Support Information

Identify crystal structures by a new paradigm based on graph theory for building materials big data

Mouyi Weng‡, Zhi Wang‡, Guoyu Qian‡, Yaokun Ye, Zhefeng Chen, Xin Chen, Shisheng Zheng, Feng Pan*

School of Advanced Materials, Peking University, Shenzhen Graduate School, Shenzhen 518055, People's Republic of China.

Corresponding authors: panfeng@pkusz.edu.cn

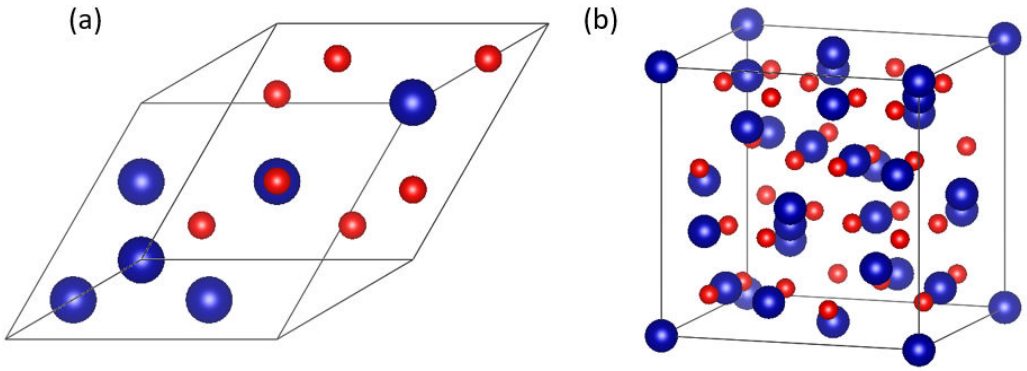

Figure S1. (a) The primitive cell for spinel Co_3O_4. (b) The conventional unit cell for spinel Co_3O_4. Cobalt atoms was plotted in blue ball and oxygen atoms was plotted in red.

Exchange-Correlation Functional	Pseudo-potential	a	b	c	α	β	γ
LDA	FHI	7.98	7.98	7.98	89.71	90.29	89.71
	SG15	7.67	7.67	7.67	89.68	90.32	89.68
	GBRV	7.63	7.63	7.63	89.66	90.34	89.66
LDA+U	FHI	8.22	8.22	8.22	88.87	91.13	88.87
	SG15	7.70	7.70	7.70	89.58	90.44	89.58
	GBRV	8.17	8.17	8.17	89.81	90.19	89.81
PBE	FHI	8.14	8.14	8.14	89.60	90.40	89.60
	SG15	7.80	7.80	7.80	89.66	90.28	90.28
	PD04	7.47	7.47	7.47	88.27	91.73	91.73
	PWM	8.09	8.09	8.09	90.02	89.98	89.98
	GBRV	8.18	8.18	8.18	89.84	90.16	90.16
PBE+U	FHI	8.35	8.35	8.35	88.81	91.19	91.19
	SG15	7.71	7.78	7.78	89.53	90.89	90.89
	PD04	8.49	8.49	8.49	89.27	90.73	90.73
	PWM	8.56	8.56	8.56	89.62	90.38	90.38
	GBRV	7.78	7.78	7.78	89.60	90.40	90.40

Table1. The differences in lattice parameters caused by exchange-correlation functionals and pseudo-potentials were investigated in our work and the results are shown in Table.1. The exchange-correlation functionals include LDA and PBE functionals. The pseudo-potentials include 2 types: Norm-Conserving Pseudopotential(FHI, SG15, PD04, PWM) and Ultrasoft Pseudopotnetial(GBRV). The lattice parameters are indicated as a, b, c and α, β, γ.

The detail of pruning algorithm.

It is well known that, the time complicity of comparing isomorphism of two graphs with n vertices is $O(n^n)$. The comparison of a complex structures is expensive. Here, we used labels on vertices for pruning. Only vertices with same labels can be exchanged during comparison. The labels are

A: the element type,
B: the distance from the center atom,
C: the degree of the vertex (how many neighbor atoms),
D: the sum of A of all neighbor atoms,
E: the sum of B of all neighbor atoms;
F: the sum of C*B of each neighbor atoms.

