Draft version June 13, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

# Dynamic Weighted Peptide Network Analysis for Characterizing and Predicting Aggregate Stability

AstroPilot[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

## ABSTRACT

Peptide self-assembly is a complex dynamic process, and characterizing and predicting aggregate stability and transitions remain significant challenges often limited by traditional coarse-grained or binary metrics. We address this by representing the peptide system as a dynamic weighted graph where nodes are peptides and edges quantify inter-peptide noncovalent contacts, weighted by type (hydrophobic, aromatic, hydrogen bonds). We analyze the temporal evolution of this network using graph theoretical metrics. Using molecular dynamics simulations of KYFIL pentapeptides, we studied aggregate behavior from 100 ns onwards by constructing dynamic weighted and binary graphs and calculating metrics including weighted graph Laplacian spectral properties (Fiedler value), global properties (density, connected components, largest connected component or LCC size). We correlated these graph metrics with LCC physical properties such as radius of gyration and packing score, and compared results to binary graph analysis. Our analysis reveals significant dynamic fluctuations in aggregate structure and size. Weighted graph metrics, particularly the LCC Fiedler value and density, demonstrate greater sensitivity to interaction strengths compared to their binary counterparts. Both weighted and binary graph metrics correlate significantly with LCC physical properties, indicating that the network structure effectively captures aggregate compactness. System-level analysis confirms the presence of multiple dynamic clusters. A combined graph-based order parameter for the LCC was developed, showing potential for tracking aggregate state transitions. This dynamic weighted graph analysis provides a robust quantitative framework for characterizing peptide aggregates and identifies promising metrics that can serve as sensitive indicators and potential predictive order parameters for aggregate stability and fragmentation.

*Keywords:* Molecular physics, Chemical kinetics, Chemical reaction network models, Computational methods, Theoretical techniques

## 1. INTRODUCTION

Peptide self-assembly, the spontaneous association of peptides into ordered supramolecular structures, is a ubiquitous phenomenon central to numerous biological processes and holds immense promise for applications in biotechnology and materials science. Driven by a complex interplay of noncovalent forces, including hydrophobic interactions, aromatic $\pi - \pi$ stacking, and hydrogen bonding, peptides can form diverse aggregates ranging from transient clusters to highly ordered amyloid fibrils. A deep understanding of the mechanisms governing peptide assembly pathways, as well as the factors determining the stability and dynamic behavior of the resulting aggregates, is essential for both rational design in nanotechnology and therapeutic intervention in protein misfolding diseases.

Despite significant progress facilitated by experimental techniques and molecular simulations, quantitatively characterizing and predicting the stability and dynamic transitions within peptide aggregates remains a formidable challenge. Existing analytical methods often rely on simplified representations, such as coarse-grained models that abstract away atomic details, or utilize binary metrics that merely indicate the presence or absence of an interaction without accounting for its type or strength. These approaches struggle to capture the dynamic, multi-scale nature of the assembly process and the nuanced contributions of different noncovalent forces. Consequently, they may overlook subtle structural rearrangements and interaction network reconfigurations that are critical determinants of aggregate stability, growth, fragmentation, or dissolution. There is a clear need for a more sensitive, quantitative frame-

work that can effectively probe the dynamic network of specific inter-peptide interactions.

To address these limitations, we propose a novel approach that represents the peptide self-assembly system as a dynamic weighted graph. In this framework, each peptide molecule is treated as a node, and the interactions between peptides are represented by edges. Crucially, these edges are weighted based on a quantitative measure of the inter-peptide noncovalent contacts. Specifically, the weight of an edge between two peptides is a function of the number and type of contacts they share, such as hydrophobic, aromatic, and hydrogen bonds, allowing us to differentiate the contributions of various interaction types and their collective strength. By constructing these graphs dynamically from molecular dynamics simulation trajectories, our method captures the temporal evolution of the peptide interaction network, providing a rich description of the system's state over time.

We leverage tools from graph theory to analyze the structural and connectivity properties of this dynamic weighted network (Yang & Yu 2023; Strey et al. 2024). We focus on key graph metrics, including global properties such as graph density, the number of connected components, and the size of the largest connected component (LCC), which provide insights into the overall aggregation state and compactness (Yang & Yu 2023; Strey et al. 2024). Furthermore, we analyze spectral properties derived from the weighted Graph Laplacian, such as the Fiedler value (weighted algebraic connectivity), which is sensitive to the connectivity and potential for fragmentation within the network (Pavlou et al. 2023; Strey et al. 2024). We hypothesize that these dynamic weighted graph properties serve as powerful quantitative descriptors intimately linked to the physical state and stability of the peptide aggregate.

To validate our approach and demonstrate its utility, we apply this dynamic weighted graph analysis to molecular dynamics simulations of KYFIL pentapeptides, a well-studied self-assembling system. We analyze a substantial portion of the simulation trajectory, focusing on the later stages where aggregates are formed and undergo dynamic fluctuations. We calculate the time evolution of the proposed weighted graph metrics and systematically correlate them with physical properties of the peptide aggregate, such as the radius of gyration and a derived packing score for the LCC, to confirm that the network structure effectively captures the aggregate's physical state. We also perform a comparative analysis using binary contact graphs to quantitatively assess the added sensitivity and information content provided by incorporating interaction weights. Through this comprehensive analysis, we aim to identify specific weighted graph metrics that serve as robust indicators of aggregate structural states and hold potential as predictive order parameters for tracking dynamic transitions like fragmentation or changes in compactness. This work establishes a rigorous quantitative framework for characterizing peptide aggregates based on their dynamic weighted network properties, offering enhanced insights into the complex interplay of forces governing their behavior and stability.

## 2. METHODS

The objective of this study is to quantitatively characterize the dynamic behavior and stability of peptide aggregates formed during self-assembly using a novel dynamic weighted graph analysis approach. This methodology leverages molecular dynamics simulation data to construct and analyze time-evolving peptide interaction networks, correlating network properties with physical characteristics of the aggregates. The analysis focuses on the later stages of assembly, from 100 ns onwards, where distinct aggregates have formed and undergo dynamic fluctuations.

### 2.1. *Molecular Dynamics Simulation Data*

The analysis is based on a pre-existing molecular dynamics simulation trajectory of 30 KYFIL pentapeptides in explicit solvent. The simulation data comprises a topology file (`stripped.parm7`) and a trajectory file (`stripped.nc`). The simulation has a total duration exceeding 100 ns.

The trajectory contains atomic coordinates saved at regular time intervals, providing a detailed record of the system's temporal evolution. For this study, we specifically analyzed the trajectory segment starting from 100 ns until the end of the simulation. This period was chosen to focus on the dynamic behavior and stability of established aggregates rather than initial assembly events. The simulation universe was loaded and processed using the `MDAnalysis` library, allowing access to atomic coordinates, topology information, and simulation parameters such as the time step. Each of the 30 peptide molecules, composed of the residues Lys-Tyr-Phe-Ile-Leu, was identified and treated as a distinct entity for network analysis. For subsequent contact calculations, the heavy atoms (non-hydrogen atoms) of each peptide were selected.

### 2.2. *Peptide Representation and Contact Definition*

In our graph-based framework, each peptide molecule is represented as a node. Edges between nodes signify noncovalent interactions between the corresponding peptides. To capture the nuanced nature of peptide

interactions, we defined several distinct types of noncovalent contacts based on inter-atomic distances between heavy atoms of different peptides. These contact types are:

#### 2.2.1. *Hydrophobic Contacts*

Hydrophobic contacts were defined based on proximity between heavy atoms belonging to the sidechains of hydrophobic and aromatic residues: Isoleucine (ILE), Leucine (LEU), Phenylalanine (PHE), and the aromatic part of Tyrosine (TYR).

