Towards the analysis of errors in centrality measures in perpetuated networks

Meetkumar Pravinbhai Mangroliya^{*}, Jens Dörpinghaus^{*†}, Robert Rockenfeller^{*} ^{*} University of Koblenz, Germany

Email: meetmangroliya987@uni-koblenz.de, https://orcid.org/0009-0003-3727-9527

[†] Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

Email: jens.doerpinghaus@bibb.de, https://orcid.org/0000-0003-0245-7752

Abstract—Centrality measures are essential tools for analyzing the structure and dynamics of graphs, such as knowledge or social networks. They reveal the significance and influence of individual nodes. However, their accuracy can be influenced by data quality, algorithms, and network properties. This study investigates errors in centrality measures within perpetuated networks. It focuses on network resilience and how these results may be used to develop efficient algorithms for centrality measures. It also investigates how perturbation strategies impact network resilience and predict connectivity in the perturbed network. By employing centrality measures (degree, betweenness, closeness, eigenvector), we identify critical nodes that significantly affect network connectivity and information flow. Additionally, statistical tests (Kolmogorov-Smirnov, Cramér-von Mises) assess network robustness and pinpoint critical transition points. This study, by outlining methods for error identification, quantification, and mitigation, offers valuable insights for enhancing network resilience across various domains, including infrastructure design and social network analysis.

Index Terms—Centrality Measures, Complex Networks, Social Network Analysis, Optimization on graphs

I. INTRODUCTION

In the contemporary era, networks are pervasively present. These networks may be tangible systems, such as power grids or transportation networks, or they may be abstract entities, such as networks of acquaintances or collaborations. As individuals, we are all interconnected within a network of social relationships [1], [2].

In network analysis, centrality measures provide crucial quantitative tools for assessing the importance of individual nodes within a network [3], [4], [5], [6] and are also used for AI, e.g., link prediction methods [7], [8]. Among these measures, degree centrality is particularly significant. Degree centrality quantifies the importance of a node by counting the number of direct connections it has to other nodes in the network, thus reflecting its immediate influence within the network structure [9]. To comprehend the resilience of networks to disruptions, perturbation strategies play a pivotal role [10], [4]. These strategies entail the selective or random removal of nodes or edges from a network, with the objective of studying the impact of such removals on the network's structure and properties. In particular, we focus on four removal strategies: high-degree node removal, low-degree node removal, random node removal, and random edge removal. By evaluating the change in degree centrality before and after employing these strategies, we gain insights into the relative importance of different nodes and edges in maintaining the network's overall structure and function [4], [3]. This is our first research question.

Furthermore, as a second research question, we aim to identify the critical points at which the removal of a node or edge causes the network to be fragmented and undergo significant structural changes. This information is crucial for domains that rely on network connectivity and stability, such as telecommunications and social science, as it provides insights into how these networks might be susceptible to disruption and how their resilience can be enhanced [10], [4]. By understanding these critical points, we can proactively develop strategies to improve the networks' resilience to potential failures and improve their overall robustness.

To evaluate these changes statistically, the Kolmogorov-Smirnov (K-S) and Cramér-von Mises (CvM) tests are employed [11], [12]. These non-parametric statistical tests are used to compare the degree distribution - the distribution of node degrees, representing the number of connections each node has - of a network before and after removal and show the significant change in the structure. To our knowledge, only a few studies have employed the Kolmogorov-Smirnov (KS) test to assess the differences in the degree distributions of networks [13], [11]. However, no studies have employed the Cramér-von Mises (CvM) test for this purpose. This chapter serves as the foundation for our study, laying the groundwork for exploring types of complex networks and utilizing various tools to deepen our understanding.

Following the introduction, this study is structured into four sections. The second section presents a comprehensive review of relevant literature on centrality measures, perturbation strategies, and network resilience, establishing the theoretical and empirical foundation. The third section details the methodology, outlining data collection, analysis, and statistical evaluation procedures. The fourth section presents and analyzes the experimental results, detailing the effects of different removal strategies on network properties. Finally, the fifth section presents a discussion and outlook, interpreting the results, drawing conclusions, and proposing future research directions to advance the understanding of network dynamics and resilience.

II. RELATED WORK

A. Network Perturbation

Networks are both resilient and fragile, meaning that they can withstand some perturbations but are also vulnerable to others. Node and edge removal is a technique that can be used to study how these networks respond to perturbations. Node and edge removal analysis has a rich history in network science. By systematically removing nodes or edges, we gain a deeper understanding of network robustness and efficiency. It is also worth noting that similar questions are addressed within the field of longitudinal network analysis, see [6].

Modern research examines deeper into the intricate interplay between network structure, dynamics, and resilience under various removal scenarios. Additionally, advancements in machine learning and computational techniques have allowed for the development of predictive models to estimate the impacts of node and edge removal before they are executed, enhancing our ability to formulate proactive strategies for network optimization and management. We can use mathematical models, optimization algorithms, and simulations to study the removal of nodes and edges from a network and to predict the effects of these removal actions.

Researchers have proposed and explored diverse methods for removing nodes and edges, encompassing strategies like random removal [10], [14], as well as centrality-based removals such as degree centrality and betweenness centrality [15], [16]. A prevalent attack strategy involves pinpointing critical nodes based on metrics like degree or other centrality measures and systematically removing them in descending order of importance until the network either disconnects or loses essential properties [17]. Albert's investigation into the fragmentation of random and scale-free networks employed two distinct node removal strategies: targeted removal (attacks) and random removal (errors) [10]. In the research of Smith et al. [18], the authors systematically eliminated genes from a biological network to examine the consequences on both network connectivity and behavior. Additionally, in the study of Bellingeri et al. [19], the authors performed a comparative analysis to evaluate the effects of diverse link removal strategies on various real-world complex networks. Additionally, numerous studies have examined the impact of removal strategies on complex networks across various scientific domains [20], [21], [22], [23], [24], [25]. These methodologies provide a comprehensive understanding of how the removal of nodes and edges influences the overall structure and functionality of networks.

