Understanding Academic Collaboration Networks Through Small-World Theory: Insights for High-Impact Research

Rohan Bali[1]

Abstract

In today's rapidly growing research landscape, the significance of academic collaboration cannot be overstated at all; it serves as a cornerstone for impactful innovation and advancement of research. This study dives deeper into the dynamics of academic collaboration networks through the eyes of small-world network theory, looking forward to uncovering the structural patterns that encourage knowledge exchange and enrich research impact. This network is presented as a graph where each node symbolizes each researcher, and every edge marks the co-authorship relation between two or more individual researchers. Data from OpenAlex, accessed via PyAlex, enables the creation of a detailed co-authorship network. Identified the nodes with a higher degree, highlighting their role in knowledge propagation. The insight suggests that focused support toward highly connected individual researchers can enhance research productivity. Overall, this study advances our understanding of academic collaboration networks and provides a basic framework for optimizing collaborative research environments all over the world.

Keywords: Academic Collaboration Networks, Small-World Network Theory, Co-Authorship Analysis, Knowledge Exchange, Research Impact, Watts-Strogatz Model, Network Analysis

Introduction

Academic collaboration is crucial in promoting impactful research and innovation in an exponentially increasing interconnected research landscape. This project focuses on analyzing these networks to understand interconnectivity among researchers. It aims to reveal and study the patterns that enhance effective knowledge exchange and innovation.

The concept of a small-world network was introduced by Stanley Milgram; it is characterized by short average path lengths and high clustering coefficients, which have been observed in various real-world networks, from social connections to biological systems. Later, the formal mathematical framework for small-world networks was developed by Duncan J. Watts and Steven Strogatz. Applying small-world theory to academic collaboration networks provides insights into how research communities are structured and reveals optimal patterns for knowledge transfer and collaborative productivity. This very analysis has implications for academic institutions and funding bodies looking forward to allocating resources strategically to support research areas with comparatively higher impact.

In this study, an academic collaboration network is a graph where each node represents a unique author or researcher, and each edge corresponds to a co-authorship link. To construct this network, data has been gathered using the PyAlex Python library, extracting information on Authors, Co-Authors, and Works from OpenAlex. After refining and processing the data to streamline it, unique numerical identifiers, such as 1,2,3,4, etc., have been assigned to each author in the dataset, providing a straightforward representation of researchers within the graph. This structured approach allows the building of a network that reflects real-world academic collaborators.

Methodology

The methodology majorly focuses on identifying and then analyzing small-world characteristics within academic collaboration networks. Precisely, aiming to measure the degree distribution of a graph to investigate the network's structural efficiency in knowledge transfer throughout the network. Using the Random Graph *G (n, p)* model as a comparison tool, able to evaluate the limit to which the academic collaboration network confirms small-world principles, known for facilitating quick information spread and enhanced clustering.

1. **Data Collection and Preprocessing:** Data has been collected from OpenAlex using the PyAlex python library, which extracts detailed information on authors, their co-authors, and their published works. The data is then cleaned, and each unique author is assigned a numerical identifier to construct a simplified and manageable graph structure.
2. **Construction of Graphs:** With the already preprocessed data, an empirical network graph in NetworkX was created, where each specific node represents each specific author, and every edge represents a co-author relationship. This graph acts as a foundation for further analysis by offering a mathematical and visual representation of real-world academic collaborations.
3. **The *G (n, p)* Network Model:** For creating a comparable theoretical model, use the *G (n, p)* model, which exactly represents random world characteristics. By creating a model network with parameters aligned to our academic collaboration network, we can directly compare the two graphs in terms of clustering and average path length.
4. **Comparison and Validation of Academic Collaboration Graph:** The last step includes comparing the degree distribution of our empirical academic collaboration network, the *G (n, p)* model network. This comparison will help in the validation of the presence of small-world attributes within the academic collaboration network, give us insights into the standard of clustering, and give insights into its efficiency in promoting meaningful and impactful research.
5. **Identification of Central Nodes:** By identifying the nodes with the highest degree, recognize central figures (or major figures) who play an integral role in knowledge propagation. These highly connected individual researchers or groups serve as a hub in the network, promoting rapid information or data transfer across the community.

