Information-Geometric Tracking of HOMO Decoupling via Property Dependency Networks Across Saturation Gradients

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Understanding how quantum chemical properties jointly organize—and specifically why the highest occupied molecular orbital (HOMO) energy decouples from macroscopic descriptors—is critical for mapping molecular property manifolds. We investigate this phenomenon by tracking the topological evolution of property dependency networks across molecular saturation gradients. Using the QM9 dataset, we isolate twelve non-redundant properties from stoichiometric and size effects, partitioning the molecules into discrete classes based on their fraction of sp3 carbons. For each class, we construct distance correlation networks and analyze them using persistent homology and degree centrality to quantify structural deformations in the property manifold. Our analysis reveals that the HOMO energy systematically decouples from the global spatial-thermodynamic axis, an isolation that intensifies and peaks within highly saturated regimes. Furthermore, persistence entropy measurements uncover a distinct topological phase transition at high saturation levels, marked by the robust emergence of one-dimensional cycles in the dependency network. Validated through disjoint scaffold-split Wasserstein stability, these findings establish that the degree of molecular conjugation fundamentally regulates the joint organization of quantum properties, driving a transition from hierarchical dependencies in conjugated systems to complex, multi-way relationships in saturated molecules.

1 Introduction

1.1 Motivation and background

The mapping of molecular property manifolds is a central challenge in quantum chemistry and cheminformatics. While predictive models routinely target individual molecular properties with high accuracy, the joint organizational structure of these properties—specifically how they couple and decouple across different chemical regimes—remains less understood. A particularly striking phenomenon within this domain is the systematic decoupling of the highest occupied molecular orbital (HOMO) energy from the global spatial-thermodynamic axis. Understanding the geometric and topological nature of this decoupling is critical for characterizing the intrinsic physical organization of molecular states.

Traditional analyses often rely on linear correlations that fail to capture complex, multi-way dependencies, or they conflate intrinsic quantum property relationships with macroscopic scaling effects such as molecular size and stoichiometry. Consequently, standard property prediction paradigms lack a mathematically precise characterization of how localized quantum states couple to a molecule's overarching physical organization, and how this coupling deforms across structural gradients.

1.2 Information-geometric framework

To address this gap, we introduce an information-geometric framework designed to track the topological evolution of property dependency networks. Using the QM9 dataset, we isolate twelve non-redundant quantum chemical properties from stoichiometric and heavy-atom baseline effects. This residualization ensures that the observed correlations represent intrinsic chemical dependencies rather than shared scaling with molecular size. We then partition the chemical space into discrete structural classes based on their degree of saturation, quantified by the fraction of sp^3 hybridized carbons (F_{sp^3}).

Bin	F_{sp^3} Range	HOMO Z-score	H_0 Entropy	H_1 Entropy	$W_1(H_0)$	$W_1(H_1)$
0	[0.0, 0.2)	-0.4307	2.1950	0.0000	0.5850	0.0640
1	[0.2, 0.4)	-0.7456	2.1926	0.0000	0.5478	0.0891
2	[0.4, 0.5)	-0.3799	2.1623	0.0000	0.4535	0.0783
3	[0.5, 0.6)	-1.1464	2.1383	0.0000	0.2957	0.0518
4	[0.6, 0.8)	-1.2065	2.1401	0.6089	0.3569	0.1391
5	[0.8, 1.0]	-1.1471	2.1462	0.0000	0.2792	0.0434

Table 1: Topological Metrics across Saturation Gradient

Within each saturation regime, we construct weighted, unit-invariant property dependency graphs using complete distance correlation matrices. By applying persistent homology and network degree centrality measures to these manifolds, we rigorously quantify structural deformations in the property space. This approach allows us to determine whether the HOMO energy represents a universally isolated node or if its decoupling is dynamically regulated by macroscopic topological transitions driven by the underlying molecular scaffold.

1.3 Summary of findings

Our analysis reveals that the degree of molecular conjugation fundamentally regulates the joint organization of quantum properties. As molecules transition from highly conjugated to highly saturated states, the property manifold undergoes a distinct topological phase transition. We observe that the HOMO energy systematically decouples from the spatial-thermodynamic core, an isolation that intensifies and peaks within highly saturated regimes.

