

Polymer-Unit Graph: Advancing Interpretability in Graph Neural Network Machine Learning for Organic Polymer Semiconductor Materials

Xinyue Zhang, Ye Sheng, Xiumin Liu, Jiong Yang, William A. Goddard III, Caichao Ye,^{*} and Wenqing Zhang^{*}



Cite This: *J. Chem. Theory Comput.* 2024, 20, 2908–2920



Read Online

ACCESS |



Metrics & More



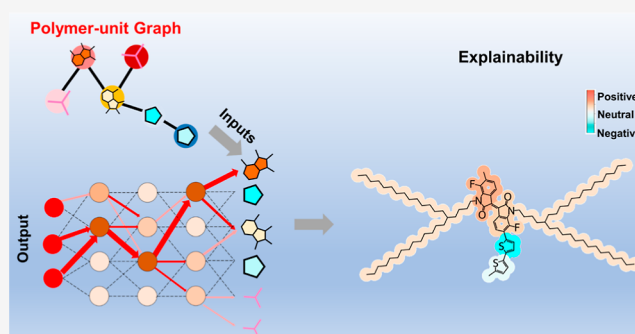
Article Recommendations



Supporting Information

ABSTRACT: The graph representation of complex materials plays a crucial role in the field of inorganic and organic materials investigations for developing data-centric materials science, such as those using graph neural networks (GNNs). However, the currently prevalent GNN models are primarily employed for investigating periodic crystals and organic small molecule data, yet they still encounter challenges in terms of interpretability and computational efficiency when applied to polymer monomers and organic macromolecules data. There is still a lack of graph representation of organic polymers and macromolecules specifically tailored for GNN models to explore the structural characteristics.

The *Polymer-unit Graph*, a novel coarse-grained graph representation method introduced in study, is dedicated to expressing and analyzing polymers and macromolecules. By incorporating the *Polymer-unit Graph* into the GNN models and analyzing the organic semiconductor (OSC) materials database, it becomes possible to uncover intricate structure–property relationships involving branched-chain engineering, fluoridation substitution, and donor–acceptor combination effects on the elementary structure of OSC polymers. Furthermore, the *Polymer-unit Graph* enables visualizing the relationship between target properties and polymer units while reducing training time by an impressive 98% and minimizing molecular graph representation models. In conclusion, the *Polymer-unit Graph* successfully integrates the concept of *Polymer-unit* into the field of GNNs, enabling more accurate analysis and understanding of organic polymers and macromolecules.



1. INTRODUCTION

Organic semiconductor (OSC) materials offer significant advantages such as remarkable flexibility, ductility, and cost-effective preparation,¹ as well as immense potential applications in fields such as photovoltaic materials,^{2,3} field effect transistors,^{4–7} light-emitting diodes,^{8,9} and more. Among the numerous physical and chemical properties associated with OSC materials, carrier mobility and molecular frontier orbital energy levels are particularly crucial. In order to develop organic polymer materials with high charge carrier mobility and appropriate molecular orbital energy levels, extensive testing of different material structures within a large parameter space is necessary. This requires continuous exploration and refinement of the initial materials to improve their properties. However, the traditional trial and error process of developing organic polymers is time-consuming. Fortunately, with the ongoing advancements in technology, researchers are actively working toward expediting this process and implementing more efficient strategies.

Deep learning is a data-driven scientific field that employs algorithms to reveal hidden relationships within data sets, offering valuable insights for experimental scientific pursuits.

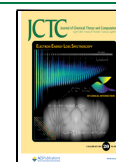
For OSCs, a significant amount of literature has been accumulated since their initial discovery in 1956,¹⁰ this vast body of literature contains valuable experimental data related to OSCs. By harnessing and fully utilizing this wealth of data, we can undoubtedly accelerate the development process of OSC materials. In recent years, the utilization of graph neural network (GNN) models has gained significant traction in the field of materials science.^{11,12} These models have been applied in various aspects, including the accurate prediction of material properties with minimal computational resources.^{13–21} By leveraging the power of GNNs, researchers can efficiently explore and analyze the structure–properties relationships, shedding light on the underlying mechanisms that govern material behavior.^{22,23} Additionally, GNNs have been

Received: December 19, 2023

Revised: March 12, 2024

Accepted: March 12, 2024

Published: March 29, 2024



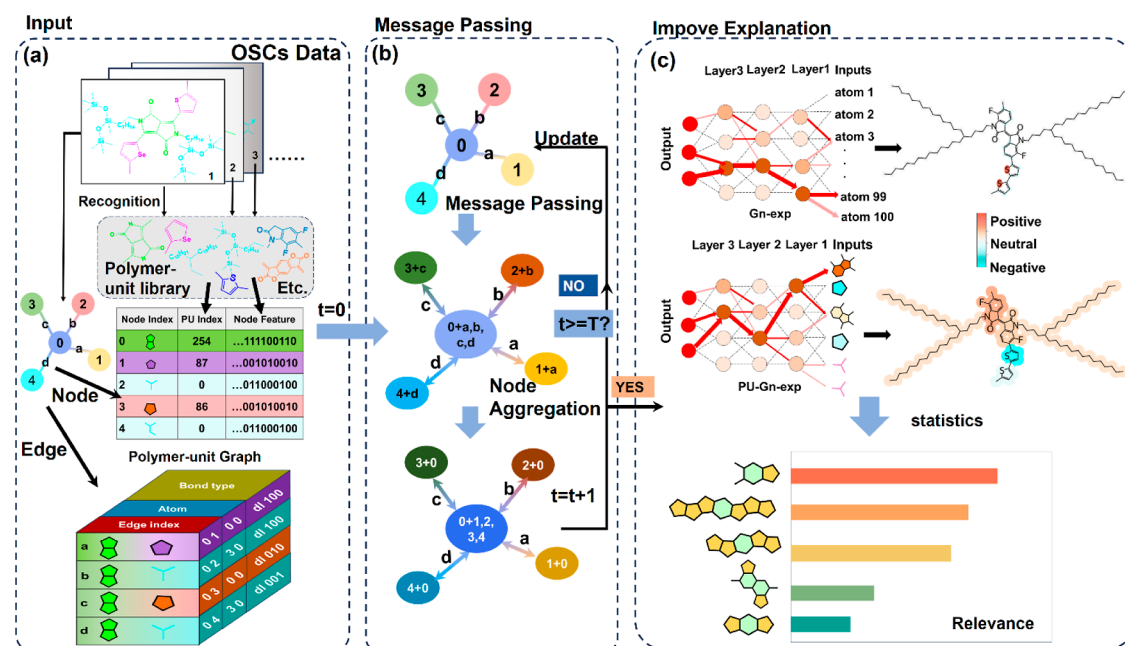


Figure 1. Overview of the application of the Polymer-unit Graph. (a) The generation diagram of the OSC Polymer-unit Graph. (b) Schematic diagram of the Message passing GNN. (c) The OSC Polymer-unit Graph has been applied to enhance the interpretability of MPNN models for acquiring structure–mobility relationships in organic polymers.

instrumental in the development of new materials that are specifically tailored to exhibit desired properties, opening up new possibilities for designing materials with targeted functionalities.^{24–31} Indeed, the crystal graph convolutional neural network (CGCNN) is a specialized GNN that is designed specifically for processing and predicting properties of crystalline materials.³² The CGCNN employs a residual structure to effectively propagate information across multiple channels, thereby addressing the issue of gradient vanishing. To avoid an increase in parameter count, expansive convolution is utilized to capture long-range dependencies. This model implements the convolutional representation of generalized crystal graphs for periodic crystal systems and enables a property prediction with DFT-level accuracy. The Messaging Passing Neural Network Framework (MPNN) was proposed by Gilmer et al.³³ in 2017. The *mol*-MPNN model is based on the MPNN framework. This model successfully predicts 13 targets in the QM9 data set, which consists of 130k small molecules with atomic numbers less than 9.³³ The application of GNNs has proven to be highly effective in addressing the challenges associated with small molecules and crystal materials.

OSC materials are mostly organic polymers and organic macromolecules. While there has been significant research on deep learning for organic polymer materials,^{34–38} the development of GNN models specifically tailored to capture the unit-structural characteristics of these organic polymers has been limited. This limitation arises from the challenges and complexities of accurately representing polymer structure in digital format,³⁹ highlights the need for tailored graph representation of organic polymers and macromolecules for exploring structural characteristics. Consequently, it becomes imperative to devise simplified methods for inputting structural representations that are well-suited for polymer GNN models.

Polymers and macromolecules are often formed by a series of reactions from a number of small molecular precursors,

namely, polymer-unit (also named as repeat units),^{35,40} in the same way that tens of thousands of English words can be formed using only 26 letters, a wide range of polymers can be generated from a limited number of polymer units through their different combinations. In 1993, Sumpter and Noid⁴¹ proposed a method of using polymer repeat units (namely, polymer-units) as nodes in neural networks (prior to the proposal of GNN), which successfully predicted nine properties for 357 different polymers with an average prediction error rate of only 3%. Thus, the effectiveness of deep learning based on polymer unit nodes is evident. Queen³⁵ et al. utilized a combination of small molecular acids and alcohols as polymerization components (a total of 17 varieties) to generate 247 different types of polyester through permutations and combinations. They conducted measurements on the glass transition temperature (T_g) and intrinsic viscosity (IV) of each polyester sample, ultimately establishing a comprehensive polyester database. The researchers also developed a model called POLYMERGNN, where each polyester is represented as an acid and alcohol pair, which is incorporated into the molecular graph and trained collaboratively as two independent GNN blocks. The prediction of T_g and IV can be achieved using POLYMERGNN by inputting the types and ratios of acids and alcohols. The D–A type OSC is a crucial constituent in the realm of OSC materials.⁴² The advantage of D–A type OSC materials lies in the ability to achieve optimal frontier molecular orbital energy levels through precise adjustment of donor and acceptor groups within the polymers.⁴³ The donor and acceptor groups, commonly referred to as small organic molecules, include well-known donor units such as *thiophene* and *selenophene*, as well as the widely used acceptor units *benzothiazole* (BT) and *pyrrole pyrrodione* (DPP). These donor and acceptor units also belong to the category of polymer-units.

