Supporting information

Multiscale graph neural network for high-density cycloalkane-based diesel and jet range biofuels properties prediction

Yanqiu Yao[1,2], Yizhuo Wang[2], Zhanchao Li[3], Jing Wang[1]*, Hong Wang[1,2]*

[1]School of Energy and Power Engineering, Xi'an Jiaotong University, Xi'an, 710054, China

[2]Frontier Institute of Science and Technology, Xi'an Jiaotong University, Xi'an, 710054, China

[3]School of Chemistry and Environmental Engineering, Sichuan University of Science & Engineering, Zigong, Sichuan, 643000, China

E-mail: hong.wang@xjtu.edu.cn

E-mail: jing.wang@xjtu.edu.cn

1 The problem of over smoothing and gradient vanishing due to GNNs deep architecture

Convolutional Neural Networks (CNNs) have impressive performance in a wide variety of fields. A key reason behind the success of CNNs is the ability to design and reliably train very deep CNN models. However, it is not yet clear how to properly train deep GNN architectures. Most state-of-the-art GNN models are no deeper than 3 or 4 layers (**ICCV.,** 2019, 16, 9266) due to the over-smoothing problem and the vanishing gradient problems. The detailed theoretical descriptions for the over smoothing problem and gradient vanishing problem are discussed as follows.

Over smoothing problem: Over smoothing is a common phenomenon in GNNs as the number of layers increases. The discussions of over smoothing problem have been reported in previous literature (**Chem. Sci.,** 2022, 13, 816; **AAAI.,** 2020, 34, 3438; **ICCV.,** 2019, 16, 9266; **PMLR.,** 2018, 80, 5453; **AAAI.,** 2018, 433, 3538). Li *et. al.* (**AAAI.,** 2018, 433, 3538) suggested that the Over smoothing problem was due to the graph convolution operations in GNNs. Using graph convolution operations, the node feature update rule for the $(l+1)$-th layer in GNNs is given by:

$$H^{(l+1)}=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{1}$$

where $H^{(l)}$ is the node feature matrix at layer l, $\tilde{A}$ is the adjacency matrix with self-loops, $\tilde{D}$ is the degree matrix of $\tilde{A}$, $W^{(l)}$ is the weight matrix at layer l, and σ is the nonlinear activation function (typically the ReLU function). The graph convolution of the GCN model is actually a special form of Laplacian smoothing as demonstrated by Li *et. al.* (**AAAI.,** 2018, 433, 3538). It is suggested that performing smoothing operations on nodes is the key mechanism for GNN to work, but nodes will converge to similar values after multiple Laplacian smoothing operations. This phenomenon is known as over smoothing. It makes the features indistinguishable and hurts the classification accuracy. Oono *et. al.* (**ICLR.,** 2020, Graph Neural Networks Exponentially Lose Expressive Power for Node Classification; **AI Open**, 2020, 1, 57) performed a more detailed analysis of the over smoothing phenomenon. They proved that if the maximum singular values of the weight matrix $W^{(l)}$ satisfies, as $s<\frac{1}{\lambda}$ the number of layers L increases, the output $H^{(L)}$ of GNNs gradually loses the information necessay to distinguish between different nodes. All node feature representations will eventually cornverge to a shared subspace. This leads to the over-smoothing issue.

Gradient vanishing problem: The vanishing gradient problem is a phenomenon where the gradients become increasingly small. It will significantly slow down or even halt the learning process, as the updates to the weights become negligible. Similar discussion has been reported in previous literature (**Chem. Eng. J.,** 2021, 414, 128817; **ICCV.,** 2015, 1026;

CVPR., 2017, 2261). As the network generates an output, the loss function(C) indicates how well it predicts the output. The network performs back propagation to minimize the loss. A back propagation method minimizes the loss function by adjusting the weights and biases of the neural network. In this method, the gradient of the loss function is calculated with respect to each weight in the network. In back propagation, the new weight(W_{new}) of a node is calculated using the old weight(W_{old}) and product of the learning rate(η) and gradient of the loss function ($\frac{\partial C}{\partial w}$).

$$W_{\text{new}} = W_{\text{old}} - \eta * \frac{\partial C}{\partial w} \tag{2}$$

With the chain rule of partial derivatives, the gradient of the loss function is represented as a product of gradients of all the activation functions of the nodes with respect to their weights. Therefore, the updated weights of nodes in the network depend on the gradients of the activation functions of each node. For the nodes with sigmoid activation functions, the partial derivative of the sigmoid function reaches a maximum value of 0.25. When there are more layers in the network, the value of the product of derivative decreases until at some point the partial derivative of the loss function approaches a value close to zero, and the partial derivative vanishes. This phenomenon is called the vanishing gradient problem. With shallow networks, sigmoid function can be used as the small value of gradient does not become an issue. When it comes to deep networks, the vanishing gradient could have a significant impact on performance. The weights of the network remain unchanged as the derivative vanishes. During back propagation, a neural network learns by updating its weights and biases to reduce the loss function. In a network with vanishing gradient, the weights cannot be updated, so the network cannot learn. Then, the performance of the network will decrease.

2 Calculation details

A molecular graph (G) was a collection of vertices (v) and edges (ε). In a molecule, $v_i \epsilon\ v$ represented the i-th atom and $e_{ij} \epsilon\ \varepsilon$ represented the chemical bond between i-th and j-th atoms. GNNs typically mapped a graph G to a vector $y_G \epsilon\ R^d$ through the message passing phase and readout phase. During the message passing phase, nodes exchanged and aggregated information by passing messages to capture the local structure of the graph. In the readout phase, a global representation was generated from the aggregated node features to comprehensively represent and analyze the entire graph.

