URC97018

The Ulam Index: Methods of Theoretical Computer Science Help in Identifying Chemical Substances

Adriana Beltran[1,2] and James Salvador[2]

[1] Center for Theoretical Research
and its Applications in Computer Science (TRACS),
Department of Computer Science, and
[2]Department of Chemistry
University of Texas at El Paso
El Paso, TX 79968
emails[1]abeltran@cs.utep.edu and
[2]james@salvador.chemistry.utep.edu.

Abstract

in this paper, we show how methods developed for solving a theoretical computer problem of graph isomorphism **are used** in structural chemistry. We also discuss potential applications **of these methods to exobiology: the search for life outside Earth.**

1 Identification of chemical substances: why we need it, why it is difficult, and what we are going to do about it

Identification of chemical substances can be reduced to a graph isomorphism problem (well known in theoretical computer science). One of the main problems of chemistry is identification of chemical substances.

In non-organic and organic chemistry, there exist experimental techniques that enable us to describe a *graph* structure of the unknown substance, i.e., to **describe which atoms it consists of, and which of these atoms are** connected by chemical bounds. In order to identify this substance, we must compare it with graphs that describe known substances.

In mathematical terms, we need to check whether an (experimentally obtained) graph is isomorphic to one of the graphs that describe known substances.

Graph isomorphism problem is known to be hard. Unfortunately, the general graph isomorphism problem is known to be hard to solve.

For some substances, different nodes correspond to different types of atoms; in this case, it is relatively easy to check whether a given molecule coincides with this substance, because we can simply identify each atom with a similar atom in the standard substance and then check whether all connections are as in the standard model.

For many other substances. however, atoms of the same type occur in different places of the structure *in* different roles; examples of such substances are organic substances and fullerenes. For these substances, we have to actually solve the difficult graph isomorphism problem.

How to solve this difficult problem: the main idea. One way of solving this problem is based on the following idea:

- It is known that to every graph, wc can assign a polynomial or several polynomials that uniquely determine this graph (i.e., the two tuples of polynomials coincide iff the graphs are isomorphic).
- Thus, to check whether the two graphs arc isomorphic, we can compare the coefficients of the corresponding polynomials.

These methods are widely used in structural chemistry; see, e.g., [8, 1, 7, 9, 10, 2, 3].

We can further compress these polynomials into *numbers* (called *indices*) that also give complete information about the graph [1 I], and compare only these numbers.

A word of warning: index methods are only heuristic. The resulting methods are, of course, only **heuristic** method, because sometimes, due to computer inaccuracy, non-isomorphic substances get erroneously identified.

How frequent are the errors? Since the index methods are purely heuristic, it is important to check how frequently the methods err.

Our numerical experiments show that these errors are extremely rare (and that, therefore, this method works really well) [4]: among all the generated graphs, only 10-5 of them got mis-identified.

2 Some technical details

The index that we use. in this work, we use an index called *Ulam index* because it originated with the ideas presented by S. Ulam in [12].

The Ulam Index is defined (and calculated) as the result of substituting the properly coded structural information of *Ulam Subgraphs* (defined in [12]) into the ***matching polynomial*** of a graph (for a definition of a matching polynomial, see, e.g., [7]; the matching polynomial is a unique and invariant representation of a graph).

We want to substitute some values into the matching polynomial and get an index. To prevent two graphs from having the same index, we differentiate between the variables that correspond to different vertices by counting the number of times that each variable representing a vertex appears in the matching polynomial. This idea is similar to the one used in the definition of the Hosoya's Z index of the graph with that vertex deleted [9].

So, the first natural idea is to use these numbers of times as values of the variables that are substituted into the matching polynomial. This first idea leads to a good index, but, unfortunately, the resulting numbers are too large and cannot be easily represented in the computer.

In order to avoid this problem, before we substitute the weights, we *normalize* them bus dividing each weight by the Hosoya's Z Index of the whole graph (i.e., by the total number of terms **in** the matching polynomial). The result of substituting these normalized weights is what we call an Ulam index.

We have a program that computes the **Ulam index.** We have developed a computer program named GRADE (Graph Recognition Algorithm Developed for Education) that computes the Ulam index. This program is used, in particular, to tutor and test students in chemical nomenclature.