Specifically, heavy atoms involved are ILE (CG1, CG2, CD1), LEU (CG, CD1, CD2), PHE (CG, CD1, CD2, CE1, CE2, CZ), and TYR (CG, CD1, CD2, CE1, CE2, CZ). A hydrophobic contact was registered between two peptides if any heavy atom from the defined hydrophobic/aromatic list in one peptide was within a cutoff distance of 5.0 Å of any such atom in the other peptide.

#### 2.2.2. *Aromatic Contacts ($\pi$-$\pi$ stacking)*

Aromatic interactions, indicative of $\pi$-$\pi$ stacking, were defined based on proximity between heavy atoms within the aromatic rings of Phenylalanine (PHE) and Tyrosine (TYR) (Das et al. 2020; Bensberg et al. 2025).

The heavy atoms considered are PHE (CG, CD1, CD2, CE1, CE2, CZ) and TYR (CG, CD1, CD2, CE1, CE2, CZ). Aromatic contact was defined to exist if any heavy atom within the aromatic ring of a PHE or TYR residue in one peptide was within 5.0 Å of any heavy atom within the aromatic ring of a PHE or TYR residue in another peptide.

While more sophisticated definitions involving ring centroids and normal vectors exist, this simpler heavy-atom proximity criterion was adopted for computational efficiency across the large trajectory dataset.

#### 2.2.3. *Hydrogen Bonds*

Inter-peptide hydrogen bonds were identified based on proximity between potential heavy atom donors and acceptors. Potential heavy atom donors include the backbone nitrogen (N), the Lysine sidechain nitrogen (NZ), and the Tyrosine hydroxyl oxygen (OH). Potential heavy atom acceptors include the backbone oxygen (O) and the Tyrosine hydroxyl oxygen (OH).

A hydrogen bond was considered to be present if the distance between a potential heavy atom donor on one peptide and a potential heavy atom acceptor on another peptide was within 3.5 Å (Modirrousta-Galian & Maddalena 2021). This definition uses the heavy atom distance as a proxy for a hydrogen bond, simplifying the calculation by not requiring explicit hydrogen atom positions or angle criteria.

For each pair of peptides and for each frame of the simulation trajectory, the number of contacts of each type (hydrophobic, aromatic, hydrogen bond) was calculated based on these criteria.

### 2.3. *Dynamic Weighted Graph Construction*

For each simulation frame $t$ within the analysis window (100 ns onwards), a weighted adjacency matrix $A_w(t)$ of size $N_p \times N_p$ was constructed, where $N_p = 30$ is the number of peptides (Silva et al. 2019; Jagvaral et al. 2022).

The element $A_w(t)_{ij}$ represents the weighted strength of interaction between peptide $i$ and peptide $j$ at time $t$. This weight was calculated as a linear combination of the counts of the different contact types between peptides $i$ and $j$ (Silva et al. 2019).

$$A_w(t)_{ij} = w_H \times C_H(t)_{ij} + w_A \times C_A(t)_{ij} + w_{HB} \times C_{HB}(t)_{ij}$$

where $C_H(t)_{ij}$, $C_A(t)_{ij}$, and $C_{HB}(t)_{ij}$ are the counts of hydrophobic, aromatic, and hydrogen bond contacts between peptides $i$ and $j$ at time $t$, respectively. The weighting factors $w_H$, $w_A$, and $w_{HB}$ were set to 1.0, 1.5, and 2.0, respectively. These weights were chosen to reflect the potential relative importance or strength of these interaction types, with aromatic and hydrogen bonds potentially contributing more specifically to ordered structures compared to generic hydrophobic contacts.

The matrix $A_w(t)$ is symmetric ($A_w(t)_{ij} = A_w(t)_{ji}$) as interactions are reciprocal. This process was performed independently for each frame in the analysis window, resulting in a time series of weighted adjacency matrices $\{A_w(t)\}$.

For comparative analysis, binary adjacency matrices $A_b(t)$ were also constructed (Shrivastava & Li 2014; George et al. 2019). In the binary graph, an edge exists between peptide $i$ and peptide $j$ if the total weighted interaction $A_w(t)_{ij}$ is greater than zero. Thus, $A_b(t)_{ij} = 1$ if $A_w(t)_{ij} > 0$, and $A_b(t)_{ij} = 0$ otherwise (George et al. 2019).

### 2.4. *Graph Theoretical Analysis*

A suite of graph theoretical metrics was calculated for each frame's weighted adjacency matrix $A_w(t)$ to characterize the dynamic interaction network (Pavlou et al. 2023; Yang & Yu 2023). These metrics were chosen to probe different aspects of network structure, connectivity, and potential stability.

#### 2.4.1. *Weighted Graph Laplacian and Spectral Properties*

The weighted Graph Laplacian matrix $L_w(t)$ was computed for each frame as $L_w(t) = D_w(t) - A_w(t)$ (Granziol

et al. 2019), where $D_w(t)$ is the weighted degree matrix. $D_w(t)$ is a diagonal matrix where the diagonal element $D_w(t)_{ii}$ is the weighted degree of node $i$, calculated as the sum of weights of all edges incident to node $i$: $D_w(t)_{ii} = \sum_{j=1}^{N_p} A_w(t)_{ij}$ (Druskin et al. 2021).

The eigenvalues of the weighted Laplacian $L_w(t)$ were computed (Drábek et al. 2018). The eigenvalues are non-negative and can be ordered as $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N_p-1}$. The second smallest eigenvalue, $\lambda_1$, is known as the Fiedler value or algebraic connectivity. For a connected graph, $\lambda_1 > 0$, and its magnitude is related to how well the graph is connected; a larger $\lambda_1$ indicates stronger connectivity and greater resistance to fragmentation. For a disconnected graph, $\lambda_1 = 0$. The Fiedler value provides a spectral measure sensitive to the overall connectivity and bottleneck structures within the weighted network.

#### 2.4.2. *Global Graph Properties*

Several global properties were calculated for the weighted graph (Sun et al. 2024):

- **Weighted Graph Density:** Defined as the sum of all edge weights divided by the maximum possible sum of weights if every pair of distinct nodes was connected with a weight of 1.0 (or scaled appropriately). More simply, it is calculated as the sum of all entries in $A_w(t)$ (excluding the diagonal) divided by $N_p(N_p - 1)$. It provides a measure of the overall interaction strength normalized by system size.
- **Number of Connected Components ($N_{cc}$):** Calculated by treating any edge with a weight greater than zero as a connection. This metric indicates the number of distinct aggregates or clusters present in the system.
- **Size of the Largest Connected Component ($S_{lcc}$):** The number of nodes (peptides) in the largest cluster. This metric provides a measure of the extent of aggregation.

These metrics were also calculated for the binary graphs $A_b(t)$ for comparison (Čokina et al. 2021; Shan et al. 2025).

#### 2.4.3. *Largest Connected Component (LCC) Analysis*

For frames where a single large aggregate (LCC) was present ($S_{lcc} > 1$), a subgraph corresponding to the LCC was extracted (Whang et al. 2015; Parroni et al. 2020). Graph theoretical metrics, including the weighted Fiedler value and weighted density, were calculated specifically for this LCC subgraph. This allowed us to probe the internal connectivity and structure of the main aggregate.

### 2.5. *Correlation with Aggregate Physical Properties*

To validate that the network structure captured by graph metrics is indeed reflective of the physical state of the peptide aggregate, we calculated physical properties of the LCC and correlated them with the corresponding graph metrics of the LCC subgraph.

#### 2.5.1. *Radius of Gyration ($R_g$)*

For each frame, the peptides belonging to the LCC were identified. The radius of gyration ($R_g$) of the LCC was calculated using the coordinates of the heavy atoms of all peptides within the LCC. $R_g$ is a standard measure of the compactness of a molecular aggregate.

#### 2.5.2. *LCC Packing Score*

A simple packing score for the LCC was defined as the ratio of its size ($S_{lcc}$) to the cube of its radius of gyration ($R_g^3$): Packing Score = $S_{lcc}/R_g^3$. This metric provides a measure of how densely the peptides are packed within the LCC, accounting for both size and spatial extent. The time evolution of this packing score was calculated.

