Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Supporting Information for:

Metal-Organic Framework Clustering through the Lens of Transfer Learning

*Gregory M. Cooper and Yamil J. Colón**

Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA, 46556

*corresponding author: ycolon@nd.edu

S1. Normalized adsorption performance in principal component space.

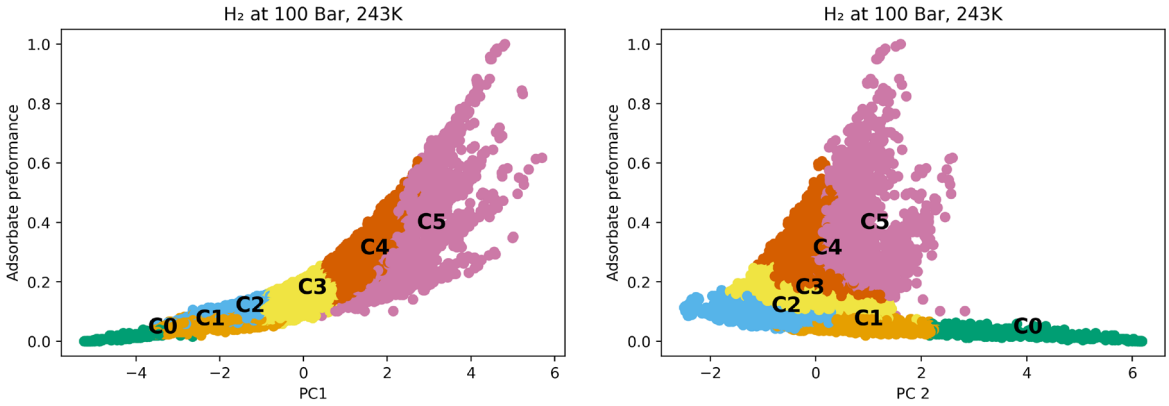

Figure S1. Normalized adsorption performance for H_2 at 100 Bar 243 K.

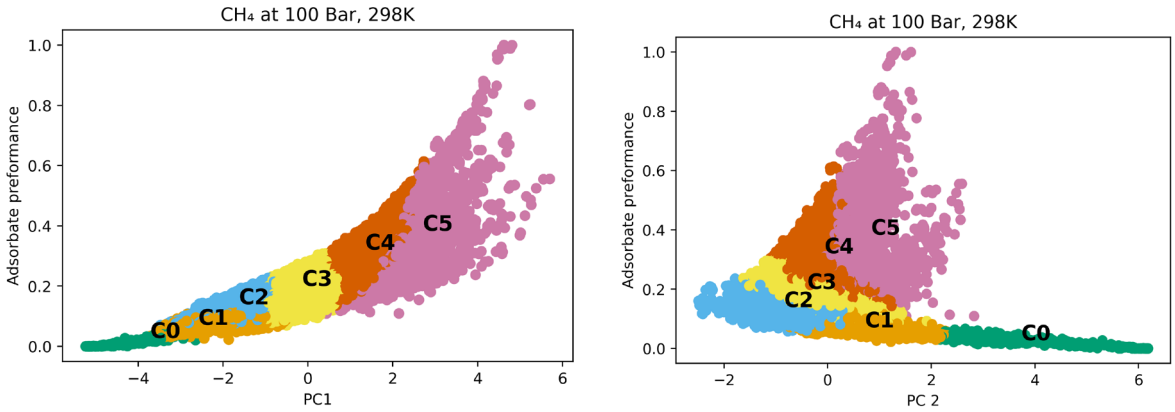

Figure S2. Normalized adsorption performance for CH_4 at 100 Bar 298 K.

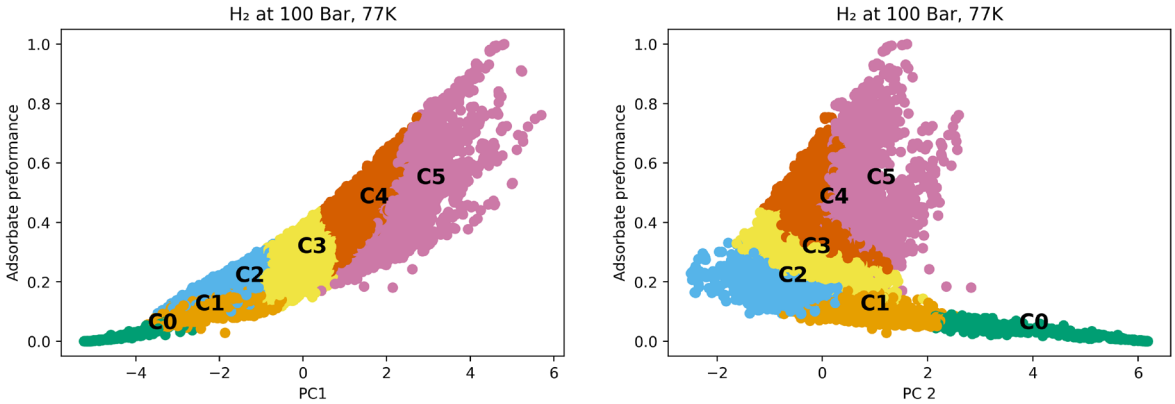

Figure S3. Normalized adsorption performance for H_2 at 100 Bar 77 K.

S2. MOFs by topology in principal component space.

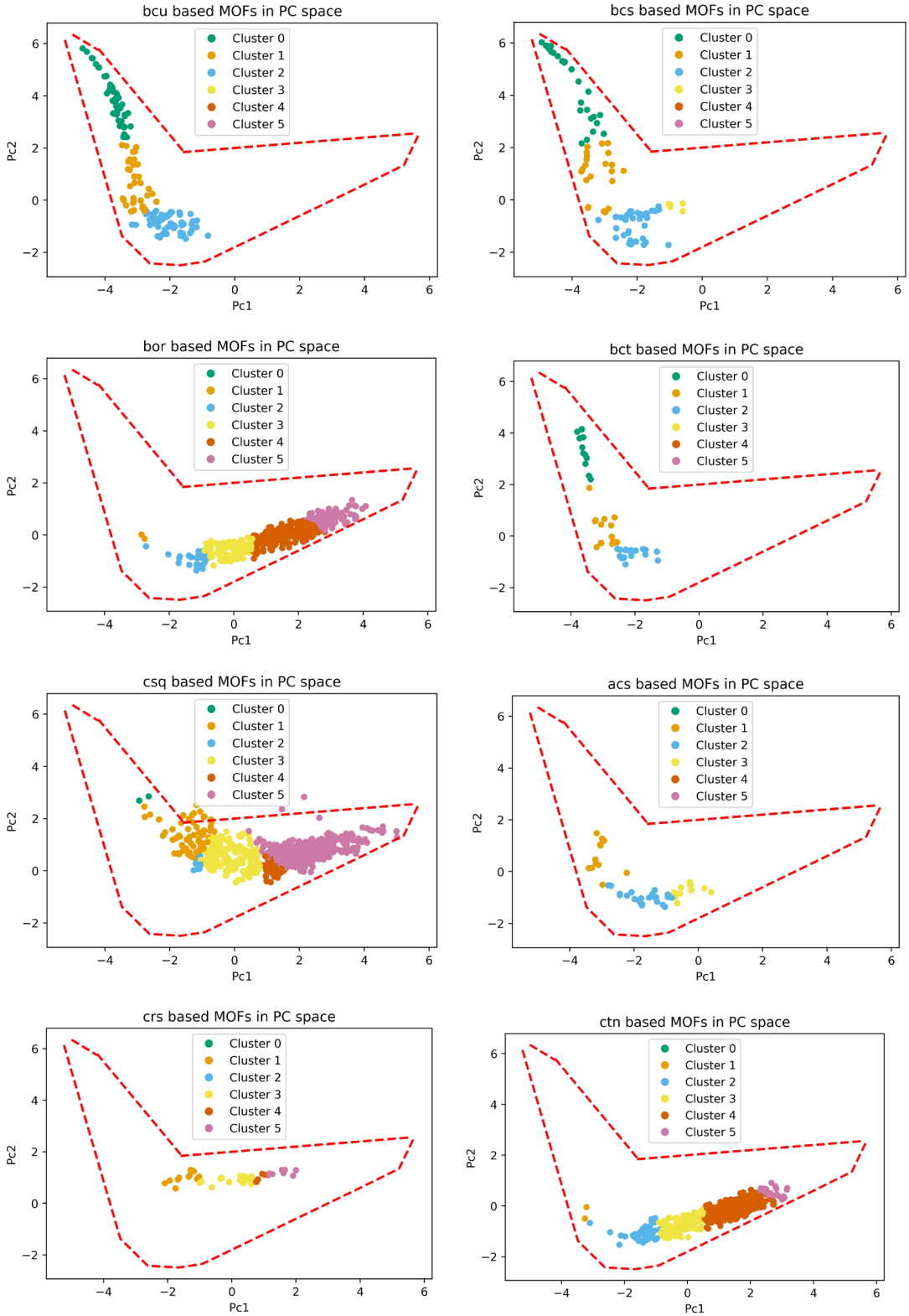

Figure S4. The first set of topologies in the PC Space.

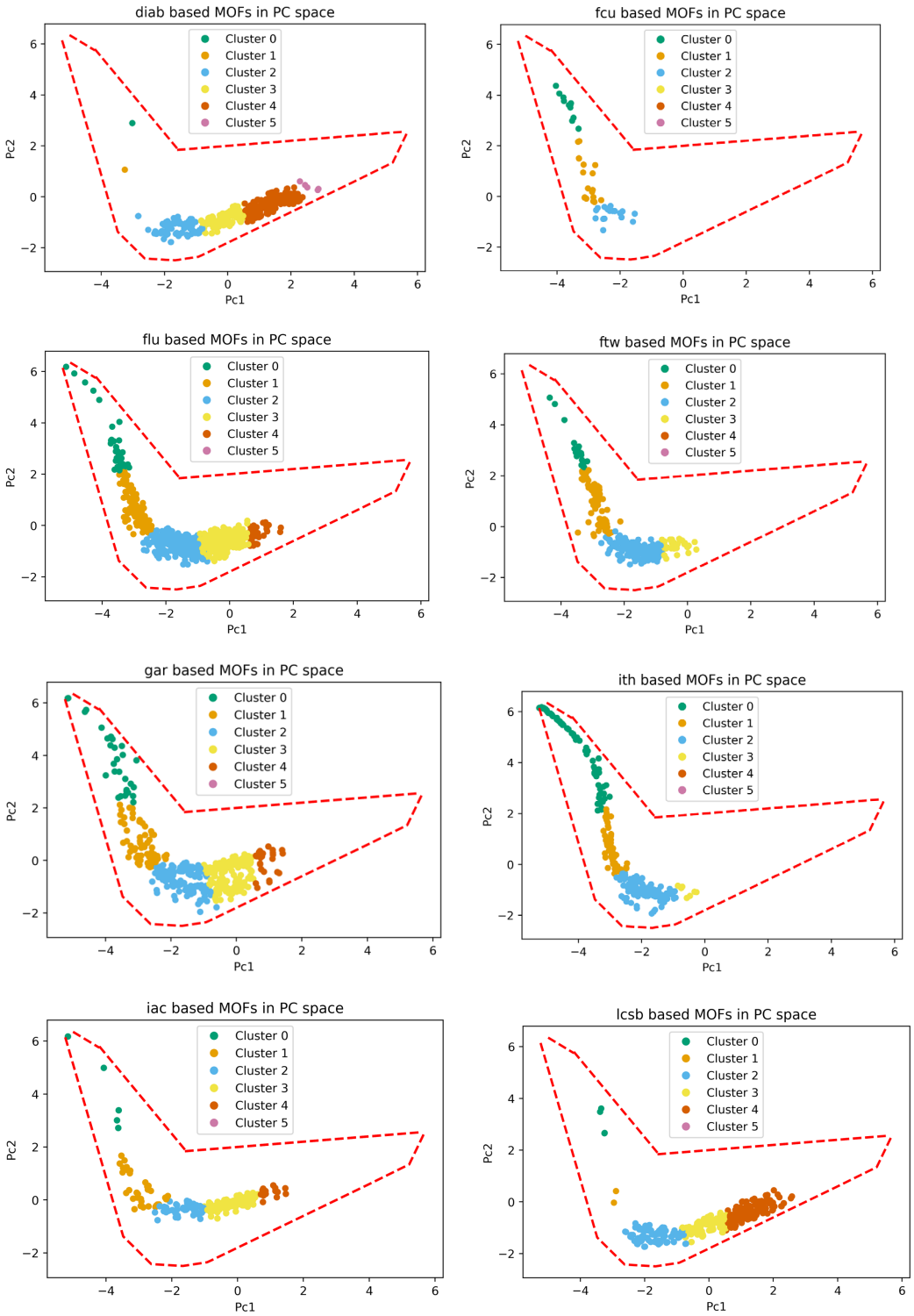

Figure S5. The second set of topologies in the PC Space.

