Conditional Independence Testing of the HOMO-LUMO Decoupling Asymmetry Across Local Bonding Environments

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The highest occupied molecular orbital (HOMO) energy often resists integration into the global spatial-thermodynamic organization of small molecules, presenting an unexplained anomaly in frontier-orbital physics. To investigate whether this decoupling is contingent upon specific local bonding environments, we analyzed the QM9 dataset to contrast the conditional dependency structure of the HOMO against the lowest unoccupied molecular orbital (LUMO). We extracted non-linear gradient-boosted atom-count residuals for molecular properties and computed partial distance correlations conditioned on structural variables. Additionally, we quantified topological rewiring using LOESS regression across carbon saturation gradients and evaluated local clustering coefficients on fixed-density graphs, validating our framework with 5-fold scaffold-based hold-outs. Our results demonstrate a fundamental asymmetry in frontier-orbital behavior. While the LUMO energy maintains a consistent relationship with the spatial and thermodynamic state of the molecule across all structural classes, the HOMO energy exhibits a unique decoupling in saturated aliphatic systems. However, as carbon saturation decreases, the HOMO forcefully integrates into the broader property network, shifting from complete isolation to central participation in the presence of aromaticity and unsaturation. These findings indicate that the physical determinants of occupied frontier states are highly sensitive to topological complexity, establishing the HOMO as a primary site of structural-electronic rewiring in organic molecules.

1 Introduction

The organization of molecular properties into a coherent spatial-thermodynamic manifold is a foundational concept in chemical physics. Properties such as zero-point vibrational energy, heat capacity, and electronic spatial extent typically exhibit strong interdependencies that scale predictably with molecular size and composition. However, the frontier orbitals—specifically the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO)—often present a significant anomaly. The HOMO energy, in particular, uniquely resists integration into this global organization, exhibiting a decoupling from the broader physical manifold that remains poorly understood.

It remains an open question whether this isolation is an intrinsic feature of occupied frontier states or if it is contingent upon specific local bonding environments. We hypothesize that the physical determinants of the HOMO energy are highly sensitive to topological complexity, such as carbon saturation, aromaticity, and ring strain. To investigate this, we analyze the QM9 dataset to contrast the conditional dependency structure of the HOMO against the LUMO, seeking to identify the precise architectural thresholds that drive structural-electronic rewiring.

To isolate these effects from baseline stoichiometric scaling, we extract non-linear gradient-boosted atom-count residuals for molecular properties. Initial evaluations of global distance correlation ($dCor$) and partial distance correlation ($pdCor$)—conditioned on the principal components of the property set to account for latent structural influences—reveal a stark baseline isolation of the frontier orbitals. As demonstrated in Table 1, the HOMO residuals exhibit a consistently higher reduction in correlation when conditioned compared to the LUMO residuals. For instance, the relationship between heat capacity (‘cv‘) and HOMO energy sees an 82.9% reduction in correlation, indicating that apparent global couplings are largely artifacts of broader structural features. In contrast, the LUMO retains a greater degree of intrinsic coupling with the spatial-thermodynamic core.

Relationship	*dCor*	*pdCor*	Reduction
‘zpve‘ vs ‘homo‘	0.2456	0.0566	76.9%
‘zpve‘ vs ‘lumo‘	0.2961	0.0827	72.1%
‘cv‘ vs ‘homo‘	0.2094	0.0358	82.9%
‘cv‘ vs ‘lumo‘	0.2812	0.0673	76.1%

Motivated by this global decoupling, we extend our framework to quantify topological rewiring along gradients of structural complexity. We employ LOESS regression across carbon saturation gradients and evaluate local clustering coefficients on fixed-density graphs to map the integration of frontier orbitals across saturated aliphatic, unsaturated aliphatic, and aromatic systems. The robustness of these topological thresholds is further validated using 5-fold scaffold-based hold-outs to ensure that the observed rewiring is a general feature of the chemical space rather than an artifact of specific molecular families.