Pearson correlation coefficients were computed between the time series of the LCC packing score and the time series of LCC graph metrics (e.g., LCC weighted Fiedler value, LCC weighted density) to assess the degree to which network properties reflect physical compactness (Nandi et al. 2023; Krishna & Rao 2024).

### 2.6. *Order Parameter Development*

Based on the analysis of individual metrics and their correlation with physical properties, we explored the development of a combined graph-based order parameter to provide a more stable and potentially predictive measure of aggregate state and transitions.

A candidate order parameter for the LCC was proposed as a product of key LCC graph metrics:

$$\mathrm{OP}_{\mathrm{LCC}}(t) = S_{lcc}(t) \times \lambda_{1,\mathrm{LCC}}(t) \times \mathrm{Density}_{\mathrm{LCC}}(t)$$

where $\lambda_{1,\mathrm{LCC}}(t)$ is the weighted Fiedler value of the LCC subgraph and $\mathrm{Density}_{\mathrm{LCC}}(t)$ is the weighted density of the LCC subgraph at time $t$ (Whang et al. 2015).

Components were normalized if necessary to ensure they contribute appropriately to the combined parameter. The temporal evolution of this combined order parameter was analyzed and compared to the dynamics of individual metrics and physical properties.

### 2.7. *Computational Implementation*

The entire analysis pipeline, from loading simulation data and calculating contacts to constructing graphs and computing metrics, was implemented using Python 3. Key libraries utilized include `MDAnalysis` for trajectory handling and atom selections (Reuter & Köfinger 2018), `numpy` for numerical operations, `scipy` for eigenvalue decomposition and graph algorithms (like connected components), and `h5py` for efficient storage of time-series data such as adjacency matrices.

The computationally intensive steps involving per-frame contact calculations and graph metric computations were parallelized across multiple CPU cores using Python's `multiprocessing` module to enhance performance. The analysis was executed on a computing cluster utilizing 128 CPUs.

The time series of weighted adjacency matrices (Gilpin 2024; García et al. 2024), binary adjacency matrices (Gilpin 2024), and all calculated graph theoretical metrics (García et al. 2024; Strey et al. 2024) and physical properties (Strey et al. 2024) were stored in HDF5 files for subsequent analysis and visualization.

### 2.8. *Comparison with Binary Graph Analysis*

To quantitatively assess the added value of incorporating interaction weights, a parallel analysis was conducted using binary graphs $A_b(t)$. For each frame, standard binary graph metrics, including binary Fiedler value, binary density, number of connected components, and size of the largest connected component, were calculated. The temporal profiles and correlation strengths of these binary metrics were compared directly against their weighted counterparts to highlight differences in sensitivity to structural fluctuations and correlation with physical aggregate properties.

## 3. RESULTS

The self-assembly process of 30 KYFIL pentapeptides was investigated over a 1335.42 ns simulation window, spanning from 100 ns to the end of the 1435.42 ns trajectory, comprising 66771 frames. This analysis employed a dynamic graph-theoretical approach, representing peptides as nodes and inter-peptide noncovalent contacts as edges. We constructed both weighted graphs, where edge weights reflect the type and count of hydrophobic, aromatic, and hydrogen bond contacts with specific weighting factors (1.0, 1.5, 2.0), and binary graphs, where edges simply indicate the presence of any contact.

### 3.1. *Overall aggregation dynamics and system-level connectivity*

The overall state of aggregation was initially characterized by the number of connected components ($N_{cc}$) and the size of the largest connected component ($S_{lcc}$). For both weighted and binary graph representations, the connectivity definition used for component analysis is binary (an edge weight $> 0$ constitutes a connection), thus these metrics are identical for both representations.

Over the analyzed trajectory segment, the system typically existed as multiple distinct aggregates, as indicated by an average $N_{cc}$ of 15.2 (standard deviation 2.28), ranging from 8 to 24 components. Correspondingly, the $S_{lcc}$, representing the size of the primary aggregate, averaged 10.67 peptides (standard deviation 3.076), varying between 2 and 19 peptides out of the total 30. The temporal evolution of $N_{cc}$ and $S_{lcc}$ is shown in the top panels of Figure 1, revealing significant dynamic fluctuations reflecting ongoing processes of aggregate formation, growth, fragmentation, and dissolution within the simulation window. These fluctuations highlight the transient nature of the aggregates and the challenge in capturing their state with simple size or count metrics alone. The frequency distribution of the weighted $S_{lcc}$ is presented in Figure 2, showing that LCC sizes between 10 and 12.5 peptides were the most frequently observed.

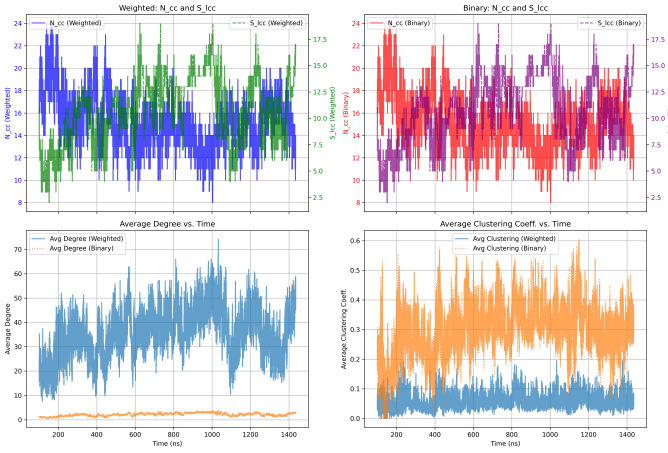


**Figure 1.** Time evolution of graph-theoretical metrics characterizing peptide aggregation from molecular dynamics simulations. Top panels show the number of connected components ($N_{cc}$) and size of the largest connected component ($S_{lcc}$) for weighted and binary graph representations, highlighting dynamic fluctuations in the overall aggregation state. Bottom panels compare average degree (Left) and average clustering coefficient (Right) between weighted and binary graphs, demonstrating the differing sensitivity of these metrics to interaction strength.

System-level connectivity was further probed by the Fiedler value ($\lambda_1$) of the entire 30-node graph. As illustrated in Figure 3 and Figure 4, the binary system Fiedler value was consistently 0.0 across all analyzed

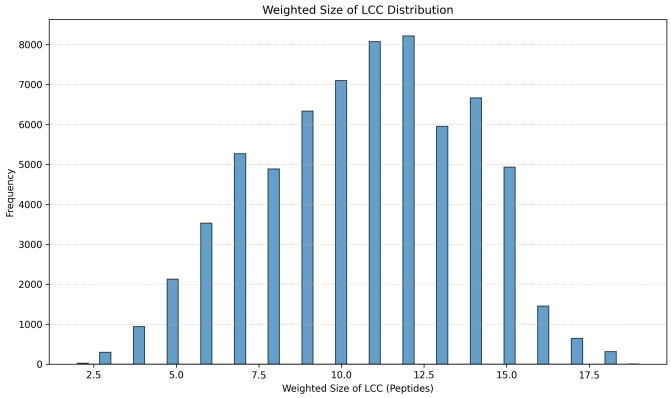


**Figure 2.** Histogram showing the frequency distribution of the metric labeled "Weighted Size of LCC (Peptides)". This distribution characterizes the range and prevalence of values for this largest connected component property observed during the simulation trajectory. The peak between 10 and 12.5 indicates the most frequently occurring values for this metric.

frames, confirming that the system, based on simple contact presence, always comprises multiple disconnected components, consistent with the average $N_{cc} > 1$. The weighted system Fiedler value was also found to be extremely close to zero (mean -5.205e-06, standard deviation 2.82e-06), further indicating that even when considering the collective strength of interactions, the entire system of 30 peptides rarely forms a single, robustly connected network. This is clearly visible in the time series plots shown in Figure 5 (top panel) and Figure 6. This underscores the importance of focusing analysis on the properties of the largest connected component, which represents the main aggregate.

