Toronto Metropolitan University
Data Science & Analytics Graduate Studies

Results Report: Network Resilience and Community Analysis

1. Introduction

This study evaluates the resilience of the Email and LastFM networks under various conditions of node addition and removal. Resilience, in this context, refers to the network's ability to maintain its structural integrity and functionality when faced with these changes. A good measure of resilience is the size of the largest connected component, as a resilient network will retain a significant portion of its nodes connected despite disruptions.

We analyze these datasets by comparing the original networks with synthetic models: Chung Lu and UPA. The Chung Lu model generates a network with the same expected degree sequence as the original but introduces randomness, allowing for the study of structural properties derived from node degrees. The UPA (Uniform Preferential Attachment) model, on the other hand, mimics the growth process of networks by adding new nodes that preferentially attach to existing ones with higher degrees, simulating the formation of scale-free networks.

Additionally, we examine the resilience of communities within these networks, identified using the Louvain method, to understand their robustness and identify key vulnerabilities. This comprehensive analysis helps in understanding how different parts of the network respond to changes and disruptions.

2. Data Preparation

2.1 Email Dataset

- **Statistics**:
  - Number of nodes: 1005
  - Number of edges: 16706
- **Degree Distribution**: Reveals that most nodes have a low degree, indicating sparse connections, while a few nodes have a very high degree, indicating highly connected hubs.
- **Department Labels Distribution**: Shows an uneven distribution across different departments, with some departments having significantly more nodes than others.

2.2 LastFM Dataset

- **Statistics**:
  - Nodes: 7624
  - Edges: 27806
- **Degree Distribution**: Shows that most nodes have a few connections, while a few nodes act as highly connected hubs, reflecting a scale-free network characteristic.

2.3 Degree Distribution Analysis

Degree distribution is a fundamental property of network analysis that describes how the connections (or degrees) of nodes are distributed across the network. In the Email and LastFM datasets, the degree distribution reveals significant insights into the network's structure and behavior. For both networks, the distribution indicates a few highly connected nodes (hubs) and many nodes with fewer connections. This pattern is characteristic of scale-free networks, where some nodes (hubs) play a critical role in maintaining network connectivity and resilience. By analyzing the degree distribution, we can identify these crucial nodes and understand the overall connectivity and robustness of the network.

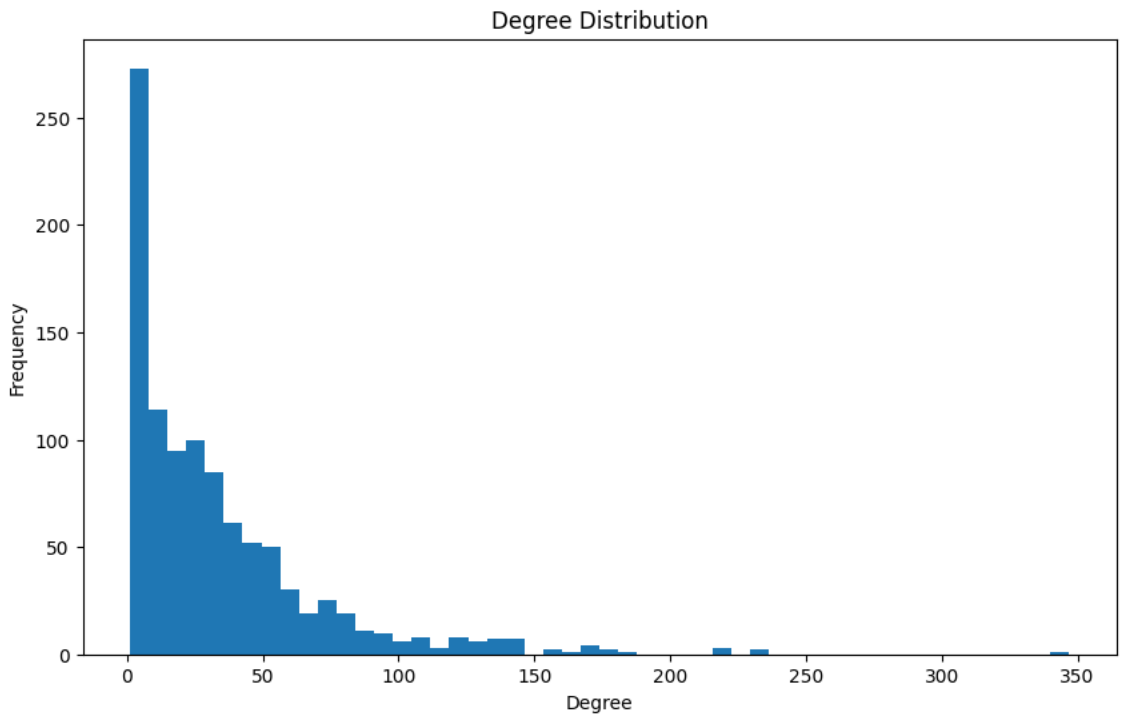


Fig 2.1

3. Community Detection

In network analysis, a community refers to a group of nodes that are more densely connected internally than with the rest of the network. To identify these communities, we employed the Louvain method, which is an efficient algorithm for optimizing modularity—a measure of the strength of the division of a network into communities.

The Louvain method iteratively groups nodes into communities in a way that maximizes modularity. High modularity indicates strong community structures, where nodes within a community have many connections, while connections between communities are fewer.

For both the Email and LastFM networks, the Louvain method identified multiple communities. The number of communities and their sizes varied, reflecting the diversity in connectivity patterns. In the Email network, for example, we detected 43 communities, ranging in size from small groups of a few nodes to larger clusters. The modularity score for the Email network was 0.42998735542482025, indicating a relatively strong community structure. Similarly, the LastFM network revealed 26 communities with a modularity score of 0.8156278462397027, highlighting robust community formations.

These community structures provide valuable insights into the network's organization and can help identify critical subgroups and potential vulnerabilities. Understanding how communities are formed and their internal connectivity is crucial for assessing the overall resilience of the network.

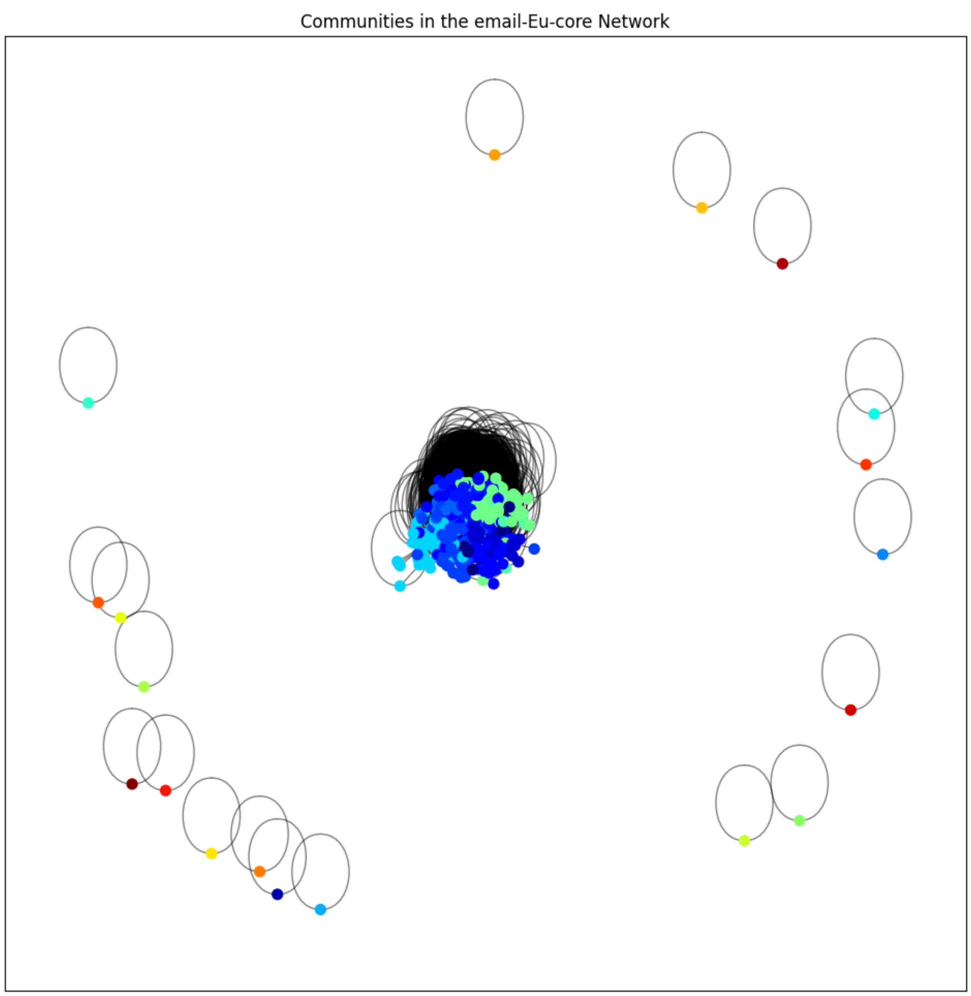


Fig 3.1

3.1 What are Self-Loops

Self-loops are edges that connect a node to itself in a network. In the context of a graph, these are edges where the start and end nodes are the same.

