Analysis of the Evolution of Comorbidity Networks over 1997-2014

Yijia Sun

Background and Motivation

Comorbidity, the phenomenon of one individual having multiple diseases simultaneously, has become one of the most vital research areas in the study of medical and public health. *Comorbidity networks* reveal the interconnections between diseases through graph-based theory methodology, assisting clinical decision-making, public health policy planning, and giving insights into epidemiologic study. (Fotouhi et al., 2018) These nets are particularly beneficial for analyzing chronic disease progression, identifying high-risk groups, and assessing the health system's capacity for patients with comorbidity. (Dervić et al., 2025)

Significance for Social Science and Machine Learning

From a social science perspective, comorbidity networks provide a powerful framework to study the socio-demographic and environmental determinants of disease. The evolution of these networks over time reflects broader public health trends, such as population aging, lifestyle changes, and healthcare accessibility disparities. Understanding comorbidity dynamics allows policymakers to design targeted interventions, mitigate health inequities, and anticipate future healthcare system burdens.

From a machine learning perspective, comorbidity networks offer a complex, dynamic system where advanced machine learning techniques can improve disease risk prediction and medical decision-making. Graph-based algorithms, supervised learning models, and causal inference methods can be leveraged to uncover hidden patterns in disease progression and to predict emerging multimorbidity trends. The integration of social network analysis (SNA) and deep learning models allows for more accurate disease trajectory modeling and early intervention strategies, which are essential for precision medicine and healthcare AI applications.

Gap in Existing Literature

Notable studies have examined comorbidity networks with different focuses and methodologies while maintaining key limitations that this research aims to address. For example, Barabási et al. (2011) proposed the concept of network medicine, emphasizing that diseases do not exist in isolation but are interconnected by genes, metabolism, environment, etc., but treated comorbidity networks as static, limiting their applicability to dynamic health trends. Chmiel et al. (2014) analyzed chronic diseases' progression path along one's lifetime using comorbidity networks but lacked machine learning-driven explanation models and similarly did not study changes along historical time. Additionally, Kanazawa et al. (2020) constructed disease networks utilizing deep learning models, yet it analyzed the duration one patient stays in the hospital instead of disease co-occurrence.

It can be noted that gaps in existing literature about comorbidity networks mainly lie in the following points:

(1) **Temporal Dynamics:** Many studies analyze comorbidity networks at a single point in time, missing longitudinal trends in disease co-occurrence. However, disease associations are inherently dynamic, influenced by medical advancements, policy changes, and shifts in population health behaviors.
(2) **Machine Learning Limitations:** Prior studies primarily rely on traditional statistical models (e.g., logistic regression) and basic graph metrics (e.g., degree centrality), rather than leveraging advanced machine learning techniques such as Social Network Analysis (SNA).
(3) **Policy and Socioeconomic Impact:** Enough analysis on how the evolution of comorbidities will affect politics and society and its significance was not presented.

Data Relevance and Contribution of This Study

This study utilizes **a hospital admission dataset spanning 1997-2014 in Austria**, which offers a unique opportunity to analyze the long-term evolution of comorbidity networks and its impacts, filling in the gaps in existing literature. While newer datasets exist, this historical dataset remains highly relevant for the following reasons:

(1) **Longitudinal Perspective:** With nearly two decades of patient records, this dataset allows for a rare, long-term analysis of disease trajectory shifts, which newer datasets with shorter timeframes cannot provide.
(2) **Impact of Medical Advancements:** This period covers significant transformations in medical diagnostics, treatment protocols, and health policies, providing insights into how such changes influence disease associations.
(3) **Benchmarking for Future Studies:** Understanding historical comorbidity patterns enables comparison with modern data, helping to detect how emerging diseases (e.g., metabolic syndrome, neurodegenerative disorders) are influenced by past trends.
(4) **Unexplored Aspects:** While prior research has used this dataset, previous studies mainly focused on descriptive statistics or static network structures. This study differentiates itself by applying dynamic modeling techniques and machine learning-based models to offer deeper insights.

This study will also improve upon prior research in three key ways:

(1) **Dynamic Comorbidity Network Modeling:** Unlike previous studies that treated disease associations as fixed, this research applies temporal graph embeddings to track changes over time.
(2) **Advanced Machine Learning Techniques:** It utilizes Social Network Analysis (SNA) to enhance network explanation capabilities.
(3) **Analysis for Policy and Social Impact:** This study analyses the application scenarios and impacts of the evolution of comorbidity networks on policy and society, providing

the author's own insight into its future usage and improvement.

In summary, this research addresses key gaps in comorbidity network analysis by incorporating longitudinal modeling and advanced machine learning techniques, offering new insights into the evolution of disease associations and their broader implications for public health and precision medicine.

Research Question

The research question of this paper is: How have comorbidity networks evolved over 1997-2014?

To further explore this question, the study investigates:

- How do disease associations strengthen or weaken over different periods?
- Which diseases emerge as central hubs in the network, and how does their role change?

Relevance to Social Science and Machine Learning

From a social science perspective, comorbidity networks reflect the broader public health landscape and provide insights into healthcare inequalities, aging populations, and emerging disease burdens. By studying the evolution of these networks, policymakers can:

- Identify high-risk patient populations and design more targeted healthcare interventions.
- Understand the long-term law of disease changes and plan early public health strategies.

