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

Dynamic, Weighted, Hierarchical Graph Analysis for Predicting Peptide Aggregate Instability and Identifying Molecular Determinants

AstroPilot[1]

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

ABSTRACT

Understanding the stability and dynamics of peptide self-assemblies is crucial for designing functional biomaterials, yet predicting aggregate instability and identifying the specific molecular interactions that govern it remains a significant challenge. Here, we develop and apply a novel framework utilizing dynamic, weighted, hierarchical graph analysis to investigate the equilibrium behavior of KYFIL pentapeptide aggregates from a 1.3 μs molecular dynamics simulation. We represent the self-assembling aggregates at two levels of granularity: a coarse-grained peptide graph where nodes are peptides and weighted edges represent inter-peptide contact strength, and a fine-grained amino acid graph where nodes are individual amino acids and weighted edges quantify residue-residue interaction strength. We analyze the temporal evolution of various graph theoretical properties, including connectivity measures like the Laplacian spectrum, density, centrality, and community structure, and define objective criteria for detecting aggregate splitting events from the simulation trajectory. Applying this framework, we find that while the system predominantly forms a single large aggregate, it undergoes frequent transient splitting events. Crucially, we demonstrate that dynamic changes in graph properties serve as predictive signatures for impending splitting events within a nanosecond timescale; specifically, decreases in coarse-grained aggregate connectivity (Fiedler value) and density, and a significant decline in the weighted sum of fine-grained residue-residue contacts bridging future fragments, precede fragmentation. Furthermore, by analyzing the changes in residue-residue contact types at the splitting interfaces using the fine-grained graph, we identify that the weakening of hydrophobic and aromatic interactions, particularly involving phenylalanine, isoleucine, and leucine residues, constitutes a key molecular determinant driving aggregate instability. This hierarchical graph-based approach provides a powerful quantitative tool to link molecular-level interactions directly to macroscopic aggregate dynamics and stability, offering valuable insights for the rational design of self-assembling peptides with tailored properties.

Keywords: Astronomy software, Detection, Astrostatistics techniques, Distributed computing, Computational methods

1. INTRODUCTION

Peptide self-assembly is a ubiquitous phenomenon in biological systems and a cornerstone for the rational design of functional biomaterials. The intrinsic ability of peptides to aggregate into well-defined nanostructures, such as fibrils, spheres, and hydrogels, holds immense promise for applications spanning medicine, engineering, and nanotechnology. However, harnessing the full potential of self-assembling peptides necessitates a deep understanding and control over the stability and dynamic evolution of the resulting aggregates. Predicting when and how these assemblies might break apart or reorganize, and identifying the specific molecular interactions that govern these processes, remains a significant scientific challenge.

While experimental techniques provide valuable macroscopic insights, molecular dynamics (MD) simulations offer an atomistic window into the dynamic world of peptide self-assembly (Huang & Shuai 2013; Krasnokutski et al. 2024). MD simulations can capture the intricate dance of peptides forming and dissociating over time, providing trajectories rich in detail (Huang & Shuai 2013). Yet, extracting quantitative, predictive metrics for aggregate instability and pinpointing the exact molecular forces responsible for breakdown from these complex, high-dimensional datasets is far from straightforward (Huang & Shuai 2013). Tradi-

tional analysis methods, often focusing on global properties like aggregate size or static contact maps, may overlook the transient structural fluctuations, local connectivity changes, and dynamic rearrangements of interactions that are critical indicators of impending instability or "weak links" within an aggregate (Huang & Shuai 2013; Liang et al. 2025). There is a clear need for advanced analytical frameworks capable of systematically bridging the gap between the dynamic landscape of molecular interactions and the macroscopic behavior of peptide assemblies, providing a quantitative basis for assessing stability and identifying its molecular determinants (Liang et al. 2025).

Here, we address this challenge by developing and applying a novel analytical framework based on dynamic, weighted, hierarchical graph analysis. Our approach is designed to systematically investigate the equilibrium behavior and instability events of peptide aggregates, using a 1.3 μs MD simulation trajectory of self-assembling KYFIL pentapeptides as a model system.

The core idea is to represent the complex, self-assembling aggregates at multiple, interconnected levels of granularity using graphs whose structure and edge weights evolve dynamically over the simulation time. Specifically, we construct graphs at two principal levels. At a coarse-grained level, the nodes of the graph represent individual peptides, and the weighted edges between them quantify the strength of inter-peptide interactions, such as the total number of atomic contacts (Airale et al. 2025,?; Coupette et al. 2025). This coarse-grained graph provides a representation of the overall aggregate structure and connectivity. Simultaneously, at a fine-grained level, we build graphs where the nodes correspond to individual amino acids within the peptides. The weighted edges in this fine-grained graph capture the strength of specific residue-residue interactions, encompassing both interactions between residues within the same peptide (intra-peptide) and, crucially, interactions between residues belonging to different peptides (inter-peptide) (Airale et al. 2025,?; Coupette et al. 2025).

By constructing these dynamic, weighted graphs for each frame of the simulation trajectory, we gain a powerful representation of how the aggregate structure and the underlying molecular interactions evolve over time (Yang & Yu 2023; Roncoli et al. 2024). We then proceed to analyze the temporal evolution of a suite of graph theoretical properties at both the coarse-grained and fine-grained levels. These properties include measures of graph connectivity, such as the Laplacian spectrum and the Fiedler value (algebraic connectivity) (Yang & Yu 2023; Roncoli et al. 2024), as well as graph density, node centrality measures, and community structure (Yang & Yu 2023). To objectively identify events where aggregates break apart, we define and apply specific criteria for detecting aggregate splitting based on changes in the connectivity of the coarse-grained peptide graph over time (Yang & Yu 2023).

The predictive power of our framework is established by correlating the dynamic changes observed in the graph properties with the detected aggregate splitting events. We investigate whether specific changes in graph characteristics systematically precede instances of aggregate fragmentation within a defined time window. Furthermore, and crucially, the hierarchical nature of our approach allows us to delve into the molecular origins of instability. By analyzing the fine-grained, residue-level interactions, particularly focusing on how the strength and type of interactions change at the interfaces between parts of an aggregate that are about to separate, we can directly link molecular-scale events to the higher-level dynamics of aggregate splitting. This enables us to identify which specific types of residue-residue interactions act as critical "weak links" or whose weakening drives aggregate instability in this system. Through this comprehensive, multi-scale analysis, we demonstrate that dynamic changes in specific graph properties serve as quantifiable predictive signatures for aggregate instability and identify the key molecular determinants governing aggregate breakdown. This hierarchical graph-based framework provides a robust and generalizable quantitative tool for dissecting the complex processes of peptide self-assembly, offering valuable insights for the rational design of self-assembling peptides with precisely controlled and predictable stability.