3.2 Why Remove Self-Loops

1. **Redundancy:** Self-loops do not contribute to the connectivity or overall structure of the network, as they do not connect different nodes.
2. **Interference with Analysis:** They can distort various network metrics, such as degree centrality and clustering coefficients, leading to misleading results.
3. **Focus on Connectivity:** For resilience analysis, the primary interest is in how nodes are interconnected. Self-loops do not impact the network's robustness or resilience.

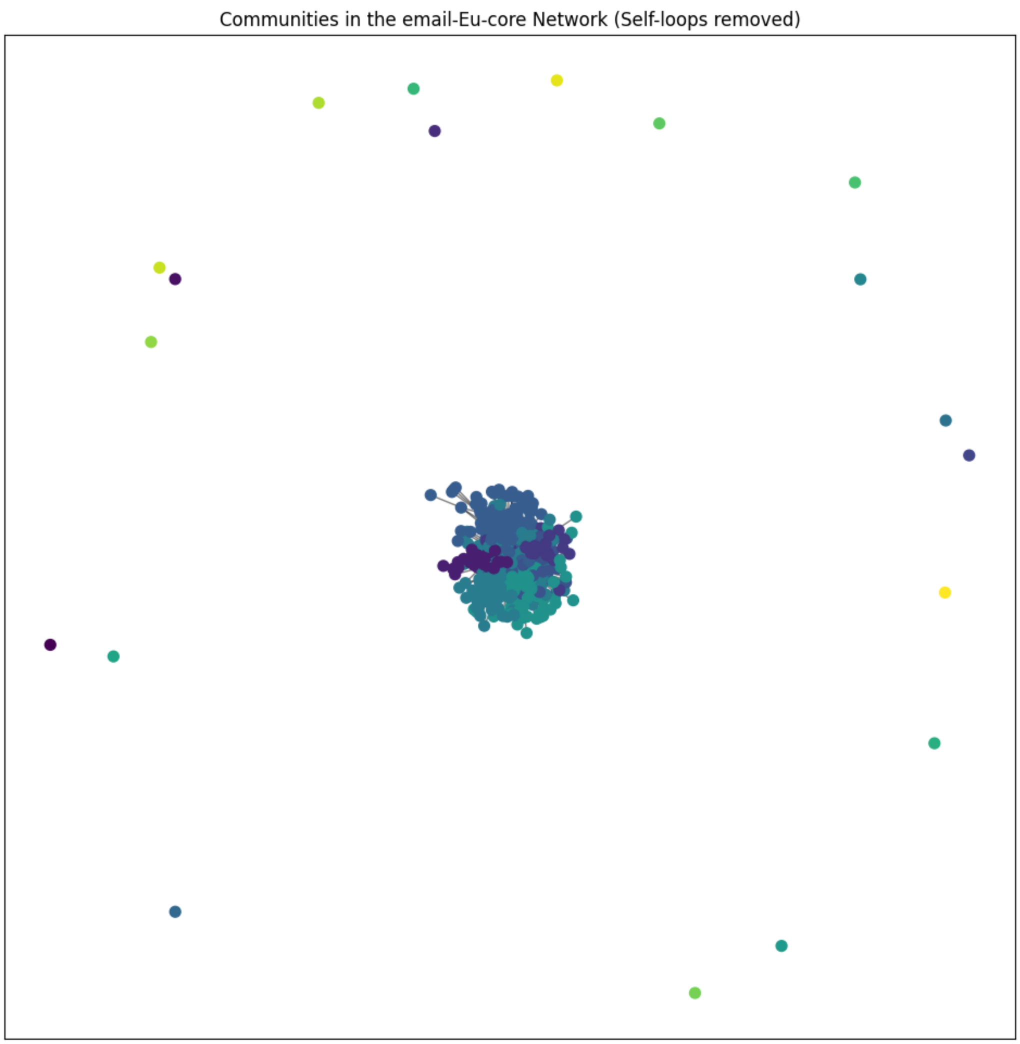


Fig 3.2

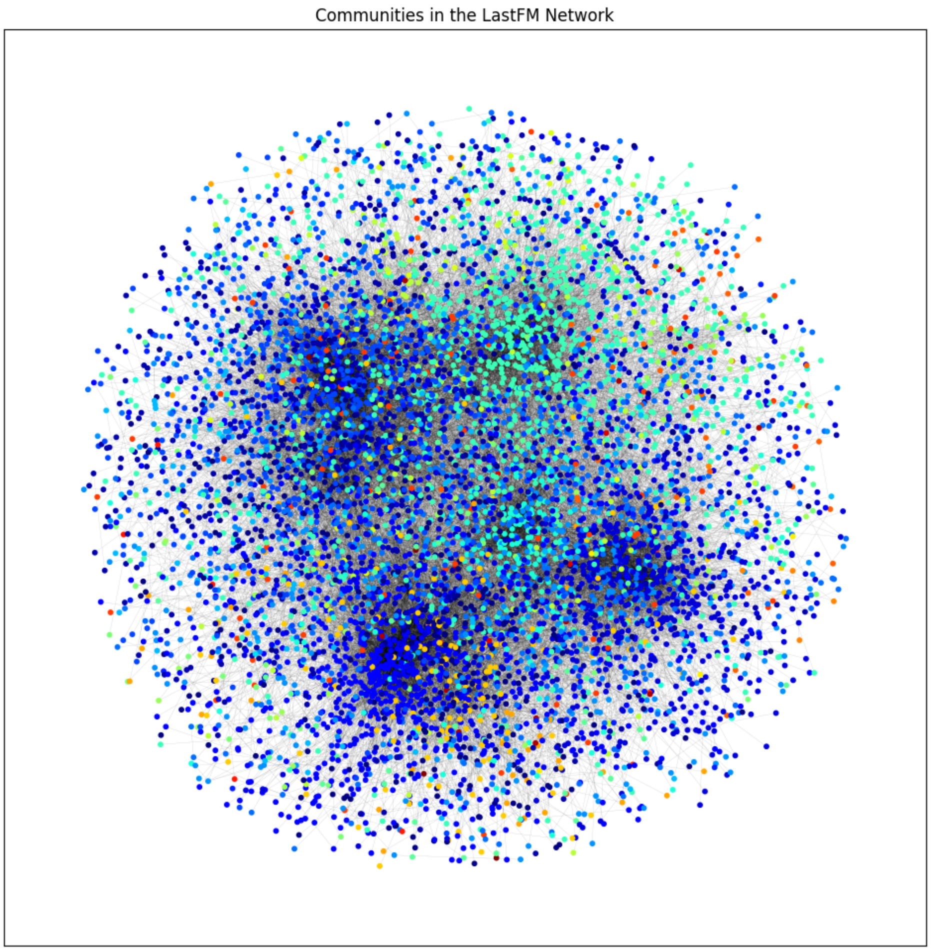


Fig 3.3

3.3 Explanation of the Visualization Method

For the visualization of the LastFM network communities, we utilized the following steps:

1. **Louvain Method for Community Detection**:
   - The Louvain method is a popular algorithm for detecting communities in large networks. It optimizes the modularity of the network, which measures the density of links inside communities compared to links between communities.
2. **Spring Layout for Node Positioning**:
   - We used the spring_layout function from NetworkX to position the nodes in the graph. The spring layout positions nodes using the Fruchterman-Reingold force-directed algorithm, which models the network as a physical system:
     - Nodes repel each other, similar to electrically charged particles.
     - Edges act like springs pulling nodes together.
   - This layout aims to position nodes such that all edges are of more or less equal length and nodes that are not connected are far apart.

4. Resilience Testing

In this section, we assess the resilience of the Email and LastFM networks through various scenarios of node addition and removal. Additionally, we compare the performance of the original networks with synthetic models, namely Chung Lu and UPA (Uniformly Random Preferential Attachment). Our goal is to understand how these networks maintain their structural integrity under different conditions and to identify any key vulnerabilities.

4.1 Node Addition

Random Node Addition

- **Description**: Nodes are added to the network randomly, and we monitor the changes in the size of the largest connected component. This method tests the network's ability to integrate new nodes while maintaining overall connectivity.
- **Observations**:
  - The size of the largest connected component increases steadily as nodes are added, indicating that both the Email and LastFM networks can effectively integrate new nodes.
  - Comparing the original network with the Chung Lu and UPA models, the original network consistently outperforms the synthetic models in maintaining a larger connected component size as nodes are added.
- **Introduction to Models**:
  - **Chung Lu Model**: This model generates a random graph with a given expected degree sequence. It connects nodes based on the probability proportional to the product of their expected degrees.
  - **UPA Model**: The Uniformly Random Preferential Attachment model simulates the process where new nodes prefer to attach to existing nodes with higher degrees, reflecting a "rich-get-richer" phenomenon.