The general description of our database：

Table S1. Number of compounds of different chemical classes in the dataset

Compound class	Density	Boiling Point	Flash Point	Viscosity/m Pa • s	VNHOC/M J/L	Cetane Number
Alkanes	106	111	98	88	102	75
Cycloalkanes	79	79	47	89	101	50
Alkene	105	106	73	97	107	25
Cycloalkenes	19	20	17	9	19	5
Alkadienes	28	28	28	20	32	4
Alkynes	23	24	15	10	18	1
Aromatics	165	167	142	120	170	48

Message passing phase：Given a graph *G*, the embedding of the *i*-th vertex at time step *t* was represented as $x_i^{(t)} \epsilon R^d$. The following graph's convolutional layer was used to update $x_i^{(t)}$ into $x_i^{(t+1)} \epsilon R^h$:

$$x_i^{(t+1)} = \sigma\left(W_1 x_i^{(t)} + W_2 \sum_{j \in \mathcal{N}(i)} x_i^{(t)}\right) \quad (3)$$

where the index i and j refer to the i-th and j-th nodes (atoms), respectively, W_1, $W_2 \epsilon R^{h \times d}$ were learnable weight matrices shared across all vertices, $\mathcal{N}(i)$ was the set of neighbors of vertex *i*, and σ contained a node-level batch normalization followed by a ReLU activation function, in which batch normalization was essential for very deep models. By using equation (3) to aggregate the neighboring messages and iterate them over time steps, vertex embeddings could gradually gather more global information on the graph.

Readout phase：The final output vector *y* from the set of vertex vectors in *G* was obtained by calculating the average of all the vertex embeddings:

$$y_G = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} x_v^{(L)} \quad (4)$$

where $|\mathcal{V}|$ was the number of vertices in the molecular graph, and *L* was the final time step. The readout phase aggregated vertex embeddings into a unified graph embedding.

Multiscale graph neural network：

Multiscale block**.** A multiscale block contained N graph convolutional layers described

by equation (5). To solve the over-smoothing and vanishing gradient problems caused by the deep architecture of GNNs, dense connections were established (the output of each layer was connected to the input of all subsequent layers) between each layer and all subsequent layers in the network. This connection scheme allowed each layer to directly access the feature maps of all preceding layers and use them as inputs, enabling the network to more fully utilize the information from earlier layers, as shown in **Figure 3(b)**. Formally, the multiscale block could be expressed as:

$$\begin{aligned} x_i^{(1)} &= \mathscr{H}(x_i^{(0)}, \Theta_1) \\ x_i^{(2)} &= \mathscr{H}(x_i^{(0)} | x_i^{(1)}, \Theta_2) \\ x_i^{(3)} &= \mathscr{H}(x_i^{(0)} \parallel x_i^{(1)} \parallel x_i^{(2)}, \Theta_3) \\ x_i^{(N)} &= \mathscr{H}(x_i^{(0)} \parallel x_i^{(1)} \parallel \cdots \parallel x_i^{(N-1)}, \Theta_N) \end{aligned} \tag{5}$$

where $\mathscr{H}$ was a graph convolutional layer described by equation (5) with parameters Θ_n(consist of W_1 and W_2) in which n represented the n-th layer, and || was the concatenate operation. The multiscale block extracted multiscale features which described the structure information of a molecule in both local and global contexts.

Transition layer. To enhance the depth of the MGNN, transition layers were employed as connectors between two neighboring multiscale blocks. The transition layer aimed to integrate the multiscale features from the previous multiscale block and reduce the channel number of the feature map. Specifically, for an input multiscale features at time step $N+1$ as $x_i^{(0)} \parallel x_i^{(1)} \parallel \cdots \parallel x_i^N \in R^{d+(N-1)h}$, the transition layer was formulated in the following manner:

$$x_i^{(N+1)} = \sigma\left(\Phi_1(x_i^{(0)} \parallel x_i^{(1)} \parallel \cdots \parallel x_i^N) + \Phi_2 \sum_{j \in \mathcal{N}(i)} (x_j^{(0)} \parallel x_j^{(1)} \parallel \cdots \parallel x_j^N)\right) \tag{6}$$

where $\Phi_1, \Phi_2 \in R^{(M/2)\times M}$ were learnable weight matrices shared across all vertices in which $M = d + (N-1)h$. By using the transition layer, the channel numbers were reduced to half of the input to save computational cost. Finally, a readout layer described by equation (4) was used to convert the whole graph to a feature vector $y_{\mathcal{G}} \in R^d$.

After obtaining the vector representation of fuel molecules, the representation was then fed into a multi-layer perceptron (MLP) to predict properties. Concretely, the MLP contained three linear transformation layers to map the combined representation into affinity score in which each linear transformation layer was followed by a ReLU

activation and dropout layer with a dropout rate of 0.1 following with the previous studies. The mean squared error (MSE) was used as the loss function.

Isomeric effects：The isomeric property variance within a group was caused by differences in topology. To explore how the structure of cycloalkane-based high-density biofuels affects their properties, we introduced the following several parameters.

The core count of a non-hydrogen vertex α was defined as

$$\alpha = \frac{Z - Z^V}{Z^V} \cdot \frac{1}{PN - 1} \tag{7}$$

where *PN* stood for period number, *Z* is the atomic number and Z^v is the valence electron number.. Considering the hydrogen atom was the reference, the value of α for hydrogen was taken as zero. $\sum \alpha / N_v$ described the molecular bulk relative to molecular size. The terms such as

$$\frac{(\sum \alpha)_P}{\sum \alpha}, \frac{(\sum \alpha)_X}{\sum \alpha}, \frac{(\sum \alpha)_Y}{\sum \alpha}$$

$$\frac{(\sum \alpha)_P}{\sum \alpha}, \frac{(\sum \alpha)_X}{\sum \alpha}, \frac{(\sum \alpha)_Y}{\sum \alpha}$$

could be used as shape parameters. $(\sum \alpha)_P$ $(\sum \alpha)_X$ $(\sum \alpha)_Y$ stood for summation of α values of the vertices that were joined to one, three, and four other vertices, respectively, in the molecular graph.