Literature Review

Academic collaboration networks have become important in understanding the dynamics of knowledge transfer. These networks built using relationships such as co-authorship, collaborative works, etc., act as rich visualizations of how researchers interact within their own and across the disciplines. This review dives deeper into the theoretical underpinning of academic collaboration networks, focusing on the building of these networks, the Erdős–Rényi random network model, and the classification of nodes with higher degrees. Additionally, data collection techniques using PyAlex Python Library to gather information about authors, co-authors, and their respective works, which play a crucial part in forming these networks, are being discussed.

1. Theory of Academic Collaboration Networks

The concept of an academic collaboration network can be traced back to the early work in bibliometrics, where different researchers looked forward to analyzing citation patterns to understand the transmission of knowledge. Price[1] introduced the concept of invisible colleges, a term that refers to a tightly interconnected group of researchers collaborating intensively within certain fields. This idea then laid a foundation for the need for the development of academic collaboration networks where each node represents an individual researcher, and an edge represents collaborative links. Over time, researchers have applied graph theory and complex network analysis to study the structure and dynamics of such networks. Academic collaboration networks are typically known for their non-random structures. Researchers frequently collaborate within small groups or clusters that are linked by a few influential researchers. These patterns represent small-world phenomena seen in many complex systems, where nodes are not directly connected but are just a few steps away. This characteristic is very relevant for such collaboration networks, where the ability to communicate with experts in their respective fields can lead to interdisciplinary breakouts.

1.1 Importance of Academic Collaboration Networks

The importance of academic collaboration networks lies in the insights and outcomes they provide into innovation, research productivity, and the structure of knowledge exchange. For instance, network analysis has shown that breakthroughs are often the result of cross-disciplinary collaborations. Also, understanding the degree of certain nodes (i.e., researchers) within this network can help identify the key influencers and facilitate resource allocation in research funding. Also, analyzing collaboration networks allows for identifying patterns that influence scientific communication, such as the role of geographic location, institutional affiliation, and available resources. These networks thus serve not only as a means of studying scientific behavior but also as an instrument for promoting effective knowledge exchange, especially in this highly interconnected world of research.

2. The $G(n, p)$ Model (Erdős–Rényi random graph model)

The $G(n, p)$ model, or the Erdős–Rényi Model, is a fundamental tool in random graph theory introduced by Gilbert in 1959 Erdős and Rényi in 1960[2][3]. This network is commonly used to study networks formed by a probabilistic approach, and it provides crucial insights into the characteristics of random network graphs. The *G (n, p)* model defines the graph by randomly assigning edges to all the possible pairs of nodes with a probability of *p.*

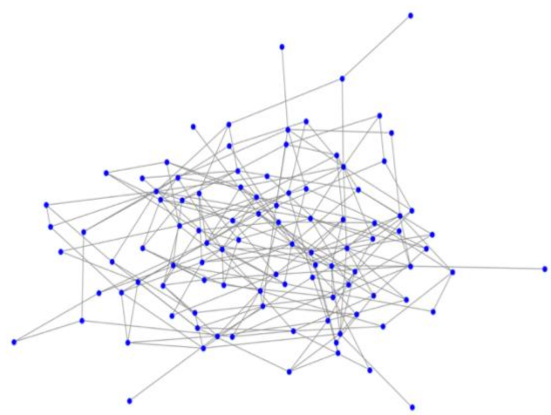


Fig. 1

Fig. 1 represents an instance of a randomly generated *G (n, p)* graph using the NetworkX library, where *n*=100 and *p*=0.05.

2.1 Key Features of a *G (n, p)* Model

2.1.1 Nodes and Edges

The graph contains *n* nodes, and an edge probability p, independent of all other nodes, connects each pair of nodes.

2.1.2 Randomness

This model assumes that the addition and removal of edges are completely independent for every pair. This probabilistic nature makes the *G (n, p)* model particularly simple yet very powerful for network behavior analysis.

2.2 Mathematical Properties

2.2.1 Degree Distribution

In *G (n, p*), the degree distribution, the degree *k* of a node (number of edges connected to a node), follows a Binomial Distribution as the addition of each edge is a Bernoulli trial with a probability *p*

$$P(\deg(v) = k) = \binom{k}{n-1} p^k (1-p)^{n-1-k}$$

For larger values of *n*, and if np→λ, the degree distribution[3] is close in on the Poisson Distribution

$$P(\deg(v) = k) \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

where $\lambda=np$ is the expected degree of a node.