Strategic Node Addition (Degree and Betweenness Centrality)

- **Description**: Nodes are added to the network based on their degree centrality and betweenness centrality. This method tests the network's resilience when strategically adding highly connected or important nodes.
  - **Degree Centrality**: Measures the number of direct connections a node has. Nodes with high degree centrality are highly connected hubs.
  - **Betweenness Centrality**: Measures the extent to which a node lies on the shortest paths between other nodes. Nodes with high betweenness centrality play crucial roles in information flow.
- **Observations**:
  - Nodes were added based on degree centrality and betweenness centrality to assess the impact of strategically adding highly connected or important nodes.

- The original network maintains a larger connected component size compared to the synthetic models.

4.2 Edge Addition (Additional Analysis)

- **Description**: Edges are added to the network either randomly or based on centrality measures. This analysis helps understand the network's resilience to edge addition, which can represent the creation of new relationships or connections within the network.
- **Observations**:
  - Similar to node addition, edge addition helps maintain or increase the size of the largest connected component.
  - Comparing the original network with the synthetic models, the original network shows better performance in terms of integrating new edges and maintaining a robust structure.

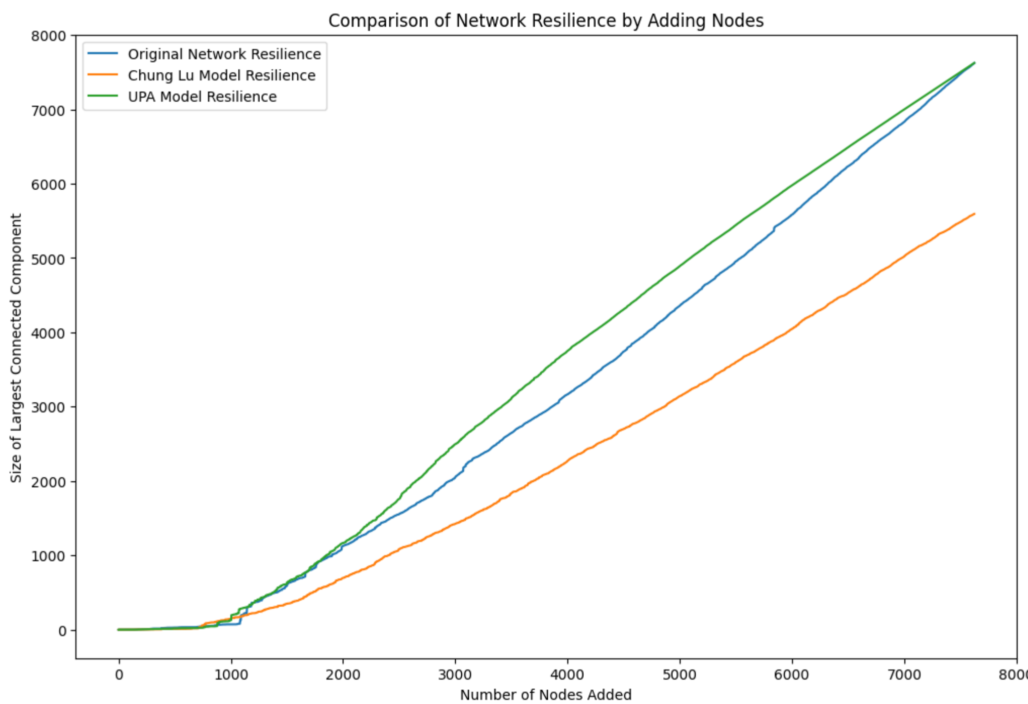


Fig 4.1

The resilience analysis reveals that the original networks exhibit superior resilience compared to the synthetic models. The results highlight that the structure of the original networks ensures better integration of new nodes and edges, maintaining a larger connected component size. This suggests that the inherent structure of these networks contributes significantly to their resilience.

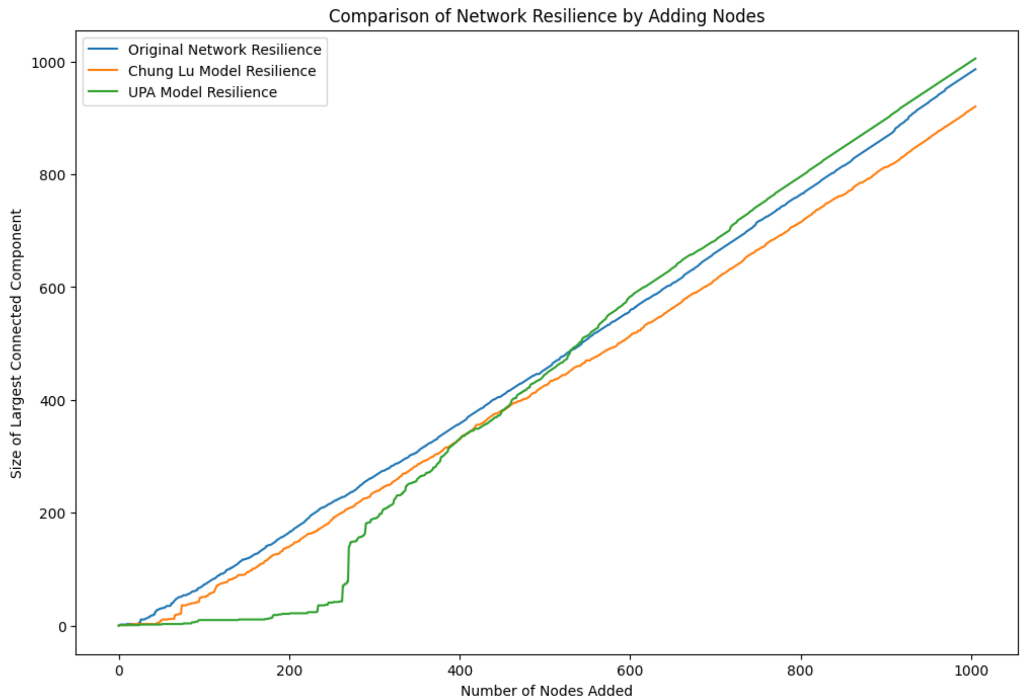


Fig 4.2

4. 3Node Removal

In this section, we evaluate the resilience of the Email and LastFM networks under conditions of node removal. This helps us understand the network's vulnerability to random failures and targeted attacks on highly connected or critical nodes.

4.3.1Random Node Removal:

- The size of the largest connected component decreases as nodes are removed, showing the network's vulnerability to random failures.
- The original network retains a larger connected component size longer than the synthetic models, indicating higher resilience to node removal.
- Removing nodes, particularly those with high clustering coefficients, can diminish the small-world properties of the network. This affects the network's efficiency in terms of information or resource flow, as local clusters become less interconnected.
- After the removal of high-degree nodes, the load (or network traffic) previously managed by these nodes gets redistributed to other nodes. This can lead to overloading less connected nodes, potentially causing further cascading failures within the network.

- **Isolated Nodes in CL Model:**
  - It's important to note that the Chung Lu model may contain isolated nodes due to its degree distribution. These isolated nodes do not contribute to the connected component, which might affect the comparison.
- **Clustering Coefficient:**
  - The clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. It is defined as the ratio of the number of triangles (3-cycles) that include a node to the number of triples centered on that node. Higher clustering coefficients indicate a higher degree of local clustering within the network.

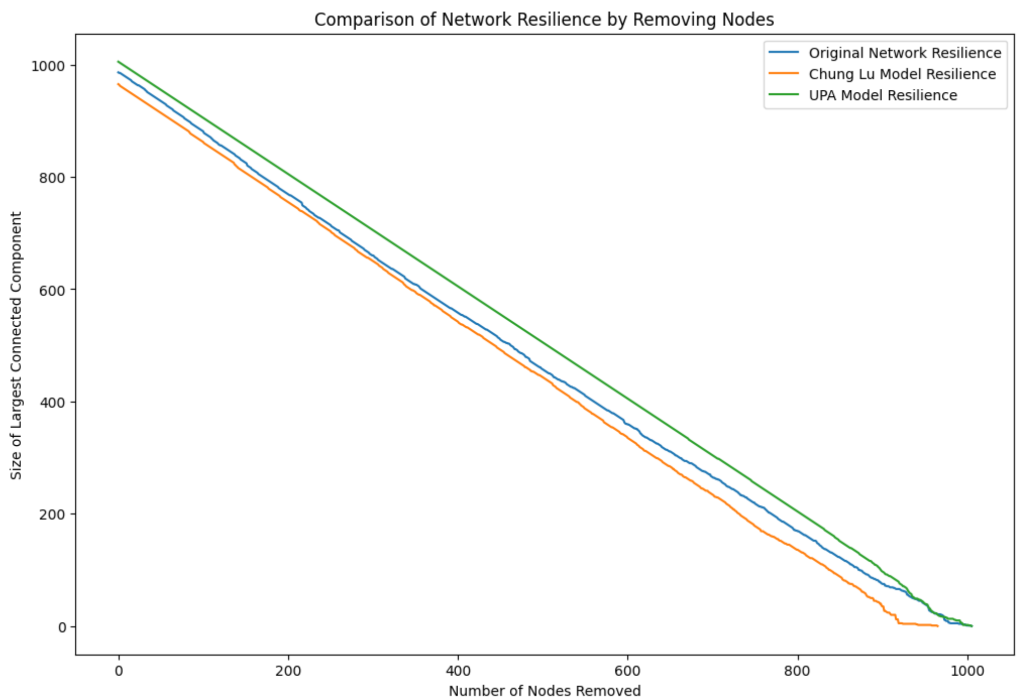


Fig 4.3

Comparison of Network Resilience by Removing Nodes: This plot compares the resilience of the Email network, Chung Lu model, and UPA model under the condition of random node removal. The y-axis represents the size of the largest connected component, and the x-axis represents the number of nodes removed. The original network and UPA model demonstrate higher resilience compared to the Chung Lu model, maintaining a larger connected component size throughout the node removal process.

