Electronic Supplementary Material (ESI) for Digital Discovery.

Supporting Information for "Molecular Graph Transformer: Stepping Beyond ALIGNN Into Long-Range Interactions"

Marco Anselmi,[a] Greg Slabaugh,[b*] Rachel Crespo-Otero[c*] and Devis Di Tommaso[a*]

[a] *School of Physical and Chemical Sciences, Queen Mary University of London, Mile End Rd, Bethnal Green, London, E1 4NS, United Kingdom.*

[b] *Digital Environment Research Institute, Queen Mary University of London, Empire House, 67-75 New Road, London, E1 1HH, United Kingdom.*

[c] *Department of Chemistry, University of College London, 20 Gordon Street, London WC1H 0AJ.*

Supporting Tables

Model	jdft2d	phonons	dielectric	log10 GVH	log10 KVH	pervoskites	bandgap	formation E	average ranking	ranking
coGN	4	4	4	2	2	1	1	1	2.375	**1**
coNGN	5	2	5	1	1	3	2	2	2.625	**2**
ALIGNN	8	3	13	3	5	2	3	3	5	**3**
MODNet (v0.1.12)	2	5	1	4	3	10	8	10	5.375	**4**
MGT	1	9	3	8	8	7	6	9	6.375	**5**
MODNet (v0.1.10)	3	7	2	5	4	11	9	11	6.5	**6**
SchNet (kgcnn v2.1.0)	7	8	10	7	7	5	11	4	7.375	**7**
DimeNet++ (kgcnn v2.1.0)	12	6	12	6	6	8	5	5	7.5	**8**
MegNet (kgcnn v2.1.0)	15	1	11	9	10	6	4	6	7.75	**9**
Finder_v1.2 structure based version	10	11	7	12	11	4	7	8	8.75	**10**
AMMExpress v2020	6	13	6	10	9	12	13	15	10.5	**11**
Finder_v1.2 composition only version	11	10	8	13	14	16	10	12	11.75	**12**
CGCNN v2019	13	14	15	11	12	9	14	7	11.875	**13**
CrabNet	9	12	9	14	13	14	12	13	12	**14**
RF-SCM/Magpie	14	15	14	15	15	13	15	14	14.375	**15**
Dummy	16	16	16	16	16	15	16	16	15.875	**16**

Table S1 table showing the ranking of each model for each task and the overall ranking for the whole MatBench dataset. The ranking of each model is determined by the Mean Average Error (MAE) obtained for each task with each model. The MAEs are shown in Table 2. The overall ranking is determined by ordering all the models in ascending order based on their average ranking.

Configuration				Test Results			
Number of Encoders	Number of MHA	Number of ALIGNN	Number of GNN	Error	Bandgap Error	HOMO Error	LUMO Error
1	1	1	1	0.2888	0.2865	0.2970	0.2830
			2	0.2828	0.2772	0.2953	0.2758
			3	0.2750	0.2690	0.2924	0.2637
			4	***0.2716***	0.2620	0.2872	0.2657
		2	1	0.2685	0.2669	0.2806	0.2580
			2	0.2617	0.2585	0.2721	0.2545
			3	0.2663	0.2636	0.2788	0.2564
			4	***0.2592***	0.2612	0.2656	0.2507
		3	1	0.2661	0.2599	0.2788	0.2597
			2	***0.2562***	0.2500	0.2684	0.2503
			3	0.2643	0.2604	0.2741	0.2583
			4	0.2661	0.2594	0.2777	0.2613
		4	1	0.2672	0.2637	0.2813	0.2566
			2	0.2658	0.2591	0.2796	0.2588
			3	***0.2641***	0.2619	0.2721	0.2583
			4	0.2657	0.2577	0.2806	0.2589
	2	1	1	0.2880	0.2841	0.3028	0.2770
			2	0.2814	0.2815	0.2921	0.2705
			3	0.2779	0.2750	0.2867	0.2719
			4	***0.2689***	0.2616	0.2803	0.2648
		2	1	0.2706	0.2673	0.2837	0.2610
			2	0.2722	0.2704	0.2841	0.2621
			3	0.2706	0.2661	0.2865	0.2591
			4	***0.2642***	0.2648	0.2739	0.2540
		3	1	0.2705	0.2666	0.2812	0.2635
			2	0.2652	0.2642	0.2768	0.2547
			3	0.2643	0.2570	0.2796	0.2562
			4	***0.2613***	0.2589	0.2713	0.2536
		4	1	0.2671	0.2628	0.2783	0.2601
			2	0.2666	0.2627	0.2746	0.2624
			3	***0.2624***	0.2574	0.2756	0.2541
			4	0.2693	0.2617	0.2843	0.2618

