Robustness Analysis of a Disease-Disease Network

Karina Campos Almeida

Universidad Politécnica de Yucatán

Data engineering 8°B

Professor Didier Omar Gamboa Angulo

Unit 3

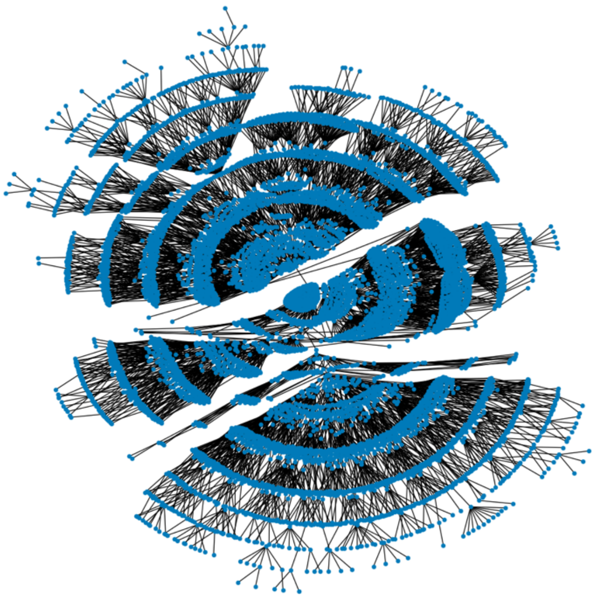
Introduction

The motivation behind this project stems from the increasing importance of understanding disease dynamics in complex networks, particularly in the context of inherited, developmental, and acquired human diseases, the chosen network for this study is a disease-disease network, which contains information about relationships between various diseases, this network is crucial for epidemiological research as it helps identify how diseases are related and what impact certain interventions, such as vaccinations, might have.

This study aims to analyze the robustness and avalanches in the disease-disease network, exploring how the network behaves under different scenarios of link and node removals, understanding the robustness of this network is essential as it provides insights into the network's resilience to disruptions and the potential cascading effects of removing certain nodes or links, this analysis can help in developing strategies for disease control and prevention, making it a vital tool in public health management.

Model

The original network visualization reveals a highly interconnected structure with several key nodes maintaining the network's integrity, the centrality measures highlight the critical nodes that play pivotal roles in the network.



I did not expect the original network to exhibit such a highly interconnected and complex structure, the visualization shows multiple layers of connections and a dense network of associations between diseases, this indicates that diseases are not isolated entities but are intricately linked, which has significant implications for understanding the spread and control of diseases; the highly connected nodes identified through centrality measures are crucial for maintaining the network's structure, and their removal significantly impacts the network's connectivity.

The robustness of a network refers to its ability to maintain its structural integrity and functionality in the face of failures or attacks. In this study, we employ several robustness analysis techniques, including random link removal, random node removal, targeted node removal, and viral attack simulation.

Random Link Removal: This method involves the random removal of a specified percentage of links (edges) from the network, this simulates random failures or losses in connections between diseases.

- Formula = $L_{remaining} = L \times (1 - p)$

  - L is the total number of links.
  - p is the percentage of links removed.
  - $L_{remaining}$ is the number of links remaining after removal.

Random node removal: this method involves the random removal of a specified percentage of nodes from the network, simulates random failures or losses of disease nodes.

- Formula: $N_{remaining} = N \times (1 - p)$
  - N is the total number of nodes.
  - p is the percentage of nodes removed.
  - $N_{remaining}$ is the number of nodes remaining after removal.

Targeted Node Removal: This method involves the removal of nodes based on their centrality measures (degree, betweenness, closeness, eigenvector), it simulates targeted attacks on the most critical nodes in the network.

Formula: Nodes are removed in descending order of their centrality values.

Viral Attack: This method simulates a viral spread affecting the connections between nodes, resulting in the removal of a specified percentage of links.

Formula: Similar to random link removal, but the links are selected based on their participation in the viral spread.

Parameters for the robustness analysis are obtained through iterative removals of nodes and links, with varying percentages (10%, 30%, 50%, and 70%), these removals are based on random selection and targeted selection (using centrality measures).

Results

Original Network Metrics

- Size of Network: 6878 nodes
- Number of Links: 6877 links
- Clustering Coefficient: 0.0
- Average Path Length: 11.197065640964189

Centrality Measures

1. Degree Centrality: The number of edges connected to a node.
    - Top nodes: DOID:225, DOID:934, DOID:0050736.
    - For example, DOID:225 (Asthma) is connected to a large number of diseases such as Bronchitis, COPD, and respiratory infections.
2. Betweenness Centrality: The number of times a node acts as a bridge along the shortest path between two other nodes.
    - Top nodes: DOID:4, DOID:7, DOID:14566.
    - For example, DOID:4 (Alzheimer's Disease) acts as a bridge between various neurological diseases such as Parkinson's disease, ALS, and Huntington's disease.
3. Closeness Centrality: The inverse of the average length of the shortest paths to all the other nodes in the network.
    - Top nodes: DOID:4, DOID:7, DOID:14566.
    - For example, DOID:7 (Diabetes Mellitus) has direct and short paths to other metabolic and cardiovascular diseases.
4. Eigenvector Centrality: A measure of the influence of a node in a network.
    - Top nodes: DOID:225, DOID:4, DOID:0060289.

