

Graph Coefficients in $\mathcal{N} = 4$ SYM via Tree Based Machine Learning

Contents

1	Motivation	3
2	Features	3
2.1	Feature Categories	3
3	Methodology	3
3.1	Modelling	3
3.2	Hyperparameter tuning	4
3.3	Feature considerations	4
4	Intra-loop modelling	4
5	Cross-loop modelling : predicting 12-loops	5
A	Complete Feature Descriptions	5

Abstract

1 Motivation

Using global graph invariants as features for the existence and values of coefficients.

2 Features

This section provides a comprehensive overview of all graph features used in our analysis. The features are extracted using two main tools: `fgraph_features_cli3.py`.

2.1 Feature Categories

The features are organized into the following categories:

- **Basic:** Fundamental graph properties (nodes, edges, degrees, density, clustering)
- **Connectivity:** Path-based metrics (diameter, radius, shortest paths, components)
- **Centrality:** Node importance measures (betweenness, closeness, eigenvector)
- **Core:** K-core decomposition metrics
- **Robustness:** Vulnerability measures (articulation points, bridges)
- **Cycles:** Cycle counting features
- **Spectral_Laplacian:** Laplacian matrix eigenvalues and related metrics
- **NetLSD:** Network Laplacian Spectral Descriptor features
- **Planarity:** Planar embedding properties
- **Symmetry:** Graph automorphism features
- **Community:** Community detection metrics
- **Motifs_3_4:** 3-node and 4-node motif counts
- **Motifs_5:** 5-node motif counts
- **Motifs_4:** 4-node induced subgraph counts
- **Spectral_Adjacency:** Adjacency matrix spectrum features
- **TDA:** Topological Data Analysis features (persistent homology)
- **Normalized variants:** Size-normalized versions of many features

A complete list of all 243 features with their descriptions and interpretations can be found in Appendix A.

3 Methodology

3.1 Modelling

We are looking at loop levels 5, 6, 7, 8, 9, 10, 11 and 12. We are interested in two broad modelling task with the following subtasks:

- **Intra-loop Modelling:** At each loop level order we average performance in predicting a loop level graph having seen previous graphs at the same loop order. i.e. training on loop order l and predicting on a holdout sample of loop order l .

- **Denominator graphs**
 - Predict contributing graphs (binary classification, 0/1)
- **f-graphs**
 - Predict contributing graphs (binary classification, 0/1)
 - Predict coefficient values (regression or multi-class classification)
- **Cross-loop Modelling (lower \rightarrow higher loops):** At loop level l , we use all loop order information $p \leq l$ to predict $l + 1$. We do this for $l = 11$ only.
 - **Denominator graphs**
 - Predict contributing graphs (binary classification, 0/1)
 - **f-graphs**
 - Predict contributing graphs (binary classification, 0/1)
 - Predict coefficient values (regression or multi-class classification)

Within each block we perform various subsets of the full feature set where applicable.

3.2 Hyperparameter tuning

We used bayesian optimisation.

3.3 Interpretability considerations

We use SHAP values to explain models.

3.4 Feature considerations

We used the following feature groups:

- All Features - `{all}` - 243
- Lowest 10 laplacian eigenvalues `{eig}` - 10
- Lowest 10 laplacian eigenvalues and all motifs of 3,4 and 5 vertices. `{eig, motifs}` - 84
- all motifs of 3,4 and 5 vertices. `{motifs}` - 74
- all spectral features (which include eigenvalues as a subset). `{spectral}` - 31
- eigenvalues, all motifs of 3,4 and 5 vertices and centrality measures. - 98 `{eig, motifs, centrality}`

We are very interested in the laplacian eigenvalues which are standard permutation invariants of graph problems as well as motifs/graphlets. The centrality measure were added as these guaranteed uniqueness from our chosen dataset.

4 Intra-loop modelling

Using 5-fold cross validation.

¹Our key observation from table 1 is that after considering all possible features, the next most performant feature space is `{eig, motifs, centrality}`. We also observe that the `{motifs}` feature space starts to become more relevant as the loop level increases.

¹For final presentation - we probably do not want to show all columns at all, instead qualitatively argue all other columns types - also need to argue for 10 eigenvalues.