2.2.2 Clustering Coefficient

The clustering coefficient[4] in *G (n, p)* depends entirely on the edge probability *p*.

$$C = p$$

representing that the likelihood of two neighbors of a node being connected is independent of the structure of the network.

3. Comparing Graphs: Academic Collaboration Network Graph vs. The *G (n, p)* Model Graph

To evaluate the impact of small-world networks in academic collaboration networks, it is crucial to compare the academic collaboration model to the *G (n, p)* model, as it is a randomly generated network with entirely different characteristics to the ideal small-world network. The comparison is done based on power-law degree distribution in both network graphs. Therefore, the comparison, in turn, helps in the validation of the academic collaboration network.

4. Classification of Nodes with Degree

Academic collaboration networks are a subset of social networks in which nodes represent individual researchers and edges signify collaborations or co-authorship. Among the metrics used to access network properties, the degree distribution has emerged as an important measure to identify the nodes that are a part of tightly knit or connected groups. A high degree is also associated with influential nodes, encouraging impactful research and creating a collaborative environment.

4.1 Degree of Nodes in the Academic Network

In a network, a node's degree represents the number of edges that link to it. In an academic cooperation network, nodes represent scholars or institutions, while edges reflect co-authorship links between the two. The degree of a node in such a network indicates how many collaborations that researcher has with others. Mathematically, the degree is defined as:

$$k_i = \sum_j A_{ij}$$

where k_i is the degree of node i and A_{ij} is the adjacency matrix element between i and j.

4.2 Types of Degree Distributions in Academic Networks

The degree distribution[9] in academic collaboration networks typically follows a power-law distribution, which suggests that most researchers have few collaborations while a small number of researchers have many collaborations (often called hubs). The power-law nature of the distribution indicates that the network is scale-free, meaning that the number of nodes with a specific degree follows the relationship:

$$P(k) \sim k^{-\gamma}$$

where *P(k)* is the probability of a randomly selected node having degree k_i and γ is a constant that ranges from 2-3 in real-world networks. The $\sim$ symbol is used here to denote that *P (k)* is proportional to $k^{-\gamma}$ asymptotically, meaning the relationship holds for large values of k, rather than being an exact equality.

4.3 Classification of nodes

Classifying nodes based on their degree provides a basic framework to understand their nodes in the network. Nodes with higher degrees mostly represent collaboration hubs, acting as centers of localized research activity. Nodes with lower degrees rather represent the bride between less connected nodes with higher degrees and the knowledge transmission between them. Their learnings about the structure of scientific networks represent the dual role of nodes with medium-degree nodes. Also, the nodes with low degree act as connectors that link distant parts of the network. This aligns with the "strength of weak ties"[10] theory, which concludes that weakly connected nodes play a crucial role in the expansion of the research network.

Results and Analysis

The project investigates academic collaboration networks using small-world network theory. The analysis covers several major dimensions of the structural property of the network, its dynamic behavior, and its implications for understanding patterns of impactful research and resource allocation. The goal is not only to summarize the findings but also to analyze their significance in a broader network.

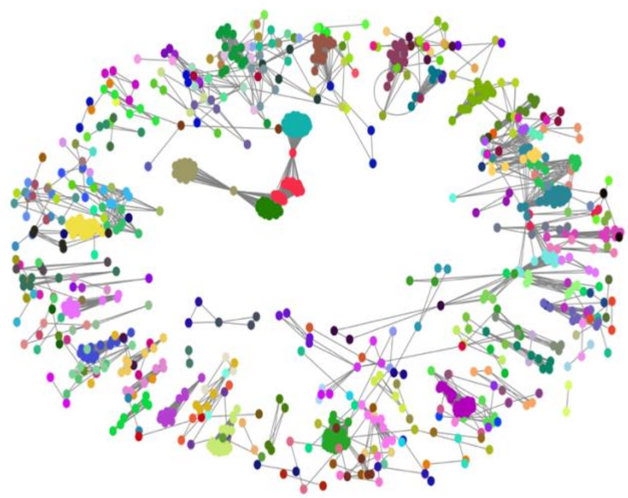
1. **Mathematical Characteristics of Academic Collaboration Network**

Metric	Value	Description
Total Number of Nodes	33981	Total unique authors or institutions in the network.
Total Number of Edges	624928	Total co-authorship links.