The system-wide densities, reflecting the overall interaction level across all possible peptide pairs ($N_p(N_p - 1)/2$), were relatively low. The weighted system density averaged 1.234 (std: 0.3281), while the binary system density averaged 0.07402 (std: 0.02169). These values are significantly lower than the densities observed within the LCC (discussed below), confirming that interactions are concentrated within specific aggregates rather than distributed uniformly across the system.

### 3.2. *Characterization of the largest connected component (LCC)*

Given its prominence as the main aggregate, the LCC was analyzed in detail using both physical and graph-theoretical metrics.

#### 3.2.1. *Physical properties of the LCC*

The physical dimensions and compactness of the LCC were quantified by its radius of gyration ($R_g$) and a derived packing score ($S_{lcc}/R_g^3$). The distribution of the

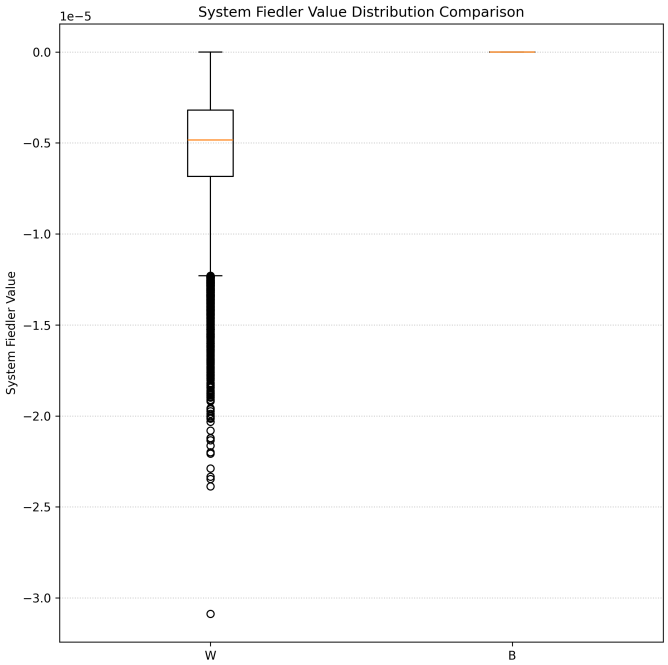


**Figure 3.** Distribution of the system Fiedler value for weighted (W) and binary (B) graphs. The binary value is consistently 0, and the weighted value is near 0, illustrating that the overall system of 30 peptides remains largely disconnected throughout the trajectory.

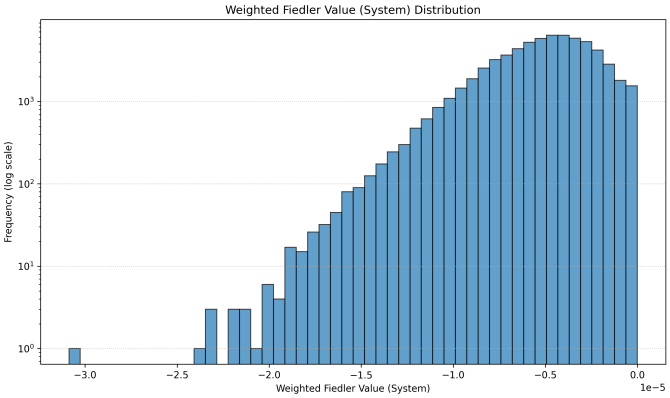


**Figure 4.** Distribution of the weighted Fiedler value for the entire system of 30 KYFIL peptides. The histogram shows that the system Fiedler value is predominantly very close to zero, indicating that the full system graph is typically disconnected or only weakly connected across the trajectory.

LCC $R_g$ is shown in Figure 7, averaging 13.28 Å (standard deviation 1.539 Å), with values ranging from 7.466 Å to 21.21 Å.

The packing score averaged 0.004531 peptides/Å$^3$ (standard deviation 0.0009041 peptides/Å$^3$), ranging from 0.000955 to 0.007483 peptides/Å$^3$. The distribu-

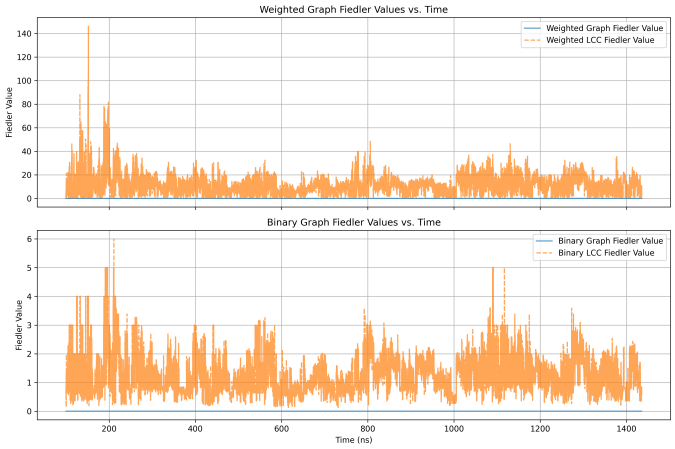


**Figure 5.** Time evolution of the Fiedler value for the entire system and the Largest Connected Component (LCC) using weighted (top) and binary (bottom) graph representations from 100 ns. System Fiedler values remain near zero, indicating the overall system is disconnected. LCC Fiedler values fluctuate significantly, showing dynamic changes in internal connectivity. Weighted LCC values exhibit larger magnitude and variability, reflecting sensitivity to interaction strengths within the aggregate.

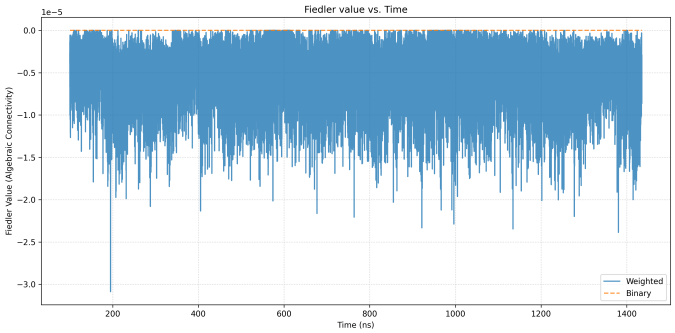


**Figure 6.** System Fiedler value over time. The Fiedler value measures the algebraic connectivity of the entire system graph of 30 peptides. The time series for both weighted (blue) and binary (orange, dashed) graph representations are shown. Both values remain close to zero throughout the trajectory, indicating that the system of 30 peptides exists predominantly as multiple disconnected aggregates.

tion of the LCC packing score is summarized by the boxplot in Figure 8 and the histogram in Figure 9.

The temporal profile of the packing score is shown in Figure 10, illustrating considerable fluctuations that reflect dynamic changes in the LCC's size, shape, and internal arrangement over time. Smaller $R_g$ values and higher packing scores correspond to more compact aggregates.

#### 3.2.2. *Graph-theoretical signatures of the LCC*

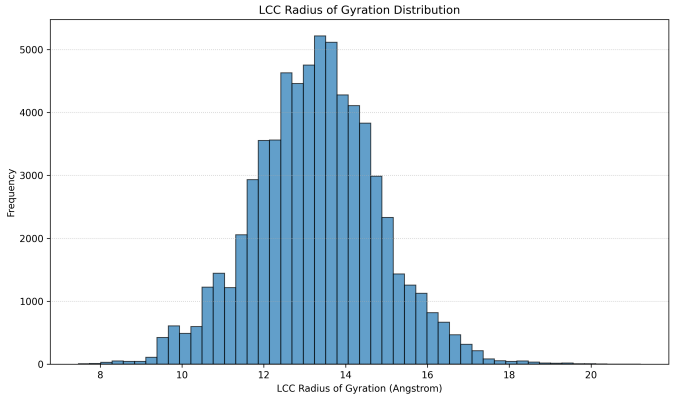


**Figure 7.** Distribution of the Largest Connected Component (LCC) Radius of Gyration (Å) from molecular dynamics simulations. The histogram shows the frequency distribution of LCC compactness over the trajectory, revealing its range and most probable values.

