



Bachelorarbeit Medizininformatik

Optimiertes Layouting biologische Netzwerke für die vergleichende visuelle Analyse

(Optimized Layouting of Biological Networks for the Comparative Visual Analysis)

Einführung

Das Universitätsklinikum Tübingen hat in laufenden Forschungsprojekten große Mengen unterschiedlicher Datensätze von Patienten gesammelt, welche ein hohes Diabetes-Risiko aufweisen (Typ II) [1]. Diese Daten sind mit herkömmlichen Methoden nur schwer zu analysieren. Unter anderem ist die Fusion und Korrelation von Informationen aus verschiedenen Quellen zur Unterstützung einer umfassenden Analyse eine ungelöste Fragestellung. Insbesondere für große Datenmengen ist die visuelle Analytik hierfür ein leistungsstarkes Werkzeug. Gleichzeitig wurden durch systembiologische Forschung zahlreiche biologische Netzwerkkarten erstellt, die eine Vielzahl zellulärer Prozesse umfassend abbilden [2]. Diese statischen Karten sollen in diesem Projekt genutzt werden, um die Patientendaten in ihrem funktionalen Kontext darzustellen. Ein besonderer Schwerpunkt liegt auf Lipidomen und mitochondrialen Systemen, die in der aktuellen Literatur noch nicht ausreichend abgedeckt sind.

Zielsetzung

Ziel der Aufgabe ist es, eine bestehendes web-basiertes Programm zur Visualisierung biologischer Netzwerke zu erweitern. Das bisher genutzte Force-Directed-Layout (siehe Abbildung 1) soll durch ein Verfahren ersetzt werden welches das gesamte Netzwerk oder bestimmte Teile davon (sogenannte Pathways) übersichtlich darstellt und so eine einfachere visuelle

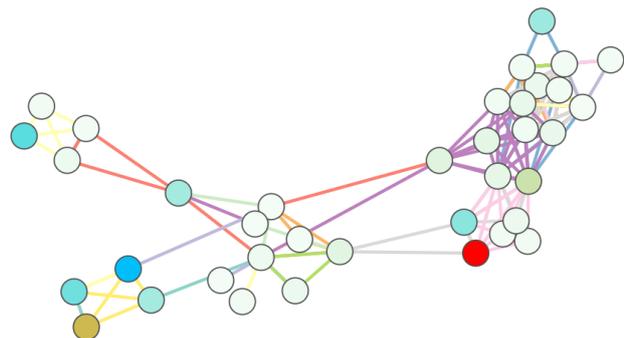


Abbildung 1: Visualisierung eines Gen-Netzwerks auf Basis der KEGG-Datenbank. Die Knotenpositionen wurden mit dem Force-Directed-Layouting der D3-Library berechnet. Die Einfärbung der Knoten zeigt den Fold Change.

Analyse ermöglicht. Hierzu können entweder Graph-Layouting-Verfahren aus der Literatur eingesetzt werden oder ein neues Verfahren entwickelt werden. Denkbar wäre beispielsweise, ein Layout basierend auf den in der KEGG-Datenbank¹ gespeicherten

¹ <https://www.kegg.jp/>



Layouts für einzelne Netzwerke zu entwickeln. Bei der Auswahl und Umsetzung eines geeigneten Verfahrens soll die Vergleichbarkeit von Netzwerken in Betracht gezogen werden. Durch ein angleichen des Layouts soll so beispielsweise der Vergleich verschiedener Messzeitpunkte für ein Individuum oder der Vergleich verschiedener Individuen ermöglicht werden.

Je nach Verlauf der Implementierung und Evaluierung und Komplexität des gewählten Verfahrens sind verschiedene, optionale Erweiterungen denkbar: bisher werden in dem bestehenden Programm nur Transcriptomics-Daten geladen und visualisiert. Um eine umfassendere Datenanalyse zu ermöglichen sollten weitere Omics-Daten (z.B. Proteomics oder Lipidomics) einbezogen werden. Des Weiteren könnte auf Basis des neu entwickelten Layouts eine vergleichende Visualisierung integriert werden. In einer früheren Arbeit an der Universität Tübingen wurde bereits ein auf Java basierendes Tool entwickelt, welches ggf. als Referenz genutzt werden kann².

Kontakt

Jun.-Prof. Dr. Michael Krone michael.krone@uni-tuebingen.de

Jun.-Prof. Dr. Andreas Dräger draeger@informatik.uni-tuebingen.de

Literatur

- [1] X. Liu *et al.*, "Quality Control of Serum and Plasma by Quantification of (4E,14Z)-Sphingadienine-C18-1-Phosphate Uncovers Common Preanalytical Errors During Handling of Whole Blood," *Clinical Chemistry*, vol. 64, no. 5, pp. 810–819, May 2018.
- [2] H.-Y. Wu, M. Nöllenburg, and I. Viola, "A Visual Comparison of Hand-Drawn and Machine-Generated Human Metabolic Pathways," presented at the EG/VGTC Conference on Visualization, 2018.

² http://www.cogsys.cs.uni-tuebingen.de/forschung/terminiert/diabetes/welcome_e.html