The pharmaceutical companies need to continuously discover and develop new drugs. Three approaches to drug discovery namely random screening, structural modification of lead compounds and rational drug design have been utilized together by most of the companies in order to accelerate the discovery of new drug molecules. The huge amount of data furnished by combinatorial chemistry and high throughput screening has updated an "old" rational technique used since the early 1960s for prediction of chemical properties: the quantitative structure-property or structure-activity relationships (QSPR or QSAR) (Gozalbes, 2002). The two classical approaches to QSAR are as follows:

(1) Techniques based on the recognition of molecular features (fragments, groups or sites) and calculation (generally by regression analysis) of the contribution of these fragments to activity, assuming the additivity of the effects. Some studies rely on the partition of the whole molecule into fragments whereas other consider groups substituting a common skeleton such as the Free-Wilson approach (Free and Wilson, 1964).

(2) Techniques based on physicochemical parameters as structural descriptors: the Hansch approach (Hansch and Fujita, 1964). The rationale of this method is the fact that biological responses of the living organisms to drugs are frequently controlled by lipophilic, electronic and/or steric properties. This approach has two disadvantages: i) when an appropriate physicochemical descriptor of a molecule is not available, this compound cannot be included in the study, and ii) it very often occurs that the complexity of the drug-receptor interactions is not sufficiently described by the physicochemical parameters. As a consequence, this technique is often limited to congeneric series, in which substitutions are introduced in a common reference structure.

An important advancement in automated computer treatment of chemical structures and QSAR has been the application of a mathematical technique, namely "Graph Theory", to chemistry (Milne, 1997). Graph theory enables the construction of tools that are very useful for a deeper understanding of chemical and physicochemical properties of molecules.
In chemical graph theory, molecular structures are represented as hydrogen-suppressed graphs, commonly known as molecular graphs, in which the atoms are represented by vertices and the bonds by edges. A molecular graph can be represented by: a sequence of numbers, a polynomial, a single number or a matrix. These representations are aimed to be unique, for a given structure. Topological matrices can be accepted as a rational basis for designing topological descriptors (TDs). In order to account for the presence of multiple or aromatic bonds, or of heteroatoms, in molecular graphs, special parameters have been developed. A general approach was developed by Trinajstic and co-workers by weighing the contributions of atoms and bonds with parameters based on the atomic number $Z$ and the topological bond order (Barysz et al., 1983). The chemists prefer to use TDs as molecular descriptors to evaluate toxicity, and predict biological activity since the TDs encapsulate structural information at the molecular level which is pertinent to the property in question and offer a simple way of measuring molecular branching, shape, size, cyclicity, symmetry, centricity and complexity (Hu et al., 2003a). The direct structural interpretation, good correlation one property, good discrimination of isomers, locally defined, generalizable to higher analogues, not trivially related to other indices, gradual change with gradual change in structures, simplicity, a correct size-dependence and not based on physico-chemical properties are some of the desirable attributes of TDs outlined by Randic (Randic, 1991a) in the view of preventing their hazardous proliferation. In recent years a large number of TDs have been reported in the literature. These are widely employed as simple numerical descriptors for quantitative comparison of physical, chemical or biological parameters of molecules in wide range of [(Q)SAR/SPR] studies. Various TDs have been found to possess different correlating abilities with molecular properties/activities of diverse nature. Various TDs and matrices associated with them have been briefly reviewed in the present chapter.

**The adjacency matrix**

Derived from the molecular graph $G$, the adjacency matrix $A$ represents the whole set of connections between adjacent pairs of atoms (Trinajstic, 1992). The adjacency matrix $A = A(G)$ of a graph $G$ with $N$ vertices is the square $BxB$ symmetric matrix whose entry in the $j^{th}$ column is defined as:
\[ [A]_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ 0 & \text{if } i = j \text{ and } e_{ij} \notin E(G) \end{cases} \]

Where \( E(G) \) is the set of the edges in a connected molecular graph \( G \), \( e_{ij} \) is the edge formed by atoms \( i \) and \( j \). The diagonal elements are zero.

**Augmented adjacency matrix**

In order to take into account the heteroatoms in the molecule, the augmented adjacency matrix was proposed by Randic (Randic, 1991b; Randic, 1991c; Randic and Dobrowolski, 1998b) replacing the zero diagonal entries of the "normal" adjacency matrix with specific values empirically obtained and characterizing different atoms in the molecule. The row sums of this adjacency matrix are local vertex invariants encoding the connectivity of each atom and its atom type; therefore they can be viewed as augmented vertex degrees.

**Edge adjacency matrix**

Derived from the molecular graph \( G \) the edge adjacency matrix \( E \) encodes the connectivity between graph edges. It is a square symmetric matrix of dimension \( B \times B \), where \( B \) is the number of bonds and is usually derived from an H-depleted molecular graph (Bonchev, 1983). The edge adjacency matrix of a graph \( G \) is equal to the adjacency matrix of the line graph of \( G: EA(G) = A(Li(G)) \) (Gutman and Estrada, 1996).

**Additive adjacency matrix**

The additive adjacency matrix \( A_{\alpha} \) is a modification of adjacency matrix. It is defined as the matrix obtained by the sum of the non-zero row elements in the adjacency matrix representing the degree of corresponding vertex (vertices adjacent to vertex \( i \)) of a molecular graph \( G \). (Gupta et al., 2001). Similarly, the matrix obtained by sum of the non-zero row elements in the adjacency matrix representing the chemical degree of corresponding vertex (vertices adjacent to vertex \( i \)) is defined as additive chemical adjacency matrix. The chemical degree of a vertex can be obtained by substituting, row elements of the adjacency matrix corresponding to heteroatom(s), with atomic weight with respect to carbon atom (Gupta et al., 2003).
**Augmentative adjacency matrix**

The augmentative adjacency matrix ($A^\alpha$) is also an modification of adjacency matrix. It is defined as the matrix obtained by the product of the non-zero row elements in the adjacency matrix represents the degree of corresponding vertex (of the vertices adjacent to vertex $i$) of a molecular graph $G$, the matrix may be defined as augmentative adjacency matrix (Dureja and Madan, 2007). In the similar way, augmentative chemical adjacency matrix is defined as the product of the non-zero row elements in the adjacency matrix representing the chemical degree of corresponding vertex (vertices adjacent to vertex $i$) of a molecular graph $G$ (Dureja et al., 2008a).

**The distance matrix**

The distance matrix (viz., vertex distance matrix) summarizes in matrix form the topological distance information between all the atom pairs in a molecular graph $G$ (Trinajstic, 1992). It is a square symmetric table, of dimension $BxB$ whose elements $[D]_{ij}$ are defined as:

$$[D]_{ij} = \min_{d_{ij}} \text{ if } i \neq j$$

$$= 0 \text{ if } i = j$$

The topological distance $d_{ij}$ is the number of edges in the shortest path between the vertices $i$ and $j$. A number of MDs make use of distance matrix and its computation can be performed with various algorithms.

Distance sum is the simplest number obtained from distance matrix which can be associated with the chemical structure is the graph distance sum. The distance sum (or vertex distance degree) is a local vertex invariant. It can be expressed as:

$$S_i = \sum_{j=1}^{n} d_{ij}$$

Where $d_{ij}$ is length of the path that contains the least number of edges between vertex $i$ and vertex $j$ in graph $G$; $n$ is the maximum possible number of $i$ and $j$ (Balaban, 1983).

The distance sum was used to define a variety of MDs including the well known Balaban index (J) and Wiener index (W).
The reciprocal distance matrix

The reciprocal distance matrix $D^{-1}$ or Harary matrix is a square symmetric AxA matrix derived from the distance matrix whose entries are defined as:

$$[D^{-1}]_{ij} = 1/d_{ij}$$

i.e. each off-diagonal element is the reciprocal of the topological distance $d$ between the considered vertices (Ivanciuc et al., 1993a; Plavsic et al., 1993) and the diagonal elements are zero by definition. The original Harary index H is calculated from this matrix as an analog of the Wiener index.

The reciprocal distance sum $RD_S$ of the $i^{th}$ vertex is a local invariant defined as the sum of the reciprocal distance matrix elements in the $i^{th}$ row.

The $D^{-1}$ or Harary matrix was successfully utilized to generate new structural descriptors and in the computer generation of acyclic graphs based on the local vertex invariants and TDs.

Generalized distance matrix

The generalized distance matrix denoted as $D^\lambda$, are derived from the distance matrix by raising its elements to any real exponent $\lambda$ (Nikolic et al., 2005). The entries are defined as:

$$[D^\lambda]_{ij} = d^\lambda_{ij} \text{ if } i \neq j$$
$$= 0 \text{ if } i = j$$

Resistance distance matrix ($\Omega$)

Resistance distance matrix was proposed by Klein and Randic (Klein and Randic, 1993) in connection with the electrical network theory, was defined as the effective electrical resistance between two vertices ($i$ and $j$) when a battery is connected across them and each graph edge is considered as a resistor taking a value of 1 ohm. The elements of resistance distance matrix i.e. $\Omega_{ij}$ are defined as:

$$[\Omega]_{ij} = 0 \text{ if } i = j$$
$$= \max\Omega_{ij} \text{ if } i \neq j$$

For acyclic graphs, resistance distances are equal to topological distances but for cyclic graphs resistance distances may be smaller than, or equal to topological distances. The resistance distance was used to generate rules to characterize molecular cyclicity and centricity (Klein and Randic, 1993).
**Conductance matrix (ρ)**

The conductance matrix (Klein and Ivanciuc, 2001) is also called as the reciprocal resistance matrix as the inverse of the resistance distance is the conductance \( \rho_{ij} \) between two vertices \( i \) and \( j \). It can be expressed as:

\[
[\rho]_{ij} = \frac{1}{\Omega_{ij}} = \sum p_{ij} [p_{ij}]^{-1}
\]

Where the sum runs over all the paths \( p_{ij} \) connecting the two considered vertices and \( p_{ij} \) is the length of the considered path \( p_{ij} \) (Klein and Ivanciuc, 2001).

**The detour matrix (Δ)**

The detour matrix \( A \) of a graph \( G \) (or maximum path matrix) is a square symmetric \( A \times A \) matrix, \( A \) being the number of graph vertices, whose entry \( i-j \) is the length of the longest path from vertex \( i \) to vertex \( j \) (Buckiey and Harary, 1990; Ivanciuc and Balaban, 1994b) The detour matrix, \( [\Delta]_{ij} \), is defined as:

\[
[\Delta]_{ij} = p_{ij} \quad \text{if} \quad i \neq j
\]

\[
= 0 \quad \text{if} \quad i = j
\]

Where \( p_{ij} \) is the detour distance which is the the number of steps in a longest path between (i.e. the maximum number of edges) vertices \( i \) and \( j \) in a molecular graph \( G \). For acyclic graphs, the detour matrix is generally identical to the distance matrix, but for cyclic graphs elements in the detour matrix may be equal to, or larger than, those of the distance matrix (Trinajstic et al., 1997c). For edge-weighted graphs, the weighted detour matrix \( ^w\Delta \) was proposed (Nikolic et al., 1996a). The off-diagonal \( i-j \) entry is defined as the maximum path weight, i.e. the maximum sum of edge weights along the path between the vertices \( i \) and \( j \) which is not necessarily the longest possible path between them.

From the detour matrix and the distance matrix, a combined matrix, called detour/distance matrix \( \Delta/D \) (or maximum/minimum path matrix), is defined as (Ivanciuc and Balaban, 1994b):

\[
[\Delta-D]_{ij} = [\Delta]_{ij} (i,j) \in E (G) \text{ if } i < j
\]

\[
= [D]_{ij} (i,j) \notin E (G) \text{ if } i > j
\]

\[
= 0 \quad \text{if} \quad i = j
\]
It is a square unsymmetric AxA matrix, where the upper triangle of the matrix contains the elements of the detour matrix (information about the longest paths) and the lower triangle contains the elements of the topological distance matrix (information about the shortest paths).

The distance/detour quotient matrix D/A is also derived from detour and distance matrices but it is a square symmetric matrix AxA whose off-diagonal entries are the ratios of the lengths of the shortest to the longest path between any pair of vertices (Randic, 1997a). It is defined as:

\[
[D/\Delta]_{ij} = \frac{d_{ij}}{\Delta_{ij}} \quad \text{if } i \neq j
\]

\[
= 0 \quad \text{if } i = j
\]

Where \(d_{ij}\) and \(\Delta_{ij}\) are the topological and detour distances between vertices \(i\) and \(j\) respectively.

The reciprocal detour matrix \((\Delta^{-1})\) may be expressed as:

\[
[\Delta^{-1}]_{ij} = \Delta^{-1}_{ij} \quad \text{if } i \neq j
\]

\[
= 0 \quad \text{if } i = j
\]

All elements equal to zero are left unchanged in the reciprocal matrix. The matrix is used to derive Harary detour indices and Hyper-Harary detour indices (Diudea et al., 1998).

**The Laplacian matrix \(L(G)\)**

Laplacian matrix is also known as admittance matrix and Kirchhoff matrix. It is a square BxB symmetric matrix, B being the number of vertices in the molecular graph \(G\), obtained as the difference between the vertex degree matrix \(V(G)\) and the adjacency matrix \(A(G)\) (Mohar, 1989a; Mohar, 1989b)

\[
L(G) = V(G) - A(G)
\]

Where \(V(G)\) is a diagonal matrix of dimension BxB whose diagonal entries are the vertex degrees of molecule atoms.

Hence, the Laplacian matrix can be represented as given below

\[
[L]_{ij} = \delta_{ij}, \text{ if } i = j
\]

\[
= -1 \text{ if } e_{ij} \in E(G)
\]

\[
= 0 \text{ if } e_{ij} \notin E(G)
\]

Where \(\delta_{ij}\) means the vertex degree of atom \(i\) and where \(E(G)\) is the set of edges of the molecular graph \(G\).
**The $\chi$ matrix**

Moreover, another connectivity matrix is obtained by weighting each bond between vertices $v_i$ and $v_j$ by the edge connectivity, $(\delta_i \cdot \delta_j)^{-1/2}$. This matrix is known as $\chi$ matrix (Randic, 1992) or more formally edge-$\chi$ matrix and is defined as:

$$[\chi]_{ij} = (\delta_i \cdot \delta_j)^{-1/2} \quad \text{if } e_{ij} \in E(G)$$

$$= 0 \quad \text{otherwise}$$

Where $\delta$ being vertex degree of the atoms (Randic, 1992).

**The Zagreb matrix**

Zagreb matrices are a generalization of the $\chi$ matrix in terms of a variable exponent $\lambda$ as:

$$[ZM_\epsilon(\lambda)]_{ij} = (\delta_i \cdot \delta_j)^\lambda \quad \text{if } e_{ij} \in E(G)$$

$$= 0 \quad \text{otherwise}$$

For $\lambda = -1/2$, the Zagreb matrix obviously reduces to the edge-$\chi$ matrix.

The Zagreb matrix can also be considered as the vertex- and edge-weighted matrices related to the vertex- and edge-connectivity matrices. They can be formulated in terms of the vertex- or edge-degrees (Janezic et al., 2007).

The vertex-Zagreb matrix, $ZM_v$, was defined as (Janezic et al., 2007)

$$[ZM_v]_{ij} = (\delta_i^2) \quad \text{if } i = j$$

$$= 0 \quad \text{if } i \neq j$$

Similarly, the modified vertex-Zagreb matrix, $^{m}ZM_v$, was defined as (Janezic et al., 2007),

$$[^{m}ZM_v]_{ij} = (1/\delta_i^2) \quad \text{if } i = j$$

$$= 0 \quad \text{if } i \neq j$$

The sum of the diagonal elements of the vertex Zagreb matrix results into the first Zagreb index ($M_1$) (Gutman and Trinajstic, 1972), while the sum of the diagonal elements of the modified vertex Zagreb matrix results into the modified first Zagreb index.
The combinatorial matrix

The two path-defined matrices named as the distance-path $D_p$ and the detour-path $A_p$ have been reported whose elements are combinatorially calculated from the classical distance (*i.e.*, distance-edge), $D$ and detour $A$ (*i.e.*, detour-edge), matrices (Brualdi and Ryser, 1991).

Matrices $D_p$ and $A_p$ allow the direct calculation of the hyper-Wiener (WW) and hyper-detour, (ww) indices, respectively.

The Wiener matrix

Wiener matrix was proposed by Randic a square matrix, denoted by $W$ is derived from the H-depleted molecular graph $G$, based on counts of external paths with respect to a considered path $i$-$j$, and defined only for acyclic graphs (Randic, 1993). The off-diagonal $w_{ij}$ entry corresponds to the number of external paths in the graph $G$ which contain the path from vertex $i$ to vertex $j$ calculated as the product of the number of vertices on each side of the path $i$-$j$, namely $N_{ixij}$ and $N_{jxij}$, including both vertices $i$ and $j$.

$$[W]_{ij} = N_{ixij} \times N_{jxij}, \quad \text{if } i \neq j$$

The Wiener matrix has been very well exploited as a source of structural invariants, useful in QSPR/QSAR.