One may also add more labels on vertices. These labels are useful in accelerate our comparing algorithm.

The detail of Mn-O structural unit in distance 1.

Towards to six coordination, we set up five kinds of geometric structures, which are regular octahedron, distorted regular octahedron, triangular prism, distorted triangular prism and other geometric structures, which account for 40%, 20%, 0%, 20% and 0%, respectively.

Similarly, for the five-coordinate structural unit, we have established the regular quadrangular pyramid, the distorted regular pyramid, the triangular biconical, the distorted triangular bifurcation, and "Others" species; and for four-coordinate structural unit, we have established the plane quadrilateral, the distorted plane quadrilateral, regular tetrahedron, distorted regular tetrahedron and "Others" species, marked as "Geometry type A", " Distorted geometry type A", "Geometry type B", " Distorted geometry type B".

The geometric structures' proportions of the three structural units are shown in the following table:

Table 1 Proportion of different geometries for three structural units (%)

Coordination number	Geometry type A	Distorted geometry type A	Geometry type B	Distorted geometry type B	Others
4	21.72	23.59	5.75	4.29	44.65
5	2.26	1.21	1.26	0.84	94.42
6	40	20	0	20	0

Some statistical results have been selected in Supporting Information, and further analysis would be done in the near future.

A special example: use GT scheme to identify MoS_2 structures.

Using GT scheme can classify different types of MoS_2 structures. Typical MoS2 structures have 2H 1T and 1T' phase. The structures are shown in Figure S2 (a), (b) and (c) respectively.

By using a normal bond length parameter, crystal structures can be converted into graphs shown in (d), (e) and (f). In these structures, 2H and 1T structures can be identified. 1T and 1T' structures are classified into a same type of structure, because the topological connections are the same of these two structures.

However, this problem can be solved by setting special rules. One way is to choose a structure unit as the vertices in graph. We can manually set octahedrons as a vertex and distorted octahedrons not. Then, the graphs of 1T and 1T' phase are different, with 1T phase have a structure unit vertex but 1T' not. The other way is to choose a suitable bond length value. The converted graphs are shown in (g), (h) and (i).

In this example, we showed different ways to convert crystal structures into graphs that come up with different results.