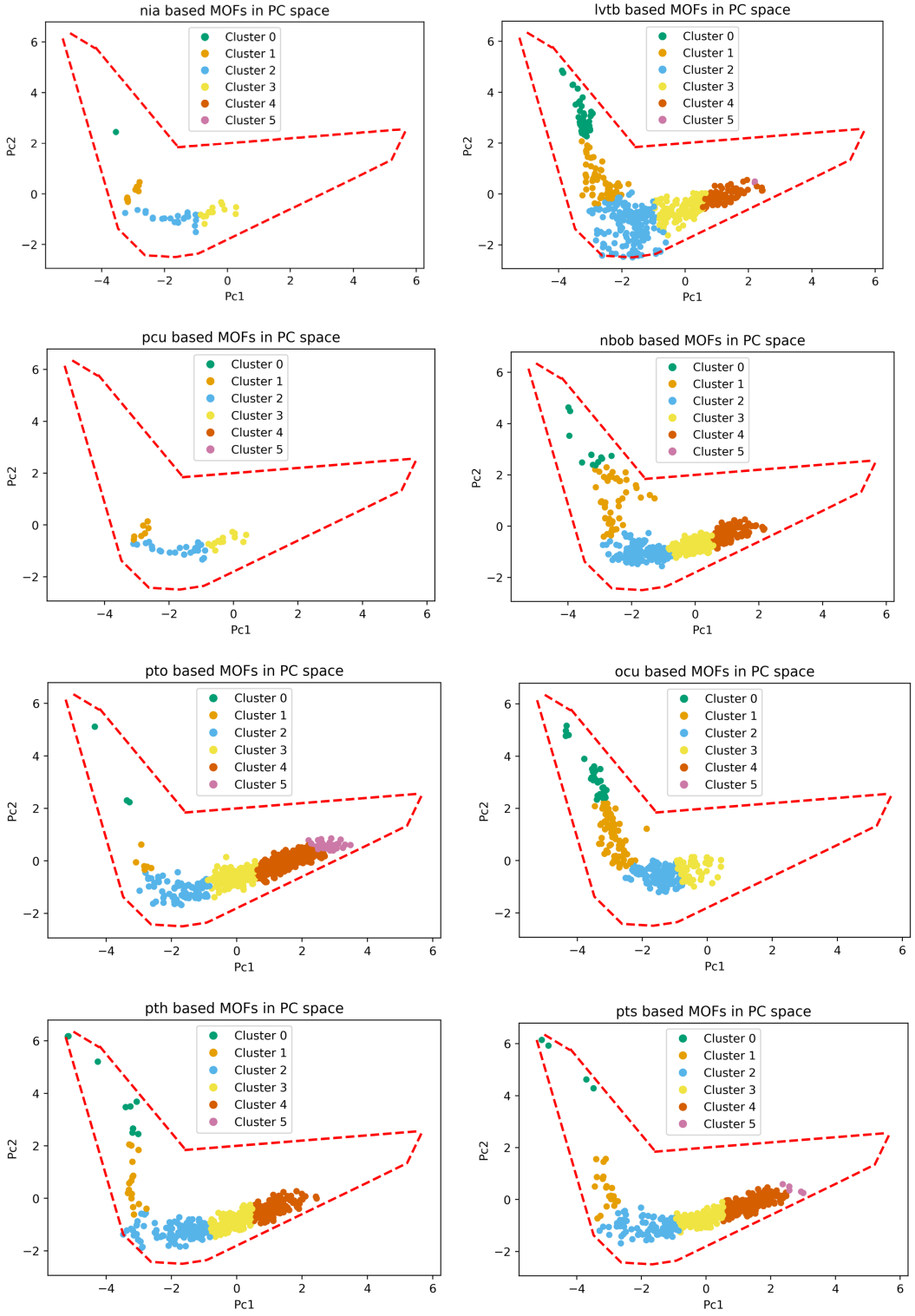

Figure S6. The third set of topologies in the PC Space.

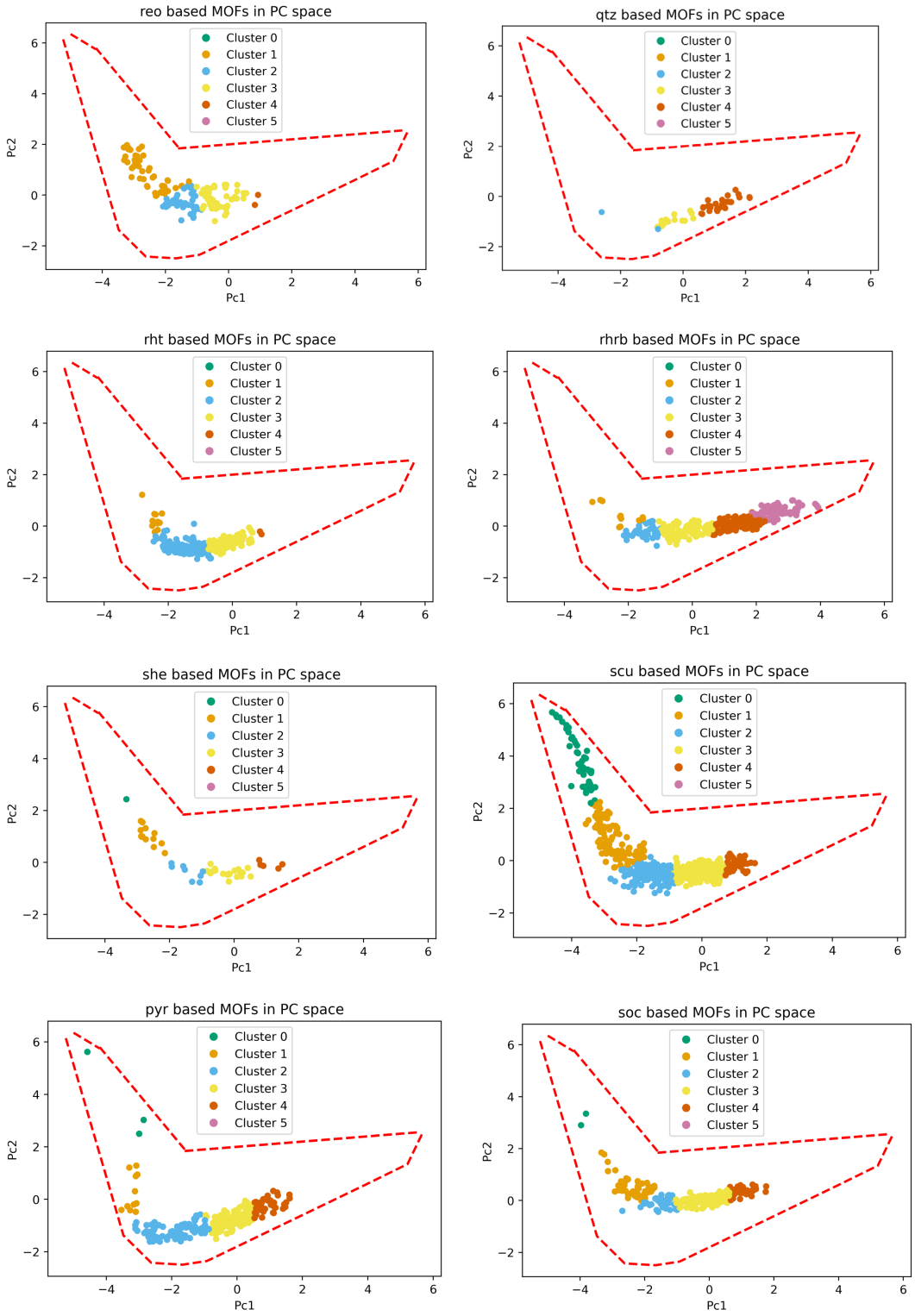

Figure S7. The fourth set of topologies in the PC Space.

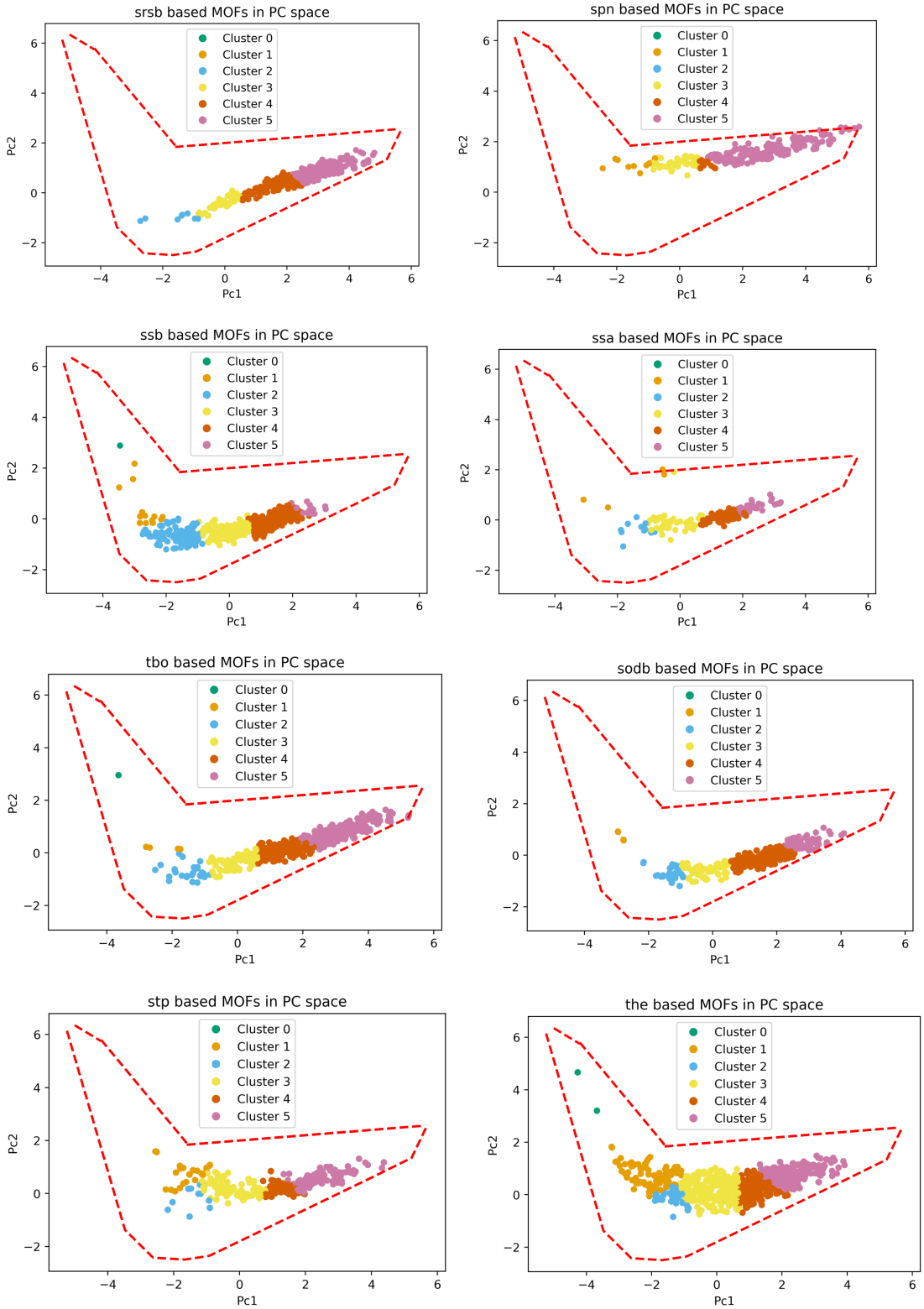

Figure S8. The Final set of topologies in the PC Space.

