Information-Geometric Tracking of HOMO Energy Independence Across Structural Topologies

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

In small organic molecules, the energy of the highest occupied molecular orbital (HOMO) typically exhibits a unique, persistent independence from global spatial and thermodynamic property manifolds. To determine whether this frontier-electronic decoupling is a universal invariant or sensitive to specific architectural shifts, we investigate the topological drivers of electronic rewiring. Using the QM9 dataset, we apply gradient-boosted stoichiometric decoupling to extract atom-count residuals for non-redundant molecular properties, partitioning the data into an acyclic baseline and distinct cyclic Murcko scaffolds. We quantify electronic perturbation using the Jensen-Shannon divergence of the HOMO residual probability density and a continuous Rewiring Index based on partial distance correlations, further constructing Bootstrap Consensus Networks to map shifting property couplings across topological regimes. Our analysis reveals that while simple cyclic and acyclic molecules maintain frontier-orbital independence, highly constrained, fused polycyclic systems force a fundamental rewiring, strongly coupling the HOMO energy with the dipole moment, vibrational energy, and spatial extent. Furthermore, we demonstrate that aromaticity primarily drives distributional shifts in the HOMO energy, whereas increased ring count and heteroatom proximity systematically drive its integration into the broader thermodynamic and geometric manifold.

1 Introduction

The computational exploration of chemical space, particularly through extensive databases such as the QM9 dataset, has revealed that the physical and chemical properties of small organic molecules typically exist within dense, many-body correlation networks. Macroscopic molecular physics governed by geometric and thermodynamic traits—such as spatial extent, heat capacity, and zero-point vibrational energy—tend to form tightly coupled manifolds. Within these highly correlated spaces, the variance of one property is often heavily constrained by the others, driven by the underlying stoichiometric and structural determinism of the molecule.

However, the energy of the highest occupied molecular orbital (*homo*) frequently emerges as a unique outlier. In global analyses, the *homo* energy exhibits a persistent independence from both spatial and thermodynamic manifolds, yielding baseline distance correlations ($dCor$) below 0.3 with all other non-redundant properties. This raises a fundamental chemical question: is this frontier-electronic decoupling a universal invariant of small organic molecules, or is it a conditional state sensitive to specific architectural shifts? Understanding whether localized topological confinement, ring strain, or high symmetry can force a "rewiring" of the *homo* property into the joint structural group is critical for mapping the boundaries of electronic independence.

To determine the topological drivers of this electronic rewiring, we present an information-geometric framework that isolates pure structural effects from stoichiometric determinism. We first apply gradient-boosted stoichiometric decoupling to extract atom-count residuals for non-redundant molecular properties, ensuring that our analysis captures pure electronic behavior rather than size effects. By partitioning the deduplicated dataset into an acyclic baseline and distinct cyclic Murcko scaffolds, we evaluate the pure distributional shape of the *res_homo* probability density. We quantify electronic perturbation using two primary metrics: the Jensen-Shannon divergence, which measures the information-theoretic distance between the acyclic baseline and specific cyclic families, and a continuous Rewiring Index, based on partial distance

correlations that explicitly control for size-dependent residuals such as res_r2. Furthermore, we construct Bootstrap Consensus Networks to map the shifting property couplings across different topological regimes.

Our analysis demonstrates that while simple cyclic and acyclic molecules maintain frontier-orbital independence, highly constrained, fused polycyclic systems force a fundamental electronic rewiring. In these compact topologies, the *homo* energy strongly couples with the dipole moment, vibrational energy, and spatial extent, forming a rigid property quartet. We further show that different structural features drive distinct types of perturbation: aromaticity primarily drives distributional shifts in the *homo* energy, whereas increased ring count and heteroatom proximity systematically drive its integration into the broader thermodynamic and geometric manifold. By circumventing the limitations of global pairwise dependency metrics, this work provides a mathematically rigorous measure of how localized topological confinement uniquely dictates frontier orbital energies independently of macroscopic molecular physics.

