

# “Topological Stress” Concept for Quantizing Chemical Graph Similarity Scores

T. Ramraj<sup>1\*</sup>, S. Syed Jaffer<sup>2</sup>, R. Prabhakar<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Coimbatore, India

<sup>2</sup>Department of Chemistry, Coimbatore Institute of Technology, Coimbatore, TamilNadu, India

<sup>3</sup>Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, TamilNadu, India

## Research Article

**Received:** 14-Nov-2023, Manuscript No. JSMS-23-120070; **Editor assigned:** 16-Nov-2023, Pre QC No. JSMS-23-120070 (PQ); **Reviewed:** 30-Nov-2023, QC No. JSMS-23-120070; **Revised:** 17-Jan-2025, Manuscript No. JSMS-23-120070 (R); **Published:** 24-Jan-2025, DOI: 10.4172/JSMS.11.1.001

**\*For Correspondence:** T. Ramraj, Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Coimbatore, India; **E-mail:** ramraj@cit.edu.in

**Citation:** Ramraj T, et al. “Topological Stress” Concept for Quantizing Chemical Graph Similarity Scores. RRJ Stats Math Sci. 2025;11:001.

**Copyright:** © 2025 Ramraj T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Graph invariants can be used to infer molecular structural properties of graphs. In this work, a new distance metric between vertices of a graph is proposed to find the structural similarity between graphs. For this, a new distance matrix (known as IE matrix) is constructed based on the inverse Euclidean distance between graph vertices. This new representation directly depicts the skeletal structure of the vertices in the n-dimensional space. In this space, minimal distance between vertices implies high topological proximity between them. Based on IE matrix, a unique signature is generated for every graph which is used to find similarity between graphs. A new concept “Topological Stress” is introduced that reveals the topological variation of a vertex with other vertices. Using this stress concept, similar topological stressed vertices between different graphs are identified and further explains the methylalkane interconversion reactions with the help of signature and dominant Eigen value. The entirety of the matrix is represented as a singularity which may serve as a unique molecular index ( $IE_G^{Pr}$ ) for chemical graphs. Similarity between graphs are calculated based on their corresponding  $IE_G^{Pr}$  values. The resultant quantified similarity score lies between 0 and 1, referring value 1 for exact isomorphic. From the computational study, it is observed that  $IE_G^{Pr}$  value is generally very low implying a low-value nomenclature of chemical graphs. Empirical study shows that the computational time required for the similarity identification procedure is scalable. From this it can be inferred that by improvising/reframing this measure a few application oriented graph matching problems can be tackled in polynomial time/near-polynomial time.

**Keywords:** Graph isomorphism; Graph signatures; Molecular topological index; Inverse euclidean matrix; Eigen values

## INTRODUCTION

This study emphasises on the fundamental entity of representing any chemical molecule and proposes a new invariant representations of chemical graphs (graph signatures) for finding similarity between graphs. In petro chemical industry, hydrocarbon is processed among which heptane (and its isomers) is one of the petrochemical entities. In this study, seven isomers of heptane are taken for consideration which vary with molecular structural topologies, boiling points and other physical properties. For each heptane isomers, a unique molecular descriptor is derived (graph signature) from which a topological index is framed. It is observed that the obtained signatures and their topological indexes are unique for individual isomers and are graph invariant. These descriptors are in turn used to check for structural similarity between heptane isomers, in polynomial time. Further investigation revealed that the signature not only depicts the structural properties but also embed some interesting features like topological stresses for each vertex which may be useful for solving sub-graph isomorphism problems. The philosophy of this work is to convey, if mathematical properties are coupled with physical properties of chemical graphs then more information can be inferred.

## LITERATURE REVIEW

Generally, mathematical properties are expressed as numbers whereas matrices are most suited to represent chemical structures. Janezic et al., listed out various representation of graph theoretical matrices used in chemical graph theory (Table 1).

**Table 1.** A set of matrices used in chemical graph theory.

Matrices used in chemistry	Brief outline
Path	<p>Matrix elements defined by shortest paths</p> $[P]_{ij} = \begin{cases} \frac{p'(i,j)}{p} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$ <ul style="list-style-type: none"> <li>• <math>p'(i,j)</math>–Total number of paths in the sub-graphs <math>G'</math> obtained by removing the edge <math>e(i,j)</math></li> <li>• <math>p</math>–Total number of paths in <math>G</math></li> </ul>
Wiener (Edge-Wiener matrix)	<p>Matrix elements are given by the product of the number of vertices on each side of path from vertex <math>i</math> to vertex <math>j</math>. A sparse symmetric square matrix (denoted by <math>W^e</math>) introduced mainly for acyclic graphs whose elements are defined by</p> $[W^e]_{ij} = \begin{cases} V_{ie}V_{je} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$ <p>Where  <math>V_{ie}</math> and <math>V_{je}</math>–The number of vertices in two sub-graphs after and edge <math>e(i,j)</math> is removed from the acyclic graph</p>
Random walk	Asymmetrical matrix determined by probability of random walk from vertex $i$ to vertex $j$
Natural Distance (ND) matrix	Distance between vertices in n-dimensional space
Hosoya matrix	<p>It is a sparse symmetric square <math>V \times V</math> matrix (denoted by <math>Z</math>) derived in the similar manner of <math>W</math> whose elements for a tree are defined as:</p> $[Z]_{ij} = \begin{cases} Z_i Z_j & \text{if vertex } i \text{ is adjacent to vertex } j \\ 0 & \text{otherwise} \end{cases}$ <p>Where  <math>Z_i</math> and <math>Z_j</math> - Hosoya <math>Z</math>-indices of the two sub-graphs after an edge <math>e(i,j)</math> is removed from the acyclic graph.</p>

Distance/Distance	<p>The distance/distance matrix<sup>i</sup> (denoted by D/D) unify the topological and topographic (geometric) information on the structure of a graph. It is an asymmetric V x V matrix in which the topographic information-<math>g(i,j)</math> is represented in upper-triangle matrix and graph-theoretical distance-<math>l(i,j)</math> (topological distance) on the lower-triangle matrix.</p> $[D^g/D^l]_{ij} = \begin{cases} g(i,j) & \text{if } i < j \\ 0 & \text{if } i = j \\ l(i,j) & \text{if } i > j \end{cases}$
-------------------	---

Table 1 represents various forms of matrices that are used in chemical graph theory. Generally molecular descriptors obtained from the above mentioned matrices are unique. So structural comparison between chemical molecules is done by comparing the indexes or signatures that are generated from these matrices, which belongs to a subset of exact graph matching problem.

From the computational study of Janezic et al., it is inferred that the classical exact graph matching algorithms employ the concepts of GI and sub-graph isomorphism which requires a bijective node mapping between two (sub) graphs which are computationally expensive (exponential time) as there is a possibility of having  $O(2^n)$  subgraphs for every graph of 'n' vertices. In agreement with that as of now, no efficient algorithm has been found to determine if two general graphs are isomorphic topologically. There are special cases of graphs like triconnected planar graphs, planar graphs, interval graphs, and trees for which GI can be solved in polynomial time. Theoretical proof of Babai, shows that there may exist quasi-polynomial time algorithms that solve GI problem. Mostly chemical graphs may not meet these special cases and therefore we cannot use those algorithms.