2. METHODS

The primary objective of this study is to develop and apply a dynamic, weighted, hierarchical graph-based framework to analyze peptide aggregate stability and identify molecular determinants of instability. This involved processing molecular dynamics simulation data, constructing time-evolving graphs at two levels of granularity, analyzing their structural and dynamical properties, defining and detecting aggregate splitting events, and correlating graph dynamics with these events to uncover predictive signatures and molecular insights. The analysis focused on the equilibrium phase of the simulation, starting from 100 ns.

2.1. *Data Source and Preparation*

The dataset for this study consists of a 1.3 μs molecular dynamics simulation trajectory of 30 KYFIL pentapeptide molecules in an aqueous environment. The

simulation data, including topology and trajectory, were loaded using the MDAnalysis library. The topology file provided structural information ('stripped.parm7'), while the trajectory file ('stripped.nc') contained atomic coordinates over time. The simulation had a time step of 20 ps.

The total trajectory comprised 65000 frames (1.3 μs / 20 ps). To focus on the equilibrium behavior of the system, the analysis was restricted to the trajectory slice starting from 100 ns. Given the 20 ps frame interval, 100 ns corresponds to frame index $100,000$ ps$/20$ ps $=$ 5000. Therefore, the analysis window included frames from index 5000 to 64999, totaling 60000 frames and covering 1.2 μs of simulation time (Radice et al. 2025).

For efficient processing, we defined atom groups for each of the 30 peptides. For each peptide, lists of atom indices were created for all atoms, heavy atoms (excluding hydrogen), and heavy atoms belonging to each of the five constituent amino acids (Lysine, Tyrosine, Phenylalanine, Isoleucine, Leucine) in sequence. This structured indexing allowed for rapid retrieval of coordinates for specific peptides or residues during contact calculations. Each peptide consisted of 76 atoms, with 38 heavy atoms. The distribution of heavy atoms per residue type was consistent across all peptides: Lys (8 heavy atoms), Tyr (11 heavy atoms), Phe (11 heavy atoms), Ile (4 heavy atoms), Leu (4 heavy atoms).

2.2. *Interaction Definition and Calculation*

Molecular interactions between peptides and residues were quantified based on the proximity of heavy atoms. A contact between two heavy atoms was defined if their Euclidean distance was less than a cutoff distance of 4.5 Å. This cutoff is commonly used in molecular simulations to identify non-bonded interactions. All interaction calculations were performed independently for each frame within the 1.2 μs analysis window.

2.2.1. *Inter-Peptide Interaction Strength*

For the coarse-grained (CG) peptide graph, the interaction strength between any pair of peptides, say peptide i and peptide j ($i \neq j$), was defined as the total number of heavy atom contacts between all heavy atoms belonging to peptide i and all heavy atoms belonging to peptide j. This count, denoted w_{ij}^{peptide}, served as the weight for the edge connecting peptides i and j. These 30×30 matrices of contact counts were calculated and stored for every frame.

2.2.2. *Residue-Residue Interaction Strength*

For the fine-grained (FG) amino acid graph, interactions were quantified at the residue level. For each frame:

- **Inter-peptide_ inter-residue contacts**: For any pair of peptides (i, j, $i \neq j$) and any pair of amino acids (residue a within peptide i, residue b within peptide j), the number of heavy atom contacts between all heavy atoms of residue a (in peptide i) and all heavy atoms of residue b (in peptide j) was calculated. This count, $w_{ab}^{\text{inter}}(i,j)$, represents the interaction strength between these specific residues across different peptides.

- **Intra-peptide_ inter-residue contacts**: For any single peptide p and any pair of distinct amino acids (a, b) within that peptide, the number of heavy atom contacts between all heavy atoms of residue a and all heavy atoms of residue b (both in peptide p) was calculated. This count, $w_{ab}^{\text{intra}}(p)$, quantifies intra-peptide residue interactions.

These residue-level contact counts formed the basis for the edge weights in the FG graph. Due to the large number of possible residue pairs (150×150 total amino acid nodes), these interactions were stored efficiently, focusing on non-zero contact counts.

2.3. *Dynamic, Weighted, Hierarchical Graph Construction*

Based on the calculated interaction strengths, dynamic graphs were constructed for each frame of the simulation trajectory (Ramos-Osuna et al. 2025).

2.3.1. *Coarse-Grained (CG) Peptide Graph*

For each frame, a CG graph was constructed with 30 nodes, each representing one KYFIL peptide. An edge was created between peptide i and peptide j if their inter-peptide interaction strength w_{ij}^{peptide} was greater than zero. The weight of this edge was set to w_{ij}^{peptide}. This resulted in a weighted adjacency matrix W^{CG} of size 30×30 for each frame, where $W_{ij}^{\text{CG}} = w_{ij}^{\text{peptide}}$ for $i \neq j$ and $W_{ii}^{\text{CG}} = 0$.

2.3.2. *Fine-Grained (FG) Amino Acid Graph*

For each frame, a global FG graph was constructed with 150 nodes, representing each of the 30 peptides' 5 amino acids. Each node was uniquely identified by its peptide ID and residue index/type within the peptide (e.g., P1_K1, P1_Y2, ..., P30_L5). Edges were created between nodes (amino acids) if their corresponding heavy atom contact count was greater than zero. The edge weights were the calculated contact counts:

- For an edge between residue a in peptide i and residue b in peptide j ($i \neq j$), the weight was $w_{ab}^{\text{inter}}(i,j)$.

- For an edge between residue a and residue b within the same peptide p, the weight was $w_{ab}^{\text{intra}}(p)$.

This resulted in a 150×150 weighted adjacency matrix W^{FG} for each frame. When analyzing specific peptide aggregates identified in the CG graph, corresponding subgraphs of the global FG graph were extracted, including only the amino acid nodes belonging to the peptides within the aggregate and all interactions between them.

2.4. *Temporal Analysis of Graph Theoretical Properties*

A suite of graph theoretical properties was calculated for the CG and FG graphs for each frame to capture the dynamic evolution of aggregate structure and interactions (Pavlou et al. 2023; Strey et al. 2024).