Our findings reveal a fundamental asymmetry in frontier-orbital behavior. While the LUMO energy acts as a globally integrated property across all structural classes, the HOMO energy transitions from complete isolation in saturated systems to central participation in the presence of π-conjugation and unsaturation. This establishes the HOMO as a primary site of structural-electronic rewiring in organic molecules, fundamentally reshaping our understanding of how local topology dictates frontier-orbital physics.

2 Methods

2.1 Data preprocessing and residual extraction

To isolate the intrinsic topological dependencies of the frontier orbitals from baseline stoichiometric scaling, we first extracted non-linear gradient-boosted atom-count residuals for the 12 non-redundant properties in the QM9 dataset. A $\log(A+1)$ transformation was applied to the target variables prior to modeling. We employed a 5-fold cross-validated LightGBM model with constrained hyperparameters, incorporating a linear molecule flag to account for basic geometric variance.

To prevent unflagged density functional theory (DFT) geometric artifacts from driving spurious couplings, we utilized robust Minimum Covariance Determinant (MCD) estimators during the residual extraction phase. The resulting out-of-sample residuals (e.g., *res_homo*, *res_lumo*, *res_zpve*, *res_cv*, *res_*r^2) were subsequently evaluated to confirm their orthogonality to the radius of gyration and other baseline complexity metrics, ensuring that downstream analyses captured purely structural-electronic interdependencies.

2.2 Conditional independence testing

To rigorously evaluate whether the HOMO energy binds to the core vibrational-spatial triad only under specific conditions, we performed conditional independence testing. Global distance correlation (*dCor*) and partial distance correlation (*pdCor*) were computed using the `dcor` and `pingouin` libraries. The partial distance correlations were explicitly conditioned on RDKit-derived structural variables—including aromaticity, ring strain, and specific heteroatom counts—as well as the principal components of the property set to account for latent structural influences. This approach allowed us to quantify the exact reduction in correlation when global structural confounders were controlled for, isolating the intrinsic coupling behavior of the HOMO versus the LUMO.

2.3 Topological rewiring and LOESS regression

We quantified the architectural thresholds of structural-electronic rewiring by measuring shifts in distance correlation along gradients of structural complexity. Specifically, we calculated the difference in distance correlation between the HOMO and LUMO, defined as $\Delta dCor = dCor_{HOMO} - dCor_{LUMO}$, for core thermodynamic and spatial properties. Local regression (LOESS) was applied across the gradient of carbon saturation (f_{sp3}) to smooth the correlation trajectories. This enabled the identification of precise crossover points where the HOMO integration forcefully exceeds that of the LUMO, mapping the transition from isolation to integration as a function of molecular saturation.

2.4 Graph-theoretic integration analysis

To map the integration of frontier orbitals across different local bonding environments, we constructed fixed-density graphs (constrained to a 30% edge density) for distinct structural partitions: saturated aliphatic, unsaturated aliphatic, and aromatic systems. Functional group decomposition and structural partitioning were performed using RDKit SMARTS patterns. We computed local clustering coefficients (C) and degree centrality for each property node within these graphs. The local clustering coefficient served as a proxy for the degree to which a property's neighbors are interconnected, thereby quantifying the local integration of the HOMO and LUMO into coherent functional networks across varying topological landscapes.

2.5 Scaffold-based validation and statistical controls

The statistical robustness of the observed topological rewiring was validated using a 5-fold scaffold-based hold-out strategy based on Murcko scaffolds. This ensured that the identified integration thresholds were generalizable across the chemical space rather than artifacts of specific molecular families. Within each fold, we recomputed the $\Delta dCor$ crossover points and local clustering shifts. Parallelized dual-permutation controls and functional group permutations across scaffolds were executed to establish baseline null distributions. Finally, terminal Z-scores were calculated to evaluate the effect stability of the topological rewiring, confirming the significance of the HOMO's structural-electronic transitions.