In computational chemistry, for dealing with molecular graphs some randomized frameworks have been proposed. In general, randomized techniques are used for fast computation and to enhance simplicity. From the previous literatures, it is observed that randomized algorithms, in some cases, leads to poor accuracy and are closely related to probabilistic analyses of the corresponding (deterministic) computational procedures. Moreover, probabilistic analyses have been used to explain the phenomenon of difficult problems (e.g., NP-hard problems). Some existing frameworks of random graphs that model chemical graphs can be found in the literatures [1]. There is a significant body of work done to representing chemical graphs as linear text strings (canonical label) and computing the structural similarity between two chemical graphs [2]. In line with literatures it is evident that for each chemical molecular graph, if a unique canonical labelling can be computed, then it would imply that efficient polynomial-time algorithms exists for GI problem. From the experimental study of Luks, it is observed for some special cases of chemical graphs, canonical labelling is well suited for solving GI problem, but may fail in some cases (Miyazaki graphs, for example) [3]. A polynomial time algorithm is proposed by Furer et al., for bounded valence graph having computational complexity as  $O(n^{\tau(d)})$ .

J.L Faulon et al., prescribed algorithms that generate unique signatures for molecules and use them as molecular descriptors for comparisons. To solve GI problem, experimental study are conducted by taking alternate ideas like representing graphs as distance matrices (based on their vertex distance from one another) and computing graph similarity by using distance metrics. Klein and DJ Randic, proposed a method that represents a graph as a resistance distance matrix. The distance between vertices in a resistance matrix is measured by electric resistance by assuming each edge carrying a unit resistance. Computational studies based on resistance matrix show that resistance matrix is more a meaningful representation [4-9]. A subset of matrices of Table 1 (say path matrices) introduced in represents the number of sub-graphs of each graph [10]. From the empirical investigation of Plavsic et al., it has been showed that Wiener index and P'/P index obtained from path matrix correlated with the graph-theoretical invariants of cycles and trees.

### Graph invariants as numerical representations

Graph invariants embeds the structural properties of Graphs. Among them, the vertex connectivity index known as Randic connectivity index and Hosoya index are often used in the structure-property-activity modelling: the Randić connectivity index<sup>iii</sup> [11]. The Hosoya index is used mostly for graphical matrices along with other molecular descriptors. de Piero et al. proposed a node-to-node mapping which is based on the powers of graph adjacency matrix [12]. Computational studies conducted by Randic et al., on molecular indexing considered a variation of distance matrix known as line distance matrix which is determined by information given in their first row. As Wiener index (molecular descriptor) is defined generally for

acyclic graphs, a few categories of methods are evolved by constructing matrices like Szeged and revised-Szeged by generalizing Wiener index to consider polycyclic structures.

Based on the findings, Gutman et al., noticed that the Szeged index does not give good regression as it does not properly reflect molecular size dependence. From the experimental study of Klein et al., it is observed that both Szeged and revised Szeged indexes give a measure of how much a graph depart from the bipartite graph. Among various topological indexes (connectivity indexes) of graphs which characterize their topologies, Balaban J index, can be considered as one of the highly discriminative molecular indexes which is simple to compute. From the critical literature review and motivated by the above studies, it is observed that the availability of finding similarity between graph that supports GI problem is quite scarce. Hence this problem has been considered for computational study. The objective of our work is to generate a unique signature for a given graph. Using this signature, a topological molecular index is computed by which similarity between graphs is presented as a quantized value.

### Proposed work

For an adjacency matrix  $A_G$  of a graph  $G$ , solving the Eigen value problem  $A_G x - \lambda x = 0$  is equivalent to identifying the set of vectors, which is used as the coordinate basis for producing the diagonal matrix of  $A_G$ . Here,  $x$  is considered as unknown Eigen vectors and  $\lambda$  are the eigenvalues of  $A_G$  which is referred as the graph spectra. This framework is similar to HMO model of quantum chemistry. Thus,  $A_G$  can be interpreted as a set of vectors which represents graph in the  $|V|$ -dimensional space, ' $|V|$ ' being the total number of vertices of  $G$ . In order to describe the distance between vertices of  $G$  in the  $|V|$ -dimensional space, an attempt is made to propose a new measure of distance in this work. In general, a graph  $G$  having  $|V|$  vertices can be represented by  $|V|$  points in  $|V|$ -dimensional space and will result in symmetric  $|V| \times |V|$  distance matrix. A new matrix (say Inverse Euclidean matrix) is proposed in which each and every row of the matrix represents the inverse of distance from one vertex to all the other vertices of  $G$ . So the new Inverse Euclidean matrix (IE) can be constructed by having  $\frac{1}{\sqrt{\sum_{x=1}^{|V|} (a_{ix} - a_{jx})^2}}$  in the  $(i, j)$  entry of IE matrix. Here the inverse of the distance is taken so as to find the similarity. When the distance is large, the similarity lessens and vice-versa.

$$IE(i, j) = \frac{1}{\sqrt{\sum_{x=1}^{|V|} (a_{ix} - a_{jx})^2}} \quad (2.1)$$

Let the sum of the  $i^{\text{th}}$  row of IE matrix be denoted as:

$$D_i = \sum_{x=1}^{|V|} IE(i, x) \quad (2.2)$$

It represents the inverse of the total distance from the  $i^{\text{th}}$  vertex to all the other vertices of  $G$ . In this work, an attempt is made to describe a new molecular atomic descriptor based on matrix row (or) column sums of the obtained IE matrix. These descriptors may be used to describe the local atomic environments in a molecule. As these descriptors reveal the local molecular structural properties they can be used in identifying the active substructure in molecular mining and Pharmacophore. Hence these molecular descriptors act like unique graph signatures for an individual molecule. Hence the signature of  $G$  can be represented as a  $|V| \times 1$  column vector as shown below:

$$S_G^* = [D_1, D_2, \dots, D_{|V|}]^T \quad (2.3)$$

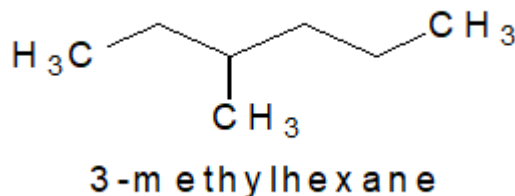
Here  $*$  denotes the set of all vertices taken for consideration. When two graphs  $G_1$  and  $G_2$  are isomorphic, the corresponding points in  $|V|$ -dimensional space will coincide, resulting in same signatures  $S_{G_1}^*$  and  $S_{G_2}^*$  with values shuffled. A set of vertices of  $G_1$  and  $G_2$  share a same value, if and only if they are node-surjective i.e., every vertex  $v_\beta \in G_2$  has at least one matching  $v_\alpha \in G_1$ . It is inferred that any formulated distance function  $d(x, y)$  in a metric space should have the following three properties:

- $d(x, y) = d(y, x) > 0$  if  $x$  not equal to  $y$ —distance is non-negative and does not depend on direction in which it is measured.
- $d(x, y) = 0$  if  $x = y$ —distance is zero only if  $x = y$ .
- $d(x, y) + d(y, z) \geq d(x, z)$ —triangle inequality.

Above three conditions have to be satisfied as the obtained IE matrix has considered the inverse of Euclidean distance between any two vertices. The first condition is satisfied as the IE matrix is a Symmetric Positive Semidefinite Matrix (PSD). All the diagonal elements of IE matrix are zeros as  $d(x, x) = 0$  and hence satisfying the second condition. A few pair of vertices

may share common neighbours resulting in some off-diagonal elements as zeros. Such pairs of vertices are termed as “Pseudo-Similar” vertices and our IE matrix does contain Pseudo-Similar vertices (Figure 1).