B. Impact on Network Connectivity and Structure

Network connectivity is essential for any network. It allows information, influence, resources, and even diseases to flow between the nodes of the network. In other words, network connectivity is what makes networks useful and powerful. Without connectivity, networks would be nothing more than a collection of isolated nodes [4]. Studying how removing nodes and edges from a network affects its ability to connect its nodes is not just a theoretical exercise rather, it can help us to understand how complex systems work and how to make them more resilient. We can identify the vulnerabilities of complex systems by identifying the nodes and edges that are essential for the network's functioning. We can also identify the types of failures that are most likely to occur in a network.

Moreover, structural changes within networks, whether through the addition or removal of nodes and edges or alterations in degree distribution, can profoundly impact their overall connectivity and efficiency [10], [26]. The effects of node and edge removal on network connectivity have been a subject of extensive research in the field of network science, various studies have explored the effects of perturbation strategies on network resilience and structure. Albert et al. [10] investigated the contrasting vulnerabilities of scalefree and random networks to node perturbation. Holme et al. [17] examined the susceptibility of complex networks to targeted perturbation, while Bellingeri et al. [27] studied attack strategies in real-world networks. Wandelt et al. [15] compared network dismantling strategies, finding node removal often outperforms edge removal, with hybrid approaches effective in specific cases. Callaway et al. [28] analyzed network resilience to failures and attacks, and Chen and Li [29] explored community detection using constrained edge-deleting strategies, demonstrating potential for increased robustness.

After reviewing the work of these researchers in this field, we aim to explore how different node perturbation strategies, including targeted and random perturbation, influence network connectivity.

However, studying the impact of node and removal on different structures and connectivity is not without challenges. The interplay between local and global network structure presents complexities in predicting network behavior [17]. Furthermore, quantifying the impact of removal strategies on network robustness requires sophisticated algorithms and computational resources [26]. Moreover, the ethical implications of removal actions in real-world networks, such as social networks, introduce ethical considerations [30]. Our research aims to navigate these challenges and contribute to a more comprehensive understanding of network connectivity and structural changes due to node and edge removal.

C. Statistical Tests in Network Analysis

In statistical hypothesis testing, the Cramér-von Mises (CvM) test and the Kolmogorov-Smirnov (K-S) test are two most popular non-parametric tests to evaluate the goodness-of-fit between two probability distributions [12].

1) Kolmogorov-Smirnov (K-S) test: The Kolmogorov-Smirnov (K-S) test is a statistical tool employed to evaluate whether two cumulative distribution functions (CDF) differ significantly [11], [12]. It quantifies the maximum vertical difference between the CDF of the dataset of original degree distribution and the CDF of reduced degree distribution of the network after removal of nodes or edges [11], [31], [32].

Let $F_{1,n}$ be the cumulative distribution function of original network with a number of nodes n and $F_{2,m}$ be the cumulative distribution function of the network after node or edge removal with a remaining number of nodes m, then the K-S statistic, denoted as $D_{n,m}$, quantifies the maximum absolute difference between these functions, given by [12]:

$$D_{n,m} = \sup_{x} |F_{1,n}(x) - F_{2,m}(x)|$$
(1)

In practice, the null hypothesis is rejected at a specified significance level α if $D_{n,m}$ exceeds a critical value or in other words, if the condition in following equation 2 satisfies [12].

$$D_{n,m} > c(\alpha) \cdot \sqrt{\frac{n \cdot m}{n+m}}$$
 (2)

The value of $c(\alpha)$ at specified significance level is tabulated in [33]. The highest vertical difference between the two distribution functions is quantified by the K-S statistic $(D_{n,m})$. A significant difference between the original and modified degree distributions is shown by a high K-S statistic and a low related p-value in the context of network analysis, suggesting that the removal approach has significantly affected the network's structure.

The KS test is particularly suitable for comparing network degree distributions as it doesn't necessitate any assumptions regarding the underlying distribution of the data [13], [31]. The K-S test is exact and straightforward to interpret, delivering dependable outcomes even when dealing with small sample sizes [34]. This precision proves advantageous, especially in situations with constrained data [35], [4].

In the analysis of an evolving social network, Kossinets and Watts employed the Kolmogorov-Smirnov (K-S) test to compare the degree distributions of the network at different time points to identify and quantify the changes in the network's structure over time [35]. Their work highlighted the potential of the K-S test for analyzing the evolution of complex networks over time.

2) Cramér-von Mises Test: The Cramér-von Mises (CvM) test is another statistical method used for comparing the empirical cumulative distribution function (ECDF) of the sample to the cumulative distribution function (CDF) of a hypothesized distribution. The Cramér-von Mises W^2 criterion is named after Harald Cramér [36] and Richard Edler von Mises [37]. According to Anderson [38], the Cramér-von Mises W^2 criterion for testing that a sample, x_1, x_2, \ldots , has been drawn from a specified continuous distribution F(x) is:

$$W^{2} = \int_{-\infty}^{\infty} [F_{N}(X) - F(x)]^{2} dF(x),$$

where $F_N(X)$ is the empirical distribution function of the sample. For a second sample, y_1, y_2, \ldots, y_M , a test of the hypothesis that the two samples come from the same (unspecified) continuous distribution can be based on the analogue of W^2 , namely

$$T = \frac{NM}{N+M} \int_{-\infty}^{\infty} [F_N(x) - G_M(x)]^2 dH_{N+M}(x),$$

where $G_M(x)$ is the empirical distribution function of the second sample and $H_{N+M}(x)$ is the empirical distribution function of the two samples together.

If the computed value of T gets higher than the tabulated critical values from [38] then the null hypothesis that the two samples come from the same distribution can be rejected in the favour of alternative hypothesis. The CvM statistic (T) provides a measure of dissimilarity between the two samples.

It is general understanding that the Cramér-von Mises (CvM) test is considered a good choice for comparing distributions with heavy tails compared to the Kolmogorov-Smirnov (K-S) test due to its sensitivity to deviations in the tails of the distribution where crucial information about the network's structure is often contained [39]. There is only few studies that have employed the K-S test as the statistical tool to assess the differences in the degree distribution of the network and none to our knowledge that have employed CvM test.

III. METHOD

In our analysis, we employ a systematic methodology to generate three types of complex networks and investigate the influence of node and edge removal strategies on centrality measures within these networks. We focus on the fundamental centrality measure—degree centrality, due to its distinct and complementary perspectives on node importance within networks. Degree centrality is computed using the NetworkX library in Python, which offers efficient implementations of various centrality algorithms [40]. This measure provides valuable insights into the structural significance of nodes within complex networks.

