3. Analyzing citation networks

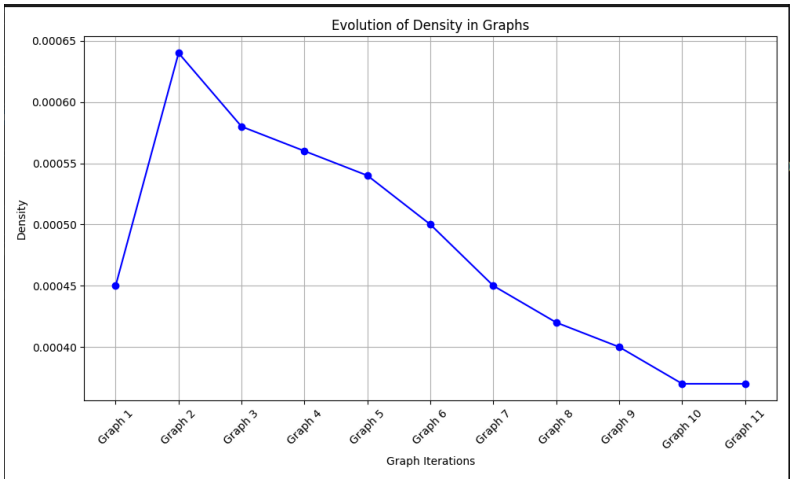
Task 1:

table.txt

Graph Iterations	Number of Nodes	Number of Edges
Graph 1	582	152
Graph 2	2117	2862
Graph 3	4421	11379
Graph 4	7276	29808
Graph 5	10481	58927
Graph 6	14013	98307
Graph 7	17736	142934
Graph 8	21739	201300
Graph 9	25775	263468
Graph 10	29878	334051
Graph 11	30558	347268

The graph evolved over the period of 11 years. These are the number of nodes and the number of edges, that got connected as the graph proceeded

1. **Density**



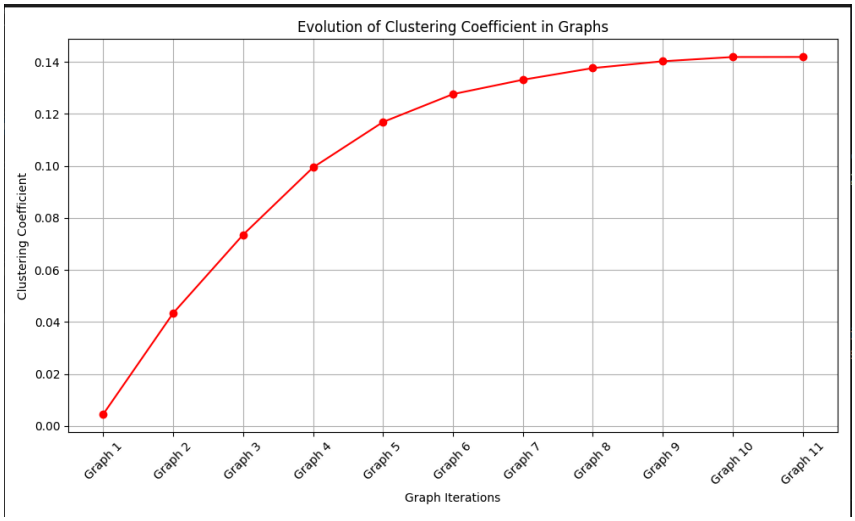
We saw a sudden rise in density from Graph 1 to Graph 2 , that was because just after the initial nodes were added , more and more paper that are based on those papers only got added , therefore an unexpected peak , after that it follows the general decreasing pattern

2. Clustering Coefficient

The clustering coefficient is a measure in graph theory that quantifies the degree to which nodes in a graph tend to cluster together. It provides insight into the local connectivity or cohesion within a network. The clustering coefficient is often used to assess the presence of clusters or communities in a graph.

$$C_i = \frac{2 \times \text{number of triangles centered on node } i}{\text{degree of node } i \times (\text{degree of node } i - 1)}$$

$$C_{\text{global}} = \frac{1}{N} \sum_{i=1}^{N} C_i$$



Graph 1:

- Low clustering coefficient suggests that, initially, research papers may not have formed tightly connected citation groups.

Graph 2:

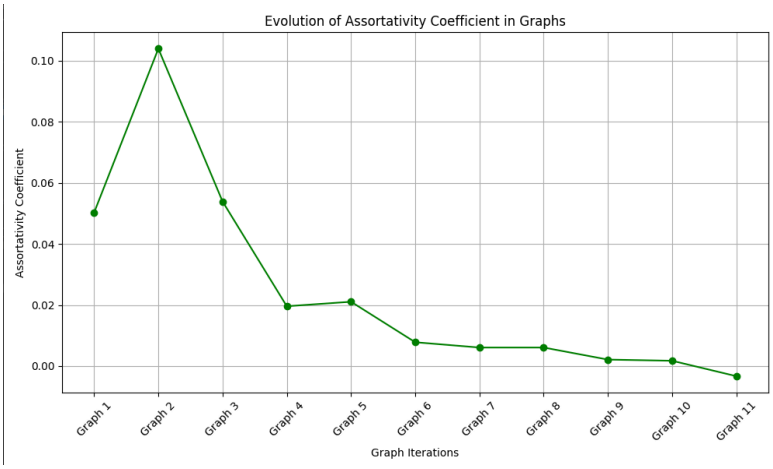
- A substantial increase in clustering coefficient indicates that, as the graph grows, papers may start forming more cohesive citation clusters or communities.

