Network-Based Systemic Risk Identification Using Aggregated Interbank Exposure Data

📖 Project Overview

This project develops a two-dimensional systemic risk framework to identify systemically important banks using:

1. Financial fragility (balance sheet risk)
2. Interbank exposure intensity (network embeddedness)

Unlike traditional ratio-based analysis, this framework integrates financial vulnerability and interconnectedness to estimate systemic importance in the banking system.

The dataset contains aggregated per-bank interbank exposures and balance sheet information.

🎯 Objective

To answer:

Which banks pose the greatest systemic risk when considering both financial fragility and interbank interconnectedness?

📊 Data Description

The dataset includes:

- Bank_Name
- Total_Assets
- Total_Liabilities
- Total_Equities
- Interbank_Lending
- Interbank_Borrowing

- Deposit_To_Interbank
- Deposit_From_Interbank
- Year

Data required extensive cleaning due to Indian-style numeric formatting.

⚙️ Methodology

1️⃣ Data Cleaning

- Removed Indian-style comma formatting (e.g., 4,75,32,50,00,000.0000)
- Converted financial columns to numeric
- Handled missing and infinite values
- Standardized financial metrics

2️⃣ Financial Fragility Index

Constructed systemic vulnerability measures:

- Leverage Ratio
- Equity Ratio
- Interbank Funding Ratio
- Log(Size)

These were standardized and combined into:

systemic_risk_index

This measures structural financial fragility.

3️⃣ Exposure-Based Network Proxy

Because bilateral bank-to-bank exposures were unavailable, a proxy measure of interconnectedness was constructed:

interbank_strength =

(Interbank_Lending + Interbank_Borrowing) / Total_Assets

This captures how embedded a bank is within the interbank system.

4️⃣ Final Systemic Importance Score

Since fragility and interconnectedness were weakly correlated ($\rho \approx 0.11$), they capture distinct risk channels.

Final systemic score:

final_systemic_score =

Standardized(systemic_risk_index)

+

Standardized(interbank_strength)

This represents:

Structural Vulnerability + Network Embeddedness

📈 Key Visualizations

- Distribution of systemic risk
- Risk map (Fragility vs Interconnectedness)
- Top 10 systemic banks
- Year-wise risk evolution
- Correlation analysis

The systemic risk map highlights banks in the upper-right quadrant as the most systemically important.

🚧 Challenges Faced

1️⃣ Data Formatting Issues

Financial values were stored in Indian comma format (e.g., 4,75,32,50,00,000), causing numeric conversion to fail and resulting in NaNs.

Solution:

- Removed commas explicitly before conversion.

2️⃣ Collapsed Ratios Due to Improper Parsing

Initial ratio calculations returned zero counts and NaNs because financial columns were not properly converted.

Lesson:

- Always inspect raw data types before modelling.

3️⃣ Network Modelling Limitations

Only aggregated exposure data was available (no bilateral exposures).

This prevented:

- True contagion modelling
- Default cascade simulation
- DebtRank implementation

Solution:

- Constructed exposure-based network proxy instead of artificial centrality modelling.

4️⃣ Centrality Metric Pitfall

Initial graph centrality results were nearly identical due to star-network structure.

Insight:

- Topology alone adds little information without bilateral edges.
- Exposure intensity is more meaningful than pure centrality in aggregated datasets.

📌 Key Findings

1. Financial fragility and interbank interconnectedness are weakly correlated.
2. Most banks exhibit low interbank exposure (< 4% of assets).
3. Systemic risk is highly skewed — a small subset of banks dominate.
4. High size does not automatically imply high interconnectedness.
5. Systemic importance arises from the combination of fragility and exposure.

🏁 Conclusion

This project demonstrates that systemic importance can be estimated even with aggregated exposure data by combining:

- Balance sheet vulnerability
- Interbank exposure intensity

The resulting two-dimensional framework provides a practical and interpretable method for identifying systemically important banks.

While true contagion modelling requires bilateral exposure data, the exposure-based proxy approach offers a defensible and scalable alternative for macroprudential risk monitoring.

🔮 Future Improvements

- Incorporate bilateral bank-to-bank exposure data (if available)
- Implement shock propagation simulations
- Apply PCA-based weighting
- Extend to dynamic multi-year network evolution
- Compare results with regulatory SIFI classification

🛠️ Tech Stack

- Python
- Pandas
- NumPy
- Scikit-learn
- NetworkX
- Matplotlib
- Seaborn

👩‍💻 Author

Anjali Joshi
Data Science | Financial Risk Modeling | Network Analysis

⭐ Final Project Status

Completed — Systemic Risk Identification Framework Implemented and Validated.