Ulam index is highly discriminating. The *Ulam Index is* a number that uniquely represents a planar graph. This index is ***highly discriminating*** in the sense that usually, non-isomorphic graphs have drastically different values of the Ulam index and therefore, even if we perform computations on real-life computers with computational inaccuracies, the resulting indices typically remain different.

In particular, as our computer experiments show, the Ulam Index differentiates all trees up to 20 vertices (there are 1,346,024 of them) and all graphs up to nine vertices (there are 274,668 of them).

3 Possible applications to space exploration

One of the major tasks for the past and future space missions to planets and other celestial bodes (such as cornets and asteroids) has been to look for life or at least for traces of the former life (see, e.g., [5, 6]). This is especially important now, when traces of life has been found in meteorites coming from Mars.

Automatic robotic missions must be able to analyze the substances that they find on the other planets and identify them.

For this identification, graph isomororphisms algorithms can be of great help.

Acknowledgments. The authors would like to acknowledge the financial support of NSF grant No. CHE-9414968 and of the NASA PACES Center. One of the authors (A. B.) is thankful to Drs. Ann Gates, Vladik Kreinovich, Luc Longpré, and Scott Starks for their valuable **comments and support.**

References

[1] J. Aihara and H. Hosoya, *Bull. Chem.* Sot. *Japan*, 1988, **Vol.** 61, pp. 2657–ff.

[2] R. A. Beezer and E. J. Farrell, "The matching polynomial of a regular graph", *Discrete Mathematics*, 1995, Vol. 137. No. 1, pp. 7–18.

[3] R. A. Beezer, E. J. Farrell, J. Riegsecker, and B. Smith, "Graphs with a minimum number of pairs of independent edges I: Matching polynomials", ***Bulletin of the Institute of Combinatorics and Its Applications, 1996*** (to appear).

[4] A. Beltran and J. M. Salvador, "The Ulam index", ***Abstracts of the Second SC'- COSMIC*** Conference *in Computational* ***Sciences, October 25–27, El Paso, TX,*** Rice University Center for Research on Parallel Computations and University of Texas at El Paso, 1996, p. 6.

[5] K. Biemann, J. Ore, l?. Toulmin, L. E. Orgel, A. O. Nier, D. M. Anderson. P. G. Simmonds, D. Flory, A. V. Diaz, D. R. Rushneck, J. E. Biller, and A. L. Lafleur, "The search for organic substances and inorganic volatile compounds in the surface of Mars", *J. Geophys.* **lies.,** 1977, Vol. 82, **pp.** 4641–4658.

[6] R. R. Christensen, D. L. Anderson, S. C. Chase, R. N. Clark, H. H. Kieffer, M. C. Malin, J. C. Pearl, J. Carpenter, N. Bandeira, F. G. Brown, and S. Silverman, "Thermal emission spectrometer experiment: The Mars Observer mission", ***J. Geophys. Res.***, 1992, Vol. 97, pp. 7719-7734.

[7] I. M. Gutman and V. R. Rosenfeld, "A Novel Approach to Graph Polynomials", ***MATCH,*** 1989, **Vol. 24, pp.** 191-ff.

[8] **H.** Hosoya, *Comp. Math. Appls.*, ***1986, Vol.*** 12B, pp. 271–ff.

[9] H. Hosoya and K. Balasubramanian, "Computational Algorithms for Matching Polynomials of Graphs from the Characteristic Polynomials of Edge-Weighted Graphs", ***Journal of Computational Chemistry, 1989,*** Vol. **10,** No. 5, pp. 698-710.

[10] M. Randic, H. Hosoya, and O. E. Polansky, "On the Construction of the Matching Polynomial for Unbranched Catacondensed Benzenoids", ***Journal of Computational Chemistry, 1989,*** Vol. 10, No. 5, pp. 683--697.

[11] J. M. Salvador, "Topological indices and polynomials: the partial derivatives", ***Abstracts of the 5th International Conference on Mathematical and Computational Chemistry, May 17-21, 1993,*** Kansas City, Missouri, p. 154.

[12] **S. M.** Ulam, ***A Collection of Mathematical Problems,*** John Wiley and Sons. New York, 1960.