Educational Resource Distribution Analysis

A Data-Driven Approach Using PCA and Clustering

Yunxi Kong

December 13, 2024

Research Background and Significance

- **Subject**: Addressing disparities in educational resource allocation.
- **Significance**: Crucial for socio-economic development and equity.
- **Problem Description**: Limited understanding of regional disparities.

Context from Literature:

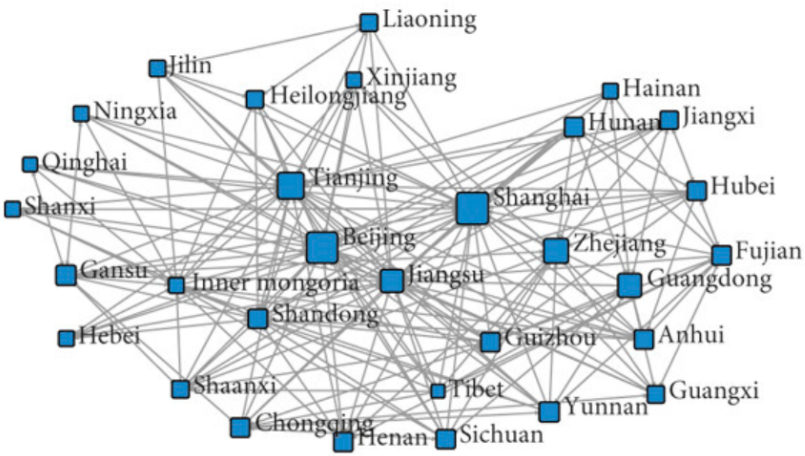
- Education equality improves human capital and promotes fairness (OECD 2012; UNESCO 2015).
- Inequalities in education exacerbate economic disparities and limit long-term prosperity.
- Policies like the "Compulsory Education Law" (2006) aim to reduce disparities but face challenges.

Challenges:

- Gini index alone fails to identify local disparities.
- Multidimensional approaches are needed to evaluate inequalities more effectively.

Knowledge Gaps and Objectives

- **SOTA**:
  - Focused on funding or teacher qualifications.
  - Lack of integrative spatial analysis.
- **Gaps**:
  - Insufficient spatial understanding.
  - Limited use of network analysis in education.
- **Objective**: Develop a data-driven framework to analyze disparities.



Source: Yu (2020)

Alex (2020)

