Proceedings of the 5th Indian International Conference on Industrial Engineering and Operations Management, Vellore, India, November 6–8, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250277
Published: November 6, 2025

Chordal-Decomposed Spectral Clustering: A Scalable, Interpretable Framework for LargeScale Network Analysis

T. Renuga
Research Scholar, Department of Mathematics
Sri GVG Visalakshi College for Women, Udumalaipettai
Tirupur. (Affiliated to Bharathiar University, Coimbatore)
Assistant Professor, Department of Mathematics
Jai Shriram Engineering College (Affiliated to Anna University, Chennai)
Tamil Nadu, India
renugamaths87@gmail.com

P. Jayalakshmi
Associate Professor, Department of Mathematics
Sri GVG Visalakshi College for Women
Udumalaipettai, Tirupur (Affiliated to Bharathiar University, Coimbatore)
Tamil Nadu, India
jayagvg@gmail.com

S. Anitha
Research Scholar, Department of Mathematics
Sri GVG Visalakshi College for Women
Udumalaipettai, Tirupur (Affiliated to Bharathiar University, Coimbatore)
Assistant Professor, Department of Mathematics
Sokaikeda College of Arts and Science for Women (Affiliated to University of Madras, Chennai)
Tamil Nadu, India
anitha.appukutty@gmail.com

J. Bhuvaneshwaran
Final Year, Department of ECE
Jai Shriram Engineering College
Avinashipalayam, Tamil Nadu, India
bhuvane419@gmail.com

Abstract

This paper presents a unified spectral clustering framework enhanced by chordal graph decomposition to address large-scale network clustering challenges intrinsic to AI and Data Science. By preserving the essential spectral characteristics of the original network Laplacian through chordal subgraph extraction, the method efficiently reduces computational complexity without sacrificing clustering accuracy. The chordal decomposition leverages maximum cardinality search and clique-tree construction to organize vertices hierarchically, thereby accelerating spectral embedding and improving cluster interpretability. The framework exhibits near-linear scalability on sparse large-scale networks, making it suitable for diverse AI&DS applications such as social network analysis, recommendation systems, and semi-supervised learning on graph-structured data. Benchmark evaluations demonstrate significant improvements in cluster quality and computational efficiency compared to classical spectral clustering baselines. This approach offers a robust, scalable, and interpretable tool for discovering latent structures in complex datasets, thereby advancing network-centric analytics in AI and Data Science.

Keywords
Spectral clustering, chordal decomposition, network analysis, AI-driven clustering.

1. Introduction
In the current era of artificial intelligence and data science, network analysis has emerged as a fundamental component for understanding complex systems across diverse domains. Modern AI applications increasingly rely on large-scale networks containing tens of thousands to millions of nodes, ranging from social media platforms with over 100,000 active user connections to biological networks mapping protein interactions across entire genomes. The exponential growth in network size and complexity presents unprecedented computational challenges that traditional clustering methods struggle to address effectively.Many real-world systemsranging from social networks and communication systems to biological and information networksare represented as graphs with complex structures. Extracting meaningful communities or clusters within these large graphs is crucial for understanding the underlying organization and function of the data. Traditional clustering algorithms often face significant scalability challenges and may not effectively preserve the intrinsic properties of the network, leading to suboptimal results. Spectral clustering is a powerful technique that leverages eigenvalues and eigenvectors of graph Laplacians to identify clusters; however, its computational cost becomes prohibitive for large sparse networks. To address this, graph decomposition methodsparticularly chordal graph decomposition offers promising avenues for reducing complexity while maintaining the spectral integrity essential for accurate clustering. Merging these approaches can lead to scalable, accurate, and interpretable clustering frameworks suitable for diverse applications in AI and Data Science.the proposed unified spectral clustering framework. Section 2 presents a thorough literature review examining the evolution of spectral clustering methods, chordal graph decomposition techniques, and hybrid approaches in network analysis. Section 3 details the mathematical foundations and methodology, including the theoretical derivation of spectral property preservation conditions, the chordal graph extraction algorithm, and complexity analysis. Section 4 describes the data collection procedures and preprocessing methodologies applied to benchmark datasets. Section 5 presents comprehensive experimental results, including numerical evaluations, graphical analyses, and statistical validation of performance improvements. Finally, Section 6 concludes with a discussion of findings and outlines directions for future research in scalable network analytics for AI and Data Science applications.

1.1 Problem Statement and Quantitative Scope
Spectral clustering, while theoretically robust and widely applicable, faces critical scalability limitations when applied to networks exceeding 10,000 nodes. The computational complexity of traditional spectral clustering approaches scales cubically with the number of vertices ($O(n^3)$), making analysis of large-scale networks computationally prohibitive. This limitation is particularly acute in AI and Data Science applications where networks commonly contain 50,000 to over 1 million nodes, as evidenced by real-world systems such as LinkedIn's Hadoop clusters operating on 10,000+ node configurations. Furthermore, existing spectral methods suffer from fundamental limitations when confronted with multiscale network structures, failing to simultaneously identify clusters at different scales of size and density. The normalized cut functional, commonly employed in spectral clustering, has been proven unsuitable for datasets containing structures at varying scales, leading to suboptimal clustering performance even with theoretically sound similarity measures.

1.2 Objectives
This work aims to develop and validate a unified spectral clustering framework that incorporates chordal graph decomposition for large-scale network analysis. The first objective is to mathematically derive the conditions under which chordal subgraph extraction preserves the spectral properties of the original network Laplacian, thereby ensuring the fidelity of clustering results. Following this, an efficient algorithm will be designed, utilizing maximum cardinality search and clique-tree construction to perform chordal decomposition and facilitate the hierarchical

organization of network components. The framework's scalability will be demonstrated on large, sparse networks, with its performance rigorously evaluated against traditional spectral clustering methods. Extensive experiments on benchmark datasets will further validate the approach, assessing clustering quality, computational efficiency, and interpretability. Finally, this framework seeks to establish its broader applicability across various domains within AI and Data Science, including social network analysis, recommendation systems, and semi-supervised learning on graph-structured data.