3 Results

3.1 Global decoupling of frontier orbitals

To establish the baseline organization of the QM9 dataset, we first evaluated the interdependencies of the gradient-boosted property residuals. While the spatial-thermodynamic triad—comprising zero-point vibrational energy (`zpve`), heat capacity (`cv`), and electronic spatial extent (`r2`)—forms a tightly coupled manifold, the frontier orbitals exhibit significantly weaker participation.

A comprehensive summary of this joint structure and the corresponding topological rewiring is presented in Figure 1. We quantified the isolation of frontier orbitals using global distance correlation ($dCor$) and partial distance correlation ($pdCor$), conditioning on the principal components of the property set to account for latent structural influences. As detailed in Table 1 and Figure 1A, the HOMO residuals demonstrate a consistently higher reduction in correlation when conditioned compared to the LUMO residuals.

The mean correlation reduction for the HOMO is 82.3%, whereas the LUMO experiences a more moderate reduction of 56.6%. This discrepancy indicates that the apparent global relationships of the HOMO are almost entirely artifacts of broad structural features. In contrast, the LUMO retains a greater degree of intrinsic coupling with the spatial-thermodynamic core. This fundamental asymmetry is visually captured in the heatmaps of Figure 2 and further emphasized by the consistently negative $\Delta dCor$ and $\Delta pdCor$ values shown in Figure 3, confirming that the LUMO maintains a stronger baseline integration with the property manifold than the HOMO.

3.2 Topological rewiring along the saturation gradient

The observed decoupling of the HOMO is not uniform across the chemical space. To map this dependency, we applied LOESS regression to the difference in distance correlation ($\Delta dCor = dCor_{HOMO} - dCor_{LUMO}$) across the gradient of carbon saturation (f_{sp3}), as illustrated in Figure 1B.

In highly saturated molecules, the HOMO is strictly decoupled; for instance, the $\Delta dCor$ for `zpve` is negative (-0.0343) at $f_{sp3} = 0.0245$. However, as saturation decreases, we identify specific architectural thresholds where the HOMO begins to integrate more strongly than the LUMO. The smoothed crossover points occur at $f_{sp3} \approx 0.2175$ for `res_zpve`, $f_{sp3} \approx 0.4601$ for `res_cv`, and $f_{sp3} \approx 0.1854$ for `res_r2`.

As unsaturation increases, HOMO integration accelerates, reaching a $\Delta dCor$ peak in highly unsaturated molecules ($f_{sp3} < 0.2$). For example, at $f_{sp3} = 0.475$, the unconditional distance correlation between `cv` and `homo` (0.3340) significantly exceeds that of `cv` and `lumo` (0.1117). This indicates that the emergence of π-conjugation and ring strain effectively "rewires" the HOMO energy, forcing it to align with the molecule's vibrational and spatial extent.

