

Analysis of Antibigram Data for Identifying Patterns of Transmission in Hospital Associated Infections

Student: Shan Akiraj
Mentor: Anil Vullikanti

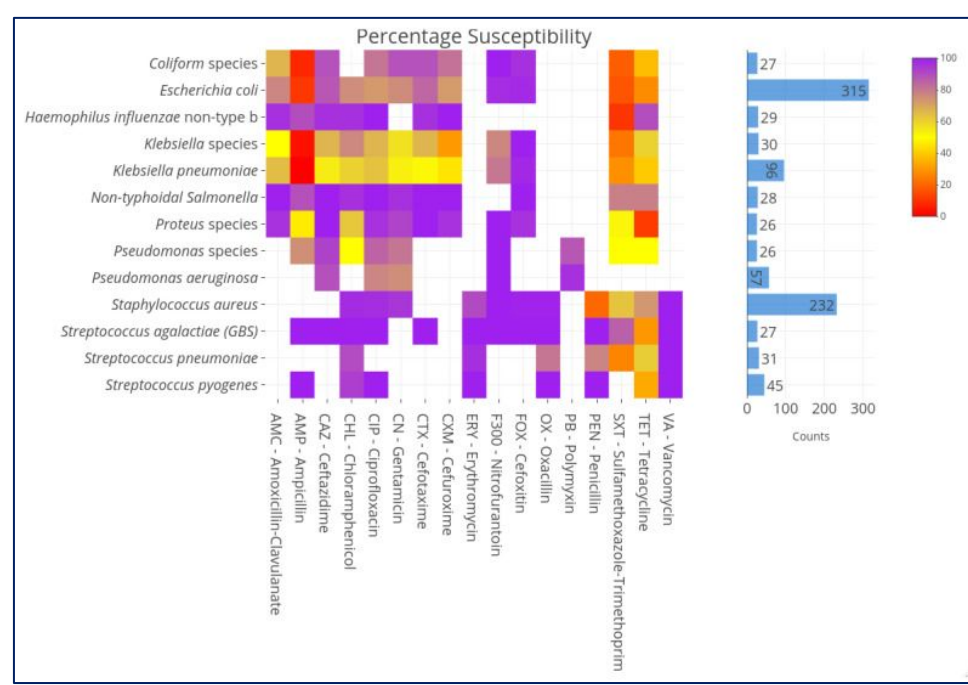
Introduction

- Antibiogram data holds significant value in healthcare. These data sets provide vital information on the susceptibility of bacterial pathogens to various antibiotics. However, this data can also be used for more than just determining effective antibiotics.

CultureType	ReportDate	Organism	Suborganism	AntimicrobialAgent	CLSI_ Interpretation	SensitivityValue
Culture, Urine	2018-11-26 16:36:00.000	ESCHERICHIA COLI	2	Amikacin	Susceptible	<=2
Culture, Urine	2018-11-26 16:36:00.000	ESCHERICHIA COLI	2	Ampicillin	Susceptible	8
Culture, Urine	2018-11-26 16:36:00.000	ESCHERICHIA COLI	2	Aztreonam	Susceptible	<=1
Culture, Urine	2018-11-26 16:36:00.000	ESCHERICHIA COLI	2	Tobramycin	Susceptible	<=1
Identification & Susceptibility	2020-09-11 15:05:00.000	BREVINDIMONAS DIMINUTIVESICULARIS	1	Ceftriaxone	Susceptible	8

Utility

- This data can also offer invaluable insights into the transmission patterns of bacterial infections.
- The overarching goal of this research is to find patterns or trends in the antibiogram data to determine which patients had the highest likelihood of transmitting to each other.

