A Knowledge Graph Approach to Discover Adjuvants for β-Lactam Antibiotic Resistance

Supervisor: Gizem Buldum
Researcher: Ethan Lucey

Project Aim
The aim of this project is to develop and validate a computational framework that uses knowledge graphs and machine learning to identify novel, non-antibiotic adjuvants capable of restoring the efficacy of β-lactam antibiotics against resistant bacteria.

Project Objectives

<u>Technical</u>
- To construct a heterogeneous knowledge graph that integrates diverse biological data including drug-protein, drug-drug and drug-bacteria interactions from public databases.
- To develop and apply a graph-based link prediction model to the knowledge graph, capable of scoring and ranking potential drug-adjuvant combinations based on their likelihood of producing a synergistic effect, evaluated using metrics such as Area Under the Receiver Operating Characteristic Curve (AUROC), Precision@K, and Mean Reciprocal Rank (MRR).
- To computationally validate the model's top-ranked predictions by searching for supporting evidence within existing scientific literature and specialised databses like DrugComb.

<u>Personal</u>
- To develop practical skills in graph-based machine learning, data integration and computational biology methodologies to support my career transition into data-intensive software engineering roles.

Description

Antimicrobial resistance (AMR) is a significant and growing threat to global public health. With the development of new antibiotics slowing due to economic and regulatory challenges, alternative strategies are urgently required to combat drug-resistant infections. One of the most effective strategies is the use of antibiotic combinations, which can offer improved efficacy at lower doses and slow the emergence of resistance. This project will focus on rescuing widely-used β-lactam antibiotics with helper compounds, or adjuvants, that inhibit bacterial resistance mechanisms.

The primary challenge is that the combinatorial space of potential drug-adjubant pairs is too vast for experimental screening. Therefore, the main focus of this project is to use computational methods to efficiently identify the most promising candidates. This will be achieved by building and analysing a knowledge graph (KG), a powerful structure for representing complex biomedical data from heterogeneous sources. By developing a link prediction model, this project will uncover previously unknown relationships between β-lactam antibiotics and potential adjuvants.

The methodology is supported by recent literature where KG-based frameworks have successfully predicted synergistic drug combinations, with some models achieving high validation hit rates in laboratory settings.

Initial Areas of Research

1. Literature review: Conduct a focused review of the provided papers and other recent literature on KG-based models for drug synergy prediction, concentrating on the specific methodologies for KG construction and link prediction.
2. Data Exploration: Explore public databases such as DrugBank, STRING, DrugComb etc. to assess data availability, file formats and access requirements for building the project's knowledge graph.

Expected Outcomes

By the end of the project, I will deliver the following:

- A well-documented, heterogeneous knowledge graph integrating data from multiple biomedical sources.
- A trained and evaluated link prediction model, assessed using AUROC, Precision@K, and MRR, with all associated code available in a project repository.
- A ranked list of novel, computationally-validated combinations of β-lactam antibiotics and adjuvants with high synergistic potential.
- A comprehensive dissertation report detailing the project's methodology, results, analysis and conclusions.

Project Plan and Timeline

June 17 - June 30: Finalise research and scoping

- Complete literature review .
- Explore and assess the required datasets from open-source data sites for access and structure.
- Finalise the specific KG schema and data integration strategy.

July 1 - July 20: Data engineering and model development

- Construct the knowledge graph by cleaning, integrating and preprocessing the data.
- Develop the core link prediction model architecture and begin initial training and hyperparameter tuning.

July 21 - August 4: Experimentation and analysis

- Complete the model training and perform a robust evaluation of its performance using AUROC, Precision@K, and MRR to assess predictive accuracy and ranking quality.
- Generate the final list of ranked predictions for novel synergistic combinations.
- Begin writing up dissertation.

August 5 - August 10: Validation and finalisation

- Conduct the computational validation of top-ranked predictions against literature.
- Complete results, discussion and conclusion sections of dissertation.
- Finalise all code, ensure its well commented and prepare the project repository.
- Proofread and format the final dissertation document with all deliverables.

August 11: Project submission