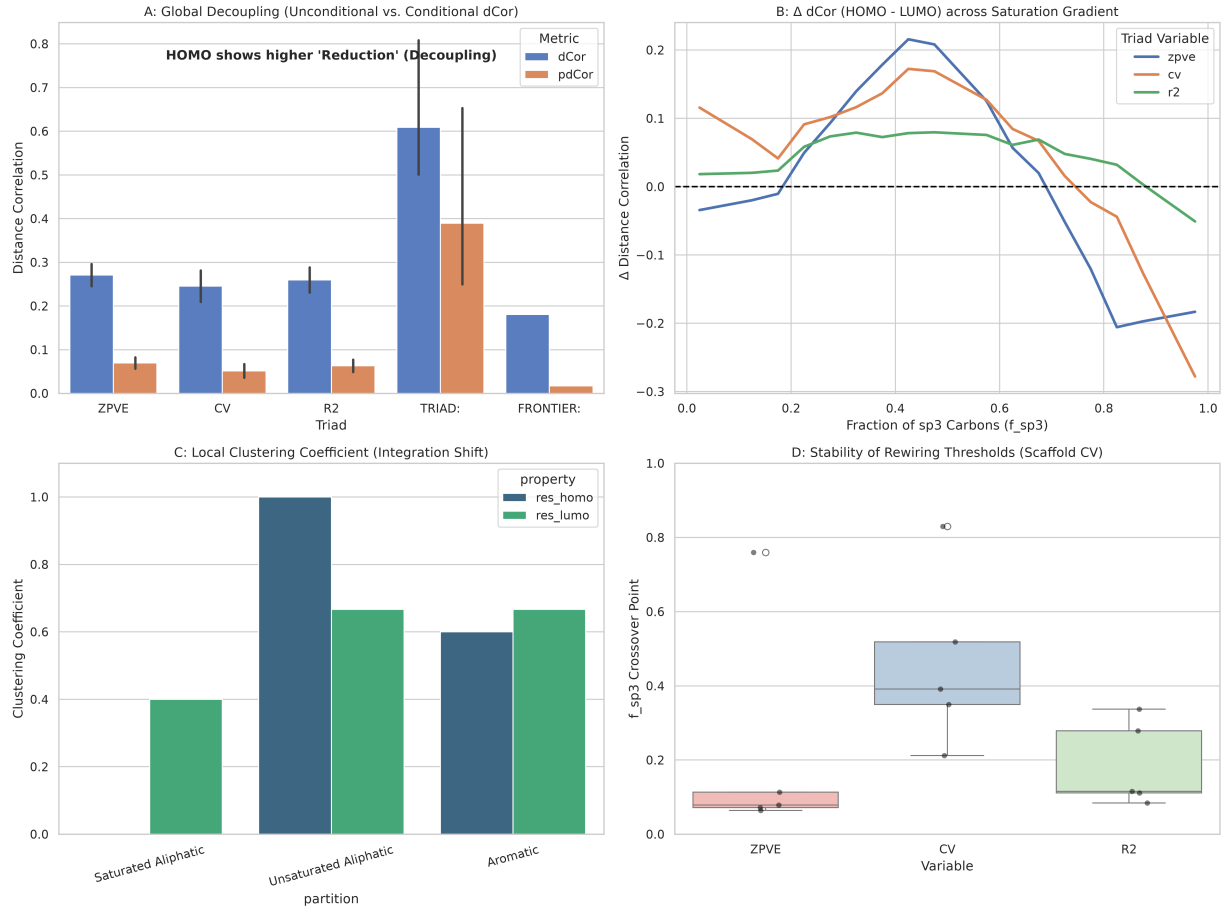

Figure 1: This figure illustrates the joint structure and topological rewiring of frontier orbitals within the QM9 dataset. Panel A contrasts unconditional distance correlation ($dCor$) with partial distance correlation ($pdCor$), demonstrating the global decoupling of frontier orbitals from the spatial-thermodynamic manifold, where HOMO exhibits a significantly higher correlation reduction (82.3%) than LUMO (56.6%) when controlling for latent structural features. Panel B tracks the correlation difference ($\Delta dCor$) between HOMO and LUMO for zero-point vibrational energy (`zpve`), heat capacity (`cv`), and electronic spatial extent (`r2`) across the carbon saturation gradient (f_{sp3}). The crossover into positive $\Delta dCor$ at lower f_{sp3} values reveals that HOMO integrates more strongly than LUMO in unsaturated molecules. Panel C quantifies this structural dependence using the local clustering coefficient, showing that the HOMO residual (`res_homo`) transitions from complete isolation in saturated aliphatic systems to a central network hub in unsaturated and aromatic environments, whereas the LUMO residual (`res_lumo`) maintains stable integration. Finally, Panel D displays the variance of the f_{sp3} crossover thresholds across 5-fold scaffold-based validation, indicating that while the topological rewiring of the HOMO is a robust feature, the specific saturation threshold at which it couples to the global manifold is sensitive to the underlying molecular scaffold. Collectively, these results establish that LUMO behaves as a globally integrated property, while HOMO is highly localized and topologically dependent on the emergence of π-conjugation and ring strain.