Loop	{eig}	{spectral}	{motifs}	{eig, motifs}	{eig, motifs, centrality}	{all}
6	0.7500	0.7292	0.5744	0.6607	0.4702	0.6905
7	0.7163	0.8078	0.7464	0.7768	0.7859	0.8590
8	0.8174	0.8469	0.8081	0.8525	0.8785	0.9064
9	0.8555	0.8839	0.8622	0.8998	0.9252	0.9456
10	0.8714	0.8990	0.8842	0.9169	0.9452	—
11	0.8827	0.9149	0.8765	0.9149	0.9478	—

Table 1: AUC scores across feature column sets and loop orders. Best value per loop in bold; second-best highlighted in red. We did not bother persuing all columns for 10 and 11 loops as these took quite some time.

5 Cross-loop modelling : predicting 12-loops

A Complete Feature Descriptions

This appendix provides a comprehensive list of all graph features with their descriptions and interpretations. Due to the large number of features, the table is split across multiple pages for readability.

Feature Name	Group	Description	Interpretation
Basic_num_nodes	Basic	Total number of nodes in the graph	The larger this number the bigger the graph is
Basic_num_edges	Basic	Total number of edges in the graph	The larger this number the more connected the graph is
Basic_min_degree	Basic	Minimum degree among all nodes	The larger this number the more connected the least connected node is
Basic_max_degree	Basic	Maximum degree among all nodes	The larger this number the more connected the most connected node is
Basic_avg_degree	Basic	Average degree across all nodes	The larger this number the more connected the graph is on average
Basic_degree_std	Basic	Standard deviation of node degrees	The larger this number the more unequal the node connections are
Basic_degree_skew	Basic	Skewness of degree distribution	Positive values mean more high-degree nodes; negative means more low-degree nodes
Basic_density	Basic	Graph density (edges/max_possible_edges)	The larger this number the more densely connected the graph is
Basic_edge_to_node_ratio	Basic	Ratio of edges to nodes	The larger this number the more edges per node the graph has
Basic_degree_entropy	Basic	Shannon entropy of degree distribution	The larger this number the more diverse the node degrees are
Assortativity_degree	Basic	Degree assortativity coefficient	Positive values mean similar-degree nodes connect; negative means opposite-degree nodes connect
Clustering_mean	Basic	Average local clustering coefficient	The larger this number the more clustered/triangular the graph is
Clustering_q10	Basic	10th percentile of clustering coefficients	The larger this number the more clustered the least clustered nodes are
Clustering_q50	Basic	50th percentile (median) of clustering coefficients	The larger this number the more clustered the typical node is
Clustering_q90	Basic	90th percentile of clustering coefficients	The larger this number the more clustered the most clustered nodes are
Clustering_frac_zero	Basic	Fraction of nodes with zero clustering	The larger this number the more tree-like the graph is
Clustering_frac_one	Basic	Fraction of nodes with clustering = 1	The larger this number the more clique-like the graph is
Degree_gini	Basic	Gini coefficient of degree distribution	The larger this number the more unequal the node degrees are
Basic_avg_degree_norm	Basic_Normalized	Average degree normalized by graph size	The larger this number the more connected the graph is relative to its size
Basic_degree_entropy_norm	Basic_Normalized	Degree entropy normalized by maximum possible	The larger this number the more diverse the node degrees are relative to maximum diversity
COEFFICIENTS	Meta	Optional coefficient or label column carried from input	Not a structural graph feature; typically used to store an external coefficient or meta-data for the graph
Unnamed: 0	Meta	Optional index/ID column carried from input	Not a structural graph feature; preserves the original row/index identifier from the input CSV