2.4.1. *CG Graph Properties*

For each frame's CG graph:

- **Basic Properties**: The number of edges and the weighted graph density were calculated. Weighted density was defined as the sum of all edge weights divided by the maximum possible sum of weights for a complete graph with the same number of nodes, using the maximum observed contact count between any pair of peptides in that frame as the potential maximum edge weight. The number of connected components, representing peptide aggregates, and the size (number of peptides) of each component were determined. The size of the Largest Connected Component (LCC) was specifically tracked.

- **Connectivity**: The weighted graph Laplacian matrix $L = D - W^{\text{CG}}$ was constructed, where D is the diagonal matrix of weighted degrees (strengths), $D_{ii} = \sum_j W_{ij}^{\text{CG}}$. The eigenvalues of the Laplacian were computed. The second smallest eigenvalue, λ_2 (the Fiedler value or algebraic connectivity), was calculated for the LCC. The Fiedler value is a measure of graph connectivity; a larger value indicates higher connectivity and robustness, while a value of zero indicates a disconnected graph. For disconnected graphs, λ_2 was computed for each non-trivial component.

- **Centrality Measures**: Weighted degree centrality (sum of edge weights connected to a node), weighted betweenness centrality, and eigenvector centrality were calculated for each peptide node to identify peptides playing key roles in maintaining aggregate structure.

- **Community Structure**: The Louvain algorithm for weighted graphs was applied to the CG graph to identify communities or sub-clusters within larger aggregates (Reid et al. 2012; Prouteau et al. 2024). Modularity was calculated as a measure of the quality of the community partition (Prouteau et al. 2024).

2.4.2. *FG Graph Properties*

For the global FG graph, or more often, for FG subgraphs corresponding to identified CG aggregates.

- **Basic Properties**: Weighted density was calculated for the FG subgraphs of aggregates.

- **Centrality Measures**: Weighted degree centrality and weighted betweenness centrality were calculated for each amino acid node. This highlighted residues with a high number or strength of contacts, particularly inter-peptide_ contacts.

- **Inter-Peptide_ Residue Interaction Analysis**: The primary focus was on analyzing the strength and types of inter-peptide_ residue-residue interactions. Statistics (sum, mean, max) of edge weights were tracked, especially for interactions between residues located at the interfaces of peptides within an aggregate.

All calculated graph properties were stored as time series data for subsequent analysis (Boniol et al. 2025).

2.5. *Aggregate Definition and Splitting Event Detection*

Peptide aggregates were operationally defined as the connected components in the CG peptide graph at any given frame.

To detect splitting events, we tracked the composition of aggregates across consecutive frames. An aggregate A at frame t was considered to have undergone a splitting event at frame $t+1$ if the set of peptides comprising A at frame t were subsequently found distributed among two or more distinct connected components at frame $t+1$. A minimum size criterion (e.g., parent aggregate size ≥ 4 peptides, resulting fragments size ≥ 2 peptides) was applied to focus on significant splitting events.

For each detected split, we recorded the frame number, the identity of the parent aggregate (list of peptides), and the identities of the resulting daughter aggregates (lists of peptides).

2.6. *Correlation of Graph Dynamics with Aggregate Stability and Splitting*

To determine if dynamic changes in graph properties serve as predictive signatures for aggregate instability,

we analyzed the temporal evolution of selected properties in a time window immediately preceding a detected splitting event.

For each splitting event occurring at frame F_{split}, a pre-splitting_ window of $N_w = 50$ frames (1 ns) was defined, covering frames from $F_{\text{split}} - N_w$ to $F_{\text{split}} - 1$.

The time series of CG properties (Fiedler value, density, average inter-peptide_ edge weight) and FG properties (sum of inter-peptide_ residue-residue contacts between the future fragments) were extracted for the parent aggregate within this window. These pre-splitting_ dynamics were compared to the behavior of the same properties during periods where aggregates of similar size remained stable for significantly longer durations. Statistical comparisons (e.g., using t-tests or Mann-Whitney U tests) were performed to assess the significance of observed changes in the pre-splitting_ windows compared to control periods.

2.7. *Elucidating Molecular Determinants of Stability*

The FG amino acid graph was crucial for identifying the molecular interactions governing aggregate stability and instability.

For each splitting event, we analyzed the inter-peptide_ residue-residue interactions that bridged the peptides forming the future fragments in the frames preceding the split. The sum of weights of these "bridging" interactions was tracked over the pre-splitting_ window.

Furthermore, we analyzed which specific types of residue-residue pairs (e.g., F-F, Y-L, K-I) contributed most significantly to these bridging interactions and whose collective strength decreased most markedly before the split. This allowed us to identify specific residue-pair interaction types that acted as critical "weak links" driving aggregate fragmentation.

The relative contributions of different interaction types (hydrophobic, aromatic, polar, charged) to the total bridging interaction strength were also quantified.

2.8. *Computational Implementation*

The analysis was implemented using Python, leveraging several key libraries (Wang 2025; Zheng et al. 2025,?). MDAnalysis was used for loading and processing the simulation trajectory. NumPy and SciPy were utilized for numerical computations, including distance calculations (‘scipy.spatial.distance.cdist‘) and eigenvalue decomposition of the Laplacian matrix (‘scipy.linalg.eigh‘). Graph construction and analysis were performed using the ‘igraph‘ library, chosen for its performance with large graphs and graph ensembles.

The frame-by-frame calculations of contacts, graph construction, and property calculation are highly parallelizable. To exploit the available 128 CPU cores, we implemented a parallel processing strategy using Python's ‘multiprocessing‘ module, distributing batches of frames to worker processes.

All calculated time series data, graph representations (weighted edge lists), and splitting event details were systematically saved using efficient formats like HDF5 for later analysis and visualization.

3. RESULTS

The analysis of the 1.3 μs molecular dynamics simulation trajectory of 30 KYFIL pentapeptides, focusing on the equilibrium phase from 100 ns onwards, employed a dynamic, weighted, hierarchical graph framework to investigate aggregate structure, stability, and the molecular determinants of instability.

3.1. *System characterization and initial aggregation dynamics*

The simulation system consisted of 30 KYFIL pentapeptides in aqueous solution. The analysis window, spanning from 100 ns to 1435.42 ns (66,771 frames at 20 ps per frame), was selected to examine the system's behavior after the initial rapid aggregation phase. Peptides were defined as aggregates based on connectivity in a coarse-grained (CG) graph where nodes are peptides and edges represent any heavy atom contact within 4.5 Å.