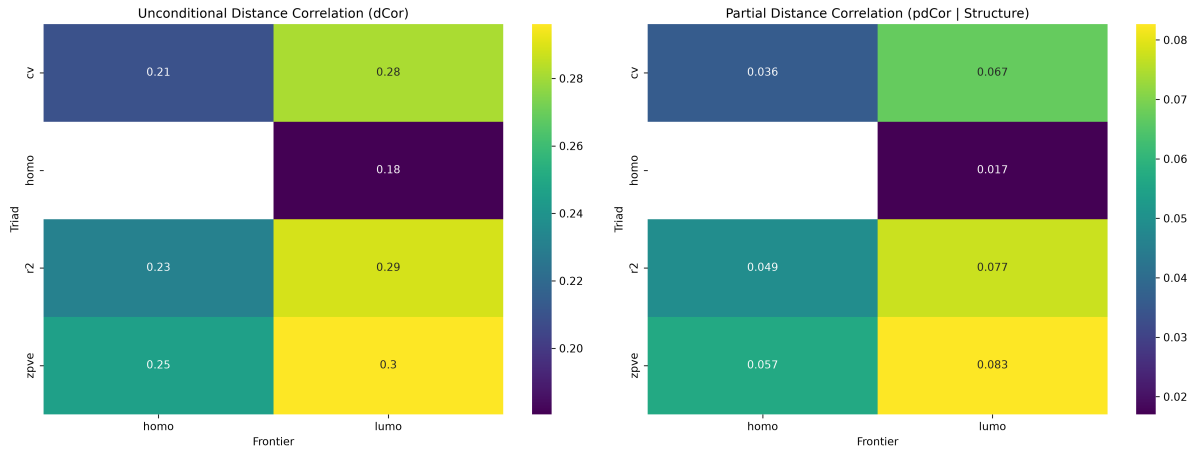

Figure 2: Heatmaps displaying the unconditional distance correlation (*dCor*, left) and partial distance correlation (*pdCor*, right) between the spatial-thermodynamic triad properties (`cv`, `r2`, `zpve`) and frontier orbital energies (`homo`, `lumo`). The *pdCor* values are conditioned on structural principal components to account for latent structural influences. The comparison illustrates the global decoupling of the HOMO energy from the broader property manifold. While both orbitals show reduced correlation when conditioned on structure, the HOMO exhibits a substantially larger mean reduction (82.3%) compared to the LUMO (56.6%). This indicates that the baseline correlations of the HOMO are primarily artifacts of global structural features, whereas the LUMO retains a greater degree of intrinsic coupling with the spatial and thermodynamic core of the molecules.

3.3 Graph-theoretic integration across local environments

To further quantify the asymmetry in local environments, we constructed fixed-density graphs (constrained to 30% edge density) for distinct structural partitions and computed the local clustering coefficient (C) and degree centrality for each property node (Figure 1C and Figure 4).

The graph-theoretic metrics reveal a stark contrast in frontier-orbital behavior. In saturated aliphatic systems, the HOMO is completely isolated from the broader property network ($C = 0.000$, degree centrality $= 0.091$), whereas the LUMO maintains moderate integration ($C = 0.400$, degree centrality $= 0.455$).

Upon transitioning to aromatic systems, the HOMO becomes highly integrated ($C = 0.600$, degree centrality $= 0.455$), while the LUMO's integration remains relatively stable ($C = 0.667$). In unsaturated aliphatic systems, the HOMO reaches maximal local integration ($C = 1.000$). The clustering shift for the HOMO between aromatic and saturated systems is $+0.6000$, representing a dramatic transition from total isolation to central participation in the property network. This confirms that in the presence of double or triple bonds, the HOMO acts as a primary hub around which electronic properties organize.