Table S2 (Part 1) table showing all the errors obtained for the ablation study. The configuration side shows the number of modules within the MGT Encoder (Figure 2a) that have been used for each test. The number of encoders has been set to 1 for the ablation study as a result of memory and time efficiency. The value highlighted in green is the lowest error obtained between all the tests. The value highlighted in red is the lowest value obtained while keeping the number of ALIGNN and GNN at 1 and changing only the number of MHA in each test (configurations X-1-1). The values in yellow are the lowest values obtained while keeping the number of MHA constant, the number of GNN at 1 and changing only the number of ALIGNN for each test (configuration C-X-1). Lastly the numbers in bold and italic are the lowest values obtained while keeping the number of MHA and ALIGNN constant and changing only the number of GNN during each test (configurations C-C-X). (The Table continues on the next page)

Configuration				Test Results			
Number of Encoders	Number of MHA	Number of ALIGNN	Number of GNN	Error	Bandgap Error	HOMO Error	LUMO Error
1	3	1	1	0.2865	0.2793	0.3019	0.2782
			2	0.2801	0.2761	0.2901	0.2742
			3	0.2792	0.2702	0.2946	0.2728
			4	***0.2766***	0.2734	0.2874	0.2690
		2	1	0.2724	0.2664	0.2858	0.2649
			2	0.2722	0.2661	0.2862	0.2643
			3	0.2705	0.2722	0.2820	0.2573
			4	***0.2675***	0.2613	0.2842	0.2570
		3	1	0.2750	0.2707	0.2841	0.2701
			2	0.2717	0.2688	0.2812	0.2651
			3	0.2720	0.2645	0.2872	0.2641
			4	***0.2692***	0.2618	0.2816	0.2642
		4	1	0.2664	0.2595	0.2772	0.2626
			2	0.2683	0.2624	0.2790	0.2635
			3	0.2701	0.2666	0.2820	0.2617
			4	***0.2664***	0.2566	0.2792	0.2634
	4	1	1	0.2877	0.2864	0.3032	0.2737
			2	0.2804	0.2777	0.2919	0.2715
			3	0.2762	0.2730	0.2895	0.2660
			4	***0.2735***	0.2799	0.2857	0.2550
		2	1	0.2713	0.2670	0.2825	0.2643
			2	0.2730	0.2732	0.2833	0.2624
			3	0.2734	0.2673	0.2901	0.2627
			4	***0.2685***	0.2720	0.2755	0.2579
		3	1	0.2684	0.2617	0.2812	0.2621
			2	0.2730	0.2660	0.2895	0.2634
			3	0.2758	0.2660	0.2883	0.2730
			4	***0.2655***	0.2577	0.2768	0.2620
		4	1	***0.2678***	0.2626	0.2815	0.2593
			2	0.2704	0.2603	0.2854	0.2653
			3	0.2702	0.2610	0.2823	0.2673
			4	0.2735	0.2677	0.2899	0.2629

Table S2 (Part 2) Continuation of Table S2 (Part 1)

Model	jdft2d	phonons	dielectric	log10 GVRH	log10 KVRH	perovskites	bandgap	formation E
ALIGNN	1.0440	0.0848	0.5286	0.2432	0.1937	0.0509	0.1412	0.0214
AMMExpress v2020	0.9581	0.1613	0.4828	0.2973	0.2207	0.3544	0.2143	0.1715
CGCNN v2019	1.1840	0.1658	0.9177	0.3044	0.2428	0.0799	0.2255	0.0335
coNGN	0.8696	0.0829	0.4816	0.2279	0.1675	0.0513	0.1288	0.0177
coGN	0.8936	0.0853	0.4733	0.2343	0.1825	0.0475	0.1183	0.0169
CrabNet	1.0966	0.1582	0.4957	0.0345	0.2585	0.7185	0.2015	0.0857
DimeNet++ (kgcnn v2.1.0)	1.1787	0.1076	0.5211	0.2694	0.1951	0.0665	0.1512	0.0234
Finder_v1.2 composition only version	1.1531	0.1337	0.4911	0.3388	0.2606	1.1400	0.1751	0.0834
Finder_v1.2 structure based version	1.1092	0.1457	0.4900	0.3095	0.2363	0.0566	0.1664	0.0341
MegNet (kgcnn v1.2.0)	1.3024	0.0826	0.5197	0.2962	0.2278	0.0622	0.1468	0.0250
MODNet (v0.1.12)	0.7980	0.0984	0.4155	0.2486	0.1869	0.1605	0.1669	0.0445
MODNet (v0.1.10)	0.8304	0.1113	0.4552	0.2486	0.1869	0.1605	0.1669	0.0445
RF-SCM/Magpie	1.2032	0.1941	0.6431	0.3537	0.2797	0.4162	0.2619	0.1158
SchNet (kgcnn v2.1.0)	1.0258	0.1119	0.5022	0.2707	0.2012	0.0604	0.1785	0.0217
MGT	0.7555	0.1120	0.4670	0.2857	0.2169	0.0638	0.1628	0.0376
Dummy	1.6177	0.9301	1.2396	0.9969	0.9880	1.0004	1.0071	0.9996

