Conditional Sensitivity Analysis of HOMO Decoupling via Partial Correlation Networks

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The highest occupied molecular orbital (HOMO) often exhibits an unexplained orthogonality to the global thermodynamic and geometric states of small molecules. To identify the structural conditions that break this decoupling, we perform a conditional sensitivity analysis to isolate chemical architecture from stoichiometric determinism. Using the QM9 dataset, we partition molecules into aromatic-rich and highly saturated, aromatic-free regimes. We apply compositional residualization and estimate sparse precision matrices via Gaussian Graphical Models to construct partial correlation networks, which allow us to measure the topological distance between the electronic frontier and macroscopic response properties. Our network analysis reveals that aromaticity acts as a structural catalyst for re-entanglement. In saturated molecules, the joint property space is highly ill-conditioned, and the HOMO remains topologically isolated. Conversely, the introduction of aromatic rings stabilizes the global correlation structure and significantly reduces the network path length between the HOMO and core thermodynamic properties, such as polarizability and heat capacity. Validated against shuffled null models, these findings demonstrate that HOMO decoupling is not a uniform feature of the molecular property space, but a conditional state strictly governed by molecular topology and delocalized electron systems.

1 Introduction

The highest occupied molecular orbital (HOMO) plays a fundamental role in determining chemical reactivity, ionization potential, and electron transport. However, in large-scale analyses of small molecules, such as those comprising the QM9 dataset, the HOMO energy often exhibits an unexplained orthogonality to the global thermodynamic and geometric states of the molecule. While macroscopic properties such as zero-point vibrational energy, heat capacity, and spatial extent form tightly coupled manifolds, the electronic frontier frequently remains topologically isolated. Understanding whether this decoupling is a universal feature of the molecular property space or a conditional state governed by specific structural architectures remains a critical open question.

Previous global analyses tend to average out local functional effects, potentially masking the structural conditions that break this decoupling. We hypothesize that the orthogonality of the HOMO is not uniform but is instead strictly governed by molecular topology and the presence of delocalized electron systems. Specifically, we propose that rigid aromatic confinements and delocalized π-electrons act as structural catalysts that force the HOMO to re-entangle with the macroscopic thermodynamic state.

To isolate the effects of chemical architecture from stoichiometric determinism, we perform a conditional sensitivity analysis. We apply compositional residualization to remove extensive size effects and basic atom-count dependencies, ensuring that any observed rewiring is due to molecular topology rather than basic chemical composition. By partitioning the molecular space into aromatic-rich and highly saturated, aromatic-free regimes, we can directly compare distinct chemical architectures. We then estimate sparse precision matrices via Gaussian Graphical Models (GGMs) to construct partial correlation networks. This targeted statistical framework replaces high-dimensional black boxes with interpretable boundaries, allowing us to measure the exact topological distance between the HOMO and core macroscopic properties.

Our network analysis reveals a macroscopic reorganization of the joint property manifold upon the introduction of aromatic rings. In saturated, acyclic architectures, the joint property space is highly ill-conditioned, and the HOMO remains topologically isolated. Conversely, aromaticity stabilizes the global

Target Property	Aromatic-rich Path Length	Aromatic-free Path Length	Ratio (Free/Rich)
Isotropic Polarisability (α)	2.756	10.931	3.97x
Heat Capacity (c_v)	3.772	9.845	2.61x
Zero-point Vibrational Energy ($zpve$)	4.701	8.447	1.80x

correlation structure and significantly reduces the network path length between the HOMO and core thermodynamic properties. Table 1 highlights this targeted path length reduction across key target properties, demonstrating that aromaticity acts as a bridge that tightly couples the electronic frontier to the molecular response to geometric and thermal perturbations.

Validated against shuffled null models, these findings confirm that the observed regime changes—including a sharp decrease in the covariance condition number and a significant reduction in average network path lengths—are genuine chemical phenomena. By mapping the conditional dependencies of the molecular property space, this study provides a rigorous framework for understanding exactly which local bonding environments govern the electronic-thermodynamic divide.