**Figure 1.** Topological structure of 3-Methylhexane.

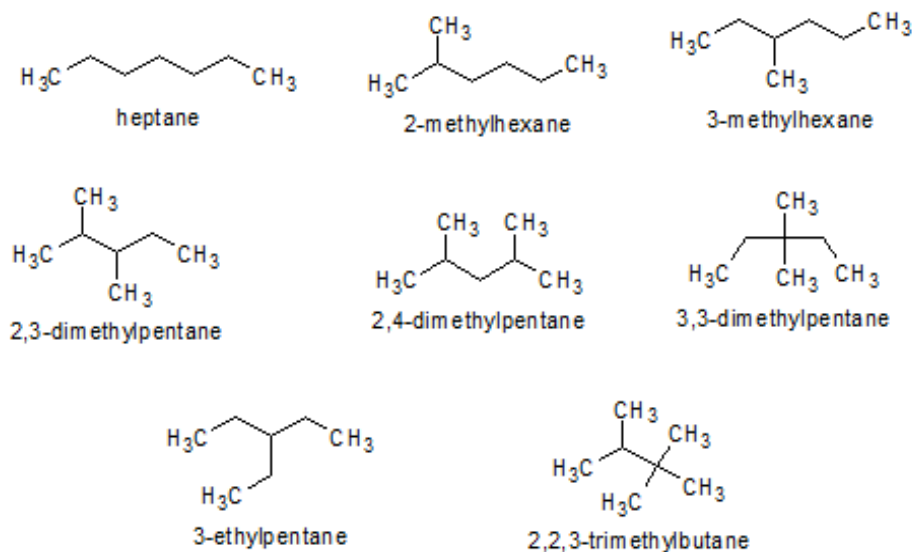


In Figure 2, the IE matrix and the natural distance matrix of 3-methylhexane graph (Figure 1.), obtained based upon Eq. (2.1) are shown. This chemical molecule is shown because as its one of the smallest trees with trivial automorphism group. The isomers of this molecule are represented and discussed in the below sections (Figure 3).

**Figure 2.** IE Matrix of 3-Methyl-hexane (left); the std. distance matrix (D)-distance entries for adjacent vertices (right).

$$\begin{pmatrix} 0 & 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{3} & 0 & 1/\sqrt{5} & 1/\sqrt{2} & 1/2 & 1/\sqrt{3} & 1 \\ 1/\sqrt{2} & 1/\sqrt{5} & 0 & 1/\sqrt{5} & 1/\sqrt{3} & 1/2 & 1/2 \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{5} & 0 & 1/2 & 1 & 1 \\ 1/\sqrt{3} & 1/2 & 1/\sqrt{3} & 1/2 & 0 & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{3} & 1/2 & 1 & 1/\sqrt{3} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 & 1/2 & 1 & 1/\sqrt{3} & 1/\sqrt{2} & 0 \end{pmatrix} \begin{pmatrix} 0 & 1/\sqrt{3} & 0 & 0 & 0 & 0 & 0 \\ 1/\sqrt{3} & 0 & 1/\sqrt{5} & 0 & 0 & 0 & 0 \\ 0 & 1/\sqrt{5} & 0 & 1/\sqrt{5} & 0 & 0 & 1/2 \\ 0 & 0 & 1/\sqrt{5} & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 0 & 0 & 1/\sqrt{3} & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \end{pmatrix}$$

**Figure 3.** Isomers of n-heptane



Devillers and Balaban, suggested to ignore hydrogen atoms in the computation of such indices as organic chemists usually do when they write a benzene ring as a hexagon.

### $IE_G^{Pr}$ and $ARS_G$

In Table 2, the inverse of the row sum for the novel  $IE$  matrices of isomers of heptane (illustrated in Figure 3), is shown. They act like unique signatures ( $S_G^*$ ) (by Equation. 2.2 and 2.3) for the specified heptane-isomers. In order to provide a quantized value for any graph, a unique individual score (called as ARS score) is generated for each graph. The Average Row Sum (ARS) score of any  $S_G^*$  is generated by the below formula:

$$ARS_G = \sum \frac{S_G^*}{|V_G|} \quad (2.4)$$

Moreover, it can also be observed that the (i,j) entry of  $IE$  matrix (calculated based upon Eq. 2.1) is same as that of

$$\frac{1}{\sqrt{\deg(v_i) + \deg(v_j) - \text{adj}(v_i, v_j)}} \quad \forall v_i, v_j \in V_G \quad (2.5)$$

with  $\deg(v_i)$ ,  $\deg(v_j)$  as degrees of vertices  $i$  and  $j$  respectively and  $\text{adj}(v_i, v_j)$  as the number of vertices adjacent to both  $i$  and  $j$ . Along with  $ARS_G$ , a new quantized value known as  $IE_G^{Pr}$  index is computed which may serve as a unique graph (molecular) topological index. This  $IE_G^{Pr}$  index computed based upon Equation. (2.6) is simple, scalable and can be used to compare two graphs directly

$$IE_G^{Pr} = \frac{\sum_{v_i \in V_G} \sum_{v_j \in V_G} IE(v_i, v_j)}{|V_G|} \quad (2.6)$$

From the experimental study (Ref. Section 3), it is found that  $IE_G^{Pr}$  index value is more or less same as that of  $\lambda_1$  and takes less computational time. It is also observed that both  $IE_G^{Pr}$  and  $ARS_G$  scores are proportional to  $\lambda_1$ . These two quantized values ( $IE_G^{Pr}$  and  $ARS_G$ ) may act as molecular descriptors for any graph 'G'.

### Quantizing the similarity between graphs based on $IE_G^{Pr}$ and $ARS_G$

In our graph comparison approach, affine invariance property plays a vital role in comparing graphs. When Graphs are represented in matrix forms, approaches like Affine Invariant Riemannian Metric (AIRM) and Log-Euclidean Riemannian Metric (LERM) may be taken for consideration as both the measures induce a Riemannian geometry; the former induces a curved geometry when the latter "flattens" the manifold by mapping into the tangent space at the identity matrix (which is Euclidean). Arsigny et al., formulated LERM metric which involves in taking logarithm for a matrix<sup>iv</sup> (covariance matrix) that represents a graph. In our case,  $IE_G^{Pr}$  and  $ARS_G$  are single values which act like topological indexes of any graph  $G$ , the LERM metric is reframed to find the similarity score. Putting it altogether, the similarity score between two graphs is quantified as

$$\text{sim}(g_t, g_i) = \begin{cases} 1 & \text{iff } g_t \text{ is isomorphic to } g_i \\ [0,1) & \text{otherwise} \end{cases}$$

The similarity between two graphs  $G_1$  and  $G_2$  can be defined as

$$D_{le}(IE_{g_t}^{Pr}, IE_{g_i}^{Pr}) = |\log IE_{g_t}^{Pr} - \log IE_{g_i}^{Pr}| \quad (2.7)$$

$$\text{sim}(g_t, g_i) = \exp^{-D_{le}(IE_{g_t}^{Pr}, IE_{g_i}^{Pr})} \quad (2.8)$$

Here  $|x|$  denotes the absolute value of  $x$ , it is observed that the row sum of the Euclidean distance matrix for the considered n-heptane isomers along with its Average Row Sum (ARS) are tabulated in the Tables 2 and 3.