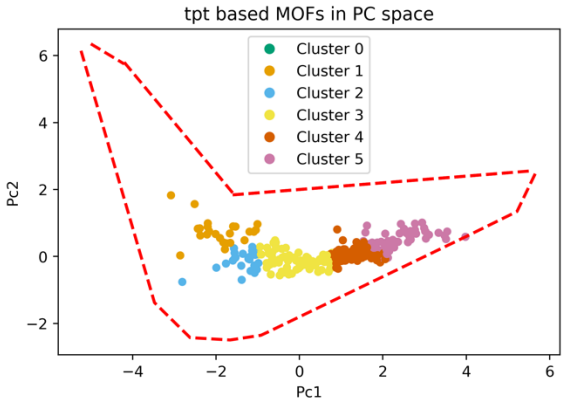

Figure S9. TPT topology in the PC space

S3. Median of topologies in principal component space

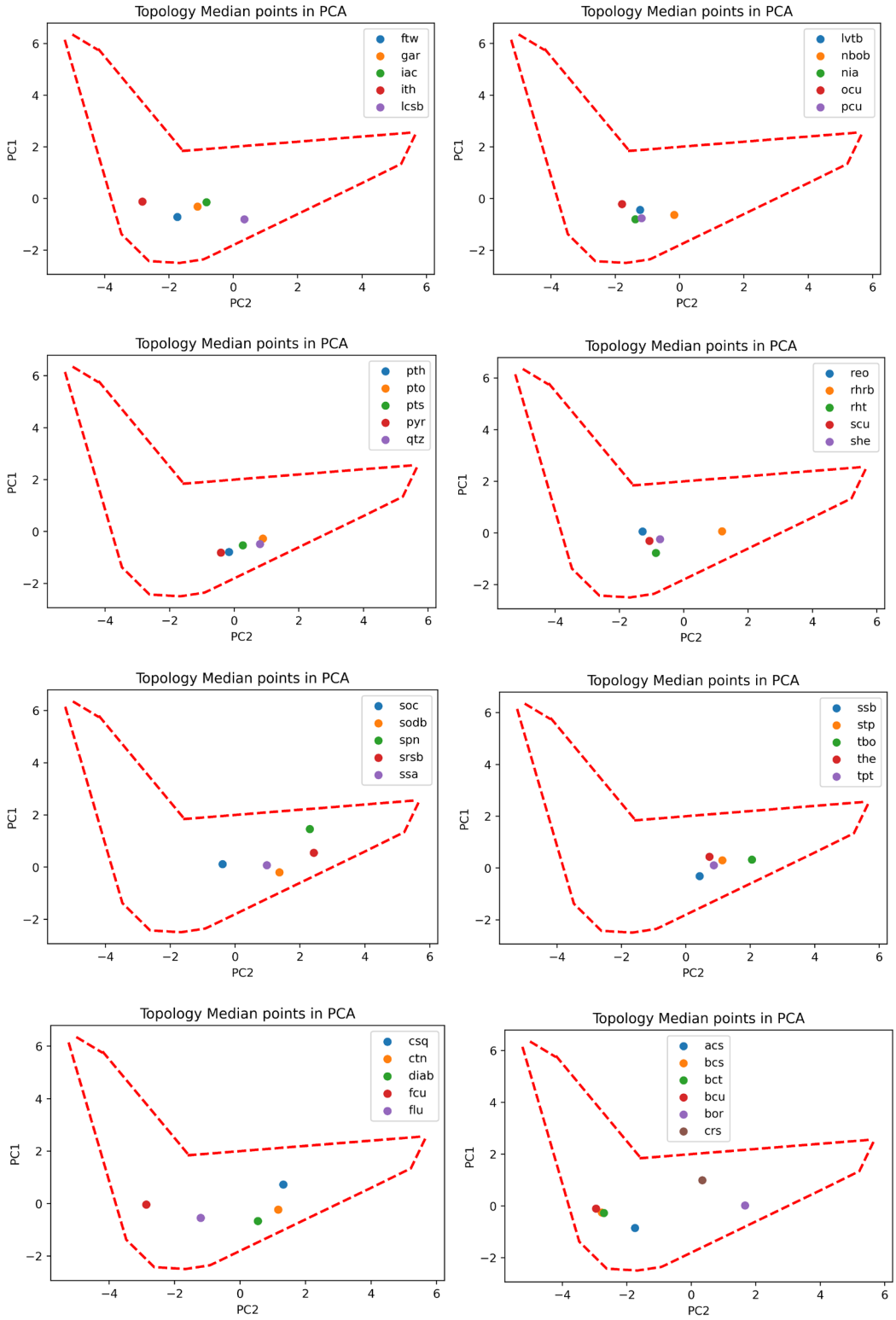

Figure S10. The topologies in the PCA space limited to 5 per plot

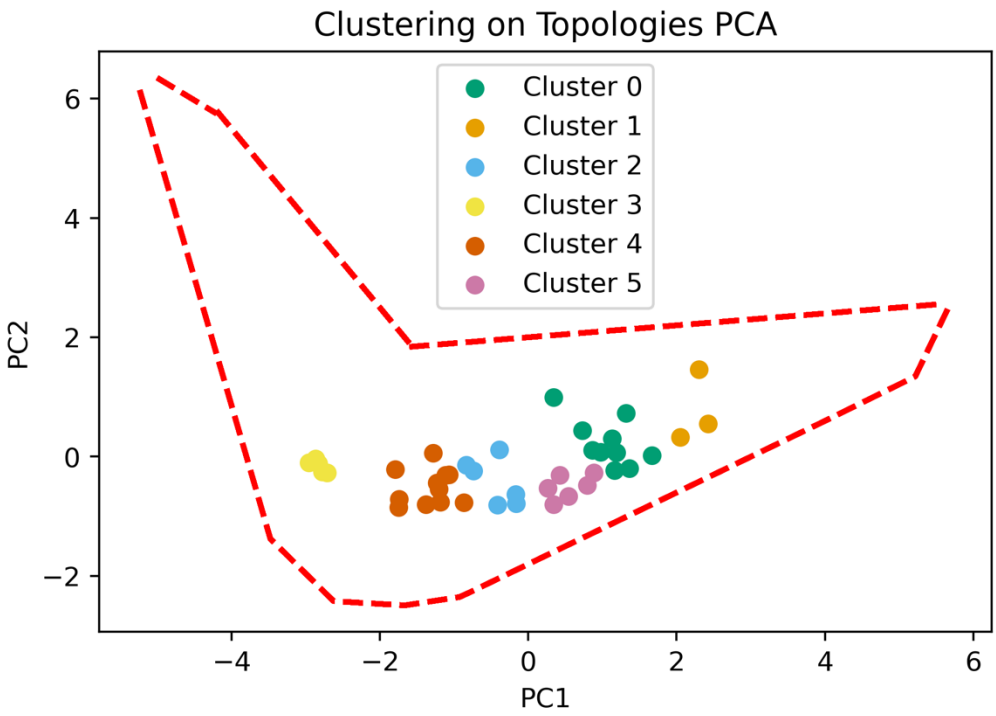

Figure S11. The whole set of topology clusters by median point. Refer to above specific topology location.

S4. Transfer learning performance in generic clusters for different base clusters

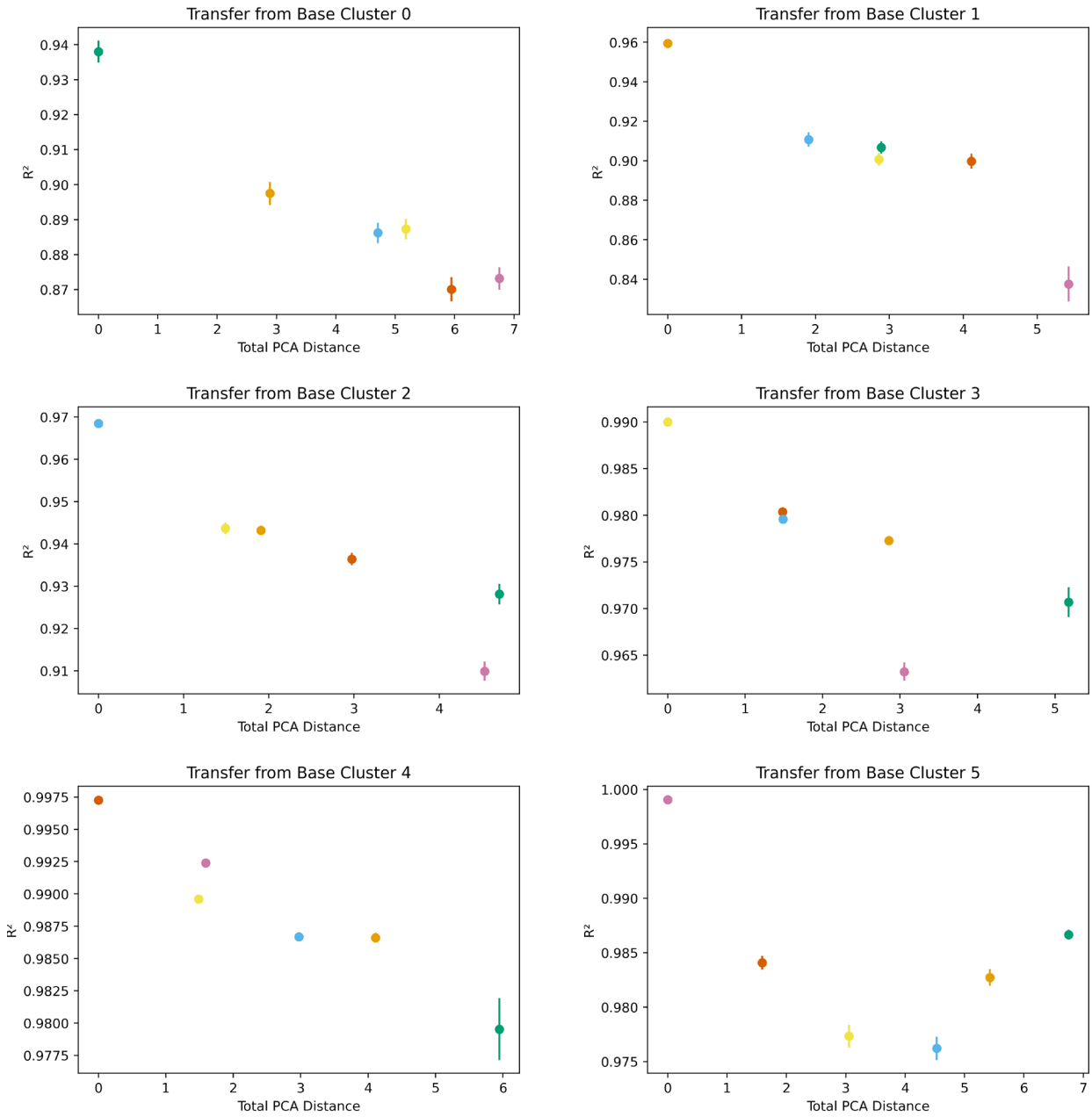

Figure S12. Generic Cluster Transfer for H_2 at 100 bar and 243 K transfer performance in terms of R^2

