.18	0.02	0.00	0.35
	p-value	0.67	**1.5e-04**	**5.9e-07**	**0.00**	**1.7e-09**	0.22	0.63	**1.3e-07**	**3.8e-16**	**5.6e-15**	**8.0e-12**	**0.01**	**1.6e-08**	**1.0e-08**	0.71	**1.2e-07**
3	Coeff.	-0.05	-1.25	-0.51	0.04	0.03	0.07	-0.18	0.92	0.15	0.00	0.01	0.17	-31.37	-0.28	0.00	5.66
	Std. Err.	0.01	0.31	0.33	0.04	0.04	0.01	0.03	0.13	0.01	0.00	0.00	0.04	3.19	0.01	0.00	0.22
	p-value	**1.3e-10**	**9.2e-04**	0.25	0.39	0.49	**1.1e-10**	**2.5e-08**	**4.2e-11**	**4.6e-68**	**4.3e-52**	**3.8e-44**	**2.3e-05**	**5.5e-22**	**0**	0.50	**0**
4	Coeff.	-0.09	1.03	-3.86	0.21	-0.43	0.16	-0.33	-0.12	0.15	0.00	0.01	0.19	-30.78	-0.39	0.00	7.88
	Std. Err.	0.00	0.14	0.13	0.02	0.02	0.01	0.02	0.06	0.00	0.00	0.00	0.02	2.57	0.01	0.00	0.11
	p-value	**0**	**4.3e-12**	**0**	**4.8e-25**	**1.0e-57**	**0**	**0**	0.16	**0**	**0**	**0**	**2.3e-22**	**1.5e-29**	**0**	0.19	**0**

Substances

Size of HOIs		$\#$	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.04	-0.13	0.58	-0.05	0.51	-0.07	0.02	0.83	0.06	0.00	0.02	-0.08	-37.92	0.01	0.00	1.24
	Std. Err.	0.01	0.28	0.30	0.01	0.06	0.02	0.01	0.08	0.00	0.00	0.00	0.01	4.93	0.00	0.00	0.12
	p-value	**2.1e-06**	0.50	0.21	**4.3e-13**	**5.4e-10**	0.07	0.19	**2.6e-18**	**0**	**9.2e-25**	**4.4e-07**	**2.3e-22**	**3.2e-04**	0.30	**0.00**	**5.1e-17**
3	Coeff.	0.05	-0.36	0.25	-0.03	0.72	-0.11	0.01	0.61	0.04	0.00	-0.01	-0.05	-131.63	-0.02	0.00	2.13
	Std. Err.	0.00	0.07	0.07	0.00	0.03	0.01	0.00	0.02	0.00	0.00	0.00	0.00	3.56	0.00	0.00	0.04
	p-value	**6.2e-76**	**1.2e-05**	**0.01**	**0**	**0**	**1.6e-55**	**2.5e-08**	**0**	**0**	0.25	**9.8e-07**	**0**	**0**	**9.6e-93**	**0.00**	**0**
4	Coeff.	0.05	-0.26	-0.01	-0.02	0.92	-0.10	0.01	0.36	0.02	0.00	0.00	-0.03	-126.63	-0.02	0.00	1.91
	Std. Err.	0.00	0.02	0.02	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.00	1.48	0.00	0.00	0.01
	p-value	**0**	**1.1e-29**	0.62	**0**	**0**	**0**	**4.0e-38**	**0**	**0**	**8.2e-42**	**1.4e-32**	**0**	**0**	**0**	**4.9e-18**	**0**

Math.sx (Tags)