**Table 2.** Row sum of the euclidean distance matrix.

Isomers of heptane	1	2	3	4	5	6	7	ARS based on euclidean dist. matrix
H	9.3424	10.8783	10.1463	10.2925	10.1463	10.8783	9.3424	10.1466
2-M	7.8783	12.2042	9.3823	10.1962	10.8464	9.2925	7.8783	9.6683
3-M	9.4388	10.1144	11.6184	9.3823	10.9282	9.2925	8.5605	9.905
2,3-M	7.9747	11.5958	10.9282	10.1643	9.3889	7.9747	8.6569	9.5262
2,4-M	7.8284	12.2361	8.4721	12.2361	7.8284	7.8284	7.8284	9.1797
3,3-M	9.4388	9.3278	12.8352	9.3278	9.4388	7.0645	7.0645	9.2139
2,2,3-M	6.4787	12.8181	10.8884	7.9747	6.4787	6.4787	7.9747	8.4417

H: n-Heptane, 2-M: 2-Methyl Hexane, 3-M: 3-Methyl Hexane, 3-E: 3-Ethyl Pentane, 2,3-M: 2,3-Dimethyl Pentane, 2,4-M: 2,4-Dimethyl Pentane, 3,3-M: 3,3-Dimethyl Pentane and 2,2,3-M: 2,2,3-Trimethyl Butane

**Table 3.** The row sums for the novel IE matrices of n-heptane isomers.

Signatures ( $S_G^*$ ) for isomers of n-heptane							
Heptane and its isomers	1	2	3	4	5	6	7
2,4-M	3.4141	2.9472	4.8947	2.9472	3.4141	3.4141	3.4141
3,3-M	3.8535	4.2698	2.8653	4.2698	3.8535	3.8535	3.8535
3-M	3.8535	3.8095	3.1786	4.2319	3.309	4.0683	4.4923
2,3-M	3.1989	3.1928	3.477	3.7565	3.9062	3.1989	4.329
2-M	3.3613	2.9717	4.2319	3.7327	3.3863	4.0683	3.3613
3-E	3.8535	3.5932	3.4626	3.5932	3.8535	3.5932	3.8535
2,2,3-M	3.1987	3.4993	2.8743	2.5685	2.5685	2.5685	3.1987

From Table 2, it is observed that 2-Methyl hexane is comparatively more similar to 3-methyl hexane, heptane and 2,4-dimethyl pentane. This is due to the fact that the 2<sup>nd</sup> vertex of 2-methyl hexane is having more topological stress (defined below) similar to 2<sup>nd</sup> and 3<sup>rd</sup> vertices of 3-methyl hexane and 2<sup>nd</sup> and 4<sup>th</sup> vertices of 2,4-dimethyl pentane. Similarly, for 2,4-dimethyl pentane the 2<sup>nd</sup> and 4<sup>th</sup> vertices have topological stress as that of 3<sup>rd</sup> vertex of both 3-ethyl pentane and 3,3-dimethyl pentane. It is evident from the tabulated result that as the  $i^{\text{th}}$  vertex ( $V_i$ ) row sum increases the topological stress for  $V_i$  also increases.

From Table 3 specified above, it is inferred that the 2<sup>nd</sup> vertex value of 2-methyl hexane is as lower as that of the 3<sup>rd</sup> vertex of 3-methyl hexane and 2<sup>nd</sup> and 4<sup>th</sup> vertices of 2,4-dimethyl pentane which in turn leads to high similarity scores between them. Similarly, the 2<sup>nd</sup> and 4<sup>th</sup> vertices of 2,4-dimethyl pentane have lesser values same that of the 3<sup>rd</sup> vertex of both 3-ethyl pentane and 3-methyl hexane which in turn leads to high similarity scores between them which is discussed later. It leads to the fact that as the topological stress increases, it gives lower signature value. These similarity scores implicit convey that there exist one or more common subgraphs with in these graphs such that they are isomorphic.

### Graph signatures generating algorithms

Algorithm 1 along with its subroutines 2.1, 2.2 and 2.3 are deployed in C language and Matlab R2011 in a core i3 processor with 4GB internal memory. The parse() function mentioned in Algorithm 1 is deployed in python 2.7.13 to parse gxl files. For experimental study, along with heptane isomer graphs, a portion of benchmark dataset is taken from IAM Graph database.

### Algorithm 1: Graphs similarity score calculation

```

// Checks whether a test graph 'gt' has similarity score (above a threshold) or not in a given Graph Database 'GD'
// It returns 1,  $\forall g_i \in G_D$  which are isomorphic to 'gt' or any other graphs  $g_s \in G_D$  which similarity score with gt is above a given threshold.
Procedure Check_Test_Graph_Presence( $g_t, G_D, \Delta$ )
// gt - test graph
// GD - Graph Database
//  $\Delta$  - Threshold value [0,1]
[
// Parse gt and generate adjacency matrices and corresponding feature vectors
 $[Adj_{g_t}, F_{g_t}^o] = \text{Parse}(g_t)$ ;
 $S_{g_t}^* = \text{Call\_gen\_graph\_signature}(Adj_{g_t}, \text{num\_of\_nodes}(g_t))$ ;
// Import( GD ) - Imports graph data from the Specified Path
// G - A set that stores isomorphic graphs / isomers of gt
 $G = \emptyset$ 
for each  $g_i \in G_D$  do
[
 $[Adj_{g_i}, F_{g_i}^o] = \text{Parse}(g_i)$ ;
 $S_{g_i}^* = \text{Call\_graph\_sign\_gen}(Adj_{g_i}, \text{num\_of\_nodes}(g_i))$ ; //subroutine 2.1
sim_score = Call compute_weight( $S_{g_t}^*, S_{g_i}^*$ ); //subroutine 2.2
if(sim_score >=  $\Delta$ ) then
if ( ! Call Compare_Features( $F_{g_t}^o, F_{g_i}^o$ ) ) then
 $G = G \cup g_i$ ;
//end for
//populate graphs whose similarity score with gt is <=  $\Delta$ 
Populate  $\Delta$  along with their characteristic features;
//end procedure

```

Assuming that all graphs are stored in a Graph database ( $G_D$ ), algorithm 1 computes similarity score between any two graphs. For a given test graph 'g<sub>t</sub>', it scans a graph database  $G_D$  and identifies a set of graphs  $g_i \in G_D$  whose similarity score is equal to or above a particular threshold value. The parse( $g_i$ ) function of algorithm 1, parses the given graph and generates its corresponding adjacency matrix and an 'o' sized vector which reveals the  $g_i$  features. The subroutine 2.1 generates a unique signature for the given graph. The subroutine 2.2 computes the similarity score between given graphs. If the score is greater than a particular threshold it is appended in a set 'G'. The auxiliary function compare\_features( $v_i$ ) compares two feature vectors of any two graphs and can be solved in polynomial time. As this paper focus more on the topological structures of graphs, the feature comparison task is not discussed in detail.

