



BiGG

Analysing Metabolic Networks in Gradoop

Anja Neumann & Christoph Georgi

Universität Leipzig

31.07.2017

Inhaltsverzeichnis

- 1 BiGG Models - Datenquelle
- 2 JSON-Konverter
- 3 Analyse mit Gradoop und Visualisierung
- 4 Ausblick

Inhaltsverzeichnis

- 1 BiGG Models - Datenquelle
- 2 JSON-Konverter
- 3 Analyse mit Gradoop und Visualisierung
- 4 Ausblick

Inhaltsverzeichnis

- 1 BiGG Models - Datenquelle
- 2 JSON-Konverter
- 3 Analyse mit Gradoop und Visualisierung
- 4 Ausblick

Inhaltsverzeichnis

- 1 BiGG Models - Datenquelle
- 2 JSON-Konverter
- 3 Analyse mit Gradoop und Visualisierung
- 4 Ausblick

BiGG Models

Datenquelle

BiGG Models

- "BiGG Models is a knowledgebase of genome-scale metabolic network reconstructions."
- mehr als 70 veröffentlichte Datensätze
- für den nicht-kommerziellen Gebrauch frei verfügbar.
- Das Hosting und die Pflege durch das System Biology Research Group an der University of California, San Diego.

BiGG Models

- "BiGG Models is a knowledgebase of genome-scale metabolic network reconstructions."
- mehr als 70 veröffentlichte Datensätze
- für den nicht-kommerziellen Gebrauch frei verfügbar.
- Das Hosting und die Pflege durch das System Biology Research Group an der University of California, San Diego.

BiGG Models

- "BiGG Models is a knowledgebase of genome-scale metabolic network reconstructions."
- mehr als 70 veröffentlichte Datensätze
- für den nicht-kommerziellen Gebrauch frei verfügbar.
- Das Hosting und die Pflege durch das System Biology Research Group an der University of California, San Diego.




BiGG Models

- "BiGG Models is a knowledgebase of genome-scale metabolic network reconstructions."
- mehr als 70 veröffentlichte Datensätze
- für den nicht-kommerziellen Gebrauch frei verfügbar.
- Das Hosting und die Pflege durch das System Biology Research Group an der University of California, San Diego.

BiGG Models

- "BiGG Models is a knowledgebase of genome-scale metabolic network reconstructions."
- mehr als 70 veröffentlichte Datensätze
- für den nicht-kommerziellen Gebrauch frei verfügbar.
- Das Hosting und die Pflege durch das System Biology Research Group an der University of California, San Diego.

Testdatensatz - 'iAB_RBC_283.json'

- Organismus: Homo sapiens
- Metabolite: 342 
- Reaktionen: 469 
- Gene: 346 
- Reaktionsorte: Cytoplasma, Interzellularraum
- Subsysteme: 42

JSON-Konverter

Gradoop-Format

EPGMGraphHead

Id	Label	Properties
----	-------	------------

POJO

DataSet<EPGMGraphHead>

EPGMVertex

Id	Label	Properties	Graphs
----	-------	------------	--------

POJO

DataSet<EPGMVertex>

EPGMEdge

Id	Label	Properties	SourceId	TargetId	Graphs
----	-------	------------	----------	----------	--------

POJO

DataSet<EPGMEdge>

EPGMVertex

Id	Label	Properties	Graphs
----	-------	------------	--------

GradoopId := UUID
128-bit

String

PropertyList := List<Property>
Property := (String, PropertyValue)
PropertyValue := byte[]

GradoopIdSet := Set<GradoopId>

https://www.scads.de/images/scads_ringvorlesung/rv-graphs-rahm.pdf

Zielformat



reaction blank node

Id : string
label : string



reaction edge

- id : string
- label : string
- source : string
- target : string
- quantity : int