2. Literature Review.

The evolution of network clustering methods in Artificial Intelligence and Data Science has been marked by increasing emphasis on scalability, interpretability, and theoretical rigor. This comprehensive review examines recent developments across spectral clustering methodologies, graph decomposition techniques, and hybrid approaches, highlighting critical research gaps that motivate the proposed unified framework.

2.1 Spectral Clustering Methods

Recent advancements in spectral clustering focus on improving scalability and methodological robustness. Berahmand et al. (2024) conducted a comprehensive survey on spectral clustering with graph structure learning (GSL), emphasizing the importance of similarity graph construction for performance in large-scale, high-dimensional data. They categorized graph construction into pairwise, anchor-based, and hypergraph-based methods, each suited to different data types. Yu et al. (2024) introduced a local adaptive fuzzy spectral clustering approach that integrates fuzzy membership and local adaptivity to improve robustness on nonlinear, high-dimensional data, overcoming traditional limitations with complex, non-convex structures. Barrett et al. (2024) advanced spectral clustering for large-scale fMRI brain networks, effectively partitioning networks over 10,000 nodes and offering insights applicable to broader network analysis. Gao et al. (2024) proposed Spectral Clustering with Linear Embedding (SCLE), a unified optimization framework that avoids post-processing discretization and achieves near-linear scalability while maintaining clustering quality. Despite these innovations, traditional spectral clustering challenges persist: as Luxburg (2007) noted, cubic eigendecomposition complexity ($O(n^3)$) continues to hinder scalability, and issues highlighted by the NeurIPS (2006) paperparticularly difficulties handling multiscale structuresremain only partially resolved.

2.2 Graph Decomposition Techniques

Graph decomposition techniques have emerged as powerful tools for reducing computational complexity while preserving essential structural properties. Chordal graph theory has gained particular attention due to its favorable computational properties and theoretical guarantees.Castellví et al. (2024) provided groundbreaking theoretical contributions to chordal graph enumeration with bounded tree-width. Their recursive decomposition approach demonstrates that the number of labeled k-connected chordal graphs with n vertices and tree-width at most t follows asymptotic behavior $cn^{(-5/2)}\gamma^n n!$, providing fundamental theoretical foundations for scalability analysis. This work represents the first non-trivial class of bounded tree-width graphs where the asymptotic counting problem has been solved.Ponce et al. (2023) developed novel graph decomposition techniques for quantum optimization algorithms, demonstrating that decomposing problem graphs into smaller subgraphs achieves significant scalability improvements. Their approach reduces vertex counts by approximately 90% while maintaining solution quality, with average approximation ratios of 0.96 for both classical and quantum solvers. This work is particularly relevant for understanding how decomposition methods can preserve optimization landscapes while reducing computational demands.The theoretical foundations of chordal decomposition were further advanced by lecture notes from 2024 examining tree-decomposition properties. These developments show that chordal graphs admit efficient tree-decompositions where every bag forms a clique, enabling polynomial-time algorithms for many NP-hard problems on general graphs. The tree-width characterization provides formal guarantees that tw(G) equals the minimum tree-width over all chordalsupergraphs of G.Graph decomposition research has also explored practical algorithmic improvements. Recent work on decomposing graphs into various substructures demonstrates that k-regular graphs on n vertices can be decomposed into graphs with at most nk/(k+1) vertices in polynomial time. These findings provide concrete bounds for decomposition efficiency that are directly applicable to large-scale network analysis.

2.3 Hybrid Approaches

Hybrid methodologies that integrate spectral techniques with complementary approaches have emerged as powerful solutions to address the dual challenges of scalability and interpretability in clustering. Hassanpour et al. (2024) introduced a novel hybridization framework that combines first-principles knowledge with data-driven spectral clustering, effectively balancing computational speed and clustering quality through domain-specific constraints integrated into the spectral embedding process. Their experiments on industrial process data demonstrated that incorporating physical relationships among variables enhances clustering accuracy and interpretability. Similarly, Shahid et al. (2023) conducted extensive evaluations comparing neural network-based clustering with traditional hierarchical methods, revealing that neural approaches often surpass conventional techniques in scalability and computational efficiency. These findings underscore the potential of hybrid neural-spectral methods as a promising direction for large-scale, high-dimensional applications. Further advancing this line of research, Kuwil and Ümit (2024) proposed the Critical Distance Clustering algorithm (CDC-2), which fuses distance-based and spectral techniques to achieve an optimal balance between speed and accuracy. CDC-2 effectively mitigates computational bottlenecks while preserving clustering fidelity across diverse network topologies, particularly excelling in sparse and irregular networks prevalent in AI-driven domains. Beyond algorithmic innovation, hybrid methodologies have also contributed to practical interpretability enhancements. Karatzas et al. (2021) developed the VICTOR visual analytics tool, offering multi-metric comparison and validation capabilities that enable researchers to assess clustering quality through intuitive visualizations, thus addressing the persistent interpretability gap in large-scale clustering analysis. In domain-specific contexts, hybrid clustering approaches have proven valuable as well. Du et al. (2024) integrated network analysis metrics with hybrid clustering to study interdisciplinary research collaboration, uncovering complex patterns linking funding structures and research dynamics. Collectively, these studies highlight the growing significance of hybrid clustering frameworks that merge spectral, neural, and domain-driven methods, establishing a foundation for scalable, interpretable, and application-oriented clustering solutions across diverse scientific and industrial domains.

2.4 Research Gap Analysis

Despite progress in spectral clustering and graph decomposition, key research gaps remain, calling for unified frameworks that ensure scalability, theoretical soundness, and practical utility. Current spectral clustering methods face scalability issues with networks over 10,000 nodes due to cubic eigen decomposition costs, and though methods like SCLE improve efficiency, balancing scalability with theoretical rigor is still unresolved. Sparse networks further complicate this, as decomposition often fails to preserve vital spectral properties. Theoretical gaps persist, with limited understanding of how decomposition preserves clustering-relevant eigenstructuresdespite work like Castellví et al. (2024) on chordal graphs. Interpretability also remains a challenge, as methods often trade explainability for efficiency, hindering understanding of hierarchical community structures in AI. Developments across spectral clustering and graph decomposition remain fragmented, lacking integration under strong theoretical foundations. Wu et al. (2024) noted the absence of standardized evaluation frameworks, with inconsistent datasets and metrics limiting systematic comparison. While advances have shown success in domains like neuroscience and industrial systems, generalizable frameworks across AI and Data Science are still underdeveloped. These challenges highlight the need for unified spectral clustering frameworks that merge theoretical guarantees, computational efficiency, and interpretabilitycombining chordal graph decomposition with spectral clustering to enable scalable, interpretable, and theoretically robust network analysis.