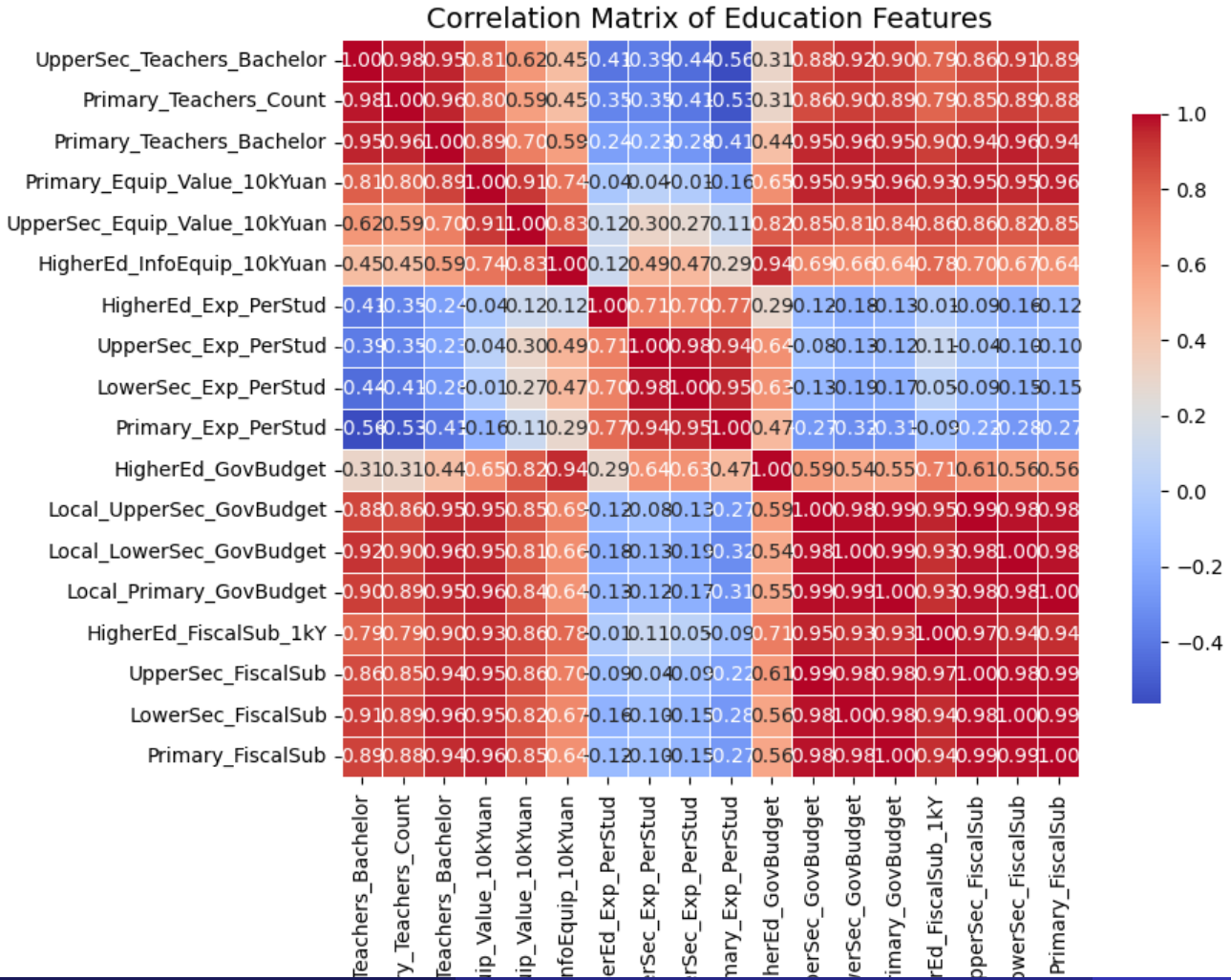
Data Description and Preprocessing

- **Data Source**: 18 indicators (e.g., funding, teacher qualifications).
- **Preprocessing**:
  - Standardization of features to ensure comparability.
  - Key Features: "Primary Teachers with Bachelor's Degrees", "Fiscal Subsidies per Student". "Value of Equipment in Primary Schools (10k yuan)" "Per Student Expenditure in High Schools" "Financial Aid for Higher Education"

地区	经度_x	纬度_x	中等教育-[illegible]	初等教育-[illegible]	初等教育-[illegible]
上海	121.4737	31.230416	-1.160852	-1.050831	-0.860361
云南	102.71	25.045806	0.39681994	0.35791452	0.04866092
内蒙古	111.76562	40.817498	-0.672864	-0.683226	-0.574313
北京	116.40753	39.90403	-1.160994	-0.981164	-0.743553
吉林	125.32599	43.896536	-0.737775	-0.728953	-0.594016
四川	104.07593	30.651651	1.11765156	0.73066344	0.51376502
天津	117.20098	39.084158	-1.16595	-1.080113	-0.976261
宁夏	106.25875	38.471317	-1.258262	-1.17736	-1.124092
安徽	117.28492	31.861184	0.56617606	0.31893021	0.27695045
山东	117.02036	36.66853	2.09730241	1.65895860	1.95291321
山西	112.5624	37.873531	0.04796797	-0.250418	-0.253307
广东	113.26653	23.132191	2.10692944	2.17222041	2.76439634
广西	108.32755	22.815478	0.30467950	0.71000001	0.11032371
新疆	87.627704	43.793026	-0.428627	-0.304763	-0.468098
江苏	118.76323	32.061707	0.83265704	0.93956111	1.74359063
江西	115.90923	28.675696	0.03862575	0.23130651	-0.025744
河北	114.46866	38.037057	1.36587535	1.53688531	1.13126056
河南	113.7536	34.765515	2.10980616	2.71818282	2.15551664

