



Unveiling Network Patterns: A Multidomain Network Analysis

Atharva Karnik 
Indiana University
Bloomington, USA
akarnik@iu.edu

Ayana Holla Pandeshwara
Indiana University
Bloomington, USA
ayholla@iu.edu

Diksha Adke 
Indiana university
Bloomington, USA
dadke@iu.edu

VV Hari Chandan
Indiana university
Bloomington, USA
hvmugan@iu.edu

Abstract—The aim of this project is to conduct a comprehensive analysis and compare the structural properties of networks from various domains to identify underlying patterns and commonalities. By examining specific network properties highlighted in our academic coursework - degree distribution, clustering coefficient, average path length, and betweenness centrality - across different types of networks (social, citation, web, and collaboration networks), we seek to group these networks based on their characteristic similarities and differences. This comparison will allow us to draw insights into how network properties vary by domain and identify potential universal principles governing network structures.

Index Terms—Graph Analysis, Python, Network Analysis

I. INTRODUCTION

In this study, we undertake an exploratory analysis of 16 real-world networks selected from four distinct domains: Collaboration, Retweet, Email, and Biology, with each domain contributing four unique networks. Our analysis is centered on examining a comprehensive set of structural properties such as node and edge counts, density of connections, degree metrics (average, maximum, minimum), assortativity, and various measures related to triangles (number, average, maximum), alongside clustering coefficients (average and fraction of closed triangles). We also investigate the structural integrity of these networks through metrics like the maximum k-core and the estimated size of the largest fully connected subgroup. To categorize these networks based on their structural characteristics, we employ hierarchical and K-means clustering techniques. This methodological choice allows us to explore potential groupings and patterns that might not be immediately apparent, aiming to uncover any universal properties or distinct behaviors that transcend the individual network types. The goal of this project is not predefined but rather to navigate through the data in an exploratory manner, using advanced clustering techniques as a way to derive meaningful insights from complex network structures. This approach helps in advancing our understanding of how different network properties influence the overall network behavior across various domains, potentially contributing to more generalized theories of network dynamics.

In the field of network science, understanding the structural properties of various networks can offer profound insights into their functionality and behavior. This is particularly relevant

in today's interconnected world, where networks can represent anything from biological systems and social interactions to technological and information ecosystems. Each type of network, whether it involves social media interactions, protein interactions in a cell, communication patterns in organizations, or collaborations in scientific research, has unique characteristics that can significantly influence its efficiency, resilience, and dynamics.

The problem being solved in this project is to determine how these diverse networks can be systematically compared and analyzed to reveal underlying patterns and principles that govern their structures. This is an intriguing problem because it holds the potential to uncover universal rules that apply across various domains, potentially allowing for predictions about network behavior and influencing the design and management of future networks. Moreover, understanding these networks better can help in identifying key nodes or connections that are critical for the stability or performance of these networks, which is crucial for areas like epidemic prevention, information dissemination, and organizational efficiency.

The excitement in tackling this problem lies in its inherent complexity and the broad applicability of the findings. By employing advanced clustering techniques like hierarchical and K-means clustering, this project not only aims to categorize networks based on their structural properties but also seeks to enhance our comprehension of how similar networks are across different contexts. This cross-domain analysis is particularly fascinating because it may demonstrate that disparate networks like those of biological interactions and technological communications might share more similarities than previously thought, suggesting new ways to manage and optimize these networks. Such insights are invaluable in an increasingly networked world, making the study both a challenging and a rewarding endeavor.

II. METHODOLOGY

A. Dataset Selection and Preprocessing

Source: Utilize the Stanford Network Analysis Project (SNAP) database to select datasets. Focus on a variety of network types, including social networks, citation networks, web graphs, and collaboration networks, to ensure a comprehensive analysis across different domains.

Preprocessing Steps: The initial step involved cleaning the raw network data to remove inconsistencies and irrelevant information. This included standardizing node labels and ensuring that all edges were correctly formatted.

Attribute Standardization: All network attributes were standardized to ensure uniformity across datasets. This involved normalizing numerical attributes and encoding categorical features to prepare the data for further analysis. Removal of Unnecessary Data Non-essential data, such as isolated nodes (nodes without any edges) and any metadata not relevant to network structure analysis, were removed. This step helped in reducing noise and focusing the analysis on significant interactions within the networks.