Feature Name	Group	Description	Interpretation
Connectivity_is_connected	Connectivity	Whether graph is connected (True/False)	True means all nodes can reach each other; False means graph is fragmented
Connectivity_num_components	Connectivity	Number of connected components	The larger this number the more fragmented the graph is
Connectivity_diameter	Connectivity	Graph diameter (longest shortest path)	The larger this number the more spread out the graph is
Connectivity_radius	Connectivity	Graph radius (minimum eccentricity)	The larger this number the more spread out the graph is
Connectivity_avg_shortest_path_length	Connectivity	Average shortest path length	The larger this number the more spread out the graph is
Connectivity_wiener_index	Connectivity	Sum of all shortest path lengths	The larger this number the more spread out the graph is
Eff_diameter_p90	Connectivity	90th percentile effective diameter	The larger this number the more spread out the graph is
Ecc_mean	Connectivity	Mean eccentricity of nodes	The larger this number the more spread out the graph is
Ecc_q90	Connectivity	90th percentile eccentricity	The larger this number the more spread out the graph is
Connectivity_diameter_norm	Connectivity_Normalized	Diameter normalized by graph size	The larger this number the more spread out the graph is relative to its size
Connectivity_radius_norm	Connectivity_Normalized	Radius normalized by graph size	The larger this number the more spread out the graph is relative to its size
Connectivity_num_components_per_node	Connectivity_Normalized	Components per node	The larger this number the more fragmented the graph is per node
Wiener_mean_distance	Connectivity_Normalized	Mean distance normalized by Wiener index	The larger this number the more spread out the graph is relative to total distance
Centrality_betweenness_mean	Centrality	Mean betweenness centrality	The larger this number the more nodes act as bridges/connectors
Centrality_betweenness_max	Centrality	Maximum betweenness centrality	The larger this number the more important the most central node is
Centrality_betweenness_std	Centrality	Standard deviation of betweenness centrality	The larger this number the more unequal the node importance is
Centrality_betweenness_skew	Centrality	Skewness of betweenness centrality distribution	Positive values mean few very important nodes; negative means many moderately important nodes
Centrality_closeness_mean	Centrality	Mean closeness centrality	The larger this number the more centrally located nodes are on average
Centrality_closeness_max	Centrality	Maximum closeness centrality	The larger this number the more centrally located the most central node is
Centrality_closeness_std	Centrality	Standard deviation of closeness centrality	The larger this number the more unequal the node centrality is
Centrality_closeness_skew	Centrality	Skewness of closeness centrality distribution	Positive values mean few very central nodes; negative means many moderately central nodes
Centrality_eigenvector_mean	Centrality	Mean eigenvector centrality	The larger this number the more nodes are connected to important nodes
Centrality_eigenvector_max	Centrality	Maximum eigenvector centrality	The larger this number the more important the most influential node is
Centrality_eigenvector_std	Centrality	Standard deviation of eigenvector centrality	The larger this number the more unequal the node influence is
Centrality_eigenvector_skew	Centrality	Skewness of eigenvector centrality distribution	Positive values mean few very influential nodes; negative means many moderately influential nodes
Centrality_closeness_mean_norm	Centrality_Normalized	Mean closeness normalized by maximum	The larger this number the more centrally located nodes are on average relative to maximum
Centrality_closeness_max_norm	Centrality_Normalized	Max closeness normalized by maximum	The larger this number the more centrally located the most central node is relative to maximum

Feature Name	Group	Description	Interpretation
Core_max_core_index	Core	Maximum k -core index	The larger this number the more tightly connected the densest core is
Core_core_index_mean	Core	Mean k -core index	The larger this number the more tightly connected nodes are on average
Robust_articulation_points	Robustness	Number of articulation points (cut vertices)	The larger this number the more vulnerable the graph is to fragmentation
Robust_bridge_count	Robustness	Number of bridges (cut edges)	The larger this number the more vulnerable the graph is to fragmentation
Robust_articulation_points_per_node	Robustness_Normalized	Articulation points per node	The larger this number the more vulnerable the graph is to fragmentation per node
Robust_bridge_count_per_edge	Robustness_Normalized	Bridges per edge	The larger this number the more vulnerable the graph is to fragmentation per edge
Cycle_num_cycles_len_5	Cycles	Number of cycles of length 5	The larger this number the more 5-cycles the graph contains
Cycle_num_cycles_len_6	Cycles	Number of cycles of length 6	The larger this number the more 6-cycles the graph contains
Spectral_algebraic_connectivity	Spectral_Laplacian	Second smallest Laplacian eigenvalue (Fiedler value)	The larger this number the more connected the graph is
Spectral_spectral_gap	Spectral_Laplacian	Difference between first two Laplacian eigenvalues	The larger this number the more well-connected the graph is
Spectral_laplacian_mean	Spectral_Laplacian	Mean of Laplacian eigenvalues	The larger this number the more connected the graph is on average
Spectral_laplacian_std	Spectral_Laplacian	Standard deviation of Laplacian eigenvalues	The larger this number the more varied the connectivity patterns are
Spectral_laplacian_skew	Spectral_Laplacian	Skewness of Laplacian eigenvalue distribution	Positive values mean few highly connected components; negative means many moderately connected components
Spectral_lap_eig_0	Spectral_Laplacian	Smallest Laplacian eigenvalue	Always 0 for connected graphs; larger values indicate more disconnected components
Spectral_lap_eig_1	Spectral_Laplacian	Second smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_2	Spectral_Laplacian	Third smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_3	Spectral_Laplacian	Fourth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_4	Spectral_Laplacian	Fifth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_5	Spectral_Laplacian	Sixth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_6	Spectral_Laplacian	Seventh smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_7	Spectral_Laplacian	Eighth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_8	Spectral_Laplacian	Ninth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Spectral_lap_eig_9	Spectral_Laplacian	Tenth smallest Laplacian eigenvalue	The larger this number the more connected the graph is
Kirchhoff_index	Spectral_Laplacian	Kirchhoff index (sum of resistance distances)	The larger this number the more spread out the graph is
Spectral_kirchhoff_index	Spectral_Laplacian	Kirchhoff index (alternative name)	The larger this number the more spread out the graph is