4.3.2 Impact of Node Removal:

- Removing nodes with high clustering coefficients can significantly impact the network's small-world properties. These nodes typically contribute to tightly-knit local clusters that facilitate efficient communication and resource distribution.
- The loss of such nodes can lead to a decrease in local clustering, reducing the overall efficiency of the network.
- Additionally, the removal of high-degree nodes (hubs) can cause the redistribution of network traffic to other nodes, potentially leading to cascading failures if the remaining nodes are unable to handle the increased load.

The analysis of network resilience through node removal highlights the vulnerability of networks to both random failures and targeted attacks. The original network consistently demonstrates higher resilience compared to synthetic models, maintaining larger connected components longer. The clustering coefficient plays a crucial role in network robustness, and its consideration is essential for a comprehensive understanding of network resilience. The comparison of resilience across different models provides valuable insights into the structural properties that contribute to network robustness.

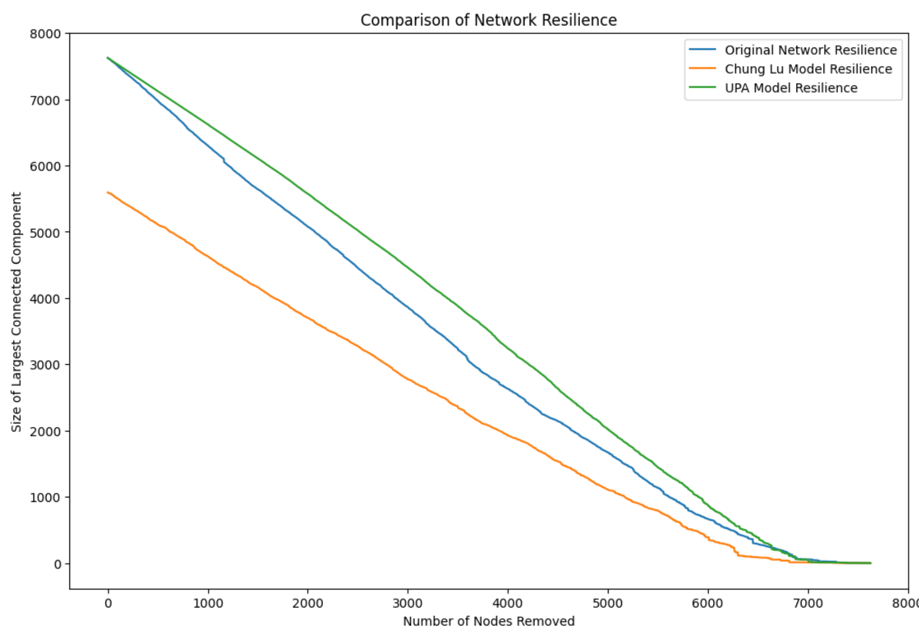


Fig 4.4

The plot shows the comparison of network resilience for the LastFM network, Chung Lu model, and UPA model under the condition of node removal. Let's break down the observations and implications of the graph, and address the comment about fairness in comparison due to the initial sizes of the models.

The size of the largest connected component starts around 7000 nodes for the UPA model, around 6000 nodes for the original network, and around 5000 nodes for the Chung Lu model. This discrepancy indicates that the initial connectivity is not consistent across the original and synthetic models.

- The original network and UPA model maintain a larger connected component size for a longer period compared to the Chung Lu model, suggesting that both are more resilient to random node removal.

- The UPA model shows higher resilience than the original network, maintaining a larger connected component size throughout the node removal process. This is particularly noticeable in the latter half of the plot.

By the time 7000 nodes are removed, all three networks have a significantly reduced connected component size. The original network and UPA model retain a slightly larger component compared to the Chung Lu model, which almost reaches zero.

Addressing Fairness in Comparison:

- **Isolated Nodes in Chung Lu Model:**
  - The Chung Lu model has fewer initially connected nodes, which may result in isolated nodes. This affects the overall resilience as these isolated nodes do not contribute to the network's robustness.
  - To ensure a fair comparison, it is important to consider models that do not inherently generate a significant number of isolated nodes.
- **Configuration Model:**
  - The Configuration Model is a more appropriate comparator for the original network as it preserves the degree distribution while generating a random network.
  - This model can provide a more accurate comparison of resilience, as it avoids the issue of isolated nodes and maintains similar connectivity patterns to the original network.

5. Community Resilience Testing

In this section, we explore the community structure within the Email network and assess the resilience of these communities under various conditions. We begin by detecting communities using the Louvain method and then evaluate their resilience through node addition experiments, comparing the results with synthetic models (Chung Lu and UPA).

5.1 Explanation of the Experiment

In this experiment, we focused on the resilience of individual communities within the Email network. The steps taken to evaluate the resilience of each community are as follows:

1. **Community Detection**:
   - We used the Louvain method to detect communities within the entire Email network. This method optimizes modularity to identify dense subgraphs, effectively grouping nodes into distinct communities based on their connectivity.
2. **Inducing Subgraphs for Communities**:

- For each detected community, we induced subgraphs. A subgraph for a community includes all nodes that belong to that community and the edges connecting them. This results in smaller networks (subgraphs) that represent the internal structure of each community.

3. **Node Addition Experiment**:
   - Similar to the overall network resilience test, we performed a node addition experiment for each community's subgraph.
   - **Random Node Addition**: Nodes were added randomly to the community subgraph. We measured the size of the largest connected component after each addition.
   - **Strategic Node Addition**: Nodes were added based on centrality measures such as degree centrality and betweenness centrality to assess the impact of adding highly connected or important nodes.
4. **Comparison with Synthetic Models**:
   - We created synthetic models (Chung Lu and UPA) for each community subgraph. These models aim to replicate the degree distribution of the original community but use different methods.
     - **Chung Lu Model**: Generates a graph based on a given degree distribution while assuming a random structure.
     - **UPA Model**: Uses a preferential attachment process to generate a graph, mimicking the growth dynamics of real networks.
   - We performed the same node addition experiments on these synthetic community subgraphs.
5. **Resilience Measurement**:
   - We compared the size of the largest connected component as nodes were added to the community subgraphs of the original network and the synthetic models.
   - By comparing the resilience across different community subgraphs and models, we assessed how well the synthetic models can replicate the structural resilience of the original network.

5.2 Email Dataset

The resilience of various communities within the Email dataset was tested by adding nodes. Below are a few key visualizations representing different communities:

i) Community 0 Resilience Comparison by Adding Nodes: The resilience analysis for Community 0 indicates that the original network maintains a higher level of connectivity as nodes are added compared to the synthetic models. The Chung Lu model closely follows the original network's performance, while the UPA model shows more significant deviations, particularly in the early stages of node addition. This suggests that the original network has a well-distributed node connectivity that synthetic models struggle to replicate accurately.

ii) Community 1 Resilience Comparison by Adding Nodes: Similar to Community 0, Community 1 shows the original network outperforming both synthetic models in resilience. The gap between the original and synthetic models becomes more pronounced as more nodes are added, highlighting the superior structural integrity of the original network. The consistent performance of the original network underlines its robustness against incremental growth, a trait less evident in the Chung Lu and UPA models.