The internal connectivity and structural robustness of the LCC were characterized using its Fiedler value ($\lambda_1$) and density, calculated for the subgraph corresponding to the LCC. The temporal evolution of the LCC Fiedler value is shown in Figure 5 (bottom panels). The distribution of the weighted LCC Fiedler value is shown in Figure 11, averaging 11.37 (std: 7.398), while the distribution of the binary LCC Fiedler value is shown in Figure 12, averaging 1.262 (std: 0.6399). The significantly larger magnitude and higher variability of the weighted Fiedler value compared to its binary counterpart demonstrate its greater sensitivity to the nuances of interaction strengths within the LCC. A higher Fiedler value indicates a more strongly connected LCC that is more resistant to fragmentation. Both metrics exhibited substantial temporal fluctuations, mirroring the dynamic nature of the LCC's internal structure and stability.

The weighted LCC density averaged 8.039 (std: 2.88), and the binary LCC density averaged 0.4959 (std: 0.1113). The boxplot comparison in Figure 13 shows that the weighted density, representing the sum of edge weights normalized by the maximum possible sum, is substantially higher than the binary density, which is simply the ratio of existing edges to possible edges. Similar to the Fiedler value, the weighted density shows greater absolute magnitude and variability, reflecting its sensitivity to the total interaction strength within the LCC. Temporal fluctuations in LCC density (not shown) indicate periods of increased or decreased internal cohesion.

### 3.3. *Correlation between graph structure and physical properties of the LCC*

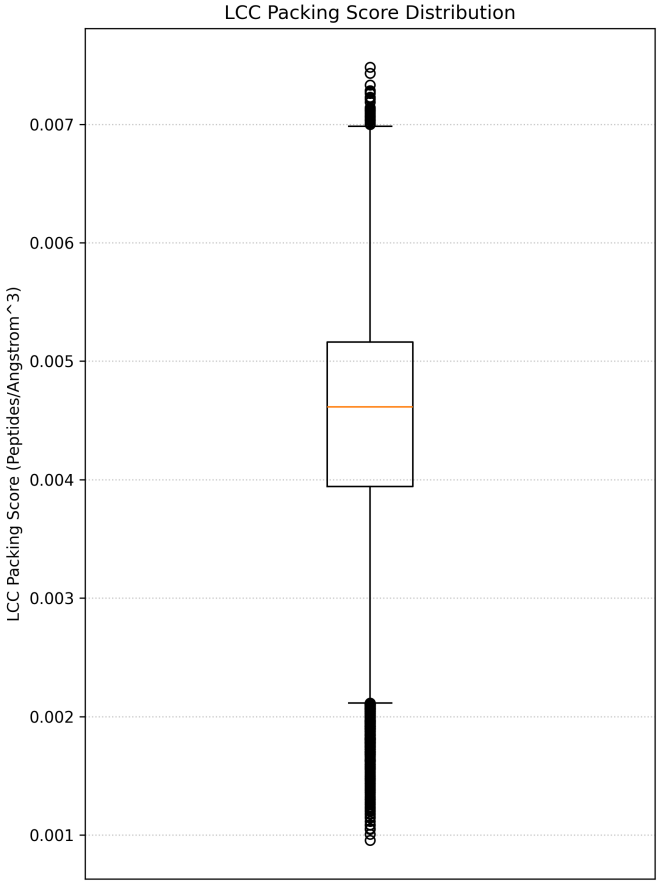


**Figure 8.** Distribution of the Largest Connected Component (LCC) packing score. The packing score, defined as $S_{lcc}/Rg^3_{LCC}$, quantifies the compactness of the LCC. The boxplot summarizes the distribution of this metric over the simulation trajectory, illustrating the dynamic fluctuations in LCC density.

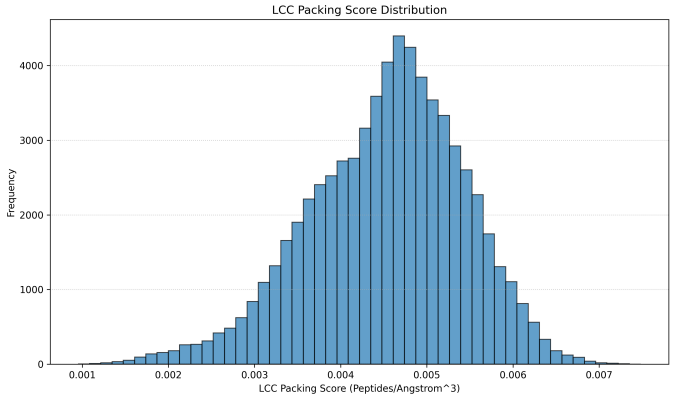


**Figure 9.** Histogram showing the distribution of the largest connected component (LCC) packing score, defined as LCC size divided by the cube of its radius of gyration ($S_{lcc}/Rg^3_{LCC}$). This metric quantifies the physical compactness of the LCC over the simulation trajectory, revealing a distribution centered around 0.0045 peptides/Å$^3$.

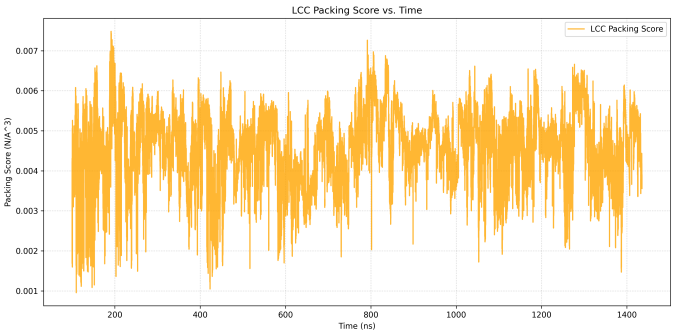


**Figure 10.** Temporal evolution of the Largest Connected Component (LCC) packing score. The fluctuations indicate dynamic changes in the compactness of the LCC aggregate over time.

To assess whether the graph-theoretical metrics capture meaningful physical characteristics of the LCC, Pearson correlation coefficients were computed between the time series of LCC graph metrics and its physical properties ($R_g$ and packing score).

Both weighted and binary LCC Fiedler values showed moderate negative correlations with $R_g$ ($r = -0.556$ for weighted, $r = -0.594$ for binary, $p < 0.001$). This indicates that LCCs with stronger algebraic connectivity, and thus greater robustness against fragmentation, tend to be more physically compact (smaller $R_g$), as illustrated for the binary Fiedler value in Figure 14.

LCC densities also correlated negatively with $R_g$. The weighted density showed $r = -0.697$ ($p < 0.001$), while the binary density exhibited a strong correlation of $r = -0.858$ ($p < 0.001$), as shown in Figure 15. This confirms that denser LCCs are more compact. Interestingly, the binary density showed a stronger correlation with $R_g$ than the weighted density, suggesting that for overall physical compactness, the topological structure of the network (how many connections exist relative to possibilities) might be a more dominant factor than the aggregate strength of interactions as defined by our current weighting scheme.

The size of the LCC ($S_{lcc}$) showed a strong positive correlation with $R_g$ ($r = 0.788$, $p < 0.001$), which is expected as larger aggregates generally occupy more space, as depicted in Figure 16.

Regarding the LCC packing score, both Fiedler values correlated positively: weighted $r = 0.631$ ($p < 0.001$) and binary $r = 0.687$ ($p < 0.001$). This indicates that more robustly connected LCCs tend to be better packed, as shown for the binary Fiedler value in Figure 17.