B. Network Property Analysis

Our methodology was primarily executed using Python, leveraging the robust capabilities of the NetworkX library for network analysis, the community-louvain library for community detection, and matplotlib and seaborn for data visualization. This section provides a detailed overview of the steps and calculations involved in analyzing the network properties.

Loading and Initial Processing of Networks: Each network was first loaded into a directed graph structure utilizing NetworkX, a powerful library designed to handle both simple and complex network structures efficiently. The process involved:

Reading Data: Each network file was parsed to extract nodes and edges. Nodes represent entities such as individuals or proteins, while edges represent the relationships or interactions between these entities.

Graph Construction: Edges were added to the graph, ensuring each interaction was represented accurately. This setup allowed us to apply various network analysis techniques directly on the graph structure.

We began our analysis by computing several fundamental properties to get an overview of each network's structure:

Node and Edge Counts: These metrics provide the most basic information about the size and complexity of the network.

Network Density: Calculated as the ratio of the number of edges to the number of possible edges between pairs of nodes, offering insights into how densely the network is connected.

Degree Metrics Analysis: Degree metrics are crucial for understanding the connectivity of nodes within the network.

Maximum, Minimum, and Average Degree: These statistics were calculated to identify nodes with the highest and lowest connectivity, as well as the overall average connectivity in the network, which helps in understanding the distribution of connectivity across nodes.

Clustering and Triangles: Clustering metrics were computed to assess the tendency of nodes to form tightly-knit groups.

Number of Triangles: We counted closed triplets within the network, which are indicative of tightly interconnected groups of three nodes.

Clustering Coefficients: Both the average clustering coefficient and the overall network transitivity were calculated. The average clustering coefficient measures the average likelihood

that two nodes with a common neighbor are connected, providing a measure of the overall cliquishness of the network.

Transitivity: similar in concept but slightly different in calculation, gives a global indication of the interconnectedness of the network.

Advanced Structural Properties - To dive deeper into the network's resilience and structural integrity, we computed:

Maximum k-core: This measure identifies the largest sub-network in which each node is connected to at least 'k' other nodes, highlighting areas of high interconnectivity.

Lower Bound of Maximum Clique: We estimated the size of the largest completely interconnected subgroup, which is crucial for identifying the most cohesive group within the network.

Community Detection and Modularity- Understanding how networks are compartmentalized into communities can reveal much about their structure:

Community Detection: Using the community-louvain method, we detected communities within the networks, which are groups of nodes more densely connected internally than with the rest of the network.

Modularity Calculation: The modularity score was computed to quantify the strength of the divisions of the network into communities, providing a numeric value that reflects the quality of the community divisions.

C. Comparative Analysis

Pattern Identification: Compare the computed properties across different networks to identify patterns, similarities, and unique characteristics. This involves statistical analysis to highlight significant deviations or commonalities in network structures.

Cluster Analysis: Employ clustering algorithms, like K-means or hierarchical clustering, to group networks based on their properties. This step helps to categorize networks into distinct types based on their structural characteristics.

Dimensionality Reduction: For visualization and to facilitate clustering, apply dimensionality reduction techniques such as PCA (Principal Component Analysis) or t-SNE (t-Distributed Stochastic Neighbor Embedding) on the network properties.

D. Visualization

Visualization Techniques: Use matplotlib and seaborn for static visualizations and potentially libraries like Plotly for interactive graphs. Visualizations will include histograms of network properties, scatter plots of networks in reduced dimensionality space, and network diagrams highlighting key structural characteristics.

Degree Distribution Visualization: By plotting the distribution of degrees across the network, we could visually assess the variance and skewness in node connectivity.

Clustering Coefficient Distribution: This helped us visualize the spread of clustering coefficients across the network, indicating areas of potential robustness or vulnerability.

Community Structure Visualization: Displaying communities helped illustrate the modular structure of the networks and the effectiveness of the community detection algorithm. These elaborate calculations and visualizations provided a multifaceted view of the networks, allowing us to conduct a thorough comparison of network properties across different domains.

III. ANALYSIS

Following the comprehensive computation of network metrics across the 16 networks, the next logical step was to consolidate all the calculated properties into a structured dataset. To achieve this, we compiled all the network properties into a CSV file, effectively creating a tabulated dataset that encompasses a wide range of structural characteristics for each network. This dataset includes metrics such as node and edge counts, density, degree metrics (maximum, minimum, and average), assortativity, number of triangles, clustering coefficients, and advanced properties like the maximum k-core and the lower bound of maximum cliques, among others.