metabolit node

- label : string
- name : string



gen node

- index : string
- name : string



gen edge

- id : string
- source : string
- target : string

logic graphs



reaction graph

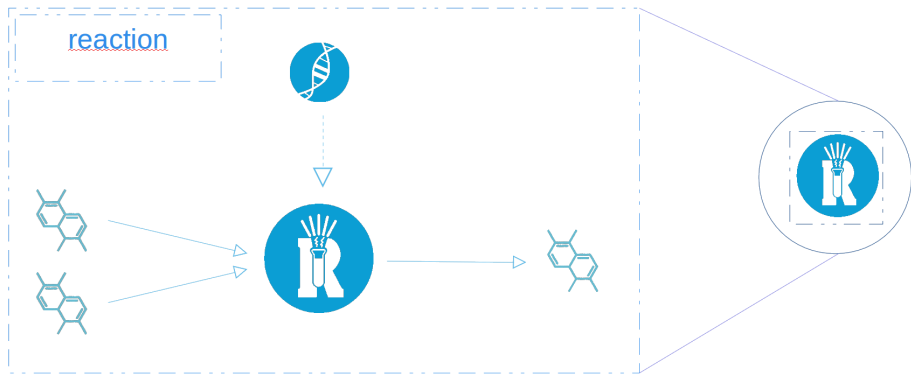


subsystem graph

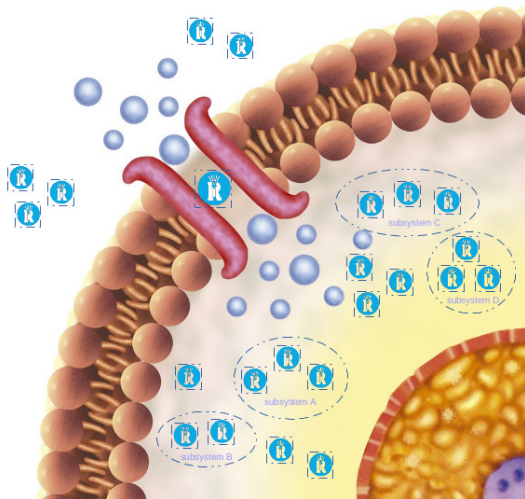


compartment graph

Zielformat



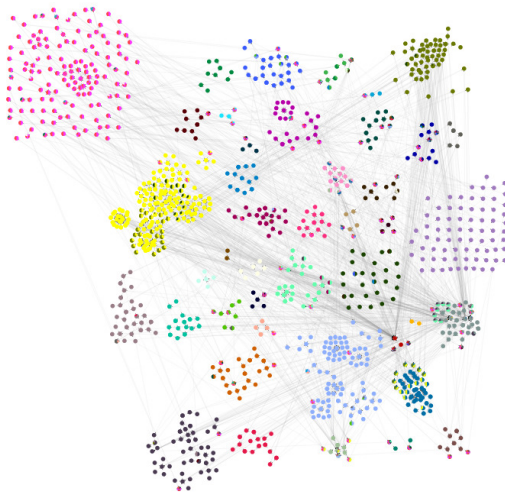
Zielformat



Analyse mit Gradoop und Visualisierung

Analyse mit Gradoop und Visualisierung

Zielformat - Gradoop Visualisierung



Visualisierung mit GradoopVis

- Entwickler: Martin Junghanns & Mohammad Ali Rostami
- Status: Entwicklungsphase
- genutzte Feature:
 - Clustering mittels ClusterId
 - Anzeigen der Labels
 - Einfärben von Knoten nach Label oder Property

Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

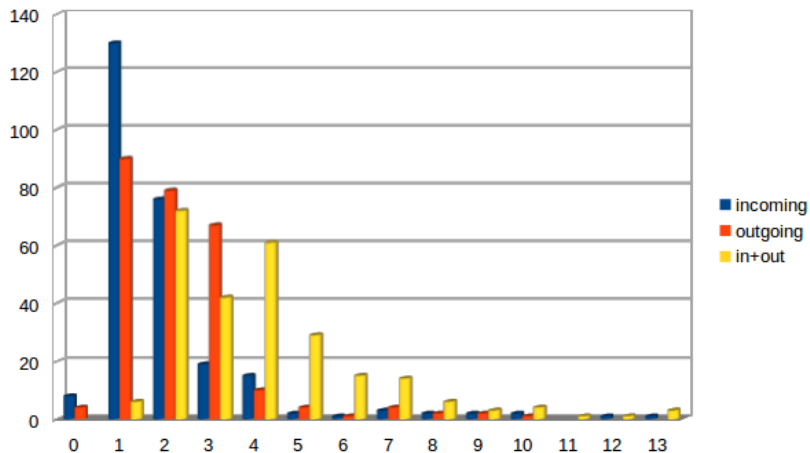
Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

Gradoop-Operationen

- Aggregation (Zählen von Kanten (2456) und Knoten (1082))
- Selektion (Subsysteme, Reaktionen)
- Grouping
- Transformation (Label, Property verändern)
- Pattern Matching (activ transport reactions)

Aggregation (Suche nach Hubs)



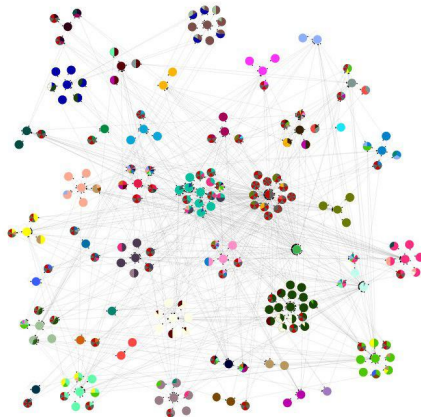
Aggregation (Suche nach Hubs)