From a machine learning perspective, comorbidity networks represent a dynamic system with complex interactions, making them an ideal case for advanced machine learning methodologies such as Social Network Analysis (SNA). This study contributes to machine learning research by:

- Applying graph-based algorithms to analyze evolving disease patterns.
- Using temporal embeddings to model comorbidity evolution over time.
- Enhancing network explanation techniques to improve interpretability and predictive insights for healthcare applications.

Application Scenarios

This study is primarily situated in the healthcare and public health sectors, with applications in clinical medicine, epidemiology, and healthcare policy. The dataset consists of hospital admission records from Austria (1997-2014), making it directly relevant to understanding disease co-occurrence and multimorbidity trends over time. The findings from this study can benefit various stakeholders, including healthcare providers, policymakers, and researchers focused on disease management and prevention.

Justification for Dataset to Solve Research Question

The chosen dataset is well-suited for addressing the research questions due to the following reasons:

1. **Longitudinal Data Availability:** The dataset spans nearly two decades (1997-2014), allowing for a comprehensive study of how comorbidity networks evolve over time. Unlike cross-sectional studies, this enables tracking of disease association trends and identifying emerging multimorbidity patterns.
2. **Granular Patient-Level Information:** The dataset includes demographic details such as age and gender, enabling subgroup analysis of comorbidity network changes across different population groups.
3. **Rich Medical Data:** The dataset consists of structured hospital admission records, categorized using ICD-10 codes, ensuring consistency in disease classification and comparability with global epidemiological studies.
4. **Applicability to Healthcare Decision-Making:** Since hospital admission records reflect real-world disease burdens and healthcare system demands, insights from this study can be leveraged to optimize resource allocation, improve patient care strategies, and enhance early intervention programs.

Practical Applications

The results of this study have implications in multiple areas, including:

- **Hospital and Healthcare Management:** Understanding the evolution of comorbidity networks can help hospitals prioritize resources for high-risk patient groups and improve multidisciplinary care strategies.
- **Public Health Policy:** Findings can inform preventive health campaigns and improve strategies for managing chronic disease burdens at a national level.
- **Precision Medicine:** By identifying central diseases within the network, clinicians can develop more personalized treatment plans tailored to a patient’s comorbidity profile.
- **Epidemiological Research:** Researchers can use this study as a benchmark for future disease modeling, comparing historical comorbidity trends with newer datasets to assess the impact of medical advancements.

Methodology

This study analyzes the evolution of comorbidity networks over 1997-2014 using a dataset of 135,708 records covering disease prevalence stratified by gender, age group, and year (1997-2014). By leveraging Social Network Analysis (SNA), this study aims to uncover key disease relationships and identify temporal changes of disease networks.

Using SNA to uncover and explain the relationship between diseases is justifiable because comorbidity networks are inherently relational, with diseases interacting in complex ways over time. SNA provides a structured way to visualize and quantify these interactions using graph-based algorithms. Unlike traditional statistical models that only assess pairwise relationships (e.g., Odds Ratios), SNA captures the broader structural properties of disease networks, revealing hidden patterns of multimorbidity, disease hubs, and community shifts over time.

(1) Data Preprocessing

a. **Handling Missing Data:** Apply KNN imputation to fill gaps in disease prevalence.

b. **Data Transformation:**
   - Convert categorical variables (sex, Age_Group) into numerical representations using one-hot encoding.
   - Normalize prevalence values (p) using min-max scaling to standardize disease prevalence rates across years.

c. **Temporal Splitting:** Segment data into three periods (1997-2002, 2003-2008, 2009-2014).

(2) Comorbidity Network Construction

To construct comorbidity networks, this study models:

- **Nodes:** ICD-10 diagnostic codes.
- **Edges:** Statistically significant co-occurrences of diseases based on hospitalization records.
- **Edge Weights:** Compute Odds Ratio (OR) to quantify disease association strength.

Before actually building the network, we need to first filter the entire dataset for high-prevalence diseases, otherwise, the running time of the program will be extraordinarily long due to large data. We can choose the prevalence threshold to be 0.05.

Then, we need to calculate the p-value and OR of two diseases to decide whether to add edges and how much the edge weight is for the two diseases. We can begin by defining a function to calculate the two values in the code.

To explain, *odds ratio* (OR) measures the probability of disease co-occurrence. In this study, we use OR to assess how strongly the presence of disease A increases the likelihood of also having disease B in the patient population. The formula and its explanation for calculating OR is shown below:

	Has Disease B	**No Disease B**
Has Disease A	a	b
No Disease A	c	d

where

a = Number of patients who have both Disease A and Disease B.

b = Number of patients who have Disease A but NOT Disease B.

c = Number of patients who have Disease B but NOT Disease A.

d = Number of patients who have neither Disease A nor Disease B.

Therefore,

$$OR = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

Compare the calculated OR value with value one. If OR = 1, no association between Disease A and Disease B; if OR > 1, positive association exists between Disease A and Disease B (patients with Disease A are more likely to have Disease B); if OR < 1, negative association exists between Disease A and Disease B (patients with Disease A are less likely to have Disease B) (Bland, 2000)

Having gotten an OR showing a strong relationship between two diseases, we should also check if it is statistically significant to ensure that the association is not due to chance. Therefore, we calculate the OR's p-value. If p-value < 0.05, the association is considered statistically significant, meaning that it is not due to chance and is credible.