This structured dataset serves multiple purposes. Primarily, it provides a standardized format for comparing and analyzing the networks across different domains. It also facilitates ease of access and manipulation of data for further statistical analysis or machine learning applications.

By organizing the data into a CSV format, we ensure that the information is not only easily readable but also ready to be used in various data analysis tools and software, enhancing the efficiency of subsequent research and analysis phases.

In the analysis of network properties, standardization of data is a crucial step to ensure that the results are comparable across different networks. For this study, we selected the Robust Scaler from the available scaling methods due to its superior ability to handle outliers, which are often present in datasets involving diverse network types like those in our study.

Understanding Robust Scaler: Robust Scaler is a scaling algorithm that is less sensitive to outliers than other scaling methods such as Standard Scaler. Unlike the latter, which removes the mean and scales the data to unit variance, Robust Scaler uses the median and the interquartile range for scaling.

Specifically, it subtracts the median from the data points and then divides by the interquartile range (IQR). The IQR is the difference between the 1st quartile (25th percentile) and the 3rd quartile (75th percentile), making this method robust to outliers since these do not influence the median or IQR as much as they would the mean and standard deviation.

$$X_{\text{scaled}} = \frac{X_i - X_{\text{Mean}}}{IQR} \quad (1)$$

Outlier Impact: Network datasets, particularly those drawn from real-world scenarios, often contain extreme values or outliers due to irregularities in data collection or inherent variability in network connectivity (e.g., a few nodes might have significantly higher connections than the rest). Methods like Standard Scaler, which rely on the mean and standard deviation, are highly sensitive to outliers, which can misrepresent

the true nature of most of the data. Robust Scaler, by using the median and IQR, reduces the influence of these extreme values.

Data Variability: Networks from different domains such as biology, social media, and email communications exhibit varied connectivity patterns and node interactions. The robustness of the Robust Scaler ensures that the scale of these differences does not distort the comparative analysis across such diverse types.

Consistency in Scale: The scaling process standardizes features by removing the median and scaling the data according to the quantity expressed by the IQR. This method ensures that the resulting feature values lie on a similar scale, providing a consistent basis for comparing and clustering networks.

Preserving Data Integrity: By avoiding the distortion caused by outliers, the Robust Scaler preserves the integrity of the network's structural properties. This is crucial for maintaining the accuracy of subsequent analyses, such as clustering, which could be adversely affected by skewed data distributions. Our study embarked on an extensive descriptive analysis of the scaled network data. This foundational step was crucial in gaining a deep understanding of the dataset's characteristics, which helped inform our subsequent analytical decisions.

Statistical Summary and Dispersion Metrics - The first phase of our descriptive analysis involved computing basic statistical summaries for each network property. This included measures of central tendency (mean) and spread (standard deviation, interquartile range) for the scaled features. Due to the application of the Robust Scaler, most features had means centered around zero, effectively mitigating the impact of outliers and facilitating a more uniform comparison across networks. The variation in standard deviations indicated differing levels of data dispersion, which was pivotal in assessing the relative variability of each network metric.

Interquartile range (IQR), which measures the spread of the middle 50% of data, had particular insights. It helped us understand how data points were clustered around median, providing insights into overall distribution shape—whether the data was tightly grouped or spread out, which in turn could influence the outcomes of clustering and PCA.

Histogram Analysis for Feature Distribution - To visually assess the distribution of each feature, histograms were generated for all scaled metrics. These histograms allowed us to observe the skewness or symmetry in the distributions of features such as 'Density' and 'Maximum number of triangles.' We noted that most network features were not normally distributed, reflecting the inherent diversity and complexity within and across network types. Features displaying extreme values or skewness were of particular interest, as they highlighted networks that either were exceptionally dense or featured a high number of triangular connections—both characteristics that could significantly influence network behavior and clustering tendencies.