2 Methods

2.1 Data curation and preprocessing

To investigate the topological dependencies of the molecular property space, we utilized the QM9 dataset. Initial data curation involved deduplicating the dataset based on SMILES strings, which removed 83 redundant entries. The deduplicated SMILES were parsed into molecular objects using RDKit, with parallel processing employed to ensure computational efficiency. To reduce trivial collinearity and focus on fundamental thermodynamic and electronic properties, seven redundant derived features were excluded from the analysis: the HOMO-LUMO gap, and the uncorrected and atom-corrected internal energy, enthalpy, and free energy at 298 K.

The rotational constant A exhibited extreme heavy-tailed outliers alongside valid zeros corresponding to linear molecules. To stabilize its variance without distorting the physical meaning of the zero values, we applied a robust rank-based inverse normal transformation. This quantile transformation mapped the distribution to a standard normal space while preserving the rank order and appropriately handling the valid zeros.

2.2 Compositional residualization

To isolate the effects of molecular topology from extensive size effects and basic stoichiometric determinism, we employed an enhanced compositional residualization strategy. For each molecule, we computed the exact counts of carbon, nitrogen, oxygen, fluorine, and hydrogen atoms, alongside explicit structural covariates for the number of heteroatoms and aromatic rings.

For each of the 12 remaining target properties, we fitted a linear regression model incorporating a second-degree polynomial expansion of the atom counts, concatenated with the explicit heteroatom and aromatic ring counts. The residuals from these models were extracted to form a new, composition-controlled property space. This procedure ensures that subsequent network analyses capture how atoms are connected rather than merely what atoms are present. Finally, we computed the Pearson correlation matrix of these residuals; any pair exhibiting an absolute correlation greater than 0.95 was flagged, and one variable from the pair was dropped to prevent singular matrices in downstream precision matrix estimation.

2.3 Targeted sub-grouping by chemical architecture

Rather than employing continuous sliding windows that might conflate diverse chemical architectures, we partitioned the dataset into two distinct, chemically interpretable regimes. The "Aromatic-rich" subset was defined as molecules containing at least one aromatic ring, while the "Aromatic-free" (highly saturated) subset comprised molecules with zero aromatic rings and a high fraction of sp^3-hybridized carbons.

To rigorously prevent size and composition from confounding the comparison between these two regimes, we performed stratified sampling based on the joint distribution of heavy atom counts and the number

of heteroatoms. This binning strategy ensured that the two subsets were statistically indistinguishable regarding molecular size and heteroatom presence, thereby isolating aromaticity as the primary structural variable driving any observed topological rewiring.

2.4 Sparse precision matrix estimation

To map the conditional dependencies between the residualized properties, we estimated sparse precision matrices (inverse covariance matrices) for each targeted subset using a Gaussian Graphical Model (GGM). The estimation was performed via cross-validated Graphical Lasso.

To ensure that the resulting network topologies were robust chemical features rather than artifacts of the regularization penalty, we conducted a sensitivity analysis on the sparsity parameter, α. The GGM was evaluated across a predefined grid of α values. We extracted the adjacency matrices and non-zero partial correlations for the optimal α selected by cross-validation, verifying that the presence or absence of key edges remained stable across the parameter grid.

2.5 Network distance and topological analysis

To quantify the decoupling of the highest occupied molecular orbital (*homo*) from the macroscopic thermodynamic state, we constructed weighted partial correlation networks from the optimal precision matrices. For any pair of properties i and j with a non-zero partial correlation ρ_{ij}, the network edge weight w_{ij} was defined as the inverse of the absolute partial correlation:

$$w_{ij} = \frac{1}{|\rho_{ij}|} \tag{1}$$

Edges corresponding to zero partial correlations were omitted, representing conditional independence. This weighting scheme ensures that stronger conditional dependencies correspond to shorter topological distances. We then applied Dijkstra's algorithm to compute the shortest path length between the *homo* node and the core response and thermodynamic properties, such as isotropic polarizability, heat capacity, and zero-point vibrational energy.