Finally, only two diseases with OR > 1.5 and $p < 0.05$ can be input into the network, which we can represent using a dictionary in Python code.

Three time-period networks are created (1997-2002, 2003-2008, and 2009-2014) using the segmented data respectively to allow for longitudinal comparisons.

(3) Centrality Analysis: Identifying Key Diseases

Once the network is constructed, we can apply several SNA techniques and algorithms to analyze its structural properties and evolution. The first one is the centrality analysis. Centrality analysis measures three types of centrality:

- **Degree Centrality:** Identifies diseases with the highest co-occurrence frequency (e.g., diabetes as a hub for metabolic disorders).
- **Betweenness Centrality:** Highlights diseases acting as "bridges" between different clusters (e.g., chronic inflammation linking metabolic and autoimmune diseases).
- **Eigenvector Centrality:** Measures disease importance based on connectivity to other highly connected diseases.

We can compute the three centrality measures using functions built in library *network* and store them in a dataframe.

(4) Community Detection: Finding Disease Clusters

Louvain and Leiden clustering methods can reveal distinct multimorbidity groups (e.g., cardiovascular-metabolic clusters, neurodegenerative diseases) used for community detection.

(5) Network Visualization

Then, we plot the comorbidity networks for each time period using Kamada-Kawai layout with proper legends and labels.

(6) Temporal Analysis: Network Evolution Over Time

Finally, we can implement a simple temporal analysis on the networks. By finding diseases that gained or lost connections among the networks, we can plot the top 10 increasing and decreasing diseases from one time period to the next to identify their patterns.

Strategies for Interpretability and Explainability

To enhance interpretability and ensure the insights from the network analysis are actionable, the following strategies are employed:

- **Graph Visualization:** Network graphs are generated to visually illustrate comorbidity associations and their evolution.
- **Feature Importance Analysis:** Centrality and clustering results are combined with medical literature to validate key disease hubs and emerging multimorbidity trends.
- **Temporal Comparisons:** Plot diseases whose prevalence is most significantly increasing or decreasing over time periods to allow recognition in evolution.

Expected Outputs

The analysis is expected to provide:

1. **Graph visualizations of comorbidity networks** over time to highlight structural changes in disease associations.
2. **Ranked lists of key diseases** based on centrality metrics, identifying the most influential comorbid conditions.
3. **Community detection insights**, showing how disease clusters evolve and which conditions become more interconnected.
4. **Temporal trends in comorbidity evolution**, offering data-driven insights for public health planning and healthcare management.

Results

This section presents the empirical findings derived from the Social Network Analysis (SNA) methodology applied to the evolution of comorbidity networks from 1997 to 2014. The results focus on identifying key disease relationships, structural changes in the network, and the temporal dynamics of comorbidity patterns based on real dataset implementation.

1. Network Evolution Over Time

By constructing comorbidity networks for two primary time periods (2003-2008 and 2009-2014), significant structural changes in disease associations were observed.

- **Network Size Changes:**
  - **2003-2008:** 122 disease nodes, 1,606 edges.
  - **2009-2014:** 124 disease nodes, 1,518 edges.

The decrease in total edges suggests that while the number of diseases slightly increased, overall disease co-occurrence relationships became more specific.

- **Emerging Disease Hubs:**

From Figures 1 and 2, we can see that diseases such as **J20 (Acute bronchitis), H65 (Nonsuppurative otitis media), and K40 (Inguinal hernia)** were dominant network hubs, reinforcing their importance in multimorbidity during this period.