Correlation Heatmap and Insights - A correlation analysis was conducted to examine the relationships between different network properties. By visualizing these correlations through

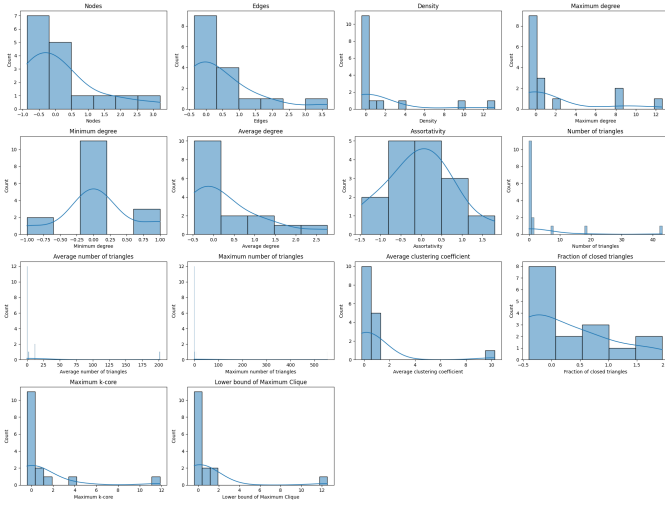


Fig. 1. Histogram Analysis

a heatmap, we identified several high positive and negative correlations. For instance, a strong positive correlation between the number of nodes and edges was intuitive, reflecting that larger networks typically have more connections. Conversely, negative correlations, such as between assortativity and degree metrics, suggested that networks with higher assortativity—tendency of nodes to connect to others that are similar—might exhibit different structural characteristics compared to those with lower assortativity.

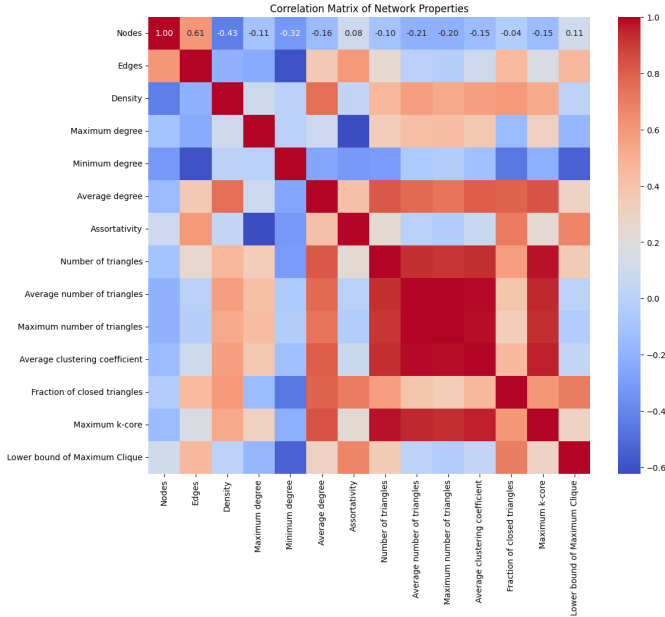


Fig. 2. Correlation Matrix

These correlations were crucial in understanding potential interactions between different network metrics and how these interactions might manifest in different types of networks. Such insights were valuable for interpreting cluster analysis results and for setting up the PCA, as they highlighted un-

derlying patterns that could dictate how network properties grouped together in reduced-dimensional space.

Exploring Pairwise Relationships - To further delve into the data, scatter plots examining pairwise relationships between key features were created. These plots, along with histograms on their diagonals, provided a dual view—showing both individual feature distributions and their relationships with one another. Observations from these plots were particularly revealing. For example, the relationship between 'Density' and 'Average Degree' varied across network types, potentially indicating that denser networks also tended to have higher average degrees, a pattern that could influence how networks cluster based on density and connectivity.

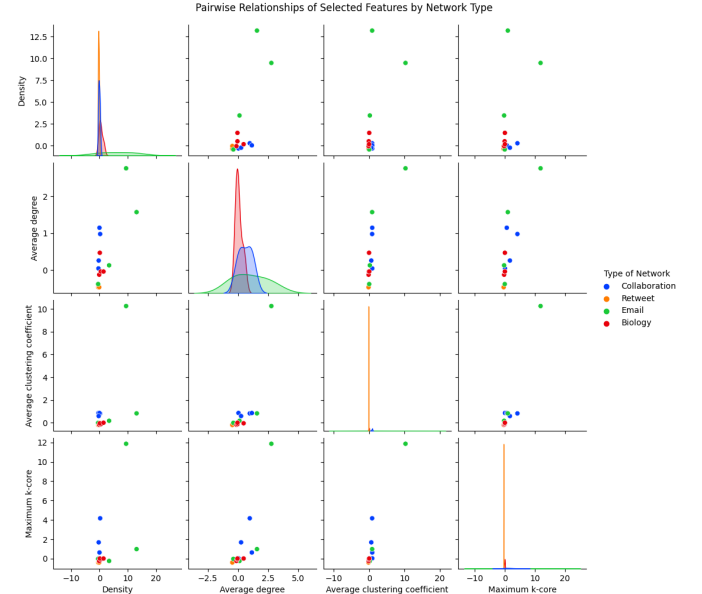


Fig. 3. Pairwise Relationships

Following the detailed descriptive analysis, Principal Component Analysis (PCA) was implemented to further refine our understanding of the dataset and to identify the underlying structure within the network properties. PCA is a powerful statistical technique used to reduce the dimensionality of a dataset while retaining as much variance as possible. This is achieved by transforming the original variables into a new set of variables known as principal components, which are linear combinations of the original variables ordered by the amount of variance they explain.