2.6 Covariance stability and null model validation

To detect macroscopic regime changes in the joint property space, the residuals within each subset were standardized to zero mean and unit variance. We then computed the condition number and the log-determinant of the empirical covariance matrix for both the Aromatic-rich and Aromatic-free groups. A significant divergence in these metrics serves as an indicator of rapid reorganization in the global correlation structure.

Finally, to validate that the observed network structures and path lengths were genuine chemical phenomena, we established a baseline using a shuffled null model. We generated a permuted dataset where the SMILES strings and their corresponding structural subset labels were randomly reassigned to the property vectors. The entire pipeline—including GGM estimation, condition number calculation, and path length computation—was repeated on this shuffled dataset. Comparing the true subsets against these null baselines confirmed that the observed re-entanglement of the HOMO is strictly governed by the underlying molecular topology rather than statistical noise inherent to the estimation process.

3 Results

3.1 Macroscopic reorganization of the property space

Following the compositional residualization and targeted sub-grouping described in the methods, we first evaluated the global stability of the joint property manifold. The transition from saturated, acyclic architectures to aromatic systems triggers a fundamental macroscopic reorganization. This shift is quantitatively captured by the condition number of the standardized residual covariance matrices.

As illustrated in Figure 1, the Aromatic-free regime exhibits a highly ill-conditioned joint property space, characterized by a condition number of 5321.76. In stark contrast, the Aromatic-rich regime demonstrates

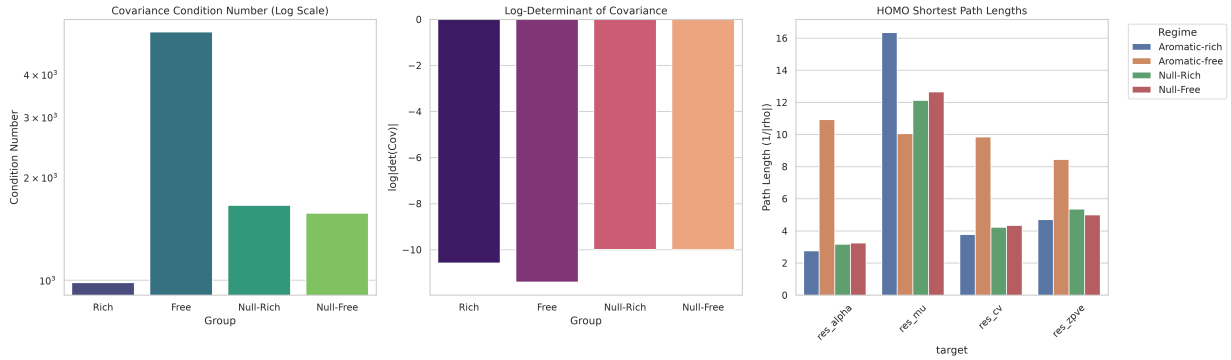

Figure 1: Macroscopic stability metrics and network topography comparing Aromatic-rich and Aromatic-free molecular regimes alongside shuffled null models. The left and center panels illustrate the covariance condition number and log-determinant, highlighting a severe instability in the Aromatic-free regime (condition number 5321.76) compared to the Aromatic-rich regime (983.16) and uniform null baselines. This divergence confirms a collapse of the global correlation structure in saturated molecules, where properties become more independent. The right panel displays the shortest path lengths from the *homo* orbital to target properties: isotropic polarisability (α), dipole moment (μ), heat capacity (c_v), and zero-point vibrational energy ($zpve$). Aromatic systems exhibit significantly shorter path lengths to α, c_v, and $zpve$, demonstrating that aromaticity acts as a structural bridge to re-entangle the electronic frontier with the macroscopic response block. Conversely, the pronounced path length to μ in aromatic systems reflects a robust conditional decoupling.

a significantly more stable global correlation structure, with the condition number dropping to 983.16. This 5.41-fold spike in instability for saturated molecules indicates a collapse of the coherent binding observed in aromatic systems, leading to a manifold where properties become more independent and topologically isolated.