In a connected molecular graph, the composite index (η) comprehensively considered bonded and non-bonded interactions. Its specific calculation process was as follows:

$$\varepsilon = -\alpha + 0.3 \times Z^v \tag{8}$$

$$\beta = \sum x\sigma + \sum y\pi + \delta \tag{9}$$

$$\gamma_i = \frac{\alpha_i}{\beta_i} \tag{10}$$

$$\eta = \sum_{i<j} \left[\frac{\gamma_i \gamma_j}{r_{ij}^2} \right]^{0.5} \tag{11}$$

where ε was the measure of electronegativity. The VEM (valence electron mobile) count (β) for non-hydrogen vertex consisted of two parts: namely sigma contribution to VEM count ($\sum \beta_s$) and non-sigma contribution to VEM count ($\sum \beta_{ns}$), which were

defined as below:

$$\beta_s = \sum x\sigma \tag{12}$$

$$\beta_{ns} = \sum y\pi + \delta \tag{13}$$

For the calculation of the VEM count β, the contribution of a sigma bond x between two atoms of similar electronegativity ($\Delta\varepsilon \leq 0.3$) was considered to be 0.5, and for a sigma bond between two atoms of different electronegativity ($\Delta\varepsilon > 0.3$), it was considered to be 0.75. Again, in the case of π bonds, contributions (y) were considered based on the type of the double bond: (a) for π bond between two atoms of similar electronegativity ($\Delta\varepsilon \leq 0.3$), y was taken to be 1; (b) for π bond between two atoms of different electronegativity ($\Delta\varepsilon > 0.3$) or for conjugated (non–aromatic) π system, y was considered to be 1.5; (c) for aromatic π system, the value of y was taken as 2. δ served as a correction factor having a value of 0.5 per atom with a loan pair of electrons capable of making resonance with an aromatic ring (*e.g.* nitrogen of aniline, oxygen of phenol, methoxide, halogens, etc.). In eq 4, the VEM vertex count (γ_i) was the i-th vertex in a connected molecular graph.

When all hetero-atoms and multiple π bonds in the molecular graph were replaced by carbon atom and single bond respectively, it corresponded to a molecular graph which might be considered as the reference alkane and the corresponding composite index value was defined as η_R. Considering functionality as the presence of heteroatoms (atoms other than carbon or hydrogen) and multiple bonds, the functionality index η_F might be calculated as $\eta_F = \eta_R - \eta$. When only bonded interactions were considered, the corresponding local composite index was written as η^{local}. It could be calculated as eq 10

$$\eta^{\text{local}} = \sum_{i<j, y_{ij}=1} (\gamma_i \gamma_j)^{0.5} \tag{14}$$

where γ_i and γ_j respectively represent the i-th and j-th nodes (atoms) in a connected molecular graph.

Similarly, the value for the corresponding reference alkane η_R^{local} could also be calculated. The local functionality contribution η_F^{local} (without considering global topology) could be calculated as follows:

$$\eta_F^{\text{local}} = \eta_R^{\text{local}} - \eta^{\text{local}} \tag{15}$$

For the calculation of branching, consideration of the local topology was sufficient. Branching was calculated with respect to theηvalue of the corresponding normal alkane (the straight-chain compound of the same vertex count obtained from the reference alkane), η_N^{local} , which might be conveniently calculated as (when $N_V \geq 3$):

$$\eta_N^{\text{local}} = 1.414 + (N_V - 3)0.5 \tag{16}$$

Finally, the branching index η_B could be calculated as:

$$\eta_B = \eta_N^{\text{local}} - \eta_R^{\text{local}} + 0.086N_R \tag{17}$$

where N_R was the number of rings in the molecular graph of the reference alkane, which could account for ring structures.

The premise of implementing the above calculations was the need to provide an adjacency matrix of the molecules. A molecule's adjacency matrix referred to a mathematical representation that described the connectivity between atoms in a molecule. In this matrix, each row and column corresponded to an atom in the molecule, and the elements of the matrix indicated whether there was a chemical bond between the atoms represented by the corresponding row and column indices. An example for the calculation of the adjacency matrix for cyclopentane was shown in **Figure S1**.

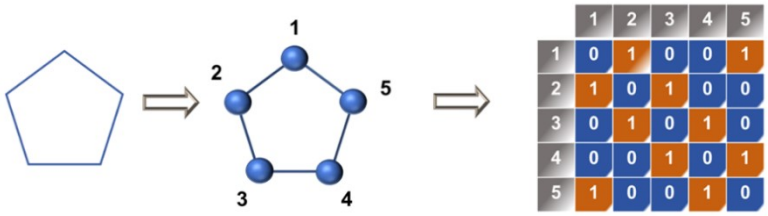

Figure S1. The adjacency matrix for cyclopentane.

3 The Hyperparameter Settings

Table S2. Search range and selected values of hyperparameters for MGNN

Hyperparameter	Search range	Responsivity
Number of multiscale blocks	[1, 2, 3, 4, 5]	3

Number of graph convolutional layers in each multiscale block	[2, 3, 4, 5, 6, 7, 8, 9, 10]	6
The hidden channel number of each graph convolutional layer	[32, 64, 96, 128, 160]	64

4 Classic graph neural network methods struggle to distinguish stereoisomers

Classic graph neural network methods struggle to distinguish stereoisomers due to the missing of the 3D information of the molecules as reported in literature (**Nat Rev Methods Primers.,** 2024, 4, 17; **Commun Mater.,** 2022, 3, 93; **Nat Mach Intell.**, 2022, 4, 127). Detailed reasons have been shown as follows:

Existing GNNs only consider the topology information of the molecules, neglecting the molecular geometry, that is, the three-dimensional spatial structure of a molecule. These works conduct self-supervised learning by masking and predicting in nodes, edges or contexts in the topology. Yet these tasks only enable the model to learn the laws of molecular graph such as which atom/group could be connected to a double bond, and lack the ability to learn the molecular geometry knowledge, which plays an important role in determining molecules' physical, chemical and biological activities. For example, the water solubility (a critical metric of drug-likeness) of the two molecules illustrated in **Figure S2** is different due to their differing geometries, even though they have the same topology .