The role of the polymer unit in the study of polymer is crucial, making it highly significant to incorporate the concept

of “polymer-unit” into GNN research within the polymer field. Recently, we proposed a Python-based script named polymer-unit-recognition script (*PURS*),⁴⁴ which is a program for recognizing the polymer-unit based on SMILES.⁴⁵ Its purpose is to identify the polymer-unit composition of polymers and macromolecules from SMILES. The Polymer-unit Fingerprint (*PUFp*) is specifically designed to enable the application of the “polymer-unit” concept in classical machine learning models such as *Random Forest*,⁴⁶ *Support Vector Machine*,⁴⁷ and *K-neighbor Algorithm*.^{48,49} The *PUFp*, however, is limited to a one dimension one-hot vector and lacks the specific properties of polymer-units (node information) as well as the connection relationships between polymer-units (edge information), making it unsuitable for application in GNN models. The GNN model excels in discerning implicit relationships from intricate inputs and is adept at handling more intricate and high-dimensional input data.^{50,51} Consequently, it enables the design of a more informative input to articulate the “polymer-unit”.

The molecular graph serves as a graph representation of the molecular structure, depicting atoms and chemical bonds as nodes and edges of the graph. The utilization of molecular graph inputs is frequently observed in MPNN models employed for small molecules.³³ However, the characteristics of polymer materials include a large number of atoms and a complex structure (with nested ring structures and various branch chains), which makes the molecular graph complex and difficult to explain. In this work, we propose a coarse-grained graph representation method called the *Polymer-unit Graph*, which introduces the concept of “polymer-unit” into molecular graphs. The polymer’s repeated units (polymer-unit) are considered as the graph’s nodes, while the interconnection between polymer-unit is regarded as the edge matrix of the graph. Node features are generated using graph embedding methods.⁵² The utilization of *Polymer-unit Graph* enhances the representation of the polymer’s structure in the GNN model, allowing for a more visualizable and explainable analysis. The utilization of the *Polymer-unit Graph* presents a fresh perspective to enhance the interpretability of GNN models. It allows for an analysis of the structure–activity relationship in OSC materials from the standpoint of polymer-units. The interaction between polymerization units in organic macromolecules and polymer materials plays a significant role in determining the macroscopic properties of these OSC materials. Therefore, analyzing the structure–activity relationship from the perspective of polymerization units holds immense scientific value in the field of such materials.^{53–57}

As Figure 1 shows, the application of *Polymer-unit Graph* in MPNN models provides promising results. The polymer unit method utilizes coarse-grained techniques, which provide greater chemical significance when applied to organic polymers and macromolecular materials. This approach forms coarse-grained units consisting of common small molecular groups and branched chains, enhancing the interpretability of these materials.

To explore the application of the polymer-unit in interpretability and visualization, an OSC *polymer-unit Graph* was integrated with the *gn-exp*⁵⁸ model, resulting in the *PU-gn-exp* model. This model is utilized to investigate the structure–property relationships of OSC materials. The *PU-gn-exp* model can effectively identify the polymer-unit that enhances carrier mobility through mechanisms such as fluorination, branch chain engineering, and donor–acceptor combination. Combin-

ing the OSC *Polymer-unit Graph* with the baseline model *mol-MPNN* to create *PU-MPNN*, and testing both models using the OSC materials database. *PU-MPNN* demonstrated a remarkable 98% reduction in training time compared to *mol-MPNN* while maintaining the same level of prediction accuracy measured by the mean square error (MSE). This showcases the efficiency of *Polymer-unit Graph* as an alternative with a similar prediction performance. This research focuses on these key innovations:

- The concept of *Polymer-unit Graph* is effectively employed in the context of GNNs.
- The interpretable model (*gn-exp*) is utilized to apply the *Polymer-unit Graph*, enabling the exploration of structure–property relationships of organic macromolecules/polymers based on polymer-units, providing valuable insights into their functionalities.
- The *Polymer-unit Graph* is an efficient technique that improves the computational efficiency of prediction models for organic macromolecules/polymers. It reduces training time to only 2% of the baseline model, benefiting researchers and practitioners in this field.

Our findings highlight the significant benefits of applying the polymer-unit concept within a GNN model for organic polymers and macromolecular materials. This approach not only enables effective analysis of the structure–property relationship but also enhances the computational efficiency. Consequently, this work facilitates in-depth mechanistic research on organic polymer materials, contributing to materials innovation in this field.

2. RESULTS AND DISCUSSION

2.1. Polymer-Unit Graph Generation. The generation of *Polymer-unit Graphs* can be partitioned into three steps (Figure 1a): (1) Identification of polymer-units; (2) Formation of the polymer-unit database; and (3) Generation of *polymer-unit Graphs* from the polymer-unit database index.

1. The *PURS* program is utilized for identification of polymer-units, wherein these polymer-units and their interconnections are determined based on the SMILES codes present in the input data sets.
2. The polymer-units identified from the OSC data set were converted into standard SMILES codes using the *Rdkit* program,⁵⁹ ensuring uniqueness and completeness of the data set. This process led to the establishment of a polymer-unit database for the OSC data set. To represent the structural information on each polymer-unit, we utilized one-dimensional vectors, which essentially serve as a graph embedding representation. The reason for utilizing graph embedding as the feature of nodes, instead of directly using the molecular graph data of the aggregate unit as node feature, is driven by its capability to avoid issues associated with dimensional inconsistencies of node features and to uphold a reasonable feature space dimension. After conducting extensive testing on numerous graph embedding methods, we identified two approaches that showed the most promising results: MACCS code⁶⁰ (166 bit), and *Genwl* graph embedding method⁶¹ (280 bit) utilizing the Weisfeiler Lehman algorithm.
3. The OSC *Polymer-unit Graph* for each data consists of nodes and edges, as shown in Figure 1a. The node

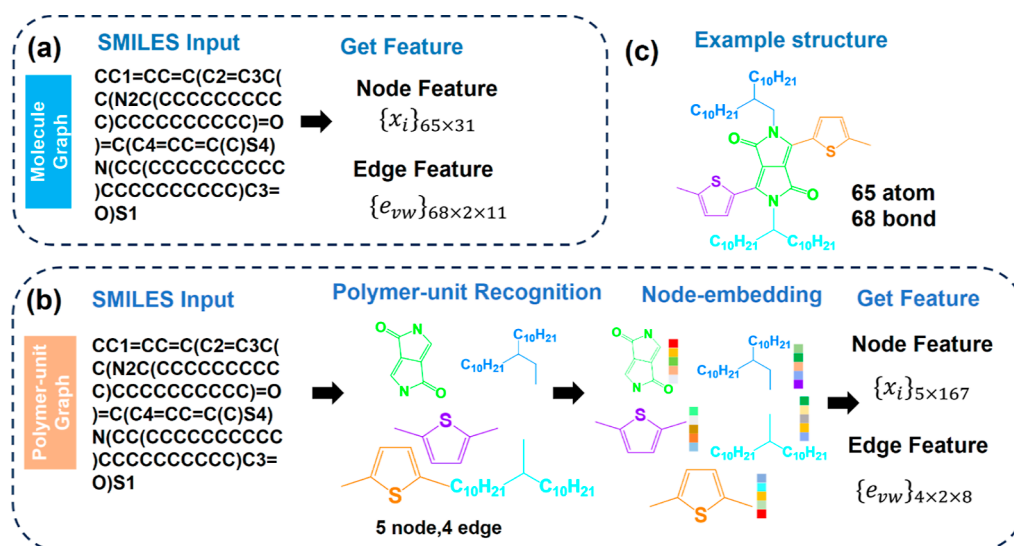


Figure 2. Comparison between the molecular graph and the polymer-unit graph. (a) The generation of molecular graph. (b) The generation of OSC polymer-unit graph. (c) Sample molecule.⁶²

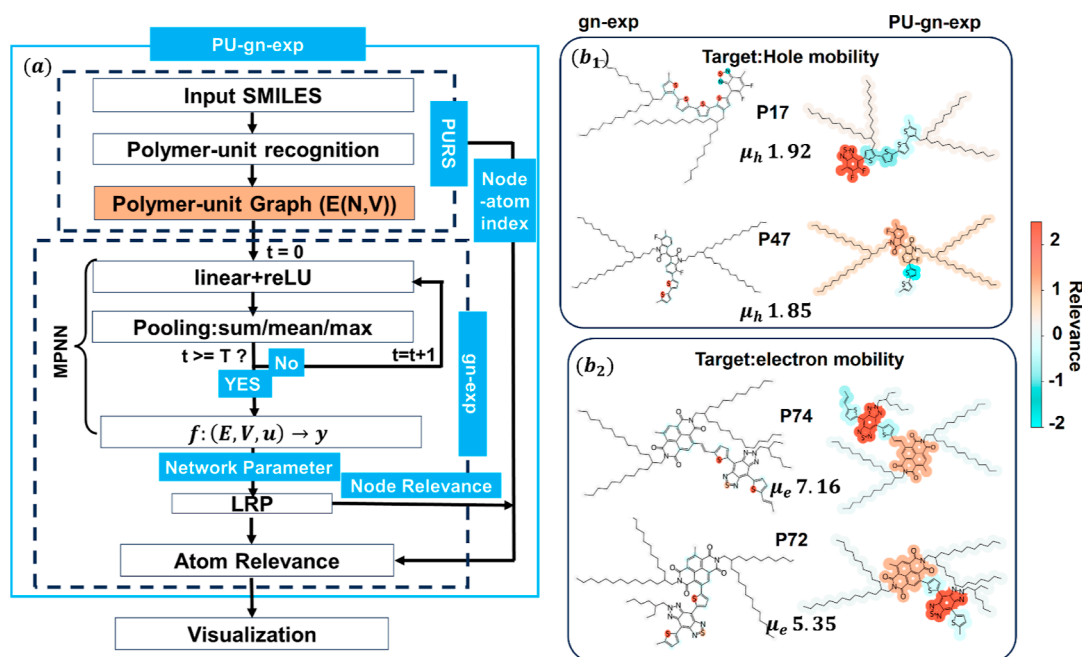


Figure 3. Framework and performance of PU-gn-exp. (a) The structure of the PU-gn-exp framework. (b) The explainability between gn-exp and PU-gn-exp is compared for carrier mobility tasks, (b1) the target is hole carrier mobility,⁶³ (b2) the target is electron carrier mobility.⁶⁴

information consists of a collection of graphs embedding the representation that corresponds to the polymer-unit indexed from the polymer-unit library. The PURS program not only recognizes polymer-unit but also identifies the topological connectivity between these units, referred to as the edge index. The edge information represents the bond between two polymer-units and is given by an 8×1 one-dimensional vector which divided into three parts. The first part indicates the index number of the connected polymer-units in the polymer-units library. The second part includes the atomic numbers of the connected atoms and their bond length. The third part is a one-hot vector that encodes stereochemical information, denoting the presence or absence of a chiral center and its type,

considering three possibilities: clockwise circular arrangement with respect to the reference atom, counter-clockwise circular arrangement with respect to the reference atom, or a nonchiral state.