To verify that these structural shifts are driven by chemical architecture rather than sample size or basic stoichiometric composition, we compared these metrics against shuffled null models ($n = 13,200$ for each group). The null models returned nearly uniform condition numbers of 1654.97 for the Null-Rich group and 1569.46 for the Null-Free group. The stark divergence of the true chemical data from these uniform baselines confirms that the regime-specific behavior is a genuine topological phenomenon rather than a statistical artifact.

3.2 Conditional independence and network sparsity

To map the specific conditional dependencies driving this macroscopic reorganization, we estimated sparse precision matrices using Gaussian Graphical Models. The resulting partial correlation heatmaps (Figure 2) reveal distinct structural break-points between the two regimes.

The true chemical networks exhibit significant sparsity, with approximately 35.56% to 37.78% of the partial correlations driven to exactly zero. In contrast, the precision matrices of the shuffled null models are 0% sparse (fully dense). This high degree of sparsity in the true data confirms that the observed conditional independencies are robust chemical constraints. For instance, even in the highly entangled aromatic regime, the *homo* orbital maintains a strict conditional decoupling from the dipole moment (μ), exhibiting a pronounced path length of 16.35. This indicates that while aromaticity re-entangles the electronic frontier with thermodynamic properties, it preserves specific, physically meaningful orthogonalities.

3.3 Topological entanglement of the electronic frontier

By converting the sparse precision matrices into weighted partial correlation networks, we quantified the topological distance between the *homo* orbital and the core macroscopic response block. Figure 3 visualizes