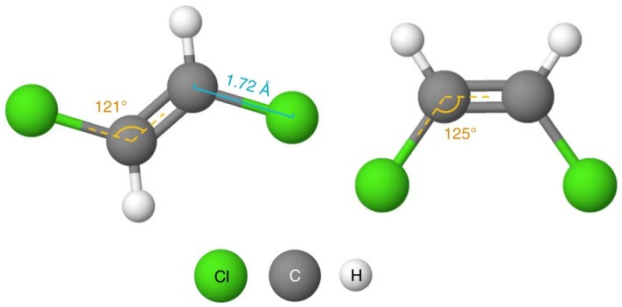

Figure S2. Comparison between two stereoisomers with the same topology but different geometries. The two chlorine atoms are on different sides in trans-1,2-dichloroethene (left) but the same side in cis-1,2-dichloroethene (right).

5 Specific Isomeric Effects

Specific Isomeric Effects on Density

As an important property of diesel and jet fuels, density directly related to energy density, combustion efficiency and storage and transportation efficiency, making it a crucial indicator for evaluating fuels performance. The different influence of the composite index (**Figure S3a**) and the branching index (**Figure S3b**) in determining the alkane density could be observed. The composite index not only considered the local topology characteristics inside the alkane molecules, but also integrated the global topology role of the overall structure, thus providing a more comprehensive structural description. The color change of the points visually showed the differences in the number of carbon atoms. The apparent stratification arose from the ring numbers of the alkane molecules (**Figure S3a**). The difference in ring numbers led to differences in molecular structure, which in turn affected the physical properties of alkanes, including their density.

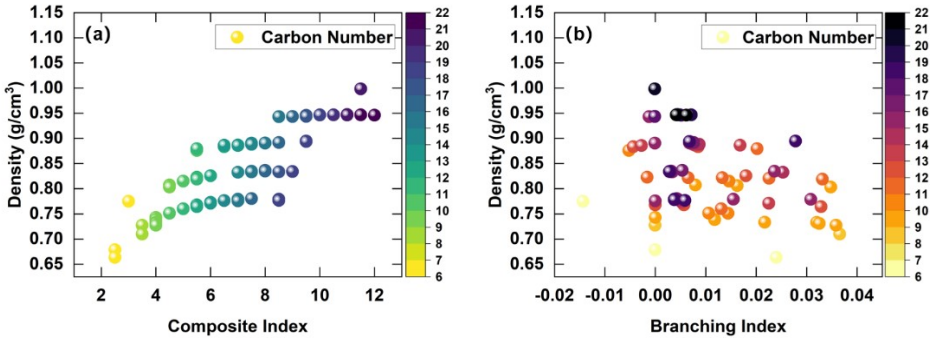

Figure S3. The relationship between density and composite and branching index of biofuels.

Similarly, the effects of the shape parameters Shape-P, Shape-Y, and Shape-X on the alkane density were analyzed in**Figure S4**. It was clearly observed that, among these topological descriptors there was a significant correlation between the alkane density and the composite index, Shape-P and Shape-X. These findings provided new ideas and methods for the regulation of alkane density.

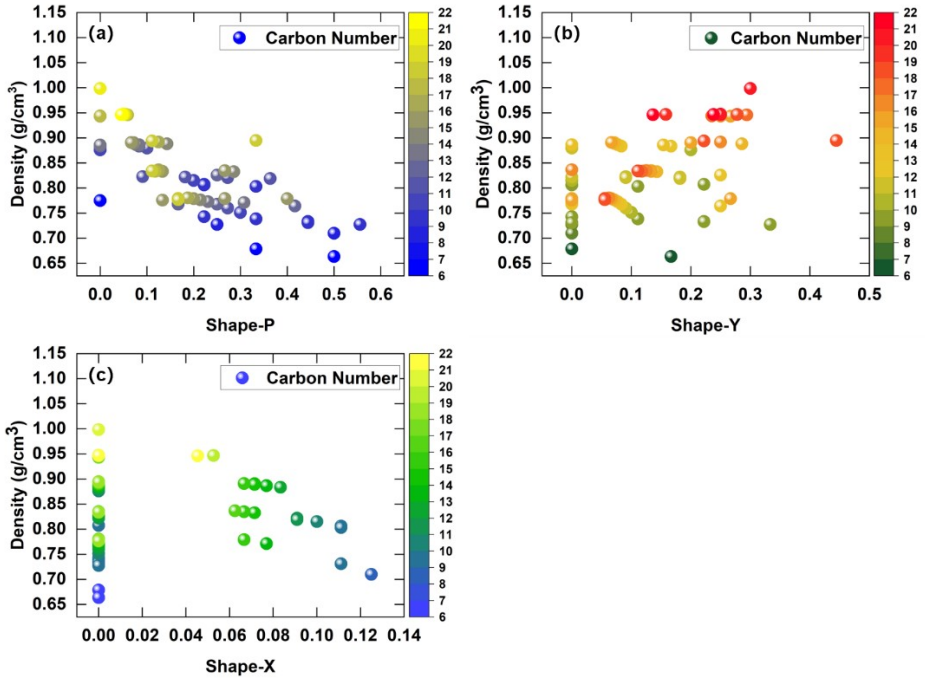

Figure S4. The relationship between density and shape parameters of biofuels.

To clarify the effect of the branching index on the density, biofuels were classified based on the total number of carbon atoms in order to observe the changing relationship between density and branching index when the number of carbon atoms was fixed (**Figure S5**). The blue dots in the figure represented cycloalkanes, while the green dots represented alkanes (chain alkanes) without cyclic structures. The red rectangular region intuitively reflected the range of density changes under different conditions. In each subgraph, the common feature of high-density biofuels was that they have a cyclic structure and a small branching index. The phenomenon that blue dots were generally above green dots strongly indicated the importance of cyclic structure for increasing the alkane fuel density. Further comparison of the changes in density of cyclic and chain alkanes in response to changes in branching index, differences could be found. For catenae with the same number of carbon atoms, the density remained largely stable and not significantly affected by changes in the branching index. However, for cycloalkanes, a decreasing density was observed in most cases as the branching index increased. This was might because the more branches, the more complex the spatial configuration of the molecule, and the more complex the molecular interactions. This might led to increased intermolecular repulsion, which reduced the molecular tightness, reducing the density of cycloalkanes. Based on the above analysis, in order to improve

the biofuels density, we could consider to improve the composite index and reduce the branching index, Shape-P and Shape-X.