S5. Representative sample of the generic cluster's performance using all metrics

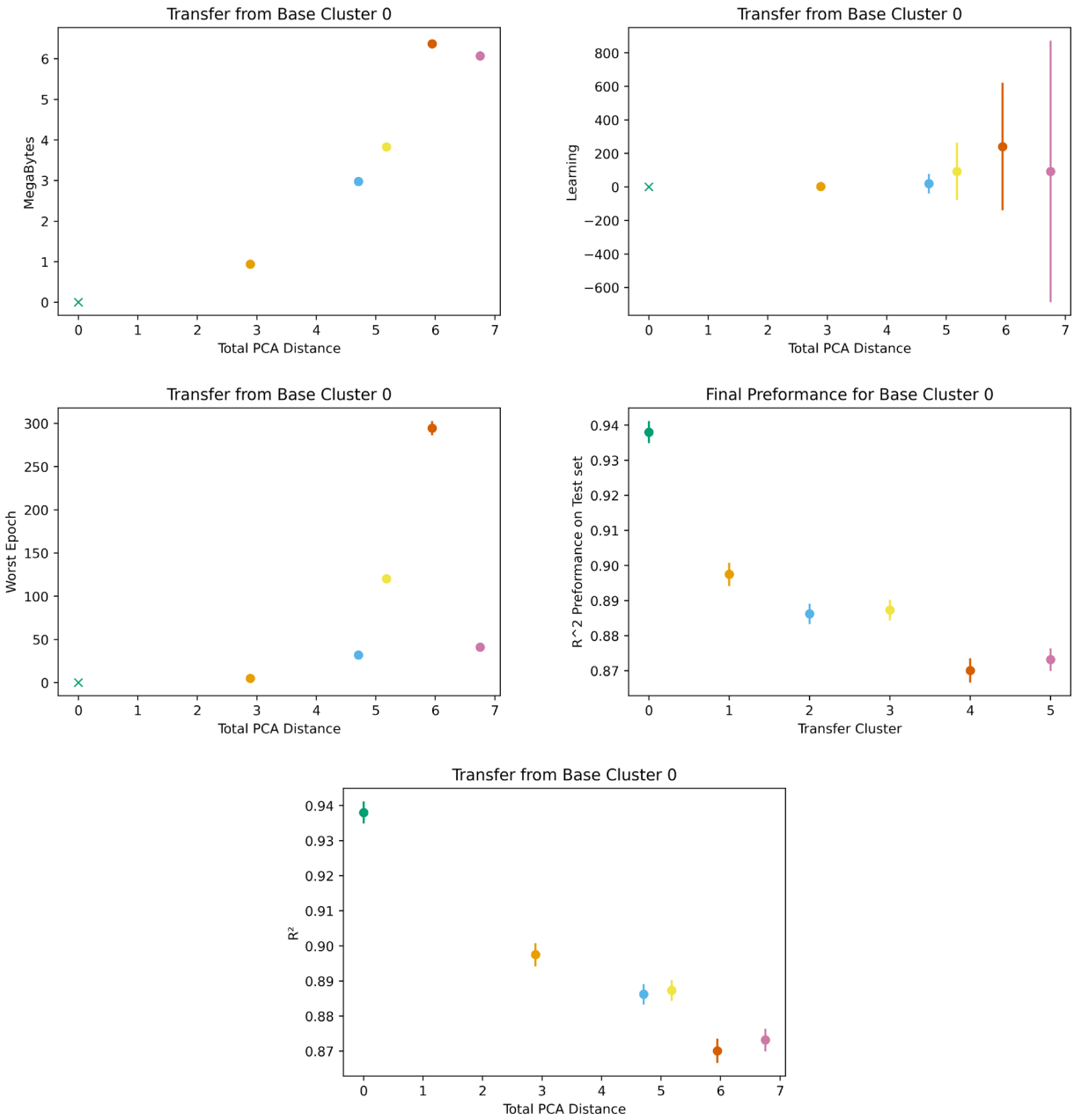

Figure S13. Generic Cluster Transfer for Cluster 0 using Total PC distance H_2 at 100 bar and 243 K transfer

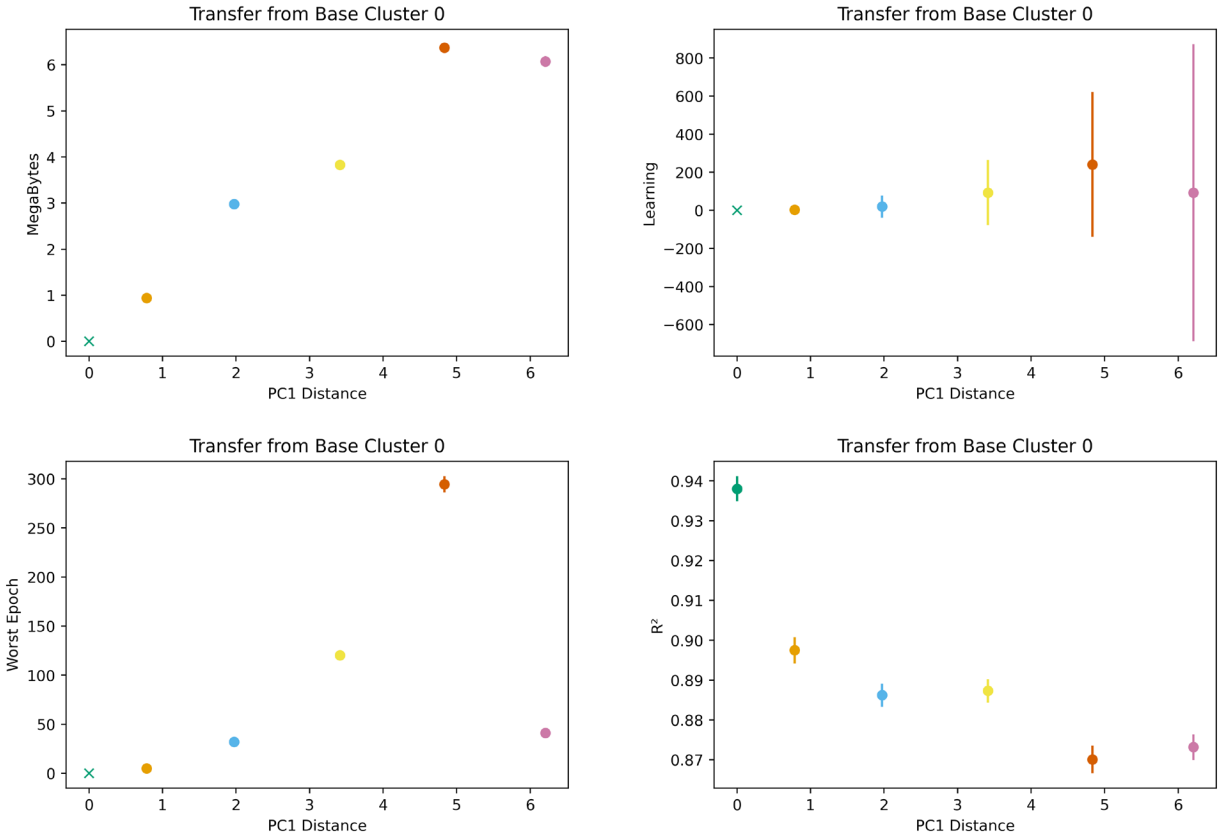

Figure S14. Generic Cluster Transfer for Cluster 0 using PC1 for H_2 at 100 bar and 243 K transfer

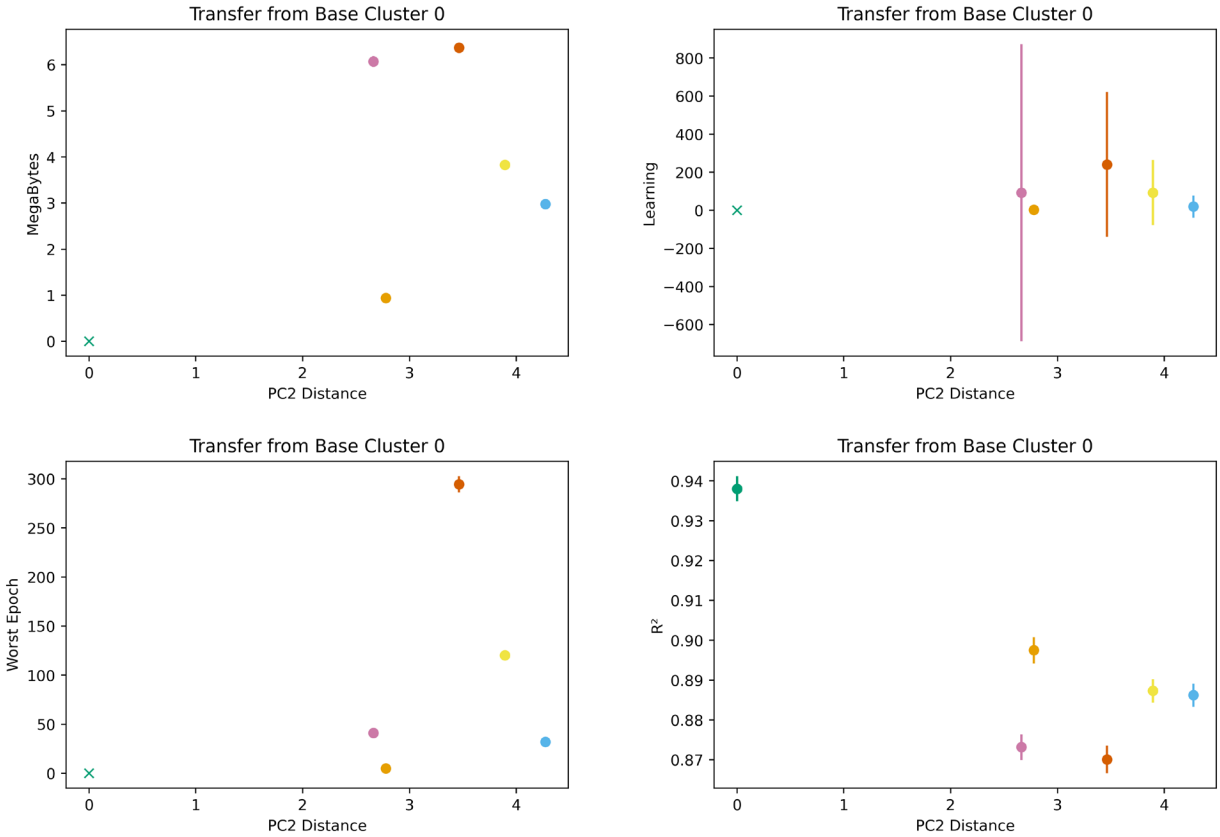

Figure S15. Generic Cluster Transfer for Cluster 0 using PC2 for H_2 at 100 bar and 243 K transfer

S6. Remaining metrics (exclusive of manuscript figures) regarding base performance of generic clusters

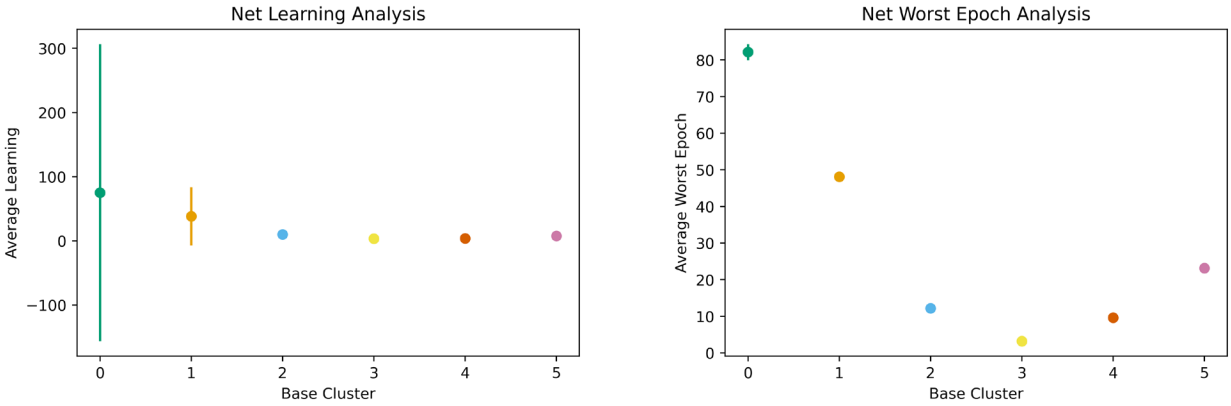

Figure S16. Worst Epoch and Learning analysis for gH_2 at 100 bar and 243 K generic transfer

S7. Representative sample of the topology cluster's performance using all metrics