The LCC densities also correlated positively with the packing score, though less strongly than the Fiedler

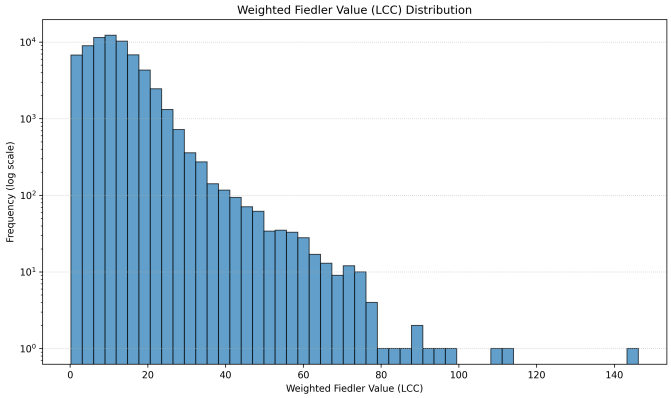


**Figure 11.** Histogram showing the distribution of the weighted Fiedler value for the Largest Connected Component (LCC). This metric quantifies the LCC's internal robustness based on weighted inter-peptide interactions. The broad distribution (log scale frequency) highlights the dynamic variability in aggregate stability observed during the simulation trajectory.

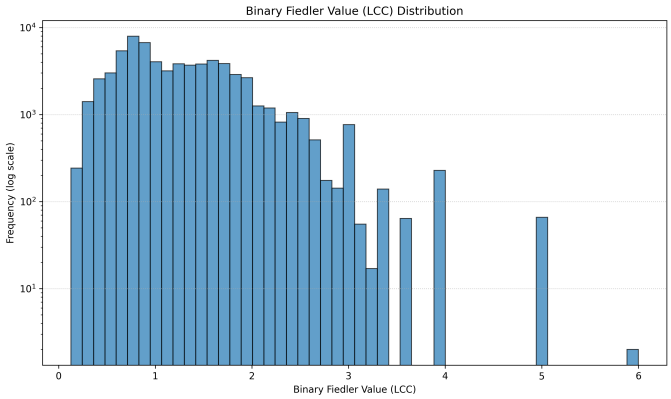


**Figure 12.** Distribution of the binary Fiedler value for the Largest Connected Component (LCC) of KYFIL peptide aggregates from molecular dynamics simulations (100-1435.42 ns). The log-scale histogram illustrates the frequency of observed binary Fiedler values, which quantify the LCC's algebraic connectivity. The distribution shows that the LCC frequently samples states with lower binary Fiedler values, indicating dynamic fluctuations in its internal robustness.

values: weighted $r = 0.229$ ($p < 0.001$) and binary $r = 0.376$ ($p < 0.001$). The correlation between $S_{lcc}$ and packing score was weak ($r = 0.144$, $p < 0.001$), suggesting that merely increasing the number of peptides in the LCC does not guarantee improved packing efficiency by this metric. These correlations demonstrate that graph-theoretical properties derived from both weighted and binary representations are effective in capturing key physical attributes of the LCC, such as compactness and packing efficiency.

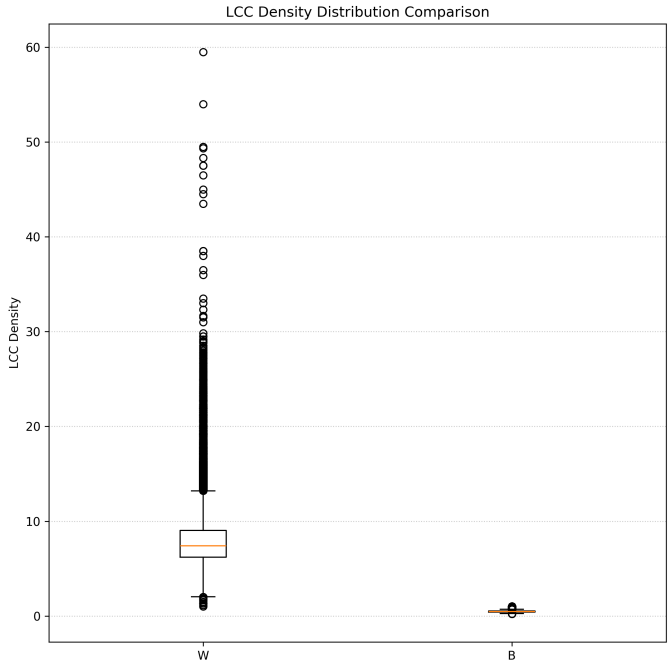


**Figure 13.** Boxplot comparison of the distributions of LCC internal density calculated using weighted (W) and binary (B) graph representations. The weighted density distribution is significantly broader and centered at a higher value than the binary density, reflecting its sensitivity to the strength and type of inter-peptide interactions. This highlights how weighted metrics capture a wider range of connectivity states within the largest aggregate compared to simple binary presence of contacts.

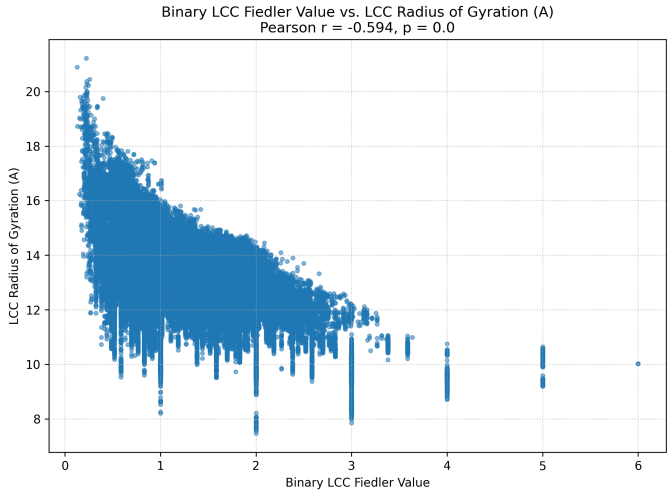


**Figure 14.** Scatter plot showing the relationship between the binary graph representation of the largest connected component (LCC) Fiedler value and its radius of gyration. Data from molecular dynamics simulations indicate a negative correlation (Pearson r = -0.594), suggesting that more robustly connected LCCs are generally more compact.

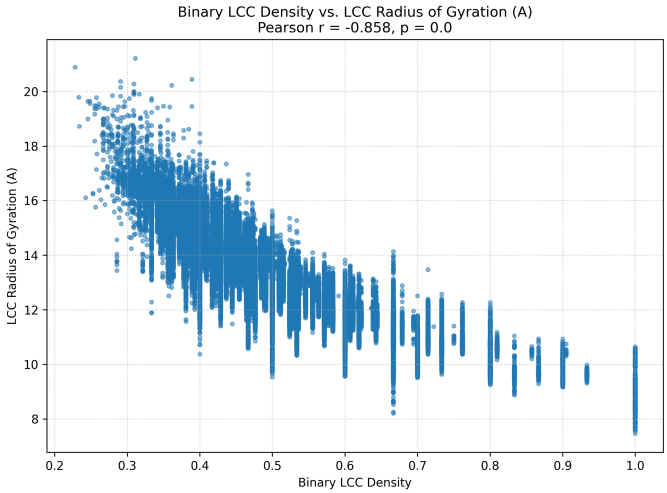


**Figure 15.** Scatter plot showing the relationship between the binary density of the Largest Connected Component (LCC) and its radius of gyration ($R_g$) over the molecular dynamics trajectory. A strong negative correlation (Pearson r = -0.858) is observed, indicating that LCCs with higher binary connectivity density tend to be more compact.