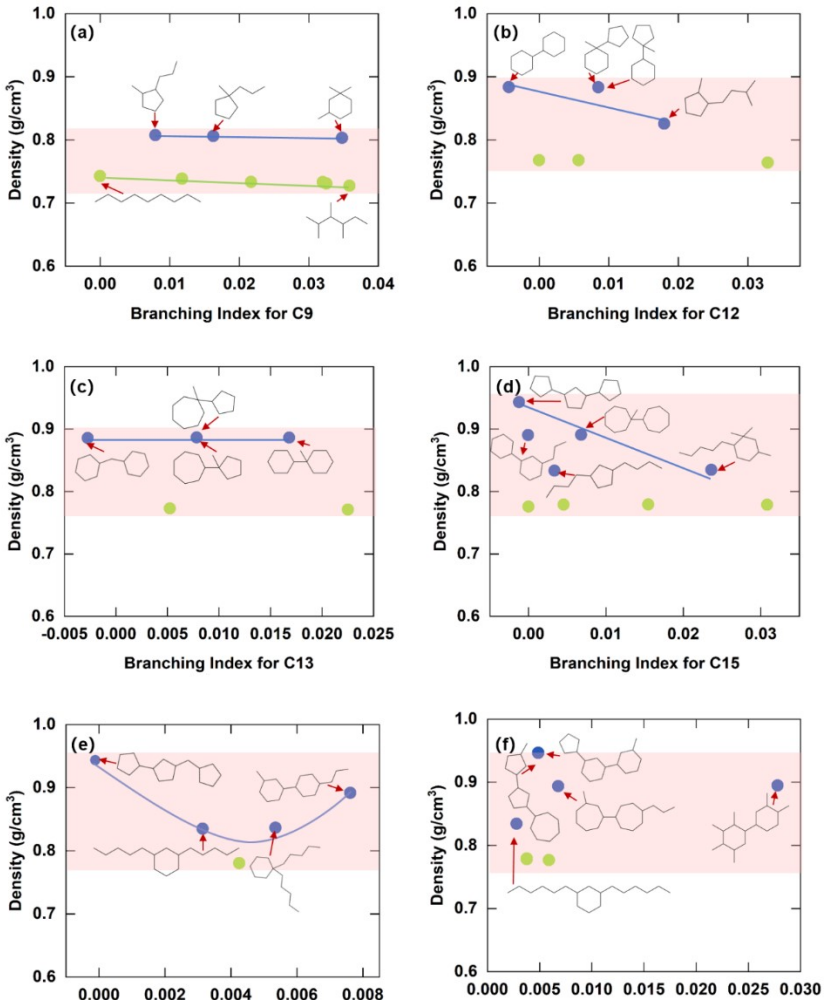

Figure S5. The effect of branching index on biofuels density

Specific Isomeric Effects on VNHOC

The volumetric heat of combustion referred to the amount of heat released when a unit volume of fuel was completely burned. This metric was typically used to measure the energy density of liquid fuels. Higher volumetric heat of combustion meant that the same volume of fuels contains more energy, which was crucial for long-distance flights of jet aircraft and the efficient operation of diesel engines. The relationship between

VHOC of biofuels and each topological descriptor was shown in **Figure S6** and **Figure S7**. Given the strong correlation between the volume combustion heat and the density of the fuel, it was observed that the trend of VHOC with each topological descriptor roughly matched the trend of the density with these descriptors. Therefore, the logic used in analyzing the density changes of biofuels could also be applied to the analysis of the volume combustion heat changes. This observation suggested that topological descriptors were not only crucial for understanding the density properties of fuels, but were also valuable in explaining their volumetric combustion thermal properties.

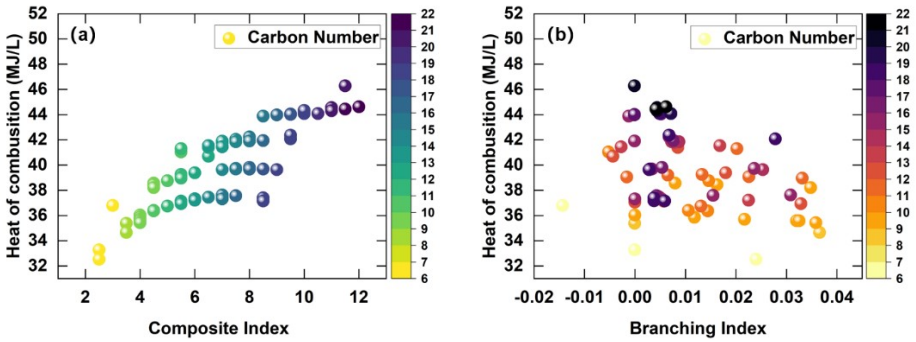

Figure S6. The relationship between VHOC and composite and branching index of biofuels.