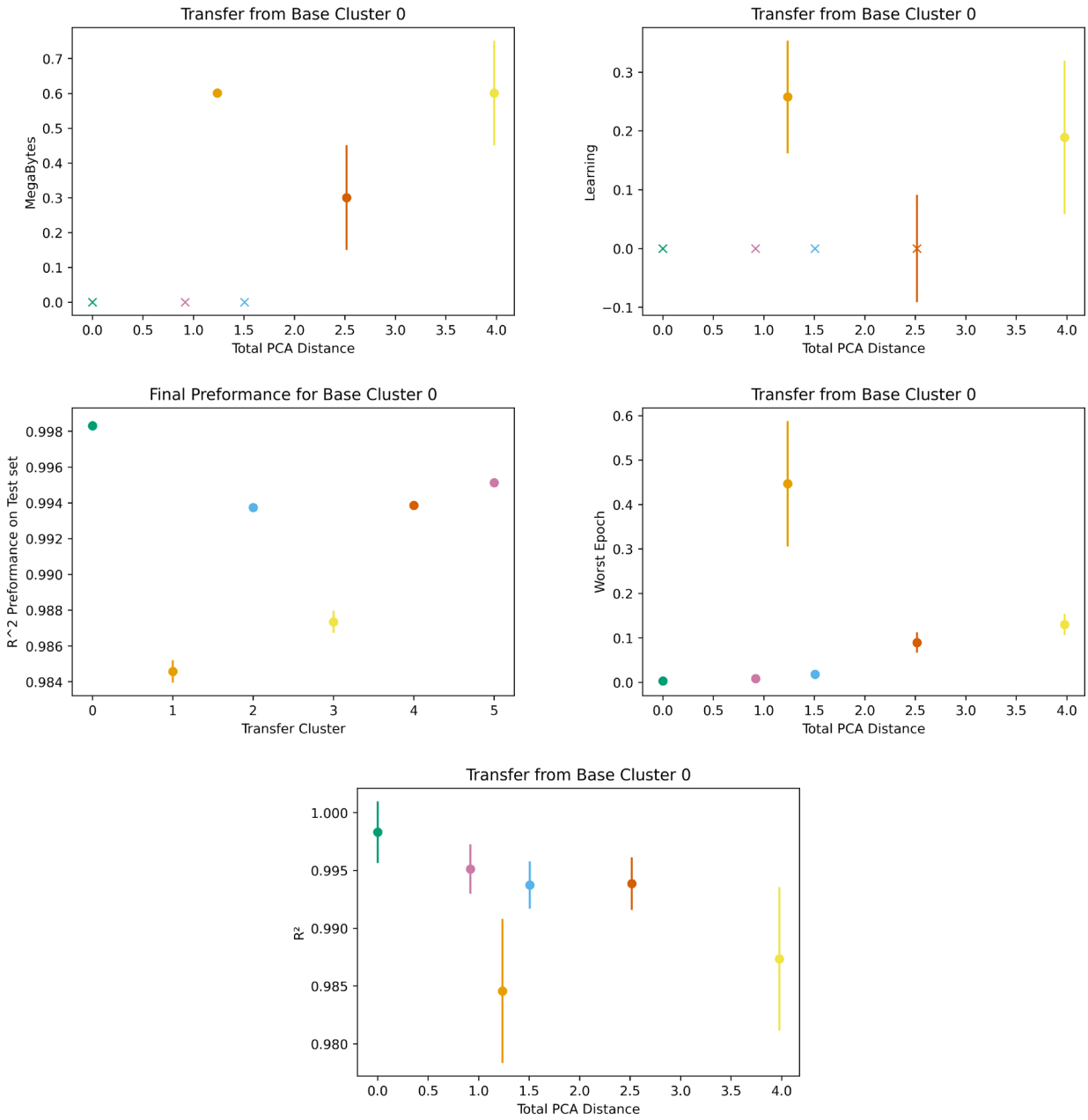

Figure S17. Topology Cluster Transfer for Cluster 0 using Total PC distance H_2 at 100 bar and 243 K transfer

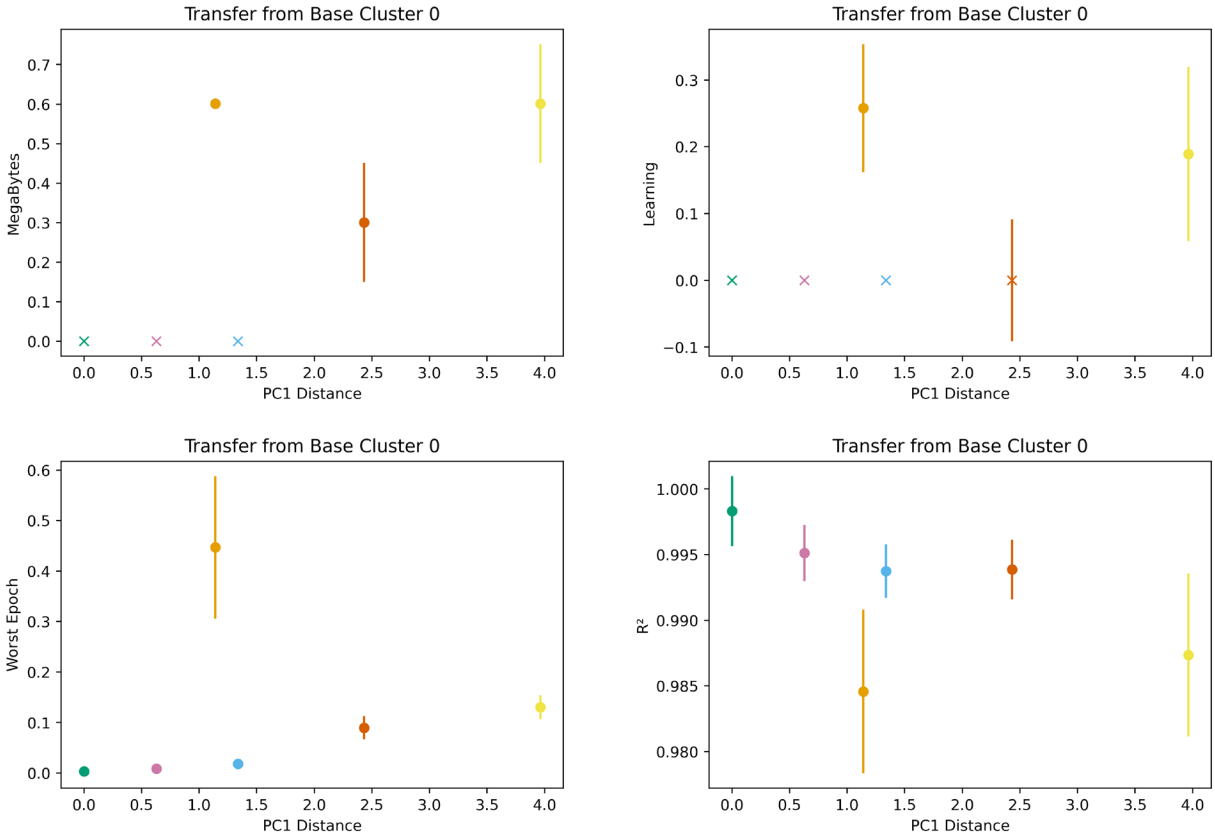

Figure S18. Topology Cluster Transfer for Cluster 0 using PC1 for H_2 at 100 bar and 243 K transfer

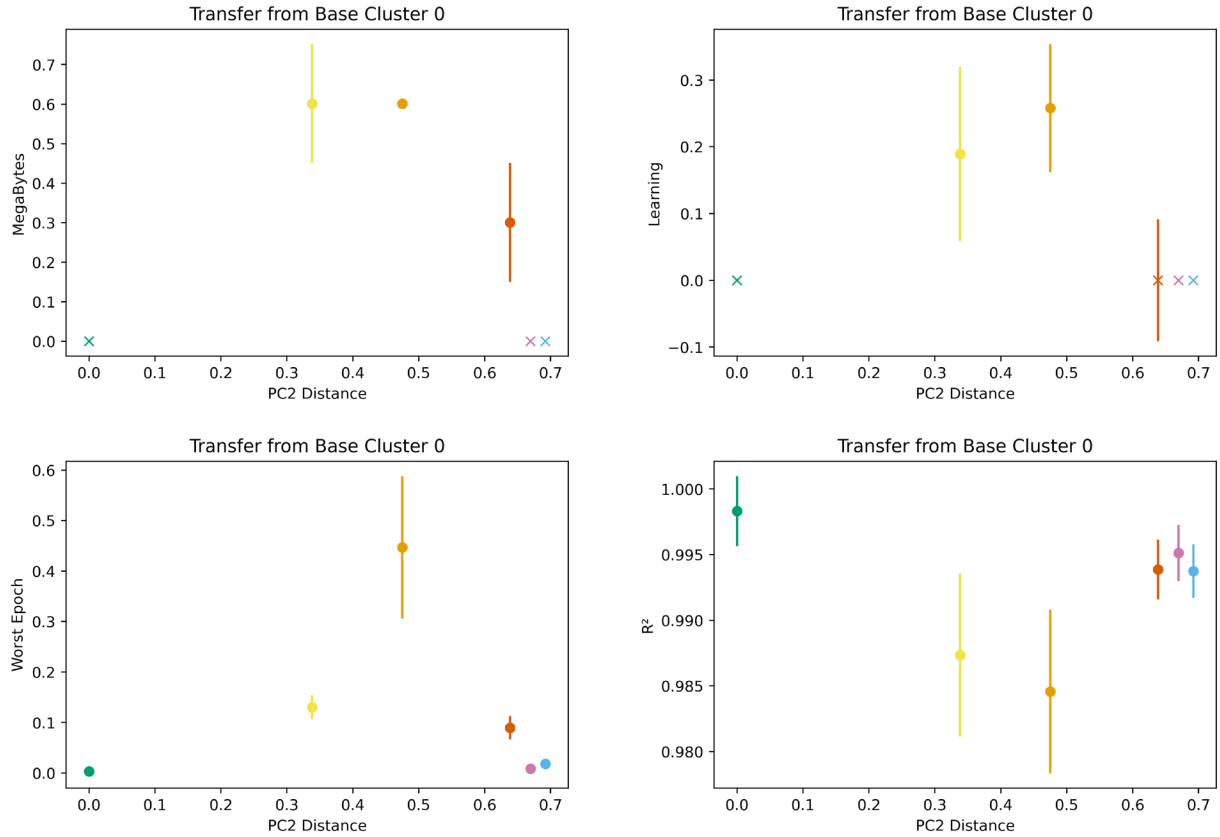

Figure S19. Topology Cluster Transfer for Cluster 0 using PC2 for H_2 at 100 bar and 243 K transfer

S8. Topology cluster metrics for all base clusters

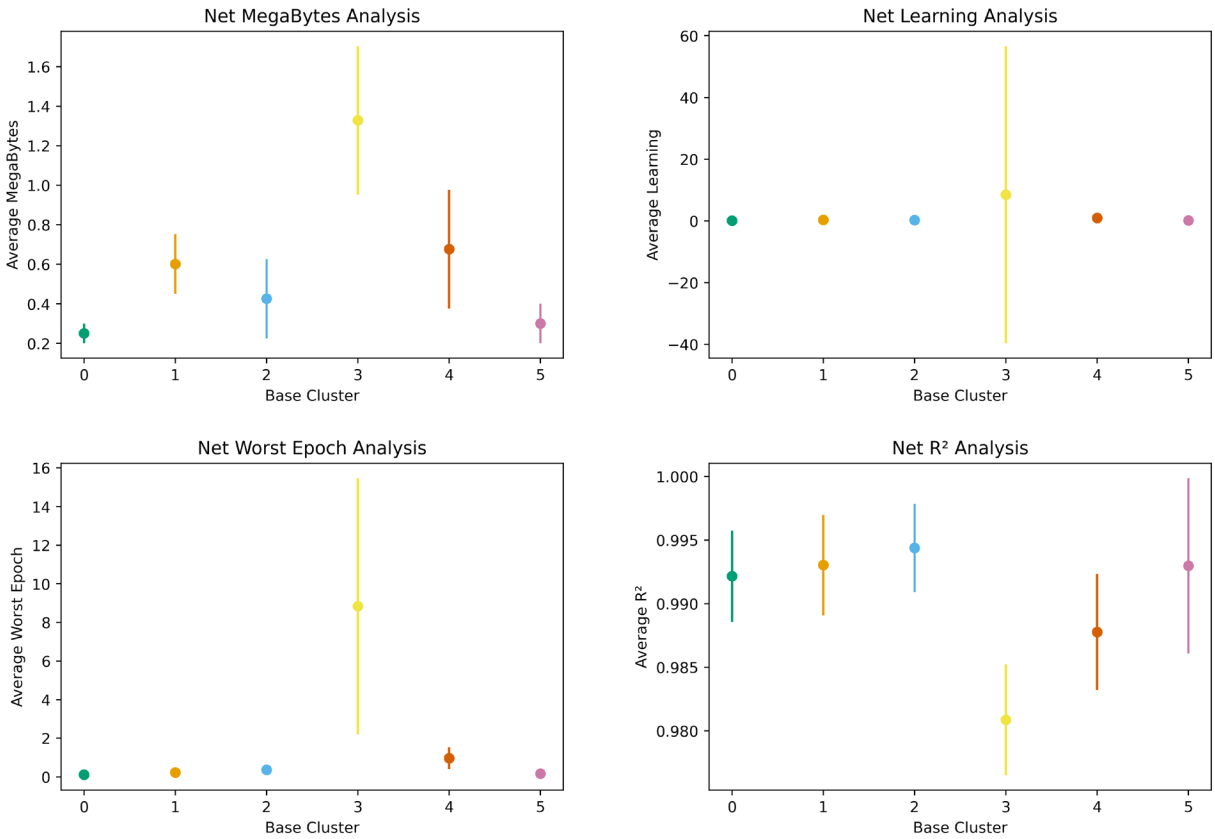

Figure S20. Topology Cluster Transfer for all base clusters using average values for H_2 at 100 bar and 243 K transfer

S9. Representative sample of generic cluster TL performance from hydrogen at 243 K to methane at 298 K