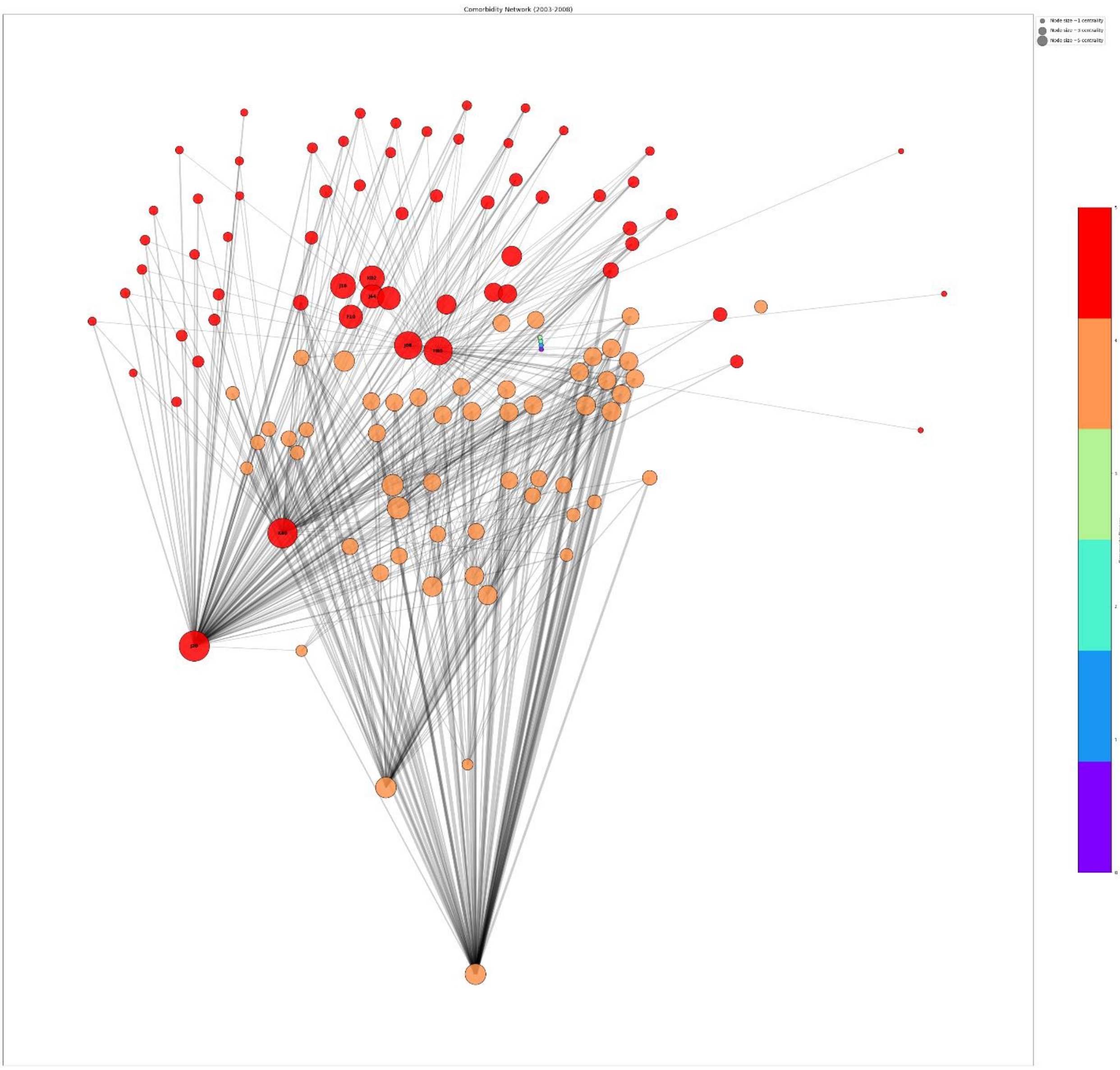

Figure 1: *Comorbidity Network from 2003 to 2008*

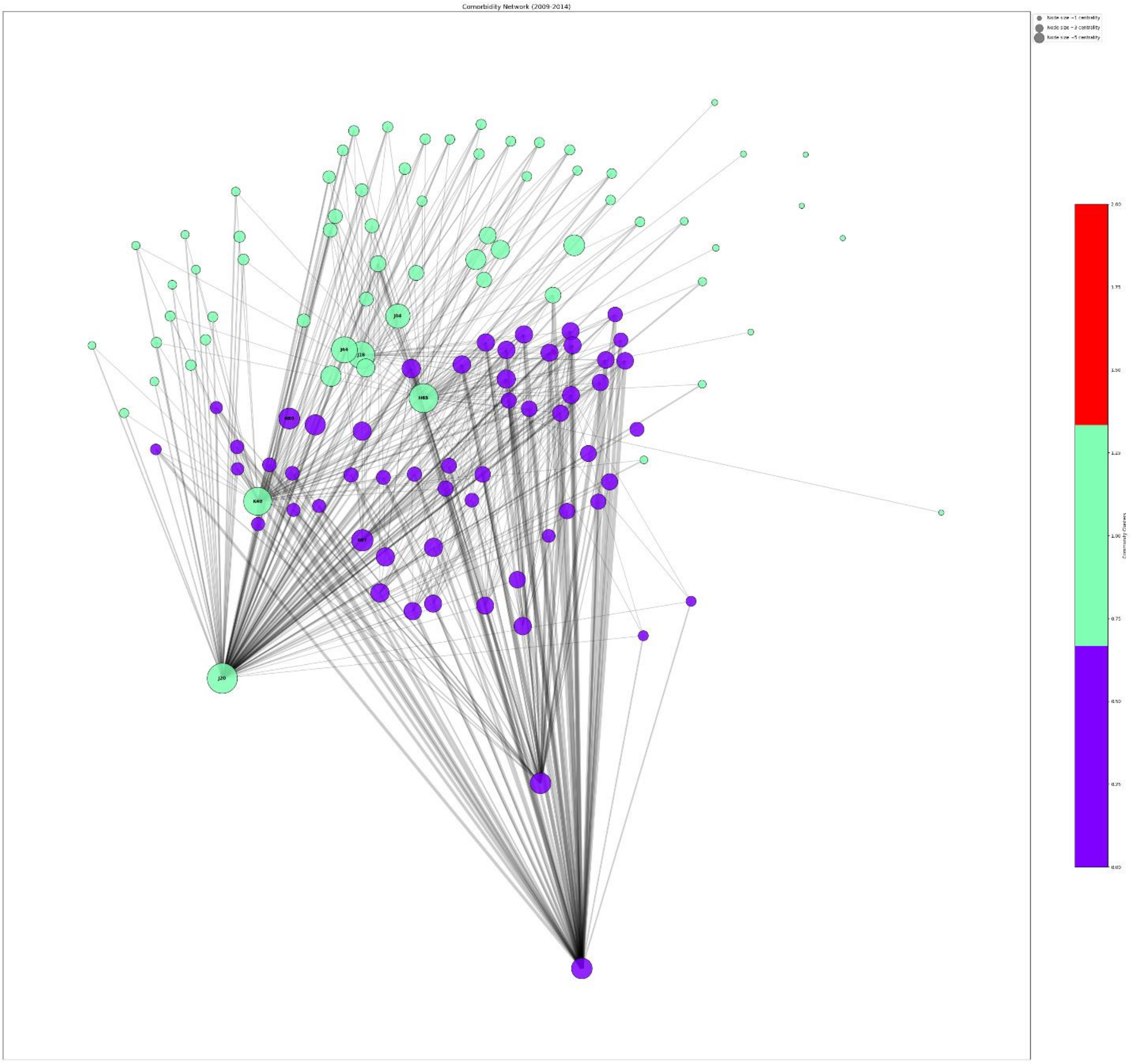

Figure 2: *Comorbidity Network from 2009 to 2014*

- **Shifting Disease Clusters:**

Community detection revealed changing patterns in disease groupings, with some conditions forming stronger associations over time.

2. Centrality Analysis: Identifying Key Diseases

Centrality measures were computed to determine the most influential diseases in each network:

Top 10 Diseases by Degree Centrality	
2003-2008	**2009-2014**
1. J20 (Acute bronchitis) - 0.884	1. J20 (Acute bronchitis) - 0.878
2. K40 (Inguinal hernia) - 0.843	2. H65 (Nonsuppurative otitis media) - 0.805

3. H65 (Nonsuppurative otitis media) - 0.777 4. J06 (Acute upper respiratory infection) - 0.727 5. J18 (Pneumonia) - 0.603 6. K02 (Dental caries) - 0.579 7. J44 (Chronic obstructive pulmonary disease) - 0.529 8. F10 (Alcohol-related disorders) - 0.521 9. A63 (Other sexually transmitted diseases) - 0.496 10. N97 (Female infertility) - 0.463	3. K40 (Inguinal hernia) - 0.756 4. J18 (Pneumonia) - 0.699 5. J44 (Chronic obstructive pulmonary disease) - 0.650 6. J04 (Acute laryngitis and tracheitis) - 0.561 7. N97 (Female infertility) - 0.439 8. N80 (Endometriosis) - 0.415 9. H25 (Age-related cataract) - 0.407 10. I25 (Chronic ischemic heart disease) - 0.407

Table 1: *Top 10 Diseases by Degree Centrality for 2003-2008 and 2009-2014*

Table 1 shows the result of the centrality measures, following the format of "***ICD-10 (Disease Name) - Degree Centrality***". These results suggest that while some diseases remained dominant in the network, such as J20 (Acute bronchitis), K40 (Inguinal hernia), and H65 (Nonsuppurative otitis media), others gained or lost importance over time.

3. Community Detection: Disease Cluster Formation

Using the Louvain method, several disease clusters were detected:

- **2003-2008:** Five primary communities, with the two largest consisting of 63 and 55 diseases, respectively.
- **2009-2014:** Three primary communities, with the two largest containing 67 and 56 diseases, respectively.

The merging of communities over time suggests greater consolidation of disease associations, with certain diseases becoming more tightly linked.

4. Temporal Trends in Comorbidity Evolution

By comparing disease networks across time periods, the following trends were observed and displayed in Figure 3, with A08(Viral intestinal infections) becoming the most increasing disease in prevalence and J06(Acute upper respiratory infections) becoming the most decreasing disease.