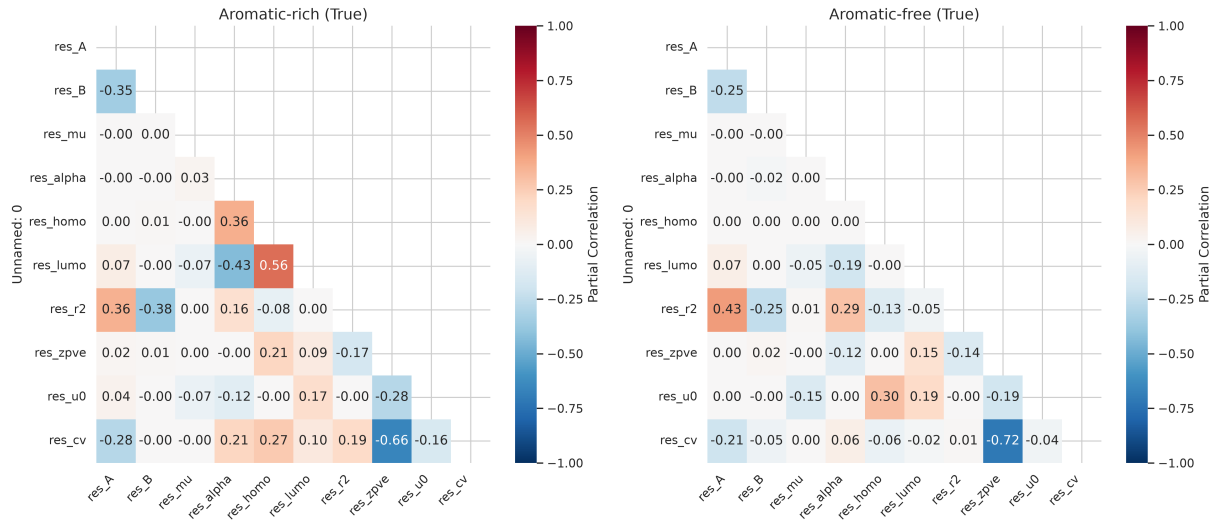

Figure 2: Partial correlation matrices for molecular properties in Aromatic-rich (left) and Aromatic-free (right) regimes, estimated via Sparse Precision Matrix Estimation. The heatmaps illustrate the macroscopic reorganization of the joint property manifold and the conditional sensitivity of the highest occupied molecular orbital (*homo*). In the saturated, Aromatic-free regime, *homo* is largely decoupled from the core macroscopic response block, exhibiting zero or weak partial correlations with properties such as isotropic polarisability (α) and heat capacity (c_v). Conversely, the Aromatic-rich regime displays stronger conditional dependencies between *homo* and these properties, translating to significantly shorter topological path lengths. The observed network sparsity (35.56% to 37.78% zero partial correlations) reflects robust chemical constraints. Comparing the two regimes reveals that aromaticity acts as a structural bridge, re-entangling the otherwise isolated electronic frontier with the global thermodynamic and geometric state of the molecule.

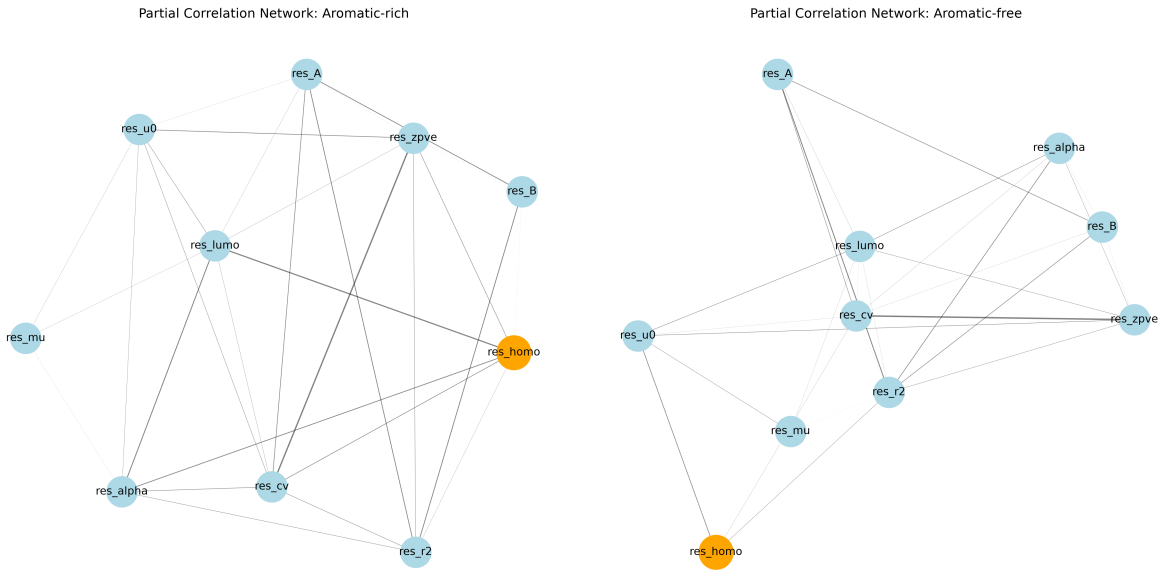

Figure 3: Partial correlation networks illustrating the conditional entanglement of the highest occupied molecular orbital (*homo*, highlighted in orange) within Aromatic-rich and Aromatic-free molecular regimes. Edge weights represent the inverse of absolute partial correlations, mapping the topological distance between properties. The graphs demonstrate that in saturated, Aromatic-free molecules, *homo* remains isolated at the network periphery with a high average path length of 9.8180. Conversely, the Aromatic-rich network reveals a macroscopic reorganization where delocalized π-electron systems act as a bridge, re-entangling *homo* with the thermodynamic and geometric core. This is evidenced by significantly reduced path lengths to isotropic polarisability (α), heat capacity (c_v), and zero-point vibrational energy ($zpve$). This structural recoupling visually corroborates the enhanced global stability of aromatic systems, which exhibit a 5.41-fold decrease in covariance condition number compared to their acyclic counterparts, while maintaining robust chemical constraints through 35.56% to 37.78% network sparsity.

these networks, highlighting the structural isolation of *homo* in saturated systems versus its integration in aromatic systems.