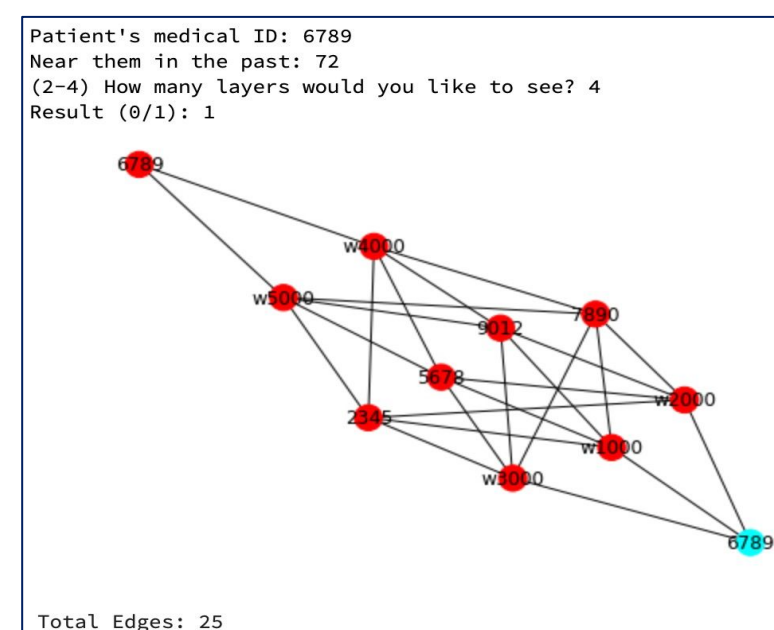
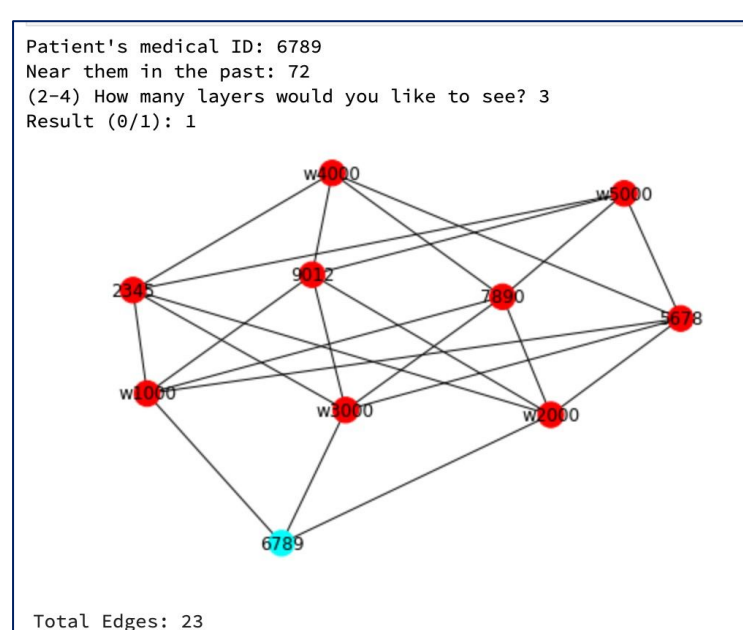


Purpose

- The ability to determine potential patient-to-patient bacterial transmission can enhance infection control strategies, prevent outbreaks, especially of multidrug-resistant organisms, and ultimately, improve patient outcomes.
- Healthcare-associated infections pose a significant public health challenge. Bacterial infections, in particular, can lead to severe health complications and even death if not correctly managed.
- By scrutinizing the similarities and differences in patient antibiograms, we aim to uncover potential transmission events and develop predictive models for future transmission scenarios.

Future Work

- Further work must be done to validate the data against an ego network to find patients with a high likelihood of having transmitted to each other.
- The KNN algorithm results must also be validated against the Hamming Distance data to determine if any patterns or trends are apparent in both datasets.
- Trends that are present in both the clustered data and the ego network create additional support to show that those specific patients were transmitters.
- Other future work includes gathering better data samples that contain more data on patient interactions as well as more accurate temporal data in order to make analysis easier.
- Future progress must also be made to use these metrics to enhance infection control strategies and improve patient outcomes.