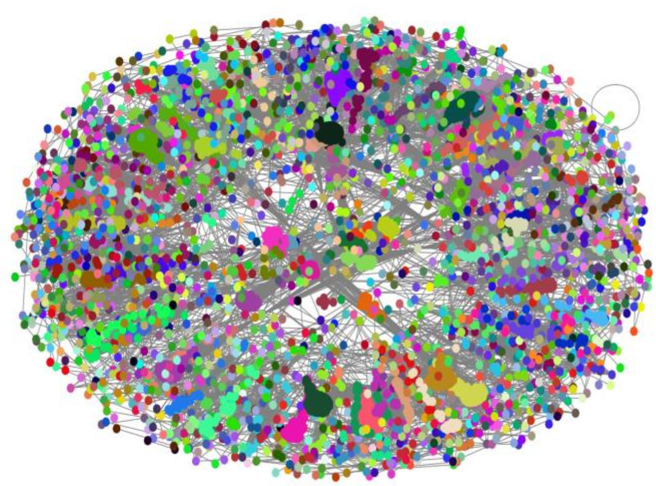
Average Degree	36.78	The average number of connections per node.
Global Clustering Coefficient	0.71	A measure of the overall clustering in the network.
Largest Connected Component Size	20872	The size of the largest subset of nodes that are mutually reachable.

2. **Academic Collaboration Network Analysis**
   I have taken a step-by-step approach to create the visual representation of the graph as follows:

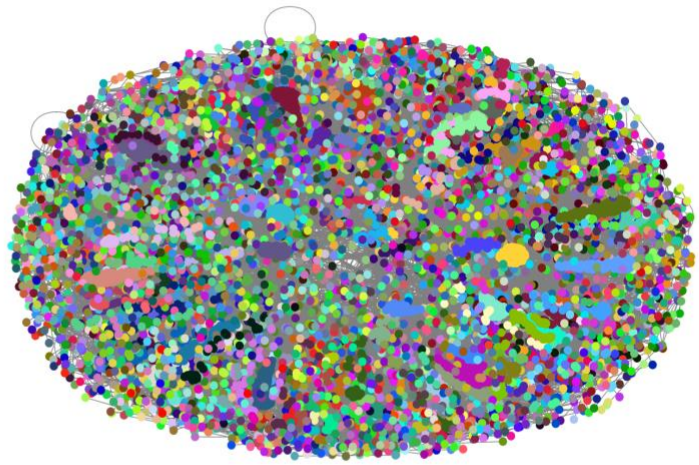
   Step 1: 1000 nodes



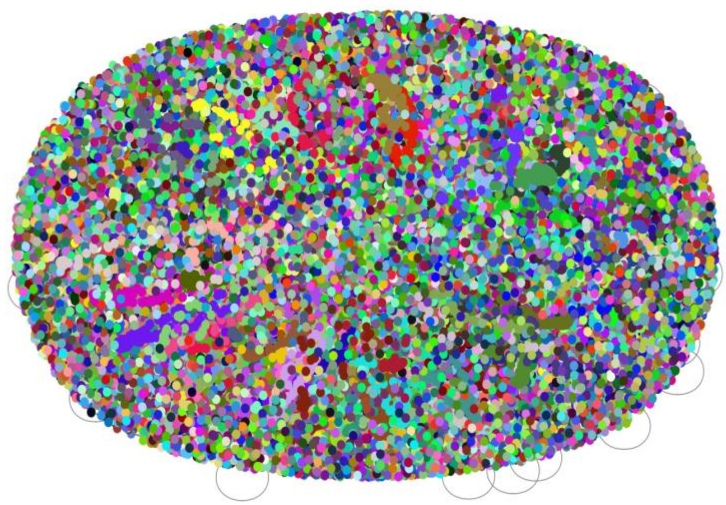
   Step 2: 5000 nodes



Step 3: 10000 nodes



Step 4: All nodes in the network



In the graph visualization, nodes that represent authors are clustered together, that is, they are grouped based on their shared works, with each work assigned a unique color. The step-by-step process of creating the graph emphasizes the increasing clarity of these clusters as the number of nodes grows.