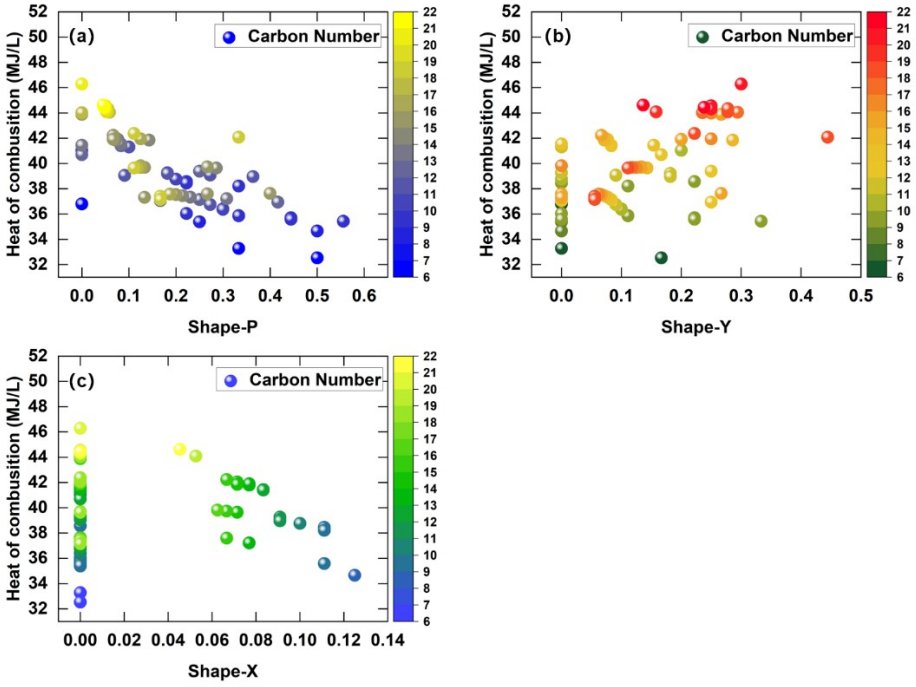

Figure S7. The relationship between VHOC and shape parameters of biofuels.

3 Specific Isomeric Effects on Boiling Point

The boiling point determined the evaporation rate and combustion characteristics of fuels. In practical applications, the boiling point of fuels needed to be optimized according to the specific application environment and performance requirements, thereby balancing the engine's cold start performance and combustion efficiency at high temperatures. For the relationship of biofuels between the boiling point and each topological descriptor showed a stronger linear correlation than the density and VHOC (**Figure S8** and **Figure S9**). This suggested that the boiling point of the biofuels was more directly responsive to the composite index. Indeed, the composite index, as a topological descriptor, had a tight correlation with the number of carbon atoms in the molecule. Therefore, when the various properties of biofuels changed with the composite index, we could find that this change was significantly similar to the trend based only on the number of carbon atoms. However, the composite index was unique in that it was not limited to reflecting the effect of the number of carbon atoms and further considering other topological structure parameters in the molecule.

In particular, the complex index was able to capture the influence of complex structural features such as the number of molecular loops. For example, with the same number of carbon atoms, alkanes with ring structure would show a higher complexity index compared to alkanes without ring structure. Thus, the complex exponent as a more comprehensive topological descriptor provided us with a more precise and comprehensive view to understand and predict the various properties of the fuel. The Shape-Y had a more pronounced effect on the boiling point compared to the Shape-P. For molecules with the same number of carbon atoms, those with a larger Shape-Y generally exhibited higher boiling points compared to those with a smaller Shape-Y.

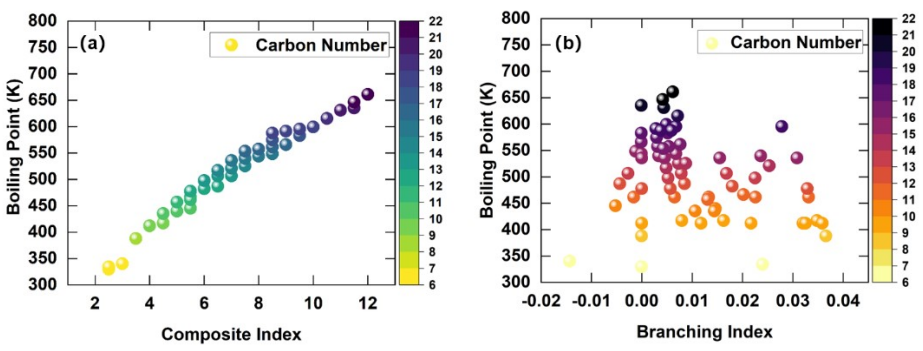

Figure S8. The relationship between BP and composite and branching index of biofuels

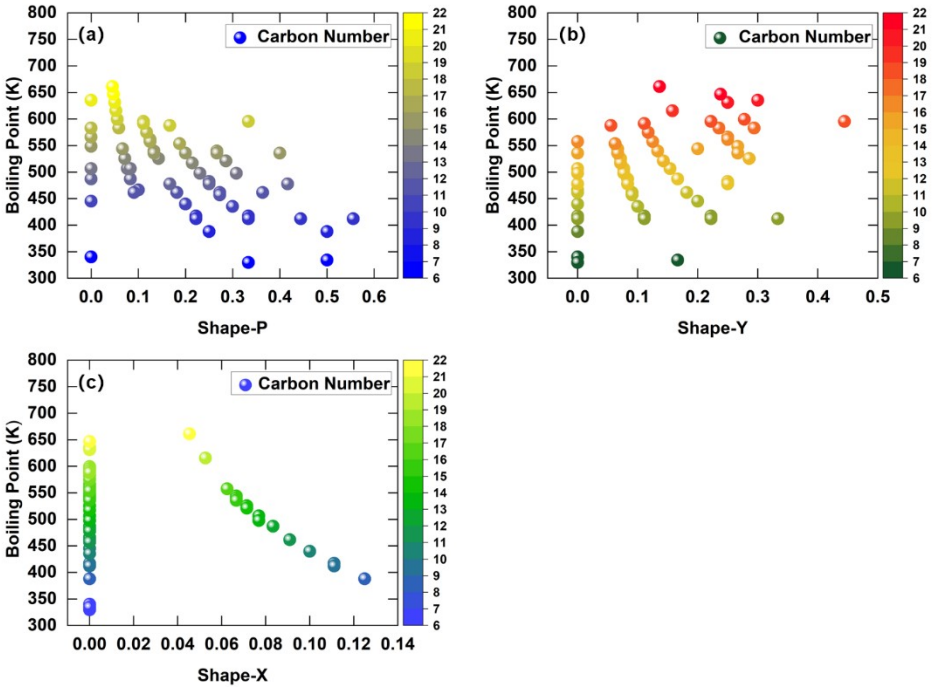

Figure S9 . The relationship between BP and shape parameters of biofuels.

