A Further Improved Structure Matching Algorithm. N. David Browna, James Haestiera, Mustapha Sadkia, Amber L. Thompsona, David J. Watkina, Mustapha Sadkia, Amber L. Thompsona, David J. Watkina. aChemical Crystallography, Inorganic Chemistry Department, University of Oxford. E-mail: dave.brown@chem.ox.ac.uk

Since publication of the classic Ullmann algorithm for subgraph isomorphism [1], a variety of works have successfully augmented the algorithm’s behaviour for graph matching, such as that presented by Cordella et al. [2]. Their paper describes a generic implementation of Ullmann with emphasis on an efficient implementation utilising linear data arrays rather than the matrices Ullmann originally specified.

As part of “Age Concern”, a joint project between the crystallography laboratories in Durham and Oxford, funded by the EPSRC (grant EP/C536282/1), Oxford team members have developed an improved version of the Cordella algorithm, to be used for chemical structure matching. We explore the problem by considering graph and chemical theory in parallel, and allowing tailoring of the algorithm’s behaviour to perform structure matching bespoke to any context via parameters specified prior to execution.

The state tree for any chemical structure matching algorithm is a search space of all possible atom-to-atom mappings between two input structures. Through parameterisation of our algorithm, we are able to winnow choices at every stage of our state search and maximise pruning of the state tree, as much as is possible for any given chemical context, according to the knowledge of the user. Pre-specified similarity conditions for both individual atoms as well as atom environments can be considered during matching, allowing ‘chemically similar’ fragments and molecules to be rapidly identified, as well as those which are ‘structurally similar’. We provide distinct definitions of ‘graph isomorphism’ and ‘chemical isomorphism’ to clarify our proposal.

Additionally, by leveraging the arbitrary mapping of certain single-degree atoms within any input chemical structure, we are able to further improve the efficiency of the original graph matching algorithm [2]. A working demonstration of the algorithm will be available at the poster presentation.


Keywords: crystallographic analysis; computing methods in crystallography; structural similarity