N-TUPLES TOPOLOGICAL AND GEOMETRIC CUTOFFS

With the purpose of establishing a relation between the topological and geometrical aspects for each group of *"N"* atoms considered and in this way take into account some short-, middle- and large-relations, two procedures are defined:

- *N-tuple Graph-theoretical cutoff* (*P*) known as "*path cutoff"*, based on the topological distance. These cutoffs are denoted as: ***lag P*** for $N = 2$, ***lag 3P*** for $N = 3$ and ***lag 4P*** for $N = 4$.
- *N-tuple Euclidean-geometric cutoff* (*L*) known as "*length cutoff"*, based on the Euclidean distance. These cutoffs are denoted as: ***lag L*** for $N = 2$, ***lag 3L*** for $N = 3$ and ***lag 4L*** for $N = 4$.

The application of one or both molecular cutoffs on the matrices $\mathbb{G}^{1}_{(F)}$, $\mathbb{GT}^{1}_{(F)}$ and $\mathbb{GQ}^{1}_{(F)}$ permits to compute the *two-, three- and four-tuple topological and geometric neighborhood quotient total (or local-fragment) spatial-(dis)similarity matrices*, $\mathbb{NQG}^{1}_{(F)}$, $\mathbb{NQGT}^{1}_{(F)}$ and $\mathbb{NQGQ}^{1}_{(F)}$, respectively. The coefficients of these novel matrix approaches are computed by multiplying the elements of the matrices $\mathbb{G}^{1}_{(F)}$, $\mathbb{GT}^{1}_{(F)}$ and $\mathbb{GQ}^{1}_{(F)}$ by a ratio obtained as the number of relations between the *N* considered atoms that present a topological and/or Euclidean-geometric distance smaller or equal to a predefined *p* and/or *l* thresholds. Then, the entries ${}^{NQ}g^{1}_{ij(F)}$, ${}^{NQ}gt^{1}_{ijl(F)}$ and ${}^{NQ}gq^{1}_{ijlh(F)}$ of the matrices $\mathbb{NQG}^{1}_{(F)}$, $\mathbb{NQGT}^{1}_{(F)}$ and $\mathbb{NQGQ}^{1}_{(F)}$ are mathematically defined as follows:

$$
\begin{aligned}
{}^{NQ}g^{1}_{ij(F)} &= g^{1}_{ij(F)} \text{ if } p_{min} \le p_{ij} \le p_{max} \;\; \textit{and / or} \;\; l_{min} \le l_{ij} \le l_{max} \\
&= 0 \text{ otherwise}
\end{aligned}
$$

$$
\begin{aligned}
{}^{NQ}gt^{1}_{ijl(F)} &= gt^{1}_{ijl(F)} \quad \textit{if } p_{min} \le p_{ij}, p_{jl}, p_{li} \le p_{max} \;\; \textit{and/or} \;\; l_{min} \le l_{ij}, l_{jl}, l_{li} \le l_{max} \\
&= \frac{2}{3} gt^{1}_{ijl(F)} \quad \begin{cases} \textit{if } p_{min} \le p_{ij}, p_{jl\,(li)} \le p_{max} \;\; \textit{and/or} \;\; l_{min} \le l_{ij}, l_{jl\,(li)} \le l_{max} \\ \textit{if } p_{min} \le p_{jl}, p_{li} \le p_{max} \;\; \textit{and/or} \;\; l_{min} \le l_{jl}, l_{li} \le l_{max} \end{cases} \\
&= \frac{1}{3} gt^{1}_{ijl(F)} \quad \textit{if } p_{min} \le p_{ij\,(jl,li)} \le p_{max} \;\; \textit{and/or} \;\; l_{min} \le l_{ij\,(jl,li)} \le l_{max} \\
&= 0 \qquad \textit{otherwise}
\end{aligned}
$$