4 Specific Isomeric Effects on Viscosity

Viscosity affected the flow ability of fuels within the fuels system. Appropriate viscosity ensured smooth transportation of fuel from the storage tank to the engine combustion chamber, which maintained the proper supply rate and ensured the engine operates normally. In the investigation into the relationship between the viscosity of biofuels and various topological descriptors (**Figure S10** and **Figure S11**), we noted that，when the carbon atom count within the biofuel molecule was low, the influence of composite index on fuel viscosity appeared to be relatively minor. However, as the number of carbon atoms escalated, the impact of these intricate factors on viscosity became increasingly pronounced, particularly at higher carbon atom counts. Concurrently, the correlation between viscosity and the shape factor Shape-P exhibited greater robustness in comparison to other properties of the biofuels under scrutiny. This relationship was non-linear. Specifically, at lower values of Shape-P, the viscosity of the biofuels decreased rapidly as Shape-P increases. However, as Shape-P continued to increase, the rate of viscosity decrease diminished progressively and eventually stabilized at a relatively low viscosity level.

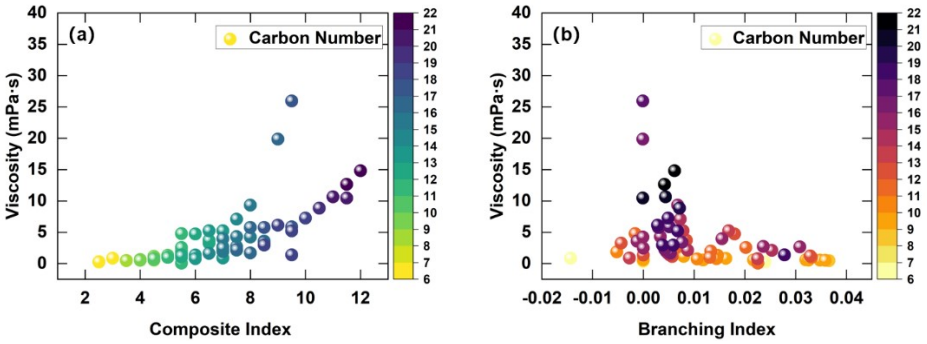

Figure S10. The relationship between viscosity and composite and branching index of biofuels.

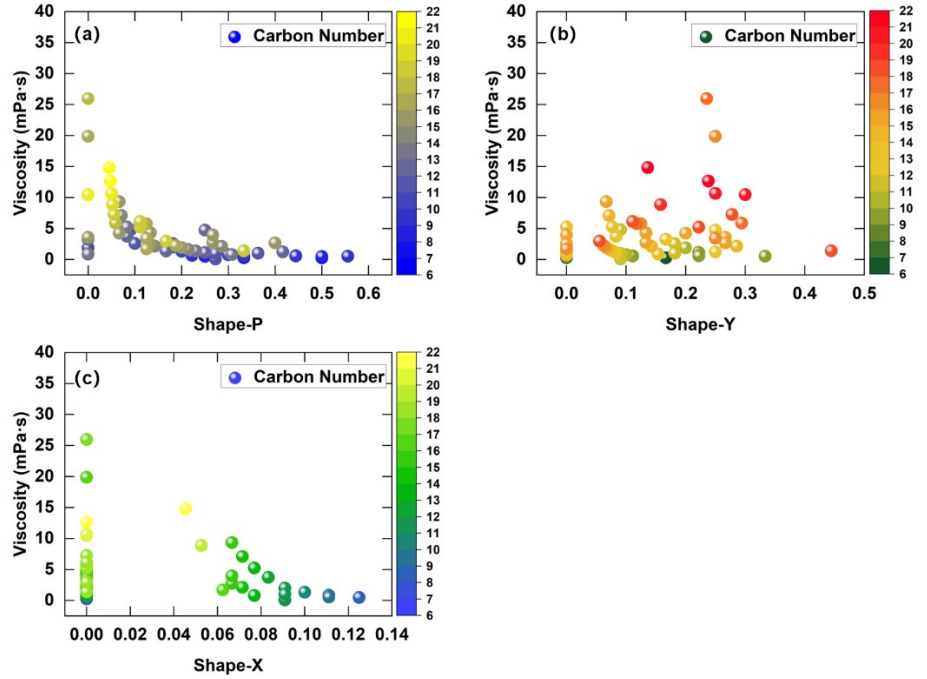

Figure S11. The relationship between viscosity and shape parameters of biofuels.

5 Specific Isomeric Effects on Flash Point

The flash point was the minimum temperature at which a liquid vaporizes and ignites with a certain ignition source under specified test conditions. Flash point was a safety index of flammable liquid storage, transportation, and use, and also a volatile index of flammable liquid. The flash point of biofuels exhibited a trend similar to that observed for the boiling point when analyzed in relation to each topological descriptor (**Figure S12** and **Figure S13**).

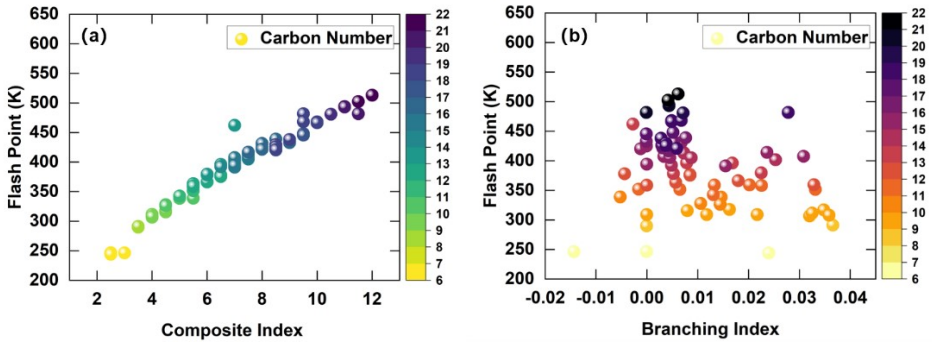

Figure S12. The relationship between FP and composite and branching index of biofuels.

