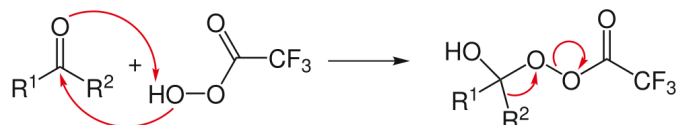


## Praca współautorstwa naszych informatyków ukazała się w Nature Communications

Reakcje chemiczne często przedstawia się w postaci schematów, na których od razu widać co dzieje się z poszczególnymi atomami.



Rys. 1. Przykładowa reakcja chemiczna.

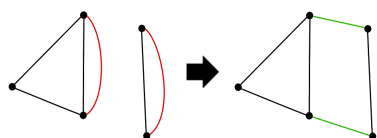
Kiedy taki schemat sprzed i po reakcji jest nieskompilowany, wtedy nawet „na oko” można dostrzec co zaszło w wyniku reakcji. To znaczy: które wiązania zostały zerwane, a gdzie powstały nowe. Często jednak takie schematy są na tyle złożone, że do stwierdzenia *co zaszło* w danej reakcji zaprzęga się komputery.

W pracy *Automatic mapping of atoms across both simple and complex chemical reactions* opisano algorytm, który umożliwia modelowanie reakcji chemicznych, będących do tej pory poza zasięgiem komputerów (np. reakcje pericykliczne i elektrocykliczne oraz przegrupowania sigmatropowe). Autorami pracy są informatycy (pięć osób związanych z naszym Wydziałem, w tym troje studentów) oraz czworo chemików: Wojciech Jaworski<sup>1</sup>, Sara Szymkuć<sup>2</sup>, Barbara Mikulak-Klucznik<sup>2</sup>, Krzysztof Piecuch<sup>1</sup>, Tomasz Klucznik<sup>2</sup>, Michał Kaźmierowski<sup>1</sup>, Jan Rydzewski<sup>1</sup>, Anna Gambin<sup>1</sup> oraz Bartosz A. Grzybowski<sup>2,3,4</sup>. W tym tekście postaramy się przybliżyć przedstawiony w pracy algorytm.

W analizie reakcji chemicznych korzysta się z grafów, którymi przedstawia się schemat cząsteczek chemicznych przed reakcją (*substraty*) i schemat wyniku reakcji (*produkty*). Żeby dowiedzieć się co zaszło **podczas** reakcji, bada się podobieństwo tych grafów. Dokładniej problem sprowadza się do izomorfizmu podgrafów obu schematów, co jest problemem *NP-trudnym*.

Jednakże, grafy przedstawiające cząsteczki mają pewne przyjemne własności, umożliwiające efektywne modelowanie bardzo wielu reakcji. Własności, o których mowa to:

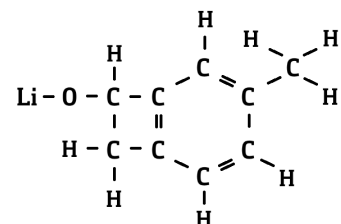
- niskie stopnie wierzchołków grafów (które zresztą zwykle są planarne),
- mała liczba wierzchołków – rozważane grafy mają zazwyczaj do 100 wierzchołków,
- zaledwie kilka spójnych składowych (substraty reakcji tworzą jeden graf, produkty drugi),
- etykietowane wierzchołki – tymi etykietami są nazwy pierwiastków chemicznych: węgiel, tlen, etc.



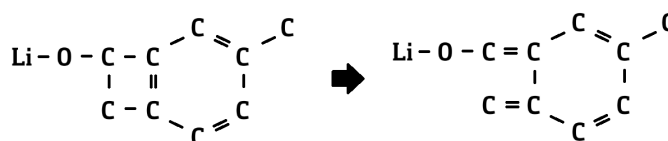
Rys. 2. Lewa strona przedstawia przykładowy graf przed reakcją, prawa po. Jeżeli usuniemy czerwone oraz zielone krawędzie to oba grafy będą izomorficzne. To znalezienie tych właśnie krawędzi jest celem algorytmu.

