Visualizing Metabolic Pathways: Enhancing Insight and Analysis

Allen Qian[1], Sha Cao [2] and Chi Zhang, PhD[3]

[1] Carmel High School, Carmel, IN, [2]Department of Medical and Molecular Genetics, [3]Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN

Abstract/Overview

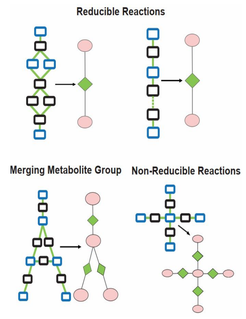
Metabolic pathways play a crucial role in maintaining the energy and essential molecules needed for the proper functioning of the human body through the regulation of processes such as digestion, respiration, hormone production, and immune response. Understanding these pathways is vital to understand human disease and enable advancements in fields such as medicine. However, the complexity of metabolic pathways often poses challenges in their analysis and visualization, as there are over 8,600 unique reactions in a metabolic network. Through the reactions and compounds provided by the KEGG Reaction Database, this study aims to address this issue by developing a methodology for reducing the complexity of metabolic pathway graphs through the identification and removal of redundant compounds and chains of compounds in these networks. By reducing the graph complexity, researchers can focus on the essential components and relationships within the pathways, uncovering key metabolic features and potential regulatory mechanisms, and enabling more efficient exploration and interpretation of these vital biological networks.

Objectives

The objectives of this research are to develop a methodology for reducing the graph complexity of metabolic pathways and to identify and remove redundant compounds within the pathway graphs. The underlying idea of network reduction is to identify network structures that can be reduced without losing major topological property of the network. We first summarized a few basic (commonly seen) structures in a metabolic network:

(1) Directed Chain: A set of reactions of one direction, each edge has exactly one input and one output.
(2) General Chain: Connected reactions that "generally" form a chain shape but may contain small branches or parallel reactions.
(3) Directed Circles: A directed circle.
(4) Directed Tree: A directed acyclic graph, and its underlying undirected graph is a tree (acyclic).
(5) Network Block: A set a highly connected reactions.
Hub node: Metabolite that involves in a lot of reactions.

We further developed a network reduction plan including the following to-be-developed functions:
(1) Chain -> Edge
(2) General chain -> Edge
(3) Circle -> Node
(4) Networks or Tree -> Node



Breadth First Search (BFS) & Depth First Search (DFS)
BFS is an algorithm that travels through a network in a way that the next node is "beside it".
DFS is an algorithm that travels through a network deep, then start back at the beginning and repeats.

Methods

To achieve the reduction of complexity, the reactions encoded in the KEGG database were first analyzed to identify patterns of redundant compounds by visualizing in Cytoscape. This is accomplished by examining the participation of compounds in similar or overlapping reactions. By comparing reaction substrates and products, redundant compounds are identified and flagged for removal.

We further propose methods to identify redundant compounds and eliminate them from the pathway map. This process aims to simplify the network by removing non-essential or duplicated compounds, leading to a more streamlined and visually manageable representation of the metabolic pathways.

For example, a **general chain** can be detected and reduced into an edge by implementing a scoring function into the BFS/DFS method.

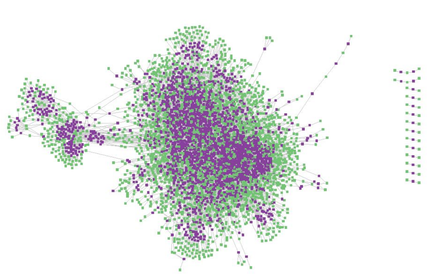


Fig 1. A unreduced representation of a human metabolic network, with green nodes representing compounds and purple representing reactions: created in Cytoscape: This network has 11014 edges.

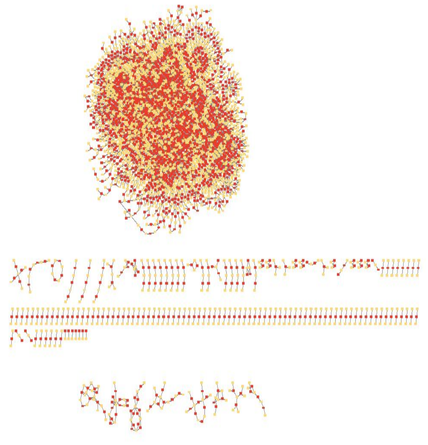


Fig 2. A reduced representation of a human metabolic network, with red nodes representing reactions and yellow representing compounds: created in Cytoscape: This network has 5686 edges

Results

The developed methods effectively reduced the complexity of metabolic pathway graphs by removing redundant compounds and chains, resulting in streamlined and comprehensible visualizations. This simplification allowed researchers to focus on essential components and relationships, revealing key metabolic features and potential regulatory mechanisms. The clearer representation enabled a deeper understanding of critical pathways and their associated compounds. However, more complicated network structures need to be handled by developing additional algorithms

The study's findings demonstrated that the proposed methodology successfully addressed the challenge of reducing graph complexity in metabolic pathway visualizations, contributing to advancements in our understanding of metabolic pathways' role in human health and disease. These outcomes hold significant implications for medicine and biomedical research, providing valuable insights for identifying potential drug targets, personalized therapies, and exploring metabolic dysregulations in various diseases.

Conclusion

In conclusion, the methodology effectively reduced graph complexity by eliminating redundant compounds and chains in metabolic pathways. The streamlined visualization revealed key features and regulatory mechanisms, enhancing our understanding of these vital processes. The approach holds promise for advancing medical research and disease exploration, providing valuable insights into potential therapeutic targets and mechanisms. Overall, the methodology contributes to unraveling the intricacies of metabolic pathways and their significance in human health.

Acknowledgements

I thank the Indiana CTSI, the Project SEED Committee, the Indianapolis Project STEM program, and Biomedical Data Research Lab at Indiana University School of Medicine for supporting this summer internship experience.

References/Resources

KEGG REACTION Database, 21 May 2021, https://www.genome.jp/kegg/reaction/. Accessed 12 July 2023.
Altman, Tomer. "A systematic comparison of the MetaCyc and KEGG pathway databases - BMC Bioinformatics." BMC Bioinformatics, 27 March 2013, https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-112. Accessed 10 July 2023.