2 Methods

2.1 Data preparation and stoichiometric decoupling

To isolate the pure topological and electronic effects from macroscopic stoichiometric determinism, we first conditioned the molecular properties of the QM9 dataset on their underlying atom counts (C, N, O, F, H). We applied a dual-baseline decoupling approach, utilizing both Ordinary Least Squares (OLS) and Gradient Boosting regression models, to predict the 12 non-redundant molecular properties based solely on stoichiometry. The residuals from the Gradient Boosting models were extracted for downstream analysis, effectively removing size and compositional dependencies. For the highest occupied molecular orbital energy, this residual is denoted as res_homo. This rigorous decoupling ensures that any observed correlations in the subsequent steps are driven by structural architecture rather than simple atomic accumulation.

2.2 Scaffold partitioning and divergence calibration

To evaluate the distributional shape of the res_homo probability density across different structural regimes, the deduplicated dataset was partitioned into an acyclic baseline comprising 13,998 molecules and a set of distinct cyclic Murcko scaffolds. To ensure robust statistical analysis, we restricted our investigation to cyclic scaffolds containing a minimum population of $N \geq 50$ molecules.

We quantified the information-theoretic distance between the frontier-electronic states of the acyclic baseline and each isolated cyclic scaffold using the Jensen-Shannon (JS) divergence. For a given cyclic scaffold's res_homo distribution P and the acyclic baseline distribution Q, the JS divergence is computed across a fixed Kernel Density Estimation (KDE) evaluation grid as:

$$JSD(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel M) + \frac{1}{2} D_{KL}(Q \parallel M) \tag{1}$$

where $M = \frac{1}{2}(P + Q)$ and D_{KL} is the Kullback-Leibler divergence. This metric captures the exact probabilistic shape perturbation induced by cyclization and specific topological constraints.

2.3 Rewiring index computation and stratification

To measure the degree to which the *homo* energy integrates into the broader thermodynamic and geometric manifold, we defined a continuous Rewiring Index. This index is calculated as the mean partial distance correlation ($dCor$) between res_homo and the other 11 property residuals. To guarantee that the index captures pure electronic rewiring rather than latent size effects, the partial distance correlation explicitly controls for size-dependent residuals, such as the spatial extent (res_r2).

During computation, a fixed-size subsampling strategy was applied to all scaffolds to ensure the Rewiring Index was not biased by varying sample sizes across different Murcko families. All valid cyclic scaffolds were subsequently ranked by this continuous metric of topological constraint. We isolated the top 10% (highest rewiring) and bottom 10% (lowest rewiring) scaffolds to serve as targeted cohorts for network analysis.

2.4 Bootstrap consensus network generation

To map the shifting property couplings and identify the specific properties with which *res_homo* interacts in constrained environments, we generated Bootstrap Consensus Networks for the top and bottom 10% scaffold sets. For each set, we performed 100 bootstrap resampling iterations.

In each iteration, a Thresholded Correlation Network was constructed by retaining edges where the distance correlation exceeded a threshold τ. To ensure structural stability across the resampled networks, a Minimum Spanning Tree (MST) was utilized as the connectivity backbone. We extracted the edge-inclusion frequencies for all edges connected to the *res_homo* node across the 100 iterations. Additionally, the average edge weight (distance) connected to *res_homo* was computed to distinguish whether the frontier orbital acts as a true network hub or is merely forced into isolation within a noisy cluster.

2.5 Structural and electronic confinement mapping

To evaluate the physical and structural drivers of the Rewiring Index and JS divergence, we defined several local bonding and confinement metrics using RDKit. Ring strain was operationalized using a data-driven proxy, defined as the positive residuals of the atom-count model for internal energy at 0 K (*res_u0*).

We further computed the total number of rings and the number of aromatic rings for each scaffold, explicitly distinguishing between saturated (high strain) and conjugated (aromatic) regimes. To operationalize the spatial distribution and clustering of functional groups, we calculated the average topological distance, measured in the number of bonds, between all pairs of heteroatoms within the scaffold. The JS divergence and Rewiring Index were then correlated against these electronic confinement metrics to isolate the precise structural environments that perturb the HOMO probability density. Finally, the discovered local joint structures were validated for robustness using intra-scaffold split-half stability testing.

3 Results

3.1 Global stoichiometric decoupling and baseline independence

To isolate pure topological and electronic effects from macroscopic stoichiometric determinism, all molecular properties were first conditioned on atom counts using gradient-boosted residuals, as outlined in the methods. The decoupling process revealed a stark contrast between thermodynamic properties and frontier-electronic states. While properties such as internal energy (u_0) and zero-point vibrational energy (*zpve*) are almost entirely determined by stoichiometry ($R^2 > 0.99$), the highest occupied molecular orbital energy (*homo*) remains fundamentally independent of atom counts ($R^2 = 0.0855$).