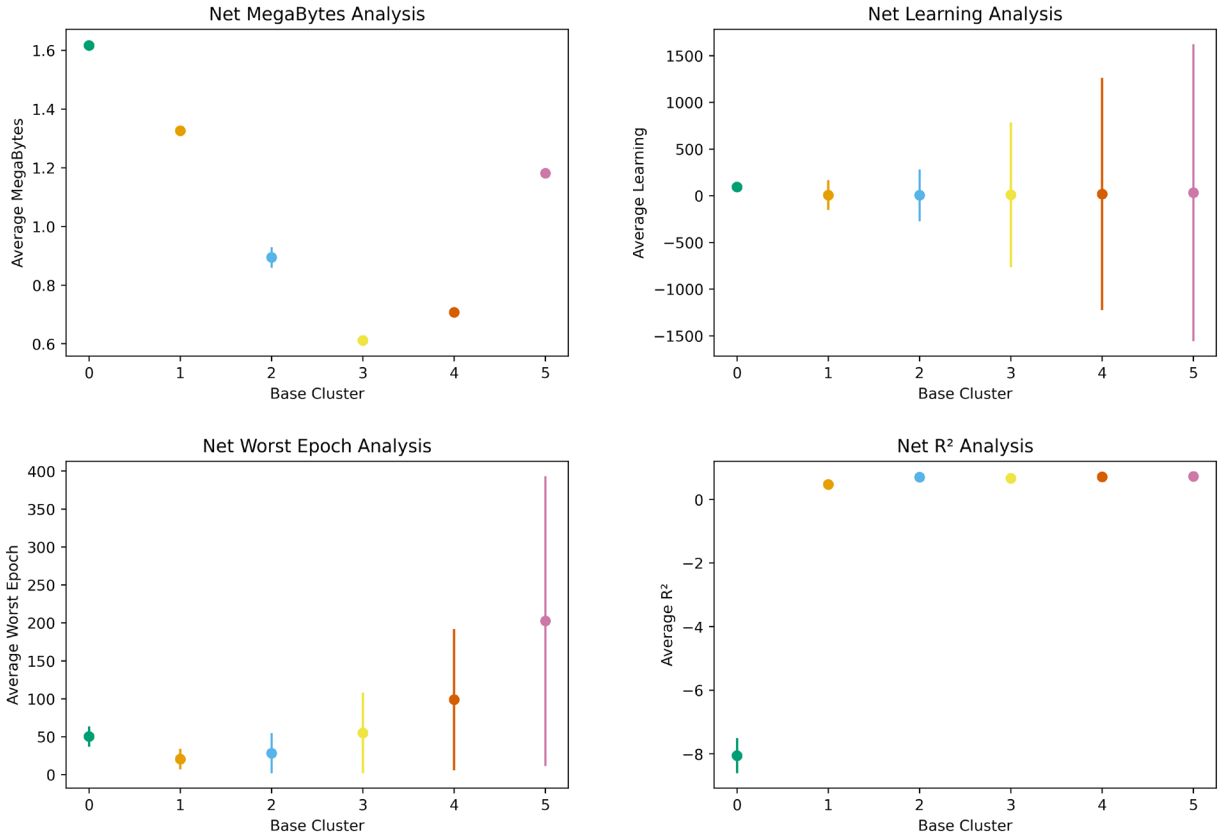

Figure S21. Generic Cluster Transfer for all base clusters using average values for H_2 at 100 bar and 243 K transfer to methane at 298 K

S10. Representative sample of topology cluster TL performance from hydrogen at 243 K to methane at 298 K

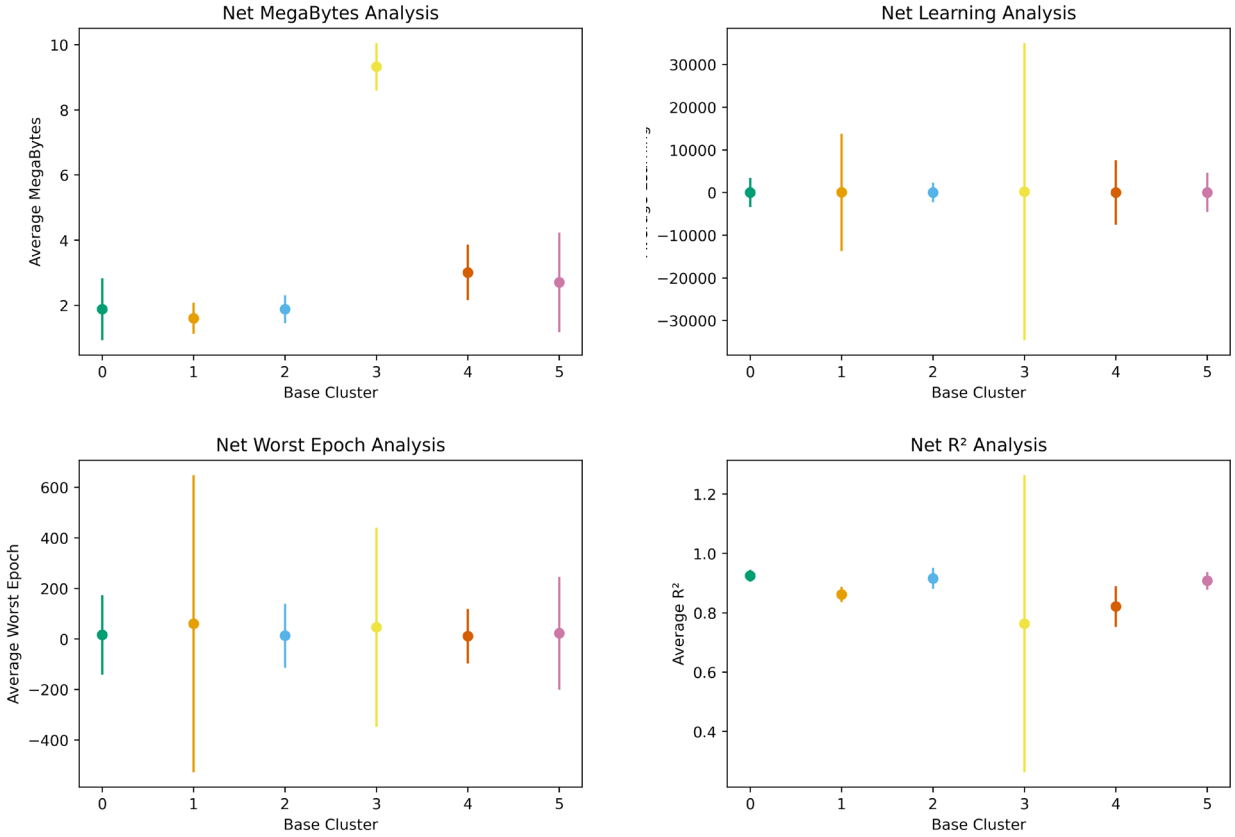

Figure S22. Topology Cluster Transfer for all base clusters using average values for H_2 at 100 bar and 243 K transfer to methane at 298 K

S11. Representative sample of generic cluster TL performance from methane at 298 K to H2 at 100 bar and 243 K

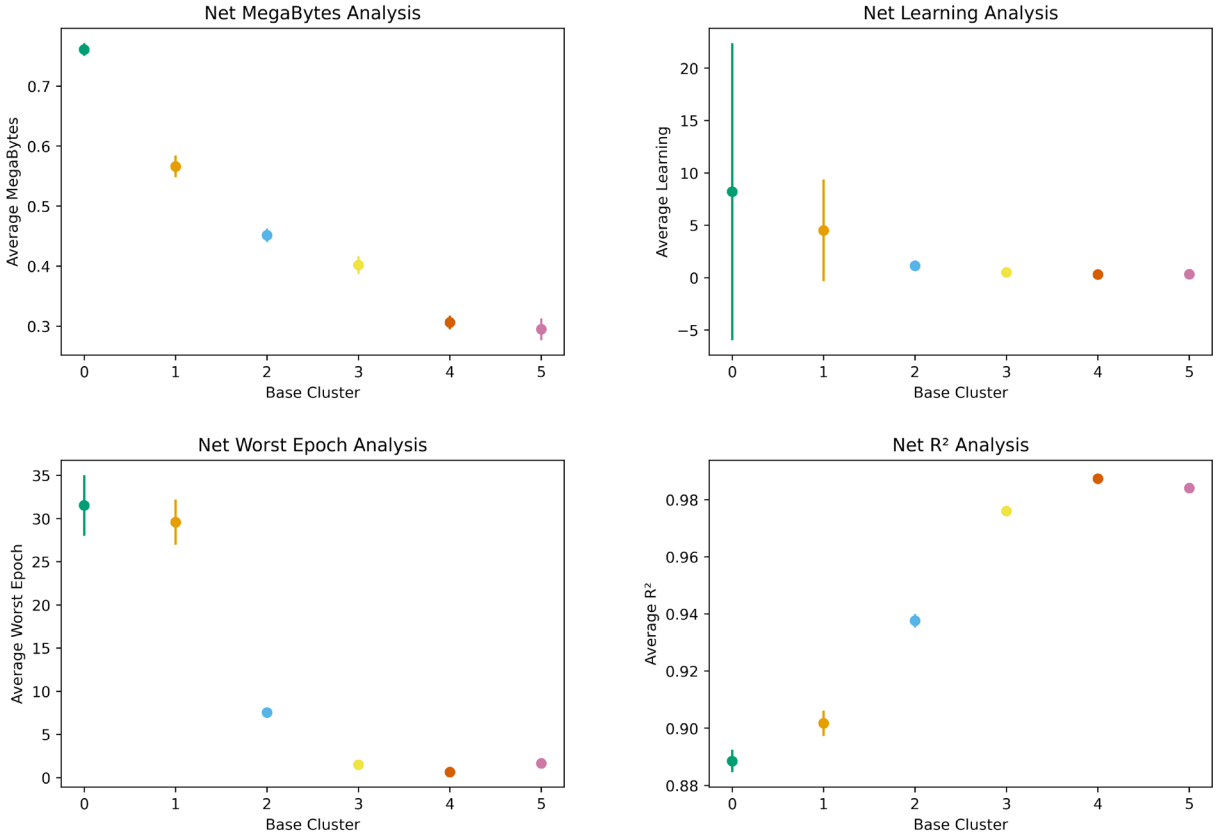

Figure S23. Topology Cluster Transfer for all base clusters using average values from methane at 298 K to H_2 at 100 bar and 243 K transfer

S12. Representative sample of topology cluster TL performance from methane at 298 K to H2 at 100 bar and 243 K

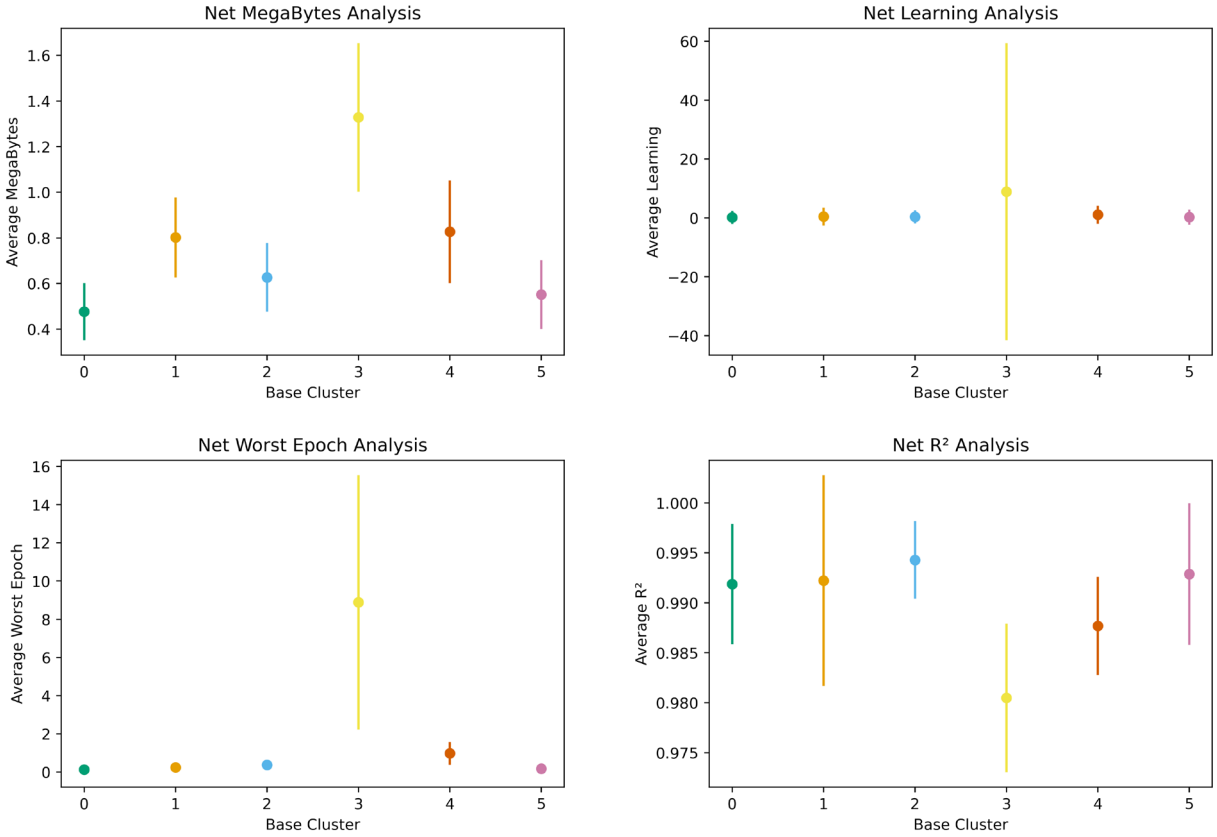

Figure S24. Topology Cluster Transfer for all base clusters using average values from methane at 298 K to H_2 at 100 bar and 243 K transfer

S13. Correlation data for generic clusters with respect to adsorption tasks

Table 1. The set of correlation data for all generic clusters partitioned by generic cluster.