Implementation of PCA: PCA was applied to the scaled network data, focusing on extracting principal components that could effectively summarize the key information:

Variance Explained by Components: The first step in our PCA implementation involved calculating the explained variance ratio for each component. This metric indicates how much of the total variance in the dataset is captured by each principal component. The analysis typically starts by looking at the cumulative variance explained as more components are considered, helping to determine how many components are

needed to adequately describe the data.

Scree Plot Visualization- To visually assess the contribution of each principal component, a scree plot was created. This plot displayed the cumulative explained variance against the number of components, which is crucial for identifying the "elbow point" where the addition of further components does not significantly increase the explained variance. This point is key to determining the optimal number of components to retain for further analysis.

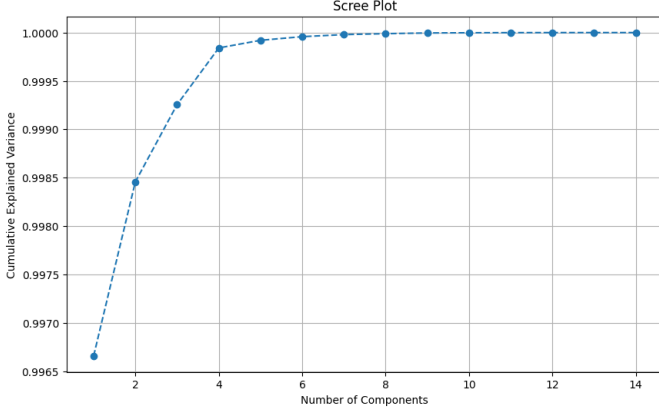


Fig. 4. Scree Plot

Variance Explained by Principal Components:

Individual Component Variance: The first principal component (PC1) emerged as highly dominant, capturing approximately 99.67% of the total variance in the dataset. This substantial variance indicated that PC1 alone encapsulates the majority of information inherent in the original features.

Subsequent Components: The second principal component accounted for an additional 0.18% of the variance, with each subsequent component contributing progressively smaller percentages. This pattern highlighted the diminishing informational returns with the inclusion of more components.

Cumulative Variance: The cumulative variance explained by the principal components quickly approached totality with the inclusion of just the first few components, indicating that the bulk of data variability could be represented effectively by these initial components. PCA loadings and their interpretations are as follows:

First Principal Component (PC1): The loadings for PC1 revealed that it heavily weights features associated with the number of triangles, specifically the Maximum number of triangles and Average number of triangles. This suggests that the primary differentiator between networks within the dataset relates to their triadic closure properties, which are central to understanding the network's cohesiveness and social structure.

Subsequent Components: The second and third principal components began to incorporate other features such as Edges, Density, and Lower bound of Maximum Clique. These components appear to capture more nuanced differences between networks that the dominant triangle-related metrics do not fully explain.

K-Means Clustering on PCA-Reduced Data Following the dimensionality reduction achieved through PCA, our next step involved employing K-Means clustering to categorize the networks into distinct groups based on their principal component scores. This clustering aimed to discern underlying patterns and group similar networks together, enhancing our understanding of the dataset's structure.