Furthermore, persistence entropy measurements uncover the robust emergence of one-dimensional cycles in the dependency network at high saturation levels, indicating a shift from hierarchical dependencies in conjugated systems to complex, multi-way relationships in saturated molecules. These topological metrics, validated through disjoint scaffold-split Wasserstein stability, are summarized in Table 1.

Ultimately, this work demonstrates that the joint structure of properties is not a static feature of the chemical space, but a dynamic manifold that rewires as delocalized π-systems are replaced by localized σ-bonds. By mapping the intrinsic geometry of these correlations, we provide a foundational perspective on how structural saturation dictates the global organization of molecular properties.

2 Methods

2.1 Data curation and preprocessing

The study utilizes the QM9 dataset, which provides quantum chemical properties for small organic molecules. To establish a robust foundation for topological analysis, the dataset was first deduplicated by SMILES strings. We isolated 12 non-redundant properties from the original 19 numeric columns. Redundant thermodynamic energies (internal energy, enthalpy, and free energy at 298 K) and their corresponding atomization energies were excluded to prevent the introduction of artificial low-dimensional structures. The defining identity gap (LUMO minus HOMO) was also removed. To manage the extreme variance of the rotational constant A, we applied a logarithmic transformation, yielding $\log(A + 1)$. Finally, nine strictly linear molecules were completely excluded from the dataset, as they represent physical outliers whose unique geometric constraints would otherwise introduce severe topological distortions into the global property manifold.

2.2 Scaffold splitting and structural stratification

To ensure rigorous validation and prevent information leakage during stability testing, a strictly disjoint Murcko scaffold split was executed on the entire dataset prior to any feature residualization. This generated two independent halves with non-overlapping molecular backbones. Both the full dataset and its split-halves were then partitioned into six discrete saturation bins based on the fraction of sp^3 hybridized carbons (F_{sp^3}).

The bins were defined to capture the gradient from highly conjugated/aromatic to highly saturated systems: $[0.0, 0.2)$, $[0.2, 0.4)$, $[0.4, 0.5)$, $[0.5, 0.6)$, $[0.6, 0.8)$, and $[0.8, 1.0]$.

2.3 Global stoichiometric residualization

Standard property correlations often conflate intrinsic quantum relationships with macroscopic scaling effects, such as molecular size and stoichiometry. To isolate the pure chemical structure effects, all 12 selected properties were residualized against a linear baseline. Regression models were trained globally on the entire dataset—rather than on individual saturation bins—to predict each property using individual atom counts (C, N, O, F, H) and the total heavy atom count (N_{heavy}) as simultaneous features. Global training is critical to this methodology, as it establishes a consistent baseline and prevents the artificial removal of structural reorganization signals across the different F_{sp^3} bins. The resulting residuals represent chemical variance strictly independent of composition and size.

2.4 Distance correlation network construction

Within each F_{sp^3} bin, as well as for each scaffold-split half, we computed complete distance correlation matrices for the 12 residualized properties. Distance correlation was selected over linear metrics (e.g., Pearson correlation) because it captures both linear and non-linear dependencies, providing a more comprehensive measure of property coupling. The 66 unique pairwise calculations per bin were parallelized to ensure computational efficiency. These correlation matrices were then inverted into symmetric distance matrices, effectively defining weighted, undirected graphs where nodes represent the property residuals and edges represent their information-theoretic distances.

2.5 Topological analysis and persistence entropy

To quantify the structural deformation of the property manifold across the saturation gradient, we applied persistent homology to the distance matrices. Using the Gudhi library, Vietoris-Rips simplicial complexes were constructed for each bin's dependency graph. We extracted 0-dimensional (0D) and 1-dimensional (1D) persistence diagrams to track the evolution of connected components and topological cycles, respectively, across all distance thresholds. From these diagrams, we calculated the Persistence Entropy (H_0 and H_1). This provided a robust, scalar measure of the network's "information density," allowing us to mathematically identify phase transitions in the complexity of property dependencies.