Feature Name	Group	Description	Interpretation
Spectral_laplacian_heat_trace_t0.1	Spectral_Laplacian	Laplacian heat trace at $t=0.1$	The larger this number the more heat spreads quickly through the graph
Spectral_laplacian_heat_trace_t0.5	Spectral_Laplacian	Laplacian heat trace at $t=0.5$	The larger this number the more heat spreads through the graph
Spectral_laplacian_heat_trace_t1.0	Spectral_Laplacian	Laplacian heat trace at $t=1.0$	The larger this number the more heat spreads through the graph
Spectral_laplacian_heat_trace_t2.0	Spectral_Laplacian	Laplacian heat trace at $t=2.0$	The larger this number the more heat spreads through the graph
Spectral_laplacian_heat_trace_t5.0	Spectral_Laplacian	Laplacian heat trace at $t=5.0$	The larger this number the more heat spreads through the graph
Spectral_laplacian_heat_trace_t0.1_per_node	Spectral_Normalized	Heat trace $t=0.1$ per node	The larger this number the more heat spreads quickly per node
Spectral_laplacian_heat_trace_t1.0_per_node	Spectral_Normalized	Heat trace $t=1.0$ per node	The larger this number the more heat spreads per node
Spectral_laplacian_heat_trace_t5.0_per_node	Spectral_Normalized	Heat trace $t=5.0$ per node	The larger this number the more heat spreads per node
Spectral_algebraic_connectivity	Spectral_Average	Algebraic connectivity over average degree	The larger this number the more connected the graph is relative to its average connectivity
Spectral_spectral_gap_rel	Spectral_Normalized	Relative spectral gap	The larger this number the more well-connected the graph is relative to its connectivity
NetLSD_mean	NetLSD	Mean NetLSD signature	The larger this number the more complex the graph structure is
NetLSD_std	NetLSD	Standard deviation of NetLSD signature	The larger this number the more varied the graph structure is
NetLSD_q10	NetLSD	10th percentile of NetLSD signature	The larger this number the more complex the simplest parts are
NetLSD_q90	NetLSD	90th percentile of NetLSD signature	The larger this number the more complex the most complex parts are
Planarity_num_faces	Planarity	Number of faces in planar embedding	The larger this number the more complex the planar structure is
Planarity_face_size_mean	Planarity	Mean face size in planar embedding	The larger this number the larger the typical face is
Planarity_face_size_max	Planarity	Maximum face size in planar embedding	The larger this number the larger the biggest face is
Planarity_num_faces_over_upper_bound	Planarity_Normalized	Faces over theoretical upper bound	The larger this number the more complex the planar structure is relative to maximum possible
Planarity_face_size_mean_norm	Planarity_Normalized	Mean face size normalized	The larger this number the larger the typical face is relative to maximum possible
Symmetry_automorphism_group_order	Symmetry	Order of automorphism group	The larger this number the more symmetric the graph is
Symmetry_num_orbits	Symmetry	Number of node orbits under automorphisms	The larger this number the more diverse the node roles are
Symmetry_orbit_size_max	Symmetry	Maximum orbit size	The larger this number the more nodes share the same role
Symmetry_aut_size_log_over_log_n	Symmetry_Normalized	Log automorphism size over $\log n!$	The larger this number the more symmetric the graph is relative to maximum possible symmetry
Symmetry_num_orbits_per_node	Symmetry_Normalized	Orbits per node	The larger this number the more diverse the node roles are per node
Symmetry_orbit_size_max_per_node	Symmetry_Normalized	Max orbit size per node	The larger this number the more nodes share the same role per node