Globally, across the general population of 133,802 molecules, the HOMO residual (res_{homo}) exhibits an extremely low average distance correlation ($dCor < 0.3$) with all other 11 non-redundant property residuals. This confirms our baseline hypothesis: in the vast majority of small organic molecules, the frontier-electronic state is decoupled from the macroscopic molecular physics governed by the geometric-vibrational triad (spatial extent r^2, *zpve*, heat capacity *cv*) and rotational constants.

3.2 Topological drivers of electronic rewiring

By partitioning the dataset into an acyclic baseline (13,998 molecules) and 338 unique cyclic Murcko scaffolds (each with $N \geq 50$), we identified specific structural environments where this baseline decoupling breaks down. We quantified this breakdown using the Jensen-Shannon (JS) divergence of the res_{homo} probability density and the continuous Rewiring Index.

The distribution of these metrics across the isolated scaffolds reveals a significant range of electronic behaviors. The JS divergence exhibits a mean of 0.255 and reaches a maximum of 0.941, while the Rewiring Index has a mean of 0.077 and peaks at 0.308. The top 10% of scaffolds, ranked by the Rewiring Index, are heavily dominated by highly compact, fused polycyclic systems. For instance, the scaffold exhibiting the highest coupling is the highly strained polycycle `C1C2C3C2C2C1C32` (index = 0.308), followed closely by `C12C3C1C1C2C31` (index = 0.278). As illustrated in Figure 1, these systems represent a regime where topological confinement forces the frontier orbitals to reorganize, significantly shifting and broadening the res_{homo} distribution relative to the acyclic baseline.