Table S3 Scaled Errors obtained on the MatBench v0.1 dataset. The errors shown are scaled errors obtained by dividing the Mean Average Error (MAE) by the Mean Average Deviation (MAD). The MAE and the MAD are obtained as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - y_i^{pred}\right| \qquad MAD = \frac{1}{N}\sum_{i=1}^{N}|y_i - \bar{y}|$$

where y_i is the ith value in the dataset, y_i^{pred} is the prediction for the ith value, $\bar{y}$ is the average value of the dataset and N is the size of the dataset.

Model	jdft2d	phonons	dielectric	log10 GVRH	log10 KVRH	perovskites	bandgap	formation E
ALIGNN	117.4213 (11)	53.5010 (1)	1.9651 (11)	0.1123 (5)	0.1106 (6)	0.0559 (2)	0.4635 (6)	0.0544 (4)
AMM Express v2020	106.5460 (6)	109.7048 (11)	1.7202 (5)	0.1277 (9)	0.1183 (9)	0.2954 (12)	0.5611 (12)	0.2602 (15)
CGCNN v2019	112.7689 (9)	141.7018 (14)	1.8976 (9)	0.1337 (11)	0.1301 (11)	0.0722 (9)	0.6771 (15)	0.0682 (5)
coNGN	95.4766 (3)	57.1375 (2)	2.0335 (15)	0.1078 (1)	0.1037 (1)	0.0590 (3)	0.4271 (2)	0.0502 (2)
coGN	101.1580 (5)	57.7099 (4)	2.0546 (16)	0.1102 (2)	0.1082 (4)	0.0554 (1)	0.3956 (1)	0.0483 (1)
CrabNet	120.0088 (12)	138.3775 (13)	1.7288 (7)	0.1604 (15)	0.1471 (14)	0.5412 (14)	0.5898 (13)	0.2544 (14)
DimeNet++ (kgcnn v2.1.0)	114.9349 (10)	80.3047 (8)	1.9936 (14)	0.1255 (7)	0.1149 (8)	0.0642 (7)	0.4720 (8)	0.0695 (6)
Finder v1.2 composition only version	120.8819 (14)	94.8514 (10)	1.7189 (4)	0.1572 (14)	0.1491 (15)	0.8831 (16)	0.4837 (9)	0.2537 (13)
Finder v1.2 structure based version	120.0917 (13)	124.0783 (12)	1.7213 (6)	0.1412 (12)	0.1318 (12)	0.0594 (4)	0.4989 (10)	0.1331 (11)
MegNet (kgcnn v1.2.0)	129.3267 (16)	57.4679 (3)	1.9871 (13)	0.1358 (10)	0.1287 (10)	0.0635 (6)	0.4715 (7)	0.0701 (7)
MODNet (v0.1.12)	96.7332 (4)	70.0669 (5)	1.6832 (1)	0.1103 (3)	0.1043 (2)	0.1277 (10)	0.4525 (4)	0.0888 (9)
MODNet (v0.1.10)	92.2288 (2)	78.2220 (7)	1.7185 (3)	0.1103 (4)	0.1043 (3)	0.1277 (11)	0.4525 (5)	0.0888 (10)
RF-SCM/Magpie	112.2660 (8)	146.2764 (15)	1.8538 (8)	0.1540 (13)	0.1454 (13)	0.3346 (13)	0.6125 (14)	0.2419 (12)
SchNet (kgcnn v2.1.0)	111.0187 (7)	76.9279 (6)	1.8990 (10)	0.1260 (8)	0.1143 (7)	0.0599 (5)	0.5172 (11)	0.0529 (3)
MGT	61.8563 (1)	92.8849 (9)	1.7111 (2)	0.1229 (6)	0.1103 (5)	0.0658 (8)	0.4470 (3)	0.0729 (8)
Dummy	126.8446 (15)	492.1533 (16)	1.9728 (12)	0.3716 (16)	0.3693 (16)	0.7424 (15)	1.5989 (16)	1.1631 (16)

Table S4 Root Mean Squared Errors (RMSE) obtained on the MatBench v0.1 dataset, together with the models' ranking for each task. The RMSE is obtained as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^{pred} - y_i)^2}$$

where y_i is the i[th] value in the dataset, y_i^{pred} is the prediction for the i[th] value and N is the size of the dataset. The RMSE because of the squaring of the errors is more sensible to larger errors than the MAE, thus, allowing for an analysis of the performance of the models that consider outliers in the predictions.