Metabolit	Anzahl Kanten	Edukt	Produkt
H ₂ O	124	114	10
CO ₂	13	3	10
ATP	81	79	2
ADP	68	1	67
Phosphate	70	8	62
H ⁺	202	44	158

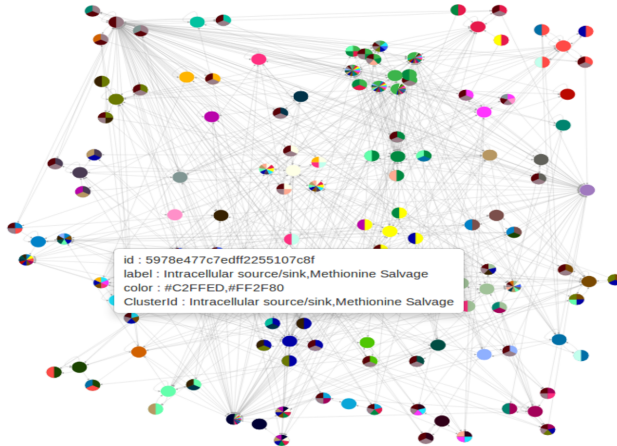
Selektion

Live-Präsentation

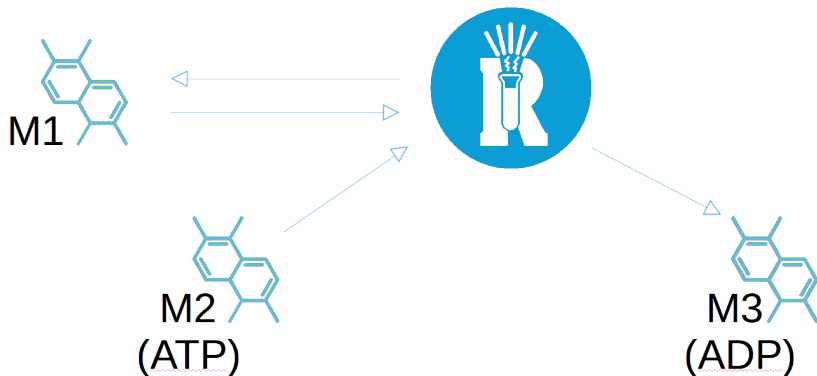
Grouping nach Subsystemen



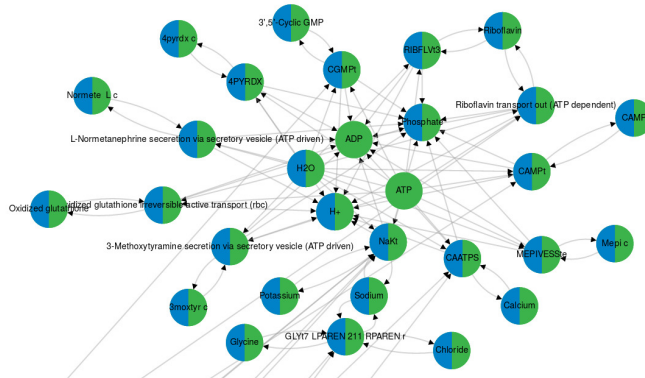
Transformation



Pattern Matching (activ transport reactions)



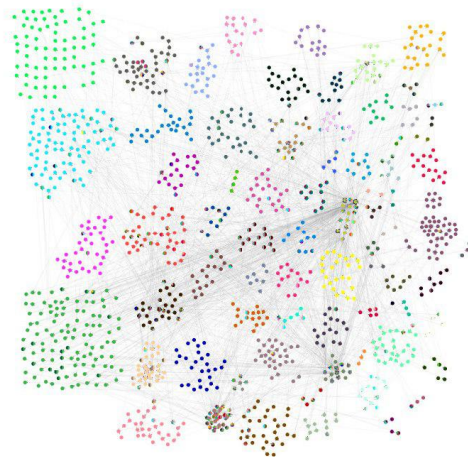
Pattern Matching (activ transport reactions)



Ausblick

- alle Bigg Model in Gradoop einlesbar + visualisierbar
- Suche nach größten Zyklen per Pattern Matching
- Suche nach häufig vorkommenden Strukturen per Frequent Subgraph Mining

Helicobacter 'iT341.json'



Danke für ihre Aufmerksamkeit

Quellen

- 'Verteilte Graphanalyse mit Gradoop' von Junghanns und Petermann (2016)
- 'graph-based data integration and analysis for big data' von Prof. Dr. Erhard Rahm (ScaDS-Ringvorlesung) (2017)
- Datenquelle: <http://bigg.ucsd.edu>