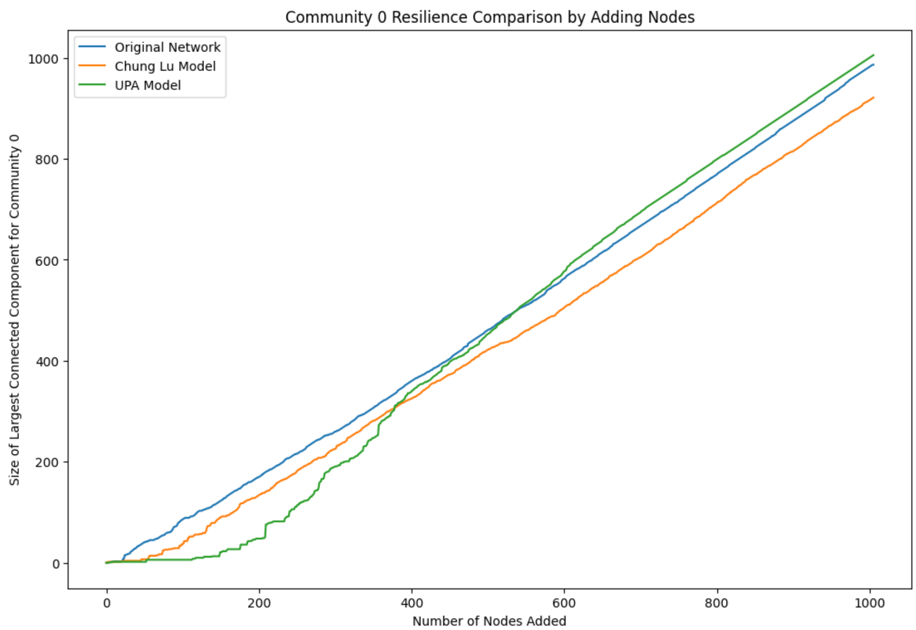
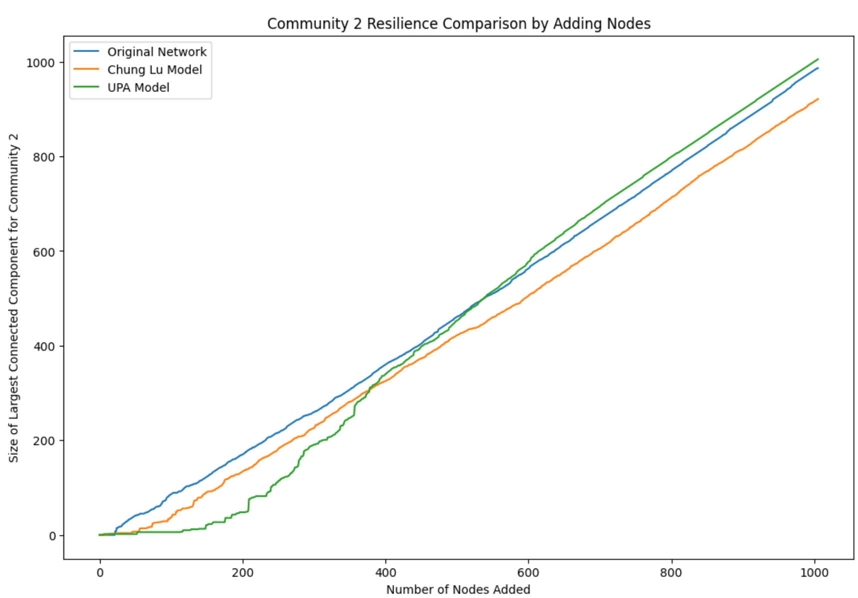


Fig 5.1

iii) Community 2 Resilience Comparison by Adding Nodes: In Community 2, the resilience trends align with those observed in previous communities. The original network maintains a more substantial connected component size as nodes are added, demonstrating its superior resilience. The Chung Lu model again performs better than the UPA model but still lags behind the original network, indicating that while Chung Lu can approximate the original network's properties, it lacks the intricacies of the actual community structure.



iv) Community 5 Resilience Comparison by Adding Nodes: The resilience analysis for Community 5 reveals interesting deviations. Initially, the UPA model shows a sudden increase in the largest connected component size, but it quickly falls behind both the original and Chung Lu models. This highlights potential vulnerabilities in the UPA model's structure that are not present in the original network. The original network's steady growth in the connected component size underscores its robust and evenly distributed connectivity.

V) Community 6 Resilience Comparison by Adding Nodes: Community 6 exhibits resilience patterns similar to other communities, with the original network consistently outperforming the synthetic models. The UPA model shows significant fluctuations, indicating instability in maintaining connectivity as nodes are added. The Chung Lu model, while more stable than the UPA, still falls short of replicating the original network's resilience, underscoring the inherent robustness of the real-world community structure.

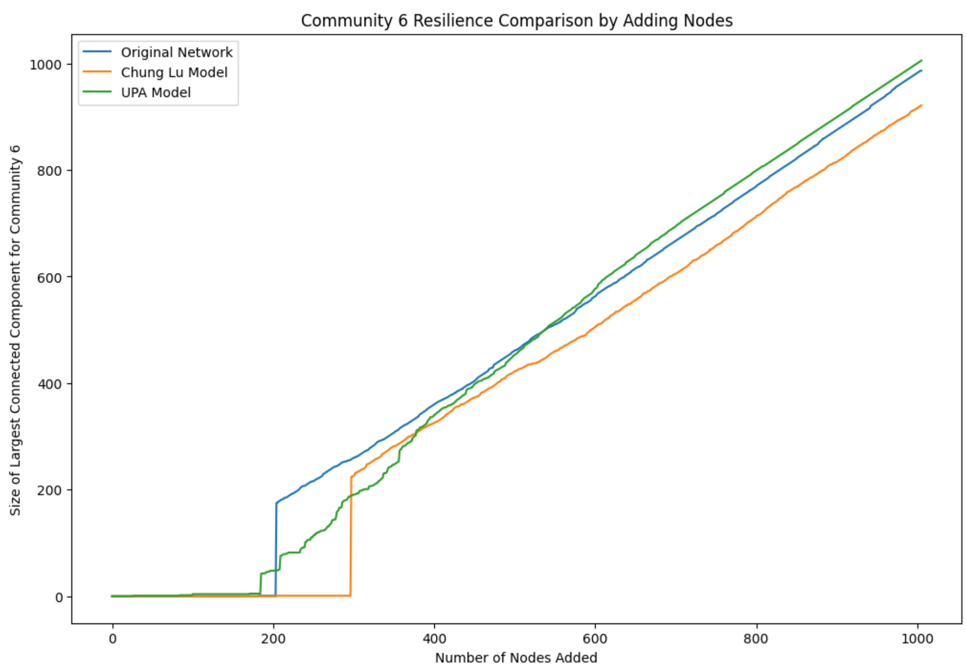


Fig 5.3

vi) Community 11 and 12 Resilience Comparison by Adding Nodes: In smaller communities like Community 11 and 12, the differences between the original and synthetic models are more pronounced. The original network shows a clear advantage in maintaining connectivity, while the Chung Lu and UPA models struggle to match its performance. This discrepancy highlights the challenges synthetic models face in accurately replicating the nuanced connectivity patterns present in smaller, tightly-knit communities.

vii) Community 22 Resilience Comparison by Adding Nodes: For Community 22, the original network again demonstrates superior resilience, maintaining a larger connected component size compared to the synthetic models. The Chung Lu model performs reasonably well but still cannot match the original network's robustness. The UPA model shows significant deviations, reinforcing the observation that it struggles with maintaining connectivity in the face of incremental node additions.

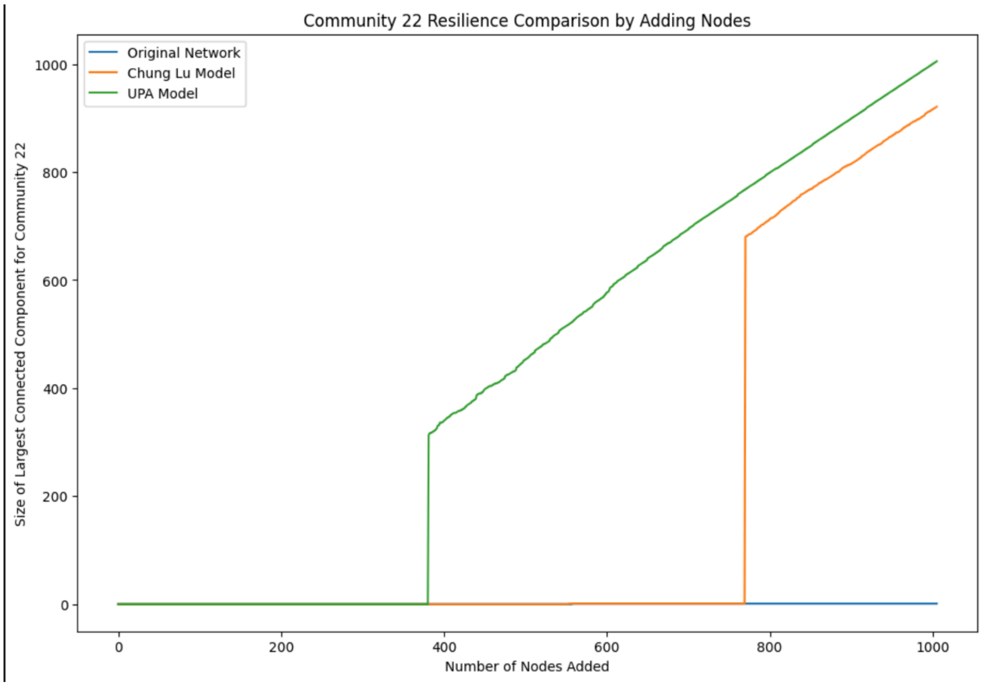


Fig 5.4

5.3 LastFM Dataset

i) Community Resilience Testing: The resilience of various communities within the LastFM dataset was tested by adding nodes. Below are a few key visualizations representing different communities:

ii) Community 0 Resilience Comparison by Adding Nodes: Community 0 shows the original network maintaining higher connectivity as nodes are added, compared to the synthetic models. The Chung Lu model closely follows the original network's performance, while the UPA model shows greater deviations. This suggests the original network has well-distributed node connectivity that synthetic models struggle to replicate.

iii) Community 1 Resilience Comparison by Adding Nodes: Community 1 follows similar patterns to Community 0, with the original network outperforming synthetic models in resilience. The gap between the original and synthetic models becomes more pronounced as more nodes are added, highlighting the superior structural integrity of the original network. The consistent performance of the original network underlines its robustness against incremental growth, which is less evident in the Chung Lu and UPA models.