2.6 HOMO isolation via degree centrality

A primary objective was to determine whether the HOMO energy systematically decouples from the global spatial-thermodynamic axis. To evaluate this, we computed the network degree centrality for the HOMO residual node within each bin-specific dependency graph. The statistical isolation of the HOMO node was quantified by calculating a Z-score relative to the distribution of degree centralities of the other 11 property nodes in the same bin. Tracking this Z-score across the F_{sp^3} index allowed us to visually and quantitatively identify regimes where the HOMO energy becomes topologically isolated.

2.7 Disjoint scaffold-split Wasserstein stability

To verify that the observed topological structures and phase transitions were intrinsic to the chemical space rather than artifacts of the specific dataset sample, we validated our findings using the independent halves generated by the Murcko scaffold split. We computed the Wasserstein distance (W_1) between the persistence diagrams of the two split-halves for each saturation bin using the Python Optimal Transport (POT) library. The W_1 distances were calculated separately for the 0D and 1D persistence diagrams, independently confirming the stability of both the network connectivity and the emergent cyclic structures across structurally distinct molecular populations.

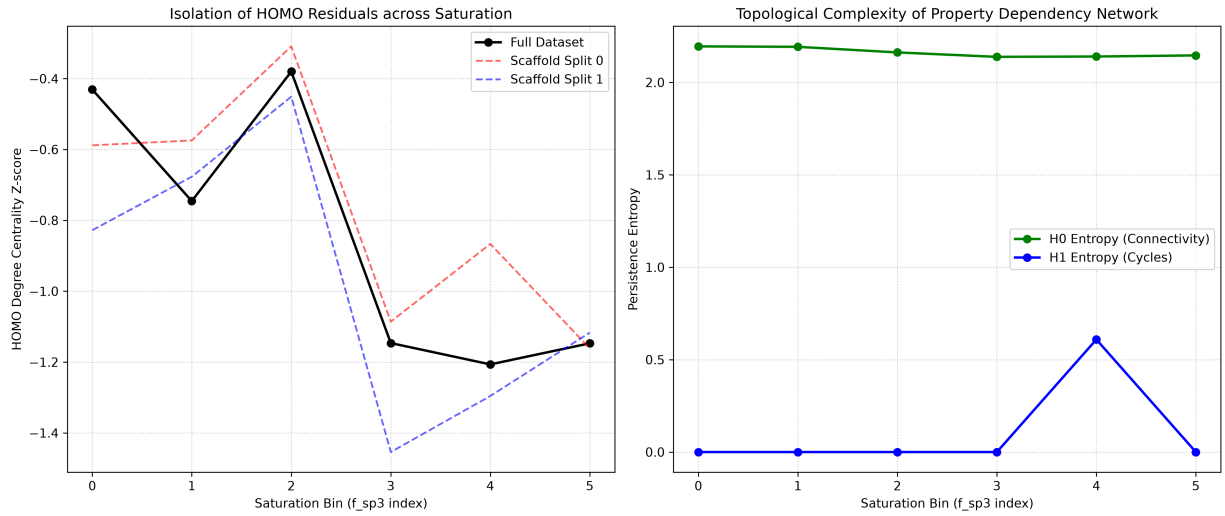

Figure 1: Evolution of the property dependency network across molecular saturation bins, quantified by the fraction of sp^3 carbons (F_{sp^3}). The left panel tracks the degree centrality Z-score of HOMO energy residuals, revealing a systematic decoupling of HOMO from the global property manifold as saturation increases, with maximum isolation occurring at bin 4. The agreement between the full dataset and disjoint Murcko scaffold splits demonstrates the structural robustness of this isolation across different molecular populations. The right panel illustrates network complexity via persistence entropy. While 0th-order connectivity entropy (H_0) remains consistently high, the 1D cycle entropy (H_1) exhibits a sharp topological phase transition at bin 4 ($F_{sp^3} \in [0.6, 0.8)$). These results demonstrate that the loss of conjugated electron systems in highly saturated molecules fundamentally rewires the property manifold, simultaneously isolating frontier orbital energies and inducing complex, non-hierarchical cyclic dependencies among the remaining properties.