Feature Name	Group	Description	Interpretation
Comm_modularity	Community	Modularity of best community partition	The larger this number the more clearly separated the communities are
Comm_count	Community	Number of communities found	The larger this number the more fragmented the graph is
Comm_size_max	Community	Size of largest community	The larger this number the more dominant the largest community is
Comm_size_gini	Community	Gini coefficient of community sizes	The larger this number the more unequal the community sizes are
Comm_internal_edge_frac	Community	Fraction of edges within communities	The larger this number the more internally connected communities are
Motif_triangles	Motifs_3.4	Number of triangles (3-cliques)	The larger this number the more triangular structures the graph has
Motif_wedges	Motifs_3.4	Number of wedges (2-paths)	The larger this number the more path-like structures the graph has
Motif_4_cycles	Motifs_3.4	Number of 4-cycles	The larger this number the more square-like structures the graph has
Motif_4_cliques	Motifs_3.4	Number of 4-cliques (K_4)	The larger this number the more tightly connected 4-node groups the graph has
Motif_triangle_edge_incidence_mean	Motifs_3.4	Mean triangles per edge	The larger this number the more triangles each edge participates in
Motif_triangle_edge_incidence_std	Motifs_3.4	Standard deviation of triangles per edge	The larger this number the more varied edge participation in triangles is
Motif_square_clustering_proxy	Motifs_3.4	Tendency to form 4-cycles relative to 2-paths	The larger this number the more square-like the graph structure is
Motif_triangle_edge_incidence_median	Motifs_3.4	Median triangles per edge	The larger this number the more triangles the typical edge participates in
Motif_triangle_edge_incidence_q90	Motifs_3.4	90th percentile triangles per edge	The larger this number the more triangles the most triangular edges participate in
Motif_triangle_edge_frac_zero	Motifs_3.4	Fraction of edges with zero triangles	The larger this number the more tree-like the graph is
Motif_triangle_edge_frac_ge2	Motifs_3.4	Fraction of edges with ≥ 2 triangles	The larger this number the more clustered the graph is
Motif_induced_K1_3	Motifs_3.4	Number of induced $K_{1,3}$ (star) subgraphs	The larger this number the more star-like structures the graph has
Motif_induced_P4	Motifs_3.4	Number of induced P_4 (path) subgraphs	The larger this number the more path-like structures the graph has
Motif_induced_C4	Motifs_3.4	Number of induced C_4 (cycle) subgraphs	The larger this number the more cycle-like structures the graph has
Motif_induced_TailedTriangle	Motifs_3.4	Number of induced tailed triangle subgraphs	The larger this number the more tailed triangle structures the graph has
Motif_induced_Diamond	Motifs_3.4	Number of induced diamond subgraphs	The larger this number the more diamond structures the graph has
Motif_induced_K4	Motifs_3.4	Number of induced K_4 (clique) subgraphs	The larger this number the more tightly connected 4-node groups the graph has
Motif_induced_connected_per_4sets	Motifs_3.4	Fraction of 4-node subsets that are connected	The larger this number the more connected 4-node groups are