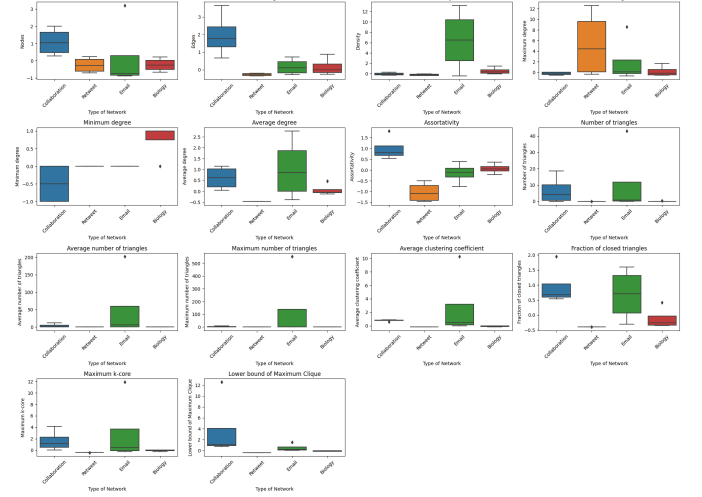


Fig. 5. Features Distribution Comparison

Determining the Number of Clusters with the Elbow Method: To identify the optimal number of clusters for the K-Means algorithm, we implemented the Elbow Method, a popular technique that involves plotting the sum of squared distances (inertia) of samples to their nearest cluster center across a range of cluster numbers. We calculated the inertia for cluster counts ranging from 1 to 9 and observed the "elbow" point—the point at which the inertia begins to decrease more gradually. This point suggests a diminishing return on the explanatory power of additional clusters and thus serves as a criterion for choosing the number of clusters. Our analysis, visualized through an Elbow Plot, indicated that the optimal number of clusters is three, as the inertia sharply tapers off after this point.

Clustering and Visualization: With the number of clusters determined, we proceeded to apply K-Means clustering to the PCA-reduced dataset. Initially, the PCA had considered components that cumulatively explained up to 99% of the variance. For practical visualization and analysis, we later adjusted this to focus on the first two principal components, which captured the most significant aspects of variance yet allowed us to visualize the data in a two-dimensional space. The clustering results were then visualized, showing how networks were grouped based on the two most informative PCA components. This visualization provided a clear, intuitive depiction of the clusters, with each network represented as a point colored according to its assigned cluster. The plot highlighted distinct groups within the data, with each cluster representing networks with similar structural characteristics as

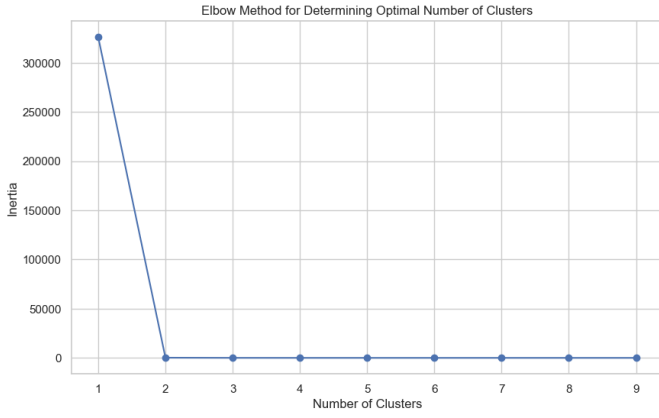


Fig. 6. Elbow Method for Optimum Cluster

captured by their PCA scores.

Interpretation and Insights- The data reveals three main clusters:

Cluster 0 (Purple): This is the most populous cluster, encompassing a wide array of networks from various domains, including Collaboration, Retweet, Email, and Biology. Networks in this cluster generally exhibit moderate values across the features measured, suggesting they represent more typical or average network structures within this dataset. The spread of networks across different types suggests that while they may vary in purpose or design, their underlying structural metrics such as node connectivity and triangle counts do not vary extremely.

Cluster 1 (Cyan): This cluster contains just one network, email-dnc, which is drastically different in its network characteristics. It stands out significantly in terms of the Density metric and the Maximum number of triangles, Average number of triangles, and Average clustering coefficient, which are exceedingly higher compared to other networks. This indicates an extraordinarily dense network structure, possibly a tightly-knit community with many triadic closures, suggesting a highly interconnected network.

Cluster 2 (Yellow): Including only CA-HepPh.txt, this cluster also characterizes a network with unique structural properties but not as extreme as Cluster 1. This network displays relatively high values in the number of triangles and lower bound of maximum clique, indicative of a robust structure with significant connectivity but less extreme than the network in Cluster 1.

Interpretation of Clusters
Cluster 0's Homogeneity: The homogeneity within Cluster 0, despite the diversity of network types, suggests that common network metrics like node degrees, edge counts, and simple connectivity measures may not vary drastically across different domains or may be influenced similarly by the scaling and PCA process. This cluster likely represents standard network forms where no extreme properties are observed.