The sample molecule (Figure 2c) illustrates the difference between the molecular graph (Figure 2a) and the OSC Polymer-unit Graph (Figure 2b). The OSC Polymer-unit Graph involves the process of identifying polymer-units and converting them into graph features, while the molecular graph directly generates a node feature for each atom. For the example molecule, the node matrix of the molecular Graph is 65×31 and the edge matrix is $68 \times 2 \times 11$, while the node matrix of the OSC Polymer-unit Graph is 5×167 and the edge matrix is $4 \times 2 \times 8$. By comparison, it can be seen that OSC Polymer-unit Graph actually encapsulates the structural

information on polymers in the form of graph embedding in the node information, taking the polymer-units as the nodes, while simplifying the connection relationship between the nodes.

2.2. Polymer-Unit Graph Enhances the Visualization and Interpretability. GNNs are commonly employed for decision-making in intricate situations, yet the rationale behind their choices often remains obscure. In 2019, Federico Baldassarre and Hossein Azizpour used the gradient method and decomposition method to investigate the interpretability of GNNs on both toy and chemical data sets, proposing the *gn-exp*, an interpretable GNN model.⁵⁸ This study represents the pioneering work in exploring the interpretability of GNNs and successfully used PyTorch's⁶⁵ automatic differentiation function to implement three distinct interpretability methods [sensitivity analysis (SA), guided backpropagation (GBP), and layer-wise relevance propagation (LRP)] on the MPNN model. Here, the *gn-exp* model is therefore selected as the baseline for investigating the application of *Polymer-unit Graph* on interpretable GNN. Additionally, we employ the LRP method because of its excellent interpretability, which has demonstrated superior performance in chemical tasks, as our chosen analytical technique.

Here, the *PU-gn-exp* model is constructed by reforming the input part of the *gn-exp* model, as illustrated in Figure 3a. Prior to inputting the polymer structure into *gn-exp*, the OSC *Polymer-unit Graph* is generated through the *PURS* program. *PU-gn-exp* utilizes a 166 bit MACCS as the characteristic value for each node. In addition, in the visualization part, since *PU-gn-exp* returns the correlation index of each node, it is necessary to assign the correlation index to each atom based on the atom-PU correspondence generated by *PURS*. The *PU-gn-exp* model follows the default setting of the baseline *gn-exp* model, which includes three layers. Each layer consists of three components: edge update aggregation, node update aggregation, and overall graph update and aggregation; these components are further explained in the *Methods Section*. Both the *PU-gn-exp* and *gn-exp* models were trained as classification models, achieving satisfactory accuracy and providing reliable analysis results (Figure S1). The target labels (carrier mobility) in this task are categorized into four groups based on the magnitude of their values (SI-Transform *gn-exp* into a classification model), following the specific classification method outlined in Table 1. The hyperparameters of *gn-exp* and *PU-gn-exp* are adjusted based on the default settings, considering the specific requirements of the training task (Tables S1 and S2).

Figure 3b displays the visualization result of the correlation between structure and mobility for OSC polymers, and the

correlation between each atom/polymer unit and the target property (mobility) is represented by a color bar. The color bar on the right illustrates the relationship between correlation and color. Positive and negative correlation values indicate whether the correlation is positive or negative. In Figure 3b, the left column shows the visualization result of the baseline *gn-exp* model on OSC data, while the right column depicts the visualization result of the *PU-gn-exp* model. Specifically, Figure 3b(1) illustrates the visualization result for hole mobility as the target attribute, and Figure 3b(2) showcases the visualization result for electron mobility as the target attribute. Here, the utilization of the OSC *Polymer-unit Graph* in *gn-exp* has significant implications. By incorporation of the *Polymer-unit Graph*, the interpretability and visualization of MPNN are enhanced. This is because the correlation between key properties like carrier mobility is found to rely more prominently on the interactions between polymer-units, rather than on atom-level interactions. To better showcase the improved interpretability of the *PU-gn-exp* model in OSC materials, we have included additional cases in Figures S3 and S4, demonstrating visual comparisons of carrier mobility between *PU-gn-exp* and its baseline model in the OSC data set. The visualization of polymer-units serves as a valuable tool for experimental scientists, enabling them to make adjustments to the composition of the OSC materials. Since polymers are primarily synthesized using these polymer-units, the visual representation facilitates a deeper understanding of the relationships between the polymer structure and its performance in terms of carrier mobility. It should be noted that visualization results for the same structure may show significant discrepancies depending on whether the smallest unit is considered as the aggregation unit or the atom. However, these discrepancies do not undermine the validity of the model. Instead, they highlight the distinct differences between PU-level structure-activity analysis and atomic-level analysis. For instance, when examining the structure P17 in Figure 3 b1, the baseline model's visual analysis revealed a positive correlation among sulfur atoms, while the visual results from *PU-gn-exp* indicated an overall negative correlation with thiophene. The presence of sulfur atoms in organic systems as electron donors can aid in mitigating charge localization,^{66,67} thus establishing a positive correlation with sulfur atoms. Conversely, an excessive amount of thiophene units can impede the crystallinity of OSC materials,⁶⁸ leading to a negative correlation.

The OSC *Polymer-unit Graph*, which is based on polymer-units, provides a more accurate representation of the structure-property relationship regarding OSC mobility. Figures 4 and 5 depict the structure-mobility relationship analysis of hole and electron mobility in the OSC data, respectively. The mobility values presented in Figures 4 and 5 are all obtained from relevant literature sources.^{1,43,62–64,69–81} The classification accuracies of visual analysis models were 81.96% (hole mobility) and 88.20% (electron mobility), respectively. In the figures, the colorbar on the right illustrates the corresponding relationship between color and correlation. Structures with a high positive correlation are represented in orange, while structures with a high negative correlation are depicted in blue. Structures with weak correlation appear to be colorless. The *PU-gn-exp* model provides an effective method for studying the structure-activity relationship, as shown in Figures 4 and 5. In Figure 5e, the visualization of a single data set may not provide a clear explanation since the polymerization units of a polymer are mostly positively correlated

Table 1. Distribution of the OSCs Data Set^a

parameter	p-type	n-type
total data number	566	275
$\mu > 10$	15	1
$10 > \mu > 4$	67	24
$4 > \mu > 1$	134	80
$1 > \mu > 0$	350	170
average atomic number	76	83
maximum atomic number	182	159

^a μ represents the carrier mobility. All OSCs data are available in the Supporting Information.

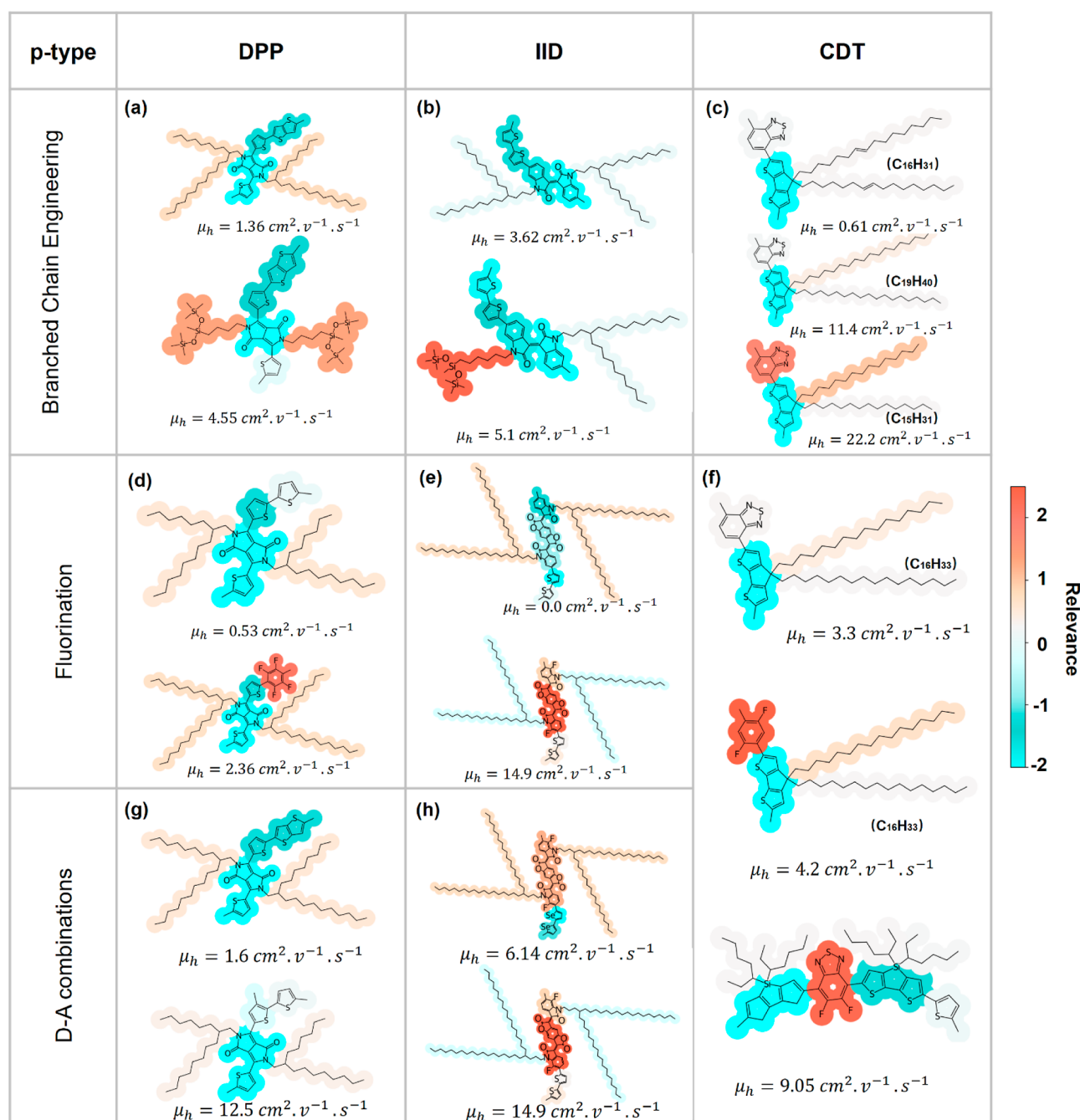


Figure 4. Present study offers a visual analysis of the correlation between mobility and molecular structure in P-type OSCs. The horizontal axis represents three representative OSC materials: diketopyrrolopyrrole (DPP), isoindigo (IID), and cyclopentadithiophene (CDT). The vertical axis corresponds to three typical effects that enhance mobility: branch chain engineering, fluorination, and donor–acceptor (D–A) combinations. (a) DPP-type OSCs of branch chain engineering effect, (b) IID-type OSCs of branch chain engineering effect, (c) CDT-type OSCs of branch chain engineering effect, (d) DPP-type OSCs of fluorination effect, (e) IID-type OSCs of fluorination effect, (f) CDT-type OSCs of fluorination effect and D–A combinations effect, (g) DPP-type OSCs of D–A combinations effect, and (h) IID-type OSCs of D–A combinations effect.