Initial exploratory analysis revealed that within this analysis window, the system predominantly existed as a single large aggregate comprising all 30 peptides. As shown in Figure 1, the aggregate size distribution is heavily skewed towards the maximum size. The average number of aggregates was 1.0 ± 0.0, and the size of the Largest Connected Component (LCC) was consistently 30.0 ± 0.0 peptides throughout the analyzed period, as illustrated by the temporal evolution in Figure 2. This high global stability suggests that the primary focus of the dynamic analysis should be on the internal dynamics and transient dissociations occurring within this single, persistent aggregate, rather than the formation and dissolution of multiple distinct clusters. The "splitting events" detected subsequently therefore represent transient fragmentation or reorganization of this main aggregate.

3.2. *Hierarchical graph construction and average properties*

Dynamic, weighted graphs were constructed for each frame within the analysis window at two levels of granularity: coarse-grained (CG) peptide graphs and fine-grained (FG) amino acid graphs. The temporal evolu-

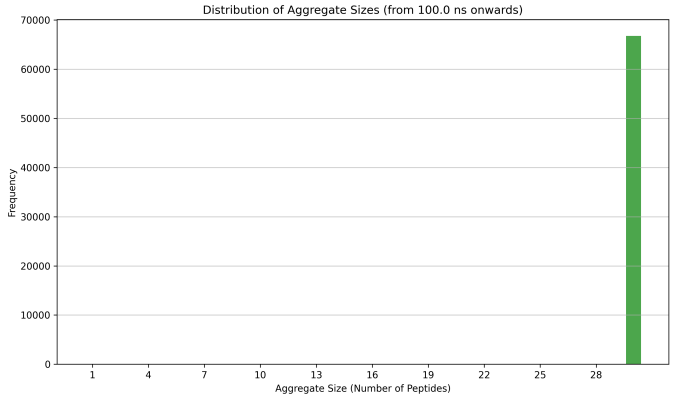

Figure 1. Aggregate size distribution of KYFIL pentapeptides from 100 ns onwards. The plot shows that the system predominantly exists as a single large aggregate comprising all 30 peptides throughout the analyzed simulation period.

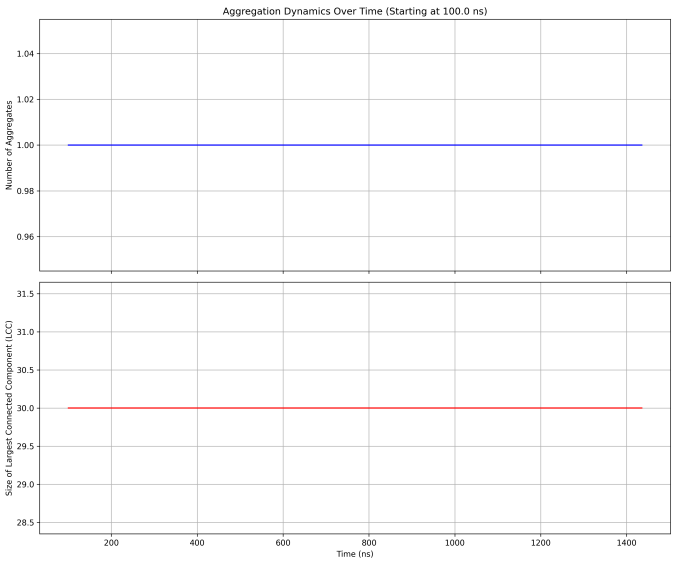

Figure 2. Temporal evolution of aggregate properties from 100 ns. Upper panel: number of aggregates (blue); Lower panel: size of the largest connected component (LCC, red). The plot shows the system consistently forms a single aggregate of all 30 peptides throughout this period, indicating its overall structural stability based on the defined contact criterion.

tion of various graph theoretical properties was calculated for these graphs to quantify structural and interaction dynamics.

3.2.1. *Coarse-grained (CG) peptide graph properties*

The CG graph represented the 30 peptides as nodes, with weighted edges corresponding to the total number of heavy atom contacts between peptide pairs. Analysis of the temporal evolution of these properties, summarized in Figure 3, over the 1335.42 ns window provided insights into the peptide-level organization of the aggregate. The average number of connected components was 1.17 ± 0.40, and the average LCC size was 28.89 ± 2.72 peptides. This confirms the general stability of the large aggregate but also indicates that transient detachment of one or two peptides occurred periodically, as reflected in the fluctuations shown in Figure 3F. The average graph density was 0.138 ± 0.024, and the average weighted density was 2.65 ± 0.32 (Figure 3A). These values suggest a moderately connected network where peptides are not in contact with all neighbors, but significant interactions exist. The average Fiedler value (second smallest eigenvalue of the Laplacian) for the LCC was 2.94 ± 2.07 (for the normalized Laplacian, 0.056 ± 0.041), with its temporal fluctuations shown in Figure 3B. The positive, non-zero values confirm the connectivity of the LCC, and their fluctuations indicate dynamic changes in its structural robustness over time. Centrality measures (weighted degree, betweenness, eigenvector centrality) displayed in Figure 3C and D suggested a relatively homogeneous distribution of importance among peptides, without strong centralization in specific nodes. Community detection using the Louvain algorithm identified an average of 5.23 ± 0.59 communities within the CG graph, with an average modularity of 0.571 ± 0.052 (Figure 3E). This significant modularity suggests that the aggregate is not a uniform assembly but possesses internal sub-structures or domains of more tightly interacting peptides.