Zadaniem algorytmu przedstawionego w przytoczonej publikacji jest znalezienie „różnic” między tym co było przed reakcją, a tym co otrzymaliśmy po. Sprowadza się to do znalezienia minimalnego zbioru krawędzi w grafie substratów i grafie produktów, których usunięcie spowoduje, że grafy będą izomorficzne. Poza znalezieniem tych krawędzi algorytm wskazuje także możliwe izomorfizmy.

Pokażemy działanie algorytmu na konkretnym przykładzie. Rozważmy cząsteczkę przedstawioną na rysunku 3. W pierwszej kolejności schemat cząsteczki należy uprościć, pomijając wszystkie atomy wodoru. To samo należy wykonać dla produktu reakcji.

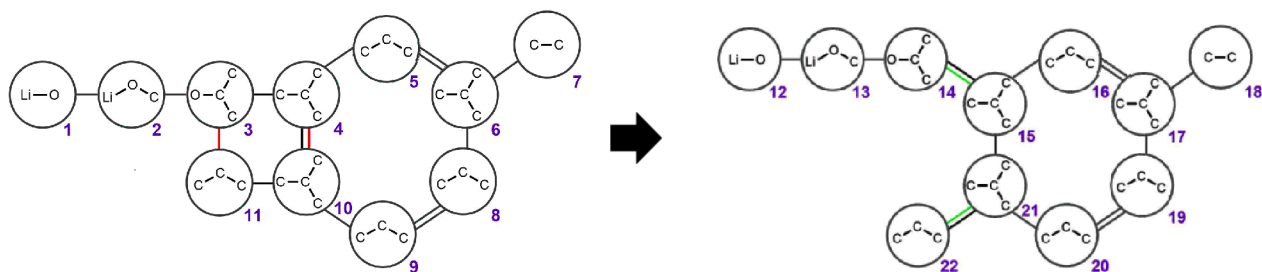


Rys.3. Na startcie algorytm dysponuje takim schematem cząsteczki przed reakcją.



Rys. 4. Uproszczone (tj. bez wodoru) schemat substratu (z lewej) oraz produktu reakcji (z prawej).

Teraz algorytm przechodzi do etykietowania grafu.



Rys. 5. Etykiety utworzone dla każdego wierzchołka oraz jego najbliższych sąsiadów (tj. etykiety stopnia 1-go).

W analogiczny sposób można etykietować wierzchołki oraz ich otoczenie stopnia 4 (tj. w otoczeniu danego wierzchołka znajdują się takie wierzchołki, że ścieżka pomiędzy nimi ma długość co najwyżej 4). Algorytm znajduje izomorfizm podgrafów, właśnie między grafami etykiet stopnia 4-go.

Czynność powtarzana jest dla otoczeń coraz mniejszego stopnia. W efekcie pozostaje niewielka liczba niedopasowanych atomów znajdująca się w *centrum reakcji* (czyli zbiorze atomów, których wiązania zmieniają się w jej trakcie). Te krawędzie, które nie mają odpowiednika po drugiej stronie reakcji należy **usunąć**. A jeśli to nie pozwoli ustalić izomorfizmu między grafami, to usuwane są kolejne krawędzie. Na rysunku 5 na czerwono oraz zielono zaznaczono krawędzie, które należy usunąć.

W pesymistycznym przypadku algorytm działa w *czasie wykładniczym* (czas działania algorytmu zależy w sposób wykładniczy od liczby wierzchołków w grafie). Jednak taka złożoność praktycznie nie występuje w rzeczywistych reakcjach. Ze względu na fakt, że w centrum reakcji zostało niewiele atomów, to można *sprawnie* przejrzeć wszystkie możliwe podziały.

Zaproponowany w pracy algorytm zaopatrzone jest także w zbiór heurystyk obsługujących istotne przypadki niepasujące do ogólnej reguły.

Wersja demonstracyjna algorytmu dostępna jest pod adresem:  
<http://mapper.grzybowskigroup.pl/marvinjs>.

Źródła rysunków:  
rysunek 1 – wikipedia  
File:Baeyer-Villiger-Oxidation-V1.svg;  
rysunek 5 – praca *Automatic mapping of atoms across both simple and complex chemical reactions*.

Publikacja w całości dostępna jest pod adresem: <https://www.nature.com/articles/s41467-019-09440-2>

Wojciech JAWORSKI