In Step 1, with 1,000 nodes, the first clusters emerge, with authors who co-authored the same work sharing the same color, although the clusters remain rather diffuse.

By Step 2, with 5,000 nodes, the network begins to form more coherent clusters, with writers within the same work being more closely related.

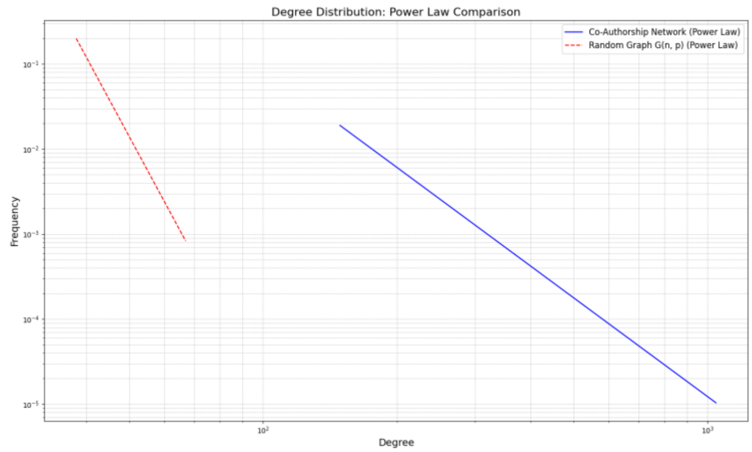
As we progress to Step 3 with 10,000 nodes, the clusters become more defined, and nodes representing authors of the same work are clustered together, resulting in a sharper demarcation between various works.
In Step 4, when all nodes are integrated, the complete network is displayed, with well-defined, colorful clusters, where each group of authors is tightly interconnected by co-authorship links within the same work, making the collaborative structure more prominent and visually distinct.

3. Outcome Analysis

Finding	**Analysis**
Nodes with High Degree	These nodes correspond to key researchers or institutions forming tight-knit collaborative groups.
Nodes with Moderate Degree	Act as bridges between groups, promoting interdisciplinary research and innovative collaboration.
Nodes with Low Degree	Connect isolated groups, facilitating the integration of diverse research areas.
Network Efficiency	The small-world properties (high clustering and short path lengths) enhance knowledge dissemination.
Key Observations	Include standout features, such as prolific researchers with degree.
Challenges in Network Analysis	Missing data or incomplete records may affect degree measures and node classification accuracy.

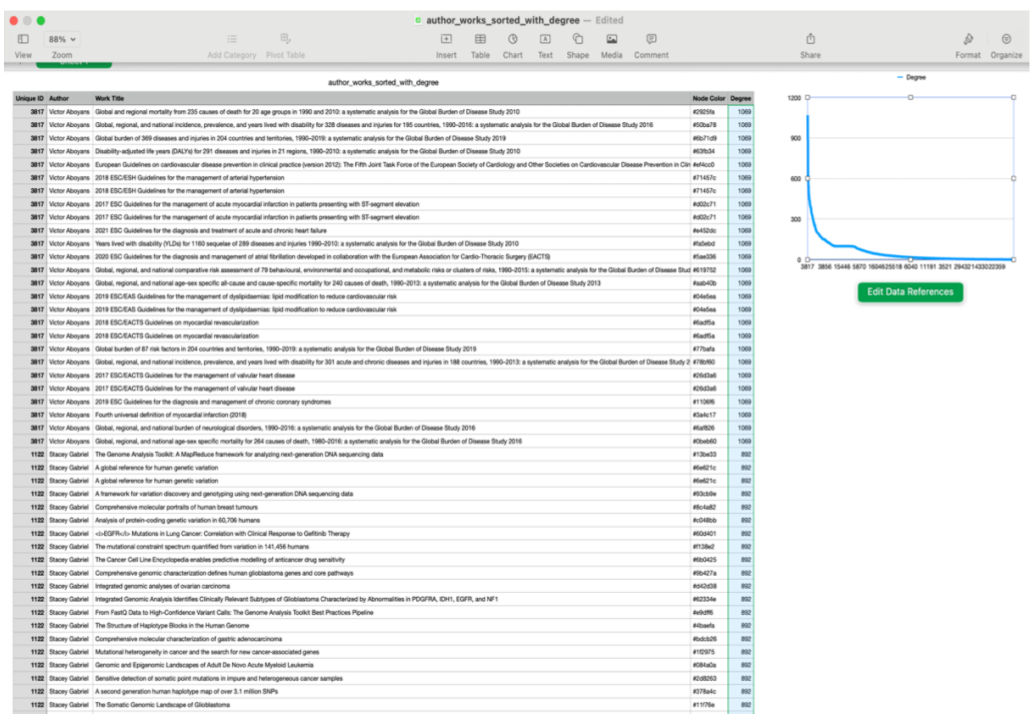
4. Validation of Academic Collaboration Network