Subroutine 2.1: Generate Graph Signature based on measure of proximity	Sub Routine 2.2: Computes Similarity score based on Log Euclidean Distance
1. Procedure graph_sign_gen ( A, n)	1. // Similarity Measure – Based on Log
2. // A – Adjacency Matrix of size $n \times n \in (0,1)$ matrix	2. //Euclidean Riemannian Distance
3. // n – number of nodes of A	3. Procedure
4. for ( i = 1 .. n) do	4. compute_weight( $S_{g_1}^*, S_{g_2}^*$ )
5. [	5. // $S_{g_1}^*, S_{g_2}^*$ – Graph signatures of $g_1, g_2$
6. for (j = 1..n) do	6. [
7. [	7. // Compute $ARS_{g_1}$ and $ARS_{g_2}$ based on Eq. 2.4
// Compute $IE(i, j)$ based upon Eq. 2.1	8. $ARS_G = \frac{S_G^*}{ V_G }$
$IE(i, j) = \frac{1}{\sqrt{\sum_{x=1}^{ V } (a_{ix} - a_{xj})^2}} \in \mathbb{R}^{n \times 1}$	9. [
8. //end for	10. // Compute $IE_{g_1}^{Pr}$ and $IE_{g_2}^{Pr}$ based on Eq. 2.6
9. //end for	10. $IE_G^{Pr} = \frac{\sum_{v_i \in V_G} \sum_{v_j \in V_G} IE(v_i, v_j)}{ V_G }$
10. for each $v_i \in V_G$ do	11. $dist_{ie} = \log(IE_{g_1}^{Pr}) - \log(IE_{g_2}^{Pr})$
11. [	12. $si_{ie} = e^{-\ dist_{ie}\ _F}$
$D_{v_i} = \sum_{x=1}^{ V } IE(i, x); D_{v_i} \in \mathbb{R}^{1 \times 1}$	13. $dist_{ars} = \log(ARS_{g_1}) - \log(ARS_{g_2})$
12. //end for	14. $score[] = \emptyset$
13. // Generate Signature	15. $score = [dist_{ie}, dist_{ars}]$
14. $S_G^* = [D_1, D_2, \dots, D_n]^T$ ; [where $S_G^* \in \mathbb{R}^{1 \times n}$ ]	16. return score;
15. return $S_G^*$ ;	17. //end procedure
16. ] //end procedure	

The subroutine 2.1 generates signature based on the equations 2.1 and 2.2 with time complexity as  $O(n^2)$ , n being the number of vertices of a graph. It can be proved  $S_G^*$  as graph invariant. The molecular indexes created from the signature is



used to find the similarity score. The execution time, computational complexity and inferences are discussed in the following sub sections. The subroutine 2.2 compute the similarity score based on both  $IE_G^{Pr}$  and  $ARS_G$ . The computational complexity of this subroutine is  $O(1)$ . So the total time taken to identify the similarity score between graphs is  $O(n^2)$ . Depending upon graph database ( $G_D$ ) size the computational time increases yielding  $O(|G_D|.n^2)$  where  $n$  being the maximum number of vertices present in a large graph  $g_i \in G_D$ .

It is also necessary to identify a subgraph within two graphs which are isomorphic. Let us consider  $S_i$  be a subgraph of  $G_i$ . We are interested in finding out  $S_1, S_2$  subgraphs of  $G_1$  and  $G_2$  respectively such that  $S_1$  and  $S_2$  are isomorphic. Initially, it is necessary to find out a vertex say  $v_1 \in S_1$  that maps to  $v_2 \in S_2$  (one-to-one correspondence). This process is repeated until all the vertices of  $S_1$  uniquely map to all the vertices of  $S_2$ . To map a set of vertices between two graphs a new term is introduced known as “Topological Stress” (TS). The topological stress of each vertex is defined as

$$TS(v_i) = \frac{\alpha}{S_G^*(v_i)} ; \in \mathbb{R}^{1 \times 1} \tag{2.9}$$

Where,

- $TS(v_i)$  – Topological stress pertaining to vertex  $v_i$  of a graph  $G$ .
- $S_G^*(v_i)$  – Signature value at the  $v_i$  component of  $G$  (i.e.  $D_{v_i}$ )
- $\alpha \in \mathbb{R}^{1 \times 1}$ , an arbitrary weight<sup>vi</sup> between  $[0,1]$ .

So based upon the topological stress values, a vertex,  $v_i \in G_1$  can be mapped to  $v_j$  of  $G_2$  by which a permutation matrix can be framed in linear time-  $O(n)$ [49] for GI problem. It is also observed that similar topological stressed vertices shows similarity not only in their structure but also in their properties (Ref. subsection 3.1). Below is an algorithm that finds one or more matching vertices of any graph for the give  $v_i \in G_i$  that exhibit similar topological stress values between different graphs.

**Sub routine 2.3: Identifying similar vertices between graph signatures**

```
// Find_Similar_Vertices – This procedure finds similar vertices // of different heptane isomers graphs which have similar and minimal topological stress.
1. Procedure find_similar_vertices( S*, UB)
2. // S* - set of signatures of all heptane isomers
3. // UB – Upper Bound i.e., tolerate level
4. // S*g1, S*g2 – signatures of Graph g1 and g2
5. [
6. for each S*gi ∈ S* do
7. [
//find minimal topological stress of S*gi
x = find_min(S*gi);
for all S*gj ∈ S* \ {S*gi} do
[
//set that stores similar vertices of x
Y = ∅;
//find minimal topological stress of S*gj
y = find_min(S*gj);
min_threshold = |x-y| + UB;
for every vertex v ∈ S*gj do
[
if( |x- S*gj(v)| ≤ min_threshold) then
Y = Y ∪ v;
end if
]//end for
populate Y;
]//end for
]//end for
8. ]//end procedure
```

Subroutine 2.3 populates a set of pairs of vertices of different graphs that have more topological stress (i.e., low score in their corresponding signatures). The time complexity of this subroutine is  $O(\frac{(|S^*| \cdot |S^*|)}{2} \cdot |v|)$  where  $|S^*|$  the total number of signatures taken for consideration. In this computational study,  $|S^*|$  value is number of heptane isomers and  $|v|$  is 7. Table 4 shows the ratio of topological stresses of each and every vertex against their high topological stress vertices for

each and every heptane isomers. These stress values are not influenced by the degree of the graph. Instead, it is highly influenced by the degree of topological variation of each vertex with others, in the sense, how far one vertex is topologically varied with the other vertex.

Table 4. Ratio of topological stresses among various vertices of heptane isomers.

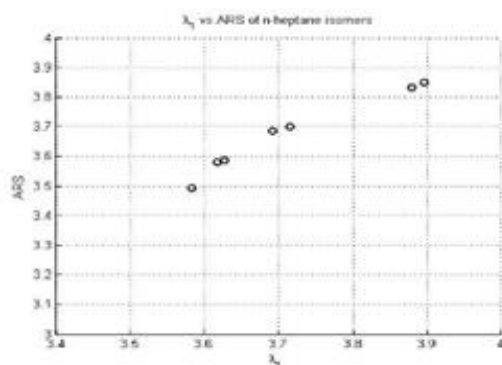
Heptane isomers	Ratio of topological stress among various vertices						
	1	2	3	4	5	6	7
2-M	1.1311	1	1.4241	1.2561	1.1395	1.369	1.1311
3-E	1.1129	1.0377	1	1.0377	1.1129	1.0377	1.1129
3-M	1.2123	1.1985	1	1.3314	1.041	1.2799	1.4133
2,3-M	1.0019	1	1.089	1.1766	1.2234	1.0019	1.3559
2,4-M	1.1584	1	1.6608	1	1.1584	1.1584	1.1584
2,2,3-M	1.2454	1.3624	1.1191	1	1	1	1.2454
H	1.1948	1	1.1261	1.0617	1.1261	1	1.1948

From Table 4, the high topological stressed vertices have value 1 whereas other have value greater than 1. For example, in case of 2-Methylhexane, as vertex 2 is highly stressed its value becomes 1, whereas for vertex 3, the value is 1.4241 giving an inference that vertex 3 is approximately 40% less stressed than vertex 2 for 2-Methyl hexane meaning they have 60% common neighbourhood.