Reciprocal Wiener matrix

Reciprocal Wiener matrix ($W_{x}^{-1}$) is a matrix (Diudea, 1997a) whose elements are the reciprocal of the corresponding Wiener matrix elements. It can be represented as

$$[W_{x}^{-1}]_{ij} = 1/[W]_{ij}$$

Edge-Wiener matrix

Edge-Wiener matrix ($W_{E}$) utilizes edges in place of vertices and can be expressed as:

$$[W_{E}]_{ij} = N_{i,Eij} \times N_{j,Eij}, \quad \text{if } i \neq j$$

Where $N_{i,Eij}$ and $N_{j,Eij}$ denote the number of vertices lying on the two sides of the path *i.e.* edge path having vertices $v_i$ and $v_j$ as the endpoints.
The Hosoya matrix

A square symmetric matrix of dimension $A \times A$, $A$ being the number of vertices in the H-depleted molecular graph $G$. The original Hosoya $Z$ matrix is defined only for acyclic graphs; its entry $z_{ij}$ is equal to the Hosoya $Z$ index of the subgraph $G'$ (Randic, 1994) obtained by erasing all edges along the path connecting two vertices $v_i$ and $v_j$. The $Z$ matrix entry $[Z]_{ij}$ corresponding to a pair of vertices $v_i$ and $v_j$ of a tree $T$ is given by:

$$[Z]_{ij} = Z(T-p_{ij}) \text{ if } i \neq j$$

$$0 \text{ if } i = j$$

Where $Z(T-p_{ij})$ is the $Z$ index of the spanning subgraph $T-p_{ij}$ obtained from $T$ by the removal of all edges along the path $p_{ij}$ connecting the vertices $v_i$ and $v_j$ (Randic, 1994).

A general definition of the Hosoya $Z$ matrix (generalized Hosoya $Z$ matrix) able to represent both acyclic and cyclic graphs is the following:

$$[Z]_{ij} = \sum_{\text{min } p_{ij}} Z(T-p_{ij})/p_{ij} \text{ if } i \neq j$$

$$Z_{ij} \text{ if } i = j$$

Where $Z(T-p_{ij})$ is the $Z$ index of the spanning subgraph (Plavšic et al., 1997).

The Szeged matrix

Since the Wiener matrix is not defined in cyclic structures, Ivan Gutman (Gutman, 1994a) has extended the meaning of $N_i$ and $N_j$ as follows:

$$N_{i,e/p} = |\{vk \in V(G); [D]_{ik} < [D]_{jk}\}|$$

$$N_{j,e/p} = |\{vk \in V(G); [D]_{jk} < [D]_{ik}\}|$$

Thus, $N_{i,e/p}$ and $N_{j,e/p}$ denote the cardinality of the sets of vertices lying closer to $i$ and $j$; vertices equidistant to $i$ and $j$ are not counted. The Szeged matrix $(SZ)$ of an H-depleted molecular graph $G$ is a square unsymmetric $A \times A$ matrix, $A$ being the number of graph vertices, whose off-diagonal entry $ij$ is the number of vertices $N_{i,(ij)}$ lying closer to the focused vertex $v_i$ (Diudea, 1997)

$$[SZ_{e/p}]_{ij} = N_{i,e/p} \cdot N_{j,e/p}$$

Where $[SZ_{e/p}]_{ij}$ are the non-diagonal entries of this matrix.
The path matrix
Path matrix is a square symmetric $B \times B$ matrix whose entry in the $i$th row and $j$th column is defined (Randic, 1991d) by the equation:

$$[P]_{ij} = \frac{p'_{ij}}{p} \quad \text{if } i \neq j \text{ and if } (i, j) \in E(G)$$

$$= 0 \quad \text{otherwise}$$

Where $p'_{ij}$ is the total number of paths in the subgraph $G' = G-(i,j)$ and $p$ is the total number of paths in $G'$. If $G$ is disjoint then the contributions of each component are added (Randic, 1991d).

The Cluj Matrix
The unsymmetric Cluj matrix (CJ) had been defined by Diudea (Diudea et al., 1996; Diudea, 1997b; Diudea, 1997c). It is a square unsymmetrical matrix following the principle of single endpoint characterization of a path. The non-diagonal entries, $[UM]_{ij}$, $M = CJ_D$ (Cluj-Distance) or $CJ_\Delta$ (Cluj-Detour), are defined as: by using either the distance or the detour concept:

$$[CJ]_{ij} = N_{i,(i,j)}$$

$N_{i,ij} = \max \left\{ v_k \mid v_k \in V(G); [D]_{ik} < [D]_{jk} \right\}$

$; (i,k) \cap (i,j) = \max \left\{ i \right\} \left\{ i,j \right\} = \min \}$

It collects the vertices lying closer to the focused vertex $i$ but out of the shortest path $(i,j)$ or, in other words, the “external” paths on the side of $i$, which include the path $(i,j)$. Several indices can be calculated on Cluj matrices, either directly by the Wiener orthogonal operator from the unsymmetric matrices or as the half-sum of entries in the corresponding symmetric matrices by the Wiener operator.

The pendent matrix
The pendent matrix (Dp) of a graph $G$ is a submatrix of distance matrix, obtained by retaining the columns corresponding to pendent vertices (Gupta et al., 1999).

$$[Dp]_{ij} = d_{ij} \text{ if } i \neq j$$

$$= 0 \text{ if } i = j$$

Where $d_{ij}$ is the length of the path that contains the least number of edges between vertex $i$ and vertex $j$ in graph $G$. 

37
The chemical pendent matrix
The chemical pendent matrix is a modified pendent matrix. The chemical distance of a terminal or pendent vertex was obtained from the pendent matrix by substituting, row elements corresponding to heteroatom(s), with atomic weight with respect to carbon atom. (Gupta et al., 1999; Goyal et al., 2010).

The Barysz distance matrix
The Barysz distance matrix (DZ) is a weighted distance matrix accounting simultaneously for the presence of heteroatoms and multiple bonds in the molecule; it is defined as (Barysz et al., 1983)

\[
[D^2]_{ij} = 1 - \frac{Z_C}{Z_i} \sum_{d_{ij}} \left( \frac{1}{B} \ast \frac{Z_C}{Z_{b1}} \ast \frac{Z_C}{Z_{b2}} \right)
\]

Where \(Z_C\) is the atomic number of the carbon atom, \(Z_i\) the atomic number of the \(i^{th}\) atom, \(B\) is the conventional bond order, the sum runs over all \(d_{ij}\) bonds involved in the shortest path between vertices \(v_i\) and \(v_j\), \(d_{ij}\) being the topological distance, and the subscripts \(b(1)\) and \(b(2)\) represent the two vertices incident to the considered bond (Barysz et al., 1983).

Distance based graph invariants
These indices employ distance matrix to characterize molecular graphs. Distance matrix for a molecular graph \(G\) is defined as a real, square, symmetrical matrix of of \(N \times N\) dimension representing the distance traversed in moving from vertex \(i\) to vertex \(j\) in \(G\). The entries in the distance matrix indicate the number of edges in the shortest path between vertices \(i\) and \(j\). In chemical graph theory, the distance matrix accounts for the “through bond” interactions of atoms in molecules.

The first MD based on the chemical graph theory to model some thermodynamic properties of hydrocarbon was introduced by Harold Wiener in 1947. H. Wiener called it as path number eventually it was referred as Wiener number. Wiener number/ path number \((W)\) can be defined as the sum of the distances between
any two carbon atoms in an acyclic molecule, in terms of carbon-carbon bonds. Wiener (Wiener, 1947a; b;c) also proposed a formula for calculating his “path number” by simply counting the number of atoms \(n_i\) and \(n_j\) on both sides of each bond as:

\[ W = \sum_{\text{all bonds}} n_i n_j \]

Where \(n_i\) and \(n_j\) are the number of vertices on each side of the bond, including vertices \(i\) and \(j\), respectively.

The Wiener index was first defined within the framework of chemical graph theory by Hauro Hosoya (Hosoya, 1971) in 1971, who pointed out that this index (he called it the Wiener number) can be calculated from the distance matrix \(D = D(G)\) of a molecular graph \(G\). This topological index, denoted by \(W\) in honour of Wiener is a number equal to the half-sum of the elements of the distance matrix:

\[ W = 1/2 \sum_{i=1}^{n} D_{ij}(G) \]

Where \(D_{ij}\) are the off-diagonal elements of \(D\), and \(n\) is the number of vertices in \(G\).

In fact, Wiener index is amongst the most widely studied indices and had become one of the central subject for the research to the theoretical chemists. There are a number of shortcomings associated with the Wiener index that includes high degeneracy, insensitivity to presence of heteroatom(s) and unsaturation of bonds. Therefore, to overcome these shortcomings, a large number of variants of this index have been reported in literature. A few variants are mentioned as below:

Reverse Wiener index, overall Wiener index, Wiener sum index, inverse Wiener-sum index, the product-form version of Wiener index, quasi-Wiener index (Kirchoff number), all path Wiener index (Wap) (Lukovits, 1998), Wiener type index from \(Z\) weighted distance matrix (Whetz), mass weighted distance matrix (Whetm), van der waals weighted distance matrix (Whetv), electronegativity weighted distance matrix (Whete) (Barysz et al., 1983), and Wiener Topochemical Index (Bajaj et al., 2004a).

Hosoya (1971) inventor of topological indices introduced \(Z\) index for characterizing the topological nature of saturated isomers of saturated hydrocarbons as:

\[ Z = \sum_{k=0}^{[n/2]} p(G,k), \]
Where \( p(G,k) \) called *nonadjacent number* and represents the number of ways in which \( k \) bonds are so chosen from graph \( G \) that no two of them are connected. \( n \) is the number of graph vertices and the Gaussian brackets represent the greatest integer not exceeding \( n/2 \). The \( Z \) index is calculated by summing the \( p(G,k) \) coefficients over all different \( k \) values. The *Hosoya Z index* depends on the molecular size as well as on branching and ring closure. Although *Hosoya Z index* has been associated with the adjacency matrix, it can be classified among the distance based topological indices due to the procedure used to calculate \( p(G,k) \) (Hosoya, 1971). Further, he proposed the use of \( Z \) index as a first sorting device for coding or retrieving the structure of the compounds with or without rings. He also stated that equal value of TDs of two given graphs is not sufficient condition for identity of structures and one must resort to finer comparison or identification (Hosoya, 1972). However, if \( Z \) values of two given graphs are different, further comparison is not needed. *Hosoya’s Z index* offers certain advantages over other graph invariants: 1) High discriminating power, 2) It can be computed from the molecular structure as well as polynomial, 3) It can be extended to matching polynomial and distance polynomial, 4) It is capable of considering the influence of electronic systems.

**Rouvray** (1973) defined the total sum of the row entries of the distance matrix as the new topological index called *Rouvray index*, denoted by \( I_{ROUV} \). It is actually twice the *Wiener index*, \( W \) (Rouvray, 1973) and expressed as:

\[
I_{ROUV} = \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} = \sum_{i=1}^{n} S_i = 2W
\]

**Hosoya et al.** (1975) proposed a generalization of \( Z \) index to account for the contributions made by unsaturated system in the structure of a molecule.

**Bonchev et al.** (1980) defined *total distance rank* \( (d_{r,i}) \) as sum of distances of vertex \( i \) is equal to the sum of all entries in the row column \( i \) of the distance matrix and is given by following:

\[
d_{r,i} = \sum_{j=1}^{n} d_{ij} = \min
\]

Where \( d_{ij} \) is the elements of distance matrix.
Balaban (1982, 1983) introduced three new topological distances based indices. This first one was named as mean square distance index, $D$, given by the expression:

$$
D^k = \left( \frac{\sum_i g_i \times i^k}{\sum_i g_i} \right)^{1/k}
$$

Where $g_i$ is number of occurrences of distances of length $i$.

This index possesses fair discriminating ability, especially for acyclic graphs, however, for polycyclic graphs it has poorer performance and higher degeneracy. The second one was called end point mean (square) distance topological index and is denoted as $D_1$. The third one was Balaban distance connectivity index (also called distance connectivity index or average distance sum connectivity) and is denoted as $J$. At the time of its inception, $J$ was one of the most discriminating molecular descriptors and its values did not increase substantially with molecular size or number of rings. Further $J$ was claimed to have the lowest degeneracy of all TIs proposed till the time of its proposal. It is defined as:

$$
J = \frac{B}{C + 1} * \sum_{(i,j) \in E(G)} \left( \sigma_i * \sigma_j \right)^{0.5}
$$

Where $\sigma_i$ and $\sigma_j$ are the vertex distance degrees of two adjacent atoms, and the sum runs over all the molecular bonds $b$; $B$ is the number of bonds in the molecular graph $G$, and $C$ is cyclomatic number i.e. number of rings in the graph. The denominator $C+1$ is a normalization factor against the number of rings in the molecule. $\sigma_i^* = \sigma_i / B$ is the average vertex distance degree.

Barysz et al. (1983) introduced Wiener-type index from $Z$ weighted distance matrix [$W(G_{vew})$]. It may be defined as the sum of all the elements in upper triangular Barysz matrix as well as all the diagonal elements thereof as per following:

$$
W(G_{vew}) = \sum d_{ii} + \frac{1}{2} \sum_i \sum_j d_{ij}
$$

Where $d_{ii} = 1 - \frac{6}{Z_i}$

Where $Z_i$ is the number of all (valence and inner shell) electrons in the atom $i$. 
The off-diagonal elements of the distance matrix for the vertex- and edge-weighted (multi) graphs are defined as:

\[ d_{ij} = \sum_{r} k_r \]

Where the summation goes over \( r \) (\( r = 1,2,\ldots \)), weighted and unweighted, bonds, while the parameter \( k_r \) is given by (Barysz et al., 1983),

\[ k_r = \frac{1}{b_r} \times \frac{36}{Z_i \times Z_j} \]

Where \( b_r \) =

\[
\begin{cases}
1 & \text{for a single bond} \\
2 & \text{for a double bond} \\
3 & \text{for a triple bond} \\
1.5 & \text{for an aromatic bond}
\end{cases}
\]

\( Z_i \) and \( Z_j \), are the numbers of all electrons in atoms \( i \) and \( j \) making up the \( r \)-bond (Barysz et al., 1983)

**Broto et al.** (1984a; 1984b) conceived autocorrelation descriptors of topological structure namely *autocorrelation of a topological structure (ATS)* and subsequently used to correlate structure and activity of two non-narcotic analgesic series of 190 Glafenines and 106 isoindomethacins. This is a spatial autocorrelation defined on a molecular graph \( G \) as:

\[ \text{ATS}_d = \sum_{i=1}^{A} \sum_{j=1}^{A} \delta_{ij} \cdot \delta_{d} \left( W_i \cdot W_j \right)_d \]

Where \( W \) is any atomic property, \( A \) is the atom number, \( d \) is the considered topological distance (i.e. the lag in autocorrelation terms), \( \delta_{ij} \) is Kronecker delta (\( \delta_{ij}=1 \) if \( d_{ij} = d \), zero otherwise).

**Randic** (1984) proposed a index termed as *molecular identification number (MID)* in terms of weighted path counts or weighted walk counts, to univocally identify a molecule by a single real number.

\[ \text{MID} = n + \sum_{i=1}^{n} p_{ij}(w_i) \]
Where $p_{ij}$ is the path of length $m$ from vertex $v_i$ to $v_j$, and $w_{ij}$ is path weight.

The sum runs over all paths of the graph; each path of length zero is given a unit weight.

Kier (1985) developed molecular shape index, $^2\kappa$, based on the count of the two-bond fragments. Kier introduced the term, $^2P_{\text{min}}$, as the count of the number of two-bond paths, which for any non-cyclic molecule is equal to

$$^2P_{\text{min}} = A - 2$$

Where $A$ is the total number of vertices in the graph.

On the other hand, $^2P_{\text{max}}$ can be calculated for any isomeric series through

$$^2P_{\text{max}} = (A - 1)(A - 2)/2$$

All non-cyclic molecules with in an isomeric series have $^2P_i$ values at or within the bounds of the minimum and maximum value of $P$

$$^2P_{\text{min}} \leq ^2P_i \leq ^2P_{\text{max}}$$

After appropriate considerations and refinements, Kier derived the final equation as per following:

$$^2\kappa = \frac{2(^2P_{\text{max}} - ^2P_{\text{min}})}{^2P_i} = \frac{(A - 1)(A - 2)^2}{^2P_i}$$

According to Kier, the $^2\kappa$ index describes the molecular shape in relationship to the star and linear graph and is normalized to the number of vertices. The Kappa shape invariants are the basis of a method of molecular structure quantification in which attributes of molecular shape are encoded into three invariants (Kappa values). These Kappa values are derived form counts of one-bond, two-bond and three-bond fragments, each count being made relative to fragment counts in reference structures which possess a maximum and minimum value for that number of vertices.

Balaban (1986) adapted average distance-sum-connectivity index, $J$, in order to account for both bond multiplicity and heteroatoms and defined two new descriptors i.e $J^x$ and $J^y$ employing fractional distance matrix. The quantities X and Y are recalculated atomic Sanderson electronegativities and covalent radii relative to carbon atom, respectively, obtained as a function of the atomic number.
Schultz (1989) put forward *Schultz molecular topological index* (MTI) by using adjacency, valence, and distance matrices. The quantity is defined by the following expression

$$MTI = MTI(\Gamma) = \sum_{i=1}^{N} \left[ v(A + D) \right]_i$$

Where $\Gamma$ is the molecular graph considered, possessing $N = N(\Gamma)$ vertices. Further $A$ is the $(N \times N$ dimensional) adjacency matrix, $D$ is the $(N \times N$ dimensional) distance matrix, and $v = (v_1, v_2, ..., v_N)$ is the $(1 \times N$ dimensional) vector of the vertex valencies (degrees) of the molecular graph $\Gamma$.