3.2.2. *Fine-grained (FG) amino acid graph properties*

The FG graph represented the 150 amino acid residues (5 residues per peptide) as nodes, with weighted edges representing heavy atom contacts between residue pairs (both inter- and intra-peptide). Analysis of the temporal evolution of the FG graph properties, shown in Figure 4, provided a higher-resolution view of the interactions driving assembly. The FG graph was significantly more fragmented than the CG graph, with an average of 13.07 ± 2.27 connected components and an average LCC size of 55.54 ± 17.22 residues (Figure 4H). This is expected, as specific residue-residue contacts are localized, and not all residues participate in a single global network. The FG LCC likely represents the most densely interacting core of residues. The average graph density was 0.0287 ± 0.0026, and the average weighted density was 0.297 ± 0.013 (Figure 4A). These are lower than the CG densities due to the larger number of possible edges in the FG graph. The average Fiedler value for the FG LCC was 1.19 ± 0.75 (for the normalized Laplacian, 0.027 ± 0.018), as shown in Figure 4B, indicating connectivity within the largest residue-

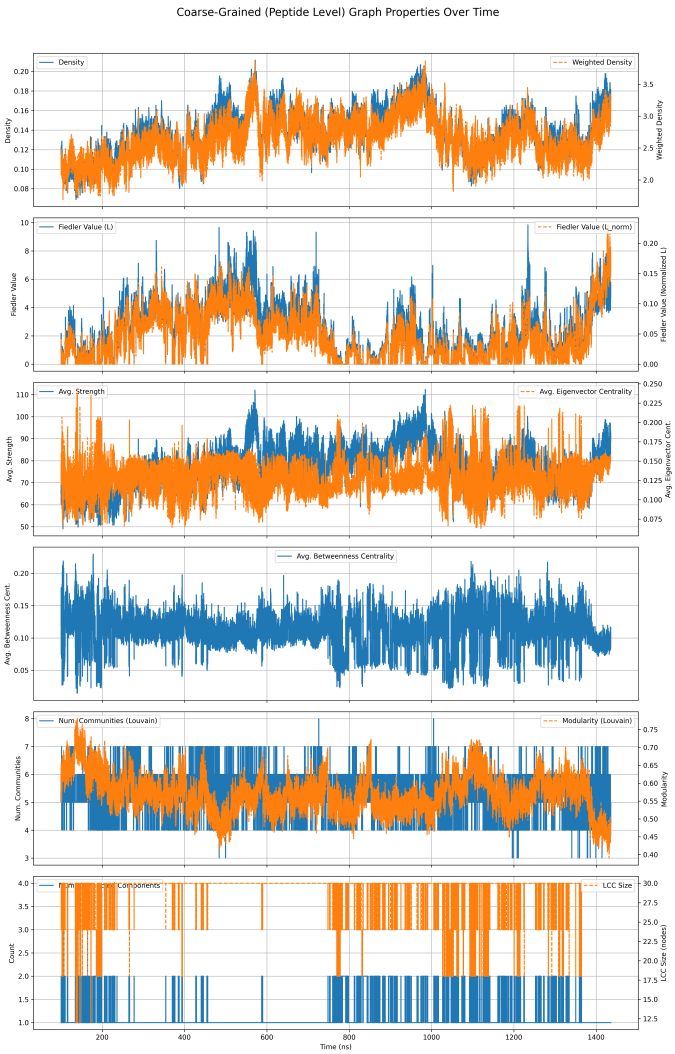

Figure 3. Temporal evolution of coarse-grained (peptide-level) graph properties during the equilibrium phase (100-1435 ns) of KYFIL pentapeptide self-assembly. Plots show the time series of (A) density and weighted density, (B) Fiedler values, (C) average strength and eigenvector centrality, (D) average betweenness centrality, (E) number of Louvain communities and modularity, and (F) number of connected components and largest connected component (LCC) size. These properties reveal the dynamic changes in the aggregate's connectivity, packing, centrality, and internal community structure, demonstrating that the system primarily forms a single large aggregate with transient dissociations.

level cluster. Centrality measures (Figure 4C and D) suggested that while average residue strength was high (44.24 ± 1.94 contacts), betweenness centrality was low, indicating localized interactions rather than extensive bridging by specific residues. Community detection on the FG graph (Figure 4E) yielded an average of 17.60 ± 1.62 communities with a very high average modularity of 0.875 ± 0.027. This extremely high modularity indicates that residue-level interactions are highly specific and form well-defined, tightly interacting clusters, likely corresponding to specific inter-peptide interfaces and stable intra-peptide motifs.

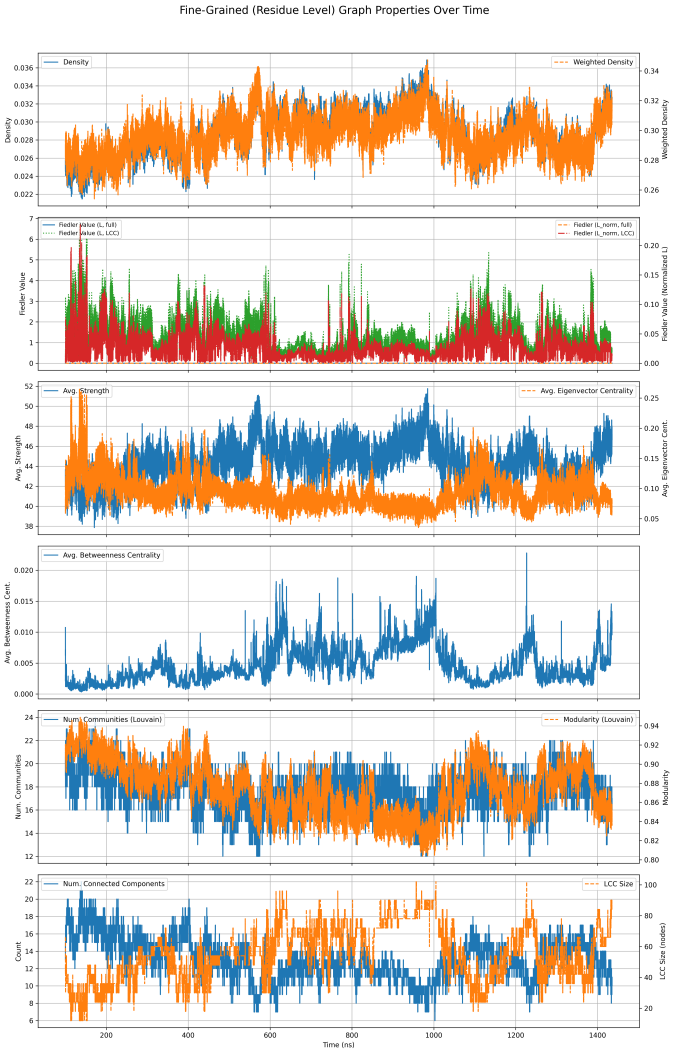

Figure 4. Temporal evolution of fine-grained (residue-level) graph properties of the peptide aggregate from molecular dynamics simulations (100-1435 ns). Time series plots show density, weighted density, Fiedler values, average strength, centrality measures, number of communities, modularity, number of connected components, and largest connected component (LCC) size. These fluctuations highlight the dynamic nature and highly modular internal structure of the aggregate's residue-level interaction network.