Graph 3 to Graph 11:

- The clustering coefficient continues to increase but at a diminishing rate. This might suggest that, as the network expands, new papers are more likely to connect with existing clusters rather than forming entirely new ones.
- It could also imply that the research topics covered by new research papers are diverse, leading to connections across different clusters, and preventing the formation of highly clustered groups.
- Alternatively, it could be influenced by the type of citation patterns in the field, where papers tend to cite works from multiple thematic areas.

3. Assortativity Coefficient

Assortativity measures the tendency of nodes with similar degrees to connect to each other in a network. In a graph, it can be positive, negative, or close to zero. Positive assortativity indicates that nodes with similar degrees are more likely to connect, while negative assortativity suggests that nodes with different degrees are more likely to connect.



Graph 1:

- A positive assortativity coefficient indicates that papers with similar numbers of references tend to cite each other. This could be due to common topics or themes that lead to interconnected citation patterns.

Graph 2:

- A substantial increase in assortativity suggests that as the graph grows, papers with similar degrees continue to connect. This might indicate the formation of citation clusters or communities with shared themes.

Graph 3 to Graph 11:

- The assortativity coefficient fluctuates but tends to decrease over time. This could suggest that, as the network expands, new papers are more likely to connect with papers of different degrees rather than primarily with papers of similar degrees.
- It might indicate that the research topics covered by new papers are diverse, leading to connections across different citation clusters, preventing a strong tendency for papers of similar degrees to connect.
- Alternatively, it could be influenced by the nature of citation patterns in the field, where papers tend to cite works from various thematic areas, and degree similarity becomes less pronounced.

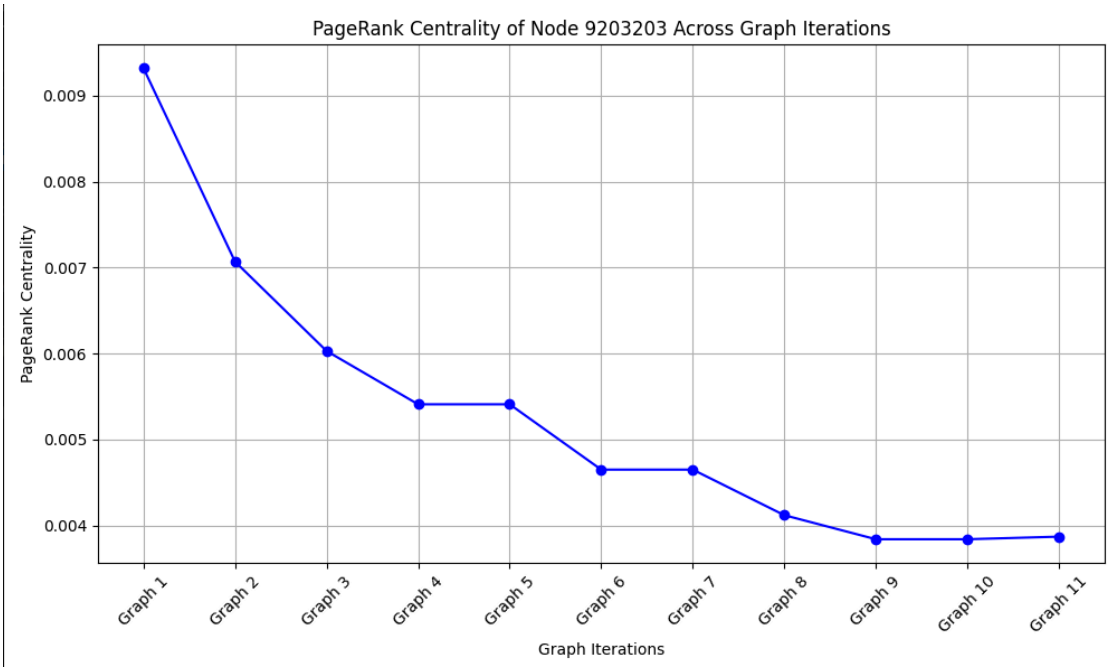
Interesting Insight:

This may suggest that original when research paper covered basic researches , they were tend to be connected similarly , but as time passed more advanced research evolved that was referred from basic papers causing an edge between basic paper(now which has more degree , because a lot of advanced papers are referring this paper now) to an advanced paper.