Hall and Kier (1990) reported *topological state or vertex topological state invariant* $S_i$, in order to capture vertex information in all the paths emanating from that vertex. The *topological state invariant* derived from topological state matrix can be calculated as:

$$S_i \equiv VS_i(T) = \sum_{j=1}^{n} [T]_{ij}$$

Where $VS$ stands for the row sum operator.

Balaban et al. (1990) used distance measure in defining a highly selective connectivity based spectrum of TDs (denoted by: $DM^k$) as:

$$DM^k = \sum_{i=1}^{n} \left[ \chi_t^i - \chi_t^i(R)^{1/k} \right]$$

Where the summation goes over all the connectivity indices of different type $t$ up to the sixth order ($m = 6$); $k$ is an integer parameter ranging from 1 to 5 ($k=1$ is the Manhattan distance, $k=1$ is the Euclidean distance). $\chi_t^i$ and $\chi_t^i(R)$ are the connectivity indices for the considered molecule and a reference molecule $R$, respectively. No degeneracy was shown by these TDs for a set of 661 alkanes between butane and dodecane (Balaban et al., 1990).

Lall (1991) proposed *Topological I-index* of a graph $G$. This is based on the topological distances from a given vertex in the edge weighted graph of the organic molecule and is defined as:
Where \( n_r \) is the number of \( r \)th kind of vertices for which \( g_r \) is the topological distance from the root in the edge weighted graph and the topological distance \( d_{ij} \) between the vertices \( i \) and \( j \) is defined as the distance associated with a minimum weight. The weights in the edge-weighted graphs correspond to \( k \) values of the Huckel parameter for the heteroatom.

Petitjean (1992) described graph theoretical shape coefficient \( I \) on the basis of topological radius and diameter as:

\[
I = D - R / R \quad 0 \leq I \leq 1
\]

The topological radius \( (R) \) is defined as the smallest vertex eccentricity in the graph and topological diameter \( (D) \) is defined as the largest vertex eccentricity in the graph. He proposed the graph-theoretical bivariate repartition of the \((R,D)\) pairs in the form of "radius-diameter diagram" and observed unexpected partitioning/classification of the compounds from the chemical abstracts services registry file in comparison to other shape coefficients (Petitjean, 1992).

Plavsic et al. (1993) introduced the Harary index \( (H) \), in the honor of Prof. Frank Harary on his 70th birthday. It is defined as the half-sum of the off-diagonal elements of the reciprocal molecular distance matrix \( D' = D^{-1} (G) \) as per the following:

\[
H = (1/2) \sum_{i=1}^{N} \sum_{j=1}^{N} (D')_{ij}
\]

Ivanciuc et al. (1993a;b) introduced the reciprocal distance matrices \( D'^{-1} \) by modifying distance matrix where each off-diagonal element is the reciprocal of the topological distance between the vertices. A local graph invariant termed as the reciprocal distance sum, \( RDS_i \), derived from this matrix was defined as:

\[
RDS_i \equiv VS_i (D^{-1}) = \sum_{j=1}^{n} d^{-1}_{ij}
\]
Ivancic et al. (1993a; b) defined two other MDs, called $RDSQ$ index and $RDCHI$ index, respectively based on a Randic-like formula as:

$$RDSQ = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta_{ij} (RDS_i \bullet RDS_j)^{1/2}$$

$$RDCHI = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta_{ij} (RDS_i \bullet RDS_j)^{-1/2}$$

The $RDSQ$ index increases with increase in both molecular size and molecular branching, while $RDCHI$ index increases with molecular size and decreases with molecular branching.

Randic et al. (1993) constructed new graph matrices by generalizing Wiener’s procedure for evaluation of Wiener numbers in alkanes and reported the sequences (higher Wiener numbers $K_W$) generated by summing the entries in the matrix for vertices at the same distances from one another. Wiener matrix was extended to consider heteroatoms. The matrix was known as Wiener matrix.

Randic (1993) introduced a modified Wiener index known as Hyper-Wiener Index, denoted as $WW=WW (G)$. However, Randic’s algorithm for computing the hyper-Wiener index could be applied only to acyclic structures. It was later shown that $WW$ can be computed for all structures as follows (Lukovits and Linert, 1994; Klein et al., 1995).

$$WW = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ (D)_{ij} + (D^2)_{ij} \right]$$

Where the summation goes over all pairs of vertices $i$ and $j$.

Randic et al. (1994a) proposed the first eigenvalue of Wiener matrix as an alternative index for the molecular branching. $JJ$ index was developed in analogy with the connectivity index from adjacency matrix. First row sums, $R_i$ were obtained and these numbers were used in formula $1/R_i R_j^{1/2}$ for all pairs of adjacent vertices.

Randic et al. (1994b) also proposed new structural invariants based upon distance/distance matrices ($DD$ matrix) for graphs embedded on two and three-dimensional grids. The first value of these matrices, $\lambda/n$ for path graphs was reported to be index of folding. The ratio $\Phi = \lambda/n$ approaches 1 (one) for geometrically linear structures while it approaches 0 (zero) as path graph is repeatedly folded.
Khadikar et al. (1995) devised Szeged index \((Sz)\) as generalization of the original Wiener index, valid both for acyclic and cyclic graphs. Wiener index is the sum of the product of the number of vertices on each side of a bond, while the Szeged index is defined as the sum of the product of the number of vertices closer to the atoms on each side of a bond (Gutman and Klavzar, 1995). Szeged index was defined as:

\[
Sz(G) = \sum_{e \in E(G)} n_1(e|G)n_2(e|G)
\]

Where the vertices \(n_1(e|G)\) is the number of vertices closer to \(i\) than to \(j\), and \(n_2(e|G)\) is the number of vertices closer \(j\) than to \(i\); vertices equidistant to \(i\) and \(j\) are not counted (Khadikar et al., 1995).

Randic and Razinger (1997) introduced molecular topographic indices, which were geometry dependent. One new way to arrive at the novel topographic indices is to design matrices, the elements of which depend on the molecular geometry. The entries in these matrices were proposed to be 3-D geometric distances between atoms and some even more general function of inter-atomic distances.

Estrada and Gutman (1996) proposed a novel molecular topological index \(MTI (E)\) based on edge-distances in molecular graphs in analogy with Schultz molecular topological index (Schultz, 1989). The edge-based version of the molecular topological index was defined as:

\[
MTI(E) = \sum_{i=1}^{n} v_i \left( A^e + D^e \right)
\]

Where \(A^e\) is the edge adjacency matrix and \(D^e\) is the edge distance matrix. \(v_i\) is the degree of the \(i\)th edge of the molecular graph \(G\). Edge-distances in a graph are equal to distances between vertices of the respective line graph. A simple relation was found between edge-distances and the distances between the vertices that are incident to the respective edges.

Linert and Lukovits (1997) introduced hyper-detour index \((\omega\omega)\) by replacing \((D)_{ij}\) by \((\Delta)_{ij}\) in equation of detour index as per the following equation:

\[
\omega\omega = (1/4) \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ (\Delta)_{ij} + (\Delta^2)_{ij} \right]/2
\]
Lukovits (1998) introduced all-path Wiener index ($W^{AP}$) as all path version of the Wiener index but with more discriminating power among cycle-containing structures. It is defined as:

$$W^{AP} = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} p_{ij} |P_{ij}|$$

Where the two outer sums on the right side run over all pairs of vertices in the graph and inner sum runs over all paths $p_{ij}$ between the vertices $v_i$ and $v_j$, and $|p_{ij}|$ denotes the length of the considered path. This index is also called the Pasareti index, because it was derived in home of Lukovits, which is located in the part of Budapest called Pasarest.

Estrada and Gutman (1996) has derived edge-type Schultz index ($E^{MTI}$) from edge adjacency matrix $E$ and edge distance matrix $ED$ as per the following equation:

$$E^{MTI} = \sum_{b=1}^{B} \left[ v \left( E + E^T \right) \right]_b$$

Where $v$ is a B-dimensional column vector whose elements are the edge degrees $\varepsilon$.

Lukovits (1996) described detour index ($\omega$) of a molecular graph $G$ in a way similar to the Wiener index; that is, the detour index is equal to the half-sum of the off-diagonal elements of the molecular detour matrix $\Delta = \Delta(G)$:

$$\omega = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} (\Delta)_{ij}$$

Where $(\Delta)_{ij}$ represents the length of a longest path between vertexes $i$ and $j$ of $G$.

Randic (1997b) described molecular descriptors as linear combination of path numbers. The coefficients determining the linear combination was defined as a function $f(p_i)$, where $p_i$ is the path of length $i$.

Diudea (1997b) reported several indices from the Cluj matrices, either as the half-sum of entries in the corresponding symmetric matrices or directly from the unsymmetric matrices as:

$$T_{\ell/p} = \sum_{u,v} (M_u)_y(M_u)_z$$
When defined on edges, $T_l^e$ is a Cluj index: denoted by $CD_e$ or $CA_e$, depending on whether it is derived from the Cluj-distance or Cluj-detour matrix (Diudea, 1997b). Similarly, when defined on paths, $T_l^p$ is a hyper-Cluj index denoted by $CD_p$ or $CD_p$. The novel hyper-index, $CD_p$, showed good correlation with the boiling points of a selected set of cycloalkanes.

Chan et al. (1998) showed that there exist algebraic connections between various molecular graphs based descriptors. Hosoya’s Z indices, Narumi’s simple topological index (product of all vertex degrees) and connectivity index of Randić exhibited algebraic connections with Wiener’s index. Such connections can be avoided if instead of molecular graph appropriately weighted variants are used.

Lukovits (1998) introduced an all path version of the Wiener index called as Pasareti index ($P$). It was defined as follows:

$$P = \sum_{i<j} \sum_{P \in \pi_{ij}} l(P)$$

Where $l(P)$ denotes the length of the path $P$, i.e., the number of edges in $P$, and where $\pi_{ij}$ is the set of all path connecting the vertices $i$ and $j$. Thus the summations in the above formula embrace all paths contained in $G$. This index is called the Pasareti index because it was derived in the home of the Lukovist that is located in the part of Budapest called Pasaret. Being an exponential function of graph size in terms of the number of its vertexes this index was transformed was into a new variant $V = V(G)$ that was called the Verhalom index:

$$V = P/k$$

Where $k$ is the total number of paths in $G$ divided by $N (N - 1)/2$. The name Verhalom index is given to this distance related index because it was originated in the chemical research center of the Hungarian academy of sciences located in the district Verhalom in Budapest.

Plavsic et al. (1998) proposed Wiener sum of the $D/\Delta$ matrix, the Wiener-sum index ($WS$) which was defined as the half-sum of the off-diagonal elements of the molecular quotient matrix $D/\Delta$:

$$WS = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} [(D)_{ij} / (\Delta)_{ij}]$$
Further, he also suggested another variant of the Wiener-sum index namely *inverse Wiener-sum index* (ws) which was defined as the half-sum of the off-diagonal elements of the molecular inverted quotient matrix $\Delta/D$:

$$ \Delta/D \equiv W_s(\Delta/D) = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} [(\Delta)_{ij}/(D)_{ij}] $$

These both indices have so far been used only for the structure-boiling point modeling of condensed benzenoid hydrocarbons.

Gupta et al. (1999) conceptualized a pendent-distance based graph invariant termed as *superpendentic index* to enhance the role of terminal vertices in [(Q)SAR/SPR] studies. This index is denoted by $\mathcal{P}$ and is defined as square root of the sum of products of non-zero row elements in the pendent matrix.

$$ \mathcal{P}(G) = \left( \sum_{i=1}^{m} \prod_{j=1}^{n} d(v_i, v_j) \right)^{1/2} $$

Where $m$ and $n$ are maximum possible number of $i$ and $j$ respectively.

Galvez et al. (2000) introduced invariants of differences of path lengths (DP$\mathit{s}$), and demonstrated that the overall sum of the inverse of the squares of the differences of distances between all pairs of vertices of the graph is a measure of the mean global kinetic energy of the electrons. These invariants were employed to predict the resonance energies and in the evaluation of biological properties such as antibacterial activities of a wide set of heterogeneous compounds.

Gutman et al. (2000) formulated *multiplicative Wiener index*, $\pi(G)$. This molecular structure descriptor is equal to the product of the distances of all pairs of vertices of the underlying molecular graph i.e,

$$ \pi(G) = \prod_{(i,j) \in \Gamma(G)} d_{ij} $$

Due to the very large numbers that are often reached by $\pi(G)$, its logarithmic version i.e $\ln \pi(G)$, seems to be more appropriate in searching for (Q)SAR/QSPR models (Todeschini and Consonni, 2009).
Espeso et al. (2000) theorized molecular indices $iZ$, based on second decomposition of Hosoya $Z$ matrix i.e $iZ$ matrices. The $iZ$ index of tree was defined by the formula:

$$iZ = \sum_{i<j} (iZ)_{ij}$$

Where the summation goes over all off-diagonal entries $iZ_{ij}$ in the upper triangle of the $iZ$ matrix of $T$. In addition to above the Hosoya hyper index $H$ and a set of molecular indices $iZ$ were defined as the sum of all off-diagonal entries in the upper triangle of the $Z$ matrix and the corresponding $iZ$ matrixes respectively.

Randic et al. (2001) proposed the variable Balaban $J$ index, the "reversed" Balaban index $l/J$ and a novel index $1/J$ derived from $J$ and $1/J$. All the variable descriptors were derived from the distance matrix and the "reversed" distance matrix and were constructed by augmenting the distance matrix by replacing the diagonal zeroes with the variables $x, y, z$ (Randic and Pomp, 2001).

Khadikar et al. (2001a, b) proposed a novel index namely Padmakar-Ivan Index and abbreviated as $PI$ Index. This index is simple to calculate and has disseminating power similar to that of Wiener and Szeged indices. This newly proposed index does not coincide with the Wiener index for acyclic trees. Thus, unlike the Sz index, the $PI$ index is different for acyclic and cyclic graphs. $PI$ index has a discriminating power similar to that of $W$ and Sz index.

Ivanciuc et al. (2001) deduced novel TDs by separating the terms of the Wiener’s polynomial in to even and odd molecular graph distances. The even and odd Wiener’s polynomial sums $WiPolE(x)$ and $WiPolO(x)$ were used as descriptors in (Q)SAR/QSPR models. The sum of Wiener polynomial terms corresponding to even graph distances and the sum of the terms corresponding to odd graph distances were defined as:

$$WiPolE(x) = \sum_{k,even} k^f \cdot x^k$$

$$WiPolO(y) = \sum_{k,odd} k^f \cdot y^k$$

Where $k^f$ is the number of pairs of vertices located at a topological distance equal to $k$, and the summations go up to the maximal distance in the graph; $x$ and $y$ are two independent variable parameters optimized during the modeling procedure.
Randic and Zupan (2001) modified the Wiener index- $W$ to $W^*$ and Hosoya index-$Z$ to $Z^*$. In the modified Wiener index-$W^*$, bond contributions were determined using the reciprocal of the product of the number of atoms on each side of a bond as:

$$W^* = \sum_{e=1}^{n_e} \left( N_{i,e} \right)^{-1} \cdot \left( N_{j,e} \right)^{-1}$$

Where $N_{i,e}$ and $N_{j,e}$ denote the number of vertices on each side of the edge $e$, including vertices $i$ and $j$, respectively, and $n_e$ is the total number of graph edges.

Similarly, Hosoya index $Z$ was modified to $Z^*$ by considering the frequency of occurrence of carbon-carbon bonds in the patterns of disjoint bonds as:

$$Z^*(x_1, x_2, \ldots) = \sum_{k} a(G,k) \cdot x_k$$

Where $x_k$ are integer weights representing the number of times each edge has appeared in all disjoint edge patterns.

Moreover, Randic and Zupan (2001) outlined a general scheme for partitioning of MDs into bond contributions for indices derived from a selection of matrices associated with molecular graphs.

Balaban et al. (2001) derived a general formula for the normalized Wiener’s index of polymers that made the calculation of the graph invariant directly from simple structural information:

$$W = d/ [3(v+o)]$$

Where $v$ is the number of atoms and $o$ is number of rings in the repeating polymer and the topological distance $d$ between the corresponding pairs of equivalent atoms in two neighboring monomer units.

Bonchev (2001) extended the approach of more detailed characterization of molecular topology by accounting for all substructures to the concept of overall distance, $OW(G)$ of a graph $G$. This index could be defined as the sum of distances in all sub graphs of $G$, as well as the sum of $e$th-order terms, $(e)OW(G)$, with $e$ being the number of edges in the sub graph. The overall distance is analyzed as a measure of acyclic and cyclic structures.

Estrada and Molina (2001) introduced novel molecular descriptors based on local spectral moments of the bond matrix. Mathematical expressions relating bond moments to linear combinations of structural fragments were derived. The novel
Patel and Cronin (2001) devised a novel descriptor, the linearity index ($L_i$), with an aim to describe molecular linearity more explicitly. This new index discriminated well linear molecules from nonlinear and branched molecules, regardless of molecular size. This index is based upon ratios of moments of inertia and the molecular weight and can be expressed as:

$$L_i = \sqrt{\frac{IM1(L)/IM2(L)/IM3(L)}{MW^2}} \times 100$$

Where $IM1$, $IM2$, $IM3$ are moments of inertia and $MW$ is molecular weight. Trial and error showed that the square root of the index provided a better discrimination of shape. The whole index was multiplied by a factor of 100 to obtain a more manageable scale.