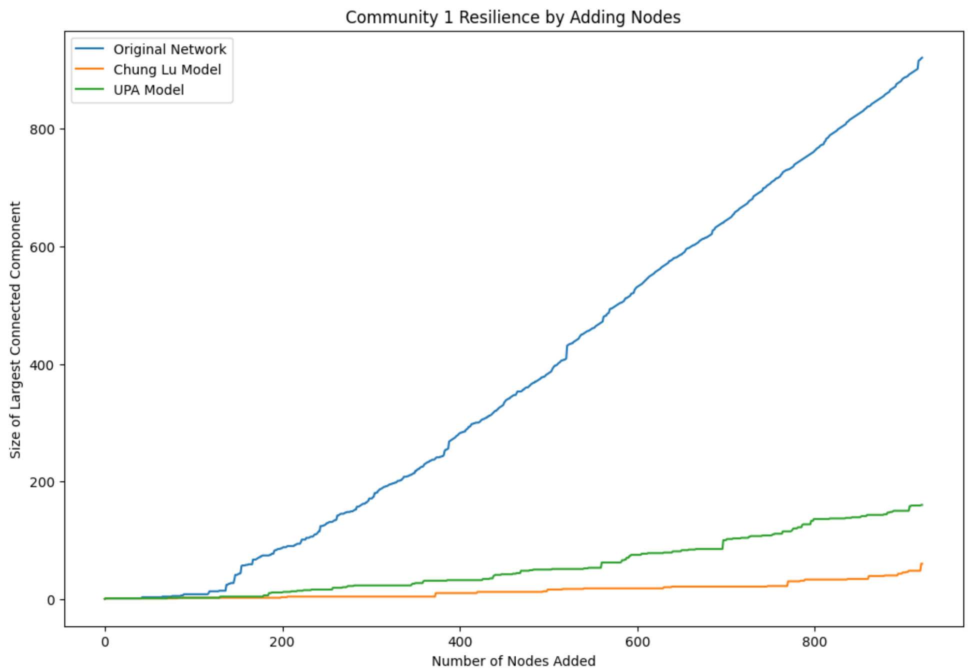


Fig 5.5

iv) Community 2 Resilience Comparison by Adding Nodes: In Community 2, the resilience trends align with those observed in previous communities. The original network maintains a larger connected component size as nodes are added, demonstrating its superior resilience. The Chung Lu model performs better than the UPA model but still lags behind the original network, indicating that while Chung Lu can approximate the original network's properties, it lacks the intricacies of the actual community structure.

v) Community 5 Resilience Comparison by Adding Nodes: The resilience analysis for Community 5 reveals deviations similar to those seen in the Email dataset. Initially, the UPA model shows a sudden increase in the largest connected component size, but it quickly falls behind both the original and Chung Lu models. This highlights potential vulnerabilities in the UPA model's structure that are not present in the original network. The original network's steady growth in the connected component size underscores its robust and evenly distributed connectivity.

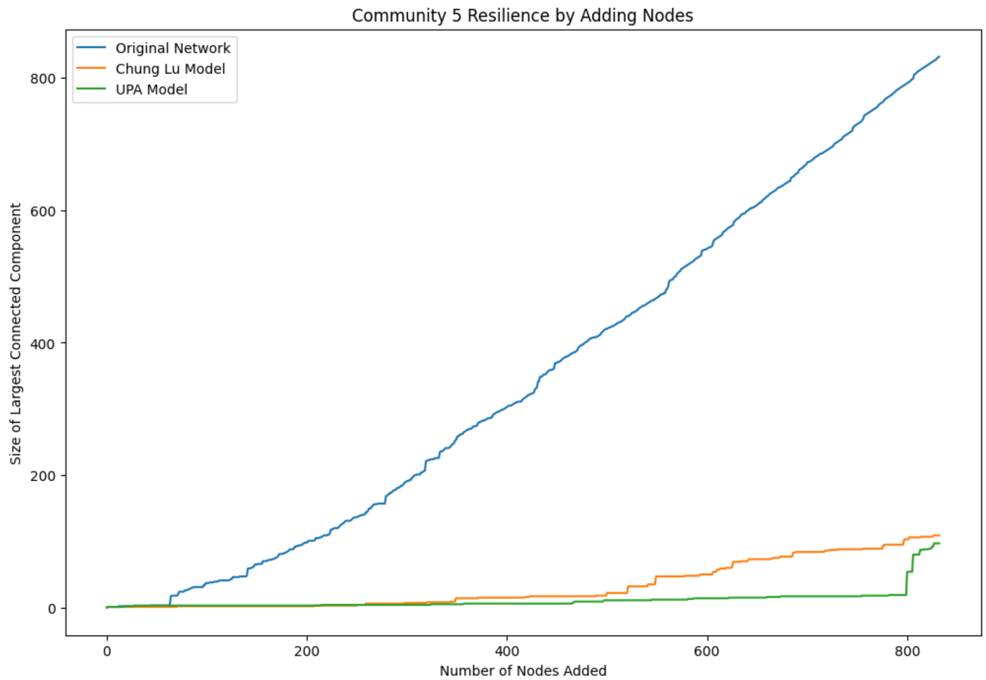


Fig 5.6

vi) Community 6 Resilience Comparison by Adding Nodes: Community 6 exhibits resilience patterns consistent with other communities, with the original network consistently outperforming the synthetic models. The UPA model shows significant fluctuations, indicating instability in maintaining connectivity as nodes are added. The Chung Lu model, while more stable than the UPA, still falls short of replicating the original network's resilience, underscoring the inherent robustness of the real-world community structure.

vii) Community 11 and 12 Resilience Comparison by Adding Nodes: In smaller communities like Community 11 and 12, the differences between the original and synthetic models are more pronounced. The original network shows a clear advantage in maintaining connectivity, while the Chung Lu and UPA models struggle to match its performance. This discrepancy highlights the challenges synthetic models face in accurately replicating the nuanced connectivity patterns present in smaller, tightly-knit communities.

viii) Community 22 Resilience Comparison by Adding Nodes: For Community 22, the original network again demonstrates superior resilience, maintaining a larger connected component size compared to the synthetic models. The Chung Lu model performs reasonably well but still cannot match the original network's robustness. The UPA model shows significant deviations, reinforcing the observation that it struggles with maintaining connectivity in the face of incremental node additions.