throughout the period. However, by comparing the structural changes between two polymers with nearly identical structures, it can be inferred that the inclusion of fluorinated *isoindigo* structure leads to a higher correlation compared to the nonfluorinated *isoindigo* structure. This inference is further supported by comparing the experimental data of the two polymers, where the polymer with the fluorinated *isoindigo* structure exhibits a higher mobility ($9.7 \text{ cm}^2/\text{V}\cdot\text{s}^{71}$ compared to the nonfluorinated *isoindigo* structure $6.67 \text{ cm}^2/\text{V}\cdot\text{s}^{82}$). This validation confirms the accuracy of the model's explanation.

Each box in the figure shows slight changes in the molecular structure, enabling the verification of the analytical rationality of the *PU-gn-exp* model through comparison.

The *PU-gn-exp* model accurately identifies the polymer-units that enhance the target property of carrier mobility. This is illustrated in Figures 4 and 5, where the analysis reveals the following key points:

1. In the column of branched-chain engineering (Figures 4a–c and 5a–c), it is evident that polymers with higher mobility have stronger correlations in their branch

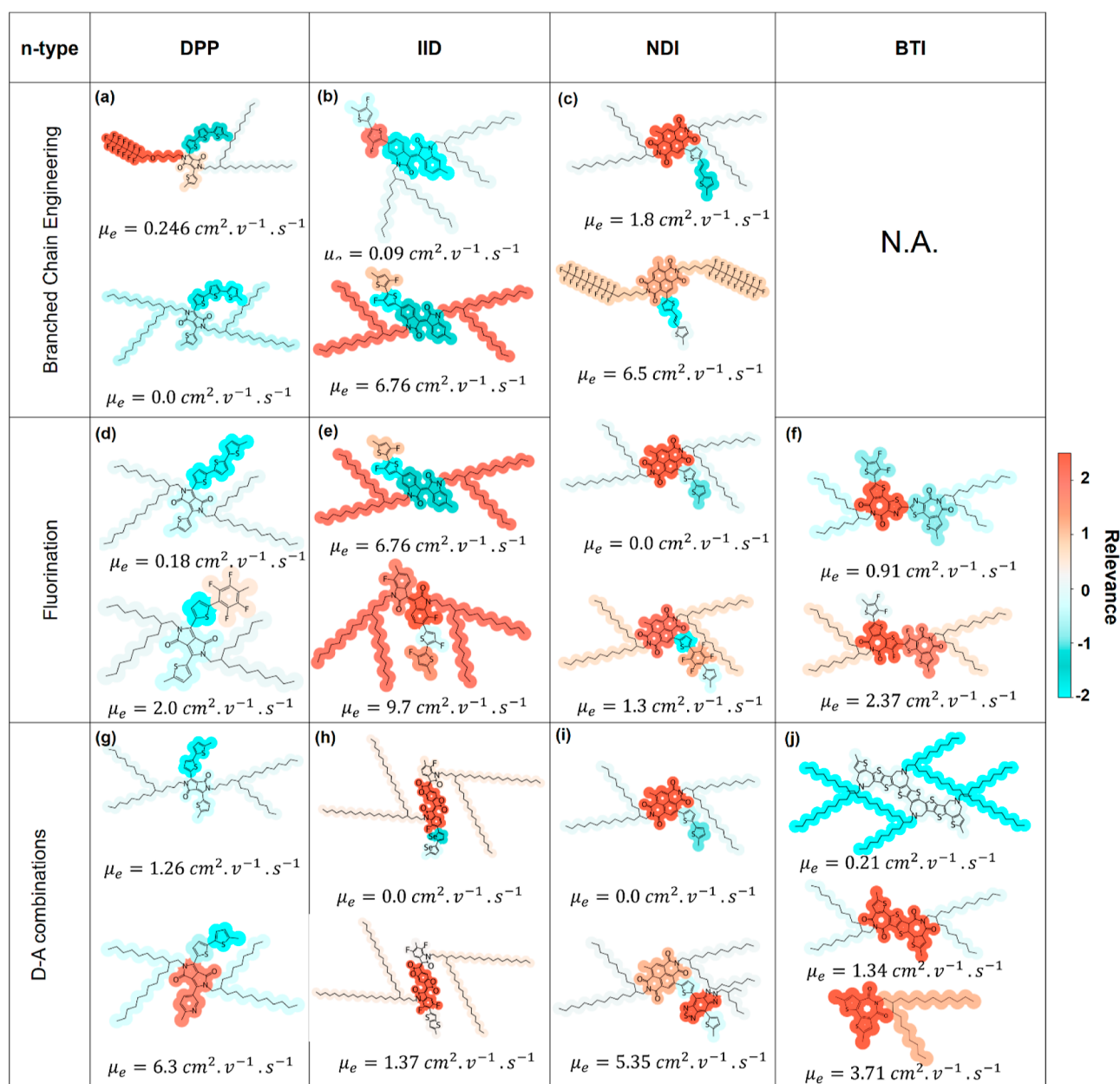


Figure 5. Present study offers a visual analysis of the correlation between mobility and molecular structure in N-type OSCs. The horizontal axis represents four representative OSC materials: diketopyrrolopyrrole (DPP), isoindigo (IID), cyclopentadithiophene (CDT), and bithiophene imide (BTI). The vertical axis corresponds to three typical effects that enhance mobility: branch chain engineering, fluorination, and donor–acceptor (D–A) combinations. (a) DPP-type OSCs of branch chain engineering effect, (b) IID-type OSCs of branch chain engineering effect, and (c) NDI-type OSCs of branch chain engineering and fluorination effect; (d) DPP-type OSCs of fluorination effect, (e) IID-type OSCs of fluorination effect, and (f) BTI-type OSCs of fluorination effect; and (g) DPP-type OSCs of donor–acceptor (D–A) combinations effect, (h) IID-type OSCs of donor–acceptor (D–A) combinations effect, (i) NDI-type OSCs of donor–acceptor (D–A) combinations effect, and (j) BTI-type OSCs of donor–acceptor (D–A) combinations effect.

chains compared to polymers within the same box (which are identical except for the branch chains). For example, in Figure 4b, the correlation of ethylene-glycol (OEG) chains is higher than that of ordinary alkyl chains, indicating a stronger relationship. The introduction of the OEG group can increase the glass transition temperature (T_g) of the polymer, improving its stability.^{83,84} Similarly, in Figure 5a, fluoroalkyl chains show a higher correlation compared to ordinary alkyl chains. The presence of alkyl fluoride chain leads to self-

assembly effects, enhancing the crystalline regularity of polymer molecules.⁶⁹

- In the column of fluorination (Figures 4d–f and 5c–f), the presence of fluorinated molecules in the polymer unit exhibits a stronger correlation, enhancing intermolecular interactions and promoting crystallinity within the polymer.⁸⁵ By incorporating fluorine atoms into the backbone structure of the polymer semiconductor, the planarity of the backbone chain is improved. This, in turn, leads to a reduction in the π – π stacking distance between the carrier chains.^{86,87}

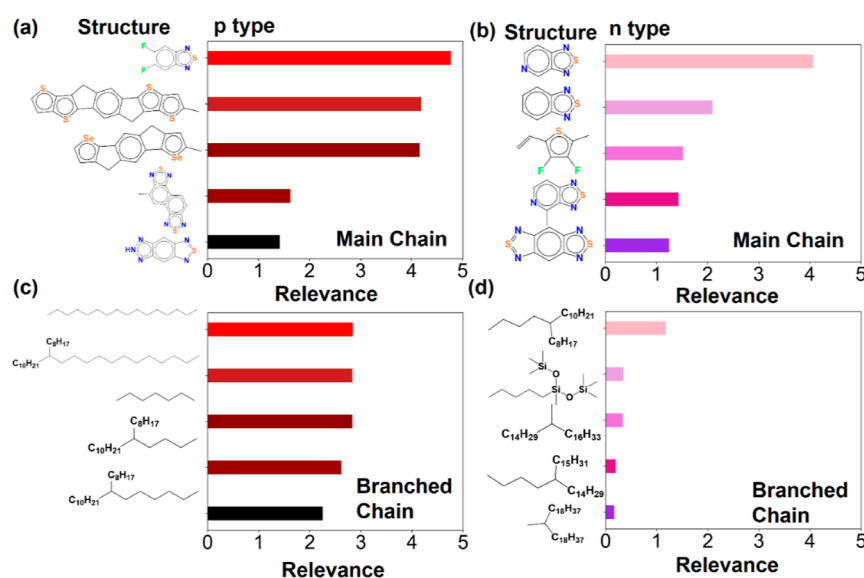


Figure 6. Statistical bar charts depicting highly correlated polymer-units. (a) The top five main chain polymer-units that exhibit the strongest mobility correlation in p-type OSCs. (b) The top five main chain polymer-units that exhibit the strongest mobility correlation in n-type OSCs. (c) The top five side chain polymer-units which exhibit the strongest mobility correlation in p-type OSCs. (d) The top five side chain polymer-units that exhibit the strongest mobility correlation in n-type OSCs.