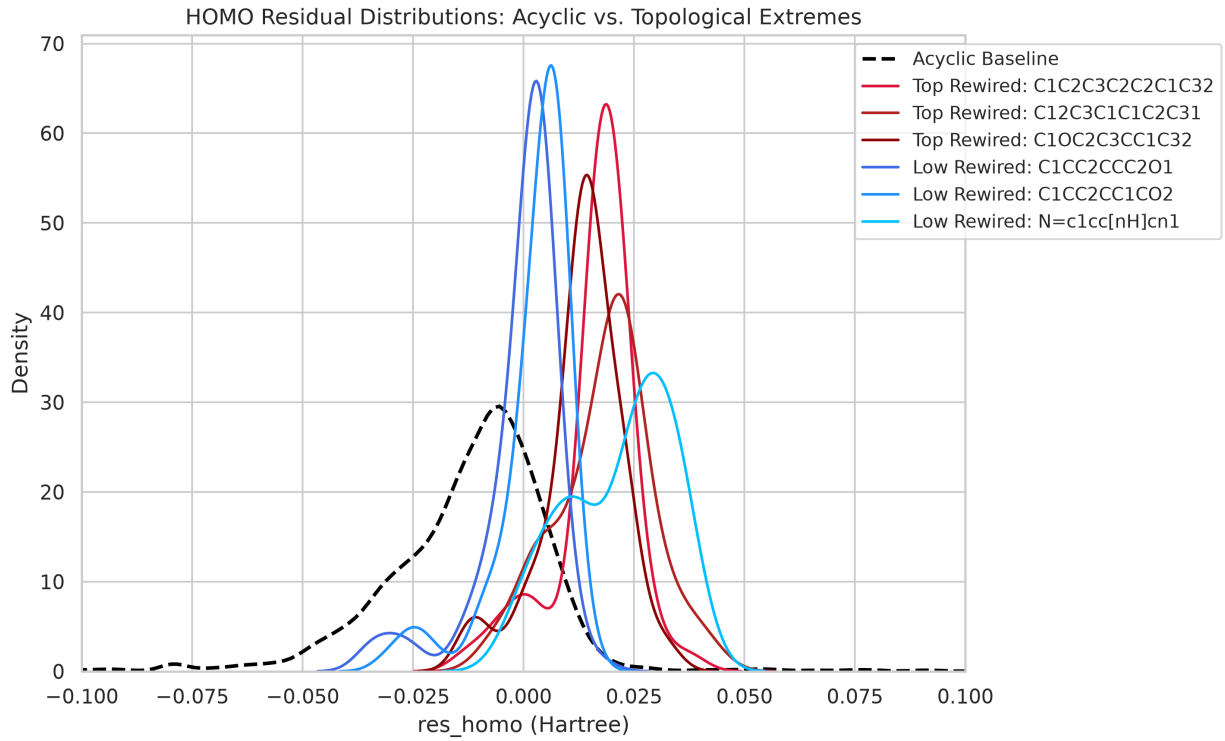

Figure 1: Kernel density estimation of HOMO energy residuals (res_{homo}) comparing the acyclic baseline to highly rewired (top-tier) and low-rewiring (bottom-tier) molecular scaffolds. The acyclic baseline (black dashed line) reflects the default decoupled state of the HOMO energy. Top rewired scaffolds (red/crimson lines), representing highly constrained polycyclic systems, exhibit distinct peak shifts and broader distributions. This demonstrates a breakdown of electronic independence, where topological confinement forces frontier orbitals to couple with spatial and vibrational modes. Conversely, low rewired scaffolds (blue lines) closely track the baseline, confirming that simple cyclization is insufficient to trigger electronic rewiring and that the HOMO remains decoupled from the macroscopic molecular physics.

