Table 1: Comparison of structural distributional shifts in the context of OOD robustness and OOD detection. We report the drop in predictive performance of the **ERM** baseline measured by *Accuracy* (left) and the quality of uncertainty estimates of the **SE** baseline measured by *AUROC* (right). The ID to OOD ratio in these experiments is 70% to 30% instead of 50% to 50% as in the paper.

	Popularity	Locality	Density
AmazonComputer	-3.80%	-12.66%	-14.04%
AmazonPhoto	-7.19%	-3.03%	-6.79%
CoauthorCS	-6.51%	-1.89%	-6.64%
CoauthorPhysics	-2.58%	-5.50%	-2.47%
CoraML	-7.56%	-13.63%	-9.49%
CiteSeer	+3.24%	+2.50%	-4.13%
PubMed	+0.46%	-0.67%	-1.76%
Average	-3.42%	-4.98%	-6.47%

	Popularity	Locality	Density
AmazonComputer	87.83	85.77	50.46
AmazonPhoto	91.80	85.78	47.27
CiteSeer	70.56	64.55	58.04
CoauthorCS	83.68	81.13	63.24
CoauthorPhysics	86.70	82.06	39.68
CoraML	82.06	84.63	76.82
PubMed	66.64	63.34	55.79
Average	81.32	78.18	55.90

Table 2: Comparison of several graph baseline models for improving the OOD robustness (left) and detecting the OOD inputs by means of uncertainty estimation (right). For each task, we report the method ranks averaged across different graph datasets (lower is better). The ID to OOD ratio in these experiments is 70% to 30% instead of 50% to 50% as in the paper.

	Popularity		Locality		Density	
	ID	OOD	ID	OOD	ID	OOD
ERM	3.0	3.4	2.9	3.1	4.3	4.3
Mixup	1.9	2.6	1.4	2.7	1.4	3.3
EERM	4.9	3.9	5.0	4.0	4.4	4.1
DANN	4.1	4.4	4.3	4.3	3.7	2.6
CORAL	4.1	4.0	4.4	4.9	4.1	3.7
DE	3.0	2.7	3.0	2.0	3.0	3.0

	Popularity	Locality	Density
SE	1.3	2.3	3.7
GPN	3.3	3.9	3.9
NatPN	5.4	4.3	3.7
DE	3.0	1.4	1.9
GPE	3.1	3.9	3.4
NatPE	4.9	5.3	4.4

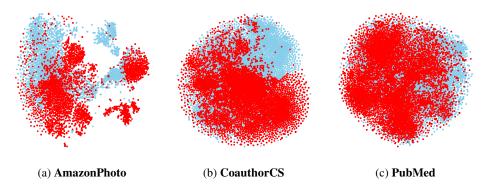


Figure 1: Visualization of distributional shift in node features induced by locality-based structural shift for three graph datasets. The picture is obtained by reducing the original node feature space into the 2D space of t-SNE representations: ID is blue, OOD is red.