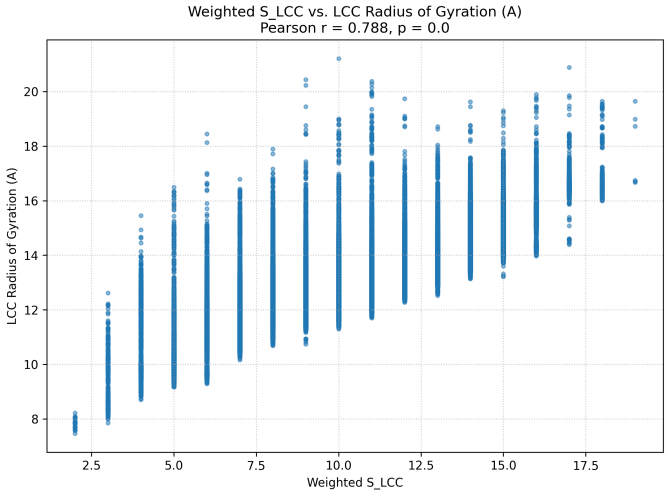


**Figure 16.** Scatter plot of LCC size ($S_{\mathrm{LCC}}$) versus LCC radius of gyration ($R_g$). A strong positive correlation ($r = 0.788$, $p < 0.001$) is observed, indicating that larger LCCs typically have a larger physical extent.

### 3.4. *Comparative sensitivity of weighted vs. binary metrics*

A direct comparison between weighted and binary graph metrics revealed key differences in their sensitivity to the underlying molecular interactions. Weighted metrics, such as weighted LCC Fiedler value and weighted LCC density, consistently showed larger absolute values and greater standard deviations over time compared to their binary counterparts (e.g., see distributions in Figures 13, 11, 12). This heightened variability reflects their sensitivity to the continuous changes in the number and types of specific noncovalent contacts within the aggregate, as opposed to binary metrics which only register the presence or absence of *any* contact. This is also evident in other metrics like the average betweenness centrality (Figure 18), where the weighted metric shows larger fluctuations. This supports our hypothesis that incorporating interaction strengths provides a more nuanced view of the dynamic interaction network.




**Figure 17.** Scatter plot showing the relationship between the binary largest connected component (LCC) Fiedler value and the LCC packing score. A positive correlation (Pearson r = 0.687, p < 0.001) indicates that LCCs with higher algebraic connectivity tend to be more efficiently packed.

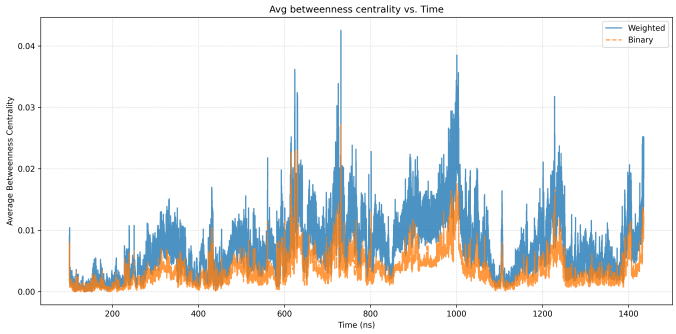


**Figure 18.** Temporal evolution of the average betweenness centrality of peptide aggregates, calculated using dynamic graph theory for both weighted (blue) and binary (orange) graph representations. The plot shows significant fluctuations over time, reflecting the dynamic nature of the aggregation process, and highlights the differences captured by the weighted and binary metrics.

However, while weighted metrics are more sensitive to interaction details, binary metrics, particularly binary

density, showed stronger correlations with the LCC radius of gyration (Figure 15). This suggests that for a macroscopic property like overall compactness, the topological structure of the network (how many connections exist relative to possibilities) might be a more dominant factor than the precise strength of each connection, under the current weighting scheme. Conversely, for the packing score, Fiedler values (both weighted and binary, Figure 17) showed stronger correlations than densities, indicating that network robustness against fragmentation is a better indicator of efficient internal packing than simple connection count or sum of weights.

### 3.5. *Combined graph-based order parameter for LCC state*

To provide a more integrated measure of the LCC's state, a combined graph-based order parameter ($OP_{LCC}$) was developed as the product of LCC size ($S_{lcc}$), LCC Fiedler value, and LCC density.

The weighted combined order parameter (w_OP_LCC_combined) averaged 975.2 (std: 814.7), while the binary version (b_OP_LCC_combined) averaged 6.619 (std: 3.912). The distribution comparison in Figure 19 shows that the weighted version exhibits a significantly larger magnitude and dynamic range, reflecting the combined sensitivity of its weighted components. The temporal evolution of the binary and weighted combined parameters are shown in Figure 20 and Figure 21, respectively, illustrating substantial fluctuations that track major changes in the LCC's size, connectivity, and density.

The w_OP_LCC_combined appears to capture a wider spectrum of LCC states due to its incorporation of interaction strengths, potentially offering a more stable and informative descriptor of aggregate stability and transitions compared to individual metrics like $S_{lcc}$ alone.

### 3.6. *Potential as indicators of aggregate stability*

The dynamic graph analysis provides a quantitative framework for characterizing the fluctuating states of peptide aggregates. Metrics like the LCC Fiedler value, which quantifies resistance to fragmentation, and the combined order parameter, which integrates multiple structural aspects, are promising candidates for serving as indicators of aggregate stability. The observed correlations between these graph metrics and physical properties like compactness validate their relevance to the physical state of the aggregate. While this study focused on characterization, the identification of sensitive metrics like the weighted LCC Fiedler value and the combined order parameter lays the groundwork for

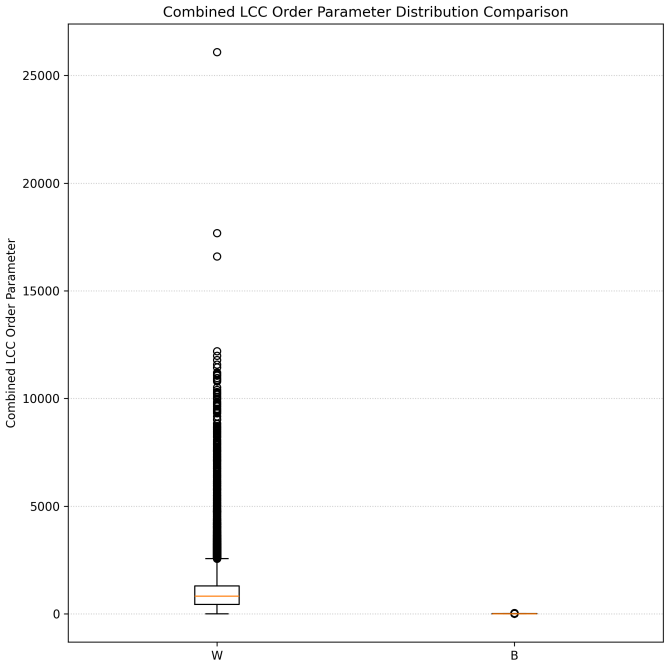


**Figure 19.** Distribution of the combined LCC order parameter for weighted (W) and binary (B) graph representations. The combined order parameter is defined as the product of LCC size, Fiedler value, and internal density. The weighted parameter shows a significantly larger range and higher values, reflecting its sensitivity to the strength of inter-peptide interactions within the aggregate.