A. Generation of Complex Networks

A random network was generated using the Erdős-Rényi (ER) graph generator from the NetworkX library [40]. The network parameters were set to include 100 nodes (n) and an edge probability of 0.2 (p) [41], [42]. This edge probability influences the network's density, with higher values resulting in denser networks and lower values yielding sparser ones. Additionally, Python's random state function was used to establish a consistent random seed (seed = 1) for reproducible network generation.

The Barabási-Albert (BA) model is widely used for generating scale-free networks [43]. This model employs preferential attachment, where new nodes connect to existing nodes with high degrees, resulting in a power-law degree distribution. To create a scale-free network using this model, we specified the number of nodes (n) and the parameter for preferential attachment (m). The parameter m dictates the number of edges added for each new node introduced, influencing the network's structure and mimicking real-world scenarios such as scientific collaborations or social interactions.

The Watts-Strogatz (WS) model is employed to generate small-world networks, characterized by both a small average shortest path length and a high clustering coefficient [44]. This model begins with a regular ring lattice where each node is connected to its k nearest neighbors. Then, with a

probability p, each link connected to a clockwise neighbor is rewired to a randomly chosen node, resulting in small-world characteristics. Varying p allows us to produce graphs with different levels of small-worldness, ranging from a regular lattice (p = 0) to a random graph with minimum connectivity constraints (p = 1).

B. Removal Strategies

In our analysis, we consider two primary node removal strategies: targeted and random. Targeted node removal involves strategically removing nodes based on specific attributes, such as high degree and low degree [10]. This approach allows us to evaluate the network's response to the intentional removal of highly connected nodes or less-connected nodes. Conversely, random node removal involves removing nodes at random, mimicking stochastic node failures or removals. Additionally, we explore edge removal strategies, focusing on the deliberate elimination of specific or random connections within the network. Random edge removal serves as a benchmark for assessing the network's resilience to random disturbances and offers insights into its robustness when faced with unpredictable edge failures [19].

C. Error Measures

To quantify the impact of node and edge removal strategies on network structures, we employ the degree distribution error as a fundamental error measure. Degree distribution error is a key metric that gauges the impact of these strategies by quantifying the dissimilarity between the degree distribution of the original network and that of the network after distinct removal strategies. We utilize the Kolmogorov-Smirnov (K-S) and Cramér-von Mises (CvM) tests to statistically evaluate the differences in degree distributions.

This approach is consistent with established methodologies in network science, providing insights into how different removal strategies impact network structure and resilience [45], [46], [13], [10], [14]. This error measure may provide insights into how different removal strategies impact network structure, resilience, and the relative importance of nodes within the network.

D. Statistical Test procedure

The data collection process using the Kolmogorov-Smirnov (K-S) and Cramér-von Mises (CvM) tests involves generating a random network, scale-free network, or small-world network to establish a baseline degree frequency distribution, which serves as reference data. This distribution is compared with the degree structure of the network after node or edge removal, executed using various strategies. The K-S and CvM tests assess whether the degree sequences of the original and reduced networks are statistically similar. The null hypothesis (H₀) posits that the degree distributions are similar, while the alternative hypothesis (H₁) suggests a difference. A significance level (α) of 0.05 is applied, balancing the risk of Type I errors and the ability to detect differences [47]. If the p-value from these tests exceeds 0.05, we conclude no significant

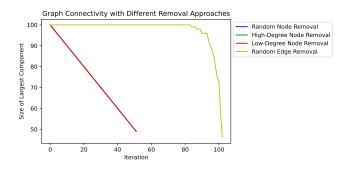


Fig. 1. Impact of Different Removal Strategies on Network Connectivity in an Erdős-Rényi Random Network

difference in degree distributions; otherwise, we reject H_0 , indicating significant dissimilarity. The results are visualized through plots and tables, elucidating the structural impacts of node or edge removal on network resilience and robustness.

IV. EXPERIMENTAL RESULTS

In this section, we examined the structural changes within random networks, scale-free networks, and small-world networks. Our approach involved applying different node and edge removal processes to observe the evolution of these networks. Specifically, for each network type, we implemented four distinct scenarios: high-degree node removal, low-degree node removal, random node removal, and random edge removal. Our primary objective is to reveal the impact of these perturbations on the networks' topology and connectivity.

A. Change in Network Connectivity

In the analysis of random networks, we employed a consistent approach to assess their structural robustness under these four distinct perturbation scenarios. Here in each iteration 1% of nodes in the node removal approach and 1% of edges in the edge removal approach are removed.

The removal of random nodes from a random network exhibited remarkable resilience, see Figure 1. Despite the stochastic nature of this process, the network displayed significant robustness. Even when a substantial fraction of nodes are removed, the structural integrity of the network is maintained, with no disconnection observed.

The same resilience was observed for the removal of highdegree and low-degree nodes, see Figure 1. The size of the largest component remained unchanged for the removal of high-degree and low-degree nodes (the respective green and yellow color line are covered by the red line in the Figure 1). The network stayed robust and connected even though the critical high-degree nodes were removed. This observation of resilience can be attributed to the random and homogeneous nature of a network.

The most striking structural robustness occurred while randomly removing edges from the network. The network's structural integrity remained totally unaffected. The network showed the robustness up to 81 iterations (81% of edge

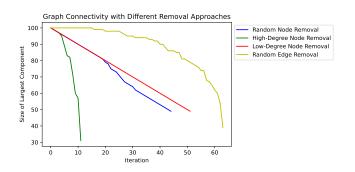


Fig. 2. Impact of different removal strategies on network connectivity in a Barabási-Albert scale-free network

removal from the original network) with the largest component of 100 nodes (original generated network with no disconnected nodes).

The network demonstrated both resilience and robustness in response to each of the four distinct types of perturbations, highlighting its ability to adapt and maintain structural integrity under a variety of challenges. The random and uniform nature of these networks contributes to their resilience under different random node and edge removal strategies. Understanding these dynamics is essential for applications in diverse fields, from information dissemination to transportation systems, where random network structures often play a vital role.

The removal of random nodes from a scale-free network demonstrated remarkable resilience, as shown in Figure 2. The network remained connected up to 21 iterations of random node removal, highlighting its robustness against the stochastic nature of this process. However, at the 35th iteration, the network faced again the disconnection, emphasizing its vulnerability to extensive random node removal.