$$
\begin{aligned}
{}^{NQ}gq^{1}_{ijlh(F)} &= gq^{1}_{ijlh(F)} \quad \textit{if } p_{\min} \le p_{ij},p_{jl},p_{lh},p_{hi} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{ij},l_{jl},l_{lh},l_{hi} \le l_{\max} \\
&= \frac{3}{4} gq^{1}_{ijlh(F)} \quad \begin{cases} \textit{if } p_{\min} \le p_{ij},p_{jl\,(lh)},p_{lh\,(hi)} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{ij},l_{jl\,(lh)},l_{lh\,(hi)} \le l_{\max} \\ \textit{if } p_{\min} \le p_{jl},p_{lh},p_{hi} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{jl},l_{lh},l_{hi} \le l_{\max} \end{cases} \\
&= \frac{2}{4} gq^{1}_{ijlh(F)} \quad \begin{cases} \textit{if } p_{\min} \le p_{ij},p_{jl\,(lh,hi)} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{ij},l_{jl\,(lh,hi)} \le l_{\max} \\ \textit{if } p_{\min} \le p_{jl},p_{lh\,(hi)} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{jl},l_{lh\,(hi)} \le l_{\max} \\ \textit{if } p_{\min} \le p_{lh},p_{hi} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{lh},l_{hi} \le l_{\max} \end{cases} \\
&= \frac{1}{4} gq^{1}_{ijlh(F)} \quad \textit{if } p_{\min} \le p_{ij\,(jl,li,hi)} \le p_{\max} \quad \textit{and/or} \quad l_{\min} \le l_{ij\,(jl,lh,hi)} \le l_{\max} \\
&= 0 \qquad \textit{otherwise}
\end{aligned}
$$

where, the coefficients g^{1}_{ij}, gt^{1}_{ijl}, gq^{1}_{ijlh} represents the relations between two, three and four atoms of a molecule and correspond to the total (or local-fragment) matrices $\mathbb{G}^{1}_{(F)}$, $\mathbb{GT}^{1}_{(F)}$ and $\mathbb{GQ}^{1}_{(F)}$, respectively. In addition, p_{XY} and l_{XY} represent the topological and Euclidean-geometric distance between two atoms of a molecule, while $[p_{\min}, p_{\max}]$ and $[l_{\min}, l_{\max}]$ constitute the used-defined topological and Euclidean-geometric intervals, respectively.

Also, other molecular cutoff procedures are proposed in order to only consider the ternary $(N=3)$ and quaternary $(N=4)$ relations between atoms of a molecule whose values are consistent with a specific multi-metric. These procedures are denominated as *N-tuple Geometric cutoff based on Multi-metrics* and its mathematical definition is as follows:

$$
\begin{aligned}
{}^{NQ}gt^{1}_{ijl(F)} &= gt^{1}_{ijl(F)} \text{ if } \mathrm{tv}_{\min} \le \mathrm{tv}_{\mathrm{ijl}} \le \mathrm{tv}_{\max} \\
&= 0 \qquad \text{otherwise}
\end{aligned}
$$

$$
\begin{aligned}
{}^{NQ}gq^{1}_{ijlh(F)} &= gq^{1}_{ijlh(F)} \text{ if } \mathrm{qv}_{\min} \le \mathrm{qv}_{\mathrm{ijlh}} \le \mathrm{qv}_{\max} \\
&= 0 \qquad \text{otherwise}
\end{aligned}
$$

where, tv_{ijl} and qv_{ijlh} are the values corresponding to the calculation of a ternary and quaternary multi-metric, respectively. In addition, $[tv_{\min}, tv_{\max}]$ and $[qv_{\min}, qv_{\max}]$ are the predefined intervals when cutoffs based on relations between three and four atoms are applied, respectively. Specifically, the ternary multi-metrics that may be used include the Triangle Area (***lag A***), Bond

Angle (***lag BA***) and Ternary (or Triangle) Perimeter (***lag TP***); while the quaternary multi-metrics that may be used include Volume (***lag V***), Dihedral Angle (***lag DA***) and Quaternary (or Quadrilateral) Perimeter (***lag QP***).

It is important to highlight that the molecular cutoffs defined for a same number of atoms could be simultaneously applied, *e.g.*: in a relation between three distinct atoms $(i \neq j \neq l)$ if any permutation of *three-tuple cutoffs* (***lag 3P***, ***lag 3L***, ***lag A***, ***lag BA*** and ***lag TP***) is used, then all the considered criteria must be fulfilled. On the other hand, also the molecular cutoffs for relations between two, three and four atoms can be concurrently applied on the same matrix representation. Therefore, on *four-tuple matrix approaches* when four distinct atoms are analyzed $(i \neq j \neq k \neq l)$ then *four-tuple cutoffs* can be applied, if three distinct atoms are analyzed $[(i = j) \neq l \neq h]$ then *three-tuple cutoffs* can be applied, and if two distinct atoms are analyzed $[(i = j = l) \neq h]$ then *two-tuple cutoffs* can be applied. Likewise, this previous strategy is employed on *three-tuple matrix approaches* when *three-tuple cutoffs* and *two-tuple cutoffs* are computed for relations between three $(i \neq j \neq l)$ and two $[(i = j) \neq l]$ distinct atoms, respectively.