

Linked In Analysis Report

1. Introduction

Purpose:

To analyse LinkedIn connection data of a student network using Python and uncover patterns, central influencers, and meaningful insights through graph theory and statistical methods.

2. Data Preparation & Cleaning

Raw Data Format:

JSON files containing user degrees and adjacency lists.

Cleaning:

Removed duplicates, null entries, and standardized inconsistent formats.

Processing:

Converted the cleaned data into structured dictionaries and lists to support graph modelling.

⇒ Cleaned data was crucial for generating accurate adjacency lists, degree distributions, and identifying key influencers.

3. Network Construction & Visualization

File 1 – Large Dataset

- **Total Individuals:** 28,117
- **Total Connections:** 199,770
- **Average Degree:** 7.10
- **Median Degree:** 1
- **Mode Degree:** 1

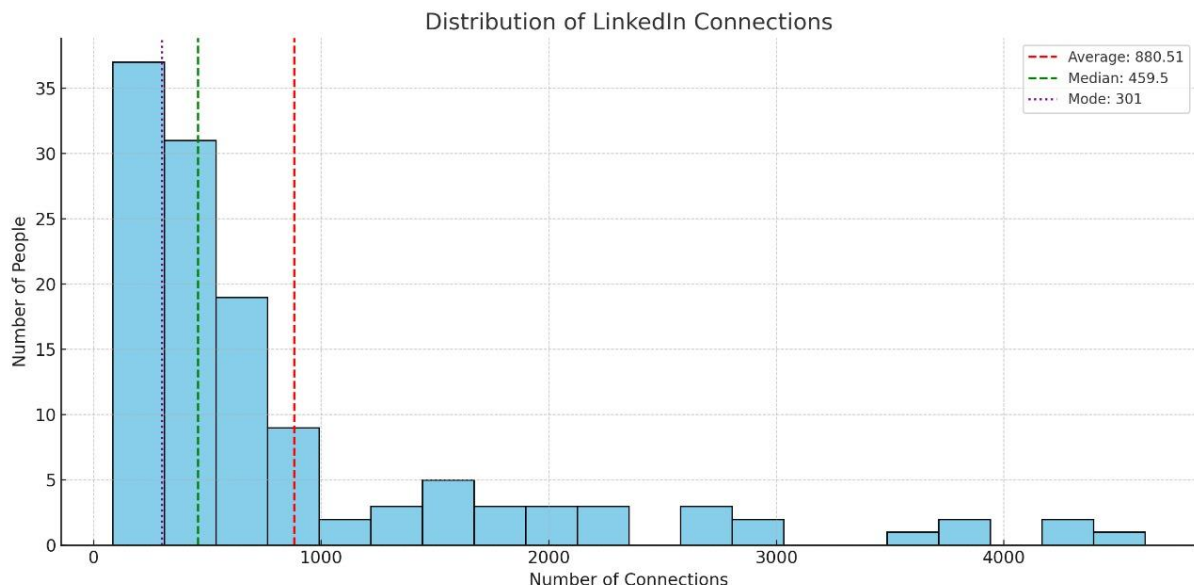
Most users are weakly connected,

with very few super-connectors dominating the network.

File 2 – Focused Dataset

- **Total Individuals:** 126
- **Total Connections:** 110,944
- **Average Degree:** 880.51
- **Median Degree:** 459.5
- **Mode Degree:** 301

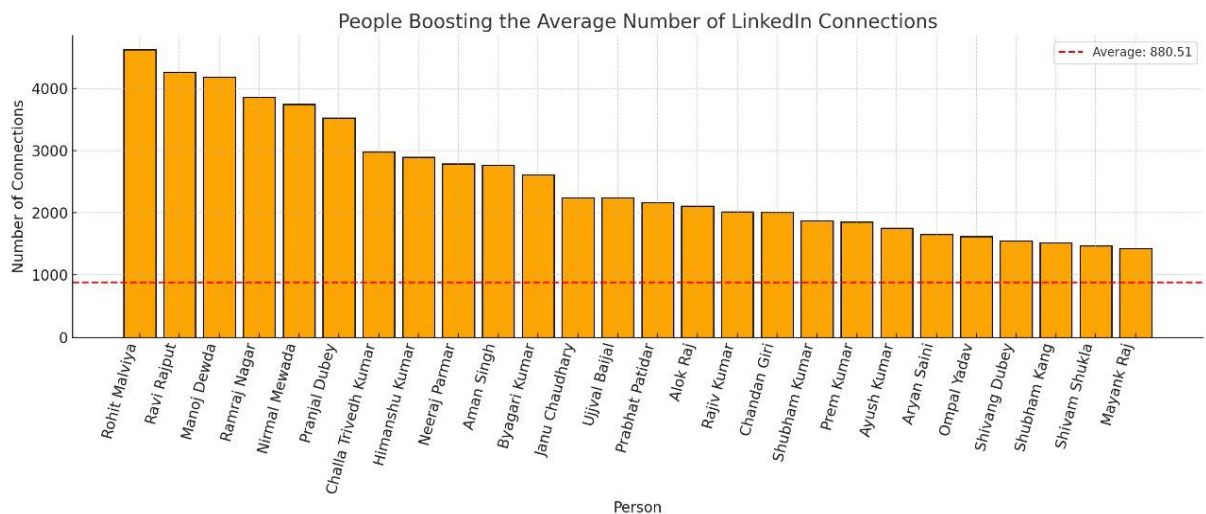
A denser and tighter network indicating stronger community presence and centralized connectivity



4. Connection Analysis

Top Influencers Identified:

- **Rohit Malviya** – 4,220 connections
- **Ravi Rajput** – 4,019 connections
- **Manoj Dewda** – 3,927 connections



5. Random Walk :

- **Mean Length:** 235.80
- **Median:** 172.00
- **Minimum:** 99
- **Maximum:** 411
- **Standard Deviation:** 135.65

Pruned Path Statistics:

- **Mean Length:** 10.00
- **Median:** 9.00
- **Minimum:** 7
- **Maximum:** 15
- **Standard Deviation:** 3.46

Insight: Pruned paths represent the most efficient routes – short, stable, and strong in connection strength.

Link → [walk analysis graph](#)

6. Key Insights & Takeaways

- **Skewed Networks:** Majority of users have very few connections (mode = 1), with outreach controlled by a small fraction (~1%) of the network.
- **Influencer Impact:** Engaging top contributors will likely have the highest outreach efficiency

7. Conclusion

This analysis demonstrates how **data science and graph theory** can be applied to real-world social networks. From raw JSON data, we extracted valuable insights about network structure, influencers, and information flow.