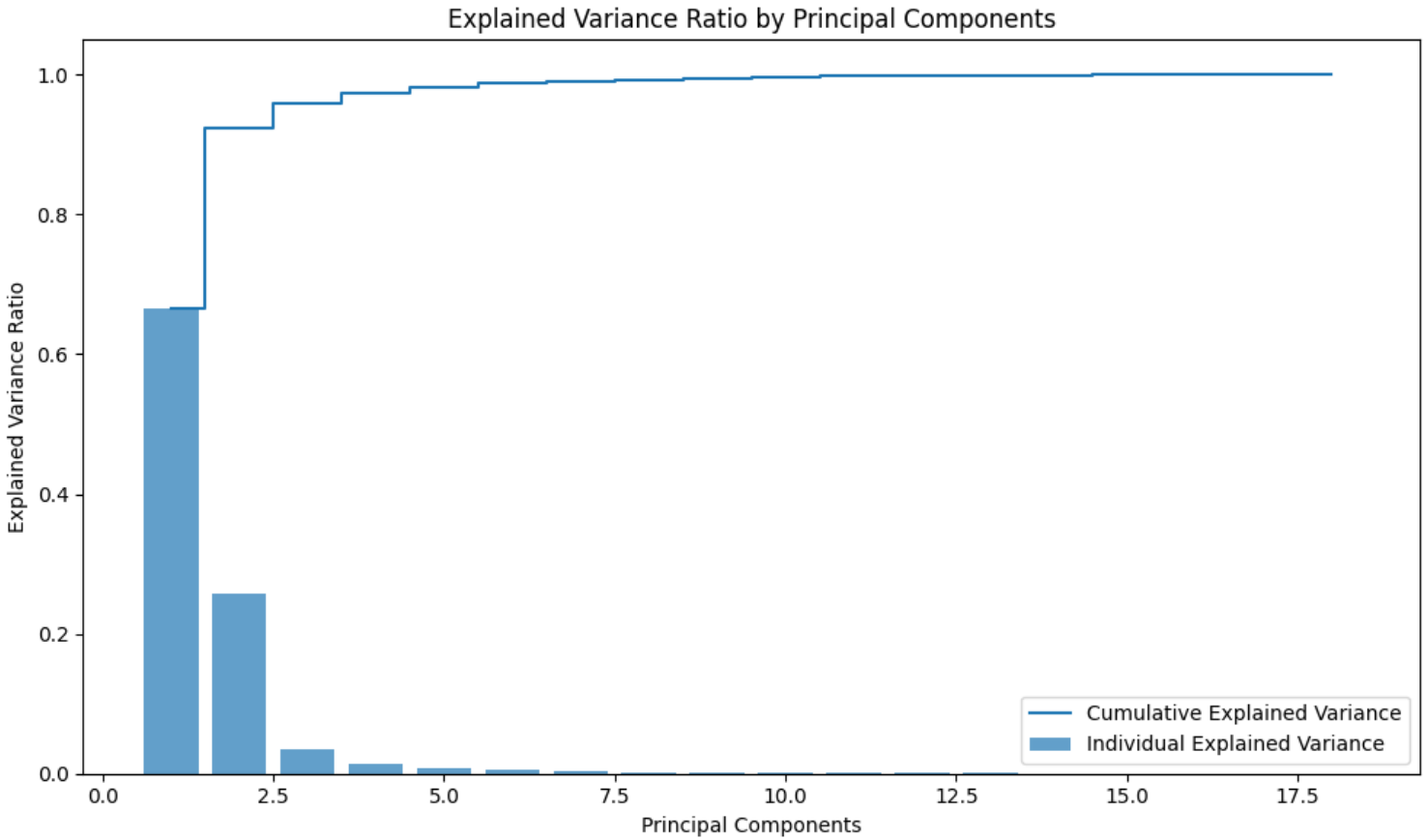
Analysis Framework

1. Dimensionality Reduction: PCA.
2. Clustering: K-means, Louvain, and Fiedler Vector.
3. Network Analysis: k-NN graph and community detection.



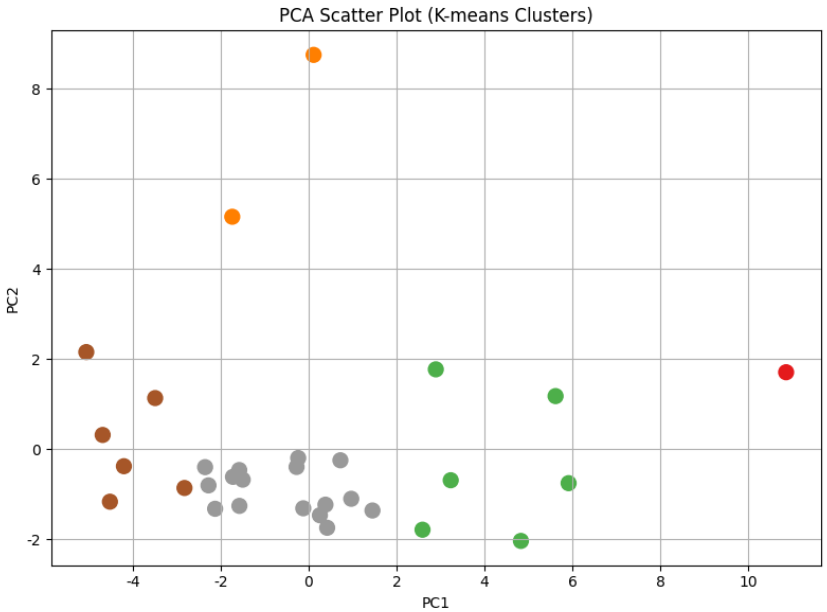
Principal Component Analysis (PCA)

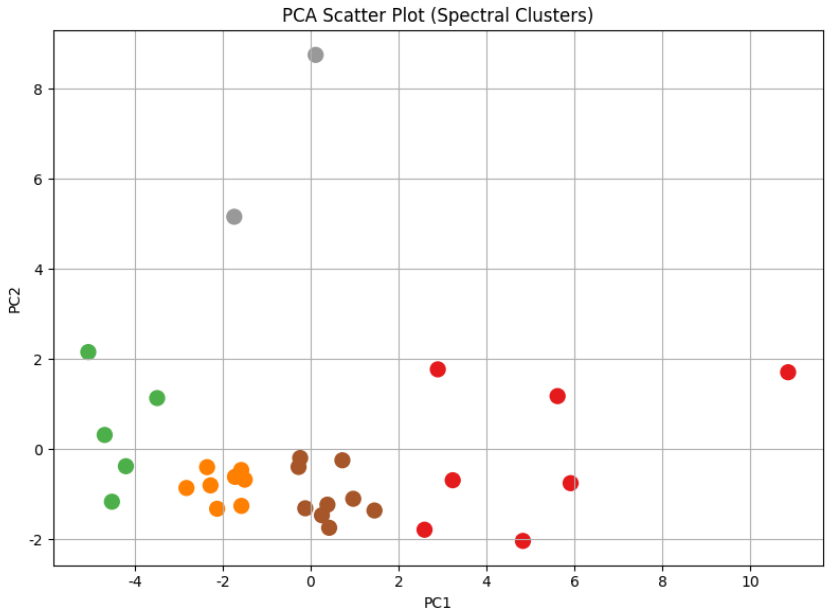
- **PC1**: Captures resource intensity (e.g., funding).
- **PC2**: Reflects resource distribution equity.
- Explained Variance: PC1 (66.7%), PC2 (23.5%).