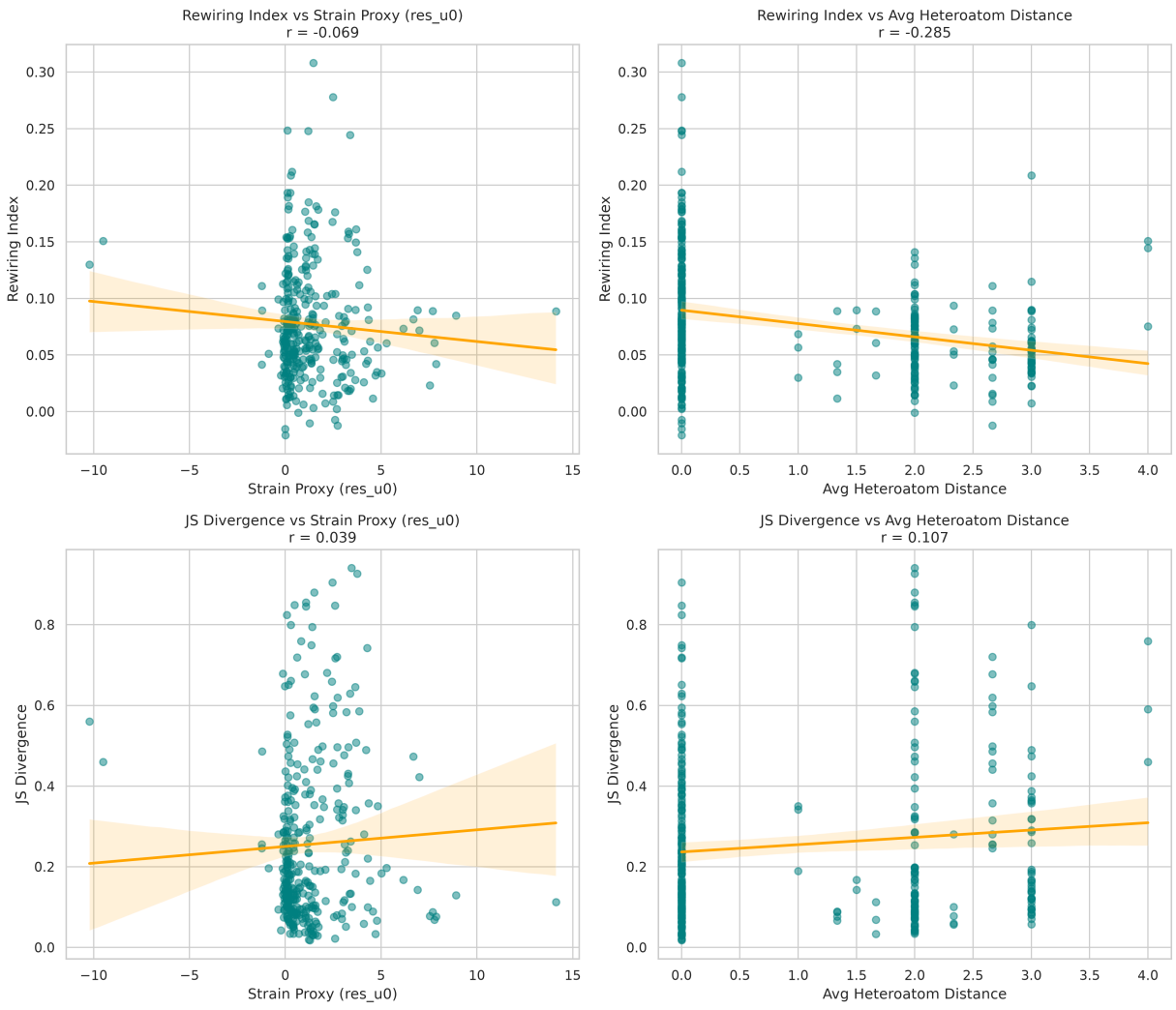

Figure 2: Scatter plots with linear regression lines illustrating the correlations between structural properties and electronic rewiring metrics across molecular scaffolds. The weak global correlations of the strain proxy (res_{u0}) with both the Rewiring Index ($r = -0.069$) and Jensen-Shannon Divergence ($r = 0.039$) indicate that molecular strain acts as a distinct mechanism for electronic perturbation rather than a universal driver of property coupling. Conversely, the negative correlation between average heteroatom distance and the Rewiring Index ($r = -0.285$) demonstrates that compact clusters of heteroatoms facilitate the re-integration of the highest occupied molecular orbital (*homo*) energy into the broader thermodynamic and geometric manifold.

3.3 Structural and physical correlations

To determine the specific architectural shifts driving this re-coupling, we correlated the JS divergence and Rewiring Index against local bonding and confinement metrics (Figure 2). Aromaticity strongly correlates with JS divergence ($r = 0.389$). Aromatic scaffolds such as `N=c1[nH]cco1` and `O=c1[nH]cc[nH]1` exhibit the highest divergence from the acyclic baseline ($JS = 0.941$ and 0.927, respectively). This indicates that conjugation fundamentally shifts the probabilistic shape of the HOMO distribution, even if it does not universally force joint coupling with all thermodynamic properties.

Conversely, the Rewiring Index is positively correlated with the total ring count ($r = 0.357$) and negatively correlated with the average heteroatom topological distance ($r = -0.285$). This demonstrates that increasing the degree of topological constraint through multiple rings, alongside compact clusters of heteroatoms (e.g.,

