Mapping Provider Networks:

An Analysis of Healthcare Provider-Organization Relationships

Jasmine Jia

Georgetown Univeristy

PPOL 5205: Advanced Modeling Techniques

Professor Brodnax

December 18th, 2024

Summary

The project revealed that the U.S. healthcare network is highly fragmented, with distinct clusters of providers forming based on shared affiliations, which is not limited by geographical location and providers' specialites. These patterns highlight opportunities to improve network cohesion and facilitate better coordination of care.

The website for this project can be found here: https://jasminejia434.github.io/ds3_provider_network_analysis/

Research Objective

This project aims to explore the structural relationships within the U.S. healthcare network. By mapping and analyzing connections between healthcare providers, identified by National Provider Identifiers (NPIs), and organizations, represented by Tax Identification Numbers (TINs), we seek to uncover the underlying patterns and dynamics of these affiliations.=

The research objective is twofold, encompassing both exploratory/descriptive and inferential approaches. In the initial phase, we will construct and examine the network structure to identify clusters of affiliations and central hubs within the healthcare system. This foundational analysis will illuminate core areas of connectivity and provide insights into the network's broader organizational framework. In the subsequent phase, the analysis will expand by integrating taxonomy data to classify providers by specialty. By joining the NPI-TIN dataset with this enriched information, the study will investigate whether the distribution of TINs per NPI varies across provider types. This step will enable a deeper understanding of how provider affiliations and specialties influence network structure, offering valuable insights into the complexities of the U.S. healthcare ecosystem.

Data

This project uses a proprietary dataset provided by Mathematica through its Healthcare Price Transparency Project. Access to this dataset was granted for the purpose of this research, and the data remains private and confidential, reflecting Mathematica's commitment to maintaining the integrity and security of sensitive information.

The dataset originates from an insurance company and initially contains 7.1 million rows of data. To facilitate interpretation and optimize computational efficiency, a 1% sample of the original dataset has been extracted for the subsequent analysis. This sampling approach ensures that the analysis remains manageable while preserving the integrity of the key insights derived from the data.

The dataset includes essential elements necessary for exploring the structural relationships within the U.S. healthcare network. It contains unique identifiers for healthcare providers (NPIs) and organizations (TINs), which serve as the foundation for mapping affiliations and connections. Additionally, the data captures shared relationships between providers and organizations, enabling a detailed examination of the network's structure and identifying clusters and central hubs. The provider's dataset also includes supplemental details, such as taxonomy classifications that categorize providers by specialty, along with relevant financial and geographic information. These features allow for a nuanced exploration of the distribution and dynamics of provider affiliations across different organizational and regional contexts.

Techniques Applied

I employed a combination of K-means clustering and network analysis to analyze the structure and connectivity of provider-organization relationships within the U.S. healthcare network. These complementary techniques allow for both the examination of the overall network structure and the identification of distinct groups of providers based on their affiliation patterns. K-means clustering complements network analysis by grouping providers into clusters based on the number and diversity of their affiliations with organizations. The centroid-based approach of K-means is particularly effective for handling large-scale data and identifying clusters of similar size, which aligns well with the goal of categorizing providers by their connection patterns. This method allowed me to identify distinct types of providers based on their roles within the network, while also ensuring computational efficiency. Alternative clustering techniques, such as density-based methods, were less suited to this context due to the variability in the sizes and densities of provider affiliation patterns.

Network analysis involved constructing two types of graphs to model relationships: a bipartite graph of NPIs and TINs, and a derived NPI-NPI graph. In the bipartite graph, NPIs (providers) and TINs (organizations) are represented as nodes, with shared affiliations forming the edges connecting them. This framework highlighted the direct relationships between providers and organizations, allowing me to identify central hubs—nodes with a high number of connections—representing key organizations or providers with extensive affiliations. These high-degree nodes play influential roles within the network, offering insights into the structural hierarchy and key stakeholders.

To further explore provider relationships, I transformed the bipartite graph into a unipartite NPI-NPI graph, where edges represent shared TIN affiliations between providers. This analysis revealed patterns of co-affiliation and allowed for the identification of tightly connected groups of providers, reflecting their shared organizational or regional connections. Additionally, I applied community detection techniques to uncover clusters within both the bipartite and NPI-NPI graphs. These clusters represented natural groupings of providers and organizations, shedding light on regional systems or distinct healthcare networks. This multi-layered network analysis provided a comprehensive understanding of the organizational layers and interaction patterns within the healthcare ecosystem.