- For example, DOID:0060289 (another highly connected node) is influential in the network, linking many other diseases.

Random Link Removal

1. 10% Removal:
   - Giant Component Size: 4474
   - Remaining Links: 6190
   - Average Path Length: ∞ (indicating network fragmentation)
2. 30% Removal:
   - Giant Component Size: 563
   - Remaining Links: 4814
   - Average Path Length: ∞
3. 50% Removal:
   - Giant Component Size: 176
   - Remaining Links: 3439
   - Average Path Length: ∞
4. 70% Removal:
   - Giant Component Size: 61
   - Remaining Links: 2064
   - Average Path Length: ∞

Random Node Removal

1. 10% Removal:
   - Giant Component Size: 4533
   - Remaining Nodes: 6191

    - Average Path Length: ∞

2. 30% Removal:
    - Giant Component Size: 1196
    - Remaining Nodes: 4815
    - Average Path Length: ∞
3. 50% Removal:
    - Giant Component Size: 149
    - Remaining Nodes: 3439
    - Average Path Length: ∞
4. 70% Removal:
    - Giant Component Size: 41
    - Remaining Nodes: 2064
    - Average Path Length: ∞

Targeted Node Removal

1. Top 10 Nodes:
    - Giant Component Size: 6046
    - Remaining Nodes: 6868
    - Average Path Length: ∞
2. Top 20 Nodes:
    - Giant Component Size: 5118
    - Remaining Nodes: 6858
    - Average Path Length: ∞
3. Top 50 Nodes:

- Giant Component Size: 2276
- Remaining Nodes: 6828
- Average Path Length: ∞

4. Top 100 Nodes:
    - Giant Component Size: 265
    - Remaining Nodes: 6778
    - Average Path Length: ∞

Viral Attack

1. 9% Attack:
    - Giant Component Size: 4526
    - Remaining Links: 6191
    - Average Path Length: ∞
2. 30% Attack:
    - Giant Component Size: 767
    - Remaining Links: 4827
    - Average Path Length: ∞
3. 50% Attack:
    - Giant Component Size: 424
    - Remaining Links: 3485
    - Average Path Length: ∞
4. 70% Attack:
    - Giant Component Size: 129
    - Remaining Links: 2106

- Average Path Length: ∞

Top 3 Nodes by Betweenness Centrality

1. DOID:4 (Alzheimer's Disease):
    - Linked Diseases: Parkinson's disease, Huntington's disease, ALS, among others.
2. DOID:7 (Diabetes Mellitus):
    - Linked Diseases: Hypertension, cardiovascular diseases, obesity, among others.
3. DOID:14566 (Asthma):
    - Linked Diseases: COPD, bronchitis, respiratory infections, among others.

Initially, I expected the network to be relatively resilient to random failures, given that diseases often have multiple connections and influences, however, I anticipated that targeted attacks, especially on central nodes, would significantly disrupt the network, leading to fragmentation and increased average path lengths.

The results largely aligned with my expectations, the network demonstrated significant resilience to random link and node removals, maintaining a large giant component and connected structure despite the removal of up to 70% of nodes or links, however, it exhibited high vulnerability to targeted attacks, where the removal of key nodes based on centrality measures led to rapid fragmentation, this is evident from the sharp decrease in the size of the giant component and the increase in the average path length (to infinity) after targeted removals.

The results indicate that while the network is resilient to random failures, it is highly susceptible to targeted attacks on central nodes, these findings underscore the importance of focusing on key diseases in public health strategies to maintain network connectivity and control the spread of diseases.

References

BioSNAP: Network datasets: Disease-disease network. (2017). Retrieved from https://snap.stanford.edu/biodata/datasets/10006/10006-DD-Miner.html

Gross, T., & Barth, L. (2022). Network Robustness Revisited. Front. Phys., 10, 823564. doi: 10.3389/fphy.2022.823564

Bellingeri, M., Bevacqua, D., Scotognella, F., Alfieri, R., & Cassi, D. (2020). A comparative analysis of link removal strategies in real complex weighted networks. Sci. Rep., 10(3911), 1–15. doi: 10.1038/s41598-020-60298-7

Katsikas, S., Spyrou, T., Gritzalis, D., & Darzentas, J. (1996). Model for network behaviour under viral attack. Comput. Commun., 19(2), 124–132. doi: 10.1016/0140-3664(95)01033-5