In the Aromatic-free regime, the *homo* node remains largely isolated from the core network, maintaining an average path length of 9.8180 to all other properties. However, the introduction of aromatic rings acts as a structural catalyst for re-entanglement, reducing the average path length to 6.8953. This 1.42-fold reduction in average path length (the global entanglement ratio) quantifies the recoupling effect driven by delocalized π-electron systems.

This re-entanglement is most pronounced when examining the shortest paths to specific thermodynamic and geometric properties. As detailed in Table 1, the topological distance from *homo* to isotropic polarisability (α), heat capacity (c_v), and zero-point vibrational energy ($zpve$) is drastically reduced in the presence of aromaticity.

The path length to polarisability experiences the most significant reduction, decreasing by a factor of 3.97 in aromatic systems. Heat capacity and zero-point vibrational energy follow similar trends, with path length reductions of 2.61x and 1.80x, respectively. These targeted path length reductions demonstrate that the orthogonality of the *homo* is not a universal constant of the molecular property space. Instead, it is a conditional state: in saturated molecules, the electronic frontier is topologically isolated, whereas aromaticity provides the necessary structural bridge to tightly couple the *homo* to the molecule's macroscopic response to thermal and geometric perturbations.

4 Conclusions

4.1 Summary of findings

This study resolves the unexplained orthogonality of the highest occupied molecular orbital (*homo*) within the QM9 dataset by demonstrating that its decoupling is not a universal constant of the molecular property space. Instead, it is a conditional state strictly governed by molecular topology. By isolating chemical architecture from stoichiometric determinism through compositional residualization and mapping conditional dependencies via Gaussian Graphical Models, we identified aromaticity as the primary structural catalyst for electronic-thermodynamic re-entanglement.

Our analysis revealed a fundamental macroscopic reorganization of the joint property manifold upon the introduction of aromatic rings. Saturated, acyclic architectures exhibit a highly ill-conditioned property space, characterized by a covariance condition number of 5321.76. In these aromatic-free systems, the *homo* remains topologically isolated at the periphery of the partial correlation network. Conversely, the presence of delocalized π-electron systems in aromatic-rich molecules stabilizes the global correlation structure, reducing the condition number by a factor of 5.41.

This macroscopic stabilization is mirrored by a significant topological recoupling. The average network path length from the *homo* to the rest of the molecular properties decreases by a factor of 1.42 in aromatic systems. As summarized in Table 1, this re-entanglement is particularly pronounced for core macroscopic response properties. The topological distance to isotropic polarisability (α), heat capacity (c_v), and zero-point vibrational energy (*zpve*) collapses dramatically, proving that aromaticity acts as a structural bridge connecting the electronic frontier to the thermodynamic core.

Crucially, the validation against shuffled null models—which exhibited uniform condition numbers and fully dense networks—confirms that these regime changes and the observed network sparsity (35.56% to 37.78%) are robust, genuine chemical phenomena rather than statistical artifacts of the estimation process. Even within the highly entangled aromatic regime, specific conditional independencies, such as the decoupling of *homo* from the dipole moment, are strictly preserved.

4.2 Broader implications

By replacing high-dimensional, global averages with interpretable topological boundaries, this conditional sensitivity analysis provides a rigorous framework for understanding the electronic-thermodynamic divide. Recognizing that the HOMO only couples to macroscopic states under specific structural conditions, such as rigid aromatic confinement, has significant implications for predictive modeling and molecular design. It highlights the danger of assuming uniform property manifolds across diverse chemical spaces. Future work can leverage these localized partial correlation networks to identify further structural break-points, ultimately guiding the targeted engineering of molecular systems where electronic and thermodynamic behaviors must be precisely co-optimized.