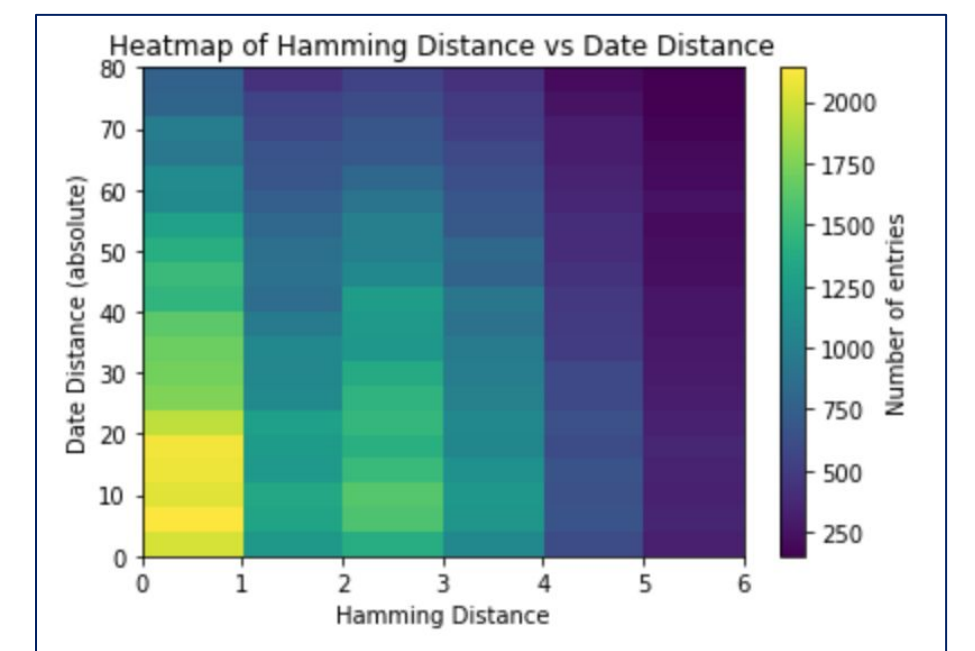
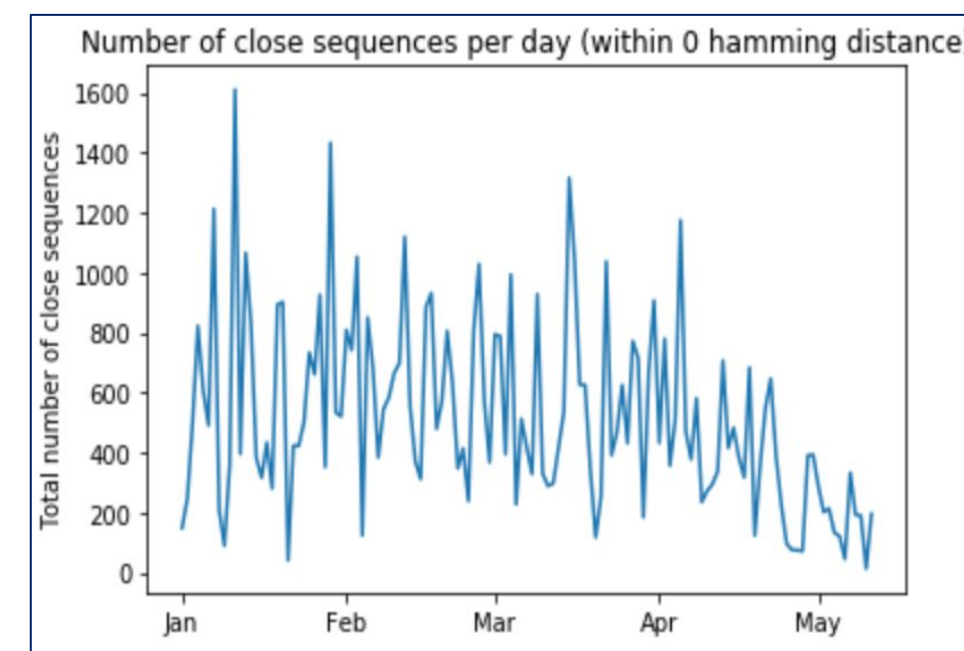
References

Darboe, Saffiatou, et al. "Using an Antibiogram Profile to Improve Infection Control and Rational Antimicrobial Therapy in an Urban Hospital in the Gambia, Strategies and Lessons for Low- and Middle-Income Countries." *Antibiotics (Basel, Switzerland)*, 21 Apr. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10135392/.

