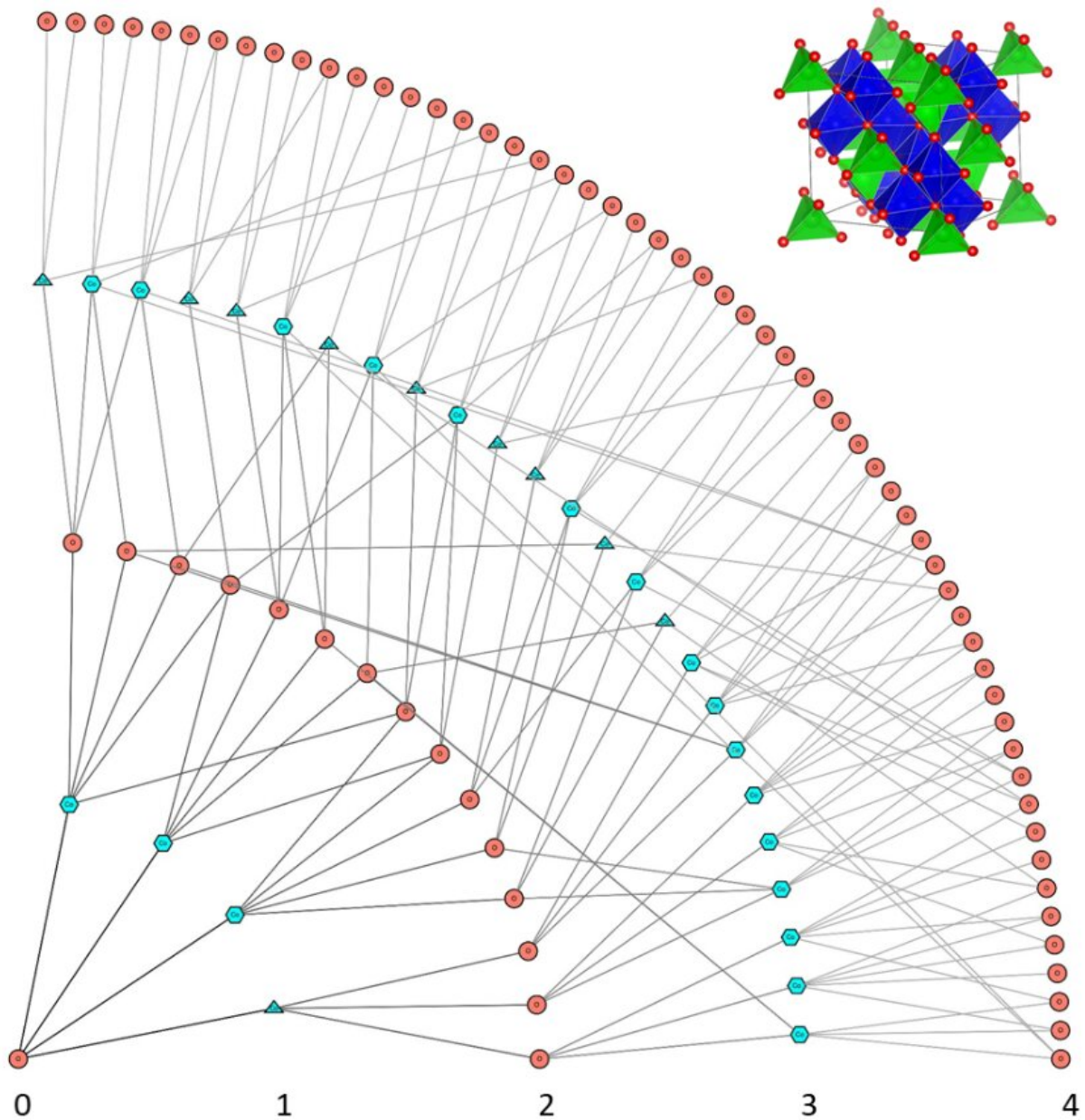# A new paradigm of material identification based on graph theory

June 14 2019

The simplified graph and the actual crystal structure (upper right) of spinel Co3O4. Credit: Science China Press

Materials Genome Initiative (MGI) and National Materials Genome Project have been launched by American and Chinese government in the past decade. One of the major goals of these missions is to facilitate the identification of materials data to speed material discovery and development. Current methods are promising candidates to identify structures effectively, but have limited ability to deal with all structures accurately and automatically in the big materials database, because different material resources and various measurement error lead to variation of bond length and bond angle.

Feng Pan and his colleagues, from Peking Univerisy Shenzhen Graduate School, propose a new paradigm based on graph theory (GT scheme) to improve the efficiency and accuracy of material identification, which focuses on processing the "topological relationship" rather than the value of bond length and bond angle among different structures.

In GT scheme, the researchers first simplify crystal structures into a graph, which only consists of vertices and edges, in which atoms are simplified as vertices and adjacent atoms with the actual chemical bonds are "connected" with edges. If the topological connections in the simplified graphs between two structures are the isomorphic, the GT scheme will consider them as one structure. By using this method, automatic deduplication for big materials database is achieved for the first time, which identifies 626,772 unique structures from 865,458 original structures.