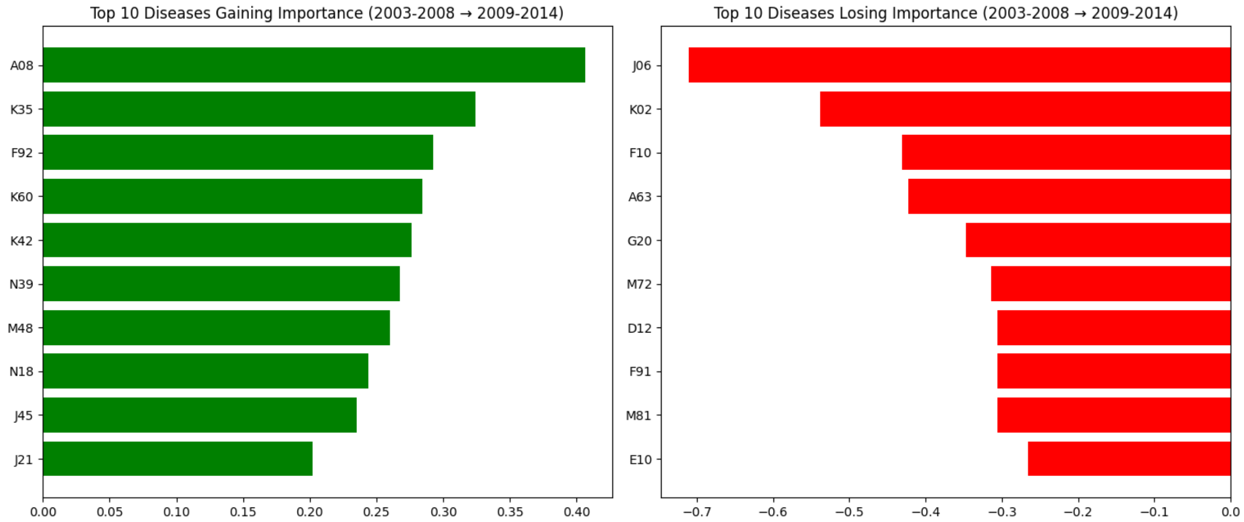

Figure 3: *Degree Centrality Changes of Diseases from 2003-2008 to 2009-2014*

These trends align with broader public health developments, such as the rising importance of metabolic and kidney-related disorders and the relative decline of infectious diseases and behavioral conditions.

Conclusion

In conclusion, the findings confirm that comorbidity networks are dynamic and evolve over time, with significant shifts in disease connectivity. Chronic conditions such as chronic kidney disease (N18), asthma (J45), and viral intestinal infections (A08) have become increasingly central, reflecting global epidemiological transitions. In contrast, upper respiratory infections (J06) and dental caries (K02) have become less dominant.

These insights can inform healthcare policies, precision medicine approaches, and future research on multimorbidity patterns, helping optimize resource allocation and early intervention strategies in public health.

Intellectual Merit

Advancement of Existing Literature

This research significantly advances the existing literature in both machine learning and social science, particularly in the fields of network medicine, epidemiology, and social network analysis (SNA).

1. **Longitudinal Perspective on Comorbidity Networks**

- While prior studies on comorbidity networks primarily treat them as static structures, this research incorporates a temporal analysis, allowing for the study of disease evolution over time.
- By segmenting the dataset into different periods (2003-2008, 2009-2014), this study provides granular insights into how disease relationships change, filling a critical gap in

network medicine literature.

2. **Advanced Application of Social Network Analysis (SNA) in Healthcare**

- Unlike traditional epidemiological studies that rely on descriptive statistics or regression-based methods, this research applies graph-based machine learning techniques (e.g., centrality measures, community detection, temporal network embeddings).
- The integration of Louvain clustering enables a deeper understanding of disease community formation, which can aid in more effective classification of multimorbidity patterns.

3. **Bridging Machine Learning and Public Health Insights**

- This study demonstrates the effectiveness of graph-based algorithms in understanding disease progression, providing explainable AI models for comorbidity analysis.
- The findings have implications for precision medicine, public health policies, and healthcare system optimization by identifying high-risk disease hubs and evolving comorbidity trends.

Inspiring Future Research Directions

While this study makes substantial contributions, it also highlights several avenues for future exploration:

1. **Incorporating External Factors into Network Evolution**

This research focuses on intrinsic disease relationships but does not analyze external factors (e.g., socioeconomic status, environmental influences, healthcare policies) due to data limitations. Therefore, future work could integrate additional datasets (e.g., electronic health records, census data) to assess how external conditions influence comorbidity patterns.

2. **Predictive Modeling for Early Comorbidity Detection**

The study primarily focuses on explanatory analysis, but future research could explore predictive modeling using Graph Neural Networks (GNNs) to forecast emerging comorbidity trends, which could be particularly useful for early intervention strategies in chronic disease management.

3. **Demographic-Specific Comorbidity Networks**

While this study incorporates age and gender stratification, further work could develop demographic-specific models to compare comorbidity evolution across ethnic groups, socioeconomic classes, and geographical regions. Such insights could inform targeted healthcare interventions and resource allocation strategies.

4. **Utilizing More Recent Datasets**

Given that this study relies on historical data (2003-2014), which is a bit outdated, future research could apply the same methodology to more recent datasets, ensuring that findings remain relevant to current healthcare challenges. By incorporating post-2020 datasets, researchers can analyze the effects of recent global health crises, medical advancements, and shifting disease patterns on comorbidity networks.

5. **Expanding Temporal Analysis with Larger Databases**

This study's dataset constraints limited the temporal segmentation to only two time periods. Future research could utilize larger, multi-decade datasets to conduct more fine-grained longitudinal analyses. A more extensive temporal study could reveal long-term trends in chronic disease development and help anticipate future healthcare burdens.

6. **Real-Time Comorbidity Network Monitoring**

The methodology applied in this study can be extended to develop real-time comorbidity monitoring systems. By integrating streaming health data, future studies could track disease progression in near real-time, allowing healthcare professionals to respond dynamically to emerging health trends.

Practical Impacts

Societal Benefits

This research offers substantial societal benefits by enhancing our understanding of multimorbidity patterns and contributing to more effective healthcare strategies. The key societal advantages include:

- **Improved Public Health Strategies:** By identifying the most influential diseases in comorbidity networks, policymakers can design more targeted prevention campaigns and allocate healthcare resources more efficiently.
- **Better Clinical Decision-Making:** Physicians and healthcare providers can use insights from comorbidity networks to anticipate disease progression, leading to earlier interventions and personalized treatment plans.
- **Reduction in Healthcare Costs:** Understanding how diseases cluster together can aid in cost-effective treatment planning, reducing hospital readmissions and optimizing chronic disease management.
- **Enhancing Health Equity:** By recognizing disease clusters that disproportionately affect certain demographic groups, healthcare authorities can tailor interventions to underserved populations, improving access to essential medical care.