3. In the case of D–A combinations, the varying combinations of donor–acceptor pairs influence the mobility outcomes by affecting the compatibility of molecular frontier orbitals and geometric factors such as planarity and steric hindrance. The OSC material efficiently injects or extracts charge from the electrode due to its suitable frontier molecular orbital energy level, while its high planarity results in a larger π – π conjugated area.⁴³ For instance, in Figure 4h, the polymer with thiophene groups shows higher mobility and correlation compared to the one with selenophene groups, indicating that the molecular frontier orbitals of thiophene align more favorably with the fluorinated IID group.⁴³ Similarly, in Figure 5j, the increased polymerization of BTI results in lower mobility, and the correlation of the BTI backbone decreases. This can be attributed to the larger BTI backbone causing increased steric hindrance, which adversely affects the morphology of OSC films.^{88,89} The examples in Figure 4h and Figure 5j can be argued to represent optimized geometric factors achieved through the selection of appropriate D–A combinations.

By considering the joint effects of branched-chain engineering, fluorination, and D–A combinations in the mobility analysis of OSC data, *PU-gn-exp* successfully identifies the structure–activity relationships of four typical polymer structures: DPP, IID, CDT, and BTI. This analysis convincingly demonstrates the analytical validity of the *PU-gn-exp* model.

2.3. Statistical Analysis Based on Polymer-Units. The OSC *Polymer-unit Graph* is composed of polymer-units as nodes. It aims to identify which polymer-unit shows a stronger correlation with the target properties. The statistical analysis focuses on the node correlation of *PU-gn-exp* model, with the mobility of electrons and holes as the labels. The correlation index obtained by the LRP algorithm is independent among the data and does not consider the influence of target values across different data. The enhancement of carrier mobility is a

crucial optimization objective in the development of OSC materials, and data with high mobility hold evident significance. Therefore, we incorporate the target value as a weight in the process of statistical correlation. The correlation Rel_i for each Polymer-unit i is calculated by

$$Rel_i = \frac{\sum_j \mu_j rel_{ji}}{num_i} \quad (1)$$

where i represents the species of PU, j represents each data point (i.e., each polymer), μ_j is the target attribute value of the data j , and num_i is the total number of PU_i in the data set.

Figure 6 illustrates the top polymer-units in the n-type and p-type OSC materials that exhibit the highest correlation to the labeled data. Based on this figure, we can categorize the highly correlated polymer-units into the following categories:

2.3.1. Benzothiazole. The benzothiazole (BTz) type structure is characterized by significant electron deficiency as well as remarkable rigidity and planarity. Moreover, BTz–DPP type D–A OSCs are particularly suitable for producing OSC films using shear coating technology. By utilizing this technique, the crystallinity of thin film OSC polymers can reach high values, up to 100 Dka, resulting in improved mobility.⁹⁰ There are three BTz polymer-units in both Figure 6a and Figure 6b.

2.3.2. Fluorinated Main Chain Polymer-Unit. The first polymer-unit in Figure 6a and the third polymer-unit in Figure 6b are both fluorinated main chain polymer-units. These fluorinated polymer-units offer distinct advantages. First, the presence of fluorine atoms enhances the van der Waals force between the ring structures, promoting a higher degree of coplanarity.¹ This enhances the charge transport properties within the polymer structure. Second, fluorine atoms possess strong electron-stabilization properties, promoting a reduction in the LUMO level. This reduction facilitates better alignment and docking of orbital energy levels for the donor–acceptor groups, resulting in improved charge transfer and overall performance of the OSC materials.⁴

2.3.3. Side Alkyl Chain. The length of the alkyl chain arm has a significant impact on the arrangement of the main chain of the polymer, thereby influencing the carrier transport performance. Figure 6c,d demonstrates the strong correlation of the side chain structure with a C₁₀ or a C₈ chain arm. To enhance the planarity of the polymer main chain, it is essential to position the branch points of the alkyl chain arms at least one or more carbon atoms away from the main chain. The polymer-units 2,4,5 in Figure 6c, as well as the polymer-units 1 and 4 in Figure 6d, all exhibit structures for which the arm branch points are positioned more than one carbon atom away from the main chain.⁶⁹

Comparing Figures 4 and 5, the positive mobility correlation of ethylene glycol (OEG) or alkyl fluoride chains is higher than that of ordinary alkyl chains. However, the statistical average in eq 1 in Figure 6 shows that the correlation of ordinary alkyl chains is higher than that of other alkyl chains. This seeming contradiction arises because Figures 4 and 5 compare individual cases, while Figure 6 is a statistical average weighted by mobility. In systems with high mobility such as BTz-cyclopentadithiophene (CDT), the side chain typically consists of ordinary alkyl chains. These results suggest that exploring the introduction of stronger atomic interaction branch chains, such as fluoroalkyl or oxyalkyl chains, in high mobility systems may be worthwhile.⁶⁹

Although there is no universally accepted standard for assessing the accuracy of the visual model, it is still crucial to compare the visual outputs of *PU-gn-exp* to those of the baseline model. The mobility structure–activity analysis diagram of the *PU-gn-exp* model and the *gn-exp* model for all data in the OSC data set has been included in the Supporting Information. Additionally, a bar chart depicting the node-correlation statistical analysis of the *gn-exp* model is available in the Supporting Information. It should be noted that the prediction accuracy of *PU-gn-exp* slightly decreased compared to the baseline model (n-type OSC: 88.2–89.6%; p-type OSC: 81.96–83.2%), see Figure S1. This does not mean that the interpretability of the structure–activity relationship based on polymerization units is lower than the molecular graph. The *gn-exp* model is a MPNN model, in which atomic-level nodes continuously converge during the MPNN process, leading to a final polyatomic structure centered on the original node atom. The fact that accurate predictions are obtained from the aggregation of atoms does not necessarily imply that an explanation based on individual atoms is more plausible. The visualization result generated by the LRP principle involves propagating the correlation responsible for producing the predicted outcome back to the input node. The explanatory comparison of OSC mobility in the *PU-gn-exp* and *gn-exp* models can be found in Figures S5 and 6.

The summary of the three types of highly correlated polymer-units in the transport mechanism of OSCs provides a comprehensive understanding of their performance. This analysis confirms the validity of the statistical analysis conducted on the highly correlated polymer-units. The *PU-gn-exp* method effectively identifies and analyzes polymer-units that exhibit a strong correlation with the label (carrier mobility). Polymer-unit Graph can not only enhance the interpretability of the MPNN model but also improve the operation efficiency of the machine learning model. The *PU-MPNN* model (See Figure S2a for details) is set up to demonstrate a significant reduction in runtime, equivalent to only 0.02 of the baseline model's (*mol-MPNN*) runtime

(Figure S2b). The decrease in training time can be attributed to the simplification of graph nodes by the OSC *Polymer-unit Graph*, thereby reducing the computational complexity involved in message passing between nodes. Additionally, the *PU-MPNN* model exhibits lower memory consumption compared to the baseline model. Specifically, when the Batch Size is set to 8, the *PU-MPNN* model can halve the memory consumption compared to the baseline. In the *PU-MPNN* model, the message passing operation primarily focuses on facilitating feature transfer between nodes while utilizing the OSC *Polymer-unit Graph* to streamline the node count and sparsify the input matrix. As a result, the inherent capability of the message passing algorithm in handling sparse matrices contributes to accelerated model performance.^{91,92}

3. CONCLUSIONS

The aim of this paper is to explore the utilization of the “*Polymer-unit Graph*” in the GNN model for analyzing the structures of organic polymer monomers and macromolecules. In terms of visualization and interpretability, our study finds that polymer materials such as OSCs exhibit enhanced interpretability in models. This is because their physical properties are primarily determined by the combination of polymer-units rather than a combination of atoms. Therefore, the *PU-gn-exp* model proves to be more effective in analyzing the structure–activity relationship of OSC polymers. In supervised learning, the *PU-MPNN* model achieves a notable reduction in training time, requiring only 2% of the training time for the baseline model while maintaining the same level of prediction accuracy. This demonstrates the successful integration of *Polymer-unit Graph* into a GNN model, resulting in improved analytical performance and operational efficiency.

Of course, the application of polymer-units in GNNs does have limitations. The OSC *Polymer-unit Graph* presented in this study is specifically designed for macromolecules and polymers with intricate structures composed of multiple polymer-units. OSC *Polymer-unit Graph* is not suitable for small molecule data due to the fact that many small molecules only consist of one or two polymer-units. Additionally, as the OSC *Polymer-unit Graph* is graph-structured data, it can only be used for GNN models. However, by leveraging the polymer unit concept in machine learning, this limitation can be overcome. In a previous study,⁴⁴ we utilized the arrangement of polymer-units as molecular fingerprints to train various classifiers, including random forests, support vector machines, and multilayer perceptron models. The classification accuracy achieved was over 80%.

The utilization of polymer-units in machine learning models opens up various avenues for exploration:

- (1) The *Polymer-unit Graph* employed in this research is specifically applied to the GNN model. However, apart from GNN, there are several other deep learning models suitable for organic macromolecules and polymers, such as the RNN model and the Transformer model. Further investigation into the application of polymer-units in these models could provide valuable insights.
- (2) In this study, polymer-units are identified based on the SMILES code rule, which is commonly used for investigating organic photoelectric materials like OFET, OLED, and OPV materials. The concept of polymer-unit can also be extended to the field of biology. Proteins, for instance, are formed through dehydration

and condensation reactions of amino acids. By identifying peptide bonds resulting from these reactions, proteins can be represented as combinations of amino acids. Exploring the use of polymer-units in the biological context could yield interesting findings.

- (3) The three-dimensional arrangement of organic materials, including aspects like chirality and coplanarity, significantly influences their properties. Therefore, incorporating the three-dimensional configuration into polymer-units for expression is a valuable endeavor. Researching ways to integrate and represent the three-dimensional aspects of organic materials within polymer-units could provide deeper insights into their structure–property relationships.