## RESULTS AND DISCUSSIONS

Figure 4, shows the plotting of dominating Eigen values ( $\lambda_1$ ) obtained from IE matrices of n-heptane isomers against ARSs. The observation here is  $\lambda_1$  is directly proportional to ARS score of the corresponding  $S_G^*$ .

Figure 4. Plotting of  $\lambda_1$  against the ARS scores of n-heptane isomers.



From this it can be inferred that ARS score of  $S_G^*$  are more close to  $\lambda_1$  and hence it exhibits some statistical significance and fine tuning this metric may serve as a simple structural interpretation. Also ARS score is proportional to the average signature vector elements. Moreover,  $IE_G^{Pr}$  score is also same as that of  $\lambda_1$  (Table 5). Besides using  $\lambda_1$  as a molecular descriptor (in some cases), from this empirical study, it is suggested to use the two descriptors defined by IE matrix namely ARS and  $IE_G^{Pr}$  (listed for the isomers of heptane in column 1 and 4 of Table 3) can be used. It is also shown by Hoyasa index, the sum of the matrix elements above the main diagonal of the standard distance matrix D is termed as the Wiener number. This leads to suggest ARS and/or  $IE_G^{Pr}$  score(s) can be used as a supportive substitute for the generalized Wiener number for heptane isomers as they are close in magnitude with  $\lambda_1$ . Computational Studies are carried out for other isomers viz., octane and nonane which shows satisfactory results.

**Table 5.**  $\lambda_1$  of IE matrices for n-heptane isomers graphs along with  $IE_G^{Pr}$  and  $ARS_G$  are listed out.

n-heptane isomers	$ARS_G$ score	$\lambda_1$	Balaban's J index	$IE_G^{Pr}$
2,2,3-M	2.9252	2.9526	3.5412	3.7669
2,4-M	3.4922	3.5823	2.9532	3.6993
3,3-M	3.8313	3.8793	3.3606	3.7441
3-M	3.849	3.896	2.832	3.6936
2,3-M	3.5799	3.6168	3.1542	3.7399
2-M	3.5879	3.6266	2.6783	3.6716
3-E	3.6861	3.6914	2.9922	3.7107
H	3.699	3.7155	2.4474	3.6473

Table 5 tabulates  $\lambda_1$  for n-heptane isomers graphs along with  $IE_G^{Pr}$  and  $ARS_G$ . From the literature it can be inferred Eigen values of a graph characterize graph's topological structure and the collection of all eigenvalues of G is denoted as the graph spectra<sup>vii</sup> (spectrum of G). Among various topological indexes which serves as the molecular descriptors of molecules (connectivity indexes), Balaban's J index is taken for consideration as it is one of the highly discriminative molecular index<sup>[13]</sup>. The third column shows Balaban's J indexes which is based on the row sums of the standard distance matrix D. It can be noticed that  $IE_G^{Pr}$  has a value equal to  $\lambda_1$  and is somewhat very nearer to Balaban's J index. Moreover, the computational time for calculating the circuit rank of any graph G in Balaban's index is high as it is necessary to find all connected components in G taking the  $O(V+E)$  time complexity.

Using the Equation (2.8), the similarity score calculated based on  $IE_G^{Pr}$  and  $ARS_G$  indexes of various isomers of heptane are tabulated in Tables 6 and 7.

**Table 7.** Similarity score between graphs based on  $IE_G^{Pr}$ .

Isomers	2-M	3-E	3-M	2,3-M	2,4-M	3,3-M	2,2,3-M	H
2-M	1	0.9895	0.994	0.9817	0.9925	0.9806	0.9747	0.9934
3-E	0.9895	1	0.9954	0.9922	0.9969	0.9911	0.9851	0.9829
3-M	0.994	0.9954	1	0.9876	0.9985	0.9865	0.9805	0.9875
2,3-M	0.9817	0.9922	0.9876	1	0.9891	0.9989	0.9928	0.9752
2,4-M	0.9925	0.9969	0.9985	0.9891	1	0.988	0.9821	0.9859
3,3-M	0.9806	0.9911	0.9865	0.9989	0.988	1	0.9939	0.9741
2,2,3-M	0.9747	0.9851	0.9805	0.9928	0.9821	0.9939	1	0.9682
H	0.9934	0.9829	0.9875	0.9752	0.9859	0.9741	0.9682	1

**Table 8.** Similarity score between graphs based on  $ARS_G$ .

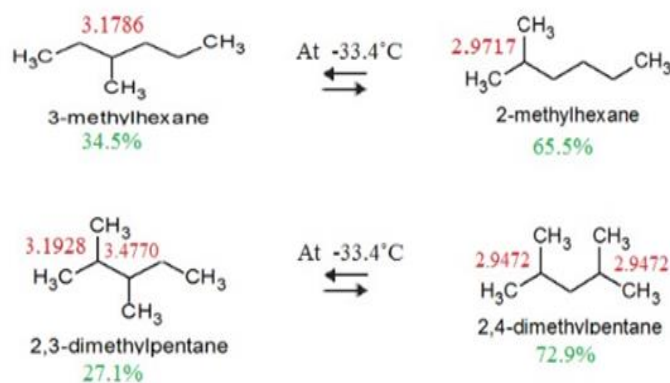
Isomers	2-M	3-E	3-M	2,3-M	2,4-M	3,3-M	2,2,3-M	H
2-M	1	0.9621	0.9327	0.9989	0.9622	0.938	0.8145	0.9621
3-E	0.9621	1	0.9695	0.9611	0.9257	0.975	0.7837	1
3-M	0.9327	0.9695	1	0.9317	0.8974	0.9943	0.7597	0.9695
2,3-M	0.9989	0.9611	0.9317	1	0.9632	0.937	0.8154	0.9611
2,4-M	0.9622	0.9257	0.8974	0.9632	1	0.9026	0.8466	0.9257
3,3-M	0.938	0.975	0.9943	0.937	0.9026	1	0.7641	0.975
2,2,3-M	0.8145	0.7837	0.7597	0.8154	0.8466	0.7641	1	0.7837
H	0.9621	1	0.9695	0.9611	0.9257	0.975	0.7837	1

From Table 8, it is observed that the similarity scores based on  $IE_G^{Pr}$  of 2-methylhexane, 3-methylhexane is 0.9940 with the second largest score of between 2-methyl hexane and 3-ethylpentane 0.9895; Similarly, the score between 2,3-dimethyl and 2,4-dimethylpentane is 0.9755, highest among the scores when 2,4-dimethylpentane is compared with other isomers [14,15]. It is evident from the above results that the  $IE_G^{Pr}$  score increases when two molecules have similar skeletal structures. Moreover, similar results referred in Table 7 show that the ARS similarity score results are less accurate compared to  $IE_G^{Pr}$  score results.

### Topological stress in interconversion of methylalkanes

From the experimental study of Maury et al., it is inferred that the interconversion of methylalkanes like 3-methylhexane and 2-methylhexane can be done using sulfuric and chlorosulfonic acids (Figure 5).

Figure 5. Methylalkanes interconversion reactions.