The temporal evolution of these CG (Figure 3) and FG (Figure 4) properties revealed that even within the globally stable 30-peptide aggregate, significant dynamic rearrangements were continuously occurring at both the peptide and residue levels.

3.3. *Detection and characterization of aggregate splitting events*

To quantify instances of transient instability, splitting events were defined as occurrences where a parent aggregate at frame t (typically the LCC) fragmented into two or more distinct connected components at frame $t + 1$, subject to minimum size criteria for parent and daughter fragments. Over the 66,770 inter-frame transitions analyzed, a total of 1184 splitting events were detected. The characteristics of these splitting events are summarized by the histograms presented in Figure 5. The average size of the parent aggregate undergoing a split was 28.98 ± 3.25 peptides (Figure 5, top left), confirming that most splits involved the main aggregate. Splits predominantly resulted in two fragments (average 2.008 ± 0.087 fragments; Figure 5, top right), with average fragment sizes of 14.44 ± 8.01 peptides (Figure 5, bottom left), indicating variability in how the aggregate partitions. The average duration of a parent aggregate with an exact peptide composition before splitting was short (0.908 ± 6.74 ns; Figure 5, bottom right), suggesting that the LCC is highly dynamic in its peripheral peptide membership, with frequent small detachments and reattachments contributing to the detected 'splitting' events.

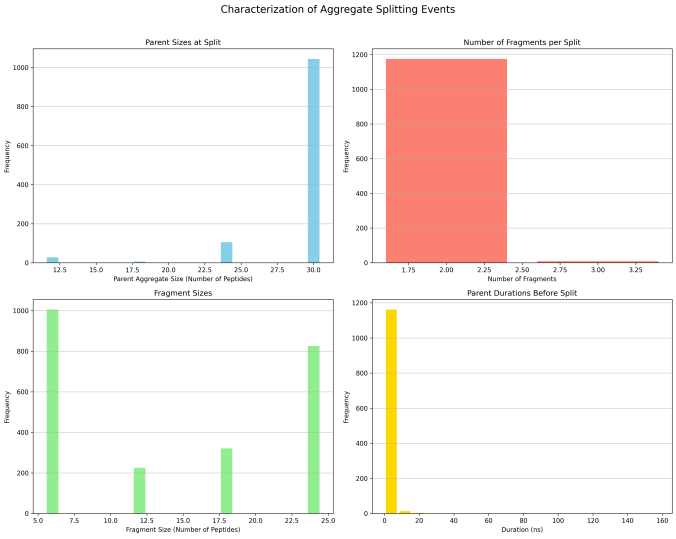

Figure 5. Histograms characterizing aggregate splitting events. The plots show the size of parent aggregates at split (top left), number of fragments per split (top right), sizes of daughter fragments (bottom left), and duration of parent aggregate composition before splitting (bottom right). The data indicate splits usually involve the largest aggregate, primarily yield two fragments, and often follow short periods of stable parent composition, demonstrating the dynamic nature of the assembly.

3.4. *Predictive signatures of aggregate instability*

To assess the predictive power of graph properties, the temporal evolution of key metrics was analyzed in a 1 ns window (50 frames) immediately preceding each splitting event and compared to control windows from stable periods.

3.4.1. *Coarse-grained (CG) graph properties*

The CG Fiedler value of the parent aggregate showed a statistically significant decrease in the 1 ns before splitting (mean 2.46) compared to control windows (mean 3.10, Mann-Whitney U p-value 4.86×10^{-44}). This trend, illustrated in Figure 6, indicates a weakening of the overall structural connectivity of the peptide network preceding fragmentation. Similarly, the CG density of the parent aggregate decreased significantly before splitting (mean 0.129) compared to control periods (mean 0.141, p-value 1.62×10^{-36}). As shown in Figure 7, this suggests a loosening of peptide packing or reduction in contact numbers leading up to the split. These results demonstrate that a decline in peptide-level connectivity (Figure 6) and density (Figure 7) serves as a quantitative precursor to aggregate splitting.

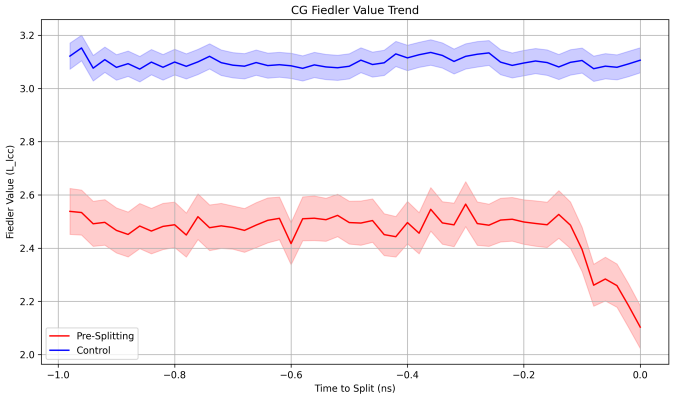

Figure 6. Average coarse-grained (CG) Fiedler value of the largest connected component (LCC) in the 1 ns window before aggregate splitting events (red) and in control windows (blue). The decrease in Fiedler value prior to splitting indicates reduced peptide-level connectivity, serving as a predictive signature.

3.4.2. *Fine-grained (FG) graph properties and bridging interactions*

The FG Fiedler value of the residue-level LCC within the parent aggregate showed a statistically significant *increase* before splitting (mean 1.42) compared to control periods (mean 1.16, p-value 1.21×10^{-13}), as depicted in Figure 8. While counterintuitive for overall weakening, this might suggest that as the peptide aggregate loosens (indicated by CG properties), the most

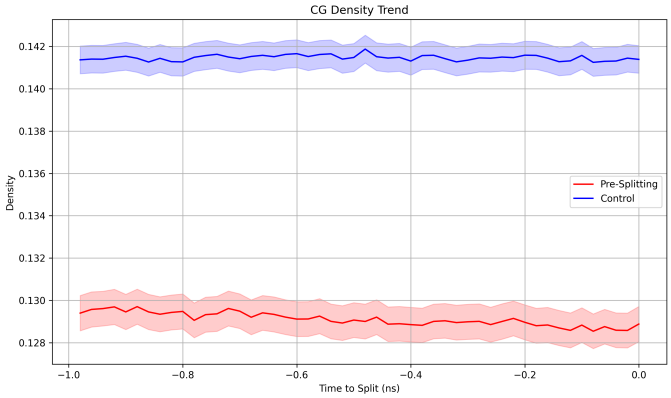

Figure 7. Temporal evolution of coarse-grained (CG) graph density in the 1 ns window before aggregate splitting events (red) and in control periods (blue). Shaded regions show standard deviation. The density significantly decreases leading up to a split, indicating a reduction in peptide-peptide contacts and loosening of the aggregate.

densely interacting residue core compacts or reorganizes, potentially stressing interfaces elsewhere.