Similarly, for the removal of high-degree nodes in a scalefree network, the network's vulnerability was evident. After just 4 iterations, the network gets disconnected, showcasing its susceptibility to the targeted removal of high-degree nodes. This disconnection resulted into 30 percent of the network as it's largest connected component after only 10 iterations.

When subject to random edge removal, a scale-free network displayed moderate resilience. It maintained connectivity up to the 15th iteration of edge removal, highlighting its ability to adapt and stay intact to some extent under this perturbation.

Conversely, a scale-free network exhibited robustness in response to low-degree node removal. The network remained connected and adaptive, demonstrating its capacity to withstand the perturbation without significant structural impact.

These findings emphasize the behavior of scale-free networks under different perturbations and emphasize the importance of understanding their vulnerabilities and strengths in real-world applications.

In the case of random node removal within a small-world network, see Figure 3, the network displayed moderate resilience to 27 iterations before facing disconnection.

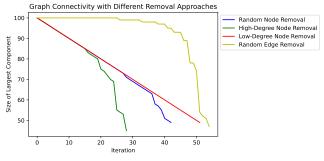


Fig. 3. Impact of Different Removal Strategies on Network Connectivity in a Watts-Strogatz Small-World Network

When subjected to high-degree node removal, a smallworld network demonstrated vulnerability, with disconnection occurring after 16 iterations. However, the network's susceptibility became apparent as early as 25 iterations, with almost a 50 percent reduction in its original size, highlighting the consequences of high-degree node removal. Conversely, low-degree node removal posed no threat to a small-world network's robustness. It remained connected throughout the removal process, even as 50 percent of the random nodes were removed, demonstrating its remarkable structural integrity.

In the analysis of random edge removal in a small-world network, the network showed moderate resilient upto around 25% of random edge removal but around 45th iteration, the network faced a significant drop, going from a 90 percent largest connected component to a 50 percent largest connected component. This dynamic response suggests that the network initially demonstrated the capability to adjust and maintain its connectivity in response to these random edge removal. However, as the edge removal increased, the network eventually reached a point where it could no longer adapt effectively and started to break down and lose its resilience.

These observations emphasize the behavior of small-world networks under various perturbation scenarios and underscore the importance of understanding their vulnerabilities and strengths in practical applications.

B. Change in Network Structure

In this subsection, we explore the impact of the node and edge removal strategies on the structure of the complex networks. We conducted a series of experiments to assess how the network's degree distribution evolves as we apply various removal strategies. These experiments were conducted on the same three types of random, scale-free, and small-world networks with the same parameters as discussed before in the methodology. Also, the same removal strategies included high-degree node removal, low-degree node removal, random node removal, and random edge removal and the removal percentage from 5% to 25%. By conducting these experiments on different network types, we aimed to gain insights into how each removal strategy influenced the degree frequency distribution of the network and, consequently, the overall

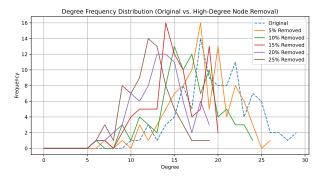


Fig. 4. Degree Frequency Distribution in Erdős-Rényi Random Network at different percentages of High-Degree Node Removal

network structure. From plot figures, the degree distribution plot where each plot illustrates a distinct removal strategy, and each color line represents the degree distribution in the corresponding network before and after a range of removal percentages. We used Non-parametric Kolmogorov-Smirnov (K-S) and Cramér-von Mises (CvM) tests with a significance level of 0.05 to statistically evaluate these changes.

1) Impact of removal strategies on Random Network: The frequency distribution plots of a random network before and after high-degree nodes removed are displayed in Figure 4. The results of the high-degree node removal strategy are summarized in Table I. The CvM and K-S statistics show a significant increase in the proportion of node removal. The CvM statistics consistently deviate from the original degree distribution, rising from 0.8888 at 5% removal to 11.0364 at 25% removal. The CvM statistics' p-values correspondingly decreased, signifying a significant deviation from the original distribution. Similar trends can be seen in the K-S statistics, which rise from 0.1984 at 5% removal to 0.79 at 25% reduction. The corresponding K-S test p-values similarly decreased to zero, confirming the significant effect of removing highdegree nodes on the degree distribution of the network. Based on these findings, the null hypothesis which states that the degree distribution of a random network is the same before and after a removal of high-degree nodes can be rejected.

The frequency distribution plots of a random network before and after the removal of low-degree nodes are shown in Figure 5. Table II provides an overview of the low-degree node removal strategy's results. As the percentage of nodes removed increases from 5% to 25%, there is a small but noticeable change in the degree distribution plot. This change is reflected in both the CvM and K-S statistics, which increase with increasing node removal. The decreasing p-values associated with the CvM and K-S statistics demonstrate that the observed changes in the degree distribution are statistically significant. As expected, the CvM test is more sensitive to the long tail of the plot than the K-S test. Given the importance of the long tail in network analysis, we can reject the hypothesis that the degree distribution remains unchanged even after removing 15% of low-degree nodes.

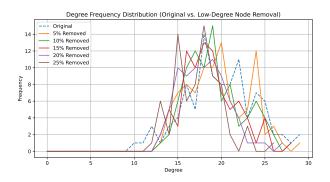


Fig. 5. Degree Frequency Distribution in Erdős-Rényi Random Network at different percentages of Low-Degree Node Removal

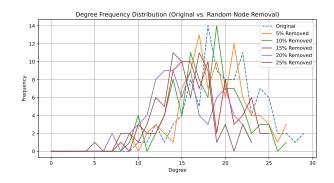


Fig. 6. Degree Frequency Distribution in Erdős-Rényi Random Network at different percentages of Random Node Removal.

The results of the random node removal method are displayed in the frequency distribution plot (see Figure 6 and Table III). At 5% removal, the CvM statistics show a slight increase to 0.1355, with a p-value of 0.4404, indicating no significant departure from the original distribution. However, as the percentage of nodes removed increases, the CvM statistics rise notably, reaching 5.5217 at 25% removal, accompanied by a very low p-value. The K-S statistics also demonstrate a progressive increase, from 0.0826 at 5% removal to 0.54 at 25% removal, with corresponding p-values diminishing significantly, reinforcing the significant deviation from the original distribution. By combining the results, we fail to reject the null hypothesis for the removal of 5% of random nodes, but we reject it for the remaining removal percentages.