Isolated Nature of Cluster 1 and 2: The isolation of email-dnc in Cluster 1 suggests a unique structure profoundly different from others, likely due to its specific use case or design,

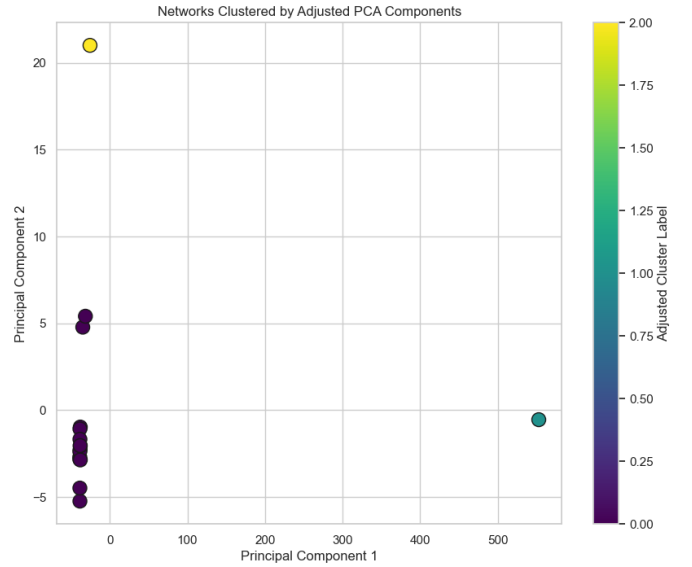


Fig. 7. Obtained Clusters

which emphasizes high density and connectivity. It could be useful in scenarios where robust information dissemination or fault tolerance is crucial. CA-HepPh.txt in Cluster 2, while also unique, suggests a high level of academic collaboration, possibly reflecting a dense citation or collaboration network common in high-energy physics.

Hierarchical Clustering Analysis

Following the PCA and K-means clustering analyses, hierarchical clustering was implemented to further explore the relationships and groupings within the dataset using a different clustering methodology. Hierarchical clustering is particularly useful for identifying nested structures and providing a visual representation of the data hierarchy through a dendrogram.

Implementation of Hierarchical Clustering- Hierarchical clustering was performed using the Ward's method, which minimizes the variance within each cluster. This method is advantageous for our analysis because it tends to create more evenly sized clusters, which can be informative when dealing with diverse network data:

Data Preparation: The clustering was applied to the scaled data, specifically using the features identified as significant in the PCA, excluding non-numerical columns like 'Network' and 'Type of Network' to focus solely on structural metrics.

Clustering Process: Using the linkage function from the `scipy.cluster.hierarchy` library, we generated a hierarchical clustering model. The Ward's method was selected to optimize the within-cluster sum of squares, aiming to form clusters that are as homogenous as possible.

Analysis of the Dendrogram: The dendrogram produced from the hierarchical clustering provides a visual representation of the clustering process:

Dendrogram Structure: The plot reveals how each network is linked at various levels of similarity, which is represented by the 'Distance' on the y-axis. This distance indicates the

level of dissimilarity between clusters, with networks joining at lower distances being more similar.

Significant Separation: Notably, the email-dnc network stands out significantly from the rest, joining the hierarchical tree at a very high distance. This indicates a substantial difference in its network structure compared to others, aligning with previous observations from PCA and K-means analyses where email-dnc was identified as an outlier due to its unique characteristics.

Cluster Formation: The majority of networks are grouped together at a relatively low distance, suggesting a higher level of similarity among them. This large cluster includes networks of various types, indicating that despite their different applications or domains, they share certain structural properties.

Hierarchical clustering complements the insights gained from PCA and K-means by providing a different perspective on network relationships. It highlights both the gradations in similarity among most networks and the stark differences of outliers.

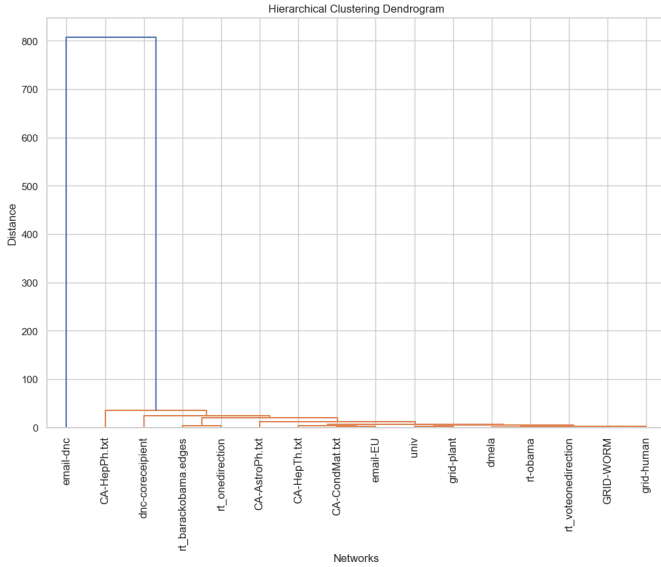


Fig. 8. Dendrogram Analysis

Outlier Identification: The clear separation of the email-dnc network reinforces its role as an outlier, which could be due to extreme values in features such as density or connectivity, making it structurally distinct from other networks in the dataset.