At  $-33.4^{\circ}\text{C}$ , the proportion of 2-methylhexane and 3-methylhexane are 65.5% and 34.5% respectively. From the signatures of the isomers (refer Table.4) the topological stresses of 2<sup>nd</sup> vertex of 2-methylhexane and 3<sup>rd</sup> vertex of 3-methylhexane are comparatively higher than other vertices, thus leading low values 2.9717 and 3.1786 respectively. Similarly, by analysing the signature of 2,4-dimethyl pentane, the 2<sup>nd</sup> and 4<sup>th</sup> vertices have more topological stresses yielding low values of 2.9472 and 2.9472 respectively. Also the signature of 2,3-dimethylpentane shows more topological stresses in 2<sup>nd</sup> vertex (3.1928) and the 3<sup>rd</sup> vertex (3.4770). It can be noted, at  $-33.4^{\circ}\text{C}$ , the proportions of 2,4 and 2,3-dimethyl pentanes are 72.9% and 27.1%. In case of both the interconversion reactions of 2-methylhexane/3-methylhexane and 2,4-dimethylpentane/2,3-dimethylpentane possessing secondary carbon atoms with similar environment. By using these molecular descriptors, we can differentiate each atomic environment which is the usefulness of subgraph mapping. It is observed that the percentage of conversion is correlated with  $\lambda_1$  obtained from the IE matrix. If  $\lambda_1$  score is less, the conversion percentage is high and vice-versa. For example, in our studies, it confirms that the interconversion reaction between 2-methylhexane and 3-methylhexane gives Eigen value 3.6266 and 3.8960 respectively. So it can be inferred that the signatures of graphs not only act like molecular descriptors but also embed topological structures along with the change of chain branching that occur during interconversion reaction of methylalkanes reversible processes which will open up new gate way for further investigations.

### Impact of signatures in complete graphs

The graph isomorphism can be solved efficiently for almost all graphs having problems only when considering special family of graphs of bounded degree. Besides this, most algorithms of GI have significant problems with the graphs originated from projective planes. The good candidates belong to this family are strongly regular graphs. In this section, applying our method we try to distinguish a few special types of hard graphs by showing some variation in  $S_G^*$  and  $IE_G^{Pr}$  score.

Table 8 tabulates signatures of complete graphs ( $K_{|V|}$ ) with  $|V|$  vertices. Here all the vertices of all graphs are dominating vertices. It is inferred from the generated signatures that, though, for example,  $K_4$  and  $K_5$  have dominant vertices, their topological stress values are different. From this, it is evident that two complete graphs never map to the same signatures

which is one of the major hurdles in graph isomorphism checking. This signature generation is the first representation of its kind, to the best of our knowledge [16].

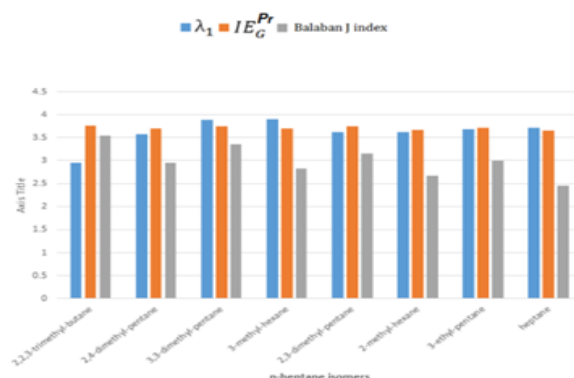
**Table 8.** Complete graphs signatures

K v	Signature										IE <sub>G</sub> <sup>Pr</sup>	ARS
K <sub>3</sub>	0.7071	0.7071	0.7071	-	-	-	-	-	-	-	1.3027	0.7071
K <sub>4</sub>	0.4714	0.4714	0.4714	0.4714	-	-	-	-	-	-	1.394	0.4714
K <sub>5</sub>	0.3536	0.3536	0.3536	0.3536	0.3536	-	-	-	-	-	1.4947	0.3536
K <sub>6</sub>	0.2828	0.2828	0.2828	0.2828	0.2828	0.2828	-	-	-	-	1.5945	0.2828
K <sub>7</sub>	0.2357	0.2357	0.2357	0.2357	0.2357	0.2357	0.2357	-	-	-	1.6909	0.2357
K <sub>8</sub>	0.202	0.202	0.202	0.202	0.202	0.202	0.202	0.202	-	-	1.7835	0.202
K <sub>9</sub>	0.1768	0.1768	0.1768	0.1768	0.1768	0.1768	0.1768	0.1768	0.1768	-	1.8722	0.1768
K <sub>10</sub>	0.1571	0.1571	0.1571	0.1571	0.1571	0.1571	0.1571	0.1571	0.1571	0.1571	1.9575	0.1571

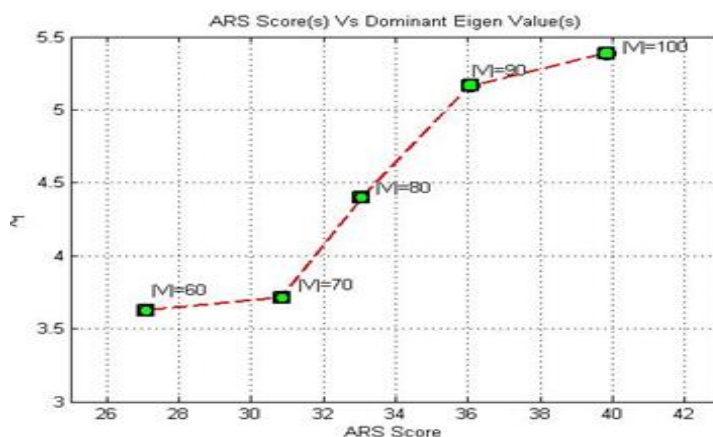
### IE<sub>G</sub><sup>Pr</sup> and ARS<sub>G</sub> scores for large graphs

From the above Figure 4, it can be seen that the IE<sub>G</sub><sup>Pr</sup> index value and  $\lambda_1$  score are relatively equal except for 2,2,3-trimethyl-butane. Moreover, the computational complexity of  $\lambda_1$  is  $O(|V|^3)$  whereas for IE<sub>G</sub><sup>Pr</sup> can be calculated in  $O(|V|^2)$  time. Figures 6 and 7, depicts the plotting of ARS scores of large graphs (sizes varying between 60 and 100) against  $\lambda_1$ . It is observed that as  $|V|$  of graph increases the ARS score also increases [17].

**Figure 6.** Comparison of  $\lambda_1$ , IE<sub>G</sub><sup>Pr</sup> index and the Balaban J index of n-heptane isomers.



**Figure 7.** Plotting of ARS Score against  $\lambda_1$  for large graphs ( $|V|$  ranging between 60 to 100).



From Figure 8, it is evident that, the execution time taken to compute leading Eigen values increase with increase in graph size and, whereas the  $IE_G^{Pr}$  computational time increases very slowly irrespective of graph size and comparatively lower than the other two. Moreover the computational time of finding  $\lambda_1$  is  $O(|V|^3)$  whereas ARS's computational complexity is  $O(|V|^2)$  which is evident from this empirical study.

**Figure 8.** Time taken for computing  $\lambda_1$ , ARS and  $IE_G^{Pr}$  for large graphs.

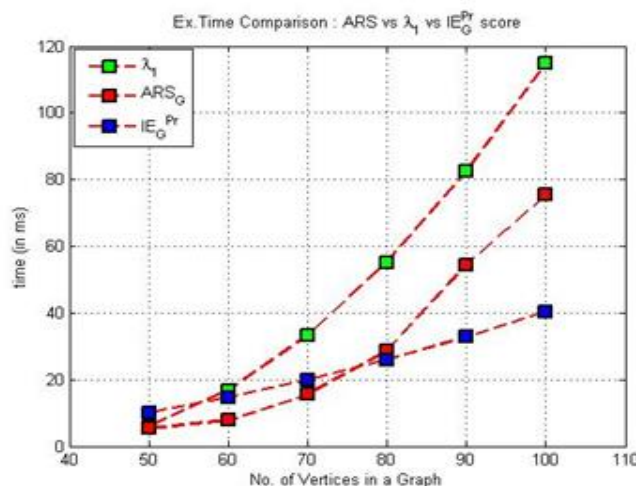


Figure 9 depicts the time taken to generate signatures for large graphs. It is observed that as the  $|V|$  size increases, the time taken also increases but in few milliseconds. This execution time can further be reduced by optimizing the working code. These promising results in execution time calculation give inference in applying these graph signatures in large graph databases mining [18].