Size of HOIs		$\#$	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.08	0.02	-0.21	0.29	-0.67	0.76	-0.37	0.52	0.00	0.00	0.00	0.01	188.79	0.00	0.00	0.19
	Std. Err.	0.01	1.01	0.91	0.01	0.07	0.03	0.01	0.03	0.00	0.00	0.00	0.00	5.32	0.00	0.00	0.07
	p-value	**8.4e-21**	0.73	0.62	**0**	**1.2e-15**	**0**	**0**	**1.1e-59**	**5.6e-33**	0.28	0.41	**3.8e-07**	**0**	**9.6e-06**	0.42	**0.04**
3	Coeff.	0.48	2.25	-3.08	0.11	0.99	-0.35	-0.07	0.41	0.00	0.00	0.00	0.00	47.36	0.00	0.00	-0.06
	Std. Err.	0.01	3.17	2.36	0.01	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00	2.64	0.00	0.00	0.05
	p-value	**0**	0.36	0.30	**1.1e-38**	**3.9e-83**	**5.4e-66**	**4.7e-14**	**0**	**1.2e-43**	**2.7e-06**	**3.7e-15**	**1.1e-06**	**3.9e-65**	**2.5e-05**	**6.3e-76**	0.20
4	Coeff.	0.08	-6.08	0.56	0.42	1.33	-0.25	-0.16	0.52	0.00	0.00	0.00	0.00	15.14	0.00	0.00	-0.13
	Std. Err.	0.01	13.93	8.97	0.01	0.06	0.01	0.01	0.02	0.00	0.00	0.00	0.00	2.98	0.00	0.00	0.05
	p-value	**2.5e-10**	0.56	0.57	**0**	**2.1e-83**	**8.5e-37**	**6.4e-67**	**0**	**4.6e-33**	0.18	**7.0e-06**	**9.4e-05**	**3.6e-06**	**1.2e-07**	**4.5e-58**	**0.03**

Ubuntu (Tags)

Size of HOIs		$\#$	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	-0.07	-2.03	1.68	0.17	-1.61	1.41	-0.25	0.54	0.00	0.00	0.00	0.01	11.16	0.00	0.00	-0.40
	Std. Err.	0.00	0.37	0.35	0.00	0.05	0.03	0.01	0.02	0.00	0.00	0.00	0.00	1.92	0.00	0.00	0.04
	p-value	**0**	**2.5e-07**	**5.5e-06**	**0**	**0**	**0**	**0**	**0**	**0**	**3.3e-38**	**7.8e-10**	**5.4e-41**	**7.3e-07**	**6.4e-72**	**4.5e-13**	**2.1e-21**
3	Coeff.	0.07	-0.34	0.50	0.21	-1.21	0.60	-0.23	0.06	0.00	0.00	0.00	0.00	-17.16	0.00	0.00	-0.16
	Std. Err.	0.00	0.68	0.56	0.00	0.04	0.01	0.00	0.01	0.00	0.00	0.00	0.00	1.87	0.00	0.00	0.03
	p-value	**1.6e-60**	0.62	0.39	**0**	**0**	**0**	**0**	**0.00**	**1.2e-66**	**1.0e-48**	**3.4e-07**	**1.3e-08**	**2.4e-16**	0.16	**5.9e-42**	**9.4e-09**
4	Coeff.	0.38	-0.16	-0.26	0.17	-0.23	0.10	-0.14	-0.06	0.00	0.00	0.00	0.00	-7.39	0.00	0.00	-0.21
	Std. Err.	0.01	2.06	1.53	0.01	0.06	0.02	0.01	0.02	0.00	0.00	0.00	0.00	2.43	0.00	0.00	0.03
	p-value	**0**	0.72	0.63	**4.2e-29**	0.12	**1.1e-04**	**0**	**0.02**	**2.2e-22**	**3.4e-34**	**3.2e-10**	0.18	**0.00**	**1.3e-05**	**1.1e-27**	**3.0e-11**

Math.sx (Threads)

Size of HOIs		$\#$	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.15	-0.29	0.42	0.04	0.06	-0.07	-0.04	-0.01	0.00	0.00	0.00	0.00	49.09	0.00	0.00	-0.06
	Std. Err.	0.00	0.03	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.00
	p-value	**0**	**8.7e-27**	**1.4e-57**	**0**	**3.7e-14**	**3.8e-38**	**0**	**0.01**	**0**	**1.3e-39**	**0**	**0**	**0**	**0**	**3.9e-59**	**1.0e-64**
3	Coeff.	0.30	-0.02	0.03	-0.02	0.67	-0.34	0.02	-0.08	0.00	0.00	0.00	0.00	2.38	0.00	0.00	0.00
	Std. Err.	0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00
	p-value	**0**	**0.01**	**2.3e-06**	**1.5e-07**	**0**	**0**	**2.5e-07**	**4.7e-95**	0.60	**3.7e-38**	**1.1e-54**	0.24	**3.7e-82**	**0.01**	**9.6e-13**	**0.00**
4	Coeff.	0.15	0.00	0.00	-0.04	0.78	-0.26	0.04	0.04	0.00	0.00	0.00	0.00	-0.04	0.00	0.00	0.00
	Std. Err.	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	p-value	0.32	0.76	0.65	0.33	0.32	0.32	0.32	0.32	**0.00**	**4.8e-04**	0.29	0.68	0.20	0.20	0.07	0.71