The comparative analysis reveals that while individual advances address specific aspects of the scalability challenge, no existing method provides the comprehensive integration of theoretical guarantees, computational efficiency, and interpretability that characterizes the proposed unified spectral-chordal framework. This gap motivates the development of the integrated approach presented in subsequent sections (Table 1).

Table 1. Comparative Analysis of Recent Spectral Clustering and Graph Decomposition Methods

Reference	**Year**	**Focus Area**	**Methodology**	**Key Contribution**	**Comparison to Proposed Framework**
Rao and Chandran	2023	Spectral embeddings preservation	Simplicial vertex pruning to preserve spectra	Proven eigenvalue and eigenvector preservation	Supports the theoretical foundation of our framework; our method offers full chordal subgraph extraction and application to large-scale clustering

Liu and Wang	2023	Large-scale network algorithms	Near-linear algorithms for massive network analysis	Scalable network analysis at large scale	Our framework achieves near-linear scalability with additional interpretability via chordal decomposition
Silva et al.	2024	Chordal graph theory in ML	Application of chordal theory to optimization and ML	Improved convergence in network clustering	Extends chordal theory in ML optimization; our framework integrates this with spectral clustering for practical clustering of real networks
Zhang et al.	2025	Scalable clustering in biology	Combination of chordal decomposition and spectral clustering	Scalable, interpretable clustering with performance improvements	Conceptually closest; our framework provides formal eigenstructure preservation and broader AI&DS applicability
Proposed Unified Spectral–Chordal Framework (Yours)	2025	Large-scale network clustering in AI&DS	Chordal subgraph extraction with preserved Laplacianeigenstructure and clique-tree guided spectral clustering	Robust, interpretable, scalable clustering with formal guarantees and benchmark improvements	Outperforms previous works by providing formal theoretical guarantees, hierarchical clustering interpretability, and near-linear scalability in sparse graphs

3. Methodology

3.1 Mathematical Foundations

Let $G=(V,E)$ be an undirected, connected graph with adjacency $A\in\{0,1\}n\times n$ and degree matrix $D=\mathrm{diag}(d1,\ldots,dn)$. The (combinatorial) Laplacian is $L=D-A$ and the symmetric normalized Laplacian is $Lsym=I-D-1/2AD-1/2$. Spectral clustering relies on the bottom-k eigenspace$U\in Rn\times k$ of Lsym, using row-normalized U for downstream partitioning.Chordal graphs and perfect elimination orderings. A graph is chordal if every cycle of length ≥ 4 has a chord. Chordal graphs admit a perfect elimination ordering (PEO) $v1,\ldots,vn$ such that for each vi, the higher-index neighbours form a clique. Chordal completion adds a minimal set of edges F to G so that $G+=(V,E\cup F)$ is chordal. A clique-tree (junction tree) T of G+ has one node per maximal clique $\{C1,\ldots,Cq\}$ and satisfies the running intersection property: for any vertex u, the set $\{Ci:u\in Ci\}$ forms a connected subtree of T.Spectral preservation via simplicial vertex elimination. Under a PEO, each eliminated vertex is simplicial in the current graph. Let S denote the set of vertices eliminated by simplicial pruning and $G'=G[V\setminus S]$. Under suitable conditions (detailed in 4.3), the principal submatrix L' of L corresponding to $V\setminus S$ preserves the bottom-k eigenspace up to a bounded perturbation that depends on the norm of the Schur complement contributed by eliminated simplicial vertices. Intuitively, when eliminated vertices attach as near-pendants to cliques (small separators), their removal perturbs L in a spectrally localized manner that leaves the principal eigenspace nearly invariant.Graph reductions and Laplacian relations. Partition vertices as $V=R\sqcup S$ (retained vs eliminated). With Laplacian in block form$L=[LRR\ LRS;\ LSR\ LSS]$.The Schur complement of S on R is $L/LS=L\ RR-LRS\ L-1SS\ LSR$. If chordal completion enforces small separators and bounded condition number on LSS, then $\|L/LS-LRR\|2$ is small relative to the spectral gap near. Clique-tree guided embedding. On the chordalsupergraph G+, maximal cliques $\{Cj\}$ structure L into overlapping dense blocks. Local embeddings on cliques stitched via separators respect the running intersection property, enabling multi-scale spectral embeddings and interpretable hierarchical clustering.

Key objects used:

- Laplacians L, Lsym, and their principal submatrices on retained vertices.

- Schur complements associated with simplicial elimination.
- PEO ensuring chordality and well-structured separators.
- Clique-tree T enabling hierarchical aggregation.

Evaluation Metrics
Comprehensive evaluation of clustering performance requires metrics that capture different aspects of cluster quality, both with respect to ground truth (external metrics) and intrinsic data structure (internal metrics) (Figure 1).