Overall, the implementation of polymer-units in deep learning models offers a broad scope for further research and exploration, encompassing different model architectures, expanding into biological contexts, and addressing the incorporation of three-dimensional configurations for a more comprehensive understanding of organic materials.

4. METHODS

4.1. OSC Data Set. The data set utilized in this study consisted of 697 OSCs data, which were obtained from relevant literature.^{1,43,62–64,69–81} The parameters for each datum point include the SMILES representing the monomer structure, electron mobility, and hole mobility of the material device, and HOMO and LUMO values obtained from DFT calculations. The distribution of this data set is presented in the following table:

4.2. Polymer-Unit-Recognition Script. For each data collected in the OSC database, the *PURS* program is used to identify the polymer-units contained within it and the connection relationship between polymer-units.⁴⁴ The methodology of polymer-unit recognition, as outlined in “Polymer-Unit Fingerprint (*PUFp*): An Accessible Expression of Polymer OSCs for Machine Learning”, is comprehensively elucidated within this document. The SMILES codes for all polymer-units are standardized using the *RDKit*⁵⁹ program. Due to the correspondence between standard SMILES code and molecular structure, identifying the uniqueness of a primitive is straightforward. Polymer-units in the OSC database are categorized into a database after eliminating duplicates.

4.3. Graph-Network-Explainability. The graph-network-explainability model was proposed by Federico and Hossein in 2019, wherein they extensively investigated the explainability of GNN models.⁵⁸ This model is composed of two components: the GNN component and the model analysis module. The GNN model is built upon the GN model proposed by Battaglia et al. in 2018.⁹³ The GN model is a GNN model that is built upon the MPNN framework. It accepts input graphs containing a node ($E = (e_k)$), edge ($V = \{V_i\}$), and graph-level information (u). The graph should be updated with the three update functions ϕ and the three aggregate functions ρ in each layer of the model:

$$e'_k = \phi^e(e_k, v_{rk}, v_{sk}, u) \quad \bar{e}'_i = \rho^{e \rightarrow v}(E'_i) \quad (2)$$

$$v'_i = \phi^v(\bar{e}'_i, v_i, u) \quad \bar{v}' = \rho^{v \rightarrow u}(E') \quad (3)$$

$$u' = \phi^u(\bar{e}', \bar{v}', u) \quad \bar{v}' = \rho^{v \rightarrow u}(v') \quad (4)$$

The symbols rk and sk denote the sender and receiver, respectively, in relation to edge k . E'_i represents all the edges of a connecting node i , E' represents the set of all the edges updated by the ϕ^e function, and V' represents the set of all the nodes updated by the ϕ^v function. The processing layers maintain the graph's structure while updating only its nodes, not its edges.

The mapping $f:(E, V, u) \rightarrow y$ can generate an output for the entire graph or for a specific node or edge. In the context of graph network explainability, all ϕ employed refers to a combination of linear transformations and *ReLU* activation functions, while all ρ denote sum/mean/max pooling operations.

The Analysis module offers three mechanisms for selection: SA, GBP, and LRP. The present work adopts the LRP principle, which demonstrates superior performance in test scenarios. The LRP principle aims to transform the output signal of the GNN into a linear combination of input signals. The LRP principle employed by graph-network explainability can be interpreted as a repeated Taylor decomposition, assuming that the correlation sum of each layer in the neural network remains constant. Consequently, the correlation sum of the output layer can be decomposed into the input correlation sum:

$$\sum R^{(x)} = \dots = \sum R^{(l)} = \dots = f(x) \quad (5)$$

The LRP principle is mainly concerned with which characteristics of the input contribute the most to the output, and it can deal with positive and negative correlations, respectively, so that the correlation between the input and the output provides more in-depth analysis. The LRP relies on back-propagation of a neural network, which involves accumulating correlations from the output to the input. This is accomplished through the utilization of the *pytorch* *autograd* module.

4.4. Density Functional Theory Calculations. The frontier molecular orbit (HOMO and LUMO) of all OSC data in this paper are calculated by DFT. The DFT is calculated using the Gaussian 09 package.⁹⁴ For geometric optimization and frequency calculations, the density functional method selected was B3LYP, and the selected basis group was 6-31G*. Because some OSC materials have heavy elements, the relativistic effect correction term *gd3bj* was used.

4.5. MACCS Fingerprint.⁶⁰ MACCS fingerprint is a System of Molecular descriptors known as the “Molecular Access System”. MACCS is a fixed length of binary bits to describe the structural characteristics of molecules, where each bit represents a molecular structure. That is, MACCS can represent this molecular structure as a set of fixed length binary codes. This article uses the MACCS conversion interface of the *RDKit* platform, which is 166 bits.

4.6. GenWL graph Embedding Method.⁶¹ The Graph embedding method is a technique for converting the graph representation to the vector representation.⁵² Weisfeiler–Lehman (WL) is a graph embedding algorithm that checks the isomorphism of graphs by iterating node labels. *GenWL* is an extended version based on the classic WL algorithm, and its improvement to the WL algorithm is to add a measure of similarity between two graphs—the tree editing distance.⁶¹

4.7. Code Available. We provide all of the code for OSC *Polymer-unit Graph* on our open-source GitHub repository, which can be found at <https://github.com/xinyue123-q/Python-based-polymer-unit-recognition-script-PURS-2.0>.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.3c01385>.

Comparison of accuracy between *gn-exp* and *PU-gn-exp*; transform *gn-exp* into a classification model; hyperparameter adjustment of *gn-exp* and *PU-gn-exp*; Polymer-unit Graph enhancing the operational efficiency of the MPNN; details for *PU-MPNN* and *mol_MPNN*; and MSE of *mol-MPNN* and *PU-MPNN* on OSC data sets (PDF)

Polymer_OSC_data sets (PDF)

Structure of polymer-units (PDF)

Baseline model and PU-gn-exp visualization results for all OSC data (PDF)

Polymer_unit_SMILES and graph embeddings feature (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Caichao Ye – Department of Materials Science and Engineering & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, PR China; Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen 518055, PR China; orcid.org/0000-0002-9197-8922; Email: yecc@sustech.edu.cn

Wenqing Zhang – Department of Materials Science and Engineering & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, PR China; orcid.org/0000-0002-2551-6375; Email: zhangwq@sustech.edu.cn

Authors

Xinyue Zhang – Department of Materials Science and Engineering & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, PR China

Ye Sheng – Department of Materials Science and Engineering & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, PR China

Xiumin Liu – Department of Materials Science and Engineering & Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen 518055, PR China; Key Laboratory of Soft Chemistry and Functional Materials of MOE, School of Chemistry and Chemical Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China

Jiong Yang – Materials Genome Institute, Shanghai University, Shanghai 200444, PR China; orcid.org/0000-0002-5862-5981

William A. Goddard III – Materials and Process Simulation Center (MSC), California Institute of Technology, Pasadena, California 91125, United States; orcid.org/0000-0003-0097-5716

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.3c01385>

Author Contributions

C.Y., W.Z., and W.A.G. formulated this project. X.Z. performed data collection, program coding, and ML analysis. Y.S., X.L., and J.Y. provided helpful discussion for program coding. X.Z. and C.Y. cowrote the manuscript. W.Z. and W.A.G. revised the manuscript. W.Z., W.A.G., and C.Y. secured the funding.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support was provided by the National Key R&D Program of China (2022YFA1203400), Natural Science Foundation of China (92163212), Guangdong Basic and Applied Basic Research Foundation (2022A1515110628), and Guangdong Provincial Key Laboratory of Computational Science and Material Design (2019B030301001). Zhang acknowledges support from the Guangdong Innovation Research Team Project (2017ZT07C062), and Goddard was supported by the US NSF (CBET 2311117). Computing resources were supported by the Center for Computational Science and Engineering at Southern University of Science and Technology.

■ REFERENCES

- (1) Nielsen, C. B.; Turbiez, M.; McCulloch, I. Recent Advances in the Development of Semiconducting Dpp-Containing Polymers for Transistor Applications. *Adv. Mater.* **2013**, *25*, 1859–1880.
- (2) Yu, G.; Gao, J.; Hummelen, J. C.; Wudl, F.; Heeger, A. J. Polymer Photovoltaic Cells—Enhanced Efficiencies Via a Network of Internal Donor-Acceptor Heterojunctions. *Science* **1995**, *270*, 1789–1791.
- (3) Li, G.; Shrotriya, V.; Huang, J. S.; Yao, Y.; Moriarty, T.; Emery, K.; Yang, Y. High-Efficiency Solution Processable Polymer Photovoltaic Cells by Self-Organization of Polymer Blends. *Nat. Mater.* **2005**, *4*, 864–868.
- (4) Shi, L.; Guo, Y.; Hu, W.; Liu, Y. Design and Effective Synthesis Methods for High-Performance Polymer Semiconductors in Organic Field-Effect Transistors. *Mater. Chem. Front.* **2017**, *1*, 2423–2456.
- (5) Nketia-Yawson, B.; Noh, Y.-Y. Recent Progress on High-Capacitance Polymer Gate Dielectrics for Flexible Low-Voltage Transistors. *Adv. Funct. Mater.* **2018**, *28*, 1802201.
- (6) Lee, W. H.; Park, Y. D. Organic Semiconductor/Insulator Polymer Blends for High-Performance Organic Transistors. *Polymers* **2014**, *6*, 1057–1073.
- (7) Lee, J. Physical Modeling of Charge Transport in Conjugated Polymer Field-Effect Transistors. *J. Phys. D: Appl. Phys.* **2021**, *54*, 143002.
- (8) Salehi, A.; Fu, X. Y.; Shin, D. H.; So, F. Recent Advances in OLED Optical Design. *Adv. Funct. Mater.* **2019**, *29*, 1808803.
- (9) Hong, G.; Gan, X. M.; Leonhardt, C.; Zhang, Z.; Seibert, J.; Busch, J. M.; Brase, S. A Brief History of OLEDs-Emitter Development and Industry Milestones. *Adv. Mater.* **2021**, *33*, 2005630.
- (10) Akamatsu, H.; Inokuchi, H.; Matsunaga, Y. Organic Semiconductors with High Conductivity. I. Complexes between Polycyclic Aromatic Hydrocarbons and Halogens. *Bull. Chem. Soc. Jpn.* **1956**, *29*, 213–218.
- (11) Gao, C.; Min, X.; Fang, M.; Tao, T.; Zheng, X.; Liu, Y.; Wu, X.; Huang, Z. Innovative Materials Science Via Machine Learning. *Adv. Funct. Mater.* **2021**, *32*, 2108044.
- (12) Cai, J.; Chu, X.; Xu, K.; Li, H.; Wei, J. Machine Learning-Driven New Material Discovery. *Nanoscale Adv.* **2020**, *2*, 3115–3130.
- (13) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.

