Matrix Representation

Four kinds of matrices (order k, k = 0-15) can be used in **QuBiLS-MAS** software, *namely* Non-stochastic (NS), Simple stochastic (SS), Double stochastic (DS) and Mutual Probabilistic (MP) matrices.

1. Non-stochastic (NS) matrix

In the earlier reports, were introduced new molecular matrices that describe changes in the electronic distribution with time throughout the molecular backbone. The nxn k^{th} non-stochastic graph–theoretic electronic-density matrix of the molecular pseudo graph ($\boldsymbol{G}$), $\mathbf{M}^{\boldsymbol{k}}$, is a square and symmetric matrix, where n is the number of atoms (atomic nuclei) in the molecule. The coefficients ${}^{k}m_{ij}$ are the elements of the k^{th} power of **M**(G) and are defined as follows:

$$m_{ij} = P_{ij} \text{ if } i \neq j \text{ and } \exists\, e_k \in \mathrm{E(G)}$$
$$= L_{ii} \text{ if } i = j$$
$$= 0 \text{ otherwise}$$

where, E(**G**) represents the set of edges of **G**. P_{ij} is the number of edges (bonds) between vertices (atomic nuclei) v_i and v_j, and L_{ii} is the number of loops in v_i.

The elements $m_{ij} = P_{ij}$ of such a matrix represent the number of chemical bonds between an atomic nucleus i and other j. The matrix $\mathbf{M}^{\boldsymbol{k}}$ provides the number of walks of length k that link every pair of vertices v_i and v_j (for duplex). For this reason, each edge in $\mathbf{M}^{\mathbf{1}}$ represents 2 electrons belonging to the covalent bond between atomic nuclei i and j. On the other hand, molecules containing aromatic rings with more than one canonical structure are represented by a pseudo graph. This happens for substituted aromatic compounds such as pyridine, naphthalene, quinoline, and so on, where the presence of pi (π) electrons is accounted for by means of loops in each atomic nucleus of the aromatic ring. Conversely, aromatic rings having only one canonical structure, such as furan, thiophene and pyrrol are represented by a multigraph. In addition, lone-pair electron(s) on heteroatoms may be considered or not in the descriptor calculations. Each pair of electrons (n) on heteroatoms is considered as loop in the pseudo graph too. Therefore, these electrons are accounted for in the main diagonal of $\mathbf{M}^{\mathbf{1}}$. Their values are 2 for oxygen atoms (four lone-pair electrons), 3 for all halogens (six lone-pair electrons).

2. Simple-stochastic (SS) matrix

The k^{th} *stochastic graph–theoretic electronic-density matrix* of G, $\mathbf{S}^k$, can be directly obtained from $\mathbf{M}^k$. Here, $\mathbf{S}^k = [{}^k\mathbf{s}_{ij}]$, is a square matrix of order n (n = number of atomic nuclei) and the elements ${}^k\mathbf{s}_{ij}$ are defined as follows:

$$ {}^k s_{ij} = \frac{{}^k m_{ij}}{{}^k SUM_i} = \frac{{}^k m_{ij}}{{}^k \delta_i} $$

where, ${}^k m_{ij}$ are the elements of the k^{th} power of M and the SUM of the ith row of M^k is named the k-order vertex degree of atom i, ${}^k\delta_i$. It should be remarked that the matrix S^k in the equation above has the property that *the sum of the elements in each row* is 1.

3. Double-stochastic (DS) matrix

The k^{th} *double stochastic graph–theoretic electronic-density matrix* of G, $\mathbf{D}^k$, can also be directly obtained from $\mathbf{M}^k$. Here, $\mathbf{D}^k = [{}^k\mathbf{d}_{ij}]$, is a square matrix of order n (n = number of atomic nuclei). It should be remarked that the matrix D^k has the property that *the sum of the elements in each row or in each column* is 1. Notice that ***S*** matrix (simple stochastic) is not symmetric, therefore, with the aim of equalize the probabilities in both senses is employed a *double-stochastic scaling*. This scaling is performed by using Sinkhorn-Knopp algorithm.

4. Mutual probability (MP) matrix

The k^{th} *mutual probability graph–theoretic electronic-density matrix* of G, P^k, can be directly obtained from $\mathbf{M}^k$. Here, $P^k = [{}^k\mathbf{p}_{ij}]$, is a square matrix of order n (n = number of atomic nuclei) and the elements ${}^k\mathbf{p}_{ij}$ are defined as follows:

$$ {}^k p_{ij} = \frac{{}^k m_{ij}}{{}^k m(S)} = \frac{{}^k m_{ij}}{\sum_{i,j} {}^k m_{ij}} $$

where, ${}^k\mathbf{p}_{ij}$ denotes the mutual probability of vertices i and j, and $m(S)$ the *sample space*. The sample space is computed by summing the all element of **M**.