BioReact: Visualization of Systems Biology Reaction Network

Haoran Yu
Zixiao Zhang
2017.12.12
Introduction

- What is a Systems Biology Reaction Network?
  - A visual representation of cause-effect relationships between biological entities
  - The relationships form a large system, resembling a directed graph
Problem

- Networks are very large, users do not know where to find datapoint of interest
- Nodes and links are cluttered, cause cognitive burden
- Related data are not close together, may take too much effort to find closely related data
Related Work: CellDesigner (Brute Force)

- Explicitly specify (x,y) coordinates of each node and link
Related Work: Topological Fisheye Views

- Reduces clutter, but is not tailored for datapoint query or identification
Related Work: Cytoscape

- For experts use only, too difficult to use for new users
Related Work: VIGOR

- Effective graph query, multiple coordinated views, filtering, and result summarization
Data Abstraction

- There are ~40000 nodes and ~100000 links in the complete network, distributed across ~600 models (static XML files).
- There are 2 categories of nodes, Species node and Reaction node.
  - The nodes are always connected alternately, i.e. Species and Reactions connect alternately forming a bipartite graph.
- Each node has additional categorical attributes:
  - id, name
- Each link has additional attributes:
  - source, target (categorical), value (quantitative)
Additional statistics were computed, such as:

- How many times does a Species.name appear in the network?
- How many times does a Species.id appear in the network?
Task Abstraction

- Present the network as node-link diagram to the user
- Allow user to query for a specific node by its name
- Allow user to browse, locate, and explore parts of the network
- Layout of the node-link diagram needs to be clear enough to show topology of the network as well as paths
Applying Technique: IPSEP-COLA

- Since the node-link diagram is directed, would be easier to preserve the flow direction of the links.
- Nodes are separated into different layers, constrained by avoiding overlaps between nodes and links
Our Network Structure

- Force Directed Layout
- Downward Layout (IPSEP-COLA)
Demo
How: Encode

- Species and Reactions: Nodes, circle and rectangle respectively.
- Links: Lines with arrow to denote the direction.
- Group: Color and Saturation.
- Value: Stroke Width of the Link.
- Specific Information: Text label triggered by clicking.
- Derived statistics: Bar chart
How: Reduce

- Query
  - By Species Name, Level, Direction
- Reduce
  - Provide multiple options to filter the searching result.
  - Reduce the complexity of the network.
- Dynamic Aggregation
How: Manipulate

- **Navigate**
  - scroll, drag to dynamically change the view.

- **Select:**
  - click the nodes in the network to get specific information.
  - click the node to make it bigger. Recover by clicking the blank area.
  - click two nodes to show path.
How: Facet

- Linked Highlighting
  - click the button on the sidebar to make the corresponding node bigger. Color to mark the button just clicked.
  - link the ordinal bar chart to the items in the search result.
Result

- A user-friendly interface.
- Two network alternatives.
- For Learner: Provide a more direct way to browse the database. Efficient for developing the knowledge topology.
- For Researcher: Cut off the tedious procedure to look up the species and reactions required in their project.
- For Administrator: Check the incomplete data or error.
Future Work and Limitation

- Optimization the user interaction by feedback. Provide a standard reaction diagram for each reaction node.
- From Vis to Data: Implement an interface to modify the database directly from the front-end.
- Traverse into more nodes in single model view.
- Integrate to a system to allow synchronized update.