3 Results

3.1 Topological evolution of the property manifold

Following the isolation of intrinsic chemical variance via stoichiometric residualization, we modeled the joint organization of the twelve non-redundant properties as weighted distance correlation networks. By tracking the persistent homology of these networks across the six saturation bins, we quantified the structural deformation of the property manifold. The persistence entropy results, detailed in Table 1, reveal a distinct shift in network complexity across the saturation gradient.

The 0th-order persistence entropy (H_0), which captures the complexity of baseline pairwise connectivity, remains consistently high across all saturation regimes, fluctuating narrowly between 2.138 and 2.195. However, the 1-dimensional persistence entropy (H_1) uncovers a topological phase transition. In highly conjugated and moderately saturated systems (Bins 0 through 3, as well as Bin 5), H_1 is effectively zero. In contrast, at Bin 4 ($F_{sp^3} \in [0.6, 0.8)$), H_1 sharply increases to 0.6089. As illustrated in the right panel of Figure 1, this transition signifies the robust emergence of non-trivial 1D cycles within the property dependency network. Chemically, this indicates that as molecules lose their conjugated electron systems, the relationships between quantum and thermodynamic properties become increasingly complex and multi-way, departing from the hierarchical dependencies observed in aromatic regimes.

3.2 Systematic decoupling of the HOMO energy

A primary objective of this study was to determine whether the HOMO energy remains integrated within the global spatial-thermodynamic manifold or undergoes systematic isolation. We evaluated this by tracking the degree centrality Z-score of the HOMO residual node relative to the other eleven properties within each bin-specific network.