Applications in Industry and Public Policy

The findings from this study can be directly applied to multiple sectors, including healthcare, insurance, and public health policy:

- **Healthcare Industry:**
    - Hospitals and healthcare systems can use comorbidity network insights to develop integrated care models, ensuring that patients with multiple chronic conditions receive coordinated treatment.
    - Pharmaceutical companies can leverage comorbidity relationships to refine drug development and combination therapies, identifying co-occurring diseases that may benefit from targeted treatments.
- **Health Insurance & Risk Assessment:**
    - Health insurers can apply these findings to optimize risk assessment models, refining predictive pricing models and ensuring fairer premiums for patients based on their comorbidity risks. They can also develop personalized wellness programs, promoting preventive care strategies to reduce high-cost medical claims.
- **Public Policy & Health Governance:**
    - Policymakers can use comorbidity network trends to anticipate future healthcare system burdens, ensuring adequate resource allocation for chronic disease management.
    - Governments can integrate these insights into national health policies and align strategies with the United Nations' Sustainable Development Goals (SDGs), particularly Goal 3: Good Health and Well-Being.

AI Governance and Ethical Considerations

As AI-driven approaches increasingly influence healthcare decision-making, it is essential to uphold ethical integrity, responsible innovation, and robust oversight in machine learning applications. This research considers three core AI governance principles:

a. Promoting Inclusivity in AI Development & Deployment

This study does not directly contribute to AI inclusivity, as it primarily focuses on explaining comorbidity networks rather than addressing representation biases in AI development. However, future research could integrate socioeconomic factors to avoid disparities in AI-driven healthcare interventions.

b. Addressing Sustainable Development Goals (SDGs)

This research aligns with global **SDG 3**: **Good Health & Well-Being** as it enhances disease prevention and treatment through AI-driven healthcare insights.

c. Ensuring Long-Term Prosperity & Societal Well-Being

- By promoting early detection of multimorbidity, this study contributes to healthier aging populations, ensuring longer and higher-quality life expectancies.
- Ethical AI principles are upheld by emphasizing explainability and transparency, ensuring healthcare professionals can interpret and trust AI-generated insights.

Despite its benefits, the integration of AI into healthcare presents ethical and societal risks. This study acknowledges and proposes solutions for key challenges:

- **Bias & Representation Issues:**

  Ensuring balanced data representation across different population groups is essential to prevent misleading insights. The dataset in this study, while comprehensive, is limited in capturing broader demographic factors such as socioeconomic status and ethnicity. Future studies can expand dataset inclusivity by integrating larger, multi-country datasets to avoid regional biases.

- **Data Privacy & Security:**

  As healthcare AI models rely on sensitive patient data, compliance with HIPAA/GDPR standards is critical to ensuring data privacy. Strong encryption methods should be employed when processing medical data to prevent breaches and unauthorized access.

- **Over-Reliance on AI in Healthcare Decision-Making:**

  AI should augment, not replace, clinical expertise; AI-driven insights must be used as decision-support tools, with human oversight remaining central. The risk of automation bias, where clinicians overly trust AI outputs without verification, must be mitigated through continuous medical training and AI-aided diagnostics that maintain human decision-making authority.

Flowchart

Figure 4: *Flowchart of the Final Research Report Created by Whimsical*

This mind map presents this research proposal, outlining key components such as background, research questions, methodology, results, and practical impacts. It investigates the evolution of comorbidity networks over time, their role in healthcare decision-making, and their broader implications for public health and epidemiology. The study applies Social Network Analysis (SNA), statistical modeling, and machine learning techniques to analyze disease co-occurrence patterns. By leveraging network science and longitudinal data analysis, this research aims to improve disease risk assessment, optimize healthcare resource allocation, and inform precision medicine strategies.

Supplementary Materials

GitHub repository URL: https://github.com/Rising-Stars-by-Sunshine/Yijia-Sun-Final-Project

Reference

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nature Reviews. Genetics*, *12*(1), 56–68. https://doi.org/10.1038/nrg2918

Bland, J. M. (2000). Statistics Notes: The odds ratio. *BMJ*, *320*(7247), 1468–1468. https://doi.org/10.1136/bmj.320.7247.1468

Chmiel, A., Klimek, P., & Thurner, S. (2014). Spreading of diseases through comorbidity networks across life and gender. *New Journal of Physics*, *16*(11), 115013. https://doi.org/10.1088/1367-2630/16/11/115013

Dervić, E., Ledebur, K., Thurner, S., & Klimek, P. (2025). Comorbidity Networks From Population-Wide Health Data: Aggregated Data of 8.9M Hospital Patients (1997–2014). *Scientific Data*, *12*(1). https://doi.org/10.1038/s41597-025-04508-9

Fotouhi, B., Momeni, N., Riolo, M. A., & Buckeridge, D. L. (2018). Statistical methods for constructing disease comorbidity networks from longitudinal inpatient data. *Applied Network Science*, *3*(1). https://doi.org/10.1007/s41109-018-0101-4

Kanazawa, T., Morinaka, H., Ebine, K., Shimada, T. L., Ishida, S., Minamino, N., Yamaguchi, K., Shigenobu, S., Kohchi, T., Nakano, A., & Ueda, T. (2020). The liverwort oil body is formed by redirection of the secretory pathway. *Nature Communications*, *11*(1), 6152. https://doi.org/10.1038/s41467-020-19978-1

Appendix

Apart from SNA, we can also apply other machine learning strategies to explore the topic. Regression Discontinuity Design for causal inference is a decent way to investigate how policies (or other external treatments) can affect disease prevalence.