Clustering Results (K-means and Spectral Clustering)

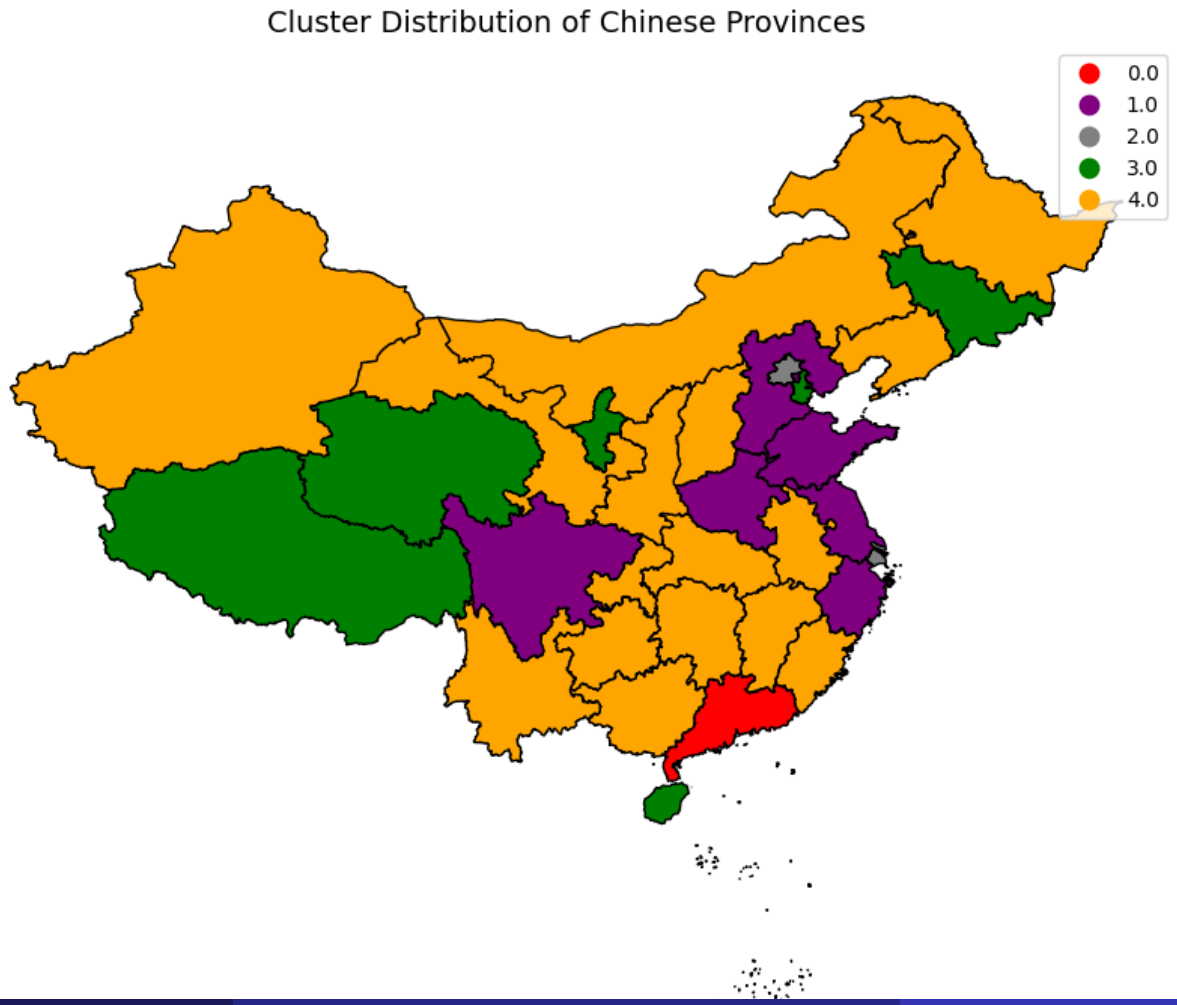
- Grouped provinces using:
  - **K-means Clustering**: Highlights significant resource disparities across five clusters.
  - **Spectral Clustering**: Validates regional characteristics via network structures.