Ubuntu (Threads)

Size of HOIs		#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	Coeff.	0.09	-0.02	0.04	0.05	0.03	-0.03	-0.06	-0.05	0.00	0.00	0.00	0.00	1.72	0.00	0.00	-0.01
	Std. Err.	0.00	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00
	p-value	**4.4e-09**	0.16	**0.01**	**1.2e-17**	**1.2e-14**	**4.1e-20**	**2.1e-25**	0.19	**0.00**	**2.6e-15**	**1.4e-16**	**2.1e-25**	0.07	**0.04**	0.06	**6.4e-05**
3	Coeff.	0.12	0.00	0.01	0.01	0.18	-0.09	-0.01	0.07	0.00	0.00	0.00	0.00	-0.98	0.00	0.00	0.00
	Std. Err.	0.01	0.01	0.01	0.00	0.03	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00
	p-value	**5.7e-06**	0.68	0.51	**7.1e-17**	**0.03**	**0.03**	**5.5e-16**	**1.7e-04**	0.29	**2.2e-21**	**1.5e-33**	0.07	**0.02**	0.31	0.34	**0.03**
4	Coeff.	0.00	0.00	0.00	0.02	-0.15	0.05	-0.02	0.06	0.00	0.00	0.00	0.00	-0.23	0.00	0.00	0.00
	Std. Err.	0.01	0.00	0.00	0.00	0.04	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00
	p-value	0.33	0.89	0.90	0.31	0.33	0.33	0.32	0.32	0.32	0.06	0.08	0.64	0.30	**0.04**	**0.05**	0.59

Table 3: Statistical significance of structural features. We report the number of datasets where each feature is significant with a given p-value in linear regression analysis.

		Persistence of HOIs																k-Node Persistence of Nodes							
Size of HOIs	p-value	#	$\frac{\#}{\cup}$	$\frac{\Sigma}{\Sigma\cup}$	$\cap$	$\frac{\#}{\cap}$	$\frac{\Sigma}{\cap}$	$\frac{\Sigma}{\#}$	$\mathcal{H}$	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l	d	w	o	c	r	$\bar{d}$	$\bar{w}$	l
2	$\leq$ **0.05**	10	6	9	11	10	9	9	11	12	10	11	11	9	8	7	12	8	7	9	8	7	10	9	9
	$\leq$ **0.01**	9	6	7	11	10	9	9	10	11	9	11	11	9	7	7	11	8	7	9	8	5	10	7	8
	$\leq$ **0.001**	9	5	7	10	10	9	8	10	10	9	10	8	8	7	6	10	7	6	7	8	4	6	5	7
	$\leq$ **0.0001**	9	4	7	10	10	9	8	9	10	9	10	8	7	7	6	10	7	6	6	8	4	4	4	6
3	$\leq$ **0.05**	11	7	7	9	10	11	10	11	7	10	11	8	11	9	9	10	9	9	6	7	8	6	5	10
	$\leq$ **0.01**	11	6	5	9	9	10	10	11	7	10	11	8	10	7	8	9	8	9	5	7	4	5	3	10
	$\leq$ **0.001**	11	4	4	9	9	10	10	10	7	10	11	8	9	6	7	7	7	8	5	7	2	3	2	8
	$\leq$ **0.0001**	11	3	2	9	9	10	10	9	7	10	11	7	8	6	7	7	5	8	5	7	2	3	2	6
4	$\leq$ **0.05**	9	5	6	9	8	9	9	8	10	9	9	8	9	9	9	8	8	10	7	8	6	4	3	8
	$\leq$ **0.01**	9	5	5	8	8	9	9	7	10	9	9	8	8	8	8	6	8	8	7	5	3	2	3	6
	$\leq$ **0.001**	9	4	5	8	8	9	9	7	9	9	9	8	7	8	8	6	7	8	5	4	3	2	2	6
	$\leq$ **0.0001**	9	4	4	8	8	8	9	7	9	7	9	8	7	8	8	6	7	8	5	4	1	2	2	5
Avg.		9.8	4.9	5.7	9.3	9.1	9.3	9.2	9.2	9.1	9.3	10.2	8.4	8.5	7.5	7.5	8.5	7.4	7.8	6.3	6.8	4.1	4.8	3.9	7.4