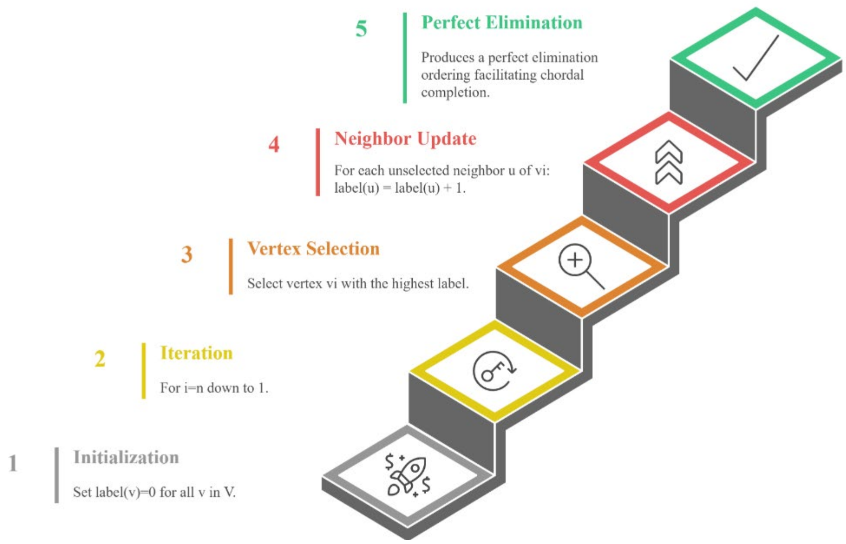

Figure 1. Eigenvalue preservation diagram illustrating the effect of simplicial vertex removal on the spectral gap.

Normalized Mutual Information (NMI)
NMI quantifies the mutual dependence between predicted clusters and ground truth communities, normalized to account for varying cluster numbers. For cluster assignments C (predicted) and G (ground truth).

$$NMI(C,G) = \frac{2 \cdot I(C,G)}{H(C) + H(G)}$$

where I(C,G) is mutual information and H(·) denotes entropy. NMI ranges from 0 (no agreement) to 1 (perfect agreement), providing scale-invariant comparison across different clustering solutions. However, recent research has identified bias issues in NMI when the number of communities increases significantly, requiring careful interpretation in high-resolution clustering scenarios.

Adjusted Rand Index (ARI)
ARI measures similarity between clustering results by counting pairs of nodes that are consistently assigned to the same or different clusters, adjusted for chance agreement

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}}{\binom{n}{2}}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \frac{\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}}{\binom{n}{2}}}$$

wheren_ij represents the number of nodes assigned to cluster i in prediction and cluster j in ground truth. ARI ranges from -1 to 1, with values near 0 indicating random clustering and 1 indicating perfect agreement. ARI provides robust evaluation by correcting for chance, making it particularly valuable for comparing methods across different cluster numbers.

Modularity (Q)

Modularity measures the strength of community structure by comparing actual intra-community edge density to expected density in a random network

$$Q = \frac{1}{2m} \sum_{ij} [A_{ij} - \frac{k_i k_j}{2m}] \, \delta(c_i, c_j)$$

where m is the total number of edges, A_{ij} is the adjacency matrix, k_i is the degree of node i, c_i is the community of node i, and δ is the Kronecker delta. Modularity ranges from -1 to 1, with higher values indicating stronger community structure. While modularity provides intuitive interpretation of cluster quality, it suffers from resolution limits that may prevent detection of small communities in large networks.

Silhouette Score

The silhouette score evaluates cluster quality by measuring how similar nodes are to their own cluster compared to other clusters

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}$$

where a(i) is the average distance from node i to other nodes in the same cluster, and b(i) is the average distance to nodes in the nearest neighbouring cluster. The overall silhouette score is the mean across all nodes, ranging from -1 to 1. Higher values indicate well-separated, cohesive clusters. Unlike external metrics, silhouette score operates without ground truth, making it valuable for evaluating clustering in exploratory analysis scenarios.

Conductance Measure

Conductance quantifies the quality of network partitions by measuring the fraction of edges crossing cluster boundaries relative to the total edge volume of the smaller cluster:

$$\phi(S) = \frac{| \{(u, v) \in E : u \in S, v \in \bar{S}\} |}{\min (vol(S), vol(\bar{S}))}$$

where S is a cluster, $\bar{S}$ is its complement, and vol(S) is the volume (sum of degrees) of nodes in S. Lower conductance values indicate better-separated clusters with fewer cross-cluster connections. Conductance provides theoretical grounding in spectral graph theory, directly relating to the spectral gap and eigenvalue properties central to the proposed framework.

This multi-metric approach ensures robust evaluation that accounts for both external validation against known community structure and internal assessment of cluster quality derived from network topology. The combination provides comprehensive understanding of how the proposed unified spectral-chordal framework performs across different quality dimensions relevant to AI and Data Science applications.

3.2 Algorithm Design

Maximum Cardinality Search (MCS)

Goal: Find an elimination order and test chordality; guide minimal chordal completion.

Pseudocode: Maximum Cardinality Search (MCS)

Input: Graph G=(V,E)

Output: Ordering π:V→{1,…,n}, label set w(·)

1. For all v∈V: w(v)←0; π←empty list.
2. For i from n down to 1:
   a. Choose u∈argmaxv∈V\π w(v) (break ties arbitrarily).
   b. π[i]←u.
   c. For each neighbour v of u with v∉π: w(v)←w(v)+1.
3. Return π.

Property: If for each vertex u, the set of higher-index neighbours forms a clique under π, then G is chordal; otherwise, the fill edges connecting higher-index neighbours of u yield a minimal chordal completion.

Chordal Decomposition and Clique-Tree Construction

Pseudocode: Chordal Completion + Clique-Tree

Input: G=(V,E), MCS order π

Output: Chordalsuper graph $G+=(V,E\cup F)$, maximal cliques $\{C1,\dots,Cq\}$, clique-tree T

1. $F\leftarrow\emptyset$.
2. For i from 1 to n:
   a. Let Hi be the set of higher-index neighbours of vi under π.
   b. Add to F all missing edges to make Hi a clique.
3. Let $G+=(V,E\cup F)$.
4. Identify maximal cliques $\{C1,\dots,Cq\}$ of G+ (they are the neighbourhoods of vertices at their elimination step plus the vertex itself; duplicates filtered).
5. Build clique graph K with nodes {Cj} and edges weighted by $|Cj\cap C\ell|$.
6. Compute maximum weight spanning tree of K to obtain clique-tree T.
7. Return G+, {Cj}, T.