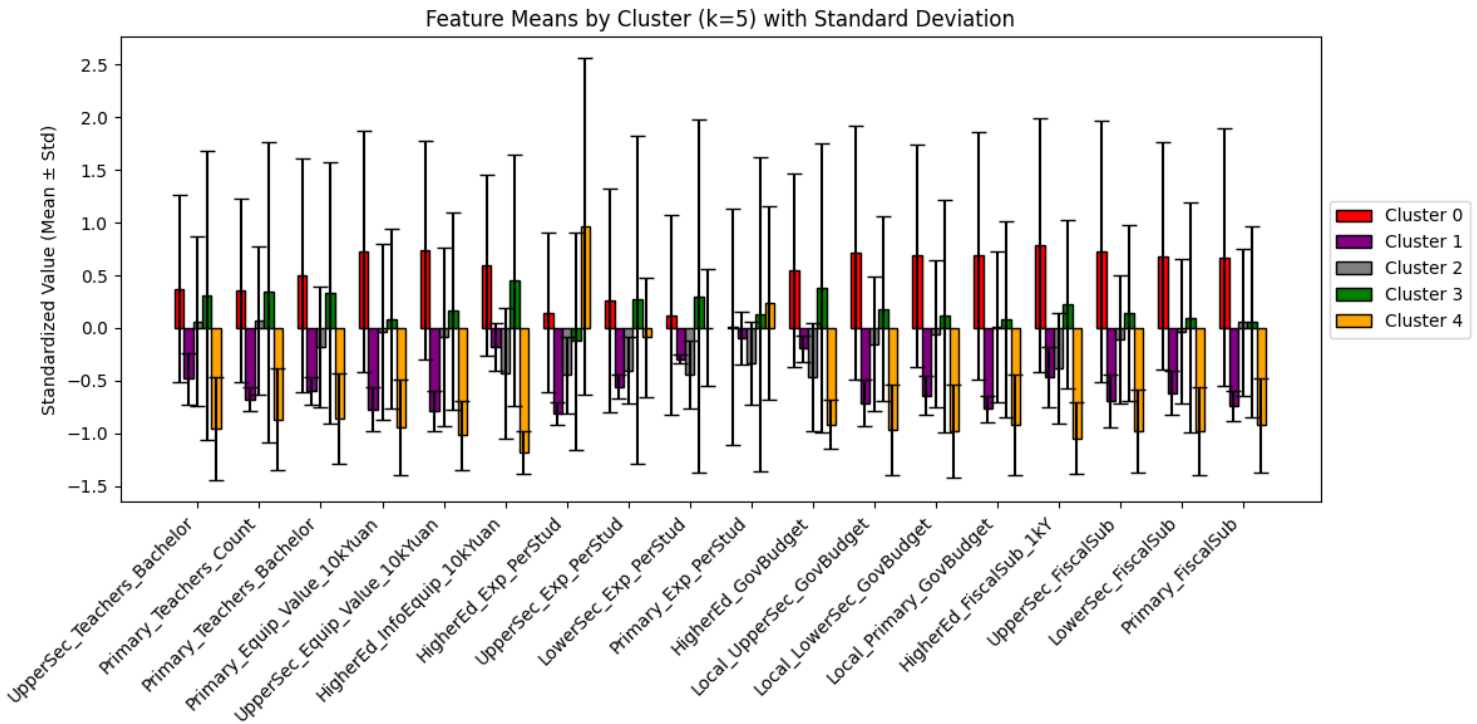
Spatial Distribution of Clusters

- Visualization of cluster results on the map:
  - Coastal provinces (e.g., Guangdong) exhibit higher resource density.
  - Central and western regions show lower resources and uneven distribution.



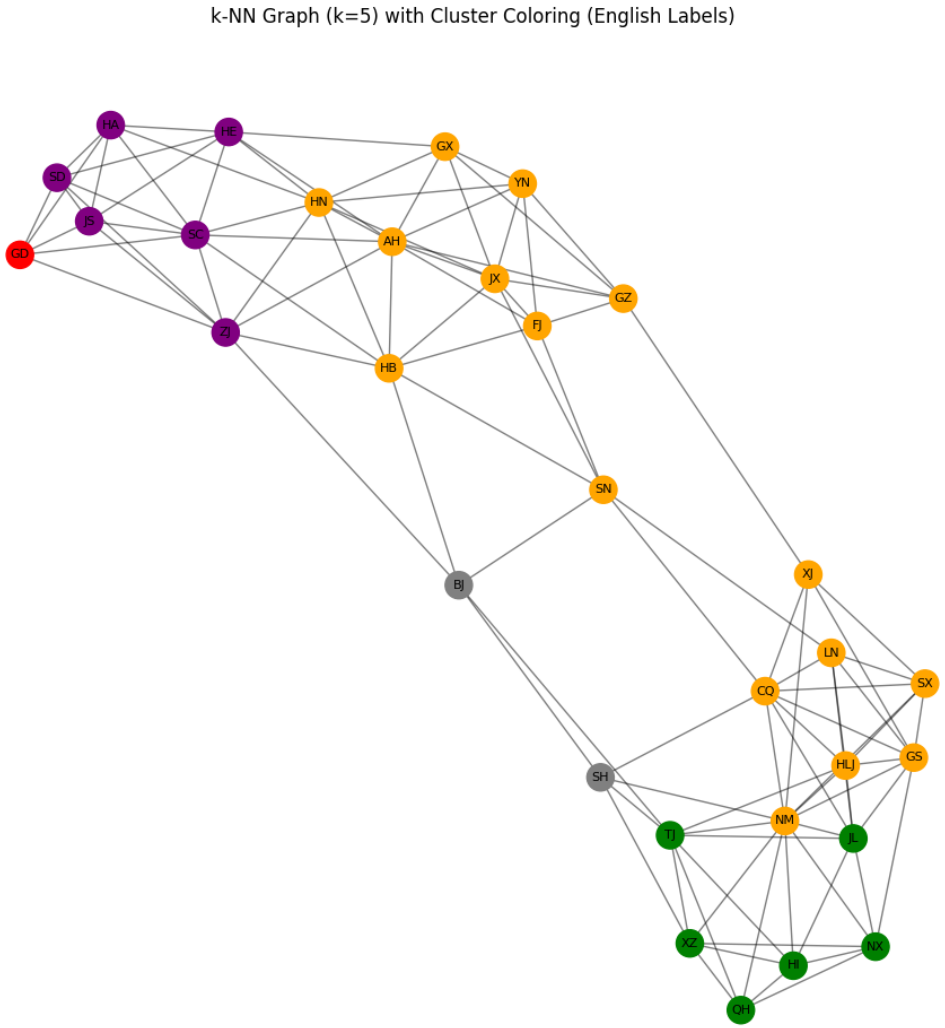
Feature Means by Cluster (with Standard Deviation)

- Analyzed feature distribution across five clusters.
  - Cluster 0 (Red): High fiscal investment but uneven distribution.
  - Cluster 2 (Gray): Low investment, requiring intervention.
- Standard deviation highlights intra-cluster variability.



