**ANALYSIS OF COMPLEX NETWORKS USING MOLECULAR SIMILARITY DATA**

ABINAYA VINOTH & BALAAPARAJITHA SENTHIL KUMAR

# Abstract

This study explores the application of complex network theory to high-dimensional molecular similarity data extracted from the `roboBohr.csv` dataset. Leveraging advanced graph-based models, we investigate structural properties, community dynamics, diffusion phenomena, and temporal behaviors within the network. Our findings provide insights into the modular architecture of molecular interactions and demonstrate parallels between classical and quantum network behaviors.

# Index

* 1. Introduction
* 2. Data Description
* 3. Network Construction
* 4. Structural Analysis
* 5. Community Detection
* 6. Dynamic Simulations
* 7. Advanced Models
* 8. Temporal and Quantum Parallels
* 9. Visual Interpretations
* 10. Conclusion

# 1. Introduction

This report presents a comprehensive analysis of complex networks derived from the high-dimensional `roboBohr.csv` dataset, containing molecular features. Using principles of graph theory and network science, we construct, analyze, and simulate dynamic processes on the resulting network. The goal is to explore structural properties, community detection, diffusion, synchronization, and other network dynamics which reflect real-world phenomena observed in both classical and quantum systems.

# 2. Data Description

The dataset includes numerical features for chemical compounds along with identifiers. From 500 randomly sampled molecules, we created a similarity-based network using cosine similarity. Edges represent strong structural or property-based similarities (threshold > 0.85).

# 3. Network Construction

We used NetworkX to build an undirected graph:  
Nodes: Molecules  
Edges: Cosine similarity > 0.85  
Graph Size: 500 nodes and ~71,000 edges.

# 4. Structural Analysis

Cliques: Over 2500 maximal cliques indicate dense subgroups  
Connectivity: The network remains largely connected with minimal isolated nodes  
Degree Distribution: Heterogeneous; a few hub nodes dominate.

# 5. Community Detection

Louvain Method: Detected 3 main communities, modularity score = 0.42  
Girvan-Newman: Hierarchical decomposition supports modular structure  
Visualizations: Circular and spring layouts showed distinct color-coded communities.

# 6. Dynamic Simulations

Laplacian Diffusion: Eigenvalue distribution shows multiple diffusion modes  
Kuramoto Synchronization: Simulated phase coupling showed partial synchronization  
Contagion Model: Activation spread over steps showed threshold-dependent phase transitions  
Cascading Failures: Targeted hub removal fragmented the network.

# 7. Advanced Models

Evolutionary Games: Strategy convergence shows localized adaptation (Prisoner’s Dilemma)  
Chaotic Dynamics: Logistic map generated bifurcated behavior  
Multilayer Networks: Potential for layered modeling of chemical/biological properties.

# 8. Temporal and Quantum Parallels

Time-dependent networks: Snapshots captured edge volatility  
Quantum Analogy: Phase coupling, entanglement simulation potential, and multilayer modeling align with quantum communication principles.

# 9. Visual Interpretations

Multiple visualizations like circular layout, rich-club analysis, and Laplacian heatmaps supported structural interpretations and temporal changes.

# 10. Conclusion

This project illustrates the effectiveness of network science in chemical data analysis. By modeling molecular similarity as a complex network, we extract structural insights and simulate processes analogous to physical and biological phenomena, including quantum synchronization. Future work may integrate GNNs or quantum graph representations.