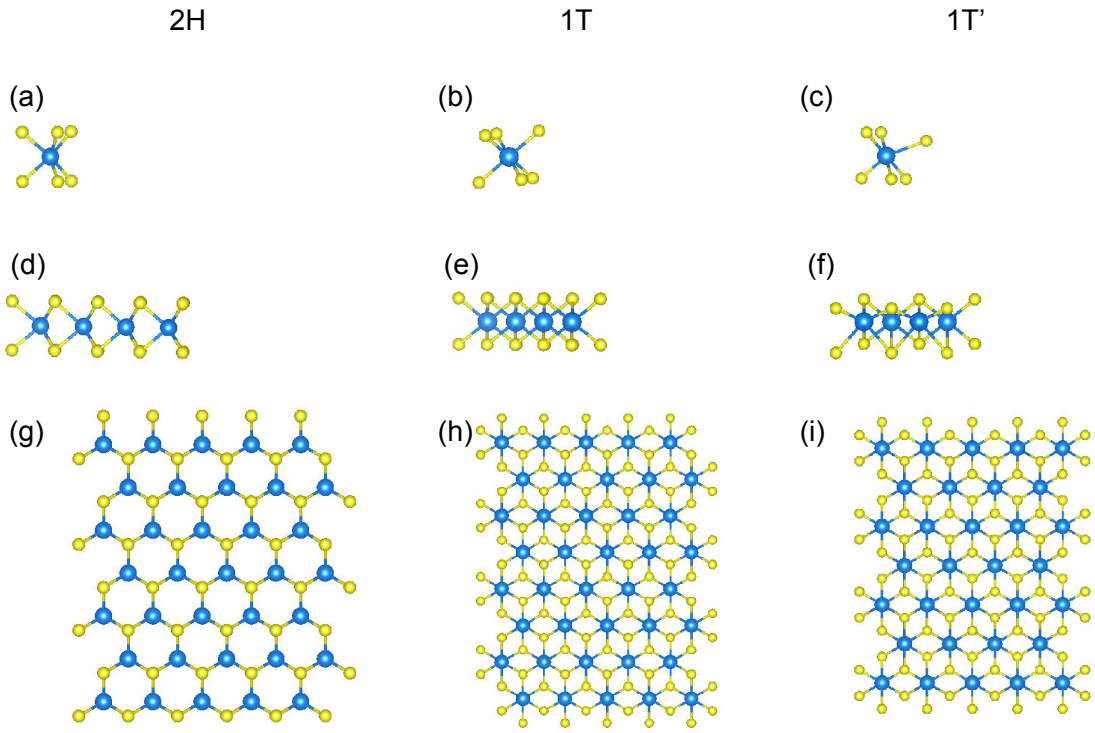

Figure S2. The lattice structures comparison of 2H MoS_2, 1T MoS_2 and 1T' MoS_2. The blue atoms are Mo element and the yellow atoms are S element. (a), (b) and (c) shows the structures of 2H MoS_2, 1T MoS_2 and 1T' MoS_2 respectively. (d), (e) and (f) show the middle pictures of (a),(b) and (c) for 2H MoS2, 1T MoS2 and 1T' MoS2 respectively, which represent the side views. (g), (h) and (i) show the right pictures of three crystal structures respectively, which represent the top views.

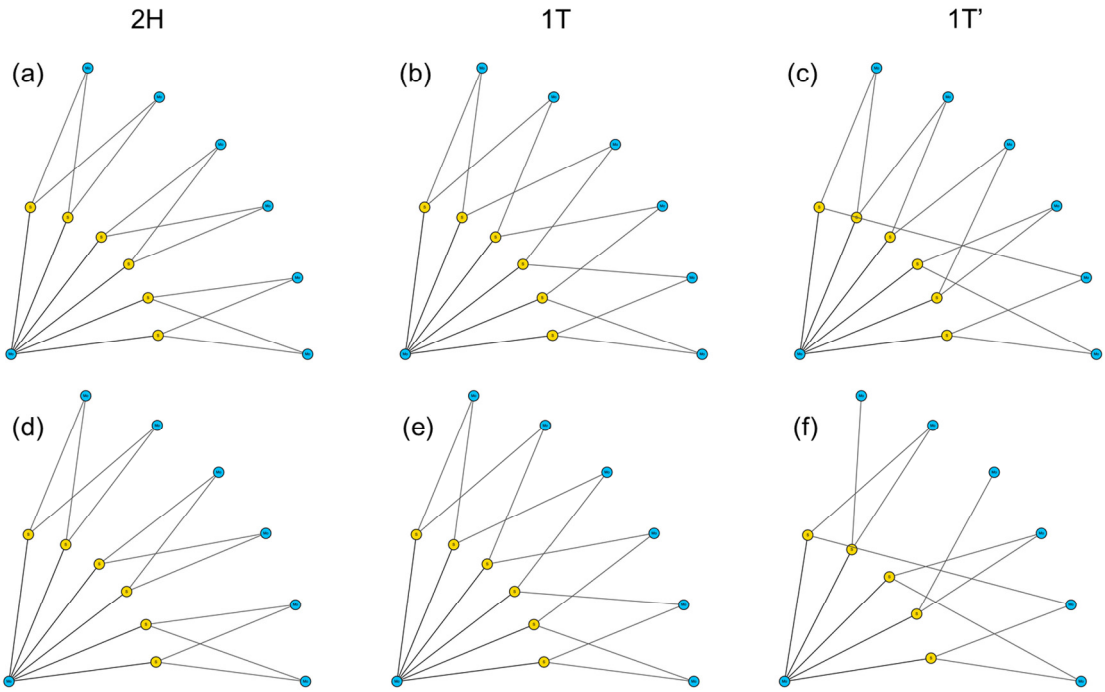

Figure S3. The graphs converted from MoS2 structures. (a), (b) and (c) shows the graphs that converted from 2H MoS_2, 1T MoS_2 and 1T' MoS_2 respectively by using normal bond length parameters. In this kind of graphs, 2H and 1T phase can be distinguished, 1T and 1T' cannot be distinguished. (d), (e) and (f) shows the graphs that converted from 2H MoS_2, 1T MoS_2 and 1T' MoS_2 respectively by using special set bond length parameters. In this kind of graphs, all three 2H, 1T and 1T' phase can be distinguished.