The results of the random edge removal strategy, and frequency distribution plot are shown in Figure 7 and Table IV respectively. The CvM statistics increase from 0.2962 at 5% removal with a p-value of 0.1390 to 6.0382 at 25% removal with an extremely low p-value, indicating the statistical significance of structure change. The K-S statistics also show a progressive rise, from 0.12 at 5% removal with a p-value of 0.4695 to 0.5 at 25% removal with a very low p-value. The results of this study indicate how removing more than 5% of random edges significantly affects the degree distribution, causing a noticeable and statistically significant

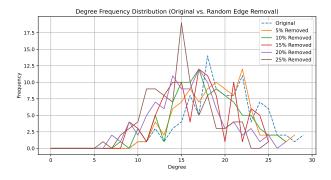


Fig. 7. Degree Frequency Distribution in Erdős-Rényi Random Network at different percentages of Random Edge Removal

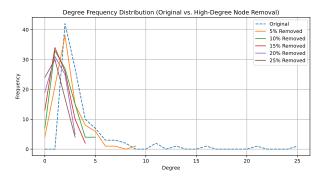


Fig. 8. Degree Frequency Distribution in Barabási-Albert Scale-Free Network at different percentages of High-Degree Node Removal

deviation from the original distribution. Consequently, for the remaining removal percentages, we reject the null hypothesis.

2) Impact of removal strategies on Scale-free Network: The results of the high-degree node removal strategy in a scale-free network are the frequency distribution plots in Figure 8 and associated statistical test results in Table V. The CvM statistics exhibit a significant increase, from 2.0114 at 5% removal to 10.1317 at 25% removal, with very low p-values consistently indicating statistical significance. Similar gradual changes are also shown by the K-S statistics, which indicate a significant shift in the degree distribution. These findings show the substantial impact of even 5% of high-degree node removal on altering the network's degree distribution in a scale-free network. Hence, we reject the null hypothesis for each removal percentage.

The results from the low-degree node removal strategy in a scale-free network, the frequency distribution plot shown in Figure 9, indicate a relatively mild impact on the degree distribution, the associated statistical results are presented in Table VI. The CvM statistics display a gradual increase from 0.0241 at 5% removal to 0.2861 at 25% removal, with p-values consistently higher than the significance value 0.05, suggesting no statistically significant deviation from the original distribution. Alongside, the K-S test demonstrates similar results, with the p-value higher than the significance value of 0.05

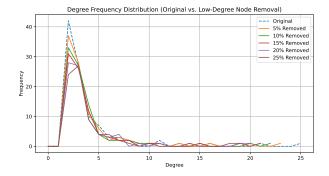


Fig. 9. Degree Frequency Distribution in Barabási-Albert Scale-Free Network at different percentages of Low-Degree Node Removal

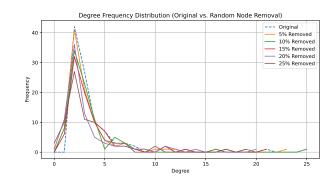


Fig. 10. Degree Frequency Distribution in Barabási-Albert Scale-Free Network at different percentages of Random Node Removal

for all removal percentages. This suggests that removing lowdegree nodes has a limited effect on the degree distribution in a scale-free network. Consequently, we fail to reject the null hypothesis.

The results from a scale-free network, as the frequency distribution plot shown in Figure 10 and the statistical results shown in Table VII, indicate that random node removal has a minimal impact on the degree distribution for removal percentages up to 15%. The CvM statistics show deviations, with p-values remaining lesser than the significance value of 0.05. However, at 20% and 25% removal, a more noticeable shift is observed, with CvM statistics reaching 0.5121 and 1.3568, respectively, and p-values dropping to 0.0371 and 0.0004, indicating a statistically significant deviation from the original distribution. Additionally, the K-S test also showed similar results up to 20% of random node removal with K-S statistics 0.15 and p-value 0.2456. This suggests that the effect of random node removal on the degree distribution becomes more pronounced as the percentage of removed nodes increases, leading to the rejection of the null hypothesis for 25% of random node removal.

The results of a scale-free network after random edge removal are illustrated in Figure 11 and associated statistical results in Table VIII. The CvM statistics progressively increase from 0.0380 at 5% removal to 1.3194 at 25% removal,

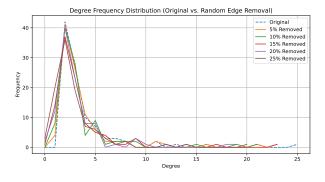


Fig. 11. Degree Frequency Distribution in Barabási-Albert Scale-Free Network at different percentages of Random Edge Removal

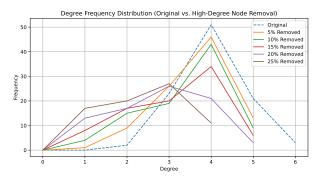


Fig. 12. Degree Frequency Distribution in Watts-Strogatz Small-World Network at different percentages of High-Degree Node Removal

with p-values decreasing significantly from 15% removal. Simultaneously, the K-S test provided similar results with a p-value below the significant value of 0.05 from 20% removal, further supporting the significant deviation from the original distribution. Here, we reject the null hypothesis at 20% and 25% removal of random edges.

3) Impact of removal strategies on Small-world Network: The frequency distribution plots in Figure 12 shows the impact of high-degree node removal on the degree distribution of a small-world network. The associated statistical results are presented in Table IX. The CvM test is more sensitive to these changes, at 5% removal, the CvM statistic is 0.5494 with a p-value of 0.0299, indicating a significant deviation from the original degree distribution. The K-S test, on the other hand, showed lower statistics and a p-value higher than the significance level of 0.05 for both 5% and 10% removal. However, after 10% removal, the K-S test also indicated a significant deviation with a p-value lower than the significance level. This suggests that the CvM test is more sensitive to changes in the degree distribution of high-degree nodes so,

including at 10% removal of high-degree nodes. Figure 13 visually represents the impact of the Low-Degree Node Removal strategy on a small-world network's degree distribution, with detailed statistical results presented in Table

we reject the null hypothesis at higher removal percentages

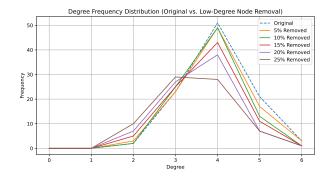


Fig. 13. Degree Frequency Distribution in Watts-Strogatz Small-World Network at different percentages of Low-Degree Node Removal