Feature Name	Group	Description	Interpretation
Motif_triangles_per_Cn3	Motifs_3.4.Normalized	Triangles normalized by $C(n,3)$	The larger this number the more triangular the graph is relative to maximum possible
Motif_4_cycles_per_Cn4	Motifs_3.4.Normalized	4-cycles normalized by $C(n,4)$	The larger this number the more square-like the graph is relative to maximum possible
Motif_4_cliques_per_Cn4	Motifs_3.4.Normalized	4-cliques normalized by $C(n,4)$	The larger this number the more tightly connected 4-node groups are relative to maximum possible
Motif_wedges_per_max	Motifs_3.4.Normalized	Wedges normalized by theoretical maximum	The larger this number the more path-like the graph is relative to maximum possible
Motif_induced_K1.3_per_Cn4	Motifs_3.4.Normalized	$K_{1,3}$ normalized by $C(n,4)$	The larger this number the more star-like the graph is relative to maximum possible
Motif_induced_P4_per_Cn4	Motifs_3.4.Normalized	P_4 normalized by $C(n,4)$	The larger this number the more path-like the graph is relative to maximum possible
Motif_induced_C4_per_Cn4	Motifs_3.4.Normalized	C_4 normalized by $C(n,4)$	The larger this number the more cycle-like the graph is relative to maximum possible
Motif_induced_TailedTriangle_per_Cn4	Motifs_3.4.Normalized	Tailed triangle normalized by $C(n,4)$	The larger this number the more tailed triangle structures are relative to maximum possible
Motif_induced_Diamond_per_Cn4	Motifs_3.4.Normalized	Diamond normalized by $C(n,4)$	The larger this number the more diamond structures are relative to maximum possible
Motif_induced_K4_per_Cn4	Motifs_3.4.Normalized	K_4 normalized by $C(n,4)$	The larger this number the more tightly connected 4-node groups are relative to maximum possible
Motif_5_cycles	Motifs_5	Number of 5-cycles	The larger this number the more 5-sided cycle structures the graph has
Motif_5_cliques	Motifs_5	Number of 5-cliques (K_5)	The larger this number the more tightly connected 5-node groups the graph has
Motif_5_cycles_per_Cn5	Motifs_5.Normalized	5-cycles normalized by $C(n,5)$	The larger this number the more 5-sided cycle structures are relative to maximum possible
Motif_5_cliques_per_Cn5	Motifs_5.Normalized	5-cliques normalized by $C(n,5)$	The larger this number the more tightly connected 5-node groups are relative to maximum possible
Motif_5_cycles_per_Kn	Motifs_5.Normalized	5-cycles normalized by complete graph	The larger this number the more 5-sided cycle structures are relative to complete graph
Motif_induced5_g_0.5	Motifs_5	Number of induced 5-node graphlet g_0	The larger this number the more g_0 structures the graph has
Motif_induced5_g_1.5	Motifs_5	Number of induced 5-node graphlet g_1	The larger this number the more g_1 structures the graph has
Motif_induced5_g_2.5	Motifs_5	Number of induced 5-node graphlet g_2	The larger this number the more g_2 structures the graph has
Motif_induced5_g_3.5	Motifs_5	Number of induced 5-node graphlet g_3	The larger this number the more g_3 structures the graph has
Motif_induced5_g_4.5	Motifs_5	Number of induced 5-node graphlet g_4	The larger this number the more g_4 structures the graph has
Motif_induced5_g_5.5	Motifs_5	Number of induced 5-node graphlet g_5	The larger this number the more g_5 structures the graph has
Motif_induced5_g_6.5	Motifs_5	Number of induced 5-node graphlet g_6	The larger this number the more g_6 structures the graph has
Motif_induced5_g_7.5	Motifs_5	Number of induced 5-node graphlet g_7	The larger this number the more g_7 structures the graph has
Motif_induced5_g_8.5	Motifs_5	Number of induced 5-node graphlet g_8	The larger this number the more g_8 structures the graph has
Motif_induced5_g_9.5	Motifs_5	Number of induced 5-node graphlet g_9	The larger this number the more g_9 structures the graph has