**Figure 9.** Computation time taken for large graph signature generation.

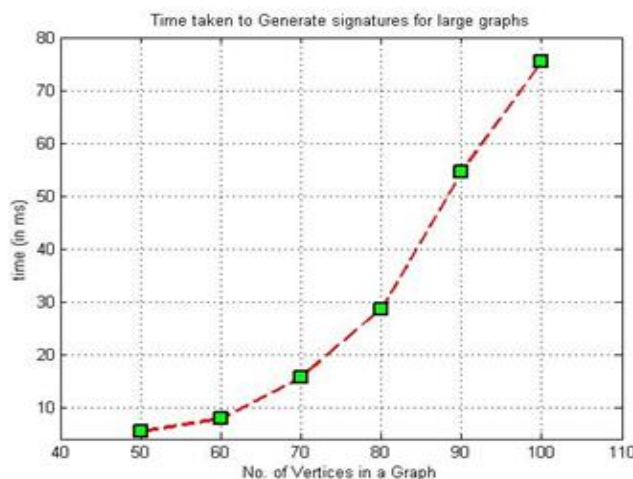
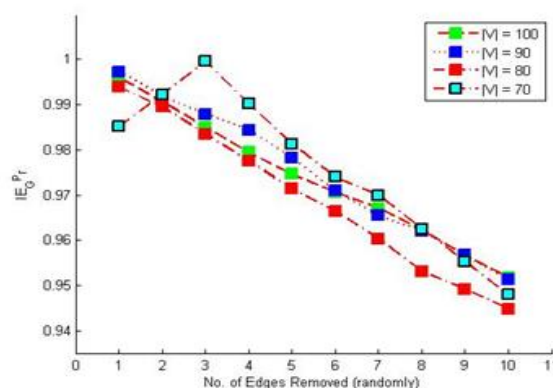


Figure 10 depicts the variations in average similarity score of large random graphs ranging  $70 \leq |V| \leq 100$ . It is observed that as number of edges removed is more, the similarity score between graphs lessen. Also there are some peaks observed. This may be due to the fact that there are some highly influential edges which may act as bridges between two or more connected components [19]. By removing those edges, the similarity value decreases drastically which leads to an assumption that the edges are also topologically stressed and finding highly stressed edges from signatures (and/or other meaningful representations) opens new direction of research study that are more useful in network analysis [20].



**Figure 10.** Variation in avg. similarity score based on removal of edges of large graphs.



## CONCLUSION

In this work, graphs are represented as objects in  $n$ -dimensional space. Based on the distance between vertices, a new distance matrix, which is referred as IE matrix is constructed. This matrix is framed by considering the inverse of the distance between every pair of vertices in the graph as the distance between vertices is large, the similarity between them lessens and vice-versa. This representation directly represents the internal fixed geometry of the vertices of graphs in the  $n$ -dimensional space. In this space, if two vertices are near, then they are very similar. Based on this matrix, new topological index values ( $IE_G^{Pr}$  and ARS) are generated for every graph. It is observed from the empirical study that among these values,  $IE_G^{Pr}$  is more meaningful and may be used as index values of molecules in computational chemistry. Moreover, based on  $IE_G^{Pr}$ , attempts are made to tackle GI problem. Empirical study shows that finding  $IE_G^{Pr}$  and identifying similarity score between graphs are simple and scalable, in the sense, they can be computed in linear time in number of edges. The obtained  $IE_G^{Pr}$  is compared with a state-of-the-art topological index which shows some improvisation in running time and accuracy. Experimental evaluations demonstrate the computational time taken to calculate  $IE_G^{Pr}$  for large graphs is very minimal (in msec) which provides a right balance between computational tractability and expressiveness of graph representation. Also, from signatures, attempts are made to identify topologically highly stressed vertices of graphs. It is observed that signatures embed topological structures along with the change of chain branching that occur during many chemical processes, like interconversion reaction of methylalkanes and the percentage of conversion is highly correlated with  $\lambda_1$ . Reliant on this  $IE_G^{Pr}$  index and the similarity score, it is recommended to imbibe other factors like bond angles between atoms, molecular density, and columnar mass density to give more expressiveness to molecular graphs by which more inferences about any chemical molecule can be made and tackling GI problem in polynomial time.

## REFERENCES

1. Barabasi AL, et al. Emergence of scaling in random networks. Science. 1999;286:509-512.
2. Kouri TM, et al. Random models and analyses for chemical graphs. Int J Found Comput Sci. 2015;26:269-291.
3. Luks EM, et al. Isomorphism of graphs of bounded valence can be tested in polynomial time. J Comput Syst Sci. 1982;25:42-65.
4. Klein DJ, et al. A theorem on Wiener-type invariants for isometric subgraphs of hypercubes. J Math Chem. 1998;17:575.
5. Ivanciuc O, et al. Intrinsic graph distances compared to Euclidean distances for correspondent graph embedding. MATCH Commun Math Comput Chem. 2001;44:251-278.
6. Palacios JL, et al. Sum rules for hitting times of Markov chains. Int J Quantum Chem. 2001;81:29.
7. Xiao X, et al. Comparison between Kirchoff index and the Laplacian energy like invariant, MATCH Commun Math Comput Chem. 2003.
8. Zhang H, et al. Resistance distance and Kirchhoff index in circulant graphs. Int J Quantum Chem. 2007;107:330-339.
9. Chen H, et al. Resistance distance and the normalized Laplacian spectrum. Discrete Appl Math. 2007;155:654-661.
10. Peliillo M, et al. Replicator equations, maximal cliques, and graph isomorphism. Neural Comput. 1999;11:1933-1955.
11. Randic M, et al. Characterization of molecular branching. J Am Chem Soc. 1975;97:6609-6615.

12. Depiero F, et al. Graph matching using a direct classification of node attendance. *Pattern Recognit.* 1996;29:1031-1048.
13. Balaban AT, et al. Highly discriminating distance-based topological index. *Chem Phys Lett.* 1982;89:399-404.
14. Randic M, et al. On representation of DNA by line distance matrix. *J Math Chem.* 2008;43:674-692.
15. Randic M, et al. Dense graphs and sparse matrices. *J Chem Inf Comput Sci.* 1997;37:1078-1081.
16. Maury LG, et al. Equilibrium in the System 2-and 3-Methylhexane and in the System 2, 3-and 2, 4-Dimethylpentane. *J Am Chem Soc.* 1954;76:5831-5833.
17. Hosoya H, et al. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull Chem Soc Japan.* 1971;44:2332-2339.
18. Arsigny V, et al. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magn Reson Med.* 2006;56:411-421.
19. Klein DJ, et al. A theorem on Wiener-type invariants for isometric subgraphs of hypercubes. *J Math Chem.* 1998;17:575.
20. Faulon JL, et al. The signature molecular descriptor. 4. Canonizing molecules using extended valence sequences. *J Chem Inf Comput Sci.* 2004;44:427-436.