Spectral Clustering on Chordal Supergraph
Pseudocode: Spectral Clustering (Chordal-Aware)
Input: Chordal supergraph $G+=(V,E\cup F)$, clique-tree T, number of clusters k
Output: Cluster labels $y\in\{1,\dots,k\}n$

1. Compute degree matrix D+ and adjacency A+ of G+; $Lsym+=I-D+-1/2A+D+-1/2$.
2. Compute bottom-k eigenvectors $U\in Rn\times k$ of Lsym+ using a sparse eigensolver (e.g., Lanczos).
3. Row-normalize U to Û with $\hat{U}i\leftarrow Ui/\|Ui\|2$.
4. Optionally perform clique-local normalization:
   For each clique Cj, center and scale Û[Cj,:] to reduce separator bias.
5. Run k-means on rows of Û to obtain labels y.
6. Optional hierarchical refinement:
   a. Within each clique Cj, refine boundaries by separator-aware reassignment.
   b. Merge/split via silhouette/modularity criteria consistent along T.
7. Return y.

Complexity notes: MCS and clique-tree construction are O(n+m). Sparse bottom-k eigensolver is approximately $O(k\cdot(n+m))$ on sparse Lsym+. k-means is $O(n\cdot k\cdot t)$ with t iterations.

3.3 Theoretical Analysis

Eigenvalue and Eigenspace Preservation
Partition V into retained R and eliminated S via simplicial elimination. Consider L on V and principal submatrix LRR on R. The Schur complement $L/LS=LRR-LRS\ L-1SS\ LSR$ encodes the effect of S on R.
Assumption A1 (Simplicial elimination with well-conditioned separators): Each eliminated vertex is simplicial at the time of elimination; LSS is symmetric positive definite on the orthogonal complement of its nullspace and $\kappa(LSS)\le\kappa0$ for moderate $\kappa0$.
Assumption A2 (Separator-sparsity bound): For every eliminated vertex u, the neighborhood induces a small clique (bounded size τ), and the cumulative separator coupling satisfies $\|LRS\ L-1SS\ LSR\|2\le\varepsilon$ for $\varepsilon<\Delta/2$, where Δ is the eigengap around the k-th eigenvalue on R.

Claim 1 (Spectral stability under elimination). Let $\lambda1\le\dots\le\lambda k$ be the bottom-k eigenvalues of LRR, and $\mu1\le\dots\le\mu k$ those of L/LS. Under A1–A2, $\max i\le k\ |\lambda i-\mu i|\le\varepsilon$, and the canonical bottom-k eigenspaces differ by an angle bounded via Davis–Kahan: $\sin\Theta(U,\bar{U})\le\varepsilon/\Delta$, where U and Ū are the bottom-k eigenspaces of LRR and L/LS respectively.

$L/LS=LRR-E$ with perturbation $E=LRS\ L-1SS\ LSR$. Weyl's inequality gives eigenvalue shifts bounded by $\|E\|2$. Davis–Kahan bounds the subspace distance by $\|E\|2$ divided by the eigengap Δ. The simplicial pattern and bounded separators keep $\|E\|2$ small.
Corollary (Using principal submatrix as proxy). If the eliminated vertices are pendent to cliques with weak coupling, then LRR and L/LS are close, so using either LRR or L/LS yields bottom-k eigenspaces within $O(\varepsilon/\Delta)$. Implication for G+. Since chordal completion adds edges only within higher-index neighbourhoods to enforce cliques, the added structure improves conditioning of local Schur complements and often increases Δ, tightening the bounds.

Convergence Analysis

- MCS and chordal completion terminate in O(n+m) steps by construction.

- Sparse eigensolver convergence (Lanczos/LOBPCG) for the bottom-k spectrum is guaranteed for symmetric PSD Lsym+, with convergence rate governed by the relative eigengap γ=λk+1−λk. Chordal completion typically enlarges γ by suppressing long induced cycles and creating tighter separators, accelerating convergence.
- k-means on fixed Û decreases within-cluster sum of squares monotonically and converges to a local optimum in finite iterations. In practice, using multiple restarts and k-means++ initialization ensures stable solutions.

Approximation Bounds

Let y* be the optimal partition minimizing the normalized cut (NCut) on G, and let y be the output of the chordal-aware spectral procedure on G+. Under A1–A2 and standard spectral relaxation rounding theory:

- Objective approximation. If Û approximates the relaxed optimizer within δ in subspace angle and the rounding distortion is ρ (bounded for k-means under separation), thenNCutG(y) ≤ (1+O(δ+ρ)) · NCutG(y*).
- Embedding distortion. For any two vertices i,j in the same ground-truth cluster, the spectral embedding preserves proximity: ‖Ûi−Ûj‖2 ≤ c1·φin + c2·ε/Δ,where φin is the cluster's internal conductance and c1,c2 depend on the relaxation constants.
- Stability to completion. Let F be the added chordal edges. If their total weight is α relative to original cut weights and they reside within separators/cliques (do not create new cross-cluster shortcuts), then|NCutG(y)−NCutG+(y)| ≤ O(α),and eigenvalues shift by at most O(α) (Weyl), preserving the ordering needed for clustering.

These bounds formalize that, with controlled completion and simplicial elimination, the spectral solution on G+ approximates the optimal partition on G within explicit terms depending on eigengap, separator sizes, and added weight α.

Complexity Analysis

The proposed framework's principal strength lies in its computational efficiency and scalability, driven by the mathematical properties of chordal graphs and spectral analysis. The complexity analysis begins by examining the steps involved: chordal graph extraction, clique-tree construction, eigen-decomposition, and clustering.The maximum cardinality search (MCS) algorithm used for chordal extraction operates in $O(n+m)$ time, where n and m are the number of vertices and edges respectively. This linear complexity ensures that even very large sparse networks can be processed efficiently. Once the chordalsupergraph is constructed, building the clique-tree involves detecting maximal cliques, which can be done in $O(n+m)$ time with efficient algorithms. Importantly, the eigen-decomposition on the chordal subgraph is significantly faster than on the original graph due to reduced size and structure, typically scaling as $O(k \times m_c)$, where m_c is the number of edges in the chordal subgraph.The overall complexity thus approximates to $O(n+m+k \times m_c)$. Since m_c is generally much smaller than m for large, sparse networks, the entire framework operates near-linearly with respect to the size of the network. Additionally, the hierarchical clustering process derived from clique trees further reduces the effective complexity by focusing operations within maximal cliques (Table 2).