Feature Name	Group	Description	Interpretation
Motif_induced_g_1_4	Motifs_4	Number of induced Path ₄ (P_4) subgraphs	The larger this number the more path-like 4-node structures the graph has
Motif_induced_g_2_4	Motifs_4	Number of induced Star ₄ ($K_{1,3}$) subgraphs	The larger this number the more star-like 4-node structures the graph has
Motif_induced_g_3_4	Motifs_4	Number of induced Cycle ₄ (C_4) subgraphs	The larger this number the more cycle-like 4-node structures the graph has
Motif_induced_g_4_4	Motifs_4	Number of induced TailedTriangle subgraphs	The larger this number the more tailed triangle 4-node structures the graph has
Motif_induced_g_5_4	Motifs_4	Number of induced Diamond subgraphs	The larger this number the more diamond 4-node structures the graph has
Motif_induced_g_6_4	Motifs_4	Number of induced Clique ₄ (K_4) subgraphs	The larger this number the more tightly connected 4-node groups the graph has
Motif_induced_g_1_4_per_Cn4	Motifs_4.Normalized	Path ₄ normalized by $C(n,4)$	The larger this number the more path-like 4-node structures are relative to maximum possible
Motif_induced_g_2_4_per_Cn4	Motifs_4.Normalized	Star ₄ normalized by $C(n,4)$	The larger this number the more star-like 4-node structures are relative to maximum possible
Motif_induced_g_3_4_per_Cn4	Motifs_4.Normalized	Cycle ₄ normalized by $C(n,4)$	The larger this number the more cycle-like 4-node structures are relative to maximum possible
Motif_induced_g_4_4_per_Cn4	Motifs_4.Normalized	TailedTriangle normalized by $C(n,4)$	The larger this number the more tailed triangle 4-node structures are relative to maximum possible
Motif_induced_g_5_4_per_Cn4	Motifs_4.Normalized	Diamond normalized by $C(n,4)$	The larger this number the more diamond 4-node structures are relative to maximum possible
Motif_induced_g_6_4_per_Cn4	Motifs_4.Normalized	Clique ₄ normalized by $C(n,4)$	The larger this number the more tightly connected 4-node groups are relative to maximum possible
Adjacency_energy	Spectral_Adjacency	Sum of absolute eigenvalues of adjacency matrix	The larger this number the more energetic/vibrant the graph is
Adjacency_estrada_index	Spectral_Adjacency	Sum of exponentials of eigenvalues	The larger this number the more communicable the graph is
Adjacency_moment_2	Spectral_Adjacency	Second moment of adjacency eigenvalues	The larger this number the more spread out the adjacency spectrum is
Adjacency_moment_3	Spectral_Adjacency	Third moment of adjacency eigenvalues	Positive values mean more high-frequency components; negative means more low-frequency components
Adjacency_moment_4	Spectral_Adjacency	Fourth moment of adjacency eigenvalues	The larger this number the more peaked the adjacency spectrum is
Adjacency_energy_per_node	Spectral_Normalized	Adjacency energy per node	The larger this number the more energetic/vibrant the graph is per node
Adjacency_energy_over_fro	Spectral_Normalized	Adjacency energy over Frobenius norm	The larger this number the more energetic the graph is relative to its total energy
Adjacency_estrada_per_node	Spectral_Normalized	Estrada index per node	The larger this number the more communicable the graph is per node
log_Adjacency_estrada_per_node	Spectral_Normalized	Log Estrada index per node	The larger this number the more communicable the graph is per node (log scale)
Adjacency_moment_2_over_avgdeg	Spectral_Normalized	Second moment over average degree	The larger this number the more spread out the adjacency spectrum is relative to average connectivity
Adjacency_moment_3_over_avgdeg3	Spectral_Normalized	Third moment over average degree cubed	The larger this number the more high-frequency components are relative to connectivity cubed
Adjacency_moment_4_over_avgdeg4	Spectral_Normalized	Fourth moment over average degree to fourth	The larger this number the more peaked the adjacency spectrum is relative to connectivity to fourth power

Feature Name	Group	Description	Interpretation
Spectral_adjacency_energy	Spectral_Adjacenc	Adjacency energy (alternative name)	The larger this number the more energetic/vibrant the graph is
Spectral_adjacency_estrada_index	Spectral_Adjacenc	Adjacency Estrada index (alternative name)	The larger this number the more communicable the graph is
Spectral_adjacency_moment_2	Spectral_Adjacenc	Adjacency second moment (alternative name)	The larger this number the more spread out the adjacency spectrum is
Spectral_adjacency_moment_3	Spectral_Adjacenc	Adjacency third moment (alternative name)	Positive values mean more high-frequency components; negative means more low-frequency components
Spectral_adjacency_moment_4	Spectral_Adjacenc	Adjacency fourth moment (alternative name)	The larger this number the more peaked the adjacency spectrum is
TDA_H0_count	TDA	Number of H0 homology features (connected components)	The larger this number the more disconnected components the graph has
TDA_H0_total_persistence	TDA	Total persistence of H0 features	The larger this number the more persistent the connectivity structure is
TDA_H0_mean_persistence	TDA	Mean persistence of H0 features	The larger this number the more stable the connectivity structure is
TDA_H0_max_persistence	TDA	Maximum persistence of H0 features	The larger this number the more stable the most persistent component is
TDA_H0_persistence_entropy	TDA	Entropy of H0 persistence distribution	The larger this number the more diverse the persistence values are
TDA_H0_mean_birth	TDA	Mean birth time of H0 features	The larger this number the later components typically appear
TDA_H0_mean_death	TDA	Mean death time of H0 features	The larger this number the later components typically disappear
TDA_H1_count	TDA	Number of H1 homology features (cycles)	The larger this number the more cyclic structures the graph has
TDA_H1_total_persistence	TDA	Total persistence of H1 features	The larger this number the more persistent the cyclic structure is
TDA_H1_mean_persistence	TDA	Mean persistence of H1 features	The larger this number the more stable the cyclic structure is
TDA_H1_max_persistence	TDA	Maximum persistence of H1 features	The larger this number the more stable the most persistent cycle is
TDA_H1_persistence_entropy	TDA	Entropy of H1 persistence distribution	The larger this number the more diverse the cycle persistence values are
TDA_H1_mean_birth	TDA	Mean birth time of H1 features	The larger this number the later cycles typically appear
TDA_H1_mean_death	TDA	Mean death time of H1 features	The larger this number the later cycles typically disappear
TDA_Betti0_at_q25	TDA	Betti number β_0 at 25th percentile filtration	The larger this number the more components exist at low filtration levels
TDA_Betti0_at_q50	TDA	Betti number β_0 at 50th percentile filtration	The larger this number the more components exist at medium filtration levels
TDA_Betti0_at_q75	TDA	Betti number β_0 at 75th percentile filtration	The larger this number the more components exist at high filtration levels
TDA_Betti1_at_q25	TDA	Betti number β_1 at 25th percentile filtration	The larger this number the more cycles exist at low filtration levels
TDA_Betti1_at_q50	TDA	Betti number β_1 at 50th percentile filtration	The larger this number the more cycles exist at medium filtration levels
TDA_Betti1_at_q75	TDA	Betti number β_1 at 75th percentile filtration	The larger this number the more cycles exist at high filtration levels