Moreover, the GT scheme has been modified to solve some advanced

problems such as identifying highly distorted structures, distinguishing structures with strong similarity and classifying complex crystal structures in materials big data. Compared with the traditional structure chemistry methods, the GT scheme can address these iusses much more easily, which enhances the efficiency and reliability of material identification.

By using this artificial intelligent technique, the researchers are trying to achieve high-throughput calculation, preparation and detection for the [materials](#) database. The GT scheme subverts the traditional material research methods and accelerates the development in material research field.
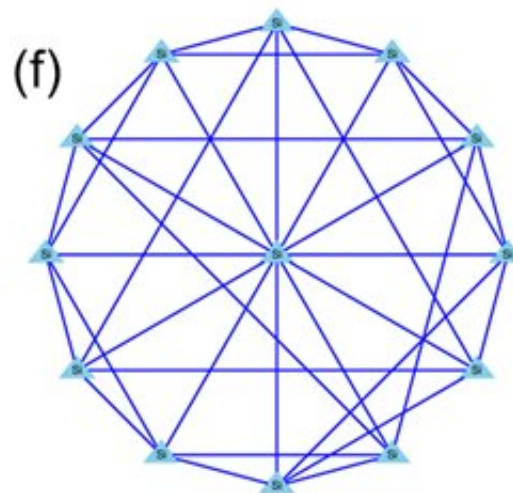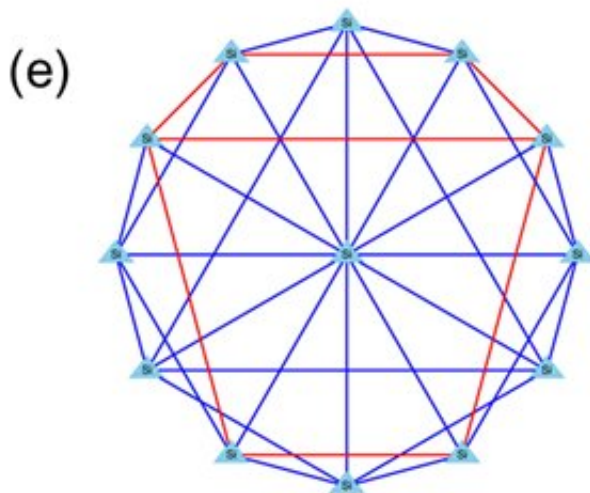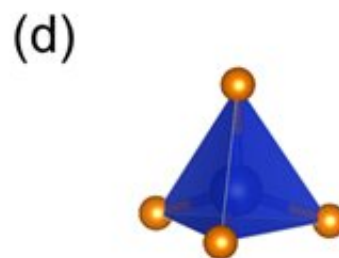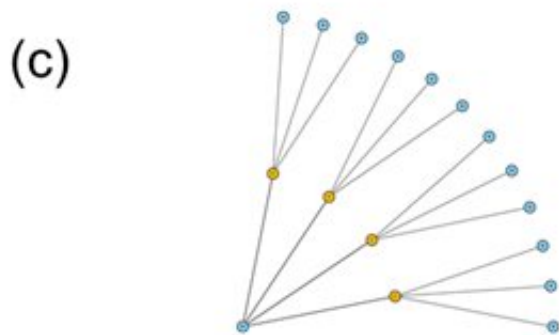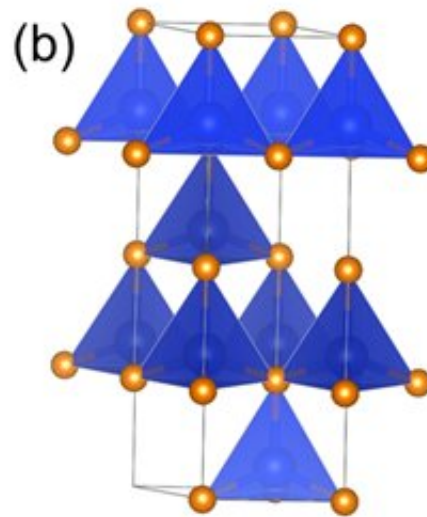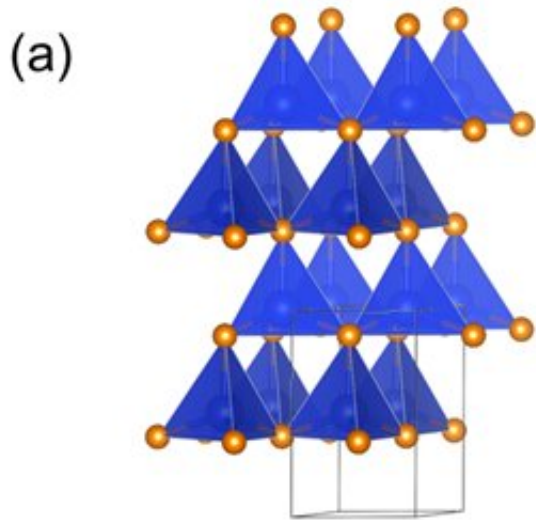
The GT scheme can distinguish 2H phase SiC from 4H phase SiC which has strong similarity. Credit: Science China Press

More information: Mouyi Weng et al, Identify crystal structures by a new paradigm based on graph theory for building materials big data, *Science China Chemistry* (2019). DOI: 10.1007/s11426-019-9502-5

Provided by Science China Press