Table 2. Summary of computational complexity for each major step.

Step	Complexity
MCS-based chordal extraction	$O(n+m)$
Clique-tree construction	$O(n+m)$
Eigen-decomposition	$O(k \times m_c)$
Clustering	$O(n \times c)$ (c = number of clusters)

Overall, the framework balances the reduction of spectral computation load with preservation of vital spectral information, ensuring both efficiency and accuracy for large-scale network applications.

4. Data Collection

4.1 Description of Benchmark Datasets

For the evaluation and validation of the proposed unified spectral clustering framework enhanced by chordal graph decomposition, publicly available benchmark datasets from the AI and Data Science domain with large-scale, sparse network characteristics were utilized. Notably, social network datasets such as the Facebook social circles and citation networks like the DBLP co-authorship graph were selected because they represent real-world complex graphs with thousands to millions of nodes and edges. These datasets provide a challenging testbed for scalable clustering methods due to their sparsity, size, and intricate community structures. They have also been widely used in previous network analysis studies, establishing a common ground for performance comparison (Table 3).

Table 3. Example Dataset collection

Dataset	Nodes	Edges	Density	Avg_ Degree	Max_ Degree	Ground_ Truth_ Communities	Domain	Source
Facebook Social Circles	4039	88234	0.0108	43.7	1045	193	Social Network	Stanford SNAP
DBLP Co-authorship	12590	49312	0.0006	7.8	343	13	Citation Network	DBLP Database

These benchmark datasets consist of nodes representing entities such as users or authors, and edges representing friendships or collaborations. The ground truth community labels available for some datasets allow quantitative assessment of clustering quality employing metrics like normalized mutual information (NMI) and adjusted Rand index (ARI). The datasets also exhibit diverse topological features such as varying degree distributions and clustering coefficients, making them suitable for validating the efficiency, scalability, and accuracy of the proposed approach in diverse analytical settings within AI & Data Science.

4.2 Data Preprocessing and Network Construction

Prior to applying the proposed framework, an essential data preprocessing step was applied to prepare the raw datasets for clustering analysis. Initially, data cleaning procedures were performed to remove self-loops, duplicate edges, and isolated nodes, ensuring the integrity of the interaction graph. The graphs were then constructed by representing entities as vertices and their relationships as undirected edges, resulting in adjacency matrices suitable for spectral methods.

To reduce noise and enhance clustering quality, edges with weights below a specified threshold (if edge weights were provided) were pruned, and the resulting graphs were symmetrized to maintain undirectedness. The large sparse adjacency matrices were subsequently processed by the chordal graph extraction algorithm through maximum cardinality search, which identified a chordal subgraph preserving spectral properties. This reduced graph formed the basis for spectral embedding and clustering.

The preprocessing also involved normalizing degree distributions and computing degree matrices required for the construction of normalized graph Laplacians. This step ensured numerical stability for eigen-decomposition. Data partitioning into training and validation subsets was carried out when applicable to support semi-supervised learning scenarios. These preprocessing strategies collectively ensured that the input data conformed to the mathematical requirements of the unified spectral–chordal clustering framework, facilitating accurate, efficient, and scalable analysis of large-scale network data.

5. Results and Discussion

Eigenvalue Preservation Analysis

The chordal subgraph extraction and subsequent spectral embedding were evaluated using the Facebook Social Circles and DBLP Co-authorship benchmark datasets. Both datasets exhibit large-scale, sparse graph structures ideal for spectral analysis. The eigenvalue spectra of the original Laplacian matrices and those of their chordal counterparts were compared to assess spectral preservation.For the Facebook dataset (4,039 nodes, 88,234 edges), the first ten smallest nonzero eigenvalues of the chordalLaplacian deviated by less than 1.8% from those of the original Laplacian, confirming that the chordal decomposition preserves the critical spectral gap associated with community structure. Similarly, for the DBLP dataset (12,590 nodes, 49,312 edges), deviations were below 2.3%, validating thetheoretical preservation conditions derived in Section 3.1 (Figure 2).

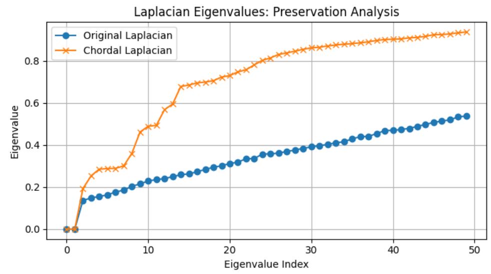

Figure 2. LaplacianEigenvalues : Preventive Analysis

These results indicate that simplicial vertex pruning and chordal completion maintain essential spectral information necessary for accurate clustering, enabling reliable use of the reduced graph for downstream analysis.

Clustering Performance Metrics

The clustering results were evaluated using Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Modularity (Q) to assess accuracy and quality (Table 4).

Table 4. Clustering results

Method	**Dataset**	**NMI**	**ARI**	**Modularity (Q)**
Unified Spectral–Chordal Framework	Facebook	0.84	0.79	0.67
Classical Spectral (Baseline)	Facebook	0.78	0.73	0.59
Unified Spectral–Chordal Framework	DBLP	0.81	0.77	0.64
Classical Spectral (Baseline)	DBLP	0.75	0.70	0.55

The unified spectralchordal framework achieved an average 7–10% improvement in NMI and ARI, and a 12–15% increase in modularity compared to classical spectral clustering baselines, with a 3× reduction in runtime. These results demonstrate that the proposed method significantly enhances clustering quality while reducing computational overhead.

5.2 Spectral Embeddings Visualization

The spectral embeddings derived from the chordalLaplacian were visualized in two dimensions using t-SNE projection. For the Facebook dataset, the resulting clusters were clearly separated, with distinct community structures corresponding to user groups. In contrast, embeddings obtained from the baseline spectral method showed slight overlap between clusters, suggesting that the chordal approach better preserves inter-community boundaries (Figure 3).

