	Size			d			$\bar{I}_{\mathbf{S}}$			$\bar{\ell}_2$		
	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ
Classification (12 datasets)	0.46	0.42	0.32	0.52	0.66	0.55	**0.92**	**0.88**	**0.73**	-0.79	-0.85	-0.66
Retrieval (15 datasets)	0.40	0.39	0.29	0.46	0.63	0.52	**0.89**	**0.89**	**0.70**	-0.71	-0.84	-0.65
Clustering (11 datasets)	0.45	0.38	0.26	0.54	0.67	0.55	**0.86**	**0.85**	**0.67**	-0.80	-0.84	-0.66
STS (10 datasets)	0.27	0.35	0.25	0.34	0.66	0.52	**0.92**	**0.82**	**0.62**	-0.70	-0.83	-0.64
Reranking (4 datasets)	0.33	0.33	0.26	0.41	0.61	0.50	**0.84**	**0.79**	**0.64**	-0.71	-0.78	-0.59
Average (56 datasets)	0.41	0.41	0.31	0.48	0.62	0.50	**0.94**	**0.90**	**0.74**	-0.77	-0.84	-0.65
Additional Classif (8 datasets)	0.41	0.62	0.47	0.43	0.64	0.55	**0.89**	**0.84**	**0.66**	-0.65	-0.72	-0.55

Table 1: Comparison with Baselines: Size of the Embedder, Dimension of the embedding output (d) and the ℓ_2 reconstruction error of the embeddings for NLP datasets.

	Size			d			$\bar{I}_{\mathbf{S}}$			$\bar{\ell}_2$		
	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ	ϱ_p	ϱ_s	τ
Absorption (8 datasets)	-	-0.21	-0.16	-	-0.43	-0.29	-	**0.89**	**0.70**	-	**-0.89**	**-0.70**
Distribution (3 datasets)	-	-0.07	-0.03	-	-0.46	-0.31	-	**0.89**	**0.70**	-	-0.86	-0.66
Metabolism (8 datasets)	-	0.06	0.03	-	-0.46	-0.34	-	**0.94**	**0.79**	-	-0.90	-0.71
Excretion (3 datasets)	-	-0.17	-0.11	-	-0.24	-0.18	-	**0.77**	**0.60**	-	**-0.77**	-0.56
Toxicity (9 datasets)	-	0.09	0.06	-	-0.49	-0.35	-	**0.92**	**0.75**	-	-0.86	-0.67
ADMET (31 datasets)	-	-0.01	0.01	-	-0.47	-0.32	-	**0.94**	**0.80**	-	-0.90	-0.72

Table 2: Comparison with Baselines: Size of the Embedder, Dimension of the embedding output (d) and the ℓ_2 reconstruction error of the embeddings for Molecular Modelling datasets.

Algorithm 1 Estimation of $\mathcal{I}_S(U \to Z)$, $\mathrm{GM}_{\mu,\mathbf{\Sigma},\mathbf{w}}$ denotes the Gaussian Mixture model with means μ, covariances $\mathbf{\Sigma}$ and weights $\mathbf{w}$.

Input: Pairs of corresponding embeddings $(z_i, u_i)_N$
Output: Information sufficiency $\mathcal{I}_S(U \to Z)$

$\mu_{\mathbf{Z}}, \mathbf{\Sigma}_{\mathbf{Z}}, \mathbf{w}_{\mathbf{Z}} \leftarrow \arg\min_{\mu,\mathbf{\Sigma},\mathbf{w}} - \sum_{i=1}^{N} \log \mathrm{GM}_{\mu,\mathbf{\Sigma},\mathbf{w}}(z_i)$

$\mu_{\mathbf{Z}|\mathbf{U}}, \mathbf{\Sigma}_{\mathbf{Z}|\mathbf{U}}, \mathbf{w}_{\mathbf{Z}|\mathbf{U}} \leftarrow \arg\min_{\mu,\mathbf{\Sigma},\mathbf{w}} - \sum_{i=1}^{N} \log \mathrm{GM}_{\mu(u_i),\mathbf{\Sigma}(u_i),\mathbf{w}(u_i)}(z_i)$

Z: $h(Z) \leftarrow \frac{1}{N} \sum_{i=1}^{N} \log \mathrm{GM}_{\mu_{\mathbf{Z}},\mathbf{\Sigma}_{\mathbf{Z}},\mathbf{w}_{\mathbf{Z}}}(z_i)$

$h(Z|U) \leftarrow \frac{1}{N} \sum_{i=1}^{N} \log \mathrm{GM}_{\mu_{\mathbf{Z}|\mathbf{U}}(u_i),\mathbf{\Sigma}_{\mathbf{Z}|\mathbf{U}}(u_i),\mathbf{w}_{\mathbf{Z}|\mathbf{U}}(u_i)}(z_i)$

Return $\mathcal{I}_S(U \to Z) \leftarrow h(Z) - h(Z|U)$

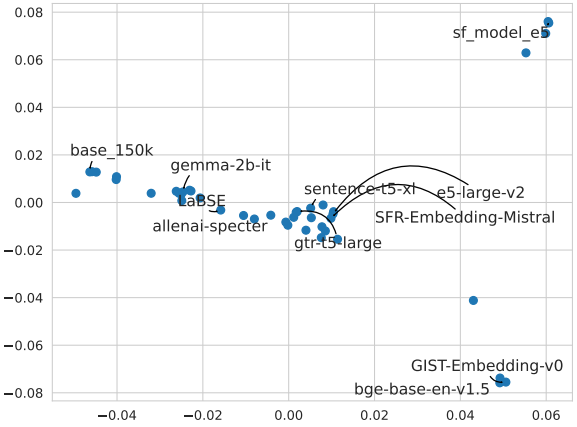

Figure 1: Spectral embedding of the embedders using our similarity measure for the MTEB dataset.