Vir Mehta

Abstract

Biomedical knowledge graphs (KGs) offer a powerful framework for representing complex biological systems by integrating genes, drugs, pathways, and diseases within a unified relational structure. This paper demonstrates the practicality and utility of a large-scale biomedical knowledge graph by illustrating its ability to investigate relationships between diseases, drugs, and genes, and to uncover new mechanistic and therapeutic insights such as drug rediscovery. Each node in the knowledge graph represents a biomedical entity (e.g., gene, protein, drug, disease), while edges encode biological or pharmacological interactions.

Multiple machine learning techniques and random-walk-based algorithms were applied to this graph and successfully identified hidden relationships between entities that do not necessarily emerge from more classical data representations. In addition, novel metapath-based algorithms utilizing biologically interpretable paths were implemented over the knowledge graph. Example metapaths include [drug → gene → disease] for identifying candidate drug repurposing targets for Alzheimer's disease, and [gene → pathway → disease] for uncovering molecular mechanisms of neurodegeneration.

This paper focuses primarily on applying these algorithms to cancer and Parkinson's Disease (PD), and trying to investigate drugs related to each condition that could be effectively leveraged for drug repurposing. Drug repurposing is particularly valuable given the high cost, labor intensity, and extended timeline associated with developing new FDA-approved therapeutics. Overall, this work highlights the power of biomedical knowledge graphs and demonstrates the potential of training machine learning algorithms on structured relational data to enable novel scientific discovery.