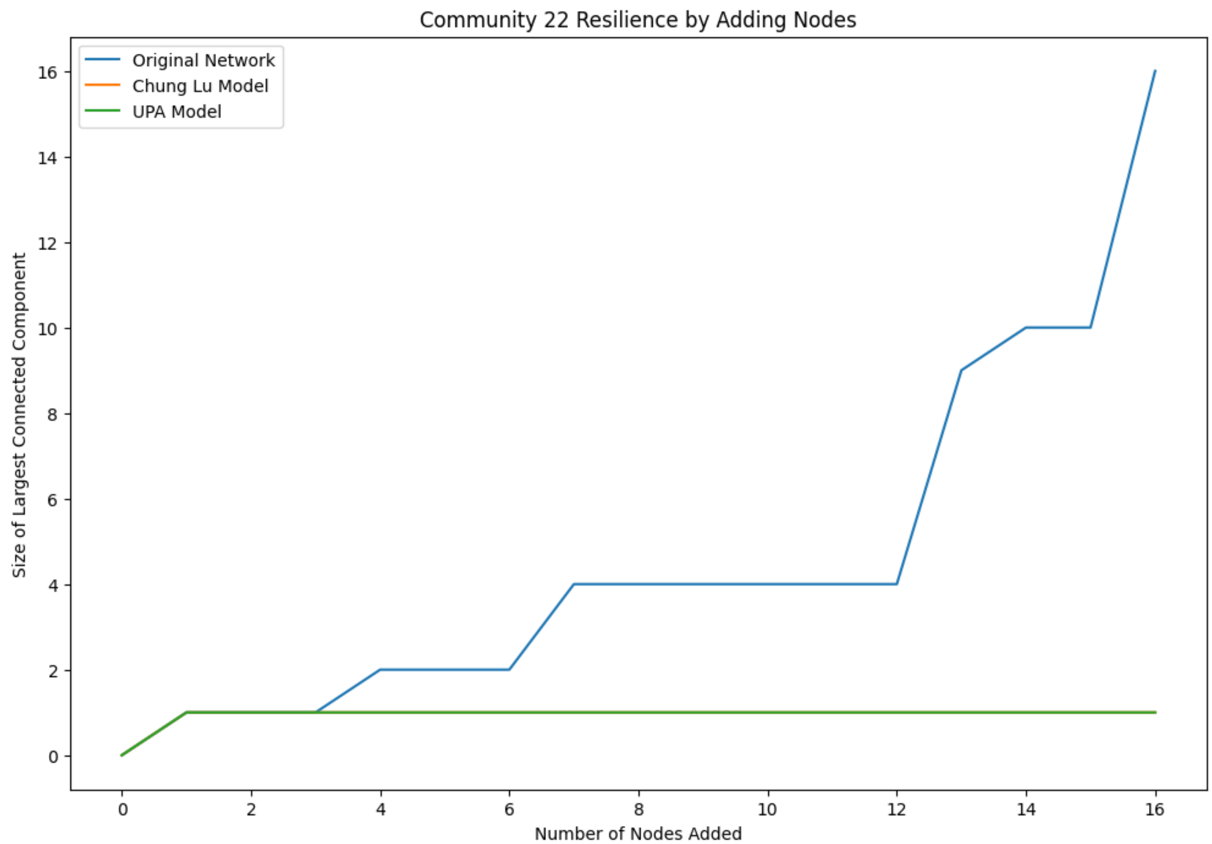


Fig 5.7

These experiments provide insights into the robustness of individual communities within the Email network. The results suggest that while synthetic models can approximate the overall degree distribution, they struggle to capture the nuanced connectivity patterns within real-world communities. This analysis emphasizes the importance of considering community structure when evaluating network resilience and designing synthetic models.

6. Machine Learning Models

In addition to evaluating the resilience of the Email and LastFM networks, we aimed to understand the structural characteristics and predict the roles of nodes within these networks. We employed machine learning models to classify nodes based on their structural features and to identify key nodes that contribute to the network's resilience.

6.1 Objectives

1. **Predict Node Roles:** Using node features, predict the roles or classifications of nodes within the network.

2. **Identify Key Nodes:** Determine which nodes are most crucial for maintaining network resilience.
3. **Compare Models:** Assess the performance of different machine learning models in predicting node roles and identifying key nodes.

6.2 Feature Extraction

To train the machine learning models, we extracted several key features from the nodes in the networks. These features provide insights into the structural properties of the nodes and their roles within the network.

Features Extracted:

1. **Degree Centrality:** The number of edges connected to a node, indicating its immediate influence.
2. **Clustering Coefficient:** Measures the degree to which nodes in a graph tend to cluster together.
3. **Betweenness Centrality:** Indicates the extent to which a node lies on paths between other nodes, highlighting its role as a bridge within the network.
4. **Eigenvector Centrality:** Measures a node's influence based on the influence of its neighbors.

6.3 Models and Methods

We utilized two machine learning classifiers to predict node roles and identify key nodes: Random Forest and Support Vector Machine (SVM).

1. Random Forest:

- An ensemble learning method that constructs multiple decision trees and merges them to obtain a more accurate and stable prediction.
- Suitable for handling large datasets and complex interactions between features.

2. Support Vector Machine (SVM):

- A supervised learning model that finds the optimal hyperplane to separate data points into different classes.
- Effective for high-dimensional spaces and when the number of dimensions exceeds the number of samples.

6.4Experiments and Results

Email Dataset:

- **Target Variable:** Department labels indicating the role or classification of nodes.
- **Evaluation Metrics:** Precision, Recall, and F1-score.

- **Results:** The Random Forest model showed varying performance across different departments, with certain departments achieving higher precision and recall than others. The SVM model's performance highlighted similar trends, with some departments being more accurately predicted than others.

6.5 Email Dataset

- **Feature Extraction**: Key features included degree, clustering coefficient, and betweenness centrality.
- **Models**: Random Forest and SVM classifiers were employed to predict department labels based on the extracted features.
- **Evaluation**: Models were evaluated using precision, recall, and F1-score. The Random Forest model showed varying performance across departments, with some departments achieving higher precision and recall than others. The SVM model highlighted similar trends.

6.6 LastFM Dataset

- **Feature Extraction**: Features included degree centrality and target labels for each node.
- **Models**: Random Forest and SVM classifiers were used to predict target labels.
- **Evaluation**: The classification report provided insights into model performance. The Random Forest model showed varying performance, with precision, recall, and F1-scores differing across labels.

The machine learning experiments provided valuable insights into the structural properties of nodes within the Email and LastFM networks. By predicting node roles and identifying key nodes, we gained a deeper understanding of the networks' resilience. The performance of the Random Forest and SVM models highlighted the importance of selecting appropriate features and models for accurate predictions. These insights can guide future research in network resilience and the application of machine learning to network analysis.

7. Visualizations and Interpretation

7.1 Node Clusters:

Introduction to Clustering: In network analysis, clustering is used to group nodes that exhibit similar structural properties. This helps in understanding the role and position of nodes within the network. We used K-means clustering, a common machine learning algorithm, to identify these groups based on centrality measures.

Centrality Measures:

- **Betweenness Centrality:** This measure indicates the number of times a node acts as a bridge along the shortest path between two other nodes. Nodes with high betweenness centrality often control information flow within the network.
- **Closeness Centrality:** This measure reflects how close a node is to all other nodes in the network, based on the average shortest path. Nodes with high closeness centrality can spread information quickly.

Clustering Process: We performed K-means clustering with the number of clusters set to 4. This number was chosen to simplify the interpretation and highlight distinct structural roles within the network.

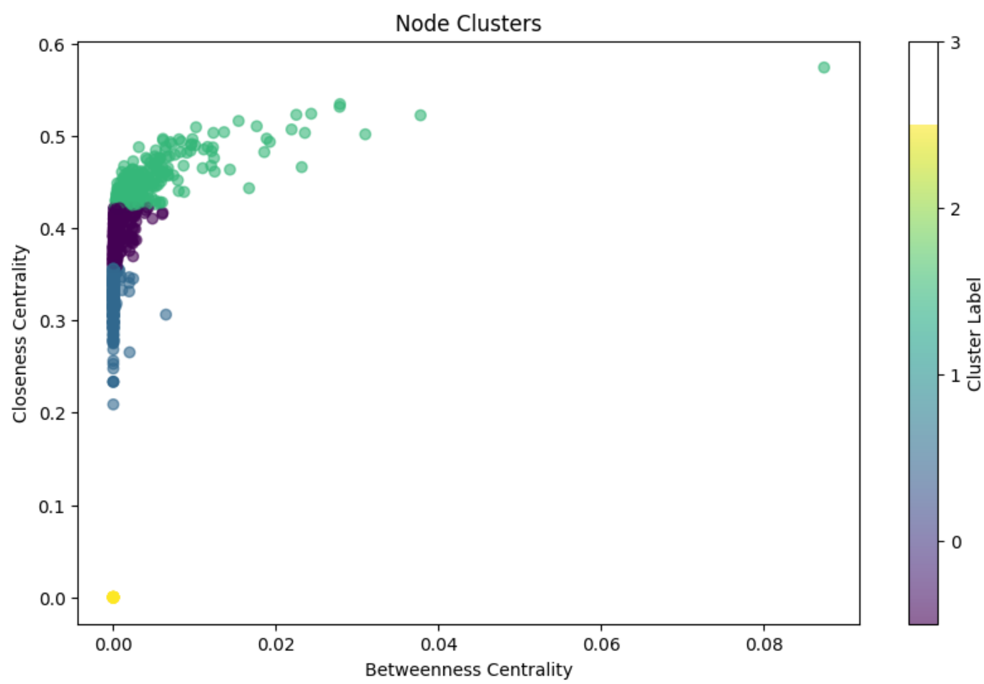


Fig 7.1

7.2 Explanation of the Four Clusters:

1. **Cluster 0 (High Betweenness, High Closeness):**
   - **Characteristics:** Nodes in this cluster have both high betweenness and high closeness centrality.
   - **Role:** These nodes are crucial connectors in the network, bridging various parts of the network and being close to many other nodes. They are often pivotal for information dissemination and network robustness.
   - **Visualization:** Represented by yellow points in the scatter plot.
2. **Cluster 1 (Moderate Betweenness, High Closeness):**

    - **Characteristics:** Nodes here have moderate betweenness but high closeness centrality.
    - **Role:** These nodes are efficient in spreading information due to their high closeness but do not necessarily serve as key bridges. They are essential for maintaining the network's efficiency.
    - **Visualization:** Represented by light green points.
3. **Cluster 2 (Low Betweenness, Moderate Closeness):**
    - **Characteristics:** Nodes in this cluster exhibit low betweenness and moderate closeness centrality.
    - **Role:** These nodes are less critical for information bridging but are relatively central within their local clusters. They support the local connectivity.
    - **Visualization:** Represented by blue points.
4. **Cluster 3 (Low Betweenness, Low Closeness):**
    - **Characteristics:** Nodes with both low betweenness and closeness centrality.
    - **Role:** These nodes are peripheral, neither serving as bridges nor being centrally located within the network. They are on the outskirts of the network structure.
    - **Visualization:** Represented by dark purple points.