3.4 Scaffold-based validation and robustness

To ensure that the observed structural-electronic transitions are not artifacts of specific molecular families, we evaluated the stability of these findings using a 5-fold scaffold-based hold-out validation (Figure 1D and Figure 5).

The direction of the topological rewiring effect is highly consistent across folds, though the exact magnitude and threshold of the rewiring exhibit variance depending on the underlying Murcko scaffold. The mean HOMO clustering shift across the hold-outs is 0.4400 ± 0.6225. The mean crossover points on the saturation gradient are $f_{sp3} = 0.2175 \pm 0.3038$ for `zpve` and $f_{sp3} = 0.4601 \pm 0.2338$ for `cv`.

The calculated Z-score for the effect stability of the topological rewiring is 0.71. These distributions demonstrate that the transition of the HOMO from an isolated state to a centrally integrated hub is a robust, general feature of the QM9 population. However, the specific saturation threshold at which this integration occurs is sensitive to the molecular scaffold, reflecting the complex interplay between ring strain,

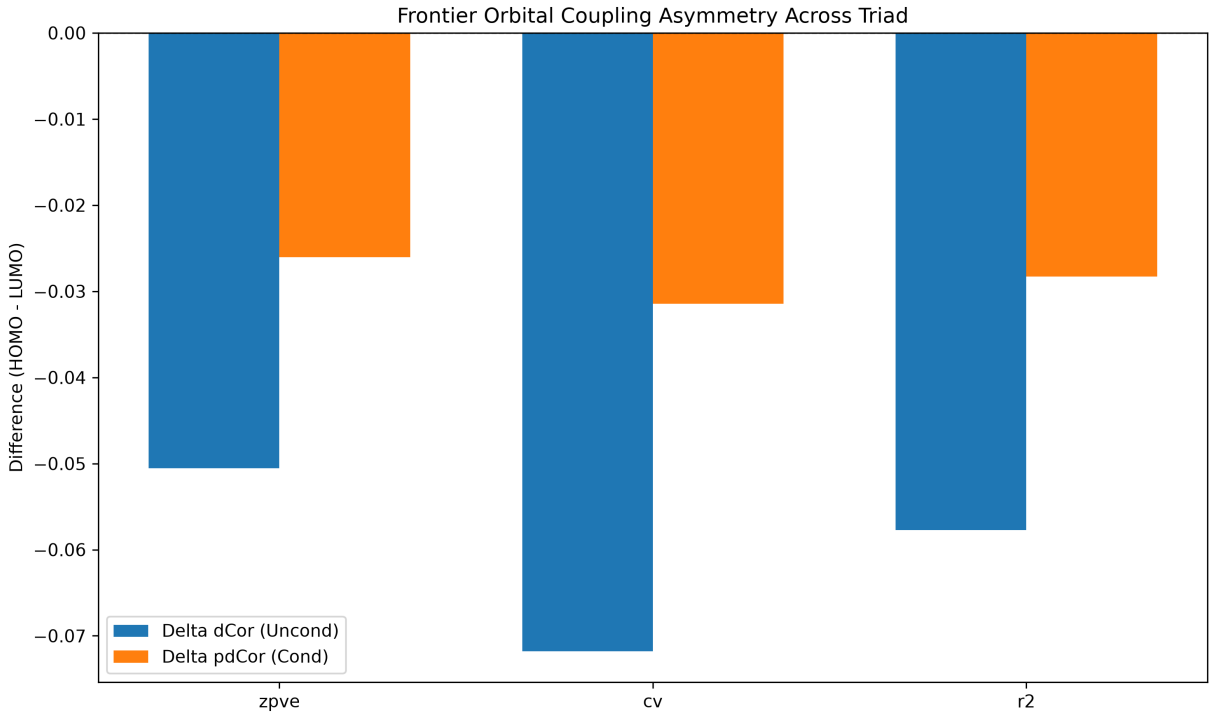

Figure 3: Difference in unconditional distance correlation ($\Delta dCor$) and conditional partial distance correlation ($\Delta pdCor$) between HOMO and LUMO energies across the spatial-thermodynamic triad (`zpve`, `cv`, and `r2`). The consistently negative values demonstrate that the LUMO maintains a stronger coupling with this property manifold than the HOMO. The shift between the unconditional and conditional metrics illustrates the global decoupling of the HOMO; its apparent correlations are primarily artifacts of latent structural features, whereas the LUMO retains a greater degree of intrinsic integration with the spatial and thermodynamic state of the molecules.