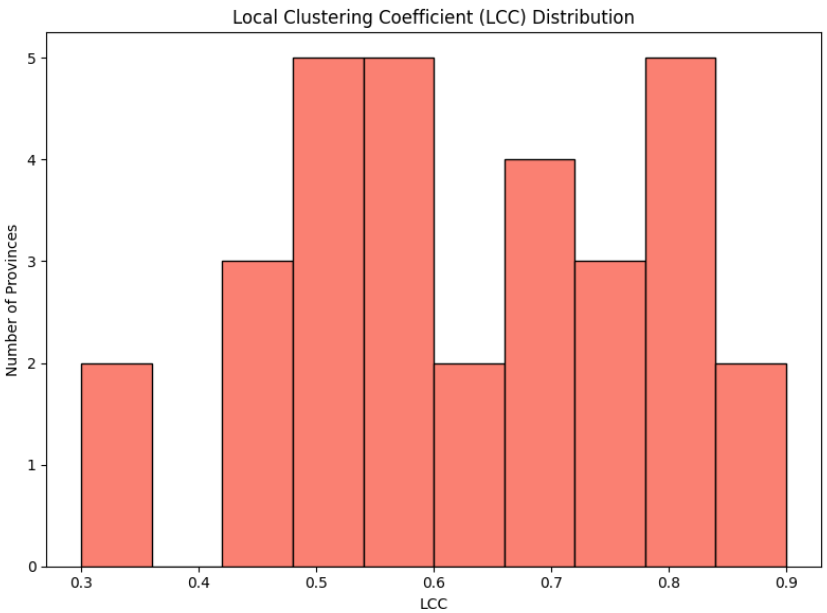
k-NN Graph and Clustering Results

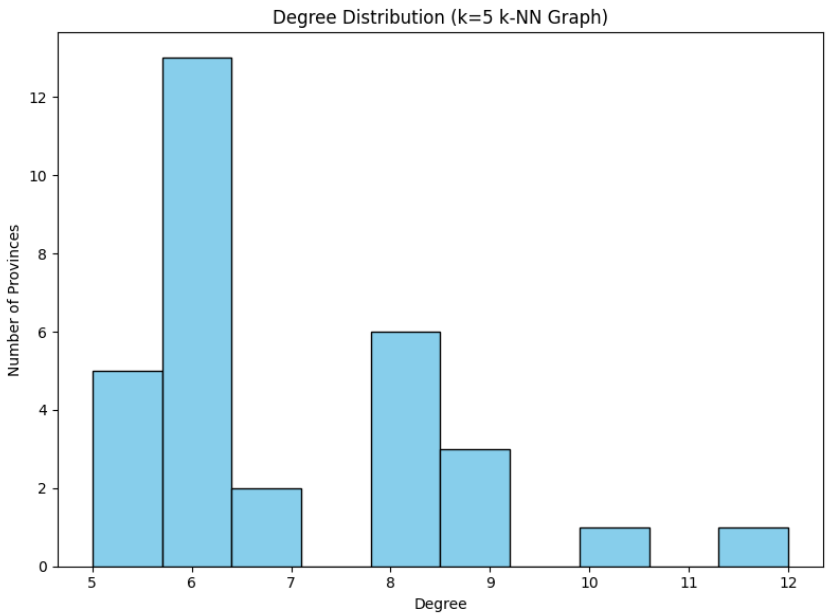
- Constructed k-NN graph to study inter-province connections:
  - High-connectivity nodes align with high-resource regions.
  - Reveals community structures within clusters.



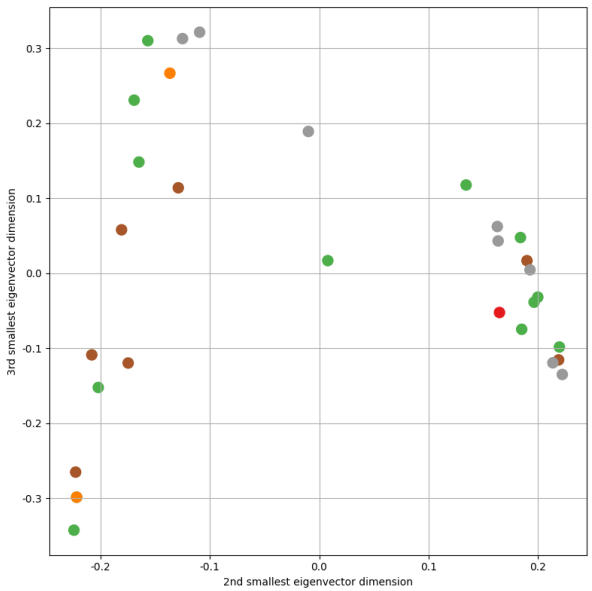
Local Clustering Coefficient (LCC) Distribution

