AI-based Pathway Visualization (AI MPW)

A. Sorokin, S. Cohen, I Goryanin

Abstract:

The explosion of high-throughput data-acquiring techniques in biology, such as -omics, alongside the rapid development of sophisticated statistical and machine learning analysis algorithms, underscores the critical need for interpretability in biological data analysis pipelines. Pathway diagrams stand as a cornerstone visualization technique for interpreting data and modeling results in biology. While the Kyoto Encyclopedia of Genes and Genomes (KEGG) [1] has historically served as a primary resource for mapping data onto pathway diagrams, its static nature and transition to a commercial platform have posed limitations. Alternative pathway collections like MetaCyc [2] and Subsystems [3] have struggled to gain attention, partly due to issues such as the overly fine-grained structure of MetaCyc pathways and the absence of visualizations in Subsystems.

Efforts to address the visualization challenge have historically focused on the development of sophisticated graph layout algorithms [4]. However, these approaches have often fallen short, producing results that are challenging to interpret and compare, particularly due to variations in layout even on the same network across different invocations.

Recent advancements in large generative models in artificial intelligence (AI) present an opportunity to tackle this challenge. We propose the development of an AI-based Pathway Visualization (AI MPW) system, leveraging a hierarchical Subsystems approach to pathways. This system aims to automatically generate static pathway diagrams linked to original network representations and open-source annotations of metabolites, reactions, and enzymes. By utilizing standard biochemical network representation formats such as Systems Biology Markup Language (SBML) [5], the AI MPW system seeks to generate aesthetically pleasing and easily interpretable pathway diagrams. Moreover, the system will incorporate automatic extraction of new pathway knowledge from literature using large language models (LLMs).

To facilitate the logical storage of pathway information, SBML to Systems Biology Graphical Notation (SBGN) [6] converters will be implemented. Additionally, the generated pathway diagrams will be available in multiple editable graphical formats such as JPEG and PNG. Future commercialization of the AI MPW system is envisioned, potentially following models similar to KEGG or as a standalone application, eliminating the need for manual drawing and enhancing accessibility to biological pathway information.

Introduction:

The rapid advancement of high-throughput data acquisition techniques in biology has revolutionized the field, enabling researchers to generate vast amounts of data. However, the interpretation of this data presents a significant challenge due to its complexity and scale. Pathway diagrams play a crucial role in interpreting biological data, providing visual representations of complex biological processes. While existing resources like KEGG have been valuable, their limitations necessitate the development of alternative approaches. In this proposal, we introduce an AI-based Pathway Visualization (AI MPW) system aimed at addressing these limitations and enhancing the interpretability of biological data.

Research Challenge:

The primary challenge addressed in this proposal is the need for improved pathway visualization techniques in biology. Existing resources like KEGG offer valuable pathway diagrams but are limited by their static nature and commercialization. Alternative resources lack widespread adoption due to various factors. Traditional approaches to pathway visualization through graph layout algorithms have not fully addressed the interpretability and comparability issues. Therefore, the challenge is to develop an AI-driven solution capable of generating aesthetically pleasing and easily interpretable pathway diagrams from standard biochemical network representations.

Data & Methodology:

The proposed AI MPW system will utilize data in the form of SBML files. Leveraging recent advancements in large generative models in AI, the system will employ techniques to automatically generate pathway diagrams. These diagrams incorporate open-source annotations of metabolites, reactions, and enzymes. Additionally, new knowledge about pathways will be automatically extracted from the literature using large language models.

RRI/Ethical Considerations:

Ethical considerations in developing the AI MPW system include ensuring data privacy and security, as well as addressing potential biases in the generated pathway diagrams. Transparency in the algorithm's decision-making process and adherence to ethical guidelines in data collection and usage will be paramount. Moreover, the system should be designed to promote inclusivity and accessibility in accessing biological pathway information.

Expected Outcome & Impact:

The expected outcome of this research is the development of an AI MPW system capable of generating high-quality, interpretable pathway diagrams. By providing an alternative to existing resources like KEGG, the proposed system has the potential to significantly impact the field of biology and bioinformatics. Enhanced pathway visualization techniques will facilitate better interpretation of biological data, leading to new discoveries and insights. Moreover, the system's open-source nature and potential for commercialization offer opportunities for widespread adoption and further development.

References:

[1] https://doi.org/10.1093/nar/gkac963

[2] https://doi.org/10.1093/nar/gkz862

[3] https://doi.org/10.1186%2F1471-2164-9-75

[4] http://biorxiv.org/lookup/doi/10.1101/2023.12.23.573191]

[5]Hucka, M., Bergmann, F. T., Dräger, A., Hoops, S., Keating, S. M., Le Novère, N., ... & Olivier, B. G. (2019). The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core. Journal of Integrative Bioinformatics, 16(2), 20190012.

[6]Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., ... & Demir, E. (2009). The Systems Biology Graphical Notation. Nature Biotechnology, 27(8), 735-741.