Mercader et al. (2001) investigated the applications of TDs based on distance and detour distance matrices. They employed some usual TDs based on both these distances to study the heat of formation of a set of hydrocarbons comprising 60 structurally diverse molecules. Surprisingly, TDs based on detour matrix yielded better correlations to predict enthalpies of formation.

Gupta et al. (2002) conceptualized a novel distance based topological index termed as eccentric distance sum index. Eccentric distance sum, denoted by $\xi^{DS}$, can be defined as the summation of product of eccentricity and distance sum of each vertex in the hydrogen suppressed molecular graph of $n$ vertices as per the following equation:

$$\xi^{DS}(G) = \sum_{i=1}^{n} E_i * S_i$$

Where $S_i$ is distance sum of vertex $i$, $E_i$ is eccentricity of vertex $i$ and $n$ is the number of vertices in graph $G$. Eccentric distance sum takes into consideration the eccentricity and distance sum of all vertices in the graph. The eccentricity $E_i$ of a vertex $i$ in a graph $G$ is the path length from vertex $i$ to vertex $j$ that is farthest from $i$ ($E_i=\max d(ij), j\in G$).
Castro et al. (2002) reported some upgraded version of the Wiener's index: a) by using sum of the bond lengths along the shortest path instead of graph theoretical distance; b) by using Euclidean distance between the respective pairs of atoms or c) by using hydrogen filled graphs. However, none of the theoretically justifiable, modifications of the Wiener index improved the applicability and value of these structure- descriptors in designing quantitative structure-property relations. They concluded that the original Wiener index - now already more than half-a-century old-is a much more valuable topological index than one would expect from its extremely simple and seemingly naive definition.

Cao and Yuan (2002) developed a new topological index $VDI_{(±1)}$ to distinguish the cis/trans isomers of cycloalkanes on the basis of Vertex degree–distance index ($VDI$) reported earlier (Cao and Yuan, 2001). They substituted the derivative matrix $S$ with $D_{mod}$ to obtain the $VDI_{(±1)}$. This new structural descriptor showed better QSPR results than with $VDI$.

Ivanciuc et al. (2002) proposed new molecular descriptors computed from the two new molecular matrices, reverse Wiener’s $RW$ and reciprocal reverse Wiener's $RRW$ matrices. The graph structural descriptors computed with the distance $D$, reciprocal distance $RD$, reverse Wiener's $RW$ and reciprocal reverse Wiener’s $RRW$ matrices were used to develop [(Q)SAR/SPR] models for boiling temperature, molar heat capacity, standard Gibbs energy of formation, vaporization enthalpy, refractive index and density.

Hu et al. (2003b) proposed a new variable index, external factor variable connectivity index (EFVCI), in which the atomic attribute was divided into two parts i.e internal and external attributes. Along with atomic attributes, the form of molecular connectivity index was used to get the external factor variable connectivity index:

$$F_{X_p}^{EFVCI} = \sum_{all\ edges} (A_i A_j, ..., A_n)^{-0.5}$$

Where $A_i$ is the attribute of carbon atom perturbed by other atoms.

This kind of index can be regarded as an extension of the molecular connectivity index by using a new atomic attribute, which makes the index flexible to different properties.
Milicevci et al. (2003) extended the Z index to general graphs and investigated the behavior of the Z index in regard to their structural characteristics such as size, branching, cyclicity, multiple edges and loops. The Z index was calculated by means of the Z counting polynomial and the matching polynomial. These polynomials were constructed using the appropriate recurrence relations. The structural behavior of the Z index for simple graphs and general graphs was tested against the total walk count index (twc) and was shown that the Z index of these graphs followed their structural changes, i.e. the Z index increased with the size, loops, multiple edges, cycles and was sensitive to branching.

Narumi (2003) defined new TDs, the bond index $B$ and the connective index $C$, by using the partition function of a graph for analyzing the statisco-mechanical aspect of the Hosoya index. The bond index $B$ was defined as:

$$B(G) = \sum_{k=0}^{m} n(G,k) = S(G,1)$$

Where $n(G,k)$ as the number of ways in which $k$ bonds are chosen from graph $G$.

The connective index $C$ was defined as:

$$C(G) = \sum_{k=0}^{m} q(G,k) = R(G,1)$$

Where $m$ is the maximum number of $k$ for molecular graph $G$.

Duchowicz et al. (2003) calculated the Gibbs free energy for a set of 60 hydrocarbons on the basis of TIs derived from the distance and detour matrices. The distance matrix considers the shortest path between any two vertices whereas the detour matrix considers longest path between any two vertices. The results showed that the TI derived from detour matrix produces better correlation to predict Gibbs free energy. They concluded that the detour matrix represents a suitable topological device to be applied in [(Q)SAR/SPR] analysis.

Yuan and Cao (2003) developed the Edge degree-Distance Index (EDI) and Sum of edges ($S_e$) based on the edge and distance of molecular graph in order to distinguish saturated and unsaturated structures. The EDI was defined as:

$$EDI = \left( \prod_{i=1}^{n} h_i \right)^{1/N}$$
Where \( h_i \) is the elements of vector \( ES (1 \times N) \) obtained by edge degree vector \( E \) multiplying derivative distance matrix \( S \):

\[
ES = (h_1, h_2, \ldots, h_n)
\]

The Sum of edges \( (S_e) \) equals to the half sum of the edge degrees \( (E_i) \) of each vertex in molecular graph as:

\[
S_e = 0.5 \sum_{i=1}^{n} E_i
\]

Yuan and Cao (2003) suggested that the combination of these indices together could represent the molecular structures not only of alkanes but also of alkenes, alkynes, and benzenoid hydrocarbons.

Randic (2004) proposed a novel graph theoretical molecular descriptor known as Wiener-Hosoya index, in view of its structural relationship to both the Wiener number \( W \) and the Hosoya index \( Z \).

Bajaj et al. (2004a) conceptualized a modification of Wiener’s index, termed as Wiener’s topochemical index \( (W_c) \) based on topochemical distance matrix:

\[
W_c = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij/c}
\]

Where \( P_{ij/c} \) is the chemical length of the path that contains the least number of edges between vertex \( i \) and \( J \) in the graph \( G \), \( n \) is the maximum possible number of \( i \) and \( J \). Wiener’s topochemical index was derived from the weighted molecular graph where each vertex was properly weighted with selected chemical/physical property. Wiener’s topochemical index was not only sensitive to the nature, number, and relative position of heteroatom but also exhibited far less degeneracy as compared to Wiener’s index.

Diudea (2006) described a new counting polynomial, called the “Omega” \( \Omega \) \((G,x)\) polynomial on the ground of quasi-orthogonal cut “qoc” edge strips in a bipartite lattice. The Omega \( \Omega \) \((G,x)\) polynomial qoc counting was defined as:

\[
\Omega(G,x) = \sum_{c} m(G,c) \cdot x^c
\]
Where \( m(G, c) \) represents the number of \( qocs \) of length \( c \). The summation runs up to the maximum length of \( qocs \) in \( G \). The polynomial is an elegant form of topological description of lattice graphs. It is related to the well-known PI index.

Lu et al. (2006) introduced a novel TD based on Wiener index, named Lu index for modeling properties of heteroatom and multiple bond containing organic compounds. By considering electronegativity of the atoms and bond length between two adjacent vertices, the Lu index was defined as:

\[
Lu = \sqrt{n} \prod_{k=1}^{g} q_k \log \left( 0.5 \sum_{i=1}^{g} \sum_{j=i}^{g} S_{ij} \right)
\]

Where \( n \) is the number of vertices in a molecular graph; \( q_k \) is the relative electronegativity value of vertex \( k \); \( S_{ij} \) represents the sum of \( v \) power of the relative bond lengths between two adjacent vertices in the shortest path of two vertices \( i \) and \( j \). The usefulness of the Lu index was demonstrated by modeling normal boiling points and molar refractions of aldehydes and ketones, and the retention indices of gas chromatography of hydrocarbons and halogenated hydrocarbons.

Balaban et al. (2007) introduced five new TIs with progressively improved performance. These indices are: (i) a Quadratic index (\( Q \)), which could be defined as the summing the squares of all numbers in the sequence one obtains \( \Sigma \sigma_i^2 \), and by dividing this sum by one plus the cyclomatic number, \( Q = \Sigma \sigma_i^2 / (\mu+1) \). (ii) On summing the square roots of all numbers in the sequence one obtains \( \Sigma \sigma_i^{1/2} / (\mu+1) \). (iii) On dividing terms in this sum by the corresponding topological distances, one obtains the Distance-reduced index \( D = \Sigma_i \{ \sigma_i^{1/2} / [i(\mu+1)] \} \). Two similar formulas define the next two indices, the first one with no square roots i.e. (iv) distance-Attenuated index: \( A = \Sigma_i \{ \sigma_i / [i(\mu+1)] \} \); and (v) the last TI with two square roots: Path-count index: \( P = \Sigma_i \{ \sigma_i^{1/2} / [i^{1/2}(\mu+1)] \} \).

Zhou and Trinajstic (2008) described lower bounds for the Kirchhoff index (Bonchev et al., 1994) in terms of its structural parameters such as the number of vertices (atoms), the number of edges (bonds), maximum vertex degree (valency), connectivity and chromatic number. The bounds of a descriptor are important information of a molecule (graph) in the sense that they establish the approximate range of the descriptor in terms of molecular structural parameters (Zhou and Trinajstic, 2008).
Iranmanesh et al. (2009) defined edge version of well known Wiener index as:

$$W_{e0}(G) = 0.5 \sum_{(e,f) \in E(G)} d_{0}(e,f)$$

Where $d_{0}(e,f) = d_{1}(e,f) + 1$ if $e \neq j$

$$= 0 \quad \text{if } e = j$$

Here, the distance between two edges is the distance between corresponding vertices in the corresponding graph $G$.

Goyal et al. (2010) conceptualized a novel highly discriminating, non-correlating pendenticity based topochemical descriptor termed as pendentic eccentricity index. It can be expressed by the following:

$$\xi^p = \sum_{i=1}^{n} \left[ \prod_{j=1}^{n} \left( P_{ij} / E_i^2 \right) \right]$$

Where $P_{ij}$ is the length of the path that contains the least number of edges between vertex $i$ and vertex $j$ in graph $G$; $E_i$ is the eccentricity of a vertex $v_i$ in $G$ and $n$ is the maximum possible numbers of $i$ and $j$. The predictive ability of this index was demonstrated through development of models for prediction of antitubercular activity of 5'-O-[(N-acyl)sulfamoyl]adenosines.

Mahmiani et al. (2010) introduced the total version of Szeged index as:

$$Sz_T(G) = \sum_{e=uv \in E(G)} \left| T_d(u) \cdot T_d(v) \right|$$

Where $T_d(u)$ is the number of vertices and edges of $G$ which are closer to $u$ than to $v$ and $T_d(v)$ is the number of vertices and edges of $G$ which are closer to $v$ than to $u$. The computation of this novel index was exemplified for some well-known graphs and in particular for zigzag nanotubes.

Diudea et al. (2011) investigated the uniqueness (discriminating ability) of a newly proposed CJN super index using (real) atomic and synthetic structures. This new descriptor distinguished all graphs uniquely and some MDs which are embedded in the super index have shown excellent correlating ability with alkanes properties.

Bruckler et al. (2011) deduced a new class of distance-based molecular structure descriptors i.e $Q$-indices with an aim to eliminate a general shortcoming of
the Wiener and Wiener-type indices, namely that the greatest contributions to their numerical values come from vertex pairs at greatest distance. The Q-index may be represented by the following:

\[ Q = \sum_{[u,v] \in \Gamma(G)} \gamma(u,v) \]

Where \( \gamma(u,v) \) depends solely on the distance \( d(u,v) \) between the vertices \( u \) and \( v \). The \( Q \)-index was also related with the Hosoya polynomial as:

\[ Q(\lambda) = 2H(G, \lambda) \]

The multiplier 2 comes from the fact that each pairwise interaction has been counted twice. Thus Q is an additive function of increments associated with pairs of vertices of \( G \).

**Deng** (2011) introduced a novel variant topological index for molecular graphs, called sum-Balaban index. For a simple and connected graph \( G \) with vertex-set \( V(G) \) and edge-set \( E(G) \), sum-Balaban index was defined as:

\[ SJ(G) = \frac{B}{C+1} \sum_{[i,j] \in \Gamma(G)} (S_i + S_j)^{-0.5} \]

Where \( S_i \) and \( S_j \) are the distance sum of the vertices \( i \) and \( j \) respectively, \( B \) the number of graph edges, and \( C \) the cyclomatic number, that is, the number of rings. The predictive ability of this index was investigated through QSPR modeling of some physiochemical properties of octanes.

**Zhang et al.** (2012a) reported \( q \)-analogs of Wiener index motivated by the theory of hypergeometric series. Some possible chemical interpretations and applications of the \( q \)-Wiener indices were also discussed.

**Doslic and Reti** (2012) investigated discriminating potential of traditional degree-based descriptors and proposed a novel \( T(G) \) index characterized by an improved discriminating potential and reduced degeneracy. The \( T(G) \) index was expressed as:

\[ T(G) = \frac{1}{|V(G)|} \sum_{[u] \in \Gamma(G)} \frac{\min(m_1(u), m_2(u))}{\max(m_1(u), m_2(u))} \leq 1 \]

This index was judged to be more efficient for discriminating between topological structures of molecular graphs than several traditional molecular indices.
Adjacency based graph invariants

These invariants are based on the consideration that the whole set of connections between adjacent pairs of atoms may be represented in a matrix form, termed as adjacency matrix. The simplest number that can be associated with chemical structure is the graph adjacency, \( A(G) \), which is the sum of all entries of the adjacency matrix of the graph. These invariants are based on the consideration that the whole set of connections between adjacent pairs of atoms may be represented in a matrix form, termed as adjacency matrix. However, this simplest topological index is extremely degenerate; it has the same numerical value for all graphs having the same number of edges. A number of attempts have been reported to express the connectivity of atoms in the molecule by more discriminating graph invariants.

Platt (1947) introduced the total adjacency of edges in a graph, known as Platt number or F index. This topological index is simultaneously a measure of the order and dimension of the molecular graph, that is of the size of the molecule and the degree of chain branching. For equal size (equal number of edges) the calculated value of this topological index is higher for the branched molecular graph. It sums over all entries of the edge adjacency matrix and defined by (Platt, 1952):

\[
F(G) = \sum_{i=1}^{A} \sum_{j=1}^{A} E_{ij} = 2N_2 = 2B_1
\]

The entries \( E_{ij} \) are equal to one if edges \( e_i \) and \( e_j \) are adjacent (the two edges thus forming a path of length two) and zero otherwise.

Leonard Spialter (Spialter, 1963; Spialter 1964), was the first to search for a graph invariant which could uniquely characterize the topology of molecular graphs. He introduced atom connectivity matrix (ACM) asserted that the characteristic polynomial of the adjacency matrix \( A(G) \) of a molecular graph \( G \) uniquely determines the topology of the molecule. It was conjectured that such a form might be the characteristic polynomial or the set of its eigenvalues (Spialter, 1964).

Gordon and Scantlebury (1964) proposed Gordon and Scantlebury’s index, also known as connection number \( (N_2) \) or Bertz branching index (BI) is the simplest graph-invariant obtained from edge adjacency matrix which considers both vertices and edges. It is the number of distinct ways a C-C-C fragment may be superimposed on the carbon skeleton (hydrogen suppressed graph). This index is based on the topological distance and is calculated as:
Further, the number of connections $N_2$ of a molecular graph is related to the $1^{st}$ Zagreb index $M_1$ and the quadratic index $Q$ by the following relationships:

$$N_2 = M_1 / 2 - A + 1 = Q + A - 2$$

Where $A$ is the number of atoms.

Morgan (1965) described the concept of extended connectivity, according to which graph vertices are ordered on the basis of their extended connectivity values obtained after a number of iterations of until constant atom ordering is obtained in two consecutive steps. The extended connectivity (or extended vertex degree), denoted as $EC_i$, of a vertex is calculated as the iterative summation of connectivities of all first neighbors as the following:

$$EC^k = \sum_{i=1}^{k} a_{ij} \text{EC}_{ij}$$

Where $a_{ij}$ are the elements of the adjacency matrix, at $k = 0$, the connectivity of each atom is simply the vertex degree $\delta$.

Gutman and Trinajstic (1972) proposed two novel TIs based on the vertex degree ($\delta$) (i.e. equal to the sum of the entries in row $i$ in the adjacency matrix) of the atoms in the H-depleted molecular graph.

The first Zagreb group index $M_1$ was defined as the sum of the squared vertex degrees (rather than a simple sum), whereas the second Zagreb group index $M_2$ is the sum over all edges of the product of the vertex degrees of the pairs of neighboring vertexes:

$$M_1 = M_1(G) = \sum_{i=1}^{n} \delta_i^2$$

$$M_2 = M_2(G) = \sum_{(i,j)} \delta_i \delta_j$$

Where $\delta_i$ is the degree (number of first neighbors) of vertex $i$ of the molecular graph and $\delta_i \delta_j$ is the weight of edge $\{i, j\}$.
The two Zagreb group parameter are strictly related to zero-order $\chi_0$ and first order $\chi^1$ connectivity indices respectively. The 1st Zagreb index $M_1$ (also called Gutman index) is also related to the Platt number $F$ and the connectivity number $N_2$ by the following relationship:

$$M_1 = F + 2(A-1) = 2^*(N_2 + A - 1)$$

Where $A$ is the number of atoms.

