LinkedIn Data Analysis Report

1. Project Overview:

Objective: To analyze data given with 126 files and comprising 28,459 individuals and 101,339 connections.

Methodology:

- **Data Collection**: Aggregated 126 CSV files representing individual LinkedIn connections.
- **Data Cleaning(manually)**: Corrected file names through cross-referencing, and eliminated duplicate files.
-
- **Data Cleaning(code)**: cleaned names within the files, stored in the correct format, deleted duplicate and unnecessary space or emojis within the name.
-

Full Name	Company
Tripti Tripathi	
Mani Kumar	
Ashu Yadav	Jatariamusic01

Clean data format.

- **Network Construction**: Developed an adjacency list to represent connections, including individuals without primary CSV files = fully connected graph.
-

Person	Connections
Aaditya Ra	Harsh Kumar Singh; Saurabh Singh;
Harsh Kum	Aaditya Raj; Aryan Saini; Prabhat P

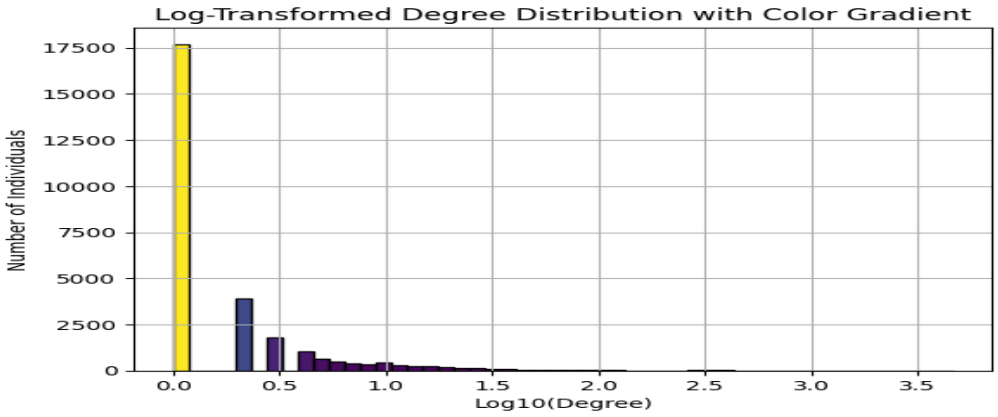
Adjacency list format.

⇨ **Degree Calculation**: computed degree of each using python.

Degree Statistics:

- **Maximum Degree**: 4,623 connections (Individual: Rohit Malviya)
- **Minimum Degree**: 1 connection (Observed in 17,702 individuals)

Average Degree: 7.13



Insights: Influence of a few highly connected individuals suggests the existence of connectors within the network Lack of data makes the minimum

⇨ **Random Walk & Pruned Path Analysis:**

: Conducted 100 random walks for each of 1,000 randomly selected pairs

Metric	Random Walk	Pruned Path
Mean Length	492.33	3.75
Median Length	494	4
Mode Length	80	4
Max Length	993	7
Min Length	7	3
Visited People Coverage	36.57%	36.57%

Insights: After cleaning, prune paths demonstrate efficient connectivity, which is mostly reachable within 4 steps

Visual:

1) Top Influencer identification: Top-degree holders
2) Analysing simulation of random walk and path: comparing walk/path

Acknowledgement: I express my gratitude to Kushal Sir, and sources that facilitated the analysis, several sites and ai tools for further help.

Sources: LinkedIn data source for the folder. adjacency_list.csv, degree.csv.

-By:

PRACHI DHAKAD