- (14) Omeel, S. S.; Louis, S. Y.; Fu, N. H.; Wei, L.; Dey, S.; Dong, R.; Li, Q.; Hu, J. Scalable Deeper Graph Neural Networks for High-Performance Materials Property Prediction. *Patterns* **2022**, *3*, 100491.
- (15) Ma, H. H.; Bian, Y. T.; Rong, Y.; Huang, W. B.; Xu, T. Y.; Xie, W. Y.; Ye, G. Y.; Huang, J. Z. Cross-Dependent Graph Neural Networks for Molecular Property Prediction. *Bioinformatics* **2022**, *38*, 2003–2009.
- (16) Jiang, S. L.; Balaprakash, P. In Graph Neural Network Architecture Search for Molecular Property Prediction. *8th IEEE International Conference on Big Data (Big Data)*; IEEE, 2020; pp 1346–1353. Dec 10–13; Electr Network.
- (17) Hao, Z. K.; Lu, C. Q.; Huang, Z. Y.; Wang, H.; Hu, Z. Y.; Liu, Q.; Chen, E. H.; Lee, C.; Assoc Comp, M. In Asgn: An Active Semi-Supervised Graph Neural Network for Molecular Property Prediction. *26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*; ACM, 2020; pp 731–739. Aug 23–27; Electr Network.
- (18) Gao, P.; Zhang, J.; Qiu, H. B.; Zhao, S. F. A General Qspr Protocol for the Prediction of Atomic/Inter-Atomic Properties: A Fragment Based Graph Convolutional Neural Network (F-Gcn). *Phys. Chem. Chem. Phys.* **2021**, *23*, 13242–13249.
- (19) Deng, D.; Lei, Z.; Hong, X.; Zhang, R.; Zhou, F. Describe Molecules by a Heterogeneous Graph Neural Network with Transformer-Like Attention for Supervised Property Predictions. *ACS Omega* **2022**, *7*, 3713–3721.
- (20) Dai, M.; Demirel, M. F.; Liang, Y.; Hu, J.-M. Graph Neural Networks for an Accurate and Interpretable Prediction of the Properties of Polycrystalline Materials. *npj Comput. Mater.* **2021**, *7*, 103.
- (21) Cai, H. X.; Zhang, H. M.; Zhao, D. C.; Wu, J. X.; Wang, L. Fp-Gnn: A Versatile Deep Learning Architecture for Enhanced Molecular Property Prediction. *Briefings Bioinf.* **2022**, *23*, 36124766.
- (22) Li, Y.; Liu, L.; Wang, G.; Du, Y.; Chen, P. Eggn: Constructing Explainable Graph Neural Networks Via Knowledge Distillation. *Knowledge-Based Systems* **2022**, *241*, 108345.
- (23) Baldassarre, F. Explainability Techniques for Graph Convolutional Networks. *Arxiv* **2019**.
- (24) Gebauer, N. W. A.; Gastegger, M.; Hessmann, S. S. P.; Müller, K. R.; Schütt, K. T. Inverse Design of 3d Molecular Structures with Conditional Generative Neural Networks. *Nat. Commun.* **2022**, *13*, 973.
- (25) Kim, K.; et al. Deep-Learning-Based Inverse Design Model for Intelligent Discovery of Organic Molecules. *npj Comput. Mater.* **2018**, *4*, 67.
- (26) Zhumagambetov, R.; Molnar, F.; Peshkov, V. A.; Fazli, S. Transmol: Repurposing a Language Model for Molecular Generation. *RSC Adv.* **2021**, *11*, 25921–25932.
- (27) Zhang, P.; Hua, X.; Wang, X.; Rui, T.; Zhang, H.; Shao, F.; Wang, D. Vsa-Cgan: An Intelligent Generation Model for Deep Learning Sample Database Construction. *IEEE Access* **2020**, *8*, 137986–138003.
- (28) Zhang, J.; Chen, H. De Novo Molecule Design Using Molecular Generative Models Constrained by Ligand-Protein Interactions. *J. Chem. Inf. Model.* **2022**, *62*, 3291–3306.
- (29) Tagade, P. M.; Adiga, S. P.; Pandian, S.; Park, M. S.; Hariharan, K. S.; Kolake, S. M. Attribute Driven Inverse Materials Design Using Deep Learning Bayesian Framework. *npj Comput. Mater.* **2019**, *5*, 127.
- (30) Nigam, A.; Pollice, R.; Krenn, M.; Gomes, G. D. P.; Aspuru-Guzik, A. Beyond Generative Models: Superfast Traversal, Optimization, Novelty, Exploration and Discovery (Stoned) Algorithm for Molecules Using Selfies. *Chem. Sci.* **2021**, *12*, 7079–7090.
- (31) Allesoe, R. L.; Lundgaard, A. T.; Hernandez Medina, R.; et al. Discovery of Drug-Omics Associations in Type 2 Diabetes with Generative Deep-Learning Models. *Nat. Biotechnol.* **2023**, *41*, 399–408.
- (32) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.
- (33) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv* **2017**, arXiv:1704.01212 [cs.LG].
- (34) Zhang, Z.; Friedrich, K. Artificial Neural Networks Applied to Polymer Composites: A Review. *Compos. Sci. Technol.* **2003**, *63*, 2029–2044.
- (35) Queen, O.; McCarver, G. A.; Thatigotla, S.; Abolins, B. P.; Brown, C. L.; Maroulas, V.; Vogiatzis, K. D. Polymer Graph Neural Networks for Multitask Property Learning. *npj Comput. Mater.* **2023**, *9*, 90.
- (36) Lu, X. X.; Giovanis, D. G.; Yvonnet, J.; Papadopoulos, V.; Detrez, F.; Bai, J. B. A Data-Driven Computational Homogenization Method Based on Neural Networks for the Nonlinear Anisotropic Electrical Response of Graphene/Polymer Nanocomposites. *Comput. Mech.* **2019**, *64*, 307–321.
- (37) Jiang, Z. Y.; Zhang, Z.; Friedrich, K. Prediction on Wear Properties of Polymer Composites with Artificial Neural Networks. *Compos. Sci. Technol.* **2007**, *67*, 168–176.
- (38) Gurnani, R.; Kuenneth, C.; Toland, A.; Ramprasad, R. Polymer Informatics at Scale with Multitask Graph Neural Networks. *Chem. Mater.* **2023**, *35*, 1560–1567.
- (39) Aldeghi, M.; Coley, C. W. A Graph Representation of Molecular Ensembles for Polymer Property Prediction. *Chem. Sci.* **2022**, *13*, 10486–10498.
- (40) Gurnani, R.; Kuenneth, C.; Toland, A.; Ramprasad, R. Polymer Informatics at Scale with Multitask Graph Neural Networks. *Chem. Mater.* **2023**, *35*, 1560–1567.
- (41) Sumpter, B. G.; Noid, D. W. Neural Networks and Graph Theory as Computational Tools for Predicting Polymer Properties. *Macromol. Theory Simul.* **1994**, *3*, 363–378.
- (42) Mullen, K.; Pisula, W. Donor-Acceptor Polymers. *J. Am. Chem. Soc.* **2015**, *137*, 9503–9505.
- (43) Kim, M.; Ryu, S. U.; Park, S. A.; Choi, K.; Kim, T.; Chung, D.; Park, T. Donor-Acceptor-Conjugated Polymer for High-Performance Organic Field-Effect Transistors: A Progress Report. *Adv. Funct. Mater.* **2019**, *30*, 143002.
- (44) Zhang, X.; Wei, G.; Sheng, Y.; Bai, W.; Yang, J.; Zhang, W.; Ye, C. Polymer-Unit Fingerprint (Pufp): An Accessible Expression of Polymer Organic Semiconductors for Machine Learning. *ACS Appl. Mater. Interfaces* **2023**, *15*, 21537–21548.
- (45) Weininger, D. Smiles, a Chemical Language and Information-System.1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.
- (46) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
- (47) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; The MIT Press, 2012; pp 492–493.
- (48) Patrick, E. A.; Fischer, F. P. A Generalized K-Nearest Neighbor Rule. *Inf. Control.* **1970**, *16*, 128–152.
- (49) Kowalski, B. R.; Bender, C. F. K-Nearest Neighbor Classification Rule (Pattern-Recognition) Applied to Nuclear Magnetic-Resonance Spectral Interpretation. *Anal. Chem.* **1972**, *44*, 1405–1411.
- (50) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444.
- (51) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90.
- (52) Qi, Z.; Wang, J.; Yue, K.; Qiao, S.; Li, J. Methods and Applications of Graph Embedding: A Survey. *Acta Electron. Sin.* **2020**, *48*, 808–818.
- (53) Robbins, J. S.; Schmid, K. M.; Phillips, S. T. Effects of Electronics, Aromaticity, and Solvent Polarity on the Rate of Azaquinone-Methide-Mediated Depolymerization of Aromatic Carbamate Oligomers. *J. Org. Chem.* **2013**, *78*, 3159–3169.
- (54) Maity, K.; Kundu, T.; Banerjee, R.; Biradha, K. One-Dimensional Water Cages with Repeat Units of (H₂O)₂₄ Resembling Pagodane Trapped in a 3D Coordination Polymer: Proton Conduction and Tunable Luminescence Emission by Adsorption of Anionic Dyes. *CrystEngComm* **2015**, *17*, 4439–4443.