X. Initially, at a 5% removal rate, the CvM statistics of 0.0314 and a p-value of 0.9754 indicated no noteworthy deviation. However, as the removal percentage increased, statistical significance became apparent, with a CvM statistic of 0.9680 and a significantly low p-value of 0.0029 at 20% removal, further decreasing at higher percentages. The corresponding K-S test at 20% removal, featuring K-S statistics of 0.1750 and a p-value of 0.1172, suggests that low-degree node removal did impact the degree distribution, although the effect was not significant. A statistically significant shift in distribution was detected at 25% removal, with K-S statistics of 0.2700 and a p-value of 0.0032. These findings demonstrate that the removal of low-degree nodes has a gradual impact on the degree distribution of a small-world network. Here, we reject the null hypothesis only at 25% removal of low-degree nodes.

The effect of the random node removal strategy on the degree distribution of a small-world network is presented in Figure 14 and the corresponding statistical results are presented in Table XI. At 5% removal, the CvM statistics of 0.2938 and a p-value of 0.1412 indicated no significant deviation from the original distribution. However, as removal percentages increased, we can see the statistical significance. Notably, at 15% removal, a CvM statistic of 2.3229 and a very low p-value highlighted a substantial deviation from the original distribution. This was supported by the corresponding K-S statistic, which also increased to 0.3265 at 15% removal, along with a p-value lower than the significance level of 0.05. The higher sensitivity of the CvM test to changes in the distribution of high-degree nodes suggests that the observed deviation is likely due to the removal of these nodes. So, We reject the null hypothesis for the higher percentage including 15% removal of random nodes.

In Figure 15, we present the effects of random edge removal strategy on a small-world network's degree distribution and the associated statistical results are presented in Table XII. At 5% removal, CvM statistic 0.2243 and K-S statistic 0.09 indicated no significant change, aligning with p-values of 0.2260 and 0.8154, respectively. However, at 10% removal and higher removal percentages, both tests illustrated progressively significant deviations from the original distribution. By 25%

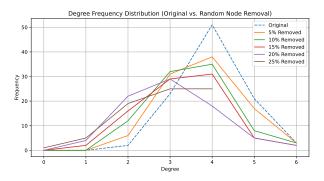


Fig. 14. Degree Frequency Distribution in Watts-Strogatz Small-World Network at different percentages of Random Node Removal

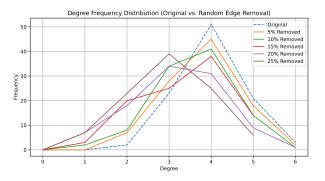


Fig. 15. Degree Frequency Distribution in Watts-Strogatz Small-World Network at different percentages of Random Edge Removal

removal, the CvM p-value and K-S p-value became very low, indicating the strategy's significant impact on the network's structural integrity. From these results, we reject the hypothesis for higher removal percentages including 10% removal of random edges.

V. DISCUSSION AND OUTLOOK

In this study, we have investigated the resilience of networks to various perturbation strategies, with a particular focus on the effects of high-degree node removal, low-degree node removal, random node removal, and random edge removal. Our research highlights the importance of understanding network resilience, particularly in the context of maintaining connectivity and stability in various real-world networks. The resilience of power grids, transportation systems, and social networks can be enhanced by gaining insights into their susceptibility to failures and the identification of critical nodes or edges whose removal could lead to significant structural changes. Furthermore, insights on critical nodes that highly influence the centrality measures for a given graph are crucial to optimize the runtime of heuristics for centrality measures in longitudinal networks and to find bounds for the possible change of node centrality measures.

The results demonstrated that random networks exhibited high resilience to connectivity disruptions, although they did lose their structural properties at a relatively low removal threshold. Scale-free networks exhibited vulnerability to highdegree node removal, but they showed resilience to other removal strategies to a moderate threshold. Small-world networks maintained connectivity up to substantial removal percentages, but they lost the structural integrity at a lower removal percentage, indicating a similar level of resilience as random networks. These findings highlight the complex relationship between node importance, network structure, and network resilience in complex networks.

Our findings demonstrate that high-degree nodes play a pivotal role in maintaining network connectivity, as their removal often leads to more significant disruptions compared to random node or edge removals. This underscores the necessity for targeted strategies to enhance network robustness, particularly by protecting or reinforcing high-degree nodes. Moreover, our study contributes to the broader field of network science by applying and validating the CvM test for degree distribution comparison, a methodology previously underutilized in this context. This development provides researchers with a novel instrument for further inquiry into the dynamics and resilience of networks.

Our study has yielded a more profound comprehension of the resilience and connectivity patterns exhibited by disparate network types. While our study offers valuable insights, it is essential to acknowledge certain limitations. Primarily, focusing on a relatively modest network comprising 100 nodes may restrict the generalizability of our findings to larger and more intricate real-world networks. Secondly, our utilization of static network analysis fails to capture the dynamic nature of real-world networks, which continuously evolve and adapt to changing conditions. Future research could enhance our understanding of network dynamics by employing temporal network analysis to investigate how networks change over time. Additionally, while our study examines various network types, the results may not be universally applicable to all network structures. Future research should address these limitations by studying larger and more intricate networks and incorporating dynamic elements. This approach would ultimately enhance the validity and applicability of our study's conclusions.

REFERENCES

- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.
- [2] J. Dörpinghaus, S. Klante, M. Christian, C. Meigen, and C. Düing, "From social networks to knowledge graphs: A plea for interdisciplinary approaches," *Social Sciences & Humanities Open*, vol. 6, no. 1, p. 100337, 2022.
- [3] S. P. Borgatti, "Centrality and network flow," *Social networks*, vol. 27, no. 1, pp. 55–71, 2005.
- [4] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [5] J. Dörpinghaus, A. Stefan, B. Schultz, and M. Jacobs, "Context mining and graph queries on giant biomedical knowledge graphs," *Knowledge* and Information Systems, vol. 64, no. 5, pp. 1239–1262, 2022.