**Bonacich (1972)** devised the eigenvector centrality, $EC_i$, of a vertex $i$, which is derived from the leading eigenvector of the adjacency matrix (Bonacich, 1972; Bonacich, 2007). The concept of centrality is related to the ability of a vertex to communicate with other vertices or to its closeness to many other vertices or to the number of pairs of vertices that need a specific vertex as intermediary in their communications. $EC_i$ is defined as the $i$th component of the eigenvector associated to the largest eigenvalue of adjacency matrix:

$$EC_i = \ell_i$$

**Lovasz and Pelikan** (1973) developed Lovasz-Pelikan index, denoted as $\lambda_1^{LP}$, (also known as leading eigenvalue $\lambda_1$) by using eigenvalues of the adjacency matrix as molecular descriptors. This eigenvalue has been suggested as an index of molecular branching, the smallest values corresponding to chain graphs and the highest to the most branched graphs. It is not a very discriminant index because in many cases the same value is obtained for two or more nonisomorphic graphs.

**Randic** (1975) introduced first connectivity index namely Randic connectivity index ($\chi$), also known connectivity index or branching index, by transforming $M_2$ into an inverse square-root function as per following equation:

$$\chi = \sum_{(i,j) \in E(G)} (\delta_i \delta_j)^{-1/2}$$

$\delta_i$ and $\delta_j$ are the vertex degrees of the atoms incident to the considered bond. It is closely related to the 2nd Zagreb index and was proposed as a measure of molecular branching.

**Kier and Hall** (1975, 1976, and 1986) defined a general scheme based on the Randic index to calculate also zero-order and higher-order descriptors, thus obtaining
connectivity indices of $m^{th}$ order, usually known as Kier-Hall connectivity indices. They are calculated as per following:

$$m \chi = \sum_{k=1}^{K} \left( \prod_{a=1}^{n} \delta_a \right)$$

$$0 \chi = \sum_{j=1}^{n} \left( \delta_j \right)^{-1/2}, \quad 1 \chi = \sum_{\text{all edges}} \left( \delta_i \delta_j \right)^{-1/2}, \quad 2 \chi = \sum_{\text{all 2-paths}} \left( \delta_i \delta_j \delta_k \right)^{-1/2}$$

Where $k$ runs over all of the $m^{th}$ order subgraphs constituted by $n$ atoms; $K$ is the total number of $m$th order subgraphs present in the molecular graph. The product is over the simple vertex degrees $\delta$ of all the vertices involved in each subgraph. Molecular connectivity index is simple, considers heteroatoms can be applied to acyclic, cyclic and aromatic molecules, have good relation with biological and physicochemical properties and possesses very high discriminating power (Kier and Hall, 1976; Kier and Hall, 1986; Kier et al., 1975).

Kier and Hall (1981; 1983) made an important extension of their approach by replacing the vertex degrees $\delta$ by the valence vertex degree $\delta^v$ in the above mentioned formulas, in order to propose novel valence connectivity indices. These indices, abbreviated as, $m \chi^v_q$, can account for the presence of heteroatom(s) in the molecules as well as double and triple bonds. Similarly, bond connectivity indices, denoted by $m \chi^b_q$, were also defined by using the bond vertex degree $\delta^b$ instead of simple vertex degree $\delta$ to specially account for multiplicity in the molecule.

Narumi and Katayama (1984) reported a simple topological index, $S$ related to molecular branching and calculated as the product of the vertex degrees, $\delta_i$:

$$S = \prod_{i=1}^{n} \delta_i$$

Where $n$ is the number of atoms.

Kier and Hall (1986) introduced benzene likeliness index (BLI) with an aim to measure the molecule aromaticity. It is calculated by dividing the first order valence connectivity index $\chi^v$ by the number of non-hydrogen bonds in the molecule and then normalizing on the benzene molecule.
Gombar et al. (1987) introduced modified connectivity indices termed as *perturbation connectivity indices*, \( m \chi_y^p \), based on perturbation delta values \( \delta^p \) and is defined as:

\[
m \chi_y^p = \left( \sum_{k=1}^{K} \left( \prod_{j=1}^{n} \delta_j^p \right)^{1/2} \right)_k
\]

Where \( k \) runs over all of the \( m \)th order subgraph of type \( y \) constituted by \( n \) atoms; \( K \) is the total number of \( m \)th order subgraphs. Perturbation delta values are obtained from valence vertex degree \( \delta^v \) by incorporating the effect of atomic environments at topological level.

Burden (1989) presented a method for generating molecular identification numbers of hydrogen-depleted structures from the eigenvalues of a connectivity matrix. The use of the method was proposed for unique numbering of the atoms in a structure and for identifying the atoms that belong to different substructures of disconnected main structure.

Yang et al. (1995) devised new topological indices, \( EA\Sigma \) and \( EA_{\text{max}} \), based upon extended adjacency (EA) matrices of molecules in which influence of factors of heteroatom(s) and multiple bonds were considered. \( EA\Sigma \) is the sum of the absolute eigenvalues of the EA matrix, and \( EA_{\text{max}} \) is the maximum of the absolute eigenvalues of the EA matrix. These indices can be calculated as:

\[
EA\Sigma = \sum_{i=1}^{n} |\lambda_i^{EA}|
\]

\[
EA_{\text{max}} = \max |\lambda_i^{EA}|
\]

These indices possess high discriminating power and correlate well with a number of physicochemical properties and biological activities of organic compounds.

Goel and Madan (1995) proposed a novel numerical graph invariant based on modification of *Randic’s molecular connectivity index* (\( \chi \)) termed as *atomic molecular connectivity index* (\( \chi^A \)), which can be calculated from adjacency topochemical matrix. It is defined as the summation of modified bond values (inverse square root of product of chemical degrees) of adjacent vertices, over all edges in the hydrogen depleted molecular graph, where summation is over all pairs of adjacent vertices of \( G \).
\[ \chi^A(G) = \sum_{i,j} \left( \frac{1}{(\delta_{ic} + \delta_{jc})} \right)^{0.5} \]

Where \( \delta_{ic} \) and \( \delta_{jc} \) are the chemical degrees of the vertices \( i_c \) and \( j_c \). The computation of \( \chi^A \) is conducted in a manner similar to that described by Kier and Hall (1976b) except that the modified valency of each vertex involved in a pair is calculated by summing up relative atomic weights of all the adjacent atoms.

**Estrada** (1995a) conceptualized edge connectivity index, \( \varepsilon \) (or bond connectivity index) a graph-theoretical invariant analogous to the classical Randic connectivity index from the edge adjacency matrix of molecular graph. The edge connectivity index, \( \varepsilon \) can be calculated as:

\[ \varepsilon = \sum_r \left( w^\delta(e_i) \cdot w^\delta(e_j) \right)^{1/2} \]

Where \( w^\delta(e_i) \) and \( w^\delta(e_j) \) are the values of edge-weighted degrees and the summation is over all r-pairs of adjacent edges.

**Estrada and Ramirez** (1996) defined a new topographic or bond order weighted edge connectivity index based on edge adjacency relationships. The elements of edge set were substituted by bond orders or more precisely valence indexes calculated from quantum chemical methods and their calculation is as follows:

\[ \varepsilon^\pi = \sum_r \left( w^\delta \pi(e_i) \cdot w^\delta \pi(e_j) \right)^{1/2} \]

Where \( \pi \) is the bond order.

**Pearlman and Smith** (1998) and **Pearlman** (1999) extended Burden approach to address searching for chemicals on large data bases. **BCUT descriptors (Burden-CAS-University of Texas eigenvalues)** were proposed by utilizing three classes of matrices whose diagonal elements correspond to 1) atomic charge-related values, 2) atomic polarizability-related values, and 3) atomic H-bond abilities.

**Lohninger** (1993) proposed modified Randic index \( \chi_{\text{mod}} \) as the sum of atomic properties, accounting for valence electrons and extended connectivities in the H-depleted molecular graph using Randic connectivity index-type formula as:
\[ \chi_{\text{mod}} = \frac{1}{2} \sum_{i=1}^{A} \sum_{j=1}^{A} \frac{Z_i}{(\delta_i \delta_j)^{1/2}} \]

Where the first sum runs over all the atoms in the molecular graph while the second runs over the first neighbours of the considered atom; \( \delta \) is the vertex degree, and \( Z_i \) is the atomic of the \( i \)th atom.

**Estrada** (1998a, 1999b) described a novel graph theoretical polynomial, \( P_e (G,x) \), counting the edge connectivity in order to account for long-range bond contributions. The *edge-connectivity counting polynomial*, \( P_e (G,x) \) is defined as follows:

\[ P_e(G,x) = \sum_{d(e_i,e_j)} \sum_{d(e_j)} (\delta(e_i) \delta(e_j)) X^{d(e_i,e_j)} \]

Where \( d(e_i,e_j) \) is the distance between edges \( e_i \) and \( e_j \).

**Estrada and Rodriguez** (2000) demonstrated the independence of edge connectivity or bond connectivity indices to other graph theoretical invariants belonging to first, second, and third generation topological indices through QSPR studies.

**Nikolic et al.** (2000) proposed symmetry-modified Zagreb invariants \( M_1 \) and \( M_2 \) by summing up only degrees (SMM\(_1\)) or edge weights (SMM\(_2\)) of symmetry nonequivalent vertices or edges of graphs. On comparing closely related symmetry-independent and symmetry-dependent complexity invariants they produced different ordering. The SMM\(_1\) and SMM\(_2\) indices are closely related to the parent \( M_1 \) and \( M_2 \) indices for \( n \)-alkanes as:

\[
\text{SMM}_1 = \frac{M_1}{2} \quad n = \text{even} \\
= \left( \frac{M_1}{2} \right) + 2 \quad n = \text{odd} \\
\text{SMM}_2 = \left( \frac{M_2}{2} \right) + 2 \quad n = \text{even} \\
= \frac{M_2}{2} \quad n = \text{odd}
\]

**Basak et al.** (2000) considered molecular surface dependent properties (boiling point and gas chromatograph retention times) and molecular volume dependent properties (molar volume and molar refraction). They found that edge connectivity invariants were appropriate for structure-molecular surface properties.
Bonchev (2001a; 2001b) attempted to assess the complexity of molecules and introduced a new topological index namely, topological complexity index (TC) by generalizing the idea of molecular connectivity of Randic, Kier, and Hall. $TC(G)$ of the graph $G$, may be defined as the sum of the total adjacencies of all $^eK$ connected subgraphs having $e$ edges and $n_i$ vertexes of degree $a_i$, including the graph itself which has $e$ edges and $k$ connected subgraphs; summarizing the information on the connectivity of vertexes in all subgraphs, the new index has the meaning of the overall connectivity of $G$:

$$TC(G) = \sum_{e=0}^{\infty} eTC(G) = \sum_{e=0}^{\infty} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \delta_j(G)$$

Where $eTC(G)$ is the topological complexity of order $e$.

Bonchev predicted physicochemical properties of alkane compounds with molecular connectivity and two probe connectivity functions- the overall connectivity versions of the second Zagreb group index, and a derivative inverse function of this index. The favorable comparisons indicate potential of the overall connectivity indices for QSAR/QSPR applications.

Lukovits and Linert (2001) applied a chiral function $F$ satisfying the condition $F(D) = F(L)$, where $D$ and $L$, denote enantiomers of the same structure, in combination with Randic atom-modified first order molecular connectivity index.

Kezele et al. (2002) tested the use of variable connectivity index in QSPR. The index yielded very good regression equations in the case of homogenous sets of molecules.

Ren (2002) derived a novel vertex degree $v_m$ for heteroatom in molecular graph on the basis of the valence connectivity $v$ of Kier-Hall. The atom-type $AI$ index and $Xu$ Index, were modified for compounds with heteroatom by replacing the vertex-degree of hetero-atom by the proposed $v_m$. The modified $Xu$ index and $AI$ index provided QSPR models for the normal boiling points (BP), Molar volumes (MV), molar refractions (MR) and molecular total surface areas (TSA) of alcohols with up to 17 non-hydrogen atoms. These physical properties were expressed as a linear combination of the individual indices related to molecular size and atom-type.
Torrens (2002) optimized a method for determining the permanent of the adjacency matrix, $per(A)$, of fullerenes. The permanent of the adjacency matrix, $per(A)$ is defined as:

$$Per(A) = \sum_{\sigma \in \Lambda_n} \prod_{i=1}^{n} \delta_{i\sigma_i}$$

Where $\Lambda_n$ denotes the set of all possible permutations of (1, 2……n).

The algorithm allows rapid computation of $per(A)$ for adjacency matrix of molecules large enough to be theoretically interesting and concluded that this aspect of chemistry could be useful for designing or predicting unknown fullerenes and their structure.

Quigley and Naughton (2002) correlated various physicochemical properties with the first order connectivity index for a series of prodrug derivatives of 5-fluorocil. The first order connectivity index correlated with the experimental log P values, molecular area and volume of all prodrugs. Comparison of cLog P values provided group contribution values while the comparison of molecular volume data provided information on the volumes of individual substituent relative to that of hydrogen.

Turker (2003a) started with the concept of $T(A)$ graphs for alternant hydrocarbons defined a novel topological index ($L$). The proposed index differentiated isomeric as well as isospectral molecules, encode alternate saturated and unsaturated systems and could be extended to heteroatom containing structures.

Cao et al. (2003) extended the application of eigenvalue of bonding orbital–connection matrix to different physicochemical properties and built the bond adjacency matrix $B_{CH}$ for each C-H bond and orbital overlapping matrix $B_{CC}$ for whole carbon skeleton in alkane molecule on the basis of polarizability effect index (PEI).

Nikolic et al. (2003) amended the original Zagreb indices through insertion of inverse values of the vertex-degrees:

$$^mM_1 = \sum_{all \, \text{vertices}}^n [1/\delta_i] [1/\delta_i]$$

$$^mM_2 = \sum_{all \, \text{vertices}}^n [1/\delta_i\delta_j] [1/\delta_i\delta_j]$$
They concluded that the modified Zagreb $M_1$ index gave a greater contribution to outer atoms than to inner atoms in a molecule. Similarly, the modified Zagreb $M_2$ index gave a greater contribution to outer bonds than to inner bonds in a molecule. This was opposite to the behaviour of the original Zagreb indices and in agreement with the chemists’ understanding that the most important contributions to the interactions between molecules that are essential for many of their physical, chemical, biological and even technological properties arise from the more exposed atoms and bonds.

Lailong and Chengjun (2004) developed a novel connectivity index $F$ on the basis of adjacency matrix and edge valency. This index reflects the chemical bond specificity of edge $i$.

$$ F = \sum (f_i, f_j, f_k, \ldots)^{0.5} $$

Where $f_i$ is the edge valency.

Bajaj (2005) refined Zagreb indices $M_1$ and $M_2$. These topochemical indices were based on topochemical adjacency matrix. The refined indices were sensitive to both the presence as well as position of the heteroatom(s) in the molecule. Zagreb topochemical indices $M_1^c$ is defined as the summation of the squares of chemical degrees over all the vertices in hydrogen suppressed molecular graph. Zagreb topochemical indices $M_2^c$ is defined as the summation of chemical weights of all edges in hydrogen suppressed molecular graph.

$$ M_1^c (G) = \sum_{i=1}^{n} (\delta_i^c)^2 $$

$$ M_2^c (G) = \sum_{ij} (\delta_i \cdot \delta_j) $$

Where $\delta_i$ is the chemical degree (number of first neighbors) of vertex $i$ of the molecular graph and $\delta_i \cdot \delta_j$ is the weight of edge $\{i, j\}$.

Zhang et al. (2007) characterized DNA by a numerical sequence by considering positions of bases and the pairs of bases in DNA. For generating the sequence invariants, the following function is used:

$$ d_r = m/n $$
Where $d_r$ is defined as the relative position parameter of a base; $m$ represents the position of this base in a sequence, $n$ represents the number of all bases in this sequence. They extracted a novel invariant (molecular connectivity index type) from the derived numerical sequences.

**Mu et al.** (2008) devised novel molecular connectivity index, denoted as $m \chi'$ based on the adjacency matrix of molecular graphs. By using delta value ($\delta'_i$) instead of the original delta value ($\delta_i$) of the molecular connectivity index, this new index was obtained.

$$m \chi'(G) = \sum_{j=1}^{n} \left( \prod_{i=1}^{m+1} \delta'_i \right)^{-1/2}$$

Where $m$ is the order of the molecular connectivity index. This index was successfully applied to predict the molar diamagnetic susceptibilities of organic compounds. Later on the converse index, denoted by $m \chi''$ was also proposed by Mu et al. (2009) as:

$$m \chi''(G) = \sum_{j=1}^{n} \left( \prod_{i=1}^{m+1} \delta'_i \right)^{y}$$

Where $y$ is a variable, whose optimal value can be found by the optimization method.