General Cluster Correlation Data					
	H2@100 bar/77K (wt%)	**H2@100 bar/243K (wt%)**	**CH4@100 bar/298 K (mg/g)**	**Pc1**	**Pc2**
Correlation for whole MOF set					
H2@100 bar/77K (wt%)	1.0	0.979	0.988	0.917	-0.067
H2@100 bar/243K (wt%)	0.979	1.0	0.990	0.872	0.054
CH4@100 bar/298 K (mg/g)	0.988	0.99	1.0	0.875	-0.052
Pc1	0.917	0.872	0.875	1.0	-0.0
Pc2	-0.067	0.054	-0.052	-0.0	1.0
Correlation for 0^{th} Cluster					
H2@100 bar/77K (wt%)	1.0	0.936	0.975	0.907	-0.884
H2@100 bar/243K (wt%)	0.936	1.0	0.943	0.856	-0.761
CH4@100 bar/298 K (mg/g)	0.975	0.943	1.0	0.893	-0.86
Pc1	0.907	0.856	0.893	1.0	-0.912
Pc2	-0.884	-0.761	-0.86	-0.912	1.0
Correlation for 1^{st} Cluster					
H2@100 bar/77K (wt%)	1.0	0.937	0.971	0.684	-0.573
H2@100 bar/243K (wt%)	0.937	1.0	0.962	0.58	-0.464
CH4@100 bar/298 K (mg/g)	0.971	0.962	1.0	0.585	-0.574

Pc1	0.684	0.58	0.585	1.0	-0.079
Pc2	-0.573	-0.464	-0.574	-0.079	1.0
Correlation for 2nd Cluster					
H2@100 bar/77K (wt%)	1.0	0.988	0.979	0.79	-0.482
H2@100 bar/243K (wt%)	0.988	1.0	0.983	0.773	-0.463
CH4@100 bar/298 K (mg/g)	0.979	0.983	1.0	0.706	-0.516
Pc1	0.79	0.773	0.706	1.0	0.009
Pc2	-0.482	-0.463	-0.516	0.009	1.0
Correlation for 3rd Cluster					
H2@100 bar/77K (wt%)	1.0	0.992	0.976	0.598	-0.63
H2@100 bar/243K (wt%)	0.992	1.0	0.963	0.657	-0.545
CH4@100 bar/298 K (mg/g)	0.976	0.963	1.0	0.515	-0.678
Pc1	0.598	0.657	0.515	1.0	0.133
Pc2	-0.63	-0.545	-0.678	0.133	1.0
Correlation for 4th Cluster					
H2@100 bar/77K (wt%)	1.0	0.991	0.986	0.746	-0.045
H2@100 bar/243K (wt%)	0.991	1.0	0.979	0.786	0.057
CH4@100 bar/298 K (mg/g)	0.986	0.979	1.0	0.732	-0.08
Pc1	0.746	0.786	0.732	1.0	0.45
Pc2	-0.045	0.057	-0.08	0.45	1.0
Correlation for 5th Cluster					

H2@100 bar/77K (wt%)	1.0	0.989	0.985	0.743	0.005
H2@100 bar/243K (wt%)	0.989	1.0	0.994	0.735	0.066
CH4@100 bar/298 K (mg/g)	0.985	0.994	1.0	0.719	-0.005
Pc1	0.743	0.735	0.719	1.0	0.413
Pc2	0.005	0.066	-0.005	0.413	1.0

S14. Comparison of model with respect to cluster 5 transfer from methane to Hydrogen tasks and the vice versa

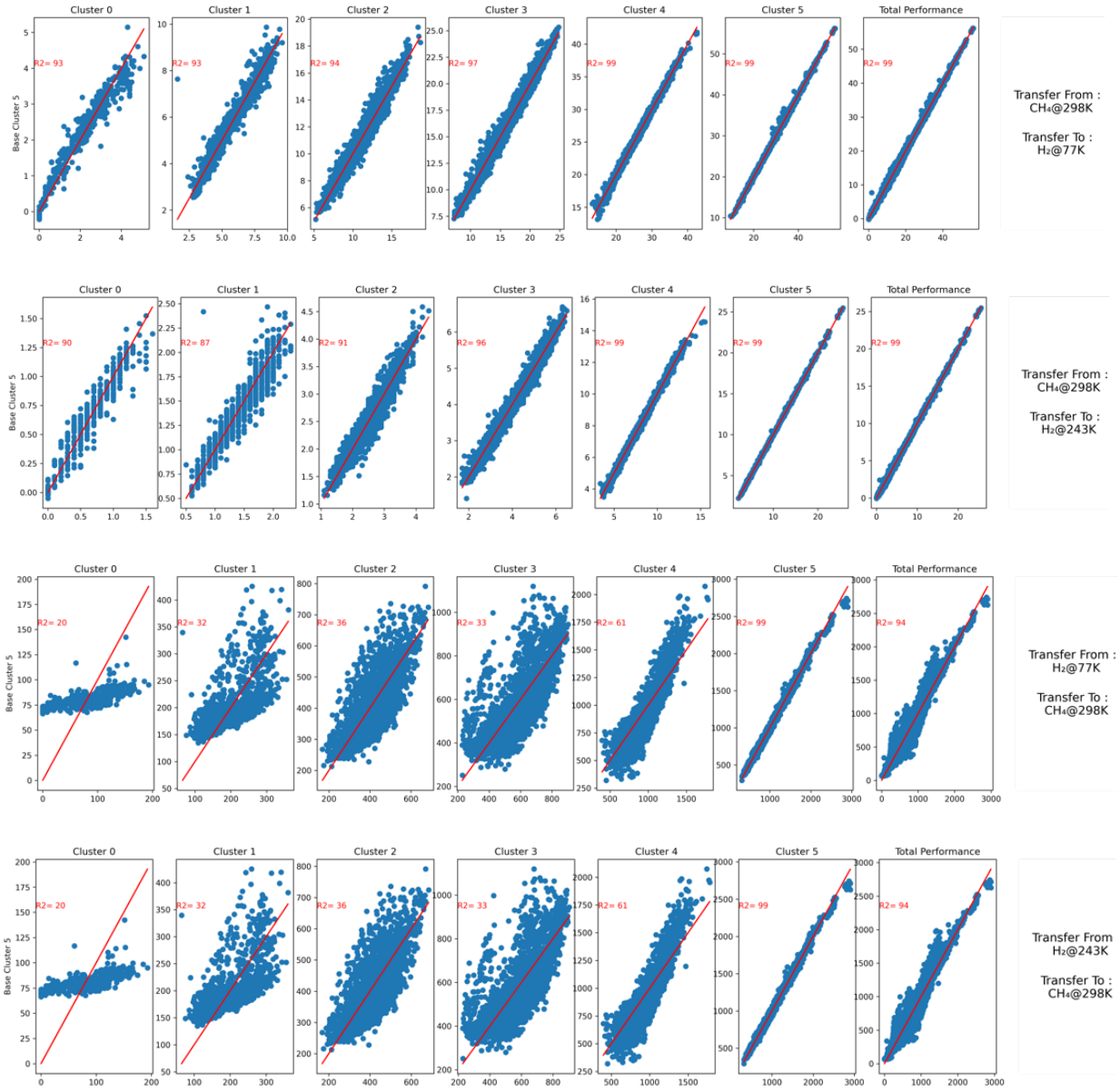

Figure S25. A comparison of transfer learning models using cluster 5. We show the performance with methane and hydrogen at different temperatures. Note that the models performed almost identically showing hydrogen adsorption temperature is not a factor in model performance.

S15. Model analysis for generic cluster adsorption transfer from H2 at 100 bar and 243 K to methane at 298 K

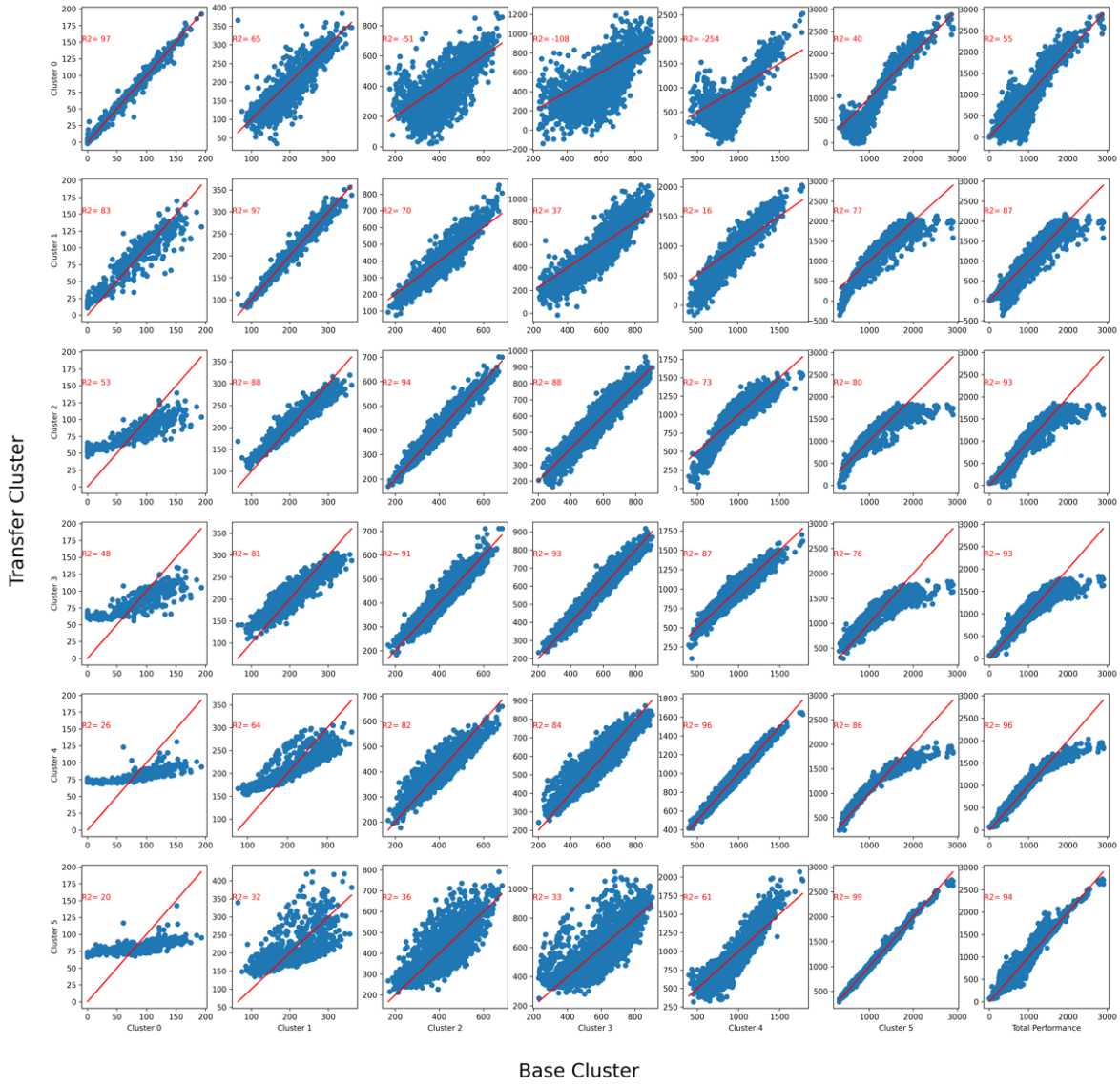

Figure S26. Analysis from H2 at 100 bar and 243 K to methane at 298 K. Note poor model performance with respect to distance of clusters.