- [7] J. Dörpinghaus, T. Hübenthal, and J. Faber, "A novel link prediction approach on clinical knowledge graphs utilising graph structures," in 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2022, pp. 43-52.
- A. Dahhani, I. Alloui, S. Monnet, and F. Vernier, "A graph matching algorithm to extend wise systems with semantic," in 2023 18th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2023, pp. 411-420.
- L. C. Freeman, "Centrality in social networks conceptual clarification," [9] Social Networks, vol. 1, no. 3, pp. 215-239, 1978.
- [10] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," Nature, 2000.
- [11] A. D. Broido and A. Clauset, "Scale-free networks are rare," Nature communications, vol. 10, no. 1, pp. 1-12, 2019.
- [12] T. B. Arnold and J. W. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions," R Journal, vol. 3, no. 2, pp. 34-39, 2011.
- A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distri-[13] butions in empirical data," SIAM review, vol. 51, no. 4, pp. 661-703, 2009
- [14] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," Reviews of Modern Physics, vol. 74, no. 1, pp. 47-77, 2002.
- [15] S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin, "A comparative analysis of approaches to network-dismantling," Scientific Reports, vol. 8, p. 13513, 2018.
- [16] E. Jenelius and L.-G. Mattsson, Resilience of Transport Systems, 05 2020.
- [17] P. Holme, B. Kim, C. Yoon, and S. Han, "Attack vulnerability of complex networks," *Physical Review E*, 2002.
- [18] J. Smith and L. Johnson, "Edge removal and network reconstruction in biological networks," PLoS ONE, vol. 10, no. 7, p. e0131180, 2015.
- [19] M. Bellingeri, D. Bevacqua, F. Scotognella, and et al., "A comparative analysis of link removal strategies in real complex weighted networks,' Scientific Reports, vol. 10, no. 1, p. 3911, 2020. [20] V. Latora and M. Marchiori, "Efficient behavior of small-world net-
- works," Physical Review Letters, vol. 87, no. 19, p. 198701, 2001.
- [21] Y. Yang, T. Nishikawa, and A. E. Motter, "Small vulnerable sets determine large network cascades in power grids," Science, vol. 358, 2017.
- [22] J. L. Caldu-Primo, E. R. Alvarez-Buylla, and J. Davila-Velderrain, "Structural robustness of mammalian transcription factor networks reveals plasticity across development," Scientific Reports, vol. 8, no. 1, pp. 1–15, 2018.
- [23] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, "Stability and topology of scale-free networks under attack and defense strategies, Physical Review Letters, vol. 94, no. 18, p. 188701, 2005.
- [24] M. Zanin and F. Lillo, "Modelling the air transport with complex networks: A short review," The European Physical Journal Special Topics, vol. 215, pp. 5–21, 2013. [25] J. Dörpinghaus, V. Weil, C. Düing, and M. W. Sommer, "Centrality
- measures in multi-layer knowledge graphs," Annals of Computer Science and Information Systems, vol. 32, pp. 163-170, 2022.
- [26] C. N. Schneider and I. Koutsopoulos, "Mitigation of malicious attacks on networks," Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems, 2011.

- [27] M. Bellingeri, D. Cassi, and S. Vincenzi, "Efficiency of attack strategies on complex model and real-world networks," Physica A: Statistical Mechanics and its Applications, vol. 414, p. 174-180, Nov. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.physa.2014.06.079
- [28] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: Percolation on random graphs," Physical Review Letters, vol. 85, no. 25, p. 5468, 2000.
- [29] X. Chen and J. Li, "Community detection in complex networks using edge-deleting with restrictions," Physica A: Statistical Mechanics and its Applications, vol. 519, pp. 181-194, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378437118315358
- [30] F. Baltar and I. Brunet, "Unintended consequences and new challenges of online social systems: Methods, challenges, and applications," Journal of Information Technology & Politics, vol. 9, no. 1, pp. 1-20, 2012.
- [31] S. Aliakbary, J. Habibi, and A. Movaghar, "Quantification and comparison of degree distributions in complex networks," in 7th International
- Symposium on Telecommunications (IST), Tehran, Iran, September 2014. M. Faloutsos, P. Faloutsos, and C. Faloutsos, "The structure of the [32] internet graph," IEEE/ACM Transactions on Networking, vol. 7, no. 3, pp. 397–412, 1999. "Table of
- [33] critical values for the two-sample https://web.archive.org/web/20130613002106/http://www.soest. test." hawaii.edu/wessel/courses/gg313/Critical_KS.pdf, memento from June 2013 in the Internet Archive. [Online]. Avail-13. https://web.archive.org/web/20130613002106/http://www.soest. able: hawaii.edu/wessel/courses/gg313/Critical_KS.pdf
- I. M. Chakravarti, R. G. Laha, and J. Roy, Engineering Statistics [34] Handbook. New York, NY: John Wiley and Sons, 1967.
- [35] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," Science, vol. 311, pp. 88-90, 2006.
- [36] H. Cramér, "On the composition of elementary errors," Skandinavisk Aktuarietidskrift, vol. 11, pp. 141-180, 1928.
- [37] R. von Mises, Wahrscheinlichkeit, Statistik und Wahrheit. Vienna Austria: Julius Springer, 1928.
- [38] T. W. Anderson, "On the distribution of the two-sample cramer-vonmises criterion," The Annals of Mathematical Statistics, pp. 1148-1159, 1962
- [39] M. A. Stephens, "Edf statistics for goodness of fit and some comparisons," Journal of the American Statistical Association, vol. 69, no. 347, pp. 730-737, 1974.
- [40] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," Proceedings of the 7th Python in Science Conference (SciPy 2008), vol. 11, pp. 11-15, 2008
- [41] P. Erdös and A. Rényi, "On random graphs, i," Publicationes Mathematicae (Debrecen), vol. 6, pp. 290–297, 1959.[42] E. N. Gilbert, "Random graphs," The Annals of Mathematical Statistics,
- vol. 30, no. 4, pp. 1141-1144, 1959.
- [43] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, pp. 509-512, 1999.
- [44] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'smallworld'networks," nature, vol. 393, no. 6684, pp. 440-442, 1998.
- [45] M. E. Newman, Networks. Oxford University Press, 2018.
- E. Estrada, "Subgraph centrality in complex networks," Physical Review [46] E, vol. 71, no. 5, p. 056103, 2005.
- [47] M. Newman, The Structure and Dynamics of Networks. Princeton University Press, 2003.