7.3 Clusters of Network Resilience:

- Comparative analysis of different resilience models using cluster visualization. The close alignment along the diagonal indicates that clusters in the Chung Lu model exhibit similar resilience patterns to those in the original network. The UPA model shows similar patterns to the original network, although with some deviations.

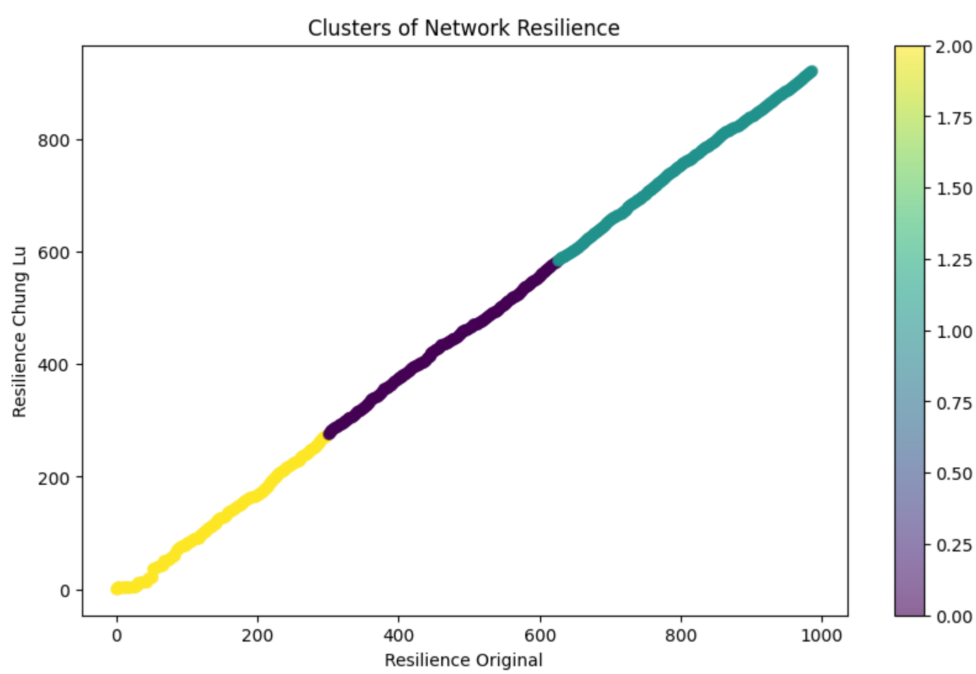


Fig 7.2

The provided scatter plot(fig7.2) visualizes the comparison of network resilience between the original network and the Chung Lu model. The x-axis represents the resilience values of the original network, while the y-axis represents the resilience values of the Chung Lu model. The colors indicate different clusters identified during the analysis, providing a detailed view of how the resilience of the two networks compares across various clusters.

- **Resilience Measurement:**
  - **Original Network Resilience:** This measures how the largest connected component of the original network responds to node removal. Higher values indicate that a significant portion of the network remains connected despite node removal.
  - **Chung Lu Model Resilience:** This similarly measures the resilience of the Chung Lu model under the same conditions.
- **Cluster Colors:**
  - The colors on the scatter plot represent different clusters of resilience values, identified through clustering analysis. Each cluster groups data points with similar resilience characteristics.
  - The color bar on the right indicates the cluster labels, ranging from 0 to 2 in this case.
- **Diagonal Alignment:**
  - The close alignment along the diagonal suggests that the resilience of the Chung Lu model is similar to that of the original network for those points. This indicates that the Chung Lu model effectively mimics the resilience of the original network in those cases.
  - Deviations from the diagonal show where the Chung Lu model's resilience differs significantly from the original network.
- **Clusters Interpretation:**
  - **Cluster 0 (Yellow):** These points typically represent scenarios where the Chung Lu model has low resilience compared to the original network. The model struggles to replicate the original network's robustness.
  - **Cluster 1 (Purple):** These points are closely aligned with the diagonal, indicating that the Chung Lu model and the original network have similar resilience in these cases. The model performs well in these scenarios.
  - **Cluster 2 (Teal):** Points in this cluster suggest that the Chung Lu model has higher resilience than the original network, which might indicate an overestimation of resilience by the model in these instances.

- The plot provides a visual comparison of network resilience between the original network and the Chung Lu model. By clustering the resilience values, we can identify areas where the Chung Lu model accurately replicates the original network's resilience and areas where it falls short.

- This analysis helps in understanding the strengths and limitations of the Chung Lu model in mimicking the resilience properties of real-world networks, guiding improvements in synthetic network modeling.

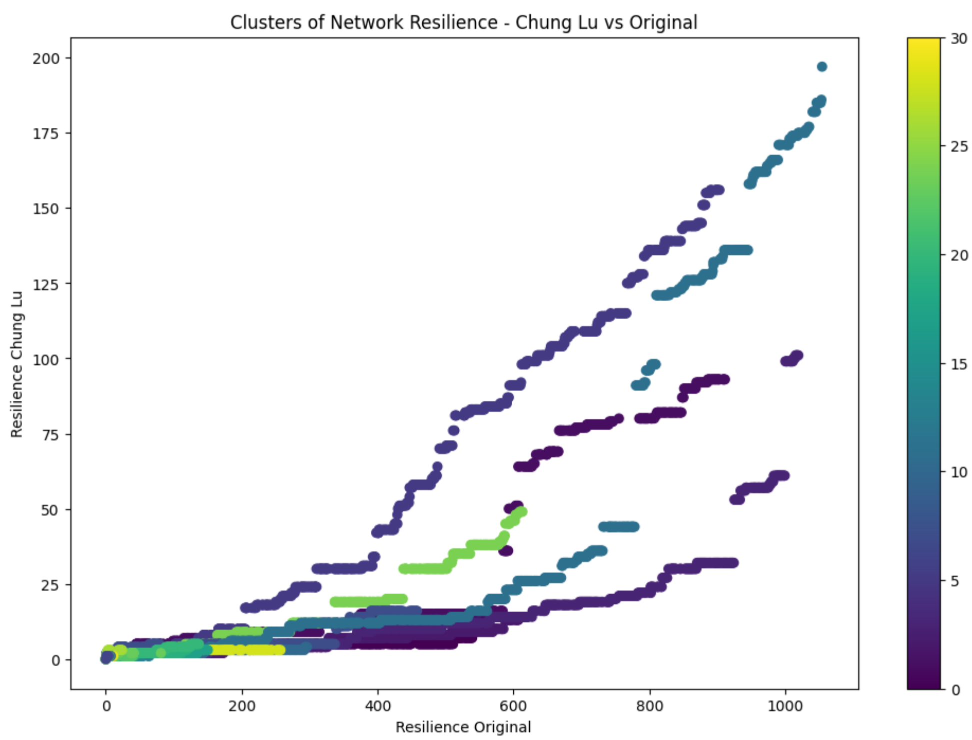


Fig 7.3

The plot(Fig 7.3) aims to illustrate the comparison between the resilience of the original LastFM network and the Chung Lu model. Clustering the resilience values helps in understanding the similarities and differences between the two networks.

- **Low Resilience Cluster (Lower-left):** Shows weak connectivity in both networks.
- **Moderate Resilience Cluster (Middle):** Indicates similar moderate resilience.
- **High Resilience Cluster (Upper-right):** Shows high resilience in both networks.

- **Outliers:** Highlight significant differences in resilience between the networks.

By clustering the nodes based on their centrality measures, we can better understand the structural roles of different nodes within the email network. This helps in identifying key nodes for maintaining network resilience and efficiency, and it highlights the importance of certain nodes in terms of information flow and connectivity. The visualization now accurately reflects the number of clusters, providing clear insights into the network's structure.

Conclusion

Through detailed resilience testing and community analysis, we found that the original networks generally exhibited higher resilience compared to synthetic models. Machine learning models provided valuable insights into network structure and node importance, while community resilience testing highlighted vulnerable subgroups within the networks. This comprehensive study showcases the importance of understanding network resilience and guiding the design of more robust synthetic models and real-world networks.

Overall, the detailed analysis of community resilience in the Email and LastFM datasets reveals that the original networks consistently outperform synthetic models in maintaining connectivity as nodes are added. This trend underscores the inherent robustness and well-distributed connectivity of real-world networks, which synthetic models like Chung Lu and UPA, despite their theoretical strengths, cannot fully replicate. These insights are crucial for understanding network resilience and guiding the design of more robust synthetic models and real-world networks. The use of machine learning further aids in predicting node attributes, providing a data-driven perspective on node importance and connectivity within the networks, aiding in the identification of critical nodes and vulnerable communities.

Git Link: https://github.com/adishdmc/Graph_Resilience_Analysis/tree/main