S16. Model analysis for generic cluster adsorption transfer from methane at 298 K to H2 at 100 bar and 243 K

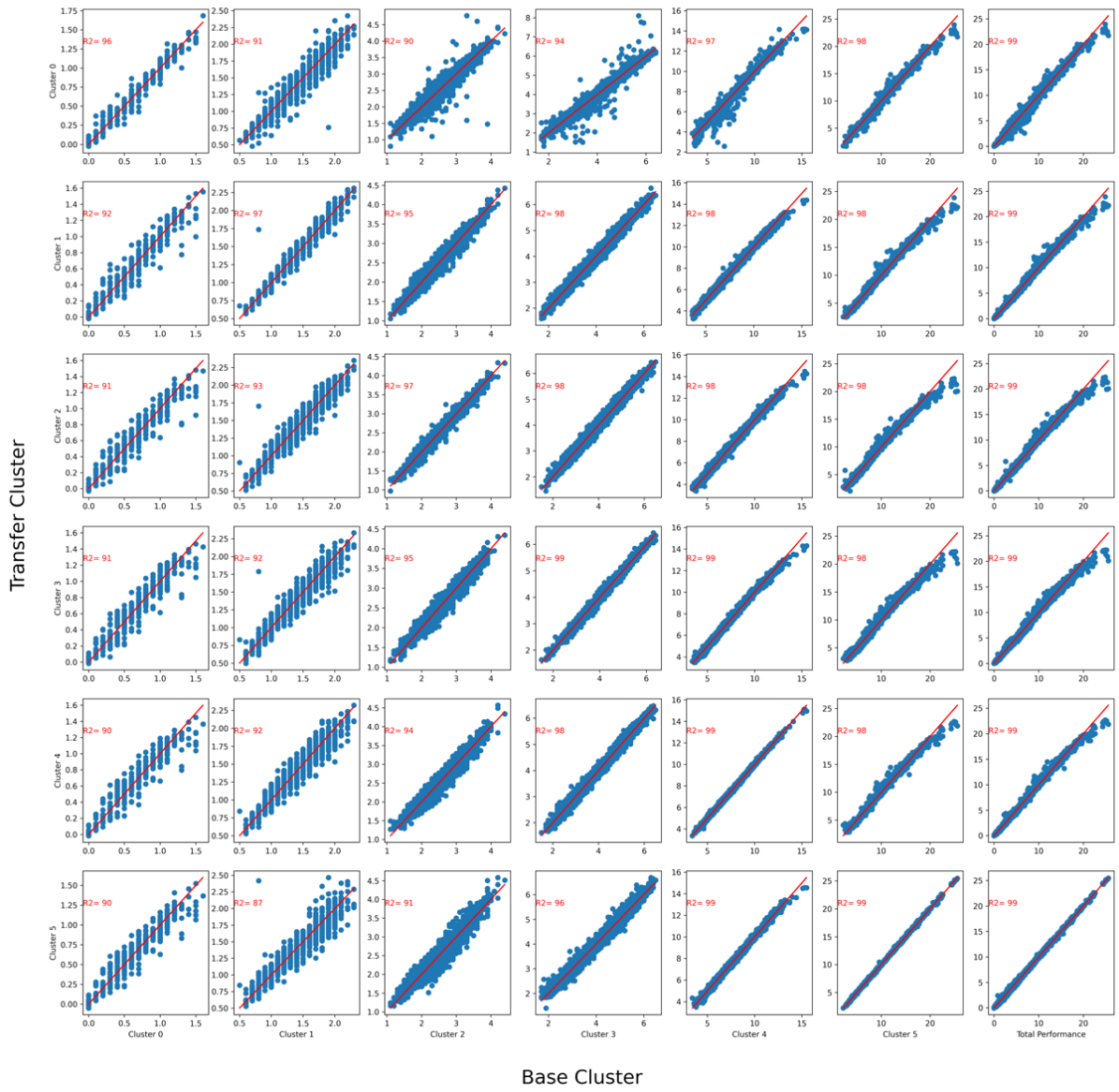

Figure S27. Analysis from methane at 298 K to H2 at 100 bar and 243 K. Note good model performance regardless of distance.

S17. Model residuals analysis from H2 at 100 bar and 243 K to methane at 298 K

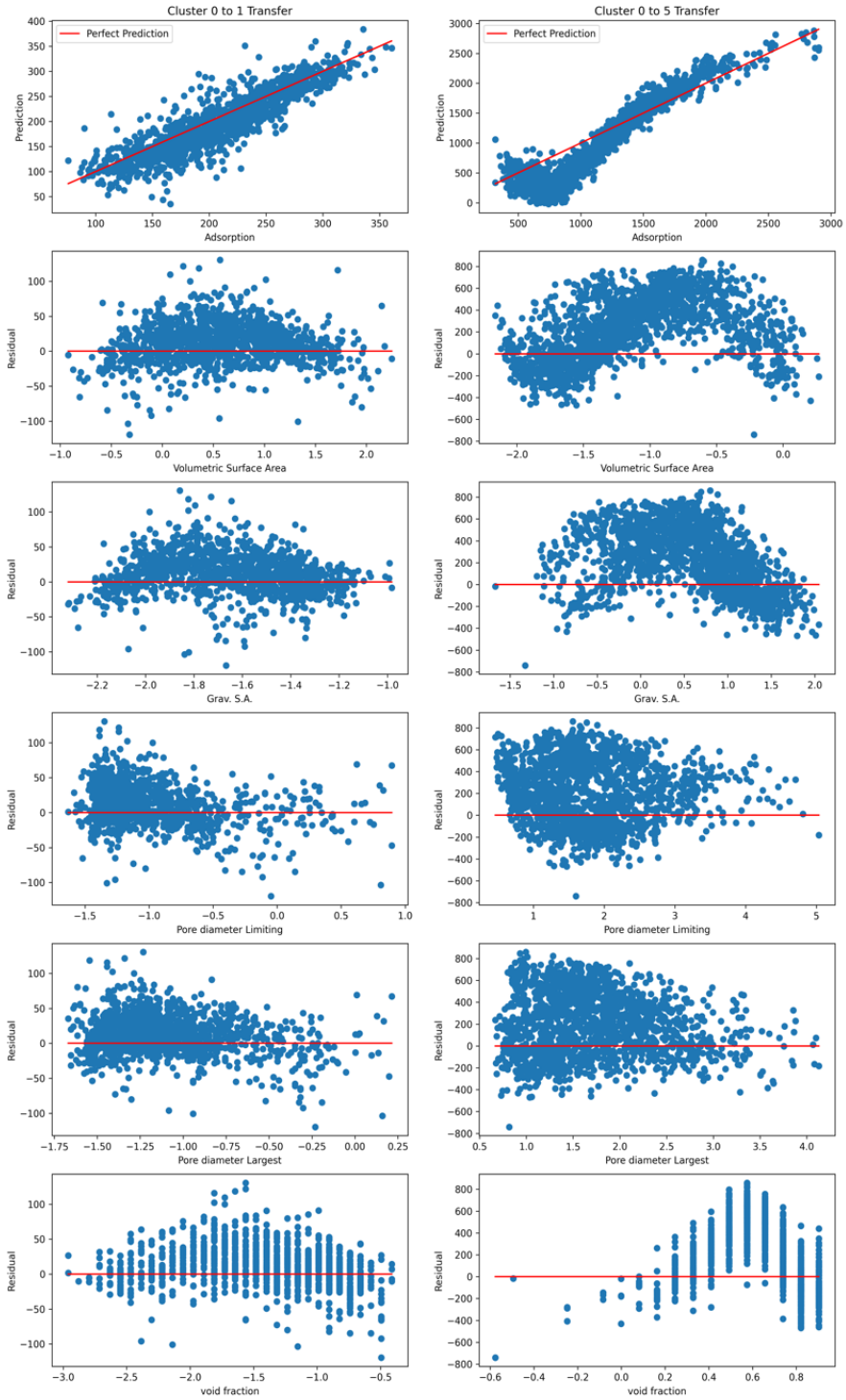

Figure S28. This shows the residuals for transfer from cluster zero to one and five. One is a close cluster, showing good results. Five is a far cluster and the residuals show patterns.

S18. Model space analysis from H2 at 100 bar and 243 K to methane at 298 K

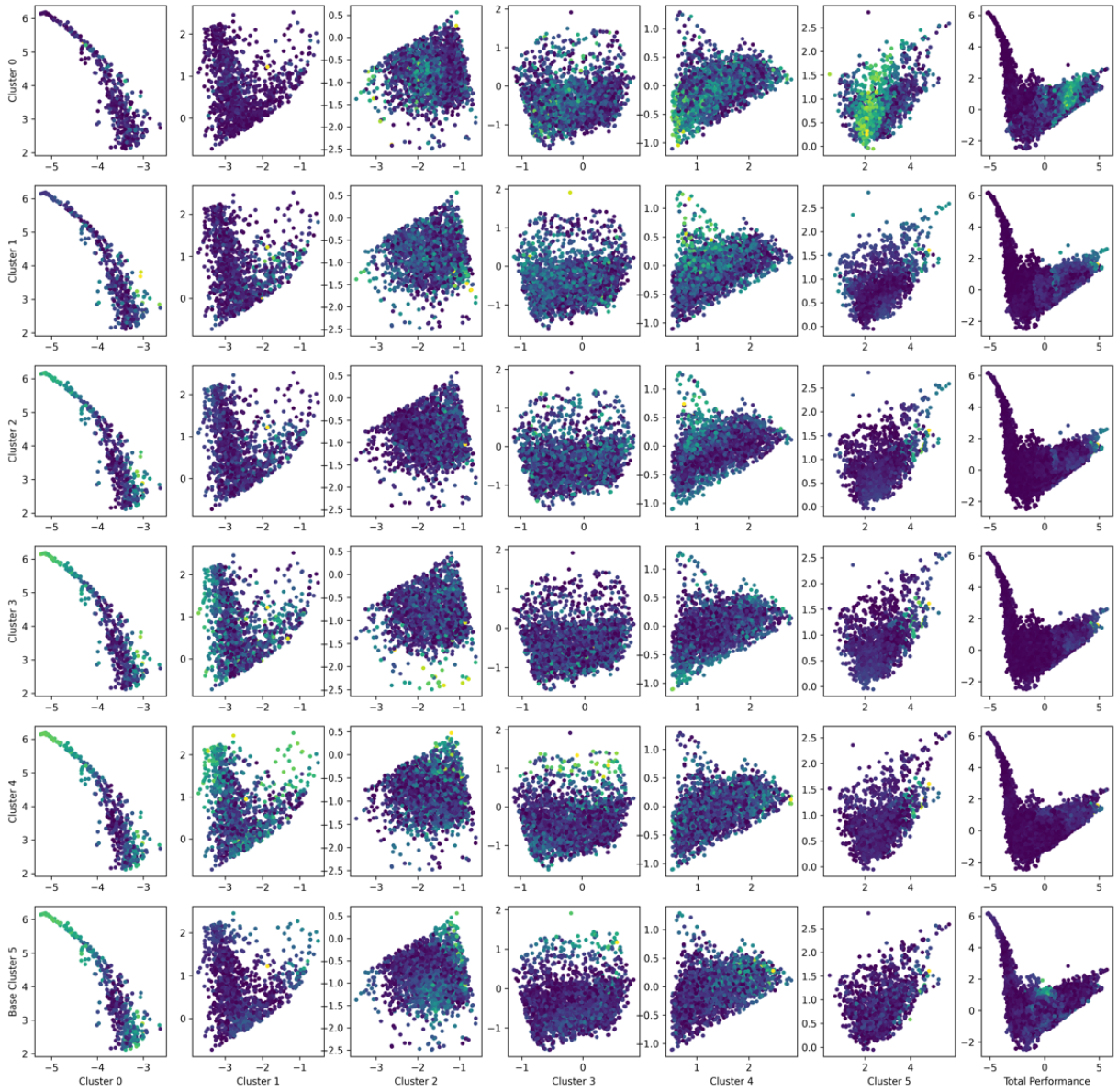

Figure S29. This shows the residuals in terms of the PCA space. Yellow denotes high residuals and dark purple represents no error.