TABLE I

ERDŐS-RÉNYI RANDOM NETWORK: STATISTICAL TEST RESULTS AFTER HIGH-DEGREE NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.8888	0.0045	0.1984	3.60×10^{-2}
10%	3.0736	5.02×10^{-8}	0.3733	2.11×10^{-6}
15%	6.2458	1.97×10^{-11}	0.5076	2.46×10^{-11}
20%	8.9809	2.73×10^{-10}	0.6575	1.09×10^{-18}
25%	11.0364	2.53×10^{-10}	0.7900	4.36×10^{-27}

TABLE II

ERDŐS-RÉNYI RANDOM NETWORK: STATISTICAL TEST RESULTS AFTER LOW-DEGREE NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.0599	0.8219	0.0721	0.9394
10%	0.2897	0.1451	0.1322	0.3430
15%	0.6399	0.0178	0.1829	0.0791
20%	1.2772	0.0006	0.2550	0.0051
25%	2.2123	4.16×10^{-6}	0.35	3.79×10^{-5}

TABLE III

ERDŐS-RÉNYI RANDOM NETWORK: STATISTICAL TEST RESULTS AFTER RANDOM NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.1355	0.4404	0.0826	0.8563
10%	0.9319	0.0035	0.2	0.0383
15%	2.8869	1.30×10^{-7}	0.3518	1.42×10^{-5}
20%	3.4546	7.19×10^{-9}	0.4025	6.09×10^{-7}
25%	5.5217	3.51×10^{-10}	0.54	5.85×10^{-12}

TABLE IV

ERDŐS–RÉNYI RANDOM NETWORK: STATISTICAL TEST RESULTS AFTER RANDOM EDGE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.2962	0.1390	0.12	0.4695
10%	1.1043	0.0014	0.23	0.0099
15%	2.5327	8.03×10^{-7}	0.33	3.21×10^{-5}
20%	4.1507	2.12×10^{-10}	0.43	1.12×10^{-8}
25%	6.0382	9.28×10^{-12}	0.5	1.00×10^{-11}

TABLE V

BARABÁSI-ALBERT SCALE-FREE NETWORK: STATISTICAL TEST RESULTS AFTER HIGH-DEGREE NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	2.0114	1.18×10^{-5}	0.2632	1.78×10^{-3}
10%	4.5115	3.88×10^{-11}	0.4444	5.88×10^{-9}
15%	7.2186	3.69×10^{-10}	0.5529	1.79×10^{-13}
20%	9.0646	3.25×10^{-10}	0.6250	8.85×10^{-17}
25%	10.1317	2.38×10^{-9}	0.7200	5.89×10^{-22}

TABLE VI

BARABÁSI-ALBERT SCALE-FREE NETWORK: STATISTICAL TEST RESULTS AFTER LOW-DEGREE NODE REMOVAL

Percentage	CvM Stat.	P-Value (CvM)	K-S Stat.	P-Value (K-S)
5%	0.0241	0.9934	0.0305	0.9999
10%	0.0840	0.6754	0.0533	0.9978
15%	0.1083	0.5505	0.0553	0.9969
20%	0.1637	0.3528	0.0700	0.9707
25%	0.2861	0.1485	0.1000	0.7523

TABLE VII Barabási-Albert Scale-Free Network: Statistical Test Results after Random Node Removal

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.0363	0.9564	0.0326	0.9999
10%	0.1855	0.2993	0.0889	0.8092
15%	0.2839	0.1508	0.0976	0.7269
20%	0.5121	0.0371	0.1500	0.2456
25%	1.3568	0.0004	0.2133	0.0354

TABLE VIII

BARABÁSI-ALBERT SCALE-FREE NETWORK: STATISTICAL TEST RESULTS AFTER RANDOM EDGE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.0380	0.9488	0.0500	0.9997
10%	0.1921	0.2849	0.0800	0.9084
15%	0.4865	0.0431	0.1500	0.2112
20%	0.7992	0.0073	0.2100	0.0241
25%	1.3194	4.42×10^{-4}	0.2100	0.0241

TABLE IX

WATTS-STROGATZ SMALL-WORLD NETWORK: STATISTICAL TEST RESULTS AFTER HIGH-DEGREE NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.5494	0.0299	0.1289	0.3528
10%	1.1332	1.19×10^{-3}	0.1911	0.0538
15%	2.3875	1.69×10^{-6}	0.2794	1.14×10^{-3}
20%	4.9056	3.68×10^{-11}	0.4500	1.29×10^{-8}
25%	8.0887	3.35×10^{-11}	0.6033	5.00×10^{-15}

TABLE X

WATTS-STROGATZ SMALL-WORLD NETWORK: STATISTICAL TEST RESULTS AFTER LOW-DEGREE NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.0314	0.9754	0.0295	0.9999
10%	0.1702	0.3356	0.0844	0.8536
15%	0.4109	0.0677	0.1029	0.6668
20%	0.9680	2.90×10^{-3}	0.1750	0.1172
25%	1.7561	4.44×10^{-5}	0.2700	3.16×10^{-3}

TABLE XI

WATTS-STROGATZ SMALL-WORLD NETWORK: STATISTICAL TEST RESULTS AFTER RANDOM NODE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.2938	0.1412	0.0926	0.7507
10%	0.6479	0.0170	0.1722	0.1050
15%	2.3229	2.36×10^{-6}	0.3265	7.45×10^{-5}
20%	3.3379	1.30×10^{-8}	0.3750	4.56×10^{-6}
25%	4.8054	2.83×10^{-11}	0.4567	1.47×10^{-8}

TABLE XII

WATTS-STROGATZ SMALL-WORLD NETWORK: STATISTICAL TEST RESULTS AFTER RANDOM EDGE REMOVAL

Percentage	CvM Statistic	P-Value (CvM)	K-S Statistic	P-Value (K-S)
5%	0.2243	0.2260	0.0900	0.8154
10%	1.0744	1.64×10^{-3}	0.2300	0.0099
15%	1.9970	1.27×10^{-5}	0.2900	4.12×10^{-4}
20%	3.4719	6.63×10^{-9}	0.3900	3.57×10^{-7}
25%	4.1784	1.84×10^{-10}	0.4200	2.75×10^{-8}