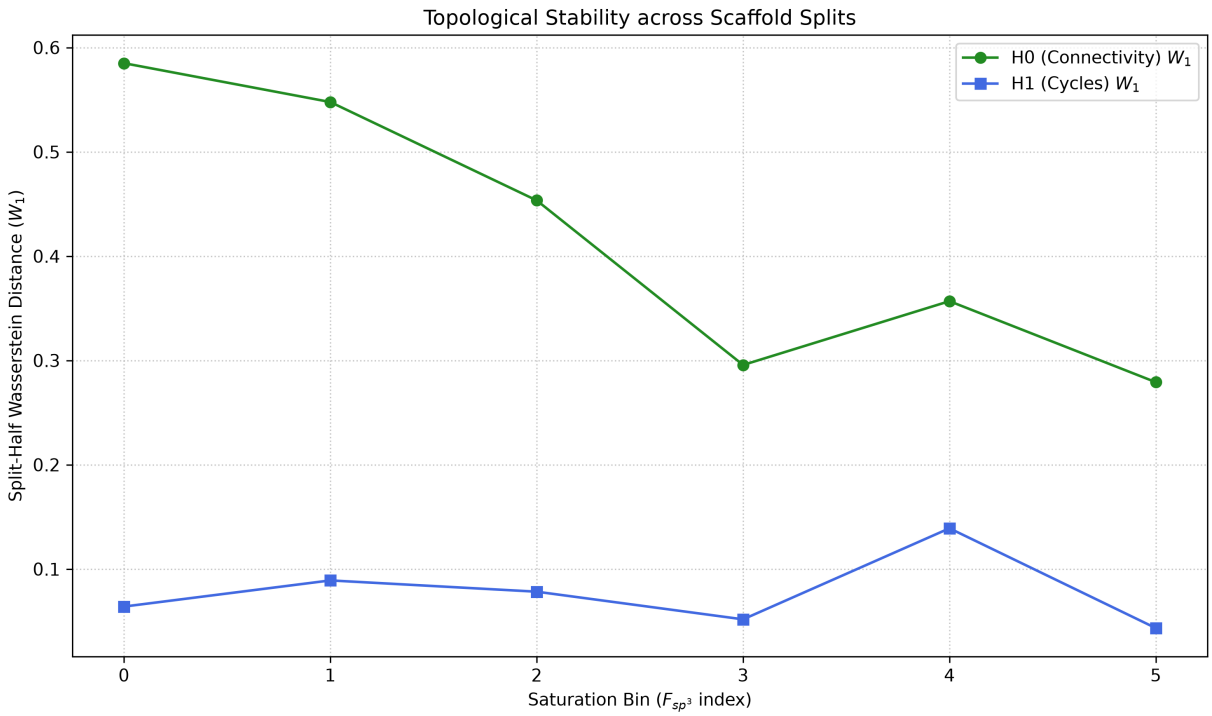

Figure 2: Split-half Wasserstein distance (W_1) evaluating the topological stability of property dependency networks across six molecular saturation bins based on the fraction of sp^3 carbons (F_{sp^3}). The distances are computed between persistence diagrams from a disjoint Murcko scaffold split for both 0D connectivity (H_0) and 1D cycles (H_1). The consistently low W_1 values for H_1 across all bins demonstrate that the emergence of cyclic property dependencies—particularly the topological phase transition observed in Bin 4—is a robust, population-wide feature rather than an artifact of specific molecular scaffolds. In contrast, the higher and more variable W_1 distances for H_0 reveal that baseline pairwise property connectivity exhibits moderate sensitivity to the inherent scaffold diversity within the dataset, altering the tightness of correlations without disrupting the overall topological phase.

Our analysis demonstrates that the HOMO energy is consistently decoupled from the global property set, with an overall mean Z-score of -0.8427. More importantly, this isolation is dynamically regulated by the degree of molecular saturation (Figure 1, left panel). In highly conjugated systems (Bin 0), the HOMO node exhibits moderate integration with a Z-score of -0.4307. As the fraction of sp^3 carbons increases, the HOMO energy systematically decouples, reaching a maximum isolation Z-score of -1.2065 in Bin 4. This trend suggests that while spatial and thermodynamic properties (such as r^2, heat capacity, and zero-point vibrational energy) maintain a tightly bound physical axis, the HOMO energy becomes governed by localized electronic features. The replacement of delocalized π-systems with localized σ-bonds effectively rewires the manifold, isolating frontier orbital energies from macroscopic molecular descriptors.

3.3 Scaffold-split stability and robustness

To ensure that the observed topological phase transition and HOMO decoupling are intrinsic features of the chemical space rather than artifacts of the specific dataset sample, we evaluated the structural stability of the networks using a disjoint Murcko scaffold split. The Wasserstein distances (W_1) between the persistence diagrams of the independent split-halves were computed for both 0D and 1D features.

As shown in Figure 2, the topological features exhibit stability under structural held-out testing. The mean W_1 distance for 1D cycles (H_1) is low at 0.0776 across all bins. Given the prominent H_1 peak of 0.6089 observed in Bin 4, this low Wasserstein distance confirms that the emergence of cyclic property dependencies