To validate the results, I cross-referenced the findings from network analysis and clustering with taxonomy classifications of providers, ensuring that the identified patterns align with known provider roles and specialties. By integrating network analysis with K-means clustering, I revealed both structural features and role-based patterns within the healthcare network, to expore insights into its organization and the behaviors of its participants.

Findings:

The application of K-means clustering provided a detailed understanding of the patterns within the healthcare network. By grouping NPIs into seven distinct clusters, the analysis revealed meaningful segmentation among providers. These clusters likely represent providers affiliated with specific organizations, those with overlapping specialties, or providers operating within similar geographic regions. The clustering analysis highlighted key distinctions, such as groups of providers frequently working under the same TIN or those reflecting regional and specialty-based concentrations.

Additionally, the NPI-only network revealed a higher level of connectivity compared to the bipartite graph. This finding indicates stronger direct affiliations among providers, which can facilitate better care coordination and resource sharing within certain groups. The clustering coefficient of 0.3916 further confirmed moderate clustering among NPIs linked to the same TIN, reinforcing the presence of tightly-knit groups with shared affiliations.

One of the most highly connected nodes identified in the network analysis is the NPI 1154094522, which demonstrates an exceptional degree of connectivity with 486 affiliations. Upon further investigation, this NPI was found to be associated with Asian Brookins, a nurse working for Quality of Life Health Service Inc. This healthcare organization operates across multiple locations and offers a wide range of medical specialties, contributing to the high connectivity of this node. Asian Brookins' extensive affiliations reflect the organizational model of Quality of Life Health Service Inc., which transcends the boundaries of a single hospital or facility to provide diverse healthcare services across multiple regions. This multi-specialty approach not only enables broad geographic coverage but also fosters significant collaboration among providers, ensuring access to a variety of services under one network.

The high degree of connectivity observed for this NPI highlights the critical role played by healthcare professionals like Asian Brookins within the network. As a central hub, this node bridges smaller clusters, enhancing overall network cohesion and facilitating coordinated care. The influence of such highly connected nodes underscores the importance of multi-specialty organizations and their affiliated providers in driving systemic healthcare integration and addressing regional disparities in access to care. This finding emphasizes the potential of leveraging highly connected nodes for targeted interventions to improve healthcare delivery and outcomes on a broader scale.

Implications

The insights from this analysis offer valuable opportunities for policy interventions. The clustering analysis points to distinct provider groupings, which could inform targeted resource allocation, especially in areas with shared specialties or geographic overlap. Policymakers could use these clusters to address disparities by incentivizing underserved clusters to connect with larger, more resource-rich networks.

The network analysis revealed significant fragmentation, highlighting the need to improve connectivity across isolated clusters. This insight could guide initiatives aimed at promoting resource sharing or forming partnerships between smaller providers and larger networks. Additionally, the sparsity of the network suggests a need for greater transparency in provider-organization affiliations. Requiring comprehensive disclosure of TIN-NPI relationships would help stakeholders make more informed decisions about healthcare delivery and access. Encouraging integration among smaller, isolated groups of providers could also enhance equity in healthcare access. By fostering collaborations with central hubs or regional networks, policymakers can ensure more equitable distribution of healthcare services across regions and communities.

Limitations and Future Considerations

This analysis has several limitations. Computational constraints necessitated sampling the dataset, which may affect the generalizability of findings. The proprietary nature of the dataset restricts external validation, and it is important to ensure that any policy applications derived from these findings prioritize equity and avoid disadvantaging smaller or independent providers. A significant consideration is that the dataset used in this analysis represents one particular insurance company. While the findings provide valuable insights into this network's characteristics, further research is required to analyze data from other insurance companies to capture a more comprehensive view of the healthcare network. Different insurers may exhibit distinct patterns of provider affiliations, network structures, and degrees of connectivity. Expanding the analysis to incorporate multiple insurance datasets would enable a broader understanding of network characteristics and their implications across the healthcare system.

Future research could also integrate longitudinal data to explore how provider-organization relationships evolve over time. Including additional variables such as patient outcomes, service utilization, and payer information would provide a more holistic view of the network. Advanced analytical techniques, such as non-linear machine learning models, could further enhance the depth and applicability of these insights.

Bibliography:

Gul, H., Al-Obeidat, F., Amin, A., Tahir, M., & Moreira, F. (2022). A systematic analysis of community detection in complex networks. Procedia Computer Science, 201, 343–350. https://doi.org/10.1016/j.procs.2022.03.046

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to data mining. Pearson Education, Inc.