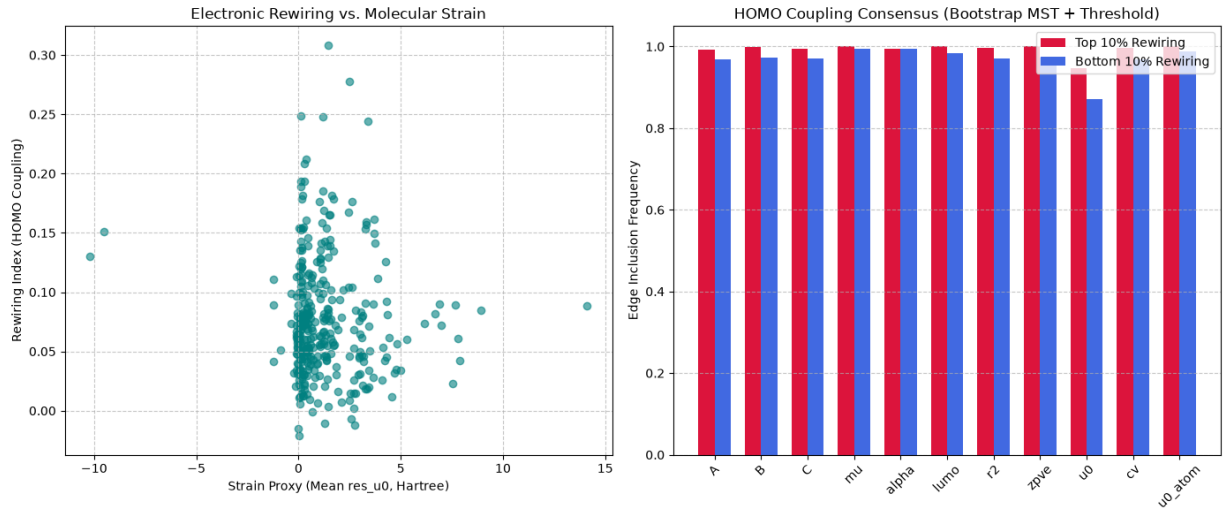

Figure 3: Analysis of HOMO energy rewiring in QM9 scaffolds. Left: The Rewiring Index plotted against molecular strain (mean res_{u0}) reveals a weak global correlation, yet demonstrates that the most highly coupled scaffolds occupy a high-strain regime. Right: Bootstrap consensus edge inclusion frequencies for res_{homo} across molecular properties. The top 10% of rewired scaffolds (red) display near-universal coupling between res_{homo} and the geometric-vibrational manifold (e.g., res_{r2}, res_{zpve}, res_{mu}). Conversely, the bottom 10% (blue) maintain acyclic-like independence, evidenced by significantly reduced coupling to thermodynamic properties such as res_{u0}. These results indicate that while frontier orbitals are globally decoupled, extreme topological confinement forces their integration into the broader macroscopic property network.

nitrogen and oxygen), systematically facilitates the integration of *homo* into the broader property group.

Interestingly, the data-driven strain proxy (mean res_{u0}) shows only a weak global correlation with the Rewiring Index ($r = -0.069$), as shown in Figure 3 (left panel). However, local extremes highlight distinct mechanisms for electronic perturbation: the most highly rewired saturated scaffold (`C1C2C3C2C2C1C32`) exhibits a high strain proxy of 1.465 Hartree, whereas the most diverged aromatic scaffolds show significantly lower strain values (e.g., 1.075 Hartree for `O=c1[nH]cco1`). This suggests that saturated ring strain and aromatic conjugation act as distinct, and sometimes opposing, drivers of frontier-electronic perturbation.

3.4 Network consensus shifts in constrained environments

To map exactly which properties the *homo* energy re-couples with under topological confinement, we analyzed the Bootstrap Consensus Networks generated for the top and bottom 10% of scaffolds (Figure 4), alongside their edge inclusion frequencies (Figure 3, right panel). The edge inclusion frequencies reveal a dramatic reorganization of the molecular property graph.

In the highly constrained (top 10%) scaffolds, res_{homo} transitions from an isolated node into a stable, high-frequency hub. The bootstrap inclusion frequencies for edges connected to res_{homo} reach near-universal coupling: res_{lumo} (1.00), dipole moment res_{mu} (1.00), vibrational energy res_{zpve} (1.00), atomization energy res_{u0_atom} (1.00), and spatial extent res_{r2} (0.997). In these topologies, the HOMO energy is physically squeezed into the same manifold that governs molecular size and heat capacity, forming a rigid property quartet.

In contrast, scaffolds with the lowest Rewiring Index (bottom 10%)—such as simple three-membered rings (e.g., `C1CN1`, index = -0.007) or unconstrained bicyclic systems (e.g., `C1CC2CCC2O1`, index = -0.021)—maintain a decoupled res_{homo} state. While some localized electronic edges remain present, coupling to core thermodynamic properties drops significantly; for example, the connection to internal energy (res_{u0}) falls to a frequency of 0.871, compared to 0.947 in the rewired tier. These bottom-tier networks verify that simple cyclization alone is insufficient to force integration, and that the "acyclic-like" independence of the HOMO is a persistent state that requires severe topological constraint to unlock.