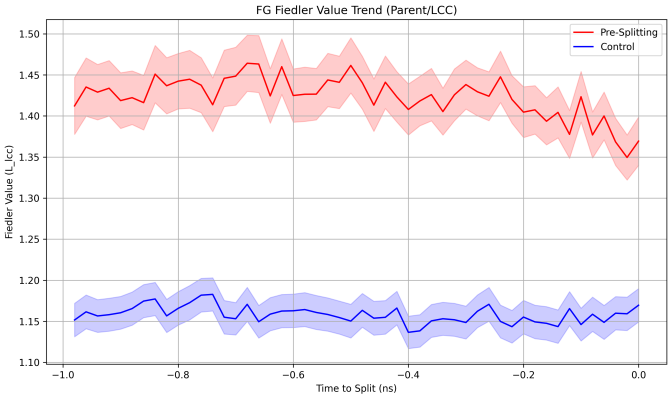

Figure 8. Temporal trend of the Fine-Grained (residue) graph Fiedler value (L_{lcc}) for the largest connected component within the parent aggregate before splitting (red) and in control windows (blue). The increased FG Fiedler value before splitting suggests internal residue-level reorganization.

Crucially, the strength of inter-peptide residue-residue interactions *bridging* the two largest future fragments was analyzed. This metric, calculated from the FG graph by summing weights of edges connecting residues in the future fragment sets, showed a clear and continuous decrease over the 1 ns window leading up to the split, as shown in Figure 9. The average bridging strength dropped significantly from the start to the end of the pre-splitting window, directly demonstrating the weakening of the specific interface that eventually ruptures. This decrease in bridging interactions (Figure 9)

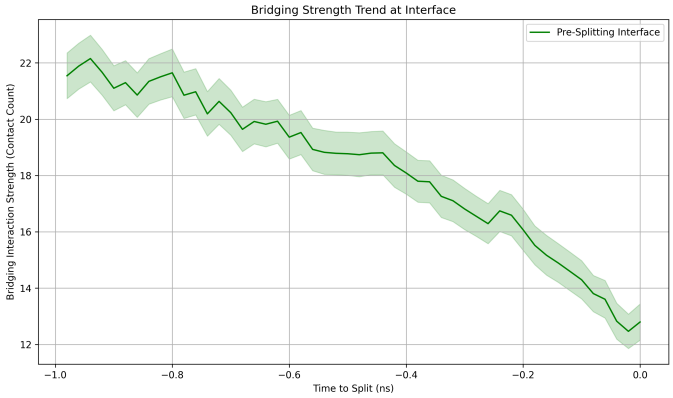

Figure 9. Average strength of residue-level bridging interactions across the interface of future aggregate fragments, plotted over the 1 ns preceding a splitting event (time = 0 ns). The clear decrease in contact count indicates weakening of the interface that ruptures, serving as a predictive signature of instability.

provides a direct, localized predictor of where and when the aggregate is likely to break.

These findings establish that dynamic changes in specific graph theoretical properties, particularly decreases in CG Fiedler value (Figure 6), CG density (Figure 7), and FG bridging interaction strength (Figure 9), are detectable precursors that can predict aggregate splitting events within a nanosecond timescale.

3.5. *Molecular determinants of stability and splitting*

To identify which specific residue-residue interactions are critical for maintaining aggregate stability and whose weakening drives splitting, the change in contact counts between different amino acid types across the splitting interface was analyzed over the 1 ns pre-splitting window. Analysis of the average change in inter-peptide residue-residue contacts between the two largest future fragments revealed a consistent trend of decreasing contact numbers for specific residue pairs, as depicted in the heatmap in Figure 10. Prominently, contacts between hydrophobic residues such as Phenylalanine (PHE_PHE, -0.4 average change), Isoleucine (ILE_ILE, -0.2), and Leucine (LEU_LEU, -0.2) showed significant reductions across the interface before splitting. Interactions involving aromatic residues, such as TYR_PHE (-0.3), also decreased. Contacts involving Lysine (LYS_TYR, -0.2; LYS_PHE, LYS_ILE, LYS_LEU) also tended to decrease, suggesting potential disruption of polar/charged interactions or cation-pi interactions at the interface.

This detailed residue-level analysis, summarized in Figure 10, indicates that aggregate splitting is primarily driven by the weakening of non-covalent interactions

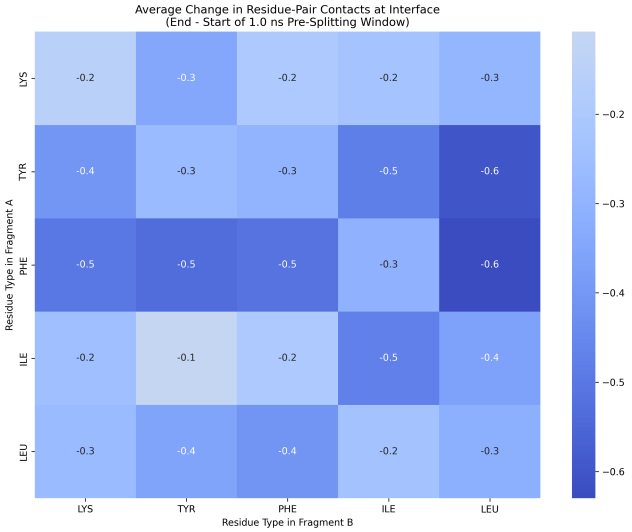

Figure 10. Average change in residue-pair contacts across the interface in the 1 ns preceding aggregate splitting. The heatmap shows the average change in heavy atom contacts (< 4.5Å) between residue types located in the two largest fragments that form after a split. Blue indicates a decrease in contacts. This highlights the weakening of inter-fragment interactions, particularly involving hydrophobic and aromatic residues (F, I, L, Y), at the interface before fragmentation.

that form the core cohesive forces of the assembly. The significant decrease in contacts between hydrophobic and aromatic residues (F, I, L, and Y) at the interface directly implies a loss of favorable van der Waals and pi-stacking interactions in the region destined to break. While the average change per specific residue pair might seem small, these values represent the average over many splitting events and many potential residue pairs at an interface. The consistent negative trend across multiple hydrophobic/aromatic combinations highlights the collective importance of these interaction types.