**Vukicevic and Furtula** (2009) constructed a novel TD based on the end-vertex degrees of edges called as ‘geometric-arithmetic’ (GA) index:

$$GA(G) = \sum_{u \neq v} \sqrt{\delta_u \delta_v} / (\delta_u + \delta_v) / 2$$

Where summation goes over all edges of graph $G$, and $\delta_u$ and $\delta_v$ are the degrees of vertices that are connected with edge $uv$. The name ‘geometric-arithmetic’ comes from the fact that $\sqrt{\delta_u \delta_v}$ and $(\delta_u + \delta_v) / 2$ are the geometric and arithmetic means, respectively, of the numbers $\delta_u$ and $\delta_v$. The predictive ability of this index was found to be better than Randic connectivity index $\chi$ in modelling some physico-chemical properties of octanes.
Furtula et al. (2010) defined augmented Zagreb index, AZI of molecular graph $G$ as:

$$AZI(G) = \sum_{ij \in E(G)} (\frac{\delta_i \delta_j}{\delta_i + \delta_j - 2})^3$$

Where $E(G)$ is the edge set of $G$, and $\delta_i$ and $\delta_j$ are the degrees of the terminal vertices $i$ and $j$ of edge $ij$. Some tight upper and lower bounds were also reported for the AZI index of a chemical tree.

Andova and Petrushevski (2011) devised variable Zagreb indices (denoted by: $^2M_1$ and $^2M_2$ in accordance with Karamata’s Inequality. The first and second variable Zagreb indices were defined as:

$$^2M_1(G) = \sum_{i=1}^{n} \delta_i^2,$$

$$^2M_2(G) = \sum_{(i,j)} \delta_i \delta_j,$$

Doslic et al. (2011) defined the average neighbor degree number, an invariant useful for measuring the diversity of vertices in molecular graph $G$, as:

$$\langle m_{avg} (G) \rangle = \frac{1}{|V(G)|} \sum_{u \in V(G)} m_{avg}(u)$$

Where $m_{avg}(u)$ is the average of degree of vertices adjacent to $u$.

The application of $m_{avg}(G)$ was investigated on the benchmark set of 18 octane isomers and found a decent correlation between $m_{avg}(G)$ and enthalpy of vaporization, as well as the standard enthalpy of vaporization.

Ghorbani and Hosseinzadeh (2012) reported new version of Zagreb indices as:

$$^\varepsilon M_2 = \varepsilon M_2 (G) = \sum_{ij \in E(G)} \varepsilon \delta_i \varepsilon \delta_j$$

$$^\varepsilon M_1 = \varepsilon M_1 (G) = \sum_{j \in V(G)} \varepsilon \delta_j^2$$

$$^\varepsilon M_2 = \varepsilon M_2 (G) = \sum_{ij \in E(G)} \varepsilon \delta_i \varepsilon \delta_j$$

Where $^\varepsilon \delta_i$ is the largest distance between $i$ and any other vertex $j$ of $G$. 
Ghorbani et al. (2012) modified Narumi-Katayama index as \( NK^* \) in which each vertex degree \( \delta_i \) is multiplied \( \delta_i \) times. They determine its basic properties and characterize graphs extremal with respect to it. This new version of \( NK \) index may be represented as:

\[
NK^* = NK^*(G) = \prod_{i=1}^{n} \delta_i^\delta_i
\]

**Adjacency-cum-distance based graph invariants**

These indices employ distance matrix as well as adjacency matrix to characterize a molecular graph. Since these indices combine the information of both adjacency matrix and distance matrix, hence, are expected to contain more considerable topological information than the other TIs derived from only single matrix.

Galvez et al. (1994) introduced new topological charge indices \( G_k \) and \( J_k \). These evaluate the charge transfer between pairs of atoms, and therefore the global charge transfer in the molecule. They defined the \( M \) matrix as:

\[
M = A \times D^*
\]

Where \( A \) is the adjacency matrix and \( D^* \) is the inverse square distance matrix.

The charge term matrix can be defined as

\[
CT_{ij} = m_{ij} - m_{ji},
\]

Where \( m \) stands for the elements of the \( M \) matrix. The diagonal entries of the matrix \( CT \) represent the topological valence of the atoms and the off diagonal entries \( CT_{ij} \) represent a measure of the net charge transferred from the atom \( j \) to the atom \( i \).

For each path of length \( k \), a topological charge index \( G_k \) was defined as:

\[
G_k = \sum_{i=N-1, j=N}^{i=N-1, j=N} \left| CT_{ij} \right| \delta(d_{ij}, D_{ij})
\]

Where \( d_{ij} \) is the topological distance between \( i\)th and \( j\)th atoms; \( \delta(d_{ij}; k) \) is a Kronecker delta function equal to 1 if \( d_{ij} = k \), zero otherwise.

The mean topological charge index \( J_k \) was defined as:

\[
J_k = G_k / (N-1)
\]

Where, the denominator \( N-1 \) is the number of edges in an acyclic molecule.

The ability of the \( G_k \) and \( J_k \) for the description of the molecular charge distribution was established by correlating them with dipole moment of a
heterogeneous set of hydrocarbons, as well as with the normal boiling temperature of alkanes and alcohols and the vaporization enthalpy of alkanes. The charge distributions descriptive ability of $G_k$ and $J_k$ was found to be higher than that shown by the molecular connectivity index and the Wiener’s index (Galvez et al., 1994).

Sharma et al. (1997) conceptualized an adjacency cum distance based topological index termed as eccentric connectivity index. It is defined as the sum total of the product of eccentricity and degree of each vertex in hydrogen suppressed molecular graph having $n$ total vertices, that is:

$$
\xi^c = \sum_{i=1}^{n} (E_i * V_i)
$$

Where $E_i$ is the eccentricity and $V_i$ is the degree of vertex in graph $G$. For a molecular graph $(G)$ if $V_1, V_2, ..., V_n$ are its vertices, the number of first neighbors of a vertex $V_i$ is the degree of this vertex and is denoted by deg $(v_i)$. The distance $d (V_i, V_j \mid G)$ between the vertices $V_i$ and $V_j$ of $G$ is length of the shortest path connecting $V_i$ with $V_j$. While eccentricity $E_i$ of a vertex $V_i$ in $G$ is the length of shortest path from $V_i$ to the vertex $V_j$ that is farthest from $V_i$ ($E_i = \max d (V_i, V_j \mid G)$). Its discriminating power was investigated with regard to physical/biological properties of molecules.

Ren (1999) developed a novel topological index based on adjacency matrix and distance matrix. It was denoted as $Xu$ index and was claimed to have high discriminant power particularly for molecular size and branching. It is defined as:

$$
Xu = n^{1/2} \log L = n^{1/2} \log (\sum_{i} v_i s_i^2 / \sum_{i} v_i s_i)
$$

Where $n$ is the number of atoms and $L$ represents the valence average topological distance calculated by vertex degree $v_i$ and vertex distance degree $S_i$ of all the atoms.

Gupta et al. (2000) conceptualized and developed adjacency-cum-path length based topological descriptor termed as connective eccentricity index ($C^\xi$). The utility of connective eccentricity index in structure-activity studies was investigated through development of models for prediction of anti-hypertensive activity of $N$-benzylimidazole derivatives. It is defined as the summation of the ratios of the degree of
a vertex \((V_i)\) and its eccentricity \((E_i)\) for all vertices in the hydrogen suppressed molecular graph as per the following equation:

\[
C^\xi = \sum_{i=1}^{n} \left( V_i / E_i \right)
\]

**Gupta et al. (2001)** proposed a novel adjacency-cum-distance based topological index called as *eccentric adjacency index*. This index was successfully used for the development of a model for the prediction of anti-HIV activity of HEPT derivatives and physicochemical properties of \(\beta\) blockers. New additive adjacency matrix was constructed for the calculation of this index. The eccentric adjacency index \(\xi^A\) is given by:

\[
\xi^A = \sum_{i=1}^{n} \left( S_i / E_i \right)
\]

Where \(S_i\) is summation of valence values of all the vertices adjacent to the concerned vertex in a hydrogen suppressed molecular graph, \(E_i\) is the eccentricity of the vertex \(i\), \(n\) is the number of vertices in graph \(G\).

**Torrens** (2001) presented a new charge transfer index, *vector semisum charge-transfer index* \((\mu_{vec})\) by modifying the charge transfer indices of Galvez et al. (1994). The \(\mu_{vec}\) was defined in terms of \(\mu\) as:

\[
\mu_{vec} = N(\mu) = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2}
\]

The molecular dipole moment \(\mu\) was defined as the sum of the edge \(\mu_{ij}\).

\[
\mu = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mu_{ij} = 1/2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} A_{ij} CT_{ij}
\]

The ability of the indices for the description of the molecular charge distribution was established by comparing them with the dipole moment of a heterogeneous set of 57 polar and 53 apolar hydrocarbons.

**Ren** (2002) utilized novel atomic level *A1 topological indices* based on the adjacency matrix and distance matrix of a graph to code the structural environment of
each atomic type in a molecule. The topological index for any atom type \( i \) in molecular graph, \( AI_i \), was defined as:

\[
AI_i = 1 + \Phi
\]

\[
\Phi = \frac{v_i s_i^2}{\sum v_i s_i}
\]

Where parameter \( \phi \) is considered as a perturbing term of \( ith \) atom, reflecting the effects its structural environment on its \( AI_i \) value; \( v_i \) is the vertex degree and \( s_i \) is the distance sum. The efficiency of Xu index and A1 index was verified by high quality QSPR/QSAR models obtained for several physical properties and biological activities of several data sets of alcohols with a wide range of non-hydrogen atoms.

**Sardana et al.** (2002) conceptualized a novel adjacency-cum-distance based topological index termed as adjacent eccentric distance sum index and investigated its discriminating power with regard to the anti-HIV activity of TIBO derivatives by developing suitable models. The proposed index offered a vast potential for structure-activity/property studies. This index is defined as the summation of the values of the product of distance sum and eccentricity and divided by the degree of the corresponding vertex for all the vertices in the hydrogen-suppressed molecular graph.

\[
\xi^{SV} = \sum_{i=1}^{n} S_i E_i / V_i
\]

Where \( S_i \) is the distance sum of the vertex \( i \), \( E_i \) is the eccentricity of vertex \( i \), and \( V_i \) is the degree over all vertices.

**Quigley and Naughton** (2002) modified the eccentric adjacency index (Gupta *et al.*, 2001) and proposed valence eccentricity adjacency index by substituting simple connectivity value with vertex valence value. The valence eccentric connectivity index can be easily calculated from additive valence adjacency and distance matrices by using following equation:

\[
\xi^{v} = \sum_{i=1}^{n} \left( S_i^v / E_i \right)
\]

Where \( S_i^v \) is the sum of vertex valences and \( E_i \) is eccentricity of vertex \( i \).
The vertex valences (incorporating the superscript $v$ to allow for calculations involving multiple bonding and heteroatoms) were defined as follows:

$$\delta_i^v = Z_i^v - h_i$$

Where $Z_i^v$ is the number of valence electrons of the vertex (atom) and $h_i$ is the number of hydrogen atoms attached to it.

Quigley and Naughton (2002) also derived a further index ($\Delta \xi$) in a manner analogous to the differential molecular connectivity index ($\Delta \chi$). This differential eccentric adjacency index ($\Delta \xi$) was expressed as:

$$\Delta \xi = \xi - \xi^v$$

They envisaged that this index will be useful in encoding further information which may be employed in structure/activity studies. Moreover, the intercorrelation between a series of physicochemical parameters and TDs for a set of β-blockers was also investigated.

Gupta et al. (2003) developed three novel eccentric-adjacency topochemical indices ($\xi_1^A$, $\xi_2^A$, and $\xi_3^A$) and concluded that the values of eccentric-adjacency topochemical indices change as connectivity of molecular structure increases from linear to cyclic molecule and branched acyclic to connected cyclic isomer.

Ren (2003) investigated the quantitative correlation between the Kovats retention indices of 33 aldehydes and ketones and their atom-type-based A1 topological indices on four stationary phases (HP-1, HP-50, DB-210, and HP-Innowax) at 50°C using four separate linear equations. The results of their studies indicated the potential of the A1 indices in structure retention index correlation.

Kumar et al. (2004) introduced reformed eccentric connectivity index to overcome its limitations owing to degeneracy, and its insensitivity towards the presence and relative position of heteroatom(s). The reformed eccentric connectivity index, termed as eccentric connectivity topochemical index ($\xi_c^A$) overcomes the limitations of eccentric connectivity index by exhibiting very low degeneracy and displaying sensitivity to both the presence and relative position heteroatom(s) without compromising with the discriminating power of eccentric connectivity index. It is defined as the summation of the product of chemical eccentricity and the chemical
degree of each vertex in the hydrogen suppressed molecular graph having \( n \) vertices, that is:

\[
\xi^c = \sum_{i=1}^{n} \left( E_{ic} \ast V_{ic} \right)
\]

Where \( V_{ic} \) is the chemical degree of vertex \( i \), \( E_{ic} \) is the chemical eccentricity of the vertex \( i \) and \( n \) is the number of the vertices in graph \( G \). The values of eccentric connectivity topochemical index were calculated from topochemical distance matrix \((D_c)\) and topochemical adjacency matrix \((A_c)\). This novel high discriminating index was used for the prediction of anti-HIV activity of 2,3-diaryl-1,3-thiazolidin-4-ones with a high accuracy of prediction.

Bajaj et al. (2004b) derived a highly discriminating topochemical descriptor, termed as superadjacency topochemical index. Superadjacency topochemical index is defined as the sum of the products of the concerned vertex chemical degree and the sum of adjacent vertex chemical degrees divided by the chemical eccentricity of the concerned vertex, over all the vertices in the hydrogen suppressed molecular graph. It is expressed as:

\[
j^{AC}(G) = \sum_{i=1}^{n} \frac{\deg(v_i) \ast S_{ic}}{E_{ic}}
\]

\[
S_{ic} = \sum \deg(V_{ic})
\]

Where \( S_{ic} \) is the sum of chemical degrees of all vertices \((V_j)\), adjacent to vertex \( i \) and \( n \) is the number of vertices in graph \( G \). The discriminating power of superadjacency topochemical index was found to be far superior when compared with that of distance based Wiener's index and adjacency based molecular connectivity index.

Bajaj et al. (2006) conceptualized a highly discriminating topological descriptor, termed as augmented eccentric connectivity index, and investigated it's discriminating power with regard to anti-HIV activity of 2-pyridinone derivatives. Augmented eccentric connectivity index defined as the summation of the quotients of the product of adjacent vertex degrees and eccentricity of the concerned vertex, for all vertices in the hydrogen suppressed molecular graph. It is expressed as:
\[ A_{EC} = \sum_{i=1}^{n} \left( \frac{M_i}{E_i} \right) \]

Where \( M_i \) is the product of degrees of all vertices \((v_j)\) adjacent to vertex \(i\), \( E_i \) is the eccentricity, and \( n \) is the number of vertices in graph \(G\). The discriminating power of the augmented eccentric connectivity index was found to be far superior when compared to that of the distance based Wiener's index and adjacency based molecular connectivity index. This index did not exhibit any degeneracy amongst all possible structures with three, four and five vertices.

Dureja et al. (2008a) proposed three new-generation, highly discriminating topochemical descriptors, termed as: superaugmented eccentric connectivity topochemical index-1 \( \left( \frac{S_{EC}}{\xi_1} \right) \), superaugmented eccentric connectivity topochemical index-2 \( \left( \frac{S_{EC}}{\xi_2} \right) \), and superaugmented eccentric connectivity topochemical index-3 \( \left( \frac{S_{EC}}{\xi_3} \right) \). These can be defined as the summation of the quotients of the product of adjacent vertex chemical degrees and powered chemical eccentricity of concerned vertex, for all vertices in the hydrogen suppressed molecular graph. It is expressed as:

\[ S_{EC} = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{M_{ic}}{E_{ic}^N} \right) \]

Where \( M_{ic} \) is the product of chemical degrees of all the vertices \((v_j)\), adjacent to vertex \(i\), \( E_i \) is the chemical eccentricity of concerned vertex, and \( n \) is the number of vertices in the graph. \( N \) is equal to 2, 3, 4 for \( \left( \frac{S_{EC}}{\xi_1} \right) \), \( \left( \frac{S_{EC}}{\xi_2} \right) \), \( \left( \frac{S_{EC}}{\xi_3} \right) \) respectively.

Dureja et al. (2009) defined four novel TDs termed as superaugmented pendent indices \( \left( \frac{S^p}{\xi_1} \right) \), \( \left( \frac{S^p}{\xi_2} \right) \), \( \left( \frac{S^p}{\xi_3} \right) \), \( \left( \frac{S^p}{\xi_4} \right) \) as the summation of quotients, of the product of non-zero row elements in the pendent matrix and product of adjacent vertex degrees; and \( N^{th} \) power eccentricity of the concerned vertex, for all the vertices in a hydrogen suppressed molecular graph. It can be expressed as:

\[ S^p = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{P_{(ij)} \cdot M_i}{E_i^N} \right) \]

Where \( P_{(ij)} \) is the length of the path that contains the least number of edges between vertex \(i\) and vertex \(j\) in graph \(G\); \( M_i \) is the product of degrees of all vertices
adjacent to vertex \( i \), The eccentricity \( E_i \) of a vertex \( v_i \) in \( G \) is the path length from vertex \( v_i \) to the vertex \( v_j \) that is farthest from \( v_i \) (\( E_i = \max d(v_i,v_j) \mid G \)) and \( n \) is the maximum possible numbers of \( i \) and \( j \). The \( N \) is 1, 2, 3 and 4 in case of \( S_{\xi}^{4-1} \), \( S_{\xi}^{4-2} \), \( S_{\xi}^{4-3} \) and \( S_{\xi}^{4-4} \) respectively. These descriptors exhibited high sensitivity towards branching, high discriminating power and extremely low degeneracy.