Feature Name	Group	Description	Interpretation
TDA_H0_count_per_node	TDA_Normalized	$H0$ features per node	The larger this number the more disconnected components exist per node
TDA_H0_total_persistence_over_diameter	TDA_Normalized	$H0$ persistence over diameter	The larger this number the more persistent the connectivity structure is relative to graph spread
TDA_H0_mean_persistence_over_diameter	TDA_Normalized	$H0$ mean persistence over diameter	The larger this number the more stable the connectivity structure is relative to graph spread
TDA_H0_max_persistence_over_diameter	TDA_Normalized	$H0$ max persistence over diameter	The larger this number the more stable the most persistent component is relative to graph spread
TDA_H0_mean_birth_over_diameter	TDA_Normalized	$H0$ mean birth over diameter	The larger this number the later components typically appear relative to graph spread
TDA_H0_mean_death_over_diameter	TDA_Normalized	$H0$ mean death over diameter	The larger this number the later components typically disappear relative to graph spread
TDA_H1_count_per_node	TDA_Normalized	$H1$ features per node	The larger this number the more cyclic structures exist per node
TDA_H1_total_persistence_over_diameter	TDA_Normalized	$H1$ persistence over diameter	The larger this number the more persistent the cyclic structure is relative to graph spread
TDA_H1_mean_persistence_over_diameter	TDA_Normalized	$H1$ mean persistence over diameter	The larger this number the more stable the cyclic structure is relative to graph spread
TDA_H1_max_persistence_over_diameter	TDA_Normalized	$H1$ max persistence over diameter	The larger this number the more stable the most persistent cycle is relative to graph spread
TDA_H1_mean_birth_over_diameter	TDA_Normalized	$H1$ mean birth over diameter	The larger this number the later cycles typically appear relative to graph spread
TDA_H1_mean_death_over_diameter	TDA_Normalized	$H1$ mean death over diameter	The larger this number the later cycles typically disappear relative to graph spread
TDA_Betti0_at_q25_per_node	TDA_Normalized	Betti0 at $q25$ per node	The larger this number the more components exist at low filtration levels per node
TDA_Betti0_at_q50_per_node	TDA_Normalized	Betti0 at $q50$ per node	The larger this number the more components exist at medium filtration levels per node
TDA_Betti0_at_q75_per_node	TDA_Normalized	Betti0 at $q75$ per node	The larger this number the more components exist at high filtration levels per node
TDA_Betti1_at_q25_per_node	TDA_Normalized	Betti1 at $q25$ per node	The larger this number the more cycles exist at low filtration levels per node
TDA_Betti1_at_q50_per_node	TDA_Normalized	Betti1 at $q50$ per node	The larger this number the more cycles exist at medium filtration levels per node
TDA_Betti1_at_q75_per_node	TDA_Normalized	Betti1 at $q75$ per node	The larger this number the more cycles exist at high filtration levels per node