These results suggest that the stability of KYFIL aggregates relies heavily on the integrity of their hydrophobic and aromatic cores. Any process that disrupts the favorable packing and contact networks of F, I, L, and Y residues is likely to render the aggregate prone to fragmentation.

3.6. *Summary of findings*

In summary, the hierarchical graph analysis provided a multi-scale quantitative description of KYFIL peptide aggregates. At the peptide level, the aggregate forms a moderately connected network with discernible internal communities, as shown by the CG graph properties (Figure 3). At the residue level, interactions are highly specific and localized, forming dense, modular clusters, evident from the FG graph properties (Figure 4). Aggregate splitting events, representing transient fragmentations of the main assembly (Figure 1, Figure 2, Figure 5), were frequent. Crucially, these events were predictable from dynamic changes in graph properties within a 1 ns window: decreases in CG Fiedler value (Figure 6) and density (Figure 7) indicated overall structural loosening, while a decrease in FG bridging contacts (Figure 9) directly pinpointed the weakening of the future rupture interface. Molecular analysis at the splitting interface (Figure 10) revealed that the loss of contacts between hydrophobic and aromatic residues (F, I, L, Y) is a key molecular determinant driving aggregate instability. This framework provides a powerful tool to link molecular interactions to macroscopic assembly dynamics and offers valuable insights for tuning peptide sequence and structure for desired material properties.

4. CONCLUSIONS

Understanding and controlling the stability and dynamic behavior of peptide self-assemblies is paramount for their successful application in biomaterials and nanotechnology. Predicting aggregate instability and identifying the underlying molecular forces driving breakdown from complex molecular dynamics trajectories has remained a significant challenge. This study addressed this by developing and applying a novel framework based on dynamic, weighted, hierarchical graph analysis to investigate the equilibrium behavior of KYFIL pentapeptide aggregates simulated over 1.3 μs.

Our approach utilized two levels of graph representation: a coarse-grained peptide graph capturing peptide-level connectivity and a fine-grained amino acid graph detailing residue-residue interactions. By analyzing the temporal evolution of graph theoretical properties, we were able to quantitatively characterize the dynamic structure of the aggregates and identify instances of transient fragmentation.

The analysis revealed that while the KYFIL system predominantly forms a single large aggregate, it undergoes frequent, transient splitting events. Crucially, we demonstrated that specific dynamic changes in graph properties serve as robust predictive signatures for these impending splitting events within a nanosecond timescale. A statistically significant decrease in the coarse-grained aggregate's connectivity, as measured by the Fiedler value of the Laplacian, and a reduction in its density were observed to precede fragmentation. Complementary analysis at the fine-grained level showed a marked decline in the collective strength of residue-residue contacts specifically bridging the future fragments, providing a localized predictor for where the aggregate would break.

Furthermore, by examining the types of residue-residue interactions that weakened at the splitting interfaces, we identified key molecular determinants governing aggregate instability in this system. The significant reduction in contacts between hydrophobic and aromatic residues, including phenylalanine, isoleucine, and leucine, directly preceding splitting events highlighted the critical role of non-polar interactions in maintaining the aggregate's cohesive structure. The weakening of these specific interactions constitutes the molecular-level event that drives the macroscopic fragmentation.

In summary, this hierarchical graph-based framework provides a powerful quantitative tool that successfully bridges the gap between molecular-level interactions and macroscopic aggregate dynamics. We have shown that dynamic changes in graph properties can predict aggregate instability and that analysis of fine-grained interactions at the nascent splitting interface can pinpoint the specific molecular forces responsible. This work offers valuable insights for the rational design of self-assembling peptides, enabling the tuning of sequence and structure to control aggregate stability and tailor material properties for specific applications.

REFERENCES

Airale, L., Longa, A., Rigon, M., Passerini, A., & Passerone, R. 2025, Simple Path Structural Encoding for Graph Transformers. https://arxiv.org/abs/2502.09365

Boniol, P., Tiano, D., Bonifati, A., & Palpanas, T. 2025, k-Graph: A Graph Embedding for Interpretable Time Series Clustering. https://arxiv.org/abs/2502.13049

Coupette, C., Wayland, J., Simons, E., & Rieck, B. 2025, No Metric to Rule Them All: Toward Principled Evaluations of Graph-Learning Datasets. https://arxiv.org/abs/2502.02379

Huang, Y., & Shuai, J. 2013, Polarization effect of zinc on the region 1-16 of amyloid-beta peptide: a molecular dynamics study. https://arxiv.org/abs/1301.1593

Krasnokutski, S. A., Jager, C., Henning, T., et al. 2024, Formation of extraterrestrial peptides and their derivatives, doi: https://doi.org/10.1126/sciadv.adj7179

Liang, T., Xu, K., Lindgren, E., et al. 2025, NEP89: Universal neuroevolution potential for inorganic and organic materials across 89 elements. https://arxiv.org/abs/2504.21286

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

Prouteau, T., Dugué, N., & Guillot, S. 2024, From communities to interpretable network and word embedding: an unified approach, doi: https://doi.org/10.1093/comnet/cnae034

Radice, D., Gamba, R., Zhu, H., & Rashti, A. 2025, AthenaK simulations of the binary black hole merger GW150914. https://arxiv.org/abs/2506.06838

Ramos-Osuna, V., Díaz-Álvarez, A., & Lara-Cabrera, R. 2025, Efficient n-body simulations using physics informed graph neural networks. https://arxiv.org/abs/2504.01169

Reid, F., McDaid, A., & Hurley, N. 2012, Partitioning Breaks Communities. https://arxiv.org/abs/1105.5344

Roncoli, A., Ćiprijanović, A., Voetberg, M., Villaescusa-Navarro, F., & Nord, B. 2024, Domain Adaptive Graph Neural Networks for Constraining Cosmological Parameters Across Multiple Data Sets. https://arxiv.org/abs/2311.01588

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

Wang, X. 2025, A Python Toolkit for Plotting Double Star Observations with 1:1 Aspect Ratio, doi: https://doi.org/10.5281/zenodo.14889911

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

Zheng, Y., Yang, Y., kun Zhang, Y., et al. 2025, PANCAKE: Python bAsed Numerical Color-magnitude-diagram Analysis pacKagE. https://arxiv.org/abs/2505.04534