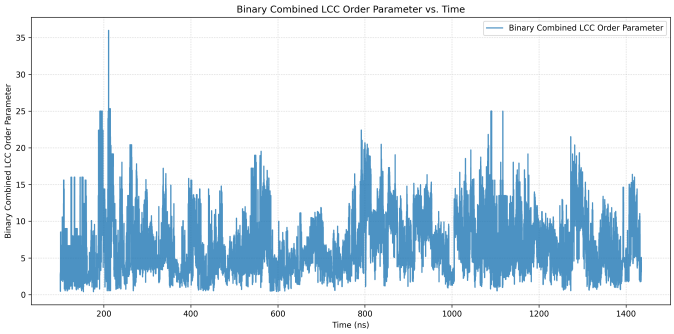


**Figure 20.** Time series of the binary combined order parameter for the largest connected component (LCC). This metric, calculated as the product of LCC size, binary Fiedler value, and binary density, illustrates the dynamic nature of the aggregate's state, reflecting fluctuations in its size, internal connectivity, and compactness over the simulation trajectory.

future studies aimed at predicting aggregate fragmentation or identifying long-lived states based on the temporal evolution of these graph properties. The distinct sensitivity profiles of weighted and binary metrics suggest that a combined approach or a careful selection of

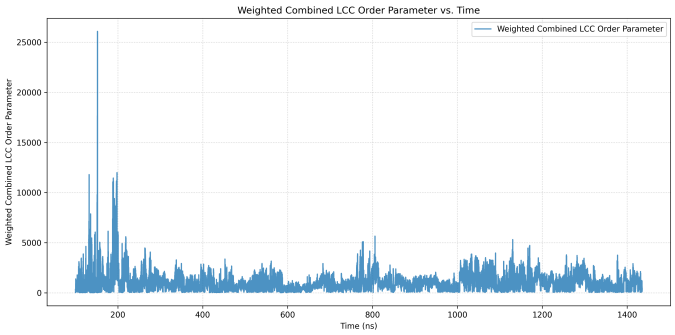


**Figure 21.** Temporal evolution of the weighted combined largest connected component (LCC) order parameter. This metric, which combines LCC size, weighted Fiedler value, and weighted density, exhibits significant dynamic fluctuations and a large range, reflecting the complex and changing state and internal stability of the primary aggregate over time.

metrics based on the specific question (e.g., overall compactness vs. internal interaction strength) is beneficial.

## 4. CONCLUSIONS

In this study, we addressed the challenge of quantitatively characterizing the complex, dynamic behavior and stability of peptide aggregates by introducing a novel approach based on dynamic weighted graph analysis. Representing peptides as nodes and their noncovalent interactions as weighted edges allowed us to capture the temporal evolution of the peptide assembly system as a dynamic network, incorporating the differential contributions of hydrophobic, aromatic, and hydrogen bond contacts.

Our analysis of molecular dynamics simulations of KYFIL pentapeptides revealed that the system exists as multiple dynamic aggregates, necessitating a focus on the properties of the largest connected component (LCC). By calculating graph theoretical metrics for the LCC subgraph, we demonstrated that both weighted and binary network properties provide valuable insights into aggregate structure. Key weighted metrics, such as the LCC Fiedler value and LCC density, exhibited greater sensitivity to the details of inter-peptide interaction strengths and types compared to their binary counterparts, reflecting the nuances of the dynamic contact network.

Crucially, we found significant correlations between these graph metrics and physical properties of the LCC, including its radius of gyration and packing score. These correlations validate that the network structure effectively captures the physical state and compactness of the aggregate. Specifically, higher LCC Fiedler values and higher LCC densities were associated with more compact and better-packed aggregates. While weighted metrics provided enhanced sensitivity to interaction details, binary density showed a strong correlation with overall compactness ($R_g$), suggesting that the topological pattern of connectivity is also a significant factor.

The dynamic weighted graph framework offers a robust quantitative means to monitor the fluctuating state of peptide aggregates over time. Metrics like the weighted LCC Fiedler value, which reflects connectivity robustness, and the proposed combined graph-based order parameter show promise as sensitive indicators of aggregate stability and potential transitions, such as fragmentation or changes in compactness.

In summary, this work establishes a rigorous, quantitative framework for characterizing peptide aggregates based on their dynamic weighted network properties. It highlights the value of incorporating interaction types and strengths into network analysis and identifies specific graph metrics that serve as powerful descriptors of aggregate structure and stability. These findings lay a strong foundation for future efforts aimed at developing predictive models for peptide self-assembly pathways and aggregate fate based on their dynamic network signatures.

## REFERENCES

Bensberg, M., Eckhoff, M., Husistein, R. T., et al. 2025, Hierarchical quantum embedding by machine learning for large molecular assemblies. https://arxiv.org/abs/2503.03928

Das, S., Lin, Y.-H., Vernon, R. M., Forman-Kay, J. D., & Chan, H. S. 2020, Comparative Roles of Charge, $\pi$ and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins, doi: https://doi.org/10.1073/pnas.2008122117

Druskin, V., Mamonov, A. V., & Zaslavsky, M. 2021, Distance preserving model order reduction of graph-Laplacians and cluster analysis. https://arxiv.org/abs/1809.03048

Drábek, P., Ho, K., & Sarkar, A. 2018, On the eigenvalue problem involving the weighted $p$-Laplacian in radially symmetric domains. https://arxiv.org/abs/1805.03512

García, C. R., Torres, D. F., Zhu-Ge, J.-M., & Zhang, B. 2024, Separating repeating fast radio bursts using the minimum spanning tree as an unsupervised methodology. https://arxiv.org/abs/2411.02216

George, S. V., Misra, R., & Ambika, G. 2019, Classification of Close Binary Stars Using Recurrence Networks, doi: https://doi.org/10.1063/1.5120739

Gilpin, W. 2024, Recurrences reveal shared causal drivers of complex time series. https://arxiv.org/abs/2301.13516

Granziol, D., Ru, R., Zohren, S., et al. 2019, A Maximum Entropy approach to Massive Graph Spectra. https://arxiv.org/abs/1912.09068

Jagvaral, Y., Lanusse, F., Singh, S., et al. 2022, Galaxies on graph neural networks: towards robust synthetic galaxy catalogs with deep generative models. https://arxiv.org/abs/2212.05596

Krishna, D., & Rao, M. S. 2024, Detecting cosmological recombination lines with a non-ideal antenna – a first step to practical realization. https://arxiv.org/abs/2411.18945

Modirrousta-Galian, D., & Maddalena, G. 2021, Of Aliens and Exoplanets: Why the search for life, probably, requires the search for water. https://arxiv.org/abs/2104.01683

Nandi, A., Pandey, B., & Sarkar, P. 2023, The correlations between galaxy properties in different environments of the cosmic web, doi: https://doi.org/10.1088/1475-7516/2024/02/012

Parroni, C., Tollet, E., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, Higher order statistics of shear field: a machine learning approach, doi: https://doi.org/10.1051/0004-6361/202038715

Pavlou, O., Michos, I., Lesta, V. P., et al. 2023, Graph Theoretical Analysis of local ultraluminous infrared galaxies and quasars, doi: https://doi.org/10.1016/j.ascom.2023.100742

Reuter, K., & Köfinger, J. 2018, CADISHI: Fast parallel calculation of particle-pair distance histograms on CPUs and GPUs, doi: https://doi.org/10.1016/j.cpc.2018.10.018

Shan, Y., Chen, J., Zhang, Z., et al. 2025, Identifying eclipsing binary stars with TESS data based on a new hybrid deep learning model. https://arxiv.org/abs/2504.15875

Shrivastava, A., & Li, P. 2014, A New Space for Comparing Graphs. https://arxiv.org/abs/1404.4644

Silva, D. H., Ferreira, S. C., Cota, W., Pastor-Satorras, R., & Castellano, C. 2019, Spectral properties and the accuracy of mean-field approaches for epidemics on correlated networks, doi: https://doi.org/10.1103/PhysRevResearch.1.033024

Strey, S.-G., Castronovo, A., & Elumalai, K. 2024, Graph Theoretic Approach Identifies Critical Thresholds at which Galaxy Filamentary Structures Form. https://arxiv.org/abs/2310.20184

Sun, Z., Ting, Y.-S., Liang, Y., et al. 2024, Knowledge Graph in Astronomical Research with Large Language Models: Quantifying Driving Forces in Interdisciplinary Scientific Discovery. https://arxiv.org/abs/2406.01391

Whang, J. J., Gleich, D. F., & Dhillon, I. S. 2015, Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion. https://arxiv.org/abs/1503.07439

Yang, D., & Yu, H.-B. 2023, A graph model for the clustering of dark matter halos, doi: https://doi.org/10.1103/PhysRevResearch.5.043187

Čokina, M., Maslej-Krešňáková, V., Butka, P., & Štefan Parimucha. 2021, Automatic classification of eclipsing binary stars using deep learning methods, doi: https://doi.org/10.1016/j.ascom.2021.100488