**Dutt and Madan** (2010) proposed new generation *superaugmented eccentric connectivity indices* (denoted by: \( S_{\xi}^{c} 4_4 \), \( S_{\xi}^{c} 5_5 \), \( S_{\xi}^{c} 6_6 \), and \( S_{\xi}^{c} 7_7 \)) along with their topochemical versions (denoted by: \( S_{ac}^{c} 4_4 \), \( S_{ac}^{c} 5_5 \), \( S_{ac}^{c} 6_6 \), and \( S_{ac}^{c} 7_7 \)) for the purpose of (Q)SAR/QSPR modeling. These indices can be expressed as:

\[
S_{\xi}^{c} \left( \sum_{i=1}^{n} \left( \frac{M_i^2}{E_i^N} \right) \right)
\]

Where \( M_i \) is the product of degrees of all the vertices \( (v_j) \), adjacent to vertex \( i \); \( E_i \) is the eccentricity; \( n \) is the number of vertices in the graph and the \( N \) is equal to 1,2,3 and 4 for \( S_{ac}^{c} 4_4 \), \( S_{ac}^{c} 5_5 \), \( S_{ac}^{c} 6_6 \), and \( S_{ac}^{c} 7_7 \) respectively. The \( S_{ac}^{c} 7_7 \) exhibited exceptionally high discriminating power of the order of \( >300000 \) for all possible structures containing only five vertices. High discriminating power amalgamated with negligible degeneracy offered proposed indices a vast potential for use in the characterization of structures, similarity/dissimilarity studies, lead identification, lead optimization, combinatorial library design and quantitative structure-activity/structure-property relationships for the prediction of various physicochemical, biological and toxicological properties so as to facilitate drug design.

**Ediz** (2010) defined a modified version of ECI, called as *Ediz eccentric connectivity index* as:

\[
E_{\xi}^{c}(G) = \sum_{v \in V} \frac{S_v}{E_i(V)}
\]

Where \( S_v \) is the sum of degrees of all vertices \( u \), adjacent to vertex \( v \); \( E_i(v) \) is the eccentricity of \( v \). The calculation of this index was demonstrated for nanostar dendrimers (Ediz, 2010; Ediz, 2011).
Goyal et al. (2011) proposed four refinements of eccentric distance sum topochemical index termed as augmented eccentric distance sum topochemical indices 1-4 (\(\xi_{1c}^{ADS}, \xi_{2c}^{ADS}, \xi_{3c}^{ADS}\) and \(\xi_{4c}^{ADS}\)) so as to significantly augment discriminating power and to reduce degeneracy. These indices were defined as:

\[
\begin{align*}
\xi_{1c}^{ADS} &= \sum_{i=1}^{n} E_{ic}^2 * S_{ic}, \\
\xi_{2c}^{ADS} &= \sum_{i=1}^{n} E_{ic} * S_{ic}, \\
\xi_{3c}^{ADS} &= \sum_{i=1}^{n} (E_{ic} * S_{ic})^2, \\
\xi_{4c}^{ADS} &= \sum_{i=1}^{n} (E_{ic} * S_{ic})^3
\end{align*}
\]

Where \(S_{ic}\) is chemical distance-sum of vertex \(i\), \(E_{ic}\) is chemical eccentricity of vertex \(i\) and \(n\) is the number of vertices in graph \(G\). These indices were successfully utilized for developing models for prediction of anti-tumor activity of bisphosphonates.

Das and Trinajstić (2011) compared the relationship between ECI and Zagreb indices (\(M_1\) and \(M_2\)) for chemical trees. Besides chemical trees, molecular graphs were also treated and the value of ECI was found greater than Zagreb index-\(M_1\) for diameter greater than or equal to 7.

Gupta et al. (2012) conceptualized highly discriminating superaugmented eccentric distance sum connectivity indices as fourth generation TDs. The topochemical versions of these indices (denoted by: \(c_{cN}^{SED_{1c}}, c_{cN}^{SED_{2c}}, c_{cN}^{SED_{3c}}\) and \(c_{cN}^{SED_{4c}}\)) was expressed by the following:

\[
\begin{align*}
S_{cN}^{c_{cN}^{SED}} &= \sum_{i=1}^{n} \left( \frac{M_{ic}}{E_{ic}^N * S_{ic}^2} \right)^{-1}
\end{align*}
\]

Where \(M_{ic}\) is the product of chemical degrees of all the vertices (\(v\)), adjacent to vertex \(i\); \(E_{ic}\) is the chemical eccentricity; \(S_i\) is the chemical distance sum of vertex \(i\) and \(n\) is the number of vertices in the graph and the \(N\) is equal to 1,2,3,4 for superaugmented eccentric distance sum connectivity topochemical indices-1, 2, 3, 4 (denoted by: \(c_{cN}^{SED_{1c}}, c_{cN}^{SED_{2c}}, c_{cN}^{SED_{3c}}\) and \(c_{cN}^{SED_{4c}}\)). These TDs were successfully employed for development of numerous models for Chk2 inhibitory activity of 2-arylbenezimidazoles through decision tree, random forest and moving average analysis.
Ediz (2012) proposed another modified version of ECI, called as *reverse eccentric connectivity index (REEC)* as:

$$ REE(G) = \sum_{v \in V} \frac{E_i(v)}{S_v} $$

Where summation goes over all vertices of graph $G$; $E_i(v)$ denotes the eccentricity of $v$ and $S_v$ is the sum of degrees of all vertices adjacent to vertex $v$. The predictive power of this index was demonstrated on some physico-chemical properties of octanes. In addition, basic mathematical properties in terms of lower and upper bounds were also investigated.

Dutt and Madan (2012b) conceptualized and developed four novel MDs termed as *superpendentic eccentric distance sum indices 1-4* (denoted by: $EDS_{p-1}, EDS_{p-2}, EDS_{p-3}$ and $EDS_{p-4}$) along with their topochemical counterparts (denoted by: $PEDS_{p-1}, PEDS_{p-2}, PEDS_{p-3}$ and $PEDS_{p-4}$). The topochemical version of these indices can be expressed as:

$$ EDS_{p-1}^{N} = \sum_{i=1}^{n} \left( \frac{P_{icjc} * M_{ic}}{E_{ic}^{N} * S_{ic}} \right) $$

Where $P_{icjc}$ is the chemical length of the path that contains the least number of edges between vertex $i$ and vertex $j$ in graph $G$. $M_{ic}$ is the product of chemical degrees of all vertices ($v_j$), adjacent to vertex $i$. $E_{ic}$ is the chemical eccentricity of concerned vertex $v_i$, $S_{ic}$ is the chemical distance sum of vertex $i$ and $n$ is the number of vertices in the hydrogen depleted graph. $N$ is equal to 1,2,3,4 for $EDS_{p-1}$, $EDS_{p-2}$, $EDS_{p-3}$ and $EDS_{p-4}$ respectively. The utility of proposed TDs was investigated through development of models for the prediction of hCRF-1 binding affinity of substituted pyrazines using decision tree and moving average analysis.

**Centric graph invariants**

The concept of graph center is based on molecular topological distances between the graph vertices. The center vertices have the smallest maximal distance to other vertices.

$$ d_{ij}^{max} = \min \text{ for } j = 1, 2, \ldots, p $$
Invariants derived from the concept of center are called centric graph descriptors and were proposed to quantify the degree of compactness of molecules by distinguishing between molecular structures organized differently with respect to their centres. The centric indices are mainly divided into two groups: indices proposed by Balaban and the indices proposed by Bonchev.

**Balaban** (1979) proposed a set of five graph invariants classified as centric invariants on the basis of sequences of numbers obtained by pruning an acyclic graph. By pruning stepwise all vertices of degree one ($\delta_i$), a vertex (center) or an edge connecting two adjacent vertices (bicenter) is obtained. Balaban developed a *Balaban centric index* ($B$), which is defined as:

$$B = \sum \delta_i^2$$

*Balaban centric index* ($B$) provides a measure of molecular branching: the higher the value of $B$, the more branched the tree. It is known as centric index because it reflects the topology of the tree as viewed from the centre. Four invariants were devised from $B$ and $M_1$ to differentiate branching from number of vertices by normalization and binormalization. Normalized and binormalized centric ($C, C'$) and quadratic invariants ($Q, Q'$) were defined as:

$$C = \frac{B - 2n + U}{2}, \quad C' = \frac{B - 2n + U}{(n - 2)^2 - 2 + U},$$

$$Q = 3V_4 + V_3, \quad Q' = \frac{2(3V_4 + V_3)}{(n - 2)(n - 3)},$$

Where $n$ is the number of vertices. $U = [I - (-1)^N]/2$, while $V_3$ and $V_4$ are vertices of degree three and four respectively.

A topological index is normalized by imposing for all graphs (regardless of $n$) the same lower bound, equal to zero for the least branched (linear) tree on all the graphs. Normalized indices, which are integers when they are derived from integer initial indices, still show a variation with $n$ and with shape and therefore topological indices are binormalized by imposing, in addition to the above lower limit, the same upper bound equal to one from the chain and star graph. In order to find the normalized quadratic index one is required to find the quadratic function of the general form.
It was found that the centric invariants parallel the ordering induced by index B, while the quadratic invariants induced ordering which parallel those due to Gutman et al’s index $M_1$ and Gordon-Scantlebury index $N_2$.

Bonchev et al. (1980) described a more complete definition of graph center called polycenter and generalized this concept of center for any connected acyclic or cyclic graph on the basis of topological distance matrix using four criteria in a specified order. The center points provide the smallest distance to any other vertex; the smallest sum of distances to all other vertices; the smallest number of times the maximum distance occurs in the distance code and constancy on repeating above three criteria to the pseudo center graph containing only the vertices selected by the previous criteria and their incident edges. The centric invariants were proposed for correlations, for estimating the isomeric chemical structures and for use in coding and computer processing of chemical structures.

The indices termed as Bonchev centric information indices were derived from distance matrix $D$ and edge distance matrix $E_D$.

Diudea et al. (1991) introduced the so-called $B$-matrix for the development of a new index by counting the vertexes in graph spheres (layers). A sphere is a list of atoms surrounding a central vertex in a given topological distance. Using spheres is advantageous in studies that investigate the influence of neighbours on a specific property of central vertex.

Diudea (1994) distinguished the sequence matrices ($SM$) and layer matrices ($LM$). A sequence matrix is a collection of walks starting from the vertices $i$ to all other $n - 1$ vertices in $G$. While a layer matrix collects the properties of vertices $u$ located in concentric shells (layers) at distance $j$ around each vertex $i$ in $G$. Further he defined two invariants based on LM matrices- centrocomplexity $x$ ($LM$) and centric invariant $c$ ($LM$).

\[
x \ [LM]_i = \left[ \frac{1}{\rho_i} \sum_{j=0}^{\xi} (lm_{ij})^{\frac{j}{\rho_i}} \right]^{-1}
\]

\[
c \ [LM]_i = \left[ \sum_{j=0}^{\xi} (lm_{ij})^{j/d_{ij}} \right]^{-1}
\]
Where $z$ is the number of digits of the max $\text{lm}_{ij}$-value in graph, $\rho_i$ is a topological property (i.e., vertex degree $k_i$, but here $\rho_i = 1$ and the exponent is +1), $\xi_i$ is the eccentricity of vertex $i$ and $\text{dsp}$ is a specified topological distance, usually larger than the diameter of graph (here $\text{dsp} = 10$, unless otherwise specified).

**Balaban** (1995) proposed regressive decremental distance sums to obtain greater discrimination between the terminal and central vertices. They are calculated from the distance sum layer matrix $LDS$ by the following:

$$x \ [LDS]_i = S_i - \sum_{j=0}^{z} \text{lm}_{ij} \cdot 10^{-j}$$

Where $S_i$ is the distance sum of the $i$th vertex. In this way, the progressively attenuated contributions due to more distant vertices are subtracted from the distance degree of the focused vertex.

**Information theory based graph invariants**

Information theory has been used in chemical graph theory for describing chemical structures and for providing good correlations between physico-chemical and structural properties. Information indices are constructed for various matrices and also for some topological indices. The advantage of such kind of indices is that they may be used directly as simple numerical descriptors in a comparison with physical, chemical or biologic parameters of molecules in structure property and activity relationships. It can also be noted that information indices normally have greater discriminating power for isomers than the respective topological indices.

**Shannon and Weaver** (1949) showed that the statistical concept of entropy can be extended beyond the thermodynamics and applied to the process of transmitting information. The basic **Shannon’s formula** to measure entropy of information in bits can be expressed as:

$$H = n \log_2 n - \sum_{i=1}^{n} n_i \log_2 n_i$$

Where $n_i$ is the probability of randomly selecting an element of the $i$th class.
One of the major consequences of Shannon’s theory was the radically new idea of viewing the structure of any kind as a communication. This study was one of the founding works in the field of information theory.

Rashevsky (1955) was the first to calculate the information content of graphs where “topologically equivalent” vertices were placed in the same equivalence class. In Rashevsky approach, two vertices $u$ and $v$ of a graph are said to be topologically equivalent if and only if for each neighboring vertex $u_i (i=1,2,\ldots,k)$ of the vertex $u$, there is a distinct neighboring vertex $v_i$ of the same degree for the vertex $v$. While Rashevsky used simple linear graphs with indistinguishable vertices to symbolize molecular structure, weighted linear graphs or multigraphs are better models for conjugated or aromatic molecules because they more properly reflect the actual bonding patterns, i.e. electron distribution.

Trucco (1956) refined the definition of topological equivalence of atoms in terms of the orbits of the automorphism group of the molecular graph. This type of molecular information content was later termed as orbit’s information index, $I_{\text{orb}}$. In the latter case, two vertices are considered equivalent if they belong to the same orbit of the automorphism group, i.e., if they can interchange preserving the adjacency of the graph.

Bonchev and Trinajstic (1977) provided information on distances to explain the molecular branching that is the critical parameter determining the relative magnitude of various molecular thermodynamic properties. Initially they used the information indices defined by Rashevsky (1955) for graphs. However, these indices were not found suitable for describing efficiently the branching properties of graphs. For reducing degeneracies, Bonchev and Trinajstic (1977) introduced an information $I_D$ on distances in a graph, considering all the matrix elements $d_{ij}$ of the distance matrix $D$ as elements of a finite probability scheme associated with the graph in question as:

$$I_D = n^2 \log_2 n^2 - n \log_2 n - \sum_{i=1}^{d(G)} 2n_i \log_2 2n_i$$
Bonchev et al. (1979) applied information theory to the problem of characterizing molecular structures and proposed information content index (I). This relation defined the information of a system, I, with n elements:

\[ I = n \log_2 n - \sum_{j=1}^{n} \log_2 n_j, \]

Where \( n \) is the number of different sets of element, \( n_j \) is the number of elements in the set \( j \) set of elements, an summation is over all sets of elements.

Basak et al. (1980) defined another information-theoretic parameter, known as structural information content (SIC), and is calculated as:

\[ \text{SIC} = \frac{\text{IC}}{\log_2 n} \]

Where IC is calculated from Shannon’s relation and \( n \) is the total number of vertices of graph.

Bertz (1981) proposed a new general information index that incorporates the information on atomic composition, information on graph connections and molecular size. The general index of molecular complexity of Bertz is given by:

\[ I_{\text{Bertz}} = I_{AC} + I_{\text{CONN}} + I_{\text{SIZE}} \]

Chirality and stereochemistry can be reflected by (I_CONN) the distribution of connections into classes of orbital equivalency.

Basak et al. (1983) defined another information-theoretic invariant, complementary information content (CIC\textsubscript{r}), which is calculated as:

\[ \text{CIC}_r = \log_2 n - \text{IC}_r \]

CIC\textsubscript{r} represents the difference between maximum possible complexity of graph (where each vertex belongs to a separate equivalent class) and the realized topological information of a chemical species as defined by IC\textsubscript{r}.

Raychaudhary et al. (1984) defined three information invariants - degree complexity (I\textsuperscript{d}), graph vertex complexity (I\textsuperscript{v}) and graph distance complexity (I\textsuperscript{D}). Graph distance complexity was found to be the only index to discriminate all the studied graphs.

Klopman and Raychoudhary (1988) described an information index namely vertex distance complexity (I\textsuperscript{d}), for the vertices of a molecular graph and used the same for qualitative evaluation of mutagenic activity of a series of non-fused ring aromatic compounds.
Skorobogatov et al. (1991) considered one more information index based on the distance matrix for structure-activity correlation studies. The information index $H_2$ was defined as:

$$H_2 = \sum_{i=1}^{k} \frac{d(i)k_i}{2W} \log_2 \left( \frac{d(i)k_i}{2W} \right)$$

Where $k_i, i = I, \ldots, k$, is the number of vertices having the distance $d(i)$.

Raychaudhury and Ghosh (2004) proposed new information-theoretical measure of similarity, $INFSIM$, based on Shannon's measure of information content of a discrete system. They used Shannon’s measure of information theoretical measure of redundancy of a system to derive the similarity measure. They also used a topological shape and size index (TSS) and a topophysical molecular descriptor (TPMD) for the study. These indices have been used to carry out molecular similarity analysis for quantitative discrimination (active/ inactive) of eleven β-lactams with respect to antibacterial activity of penicillin $G$. It was concluded that information- theoretical similarity measure, $INFSIM$, has been able to produce similarities that appear to help classify (active / inactive) the studied compounds with significant accuracy.

