

**Thesis Project Portfolio**

**Efficient Graph Representation Framework for Chemical Molecule Similarity Tasks**

(Technical Report)

**Interaction between technological and social factors in recent US pharmaceutical developments**

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Jiaji Ma

Spring, 2024

Department of Computer Science

## **Table of Contents**

Sociotechnical Synthesis

Efficient Graph Representation Framework for Chemical Molecule Similarity Tasks

Interaction between technological and social factors in recent US pharmaceutical developments

Prospectus

## Sociotechnical Synthesis

Both my technical project and STS research paper are about the drug development industry, although with different emphases. My technical project focuses on developing a framework for graph machine learning, specifically for chemical molecule graphs. This can potentially be applied in drug discovery and development settings, by enhancing the prediction of molecular properties and similarity searches. My STS research paper analyzes the relationships between technological advancements and social changes in the context of US drug development, challenging the technological determinism narrative. Although these projects address different aspects of drug development, with the technical project focusing on a potential methodological improvement for one of the relevant technologies, and the STS research paper focusing on historical and policy analysis through case studies, they are connected by their potential implications for drug development.

In my technical project, I proposed a two-stage framework to generate efficient vector representations for molecular graphs, which combines the power of the Graph Isomorphism Network (GIN) and Siamese autoencoders. This framework aims to efficiently transform graph data into a lower-dimensional space while preserving critical structural information, which is vital for tasks such as molecular similarity search in drug discovery. The first stage of the framework involves utilizing a GIN model to capture the structural and attribute information of drug molecule graphs, producing high-dimensional vector embeddings. These vectors then go through the Siamese autoencoder in the second stage. This ensures that as much useful information is retained as possible even as the dimensionality of the vector is reduced. My proposed framework is able to produce efficient representations of drug molecules and thus able to enhance the accuracy and efficiency of similarity search for drug molecules. Using machine

learning techniques, the framework can effectively predict similarity between molecules. This enables more targeted and efficient drug discovery processes. The proposed framework aims to advance the field of graph representation for similarity search, providing a powerful tool for the exploration and discovery of new compounds with desired properties. The findings provide valuable insights into the application of machine learning to graph data, specifically chemical molecule analysis. This approach demonstrates significant potential for improving the accuracy and efficiency of drug molecule similarity search for drug discovery.

My STS research paper explores the intertwined relationship between technological developments and societal changes in the context of the US drug development industry from the 1950s to the 1980s. The paper analyzes significant historical events, including the Thalidomide tragedy, the Kefauver-Harris Amendment, and the Orphan Drug Act, to highlight the cyclic and co-dependent nature of technological and social developments. In my analysis I used the theory of technological determinism, which has a reductionist nature at its core, and fails to paint a complete picture. The paper argues that technological determinism is unable to capture the complex dynamics between technological developments and social changes within this context. I argue that within the US drug development industry in this time, the relationship between technological and societal developments is not the one-directional, deterministic relationship proposed by technological determinism. Instead, there is a codependent relationship, technological developments and social changes exert influences over each other in a cyclic fashion. This is exemplified in how regulatory changes, prompted by societal changes, have reshaped technological developments and priorities in drug development, which in turn influence society. The paper reveals the insufficiency of technological determinism due to its nature and highlights the importance for a nuanced understanding of these interactions. The analysis

provides insight for policymakers and industry stakeholders in drug development. It highlights the importance of utilizing social factors in response to technological developments to ensure that social values are considered.

Working on both my technical project and report, as well as the STS research paper simultaneously this semester has provided a valuable perspective for my understanding of the drug development processes and the relevant sociotechnical landscape. This process allowed me to gain insights into the interactions between technology and society. In the process of working on my technical project, as I learned more about the project's background and potential downstream applications, specifically in drug discovery, I became interested in the subject. This not only inspired me to write my STS research paper on this topic, but it also provided me with important knowledge about the technological aspects of drug development, better facilitating my STS research. Additionally, my STS research has also helped put my technical project into better context, especially understanding its potential social implications. My technical project's focus on improving molecular similarity searches in drug discovery through machine learning has benefited from a better understanding of the social context.