The academic collaboration network, when compared to a random Erdős-Rényi graph $G(n, p)$, the academic collaboration network demonstrates a significantly different degree distribution, as validated through a power-law comparison. The power-law fit for the collaboration network reveals a heavy-tailed distribution, indicating the presence of highly connected hubs that dominate the network's structure. In contrast, the random graph lacks such hubs and follows a more uniform degree distribution.



This disparity underscores the small-world nature of the academic collaboration network, as it features a highly clustered structure and a power-law degree distribution—both hallmarks of real-world networks. The power-law comparison further validates that the academic network is driven by non-random, preferential attachment dynamics, where certain researchers play central roles in fostering collaboration and enhancing the efficiency of knowledge transfer and innovation.

5. Arrangement of Data (as per degree of each node) in Descending Order



Data is arranged in descending order depending on each node's degree, which provides crucial insights into the academic collaboration network's structure. By ordering the nodes based on their degree, we may identify writers who are deeply engaged in their local communities, indicating substantial collaborations between their research groups. A high degree indicates writers who frequently collaborate with the same group of colleagues, implying close-knit academic communities. Authors with lower degree, on the other hand, are more likely to have a more diverse and less interrelated group of collaborators. This approach enables the identification of significant authors who are crucial to these dense collaboration networks, as well as those whose work crosses multiple, probably less coherent, groups. By analyzing this sorted data, we acquire a better grasp of the collaborative dynamics within the academic community, emphasizing both the most prominent researchers and those who operate on the outskirts of tightly knit networks.

Discussion

The investigation of academic collaboration networks using small-world network theory has revealed important insights into the dynamics of research. By applying small-world characteristics to the structure of academic collaboration networks, such as short average path lengths and high clustering coefficients, this study discovered important patterns that can be used to guide strategic resource allocation and research environment optimization.

One of the study's main discoveries is the discovery of nodes with a high degree. These nodes, which frequently represent highly connected scholars, serve as important hubs in the network, facilitating the flow of knowledge and information. This observation is consistent with earlier research, which suggests that high-degree nodes are frequently at the center of scientific advances. These important scholars not only contribute to their respective professions but also bridge gaps between disciplines, hence increasing cross-disciplinary innovation.

Furthermore, the comparison of the empirical academic cooperation network to the theoretical $G(n, p)$ model using the power law revealed the presence of small-world traits in the academic collaboration network. The higher clustering coefficient and shorter average path lengths observed in the academic network as compared to the random graph indicate that academic connections are more tightly knit and efficiently structured for knowledge distribution. This study lends support to the concept that academic collaboration networks have small-world characteristics that promote rapid information flow and innovation.

Node categorization using degree distribution has also shown to be an effective approach for identifying significant participants in the network. Researchers with high degree are critical for developing localized collaboration, whereas those with moderate to low degree frequently act as bridges between different regions of the network. This dual purpose of nodes confirms the "strength of weak ties" idea, which states that links between low-degree nodes might improve network connectivity by connecting distant research communities. Identifying and supporting these critical nodes can have a significant impact on the overall productivity and innovation of the academic network.

This study also discusses the practical consequences of the findings for funding agencies and academic institutions. By focusing resources on high-degree nodes or hubs, universities might potentially improve collaboration within research groups that are already making large contributions to innovation. Furthermore, recognizing the importance of less-clustered nodes as bridges enables more focused financing methods, which can assist in connecting isolated research groups, stimulate multidisciplinary research, and accelerate the diffusion of new knowledge.