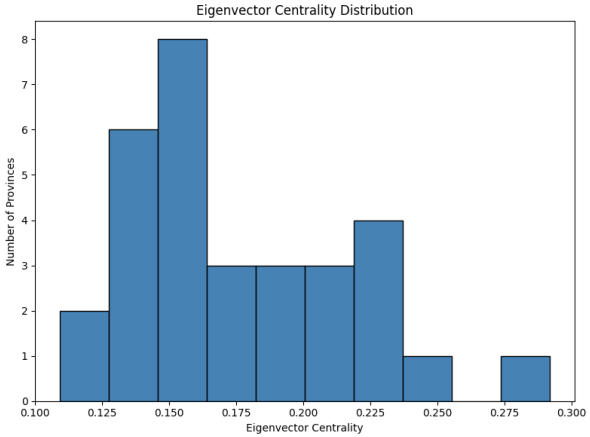
- LCC analysis reveals:
  - High LCC values indicate tight-knit communities.
  - Regions with low LCC may require enhanced interconnectivity.
- Provides insights into the structural cohesion of the network.

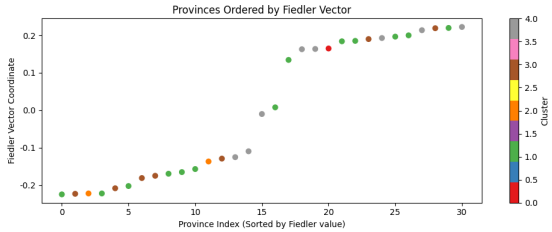


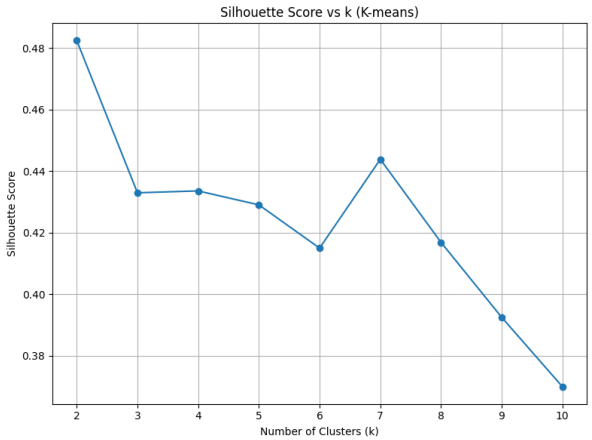


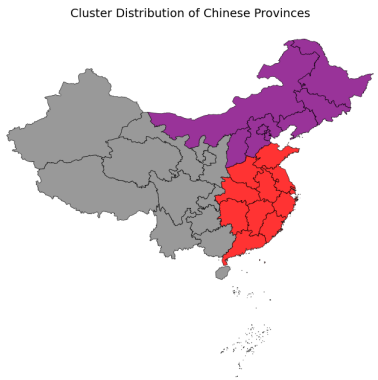
Overview of Results with Six Images

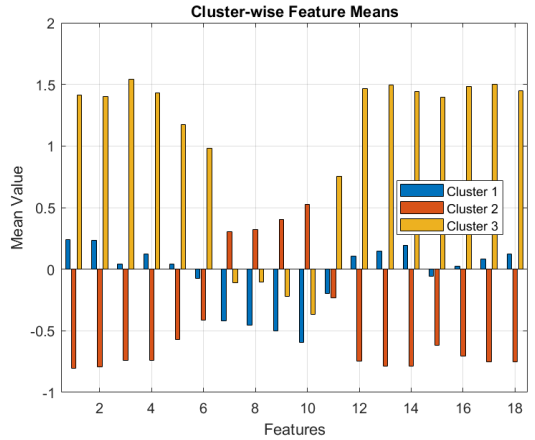




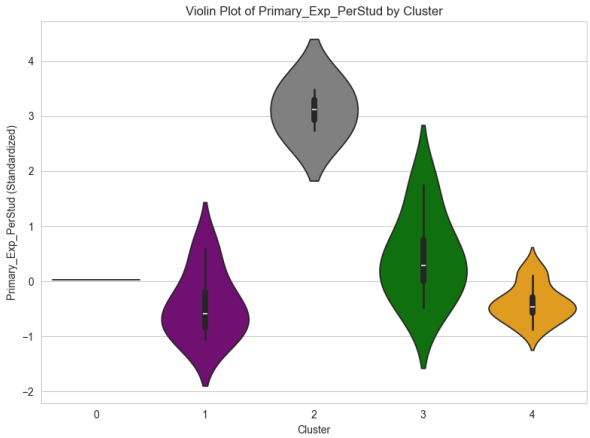


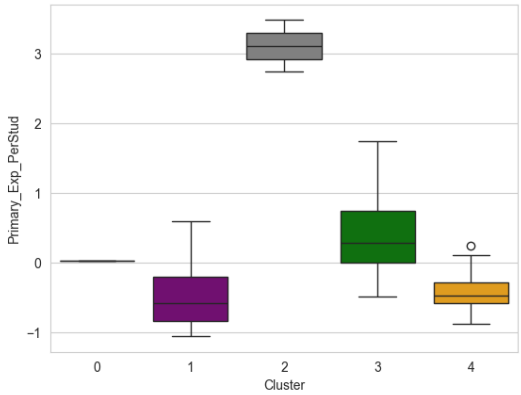


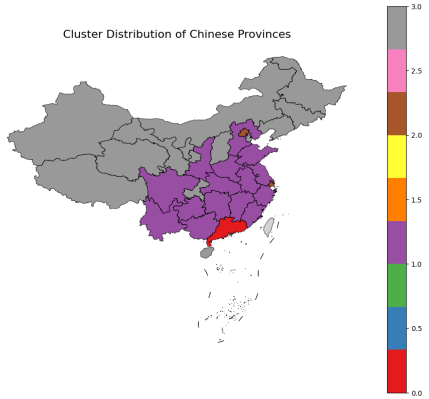


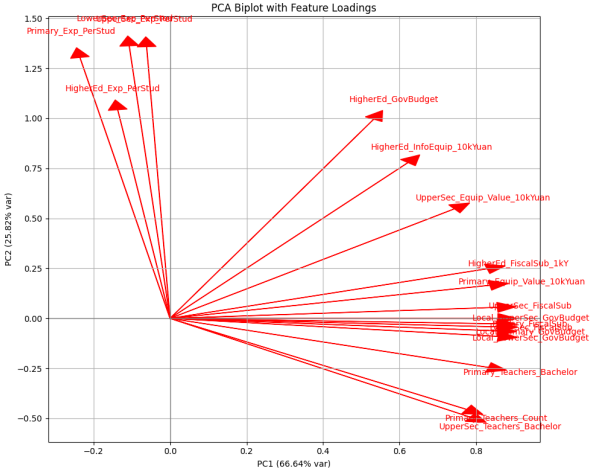


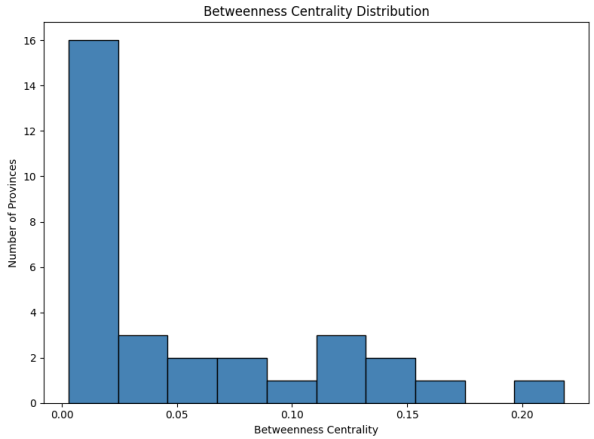
Overview of Results with Six Images

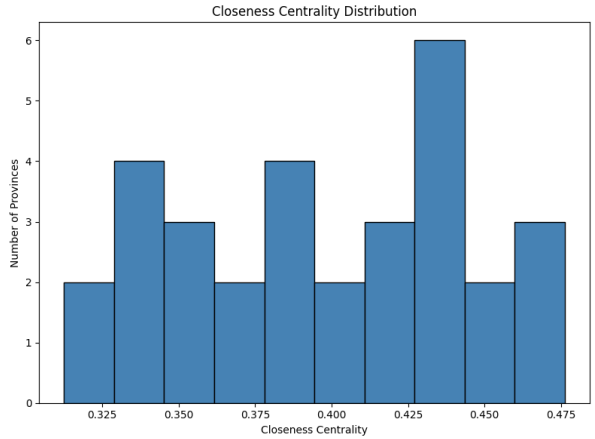












Policy Implications

- **Insights**:
  - Regions with low resources require targeted interventions.
  - High-resource regions can serve as benchmarks.
- **Recommendations**:
  - Increase fiscal investment in underperforming regions.
  - Promote balanced teacher qualifications.

References

- Chen, M., Liu, W., Lu, D., Chen, H., & Ye, C. (2018). Progress of China's new-type urbanization construction since 2014: A preliminary assessment. *Cities, 78*, 180–193.
- Chen, M., Gong, Y., Lu, D., & Ye, C. (2019). Build a people-oriented urbanization: China's new-type urbanization dream and Anhui model. *Land Use Policy, 80*, 1–9.
- Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews, 32*(1), 7–34.
- Deutsch, J., & Silber, J. (2005). Measuring multidimensional poverty: An empirical comparison of various approaches. *Review of Income and Wealth, 51*(1), 145–174.
- Fan, C. C., & Sun, M. (2008). Regional inequality in China, 1978–2006. *Eurasian Geography and Economics, 49*(1), 1–18.
- Guo, G. (2006). Decentralized education spending and regional disparities: Evidence from Chinese counties 1997–2001. *Journal of Chinese Political Science, 11*(2), 45–60.
- Gyatso, S., Postiglione, G. A., & Jiao, B. (2005). Education in rural Tibet: Development, problems and adaptations. *China: An International Journal, 3*(1), 1–23.

Conclusion and Future Work

- **Summary**:
  - Clustering and network analysis reveal significant disparities.
  - Combined PCA and Louvain provide actionable insights.
- **Future Work**:
  - Integrate socio-economic data for deeper analysis.
  - Explore dynamic changes in resource allocation.

Thank you! Questions?