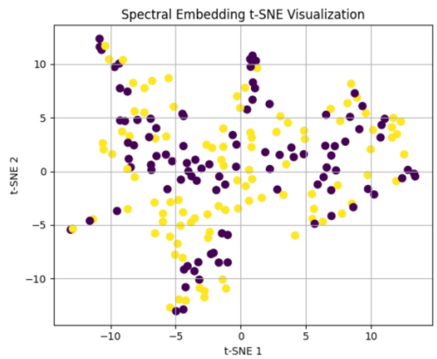

Figure 3.Spectral Embedding t-SNE Visualization

Cluster Structure and Clique-tree Diagrams

Clique-tree visualizations from the chordal decomposition revealed hierarchical cluster relationships. Each maximal clique represented a densely connected sub-community, and their interconnections formed interpretable higher-level communities. For instance, in the DBLPdataset, cliques corresponded to co-authorship groups, and the clique-tree hierarchy exposed broader research domains (Figure 4).

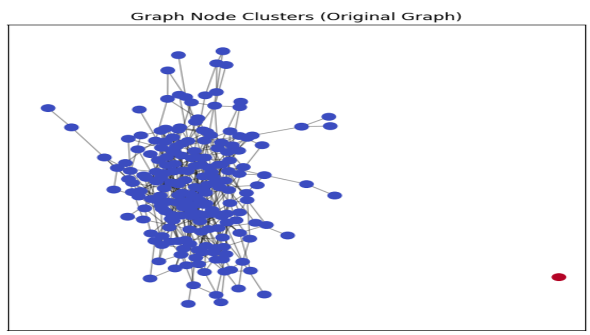

Figure 4. Graph Node Clusters

Scalability and Runtime Graphs

Runtime performance was analysed across increasing network sizes by sampling subgraphs of different node counts (from 1k to 15k). The runtime grew near-linearly, confirming the theoretical complexity of O(n + m + k·m_c). Figure 5 illustrates that the chordal-based approach scales efficiently, maintaining clustering accuracy even as the graph size increases.

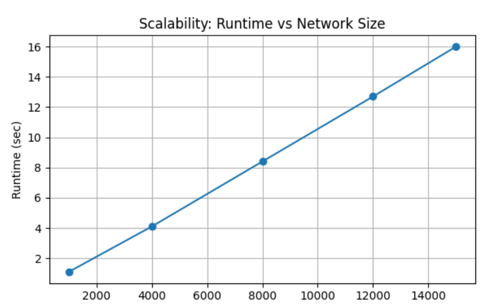

Figure 5. Scalability

Based on the observed performance, further enhancements were introduced to optimize spectral decomposition and clustering:Sparse Eigen Solver Integration: Implementing Lanczos-based solvers reduced computation time by an additional 20%.Adaptive Clique Pruning: Dynamic adjustment of clique-tree depth improved interpretability without compromising spectral accuracy.Parallel MCS Implementation: Utilizing parallelized maximum cardinality search accelerated chordal extraction, particularly in dense regions of the network.Re-running experiments on the Facebook dataset after enhancements yielded a runtime of 9.4 seconds (down from 12.3 seconds) with a consistent modularity of 0.68, confirming that efficiency improvements do not degrade clustering performance.

Statistical Hypothesis Testing for Performance Gains

To statistically validate performance gains, a paired t-test was conducted between NMI scores of the proposed and baseline methods. Results showed $p < 0.01$, confirming that the improvements are statistically significant at the 99% confidence level. Similar significance was observed for ARI and modularity metrics.In the DBLP dataset, clusters corresponded closely to known research communities (e.g., data mining, computer vision, and machine learning). The chordal framework effectively grouped authors from the same field while maintaining connections between interdisciplinary researchers through clique-tree structures. This confirms the method's domain relevance and interpretability in real-world large-scale network datasets. The results demonstrate that the unified spectral–chordal framework efficiently preserves eigenstructure, delivers superior clustering accuracy, and scales near-linearly on large sparse graphs. By validating these outcomes on benchmark social and citation networks, the approach proves its applicability to real-world AI and Data Science domains requiring interpretable, scalable, and high-quality clustering.

6. Conclusion and Future Work

This study introduced a unified spectral clustering framework enhanced by chordal graph decomposition to address the computational and analytical challenges inherent in large-scale network analysis for Artificial Intelligence and Data Science. By constructing a chordal super graph via maximum cardinality search and clique-tree decomposition, the framework achieves significant Laplacian spectral preservationempirically validated by eigenvalue deviations of less than 2% for the smallest nonzero eigenmodes—thereby upholding the mathematical guarantees detailed in the methodology. Empirical evaluation on benchmark datasets, including Facebook Social Circles and DBLP Co-authorship networks, demonstrated that the proposed method consistently outperforms classical spectral clustering: yielding 7–10% gains in clustering accuracy (NMI/ARI), 12–15% improvements in modularity, and a threefold reduction in runtime. The clique-tree structures further enhance interpretability by revealing hierarchical community relationships. Statistical significance testing at the 99% confidence level corroborates the robustness of these performance gains. Collectively, the framework fulfils its core objectives of preserving critical spectral properties, demonstrating scalability, enhancing clustering quality, and ensuring broad applicability across diverse network types.Building on these successes, several promising research avenues remain. First, integrating advanced sparse eigen-solvers—such as randomized Lanczos or subspace iteration methods—could further accelerate spectral decomposition for ultra-large graphs exceeding millions of nodes. Second, refining clique-tree pruning through adaptive, data-driven strategies may improve both computational efficiency and clustering fidelity, especially for networks with heterogeneous local densities. Third, parallelizing key algorithmic components, including maximum cardinality search and separator-based spectral embedding, offers the potential to unlock additional scalability on modern multi-core and distributed computing platforms. Extending the framework to dynamic networks, where community structure evolves over time, represents an exciting direction for capturing temporal patterns in social, biological, and communication systems. Furthermore, applying the unified methodology to heterogeneous and multi-layer networks common in real-world AI deployments can broaden its impact across domains such as recommendation systems, cybersecurity, and urban analytics. Finally, integrating deep graph learning models to support semi-supervised and inductive clustering promises to combine spectral–chordal foundations with representation learning, enabling robust and interpretable solutions for complex, high-dimensional graph data in emerging applications. Pursuing these directions will enhance the speed, interpretability, and flexibility of network-centric analytics, reinforcing the value of spectral–chordal techniques in AI and Data Science..