Conclusion

Finally, this study has shown how small-world network theory can help us understand the structure and dynamics of academic collaboration networks. By applying graph theory to real-world collaboration data, we discovered the structural qualities that encourage knowledge exchange and innovation. Academic collaboration networks exhibit small-world traits, such as strong clustering and low average path lengths, implying that they are effectively constructed for efficient information dissemination and collaborative output.

The identification of central nodes with high degree provides a road map for institutions and funding organizations to efficiently target resources, assisting researchers who play critical roles in fostering innovation. Furthermore, the examination of nodes with moderate to low degrees highlights the significance of weak linkages and their ability to connect distant research communities.

Finally, the findings of this study provide a framework for future research on optimizing academic collaboration networks. Understanding the underlying structural trends that contribute to high-impact research enables institutions and funding organizations to develop more effective and strategic conditions for academic collaboration, stimulating innovation and driving progress in a variety of research domains.

References

[1] Price, D. J. (1965). *Networks of Scientific Papers*. Science, 149(3683), 510-515.

[2] Gilbert, E. N. (1959). *Random Graphs*. Annals of Mathematical Statistics.

[3] Erdős, P., & Rényi, A. (1960). *On the Evolution of Random Graphs*. Publications of the Mathematical Institute of the Hungarian Academy of Sciences.

[4] Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.

[5] Bollobás, B. (2001). *Random Graphs*. Cambridge Studies in Advanced Mathematics.

[6] Watts, D. J., & Strogatz, S. H. (1998). *Collective dynamics of 'small world' networks*. *Nature*, 393(6684), 440–442.

[7] Newman, M. E. J. (2000). Models of the small world. *Journal of Statistical Physics*, 101(3-4), 819-841.

[8] Lambitte, J., & Panzarasa, P. (2009). Measuring the structure of academic collaboration networks. *The European Physical Journal B*, 69(2), 271-278.

[9] Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.

[10] Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.