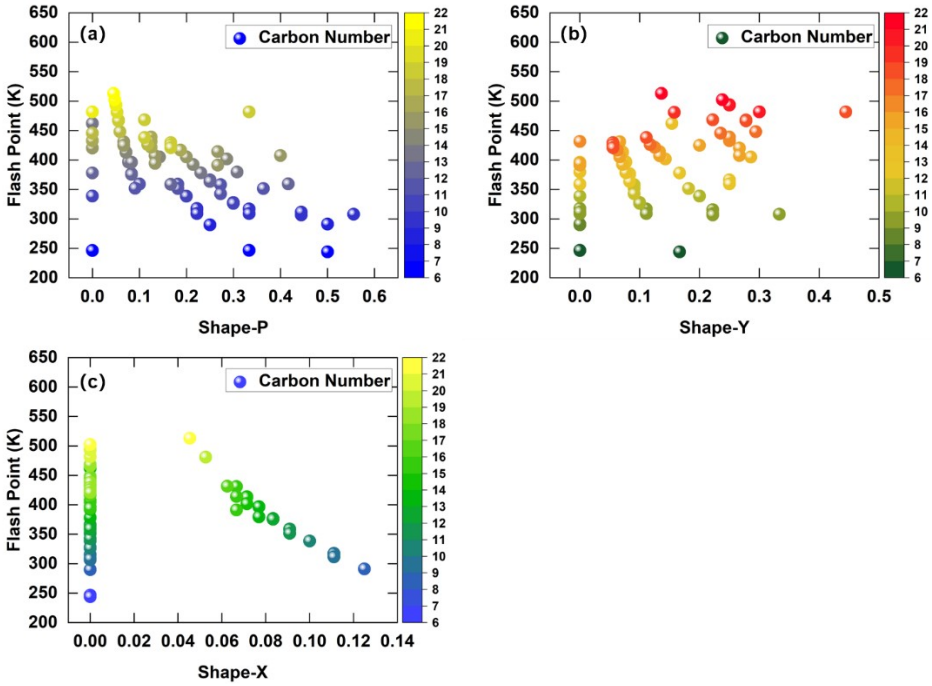

Figure S13. The relationship between BP and shape parameters of biofuels.

6 Specific Isomeric Effects on Cetane Number

Cetane number was an indicator of the spontaneous combustion of diesel fuel in a diesel engine. The higher the cetane number, the better the combustion performance of diesel, the less prone to knock and other abnormal combustion phenomena. The influence of topological descriptors on the cetane values of biofuels was notably complex (**Figure S14** and **Figure S15**), making it challenging to identify a direct correlation between any single topological descriptor and cetane values.

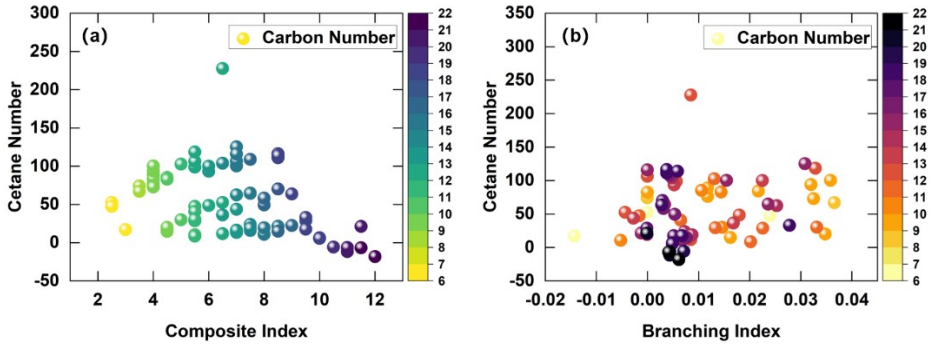

Figure S14. The relationship between CN and composite and branching index of biofuels.

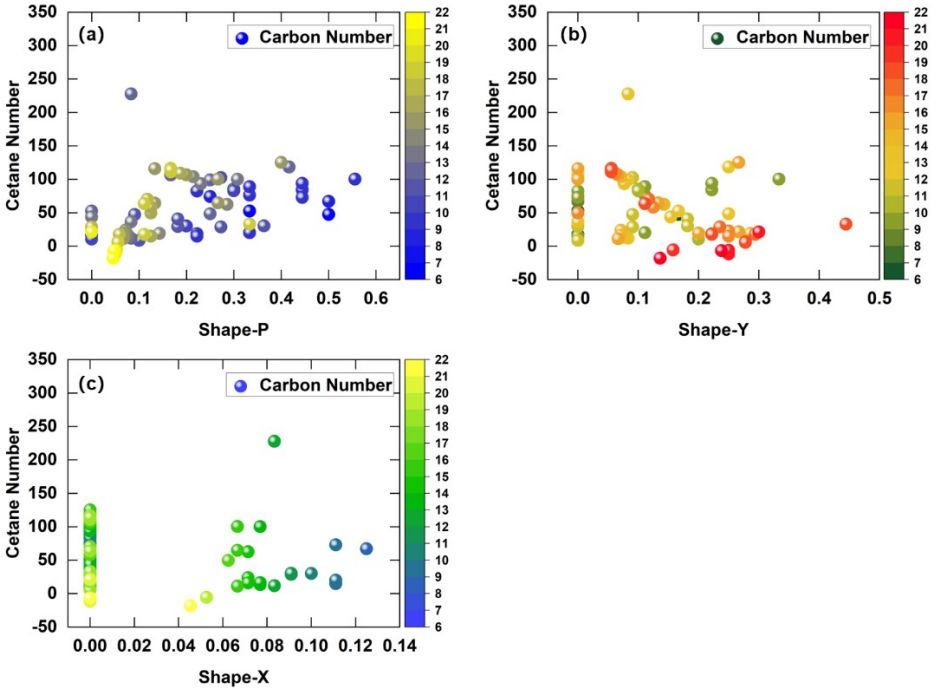

Figure S15. The relationship between CN and shape parameters of biofuels.