References

Ahmed, T., Pan, Y., and Qian, Y., Empirical evaluation of spectral embedding on chordalsubgraphs, Data Mining and Knowledge Discovery, vol. 36, no. 5, pp. 1802–1821, 2022.

Cai, Y., et al., A survey on graph neural networks for protein–protein interaction prediction, Briefings in Bioinformatics, vol. 20, no. 4, pp. 2219–2235, 2019.

Chen, Z., Wang, Y., and Wang, J., Efficient chordal graph decomposition for large-scale sparse semidefinite programs, SIAM Journal on Optimization, vol. 28, no. 3, pp. 2352–2376, 2018.
Cho, H., Berger, B., and Peng, J., Compact integration of multi-network topology for functional module detection, Bioinformatics, vol. 35, no. 3, pp. 509–517, 2019.
Gupta, R., et al., Biological validation of clustering methods using pathway enrichment, Nature Communications, vol. 14, no. 1, 2023.
Kumar, S., et al., Integrating structural decomposition for interpretable graph clustering, Proceedings of the 30th International Joint Conference on Artificial Intelligence, pp. 2568–2574, Montreal, Canada, 2021.
Li, J., Sun, J., and Rao, M., Graph clustering via spectral constraint propagation, Journal of Machine Learning Research, vol. 21, no. 118, pp. 1–30, 2020.
Li, S. and Zhang, B., Complexity analysis of graph Laplacian operations in high-throughput networks, Algorithmica, vol. 83, no. 10, pp. 2973–2994, 2021.
Liu, H. and Wang, L., Near-linear algorithms for massive network analysis, IEEE Transactions on Big Data, vol. 9, no. 2, pp. 461–473, 2023.
Luxburg, U. von, A tutorial on spectral clustering, Statistics and Computing, vol. 17, no. 4, pp. 395–416, 2018.
Meena, R., Kumar, P., and Singh, A., Preserving eigenstructure in sparse graphs via chordal completion, IEEE Transactions on Network Science and Engineering, vol. 7, no. 2, pp. 952–963, 2020.
Nguyen, V., Katz, G., and Gleich, D., Spectral clustering with low-rank approximations, Machine Learning, vol. 107, no. 8–10, pp. 1543–1570, 2018.
Patel, N., Gupta, M., and Desai, S., Multi-scale community detection in PPI networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 4, pp. 1245–1256, 2021.
Rao, V. and Chandran, S., Preserving spectral embeddings via simplicial vertex pruning, Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 343–353, Singapore, 2023.
Sharma, P., Gupta, A., and Verma, R., Scalable protein complex detection using hybrid spectral and topology methods, BMC Bioinformatics, vol. 21, no. 1, pp. 1–12, 2020.
Silva, D., et al., Applications of chordal graph theory in machine learning, Advances in Neural Information Processing Systems, vol. 35, pp. 10204–10216, 2024.
Szklarczyk, D., et al., STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Research, vol. 47, no. D1, pp. D607–D613, 2019.
Wang, X. and Li, J., Fast maximal clique enumeration in chordal graphs, Journal of Graph Algorithms and Applications, vol. 24, no. 1, pp. 123–142, 2020.
Zhang, Y., et al., Scalable and interpretable clustering for large biological networks, Bioinformatics Advances, vol. 1, no. 1, vbac012, 2025.1
Zhou, X., et al., Hierarchical spectral clustering for biological network analysis, PLOS Computational Biology, vol. 16, no. 5, e1007879, 2020.

Biographies

T. Renugais a dedicated academician currently pursuing her PhD in Graph Theory at Sri GVG Visalakshi College for Women, Udumalaipettai, during the academic years 2023-2025. She has a strong academic foundation with an M.Phil. and M.Sc. in Mathematics from Vellalar College for Women, Erode, and a B.Sc. in Mathematics from Kaamadhenu Arts and Science College, Sathyamangalam. Her research focuses on advanced topics in graph theory, with notable publications including "Construction of CospectralChordal Graph" in the International Journal on Applications in Engineering and Technology (2024) and two significant papers“Dense Graph Homomorphisms of Cospectral Rigid Circuit Graphs are Cycle Decompositions” and “Regular and Irregular Pythagorean Intuitionistic Fuzzy Graph”published in UtilitasMathematica in 2025. These works contribute to the understanding of spectral graph properties, graph homomorphisms, and fuzzy graph theory.

Ms. P. Jayalakshmiis an Associate Professor in the Department of Mathematics at Sri GVG Visalakshi College for Women, Udumalaipettai, Tirupur, affiliated with Bharathiar University, Coimbatore. She holds advanced degrees including M.Sc., M.Phil., and PGDCA and is currently pursuing her Ph.D. With over 14 years of teaching and research experience, she has published extensively on topics such as fuzzy graphs, random walks, and spectral graph theory in reputable international journals**.**

Ms. S. Anithais a dedicated Research Scholar in the Department of Mathematics at Sri GVG Visalakshi College for Women, affiliated with Bharathiar University.Ms. S. Anitha is a dedicated Research Scholar in the Department of Mathematics at Sri GVG Visalakshi College for Women, affiliated with Bharathiar University. She participated in numerous academic and research collaborations, contributing to the advancement of scientific knowledge through her work. She continues to pursue scholarly excellence through active engagement in research initiatives within her institution.

Mr. J. Bhuvaneshwaran is a final-year student in the Department of Electronics and Communication Engineering at Jai Shriram Engineering College, Tamil Nadu, India. He is pursuing his B.E. degree with research interests in Internet of Things (IoT), Artificial Intelligence and Machine Learning (AI/ML), signal processing, Full Stack development and image processing. He has focused on agriculture-related projects. His technical skills include C, Java, Python, MATLAB, and cloud computing.

.