Appendix

Input:

```
import csv
import networkx as nx
import matplotlib.pyplot as plt
import random  # For generating random colors
import pandas as pd  # Importing pandas for degree count and plotting
from pyalex import Works

# Initialize variables
data = []
unique_id = 0  # Start unique identifier from 0
author_work_mapping = {}  # To map authors to their unique IDs and works
work_to_authors = {}  # To map works to their respective authors
node_colors = {}  # To store node colors
work_colors = {}  # To store colors for each work

# Set up the search and pagination parameters
pager = Works().search_filter(title="")  # Search filter with an empty title for all works

# Loop through pages and collect data
for page in pager.paginate(per_page=100):  # Set the number of results per page
    for work in page:
        work_title = work.get("title", "N/A")
        authors = work.get("authorships", [])

        # Generate a unique color for this work if not already generated
        if work_title not in work_colors:
            work_colors[work_title] = "#%06x" % random.randint(0, 0xFFFFFF)

        # Track authors for this specific work
        current_authors = []

        for author in authors:
            if author.get("author"):
                author_name = author["author"]["display_name"]

                # Assign a unique ID if the author is new
                if author_name not in author_work_mapping:
                    author_work_mapping[author_name] = {
                        "id": unique_id,
                        "works": []
                    }
                    unique_id += 1
```

```
            # Add the author to the current work's authors
            current_authors.append(author_work_mapping[author_name]["id"])

            # Add the work to the author's list of works
            author_work_mapping[author_name]["works"].append(work_title)

            # Assign the latest color to the node
            node_colors[author_work_mapping[author_name]["id"]] = work_colors[work_title]

        # Map the current work to its authors
        work_to_authors[work_title] = current_authors

        # Stop if we reach the desired number of rows
        if len(author_work_mapping) >= 40000:
            break
    if len(author_work_mapping) >= 40000:
        break

# Generate edges for the graph by connecting authors with the same work
edges = []
for work, authors in work_to_authors.items():
    for i in range(len(authors)):
        for j in range(i + 1, len(authors)):
            edges.append((authors[i], authors[j]))

# Create the graph
G = nx.Graph()
G.add_edges_from(edges)

# Analyze graph properties
total_nodes = G.number_of_nodes()
total_edges = G.number_of_edges()
average_degree = sum(dict(G.degree()).values()) / total_nodes
global_clustering_coefficient = nx.transitivity(G)  # Global clustering coefficient

# Calculate the degree for all nodes
node_degrees = dict(G.degree())

# Categorize nodes based on degree ranges
high_degree_nodes = len([node for node, deg in node_degrees.items() if deg > 10])
moderate_degree_nodes = len([node for node, deg in node_degrees.items() if 5 <= deg <= 10])
low_degree_nodes = len([node for node, deg in node_degrees.items() if deg < 5])

# Largest connected component size
largest_cc = max(nx.connected_components(G), key=len)
largest_connected_component_size = len(largest_cc)
```

```
# Output results
print(f"Total Number of Nodes: {total_nodes}")
print(f"Total Number of Edges: {total_edges}")
print(f"Average Degree: {average_degree:.2f}")
print(f"Global Clustering Coefficient: {global_clustering_coefficient:.2f}")
print(f"Number of High Degree Nodes (greater than 10): {high_degree_nodes}")
print(f"Number of Moderate Degree Nodes (between 5 and 10): {moderate_degree_nodes}")
print(f"Number of Low Degree Nodes (less than 5): {low_degree_nodes}")
print(f"Largest Connected Component Size: {largest_connected_component_size}")

# Save author-work data to CSV, sorted by degree
output_file = "author_works_sorted_with_degree.csv"
sorted_authors = sorted(
    author_work_mapping.items(),
    key=lambda x: node_degrees.get(x[1]["id"], 0),  # Sort by the degree of the author's node
    reverse=True  # Descending order
)

# Writing to CSV
with open(output_file, mode="w", newline="") as file:
    writer = csv.writer(file)
    writer.writerow(["Unique ID", "Author", "Work Title", "Node Color", "Degree"])  #
Changed to Degree
    for author, details in sorted_authors:
        for work in details["works"]:
            # Get the color assigned to this work
            node_color = work_colors.get(work, "#FFFFFF")  # Default to white if work color not
found
            # Get the degree for the author
            degree = node_degrees.get(details["id"], 0)  # Default to 0 if not found
            writer.writerow([details["id"], author, work, node_color, degree])

print(f"CSV file created: {output_file}")

# Visualize the collaboration network graph
plt.figure(figsize=(12, 8))
pos = nx.spring_layout(G)  # Position nodes using a force-directed layout
nx.draw(
    G,
    pos,
    node_size=5,
    node_color=[node_colors.get(node, "#FFFFFF") for node in G.nodes()],
    edge_color="lightgray",
    with_labels=False
)
```

```
plt.title("Collaboration Network (Subset)")
plt.show()

# --- Comparison of Degree Distributions --- #
deg_sequence = [d for _, d in G.degree()]

# Use pandas to count degree frequencies and plot the distribution
deg_count = pd.Series(deg_sequence).value_counts().sort_index()  # Count the frequency of each degree

plt.figure(figsize=(10, 6))
deg_count.plot(kind='line', label="Co-Authorship Network")

# Generate degree distribution for a random graph G(n, p) with a similar number of nodes and edges
p = 2 * total_edges / (total_nodes * (total_nodes - 1))  # Estimate probability for G(n, p)
random_graph = nx.erdos_renyi_graph(total_nodes, p)
random_deg_sequence = [d for _, d in random_graph.degree()]

# Use pandas to count degree frequencies for the random graph
random_deg_count = pd.Series(random_deg_sequence).value_counts().sort_index()

random_deg_count.plot(kind='line', label="Random Graph G(n, p)", color='r')

# Add labels and title
plt.xlabel('Degree')
plt.ylabel('Frequency')
plt.title('Degree Distribution Comparison')
plt.legend()
plt.show()
```

Output:

```
(base) rohanbali@Rohans-MBP-2 ~ % /usr/bin/python3 "/Users/rohanbali/Desktop/Small World Networks/FINAL PROJECT/ResultPROJECT.py"
/Users/rohanbali/Library/Python/3.9/lib/python/site-packages/urllib3/__init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'LibreSSL 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
  warnings.warn(
Total Number of Nodes: 33981
Total Number of Edges: 624928
Average Degree: 36.78
Global Clustering Coefficient: 0.71
Number of High Degree Nodes (greater than 10): 19060
Number of Moderate Degree Nodes (between 5 and 10): 5674
Number of Low Degree Nodes (less than 5): 9247
```

```
Largest Connected Component Size: 20872
CSV file created: author_works_sorted_with_degree.csv
```