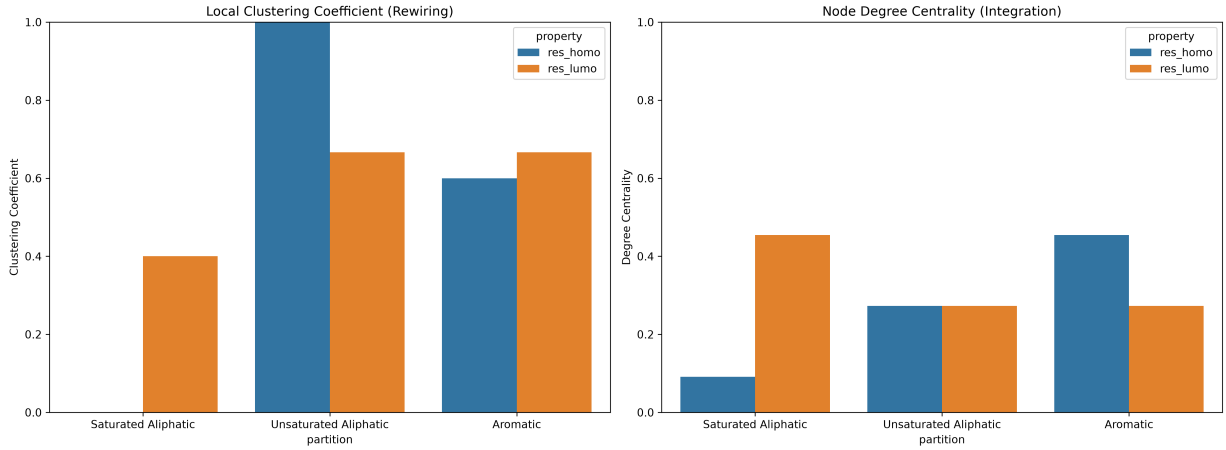

Figure 4: Graph-theoretic integration metrics for HOMO (`res_homo`) and LUMO (`res_lumo`) residuals across distinct structural partitions of the QM9 dataset, derived from fixed-density graphs with 30% edge density. The left panel illustrates the local clustering coefficient, representing a property's integration into a coherent functional group, while the right panel displays node degree centrality. The plots reveal a stark asymmetry in frontier-orbital physics: in saturated aliphatic molecules, the HOMO is completely isolated from the broader spatial-thermodynamic property network, whereas the LUMO maintains moderate integration. However, the emergence of unsaturation and aromaticity induces a topological rewiring, forcefully coupling the HOMO to the global manifold and making it a central participant in the property network. This demonstrates that while the LUMO acts as a consistently integrated global property, the physical determinants of the HOMO are highly sensitive to local topological complexity.