Nested Structures: The dendrogram also suggests potential nested relationships within the data that could be explored further. For instance, sub-clusters within the main cluster could be examined to understand finer nuances in network structure.

IV. CONCLUSION

This project embarked on an exploratory journey through the structural properties of 16 diverse networks, leveraging a combination of statistical techniques and machine learning methods.

Through the application of descriptive statistics, PCA for dimensionality reduction, K-means clustering, and hierarchical clustering, we gained deep insights into how these networks vary and relate across multiple domains. Some of the key findings were:

Dominance of Triangle Metrics: PCA revealed that triangle-related metrics, such as the number of triangles and the average number of triangles, are crucial in differentiating network structures, accounting for the majority of the variance within the data.

Cluster Insights: K-means clustering, guided by the PCA results, identified distinct groups of networks that share similar properties, highlighting the structural uniqueness of certain networks like email-dnc which displayed exceptionally high connectivity.

Hierarchical Relationships: Hierarchical clustering provided a visual depiction of the relationships and similarities between networks, confirming the outlier status of email-dnc and illustrating a general similarity among other networks.

These analyses not only demonstrate the practicality and power of advanced analytical techniques in understanding complex network data but also contribute to the broader field of network science by offering insights that are applicable across various types of networks.

V. FUTURE SCOPE

The findings from this project open several avenues for further research and application:

Deeper Network Property Analysis: Future studies could explore additional network properties not covered in this analysis, such as centrality measures, which could provide further insights into the importance of individual nodes within networks.

Temporal Dynamics: Incorporating a temporal dimension to analyze how these networks evolve over time would provide a dynamic view that could help in understanding the changing nature of these networks.

Cross-Domain Comparative Studies: Expanding the dataset to include more networks from each domain could enhance the robustness of the clustering results and allow for a more detailed comparative study across different domains.

Network Modeling and Simulation: Using the insights gained from this analysis, models could be developed to simulate network growth or predict the impact of network changes on overall structure and function.

Advanced Machine Learning Models: Applying more sophisticated machine learning models like deep learning could uncover more complex patterns and relationships in network data.

Application-Specific Analysis: Tailoring the analysis to specific application needs, such as cybersecurity, social media analytics, or biological data analysis, could provide actionable insights specific to those fields.

Integration with Graph Neural Networks: Leveraging recent advances in graph neural networks could provide novel

methods for network data analysis, potentially improving the accuracy and applicability of the findings.

REFERENCES

- [1] V. Di Massa, G. Monfardini, L. Sarti, F. Scarselli, M. Maggini and M. Gori, "A Comparison between Recursive Neural Networks and Graph Neural Networks," The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 2006, pp. 778-785, doi: 10.1109/IJCNN.2006.246763.
- [2] B. Knyazev, Graham W. Taylor, Mohamed R. Amer, "Understanding Attention and Generalization in Graph Neural Networks", <https://doi.org/10.48550/arXiv.1905.02850>
- [3] T. C. Hsu, Y. H. Hsieh and Che-Lin, "Utilizing Graph Neural Networks for Breast Cancer Prognosis Prediction with High-dimensional Genomic Data," 2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-4, doi:10.1109/EMBC40787.2023.10340045
- [4] B. Liu et al., "Nowhere to Hide: Online Rumor Detection Based on Retweeting Graph Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 4887–4898, April 2024. <https://doi.org/10.1109/TNNLS.2022.3161697>
- [5] P. Lubarski and M. Morzy, "Measuring the Importance of Users in a Social Network Based on Email Communication Patterns," 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 2012, pp. 86-90, doi:10.1109/ASONAM.2012.24.
- [6] D. Cummings and M. Nassar, "Structured Citation Trend Prediction Using Graph Neural Networks," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 3897-3901, doi: 10.1109/ICASSP40776.2020.9054769.