Current Work

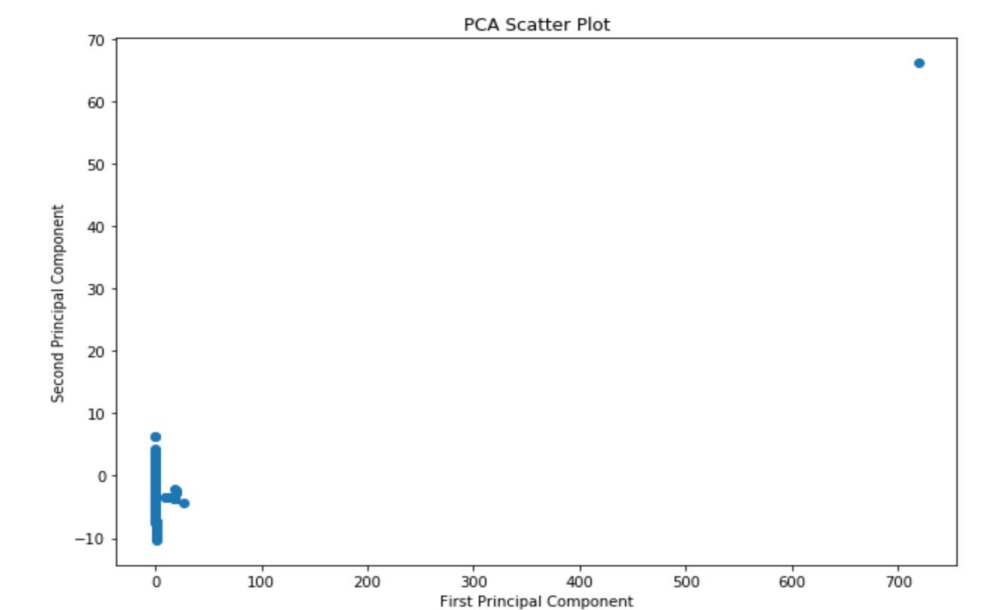
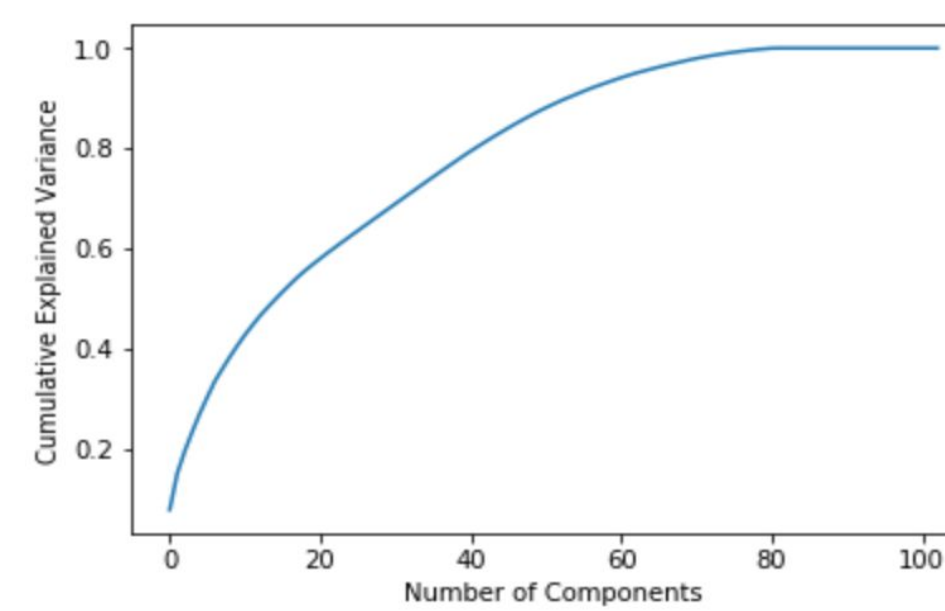
Hamming Distance Calculations

- To determine which patients had a high likelihood of transmitting to one another, we must determine the similarities in their antibiogram data. For this purpose, we use a metric known as the Hamming distance.
- The Hamming distance assesses vector similarity by calculating the amount of positions between two vectors in which corresponding positions have different values. The hypothesis is that patients with the lowest Hamming distances were likely to have transmitted to one another.



PCA for Dimensionality Reduction

- There are some drawbacks to using Hamming distance for calculations, and since the dataset had so many features per sample, PCA was used for dimensionality reduction. Below are the results of the PCA. Around 80 components were needed to represent 95% of the variance in the data.



Clustering - KNN

- After PCA was performed on the data, the dataset was then processed to account for report date times and gathered patients who were infected within the same 72 hour window.
- The K-nearest Neighbors algorithm was performed on this patient data and allowed the identification of the five nearest patients to each individual, based on their unique antibiogram data. These patients were then recorded as potential transmitters in a comprehensive dataset. This data-driven approach enabled the successful identification of potential transmitter patients for each individual, providing critical information on how infections might spread within a healthcare facility setting.
- Using this data we can develop contact networks to identify how and why these patients possibly transmitted to each other and use this information to further progress infection control strategies.

Network Graph of Patients and Transmitters