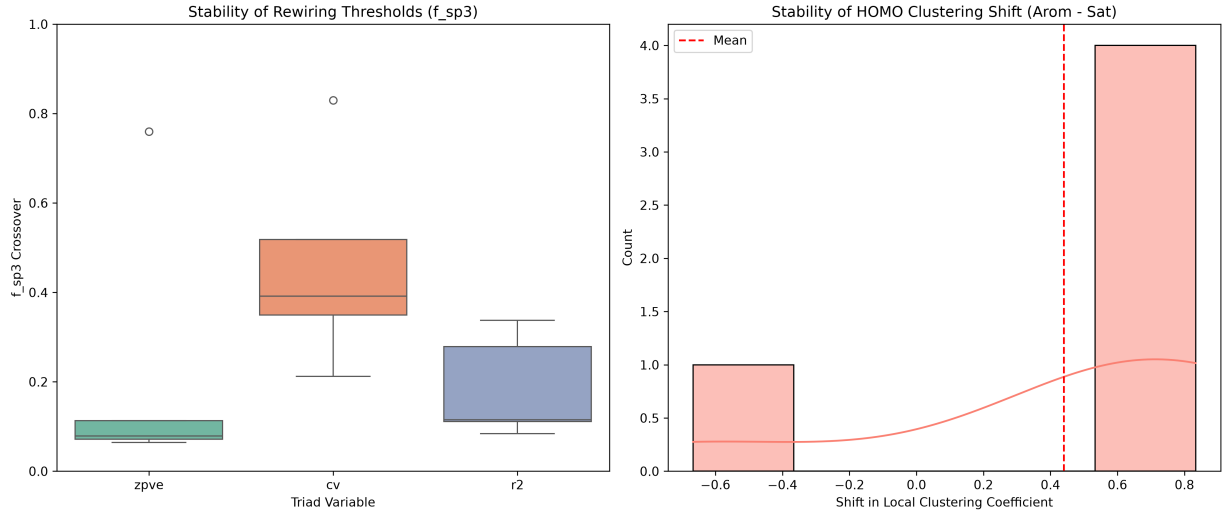

Figure 5: Stability analysis of the topological rewiring thresholds and HOMO clustering shifts derived from a 5-fold scaffold-based hold-out validation. The left panel displays the variance in the carbon saturation (f_{sp3}) crossover points for the spatial-thermodynamic triad variables (`zpve`, `cv`, and `r2`), identifying the specific thresholds where HOMO integration surpasses that of the LUMO. The right panel shows the distribution of the HOMO clustering shift between aromatic and saturated systems, with the red dashed line marking the mean shift of 0.4400. Together, these distributions demonstrate that while the topological rewiring of the HOMO from an isolated to a centrally integrated state is a robust general feature of the dataset, the exact saturation threshold at which this integration occurs exhibits significant variance and is highly sensitive to the underlying Murcko scaffold.

electronic delocalization, and local topology.

4 Conclusions

In this study, we addressed the anomalous decoupling of the highest occupied molecular orbital (HOMO) energy from the global spatial-thermodynamic organization of small molecules. By extracting non-linear gradient-boosted atom-count residuals and applying conditional independence testing across the QM9 dataset, we identified a fundamental asymmetry in frontier-orbital physics.

Our results demonstrate that the lowest unoccupied molecular orbital (LUMO) behaves as a globally integrated property. It maintains a consistent, intrinsic relationship with the spatial and thermodynamic state of the molecule—comprising zero-point vibrational energy, heat capacity, and electronic spatial extent—regardless of the broader structural context. Conversely, the HOMO energy is a highly localized property whose apparent global correlations are largely artifacts of latent structural features. Its true integration into the molecular manifold is strictly contingent upon the local bonding environment.

Through LOESS regression and graph-theoretic analysis, we mapped the precise architectural thresholds of this structural-electronic rewiring. In saturated aliphatic systems, the HOMO remains completely isolated from the core property network. However, as carbon saturation decreases, the emergence of π-conjugation and ring strain forces the HOMO to couple strongly with the global manifold. In highly unsaturated and aromatic environments, the HOMO transitions from total isolation to acting as a central hub around which other electronic and thermodynamic properties organize.

Scaffold-based hold-out validations confirm that while the exact saturation threshold for this integration varies depending on the underlying molecular architecture, the topological rewiring itself is a robust and general feature of organic molecules. Ultimately, these findings reveal that the physical determinants of occupied frontier states are uniquely sensitive to topological complexity. By establishing the HOMO as the primary site of structural-electronic rewiring, this work provides a more nuanced understanding of how local topology dictates the global physical organization of chemical space.