- (55) Honzl, J.; Ulbert, K.; Hadek, V.; Tlustáková, M.; Metalova, M. Organic Semiconductors-Donor-Acceptor Complexes of Conjugated Bases with a Repeating Structural Unit. *J. Polym. Sci., Part C: Polym. Symp.* **1967**, *16*, 4465–4481.
- (56) Honzl, J.; Ulbert, K.; Hadek, V.; Tlustáková, M. Organic semiconductors: donor–acceptor complexes of conjugated based with a repeating structural unit. *Chem. Commun.* **1965**, *0*, 440–441.
- (57) Hollomon, M. G. *Synthesis and Characterization of Oligomeric/Polymeric Assemblies as Organic Based Molecular Magnetic Materials*; North Carolina State University ProQuest Dissertations Publishing, 1998.
- (58) Federico Baldassarre, H. A. Explainability Techniques for Graph Convolutional Networks. *arXiv* **2019**, arXiv:1905.13686 [cs.LG].
- (59) Landrum, G. Rdkit: Open-Source Cheminformatics from Machine Learning to Chemical Registration. *Abstracts of Papers of the American Chemical Society*; American Chemical Society, 2019; Vol. 258, p 7727.
- (60) Polton, D. J. Installation and Operational Experiences with Maccs (Molecular Access System). *Online Review* **1982**, *6*, 235–242.
- (61) Schulz, T. H. A Generalized Weisfeiler-Lehman Graph Kernel. *arXiv* **2021**, arXiv:2101.08104v1 [cs.LG].
- (62) Li, Y.; Sonar, P.; Murphy, L.; Hong, W. High Mobility Diketopyrrolopyrrole (Dpp)-Based Organic Semiconductor Materials for Organic Thin Film Transistors and Photovoltaics. *Energy Environ. Sci.* **2013**, *6*, 1684–1710.
- (63) Hu, W.; Feng, L.; Gu, P.; Dong, H.; Yao, Y. High-Mobility Polymeric Semiconductors. *Chin. Sci. Bull.* **2015**, *60*, 2169–2187.
- (64) Sui, Y.; Deng, Y.; Du, T.; Shi, Y.; Geng, Y. Design Strategies of N-Type Conjugated Polymers for Organic Thin-Film Transistors. *Mater. Chem. Front.* **2019**, *3*, 1932–1951.
- (65) Paszke, A.; Gross, S.; Massa, F. Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems, NeurIPS Proceedings* **2019**, *32*, 8026–8037.
- (66) Mladenovic, M.; Vukmirovic, N. Charge Carrier Localization and Transport in Organic Semiconductors: Insights from Atomistic Multiscale Simulations. *Adv. Funct. Mater.* **2015**, *25*, 1915–1932.
- (67) Manurung, R.; Troisi, A. Screening Semiconducting Polymers to Discover Design Principles for Tuning Charge Carrier Mobility. *J. Mater. Chem. C* **2022**, *10*, 14319–14333.
- (68) Piyakulawat, P.; Keawprajak, A.; Jiramitmongkon, K.; Hanusch, M.; Wlosnewski, J.; Asawapirom, U. Effect of Thiophene Donor Units on the Optical and Photovoltaic Behavior of Fluorene-Based Copolymers. *Sol. Energy Mater. Sol. Cells* **2011**, *95*, 2167–2172.
- (69) Yang, Y.; Liu, Z.; Zhang, G.; Zhang, X.; Zhang, D. The Effects of Side Chains on the Charge Mobilities and Functionalities of Semiconducting Conjugated Polymers Beyond Solubilities. *Adv. Mater.* **2019**, *31*, No. e1903104.
- (70) Wadsworth, A.; Chen, H.; Thorley, K. J.; Cendra, C.; Nikolka, M.; Bristow, H.; Moser, M.; Salleo, A.; Anthopoulos, T. D.; Sirringhaus, H.; et al. Modification of Indacenodithiophene-Based Polymers and Its Impact on Charge Carrier Mobility in Organic Thin-Film Transistors. *J. Am. Chem. Soc.* **2020**, *142*, 652–664.
- (71) Paterson, A. F.; Singh, S.; Fallon, K. J.; Hodsdon, T.; Han, Y.; Schroeder, B. C.; Bronstein, H.; Heeney, M.; McCulloch, I.; Anthopoulos, T. D. Recent Progress in High-Mobility Organic Transistors: A Reality Check. *Adv. Mater.* **2018**, *30*, No. e1801079.
- (72) Ortiz, R. P.; Yan, H.; Facchetti, A.; Marks, T. Azine- and Azole-Functionalized Oligo and Polythiophene Semiconductors for Organic Thin-Film Transistors. *Materials* **2010**, *3*, 1533–1558.
- (73) Nielsen, C. B.; McCulloch, I. Recent Advances in Transistor Performance of Polythiophenes. *Prog. Polym. Sci.* **2013**, *38*, 2053–2069.
- (74) Li, M.; Wang, J.; Xu, W.; Li, L.; Pisula, W.; Janssen, R. A. J.; Liu, M. Noncovalent Semiconducting Polymer Monolayers for High-Performance Field-Effect Transistors. *Prog. Polym. Sci.* **2021**, *117*, 101394.
- (75) Li, M.; An, C.; Pisula, W.; Mullen, K. Cyclopentadithiophene-Benzothiadiazole Donor-Acceptor Polymers as Prototypical Semiconductors for High-Performance Field-Effect Transistors. *Acc. Chem. Res.* **2018**, *51*, 1196–1205.
- (76) Lei, T.; Wang, J. Y.; Pei, J. Design, Synthesis, and Structure-Property Relationships of Isoindigo-Based Conjugated Polymers. *Acc. Chem. Res.* **2014**, *47*, 1117–1126.
- (77) Kim, Y.; Lim, E. Development of Polymer Acceptors for Organic Photovoltaic Cells. *Polymers* **2014**, *6*, 382–407.
- (78) He, K.; Kumar, P.; Yuan, Y.; Li, Y. Wide Bandgap Polymer Donors for High Efficiency Non-Fullerene Acceptor Based Organic Solar Cells. *Mater. Adv.* **2021**, *2*, 115–145.
- (79) Facchetti, A. π -Conjugated Polymers for Organic Electronics and Photovoltaic Cell Applications. *Chem. Mater.* **2011**, *23*, 733–758.
- (80) Chen, H. Recent Advances in High-Mobility Polymeric Semiconductor Materials. *Chin. J. Org. Chem.* **2016**, *36*, 460–479.
- (81) Brebels, J.; Manca, J. V.; Lutsen, L.; Vanderzande, D.; Maes, W. High Dielectric Constant Conjugated Materials for Organic Photovoltaics. *J. Mater. Chem. A* **2017**, *5*, 24037–24050.
- (82) Kim, M.; Ryu, S. U.; Park, S. A.; Choi, K.; Kim, T.; Chung, D.; Park, T. Donor-Acceptor-Conjugated Polymer for High-Performance Organic Field-Effect Transistors: A Progress Report. *Adv. Funct. Mater.* **2020**, *30*, 143002.
- (83) Stanciu, M. C.; Nichifor, M. New Degradable Polyesters from Deoxycholic Acid and Oligo(Ethylene Glycol)S. *Polym. Int.* **2013**, *62*, 1236–1242.
- (84) Shimomoto, H.; Shimizu, K.; Takeda, C.; Kikuchi, M.; Kudo, T.; Mukai, H.; Itoh, T.; Ihara, E.; Hoshikawa, N.; Koiwai, A.; et al. Synthesis of Polymers with Densely-Grafted Oligo(Ethylene Glycol)S by Pd-Initiated Polymerization of Oxyethylene-Containing Diazoacetates. *Polym. Chem.* **2015**, *6*, 8124–8131.
- (85) Lu, Y.; Ding, Y.; Wang, J.; Pei, J. Research Progress in Isoindigo-Based Polymer Field-Effect Transistor Materials. *Chin. J. Org. Chem.* **2016**, *36*, 2272–2283.
- (86) Tao, X.; Liu, Y.; Du, L.; Yan, Y.; Wu, Z.; Zhao, Y.; Guo, Y.; Chen, H.; Liu, Y. Regulation of the Backbone Structure and Optoelectrical Properties of Bis-Pyridal 2,1,3 Thiadiazole-Based Ambipolar Semiconducting Polymers Via a Fluorination Strategy. *J. Mater. Chem. C* **2021**, *9*, 15083–15094.
- (87) Cao, F.-Y.; Huang, P.-K.; Su, Y.-C.; Huang, W.-C.; Chang, S.-L.; Hung, K.-E.; Cheng, Y.-J. Forced Coplanarity of Dithienofluorene-Based Non-Fullerene Acceptors to Achieve High-Efficiency Organic Solar Cells. *J. Mater. Chem. A* **2019**, *7*, 17947–17953.
- (88) Wang, Y.; Guo, H.; Harbuzaru, A.; Uddin, M. A.; Arrechea-Marcos, I.; Ling, S.; Yu, J.; Tang, Y.; Sun, H.; Lopez Navarrete, J. T.; et al. (Semi)Ladder-Type Bithiophene Imide-Based All-Acceptor Semiconductors: Synthesis, Structure-Property Correlations, and Unipolar N-Type Transistor Performance. *J. Am. Chem. Soc.* **2018**, *140*, 6095–6108.
- (89) Wei, G.; Zhang, X.; Li, J.; Bai, W.; Ye, C.; Zhang, W. Intrinsic Charge Transport for Dtzti-Based All-Acceptor Homopolymer N-Type Organic Semiconductors: Roles of Conjugation Length and Orbital Delocalization. *J. Phys. Chem. C* **2023**, *127*, 4273–4282.
- (90) Khim, D.; Luzio, A.; Bonacchini, G. E.; Pace, G.; Lee, M.; Noh, Y.; Caironi, M. Uniaxial Alignment of Conjugated Polymer Films for High Performance Organic Fieldeffect Transistors. *Adv. Mater.* **2018**, *30*, 1705463.
- (91) Wei, Y.; Zhao, M.-M.; Zhao, M.; Lei, M.; Yu, Q. An Amp-Based Network with Deep Residual Learning for Mmwave BeamSpace Channel Estimation. *IEEE Wirel. Commun.* **2019**, *8*, 1289–1292.
- (92) Huang, S.; Qiu, D.; Tran, T. D. Approximate Message Passing with Parameter Estimation for Heavily Quantized Measurements. *IEEE Trans. Signal Process.* **2022**, *70*, 2062–2077.
- (93) Battaglia, P. W.; et al. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**, arXiv:1806.01261v3 [cs.LG].
- (94) Frisch, M. J.; T, G. W.; Schlegel, H. B. *Gaussian 09*, revision D.01, 2013.