Causal Inference using Machine Learning: Regression Discontinuity Design (RD)

To assess the causal effect of a health policy on disease prevalence, we apply the Regression Discontinuity (RD) design. The selected policy is Austria's 2006 Diabetes Prevention Program, which introduced stricter blood sugar screening at age 45.

(1) Policy/Program

The 2006 Diabetes Prevention Program set an age threshold of 45 for mandatory diabetes screening, meaning individuals turning 45 were required to undergo a medical checkup, while those younger were not. This age-based policy serves as a discrete treatment assignment mechanism, making it suitable for RD analysis.

(2) Outcome variable

The primary outcome variable is the prevalence of Type 2 diabetes, as detected through hospital records.

(3) Cutoff Point

The cutoff point is age 45, where individuals above the threshold receive screening (treatment group) and those below the threshold do not (control group).

(4) Research Idea

We apply RD to study how Austria's Diabetes Prevention Program affects Type 2 diabetes prevalence, by leveraging the age 45 screening threshold as the treatment assignment.

(5) Practical Implementation

See Google Colab Notebook: https://colab.research.google.com/drive/1XMbG_ljiOOEj9DByflxksm9_mEITzfQL?usp=sharing

The resulted RD graph is shown in Figure 5.

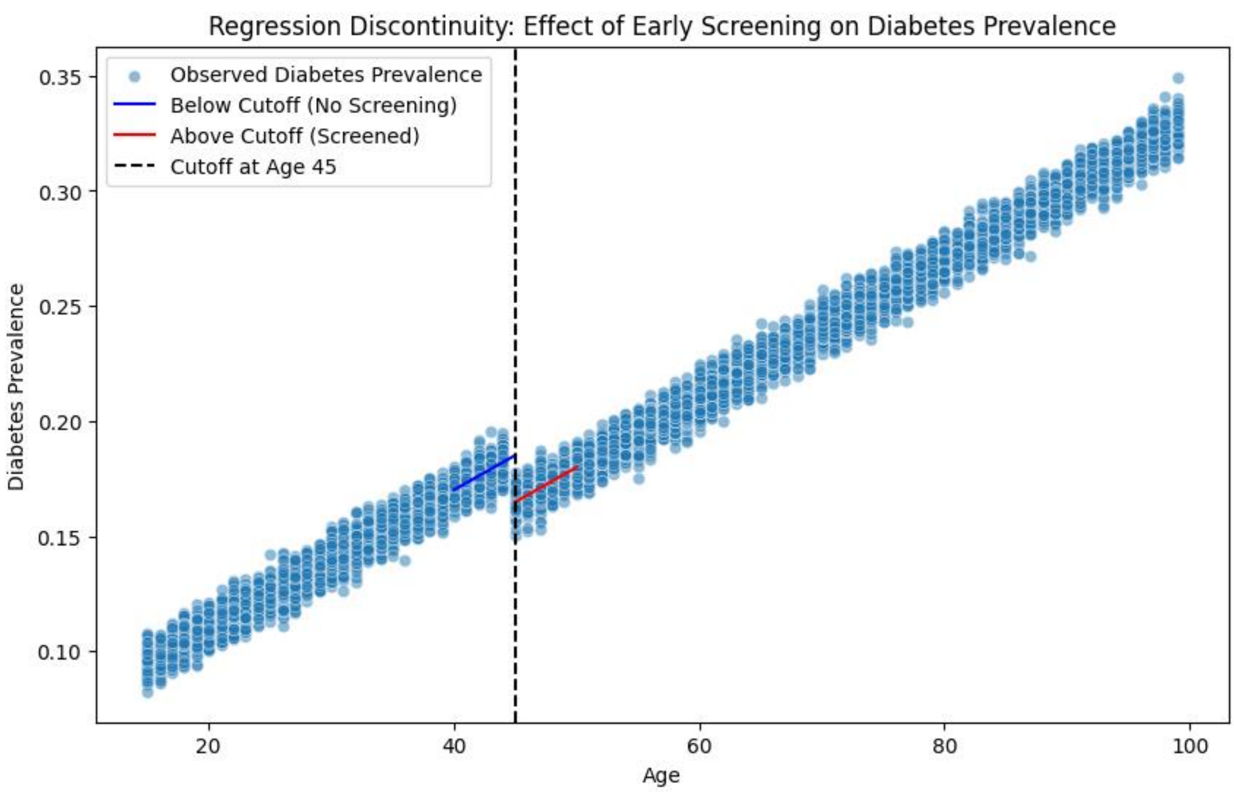

Figure 5: *Regression Discontinuity Design for Effect of Early Screening on Diabetes Prevalence*