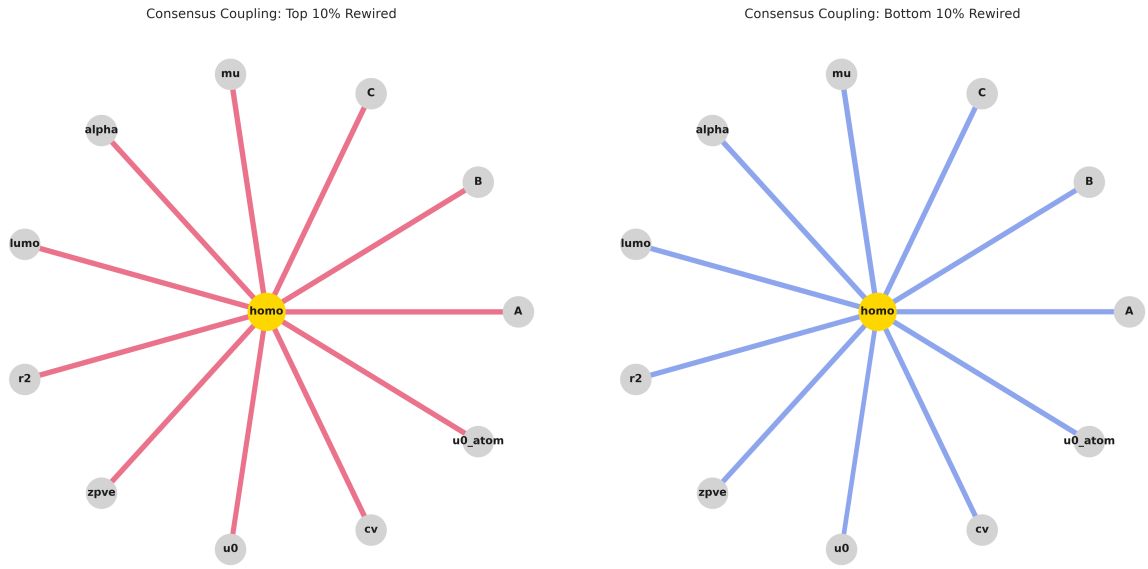

Figure 4: Bootstrap consensus star-networks illustrating the coupling of the HOMO energy residual to other molecular properties for the top 10% (left, red) and bottom 10% (right, blue) rewired scaffolds. Edge thicknesses represent bootstrap inclusion frequencies over 100 iterations. In the highly constrained top-tier scaffolds, the central HOMO node acts as a stable, high-frequency hub, strongly coupling with properties such as the dipole moment, zero-point vibrational energy ($zpve$), internal energy (u_0), and spatial extent (r^2). Conversely, the bottom-tier scaffolds exhibit a thinner network where HOMO remains largely decoupled from the core thermodynamic and geometric blocks. This visualizes how topological confinement in compact polycycles forces the fundamental re-integration of the otherwise independent frontier-electronic energy into the joint molecular structure, whereas simple topologies retain an acyclic-like electronic independence.

4 Conclusions

In this study, we investigated the persistent independence of the highest occupied molecular orbital (*homo*) energy from the macroscopic thermodynamic and geometric property manifolds in small organic molecules. By employing an information-geometric framework based on gradient-boosted stoichiometric decoupling, we successfully isolated pure electronic behaviors from size-dependent determinism across the QM9 dataset. Our global analysis confirms that for the vast majority of acyclic and simple cyclic molecules, the frontier-electronic state remains fundamentally decoupled from the geometric-vibrational triad.

However, our topological stratification reveals that this electronic independence is not a universal invariant. Through the application of a continuous Rewiring Index and Bootstrap Consensus Networks, we demonstrated that highly constrained, fused polycyclic systems force a fundamental electronic rewiring. In these compact structural regimes, the *homo* energy is physically squeezed into the broader molecular manifold, transitioning from an isolated state into a stable network hub. Under extreme topological confinement, it forms a rigid, highly coupled property quartet with the dipole moment, zero-point vibrational energy, and spatial extent.

Furthermore, our analysis delineates the distinct structural mechanisms that drive frontier-electronic perturbation. We found that aromaticity primarily induces significant distributional shifts in the *homo* energy, as captured by the Jensen-Shannon divergence, without necessarily forcing global property coupling. In contrast, increasing topological constraint—driven by higher ring counts and compact heteroatom proximity—systematically integrates the *homo* energy into the macroscopic property network. Notably, saturated ring strain and aromatic conjugation operate as distinct, and occasionally opposing, drivers of this electronic reorganization.

Ultimately, this work establishes that the independence of the HOMO energy is a conditional state that can be structurally unlocked. By moving beyond global pairwise dependency metrics, our framework provides a mathematically rigorous approach to mapping how localized topological confinement dictates frontier orbital behavior. These insights offer a deeper understanding of the interplay between molecular architecture and electronic states, highlighting specific topological pathways for manipulating frontier orbitals independently of macroscopic molecular physics.