4. **PageRank Centrality:**

PageRank centrality is a measure used in network analysis to assess the importance or influence of nodes within a graph. Originally developed by Larry Page and Sergey Brin as part of the Google search engine algorithm, PageRank assigns a numerical weight to each node in a directed graph based on the structure of links between nodes. The basic idea is that a node's importance is influenced by the importance of nodes pointing to it.

PageRank Centrality of node 9203203:



Emerging Research Topics:

- Over time, new and innovative research topics emerge, shifting the focus of academic inquiry.
- Node 9203203, which initially held high centrality, experiences a decrease as attention turns to newer subjects.

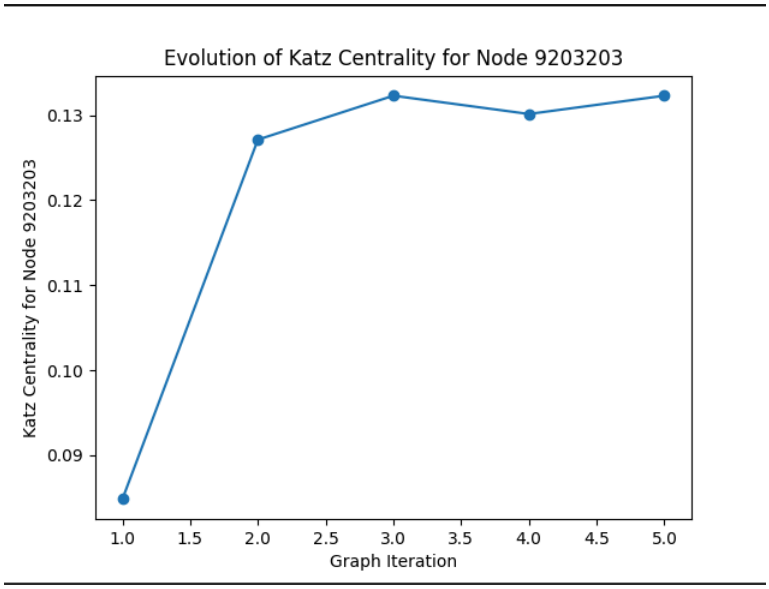
Here is the list of top 5 Page Rank Nodes :

	Graph Iteration	Top Node 1	Top Node 2	Top Node 3	Top Node 4	Top Node 5
0	1	9203206	9203203	9203220	9206242	9204220
1	2	9203203	9206203	9206242	9203206	9204220
2	3	9203203	9206203	9303255	9310316	9206242
3	4	9303255	9206203	9303255	9310316	9206242
4	5	9303255	9206203	9303255	9310316	9206242
5	6	9303255	9206203	9206242	9310316	9204220
6	7	9303255	9206203	9206242	9310316	9204220
7	8	9303255	9206203	9206242	9310316	9204220
8	9	9303255	9206203	9206242	9310316	9204220
9	10	9303255	9206203	9206242	9310316	9204220
10	11	9303255	9206203	9206242	9310316	9204220

5. **Katz Centrality**

   Katz centrality is a measure of the relative influence or importance of a node within a network. It considers both direct connections and indirect paths through other nodes. Nodes with higher Katz centrality scores are not only well-connected but also have connections to nodes that, in turn, have connections to many others.

Katz Centrality of node 9203203:



Insights:

Initially, when the node was introduced, it held a moderate Katz centrality value. However, the sudden surge in centrality indicates a rapid influx of advanced research papers. Initially, when the node was introduced, it held a moderate Katz centrality value. However, the sudden surge in centrality indicates a rapid influx of advanced research papers referencing this particular node. As the field matured and subsequent papers delved into related topics, the centrality values stabilized, reflecting a consistent influence.

However, with the saturation of research on the initial theme, node 9203203 faced challenges in maintaining its position within the top 5. The competitive landscape intensified, and the node gradually faded from prominence in later iterations as new research topics emerged, demonstrating the dynamic nature of academic networks.ncing this particular node. As the field matured and subsequent papers delved into related topics, the centrality values stabilized, reflecting a consistent influence.

However, with the saturation of research on the initial theme, node 9203203 faced challenges in maintaining its position within the top 5. The competitive landscape intensified, and the node gradually faded from prominence in later iterations as new research topics emerged, demonstrating the dynamic nature of academic networks.

A comparison on Node 9203203

Katz Centrality:

- *Adaptive Nature:* Katz centrality considers both direct and indirect connections. The sudden peak indicates a quick adaptation to a surge in citations, while stabilization suggests adaptability to ongoing research trends.
- *Saturation in Research:* The subsequent decrease in Katz centrality can be attributed to saturation in research on the initial theme. As newer topics emerge, the node faces competition and gradually fades.

PageRank:

- *Algorithm Emphasis:* PageRank emphasizes the importance of nodes with high-quality inbound links. Node 9203203, with consistently high PageRank centrality, is consistently cited by influential papers.
- *Resilience:* The node's continued presence in the top 5 PageRank centralities suggests resilience to changing research dynamics. Its enduring influence contributes to its high PageRank centrality.