Varmuza et al. (2009) proposed new family of topological information indices based on the full neighborhood of all atoms. They considered each atom of a molecular structure as a subsystem and for each atom the complete neighbourhood was characterized by an information functional $f_i$, based on the number of atoms in all spheres around the atom. The properties of all atoms were normalized to a sum of one (a probability-like measure, $p_i$) from which the information entropy was calculated. The entropy was scaled by the number of atoms in the structure to give a molecular descriptor E.

$$E = a \left( ln \ n + \sum_{i=1}^{n} p_i \ ln \ p_i \right)$$

Where $p_i$ is “normalized” probabilities and calculated as:

$$p_i = f_i / \sum_{j=1}^{n} f_i$$
For each subsystem the value $f_i$ of an invariant is calculated based on the complete neighborhood. The values of the invariants are normalized to give “probabilities” $p_i$ that are combined to an entropy measure $E$.

Dehmer et al. (2010) derived entropic measures to calculate the information content of vertex- and edge labeled graphs and investigated the influence of proposed TDs on the prediction performance of the underlying graph classification problem. They demonstrated that the application of entropic measures to molecules representing graphs is useful to characterize such structures meaningfully and such methods might be valuable for solving problems within biological network analysis.

Dehmer et al. (2012) evaluated the uniqueness of several information-theoretic measures for graphs based on so-called information functionals and compared the results with other information indices and non-information-theoretic measures such as the well-known Balaban $J$ index. They found that one of the information measures for graphs using the information functional based on degree-degree associations outperformed the Balaban $J$ index tremendously. These results were obtained by using nearly 12 million exhaustively generated, non-isomorphic and unweighted graphs.

**Miscellaneous graph invariants and approaches**

Moran (1948) introduced a coefficient $I$ to measure spatial autocorrelation as:

$$I = \frac{(n / k_1) \sum (x_i - \bar{x})(x_j - \bar{x})}{\sum (x_j - \bar{x})^2}$$

$I$ can take both positive and negative values and is close to zero when there is no spatial autocorrelation.

Geary (1954) suggested the contiguity ratio, $c$, on the basis of the squared differences between contiguous areas:

$$c = \frac{(n - 1 / 2k_1) \sum (x_i - x'_i)^2}{\sum (x_i - \bar{x})^2}$$

Where $n$ is the number of areas, $x_i$ is the value for area $t$, $\bar{x}$ is the mean of all the values, $k_i$ is the number of areas connected to area $t$ and $k_1 = \sum k_i$ is twice the sum of all connections.
Wiberg (1968) proposed bond index to measure the multiplicity of bonds between two atoms. It was defined as the sum of the squares of the bond orders \( p_{jk} \) between any one atomic orbital and all other orbitals in a molecule. It is two times the charge density in that orbital \( p_{ij} \) less the square of the charge density:

\[
\sum_k p_{jk}^2 = 2p_{ij} - p_{ij}^2
\]

For a unit charge density, the value is 1, whereas it goes to zero for \( p_{ij} = 2 \) (a non-bonded pair) or for \( p_{ij} = 0 \) (an empty orbital). Correspondingly, the sum of the squares of the bond orders to an atom corresponds to the number of covalent bonds formed by that atom, corrected for the ionic character in each bond (Trindle, 1969).

Gutman and Randic (1977) suggested a way for characterization of branching using the concept of comparability of functions. They suggested that the structure having an identical distribution of valencies should not be discriminated.

Moreau and Broto (1980a, b) derived 2D-autocorrelation descriptors from the molecular graph weighted by atom physicochemical properties (i.e. the atom weightings \( w_i \)). The spatial autocorrelation was then evaluated by considering separately all the contributions of each different path length (lag) in the molecular graph, as collected in the topological distance matrix. The total spatial autocorrelation at lag \( k \) \( ATS_k \) was obtained by summing all the products \( w_i \cdot w_j \) of all the pairs of atoms \( i \) and \( j \), for which the topological distance equals the lag as:

\[
ATS_k = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i \cdot w_j \cdot \delta(k; d_{ij}) \quad k = 0, 1, 2, ..., d
\]

Where \( w \) is any atomic property; \( N \) is the number of atoms in a molecule; \( k \) is the lag and \( d_{ij} \) is the topological distance between atoms \( i \) and \( j \), \( d \) is the topological diameter, i.e. the maximum topological distance in the molecule, and \( \delta \) is a Dirac-delta function defined as:

\[
\delta(k; d_{ij}) = 1 \text{ if } d_{ij} = k \\
0 \text{ otherwise}
\]

The autocorrelation \( ATS_0 \) defined for the path of length zero was calculated as:

\[
ATS_0 = \sum_{i=1}^{N} w_i^2
\]
i.e. the sum of the squares of the atomic properties. Typical atomic properties that can be considered are atomic masses, polarizabilities, charges, and electronegativities (Broto et al., 1984).

**Mekenyan and Bonchev** (1986) developed a new methodology: the *Optimal Approach based on Structural Indices Set (OASIS)*, by combining geometrical, topological and quantum-chemical based descriptors and also developed a computer software for the same. This methodology was reported as a second generation QSAR approach and was viewed as a generalized Hansch approach. It deals with molecules as a whole and was developed for QSAR/ QSTR studies of structurally related compounds (Bonchev et al., 1993).

**Klopman and Raychaudhary** (1988) introduced an artificial computer program called *Computer-Automated Structure Evaluation (CASE)* technique, based on topological descriptors consisting of linear/ one branch molecular fragments composed of 3-16 heavy atoms along with the hydrogen atoms attached to them. Earlier, Raychaudhary in his PhD. thesis reported a distance based topological index called *graph distance complexity* \( (H^d) \). It is obtained from the value of an information-theoretical index called *vertex distance complexity* \( (V^d) \) of the vertices in a graph.

**Pal et al.** (1988) proposed a novel topochemical arrived unique scheme \( (TAU) \) in molecular connectivity concept. It is based on nuclear and electronic features of the atoms constituting the molecular graph. This scheme defines the molecular graph in terms of sets of the vertex weights \( (V) \) and edge weights \( (E) \). Three molecular indices namely *skeletal index* \( (T_R) \), *functionality index* \( (F) \), and *branchedness index* \( (B) \), were derived from it. It was also used for development of QSAR model for inhibition of *M. tuberculosis* by substituted bromophenols (Pal et al., 1989).

**Kier and Hall** (1990) introduced new TIs called the electrotopolgical state \( (E) \) indices based on graph invariants for each atom in the molecule. The E-state variable encodes the intrinsic electronic state of the atom as influenced by electronic environment of all other atoms within the topological framework of the molecule.

**Bonchev et al.** (1992) utilized graph topological extrapolation method for the modeling of polymer properties \( (TEMPO) \) that was based on the graph topological description of the polymer elementary units by means of the normalized *Wiener number*
represented as a polynomial of degree 3 with respect to the number of atoms. The method was applied to the calculation of p-electron energies and energy gaps of various conjugated polymers, as well as to the assessment of the melting point, density, refractive index, and specific rotation of some industrially produced polymers.

Mekenyan et al. (1993) applied optimized approach based on Structural Index Sets to model the bronchospasmolytic activity and toxicity of theophylline derivatives. The opposite influence of graph topology on bronchospasmolytic activity and toxicity was established.

Galvez et al. (1995) demonstrated that by an adequate choice of topological descriptors it is possible to not only predict different pharmacological activities but also to design new active compounds, including lead drugs, in several therapeutical scopes, with a surprising level of efficiency, especially considering the simplicity of the calculations. They concluded that in spite of its limitations, molecular topology ought to be considered not just as an excellent tool for molecular and drug design but as a real alternative approach to the study of chemical bonds, whose theoretical physicochemical basis is still to be developed.

Gealy et al. (1996) compiled a database of chemicals causing allergic contact dermatitis (ACD) to discern structure-activity relationships. Clinical reports represented a considerable fraction of the data. The model was used to predict the activity of chemicals identified from clinical case reports.

Basak et al. (1997) introduced a Hierarchical QSAR approach using topostructural, topochemical and geometric parameters.

Hu and Xu (1997) devised a new topological index from an all-paths method. This new topological index displayed high discriminating power for various kinds of organic compounds such as alkane trees, complex cyclic or polycyclic graphs, and structures containing heteroatoms and thus used as a molecular identification number (MID06) for chemical documentation.

DeJulian-Ortiz et al. (1998) proposed chiral topological invariants that can differentiate pharmacological activity between pairs of enantiomers. Chiral behavior is not associated with 3-dimentional geometry but is related to symmetry, which allowed
handling of chiral atoms by weighted graphs. The 50% inhibitory concentration values of D₂ dopamine receptor and the σ receptor for a group of 3-hydroxy phenyl piperdines were predicted.

**Cercos-Del-Pozo et al.** (1999) designed new compounds showing hypolipidemic activity using computer-aided method based on molecular topology and QSAR analysis. Connectivity functions were used to design three drugs that were tested for hypolipidemic property in rats.

**Palyulin et al.** (2000) proposed *molecular field topology analysis (MFTA)*, a new method of QSAR analysis for organic compounds. This method involved the topological superposition of the training set structures and the construction of a molecular supergraph (MSG). This enabled the creation of the uniform descriptor vectors based on the local physicochemical parameters (atom and bond properties) of the molecules.

**Ivanciuc et al.** (2000) defined some new topological indices to investigate the partial ordering of heptanes and octanes. They used newly defined molecular graph operators for the computation of families of topological indices.

**Esrtrada** (2001) developed a generalized molecular graph matrix. An invariant defined on the basis of this matrix generalizes some of the classical topological indices. Thus *Wiener index, Zagreb group parameters, Balaban J index, Harary H number, Randic index* and *valence connectivity index* were part of an infinite set of molecular descriptors that could be derived from the same invariant.

**Golbraikh et al.** (2001) introduced several series of novel chirality descriptors of chemical organic molecules. These included *modified molecular connectivity indices, Zagreb group indices, extended connectivity and overall connectivity*, and *topological charge indices*. These modified descriptors made use of an additional term called chirality correction, which was added to the vertex degrees of asymmetric atoms in a molecular graph. These novel chirality descriptors were implemented in a QSAR study of a set of ecdysteroids with a high content of chiral and enantiomeric compounds. The results compared favorably with those obtained with the comparative molecular field analysis.
Kauffman (2001) used experimental IC$_{50}$ data for 314 selective cyclooxygenase-2 (COX-2) inhibitors to develop quantitative and classification models as a potential screening mechanism for larger libraries of target compounds. Numerical descriptors encoding solely topological information were calculated for all structures and were used as inputs for linear regression, computational neural network and classification of analysis routines. High degrees of predictions were obtained.

O’Brien and Popelier (2001) presented a novel molecular similarity method called quantum topological molecular similarity (QTMS) based on the topology of the electron density. QTMS is able to suggest a molecular fragment that contains the active centre or the part of the molecule that is responsible for the QSAR.

Cao and Yuan (2001) proposed three novel topological indices: OEI (odd – even index), VDI (vertex degree-distance index), and RDI (ring degree-distance index) and then carried out multiple regression analysis with these indices against the boiling points of paraffins and cycloalkanes.

Mattioni (2002) developed mathematical models to find QSAR relationships that correlate chemical structure and inhibition towards carbonic anhydrase isoenzymes. Good prediction rates were obtained with these models (Mattioni and Jurs, 2002).

Golbraikh et al. (2002) introduced several series of novel ZE-isomerism descriptors derived directly from two dimension molecular topology. The applicability of ZE-isomerism descriptors to QSAR analysis was demonstrated in the studies of a series of 131 anticancer agents inhibiting tubulin polymerization.

Junkes et al. (2003) proposed a new semi-empirical topological index denoted as $I_{ET}$, for the prediction of retention indices for a diverse set of organic compounds i.e., alkanes, alkenes, esters, ketones, aldehydes, and alcohols. The index was based on the hypothesis that the chromatographic retention is due to the interaction of each atom of the molecule with the stationary phase and consequently the value of the index is reduced by steric effects from its neighbors. The results indicated that the semi-empirical index had excellent predictive power using a single descriptor for a large data set of organic compounds.
Torrens (2003) applied new valence topological charge-transfer indices to the calculation of dipole moments. The model was generalized for molecules with heteroatoms. In valence topological charge-transfer indices terms, the presence of each heteroatom was taken into account by introducing its electronegativity value in the corresponding entry of the main diagonal of the adjacency matrix. The ability of the indices for the description of the molecular charge distribution was established by comparing them with the dipole moments of a set of homologous series of phenyl alcohols.

Ivanciuc (2004) presented a new application of topological indices in computing similarity matrices that were subsequently used to develop QSAR/QSAR models. The similarity matrices were computed using four similarity indices, namely the Cosine, Dice, Richards, and Good similarity indices. The similarity matrices were used to develop multilinear regression QSAR models of the anticonvulsant activity of 30 phenylacetanilides. The results showed that similarity matrices derived from molecular graph descriptors could provide the basis for the investigation of [(Q)SAR/QSPR] relationships.

Garcia et al. (2005) developed an algorithm for the generation of molecular graphs with a given value of the Wiener index. The selection of parameters as the interval of values for the Wiener index, the diversity and occurrence of atoms and bonds, the size and number of cycles, and the presence of structural patterns guide the processing of the heuristics generating molecular graphs with a considerable saving in computational cost. The modularity in the design of the algorithm allows it to be used as a pattern for the development of other algorithms based on different topological invariants, which allow for its use in areas of interest, say as involving combinatorial databases and screening in chemical databases.

Gutierrez-Oliva et al. (2006) analyze the application of the core-valence bifurcation (CVB) index and bond order index by considering a series of doubly hydrogen-bonded complexes. Their values are seen to be linearly related to bond energies estimated through a bond-energy-bond-order relationship; also, the mean value of the topological index appears to be related to the complexation energy computed by methods based on density functional theory.
Estrada and Matamala (2007) proposed the use of the generalized topological indices (GTIs), which account for several of the classical TIs in one single graph invariant. GTIs represent points in a six-dimensional space of topological parameters, which can be optimized for describing a specific property. The situation shows some resemblance with the geometry optimization procedures used to minimize molecular energy. Using this approach, it was observed that GTI have improved QSARs by reducing the standard deviation by almost 50%. In addition, the current approach permits the illustration of the similarities and differences among the different descriptors studied, indicating possible directions for searching new optimal molecular descriptors.

Chekmarev et al. (2008) extended the application of the shape signatures methodology to the domain of computational models for cardiotoxicity. They applied Shape Signatures method to generate MDs for use in classification techniques such as k-nearest neighbors (k-NN), support vector machines (SVM), and Kohonen self-organizing maps (SOM). The performances of these approaches were assessed by applying them to a data set of compounds with varying affinity toward the 5-HT<sub>2B</sub> receptor as well as a set of human ether-a-go-go-related gene (hERG) potassium channel inhibitors.

Burden et al. (2009) described charge fingerprints as a new type of universal descriptors for building good (Q)SAR/QSPR models of a diverse range of physicochemical and biological properties. The atomistic and charge fingerprint descriptors were found more successful than the eigenvalue descriptors in building (Q)SAR models on their own. They have suggested that universal descriptors will be useful for modeling large data sets as well as for screening large virtual libraries.

Verma and Hansch (2011) reviewed the various application/use of $^{13}$C-NMR chemical shift as (Q)SAR/QSPR descriptor. Their detailed investigation indicated that the $^{13}$C-NMR chemical shifts are sufficiently rich in chemical information and are able to encode the structural features of the molecules contributing significantly to their biological activity, chemical reactivity, or physical characteristics. They proposed $^{13}$C-NMR chemical shifts as promising descriptor in classical (Q)SAR/QSPR modelling studies.
Rabal and Oyarzabal (2012) developed a novel descriptor (LIR1f) accounting for ligand-receptor interactions to define and visually explore biologically relevant chemical space. It converts structural information into a one-dimensional string accounting for the plausible ligand-receptor interactions as well as for topological information. This descriptor was proposed with an aim to enable the clustering, profiling, and comparison of libraries of compounds from a chemical biology and medicinal chemistry perspective. The ligand receptor application of LIR1f was demonstrated with four reported compound data sets associated with four different target families.

Nie et al. (2012) proposed a novel TD \( \acute{ED}_m \) by introducing the bond angle into hidden hydrogen graph of molecules and using the geometric distance instead of the sum of bond length between two vertices. The \( \acute{ED}_m \) was derived from ionicity index matrix \( Q \) (a subtype of distance matrix), and branching degree matrix \( \acute{G} \) as:

\[
\acute{ED}_m = \acute{G}_m Q (m=1,2,3)
\]

The utility of \( \acute{ED}_m \) was also demonstrated through development of high quality QSPR models of 44 cis-trans isomers for alkenes. The \( \acute{ED}_m \) described the molecular structure more accurately, and realizes unique characterization to cis-trans isomers.

Review of literature reveals that though thousands of MDs of diverse nature have been reported but only a small fraction of them have been successfully utilized in (Q)SAR, QSPR, QSTR and QSPkR.

As a consequence, there is a strong need to develop novel molecular descriptors having very high discriminating power, low degeneracy and non correlation with the existing descriptors but showing good correlation with at least one biological activity/property of interest so as to accelerate the drug discovery process.