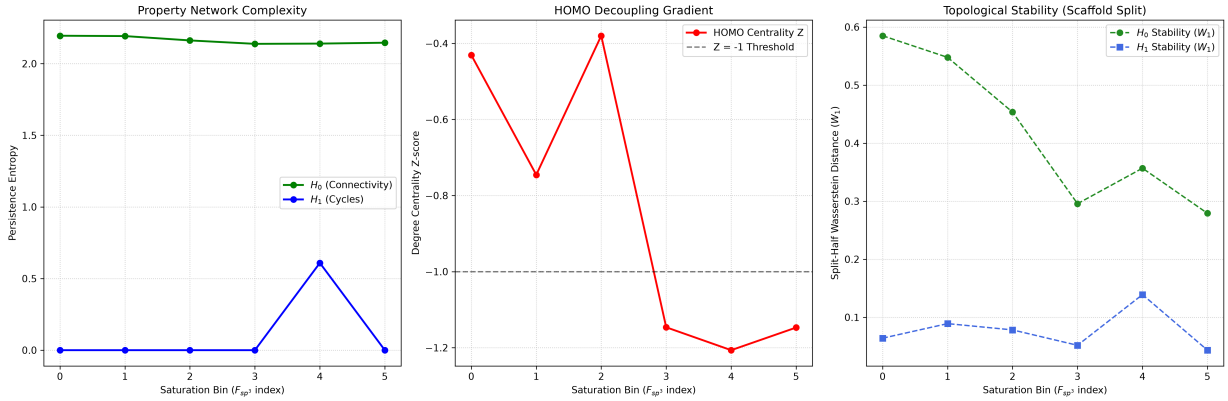

Figure 3: Evolution of property network topology and HOMO decoupling across six molecular saturation bins based on the fraction of sp^3 carbons (F_{sp^3}). The left panel displays persistence entropy, revealing a topological phase transition at Bin 4 ($F_{sp^3} \in [0.6, 0.8)$) where non-trivial 1D cycles (H_1) emerge, indicating that property relationships become increasingly complex and multi-way in highly saturated molecules while baseline connectivity (H_0) remains high. The middle panel illustrates the systematic isolation of the HOMO energy from the global property manifold; its degree centrality Z-score drops significantly as saturation increases, reaching maximum decoupling at Bin 4 as delocalized π-systems are replaced by localized σ-bonds. Finally, the right panel confirms the robustness of these findings via a disjoint Murcko scaffold split, where the low Wasserstein distance (W_1) for H_1 demonstrates that the emergence of property cycles is a stable feature of the molecular population, while the moderate variation in H_0 reflects inherent scaffold diversity.

is a robust, population-wide phenomenon rather than a scaffold-specific anomaly.

Conversely, the 0D connectivity features display moderate sensitivity to the scaffold split, with a mean W_1 distance of 0.4197. This variance reflects the inherent structural diversity of the molecular backbones in the QM9 dataset, which can subtly alter the tightness of pairwise property correlations. However, as summarized in the comprehensive overview in Figure 3, these baseline connectivity variations do not disrupt the overarching topological phase transition nor the systematic isolation of the HOMO energy. Together, these validated metrics confirm that structural saturation acts as a fundamental regulator of the global organization of molecular properties.

4 Conclusions

In this study, we introduced an information-geometric framework to investigate the joint organizational structure of quantum chemical properties. By tracking the topological evolution of property dependency networks across molecular saturation gradients, we demonstrated that the relationships between molecular properties are not static. Instead, they form a dynamic manifold that fundamentally rewires as molecules transition from highly conjugated to highly saturated states.

4.1 Summary of key findings

Our analysis yielded three primary conclusions regarding the behavior of the property manifold. First, we identified a systematic decoupling of the highest occupied molecular orbital (HOMO) energy from the global spatial-thermodynamic axis. While the HOMO energy exhibits moderate integration in aromatic systems, its isolation intensifies with increasing saturation. This decoupling reaches its peak (with a degree centrality Z-score of -1.2065) in highly saturated regimes ($F_{sp^3} \in [0.6, 0.8)$), demonstrating that frontier orbital energies become increasingly governed by localized electronic features rather than macroscopic molecular descriptors.

Second, persistence entropy measurements revealed a distinct topological phase transition that coincides with this maximum HOMO isolation. As detailed in Table 1, the robust emergence of non-trivial one-

dimensional cycles ($H_1 \approx 0.61$) at high saturation indicates a fundamental shift in network complexity. This transition signifies a departure from the hierarchical property dependencies characteristic of conjugated systems, moving toward complex, multi-way relationships in saturated molecules.

Third, the application of disjoint scaffold-split Wasserstein stability confirmed that these topological deformations are robust, population-wide phenomena. The exceptionally low Wasserstein distances for 1D cycles across structurally distinct molecular backbones validate that the observed phase transition is an intrinsic feature of the chemical space, rather than an artifact of the specific dataset sample or scaffold distribution.

4.2 Implications for molecular property modeling

These findings provide a mathematically precise characterization of how localized quantum states couple to a molecule's overarching physical organization. The replacement of delocalized π-systems with localized σ-bonds does not merely alter individual property values; it fundamentally reorganizes the dependency structure of the entire property manifold. Consequently, this work suggests that future predictive models and molecular design strategies must account for the structural regulation of property relationships. By mapping the intrinsic geometry of these correlations, we offer a foundational perspective that moves beyond isolated property prediction toward a holistic understanding of molecular property manifolds.