Visualization of time-series biological data with spatial and non-spatial features embedded

Chihua Ma
July 22, 2015
• Why time series data?
  Most real world problems are dynamic in nature.

• Why spatial data?
  Those are the properties of biological systems. Such as the physical locations (coordinates) of neurons in networks.
Time series data visualization

- Time-To-Time Mapping
- Animation
- Time-To-2D Space Mapping (Timeline)
- Line graph: ThemeRiver
- , radial-based layout: Spiral Graph
- , small multiples
- Time-To-3D Space Mapping
- Space-time cube
Time series data visualization

Time-To-Time Mapping

Time-To-Space Mapping

2D Space

3D Space

Animation

Linear representation

Heat maps

Circular design

Tree-like diagrams

Layers

Space-time cube

Secrier and Schneider. Visualizing time-related data in biology, a review. Briefings in bioinformatics (2013)
Secrier and Schneider. Visualizing time-related data in biology, a review. Briefings in bioinformatics (2013)
Representing time at different levels

• At the molecular level
• **At the gene level: linear methods and heat maps**
• **At the network level: animation**
• At the cellular level
• At the organismal level
• At the population level
• At the evolutionary scales
Spatial and non-spatial data visualization

- **Spatial**
  - Geospatial
  - Conceptual
    - Gene expression
    - Protein structures
    - Networks
    - ...

- **Non-spatial**
  - Abstract
    - Probability value
    - Cluster
    - ...

- **Multiple and coordinated views**
Multiple and coordinated views


Brain Viewer. [http://gallantlab.org/brainviewer/huthetal2012/]
Case Study 1: Dynamic mouse brain networks

• Problem

We collaborate with domain scientists from neuroscience and computational biology who use the approach of dynamic network analysis to explore the change in functional connections and community identities over time within the mouse brain.

A dynamic community is defined as a time-series of sets of neurons that have similar functional behaviors.
Data Processing

A time series of neuron activity in the mouse brain

A time series of correlation networks

Dynamic communities
CommDy

CommDy is a method of detecting dynamic communities.

Two community identification codes:
• **Home Community** identifying the community that the neuron belongs to;
• **Temporary Community** identifying the community that the neuron currently visits.
Example for CommDy
Data features

• Time: ~1,000 time steps

• Spatial data: the coordinates of nodes (a set of neurons)

• Non-spatial data: pixel intensity, node degree, community identifications, network size, etc.
Visual encodings

- Time: linear representations and animation
- Spatial data: 2D mapping (map the nodes onto the brain slice images)
- Linked views
# SwordPlots

<table>
<thead>
<tr>
<th>Sword Parts</th>
<th>Sword Pommel</th>
<th>Sword Body</th>
<th>Upper Cross-guard</th>
<th>Lower Cross-guard</th>
<th>Sword Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation</td>
<td>Current Time Status</td>
<td>Raw pixel's value</td>
<td>Temporary Community</td>
<td>Home Community</td>
<td>——</td>
</tr>
<tr>
<td>Visual Encoding</td>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
<td><img src="image3.png" alt="Diagram" /></td>
<td><img src="image4.png" alt="Diagram" /></td>
<td><img src="image5.png" alt="Diagram" /></td>
</tr>
<tr>
<td>Interaction</td>
<td>change location</td>
<td>——</td>
<td>——</td>
<td>display detail panel</td>
<td>change size</td>
</tr>
</tbody>
</table>
Space attribute cube
Case Study 2: Probability distributions at states in the FFL network motif

• Problem
Our domain scientists develop numerical methods for the simulation of biochemical networks. They need help with either journaling a set of simulations or exploring the simulation itself through visualizations.

<table>
<thead>
<tr>
<th># of Protein 0</th>
<th>...</th>
<th># of Protein N</th>
<th>Copies of Gene 0</th>
<th>...</th>
<th>Copies of Gene M</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>...</td>
<td>30</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0.000001</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10</td>
<td>...</td>
<td>100</td>
<td>0</td>
<td>...</td>
<td>1</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\sum P = 1$</td>
</tr>
</tbody>
</table>
Tasks

• Find the number of probability peaks
• Find the locations of peaks
• Describe the shapes of peaks
• Track how the peaks change over time
• Track how the peaks change over different system settings
• Comparison
• ...
Data features

• Time: 2,000 ~ 20,000 time steps
• Spatial data: corresponding states
• Non-spatial data: probability values, copies of genes, etc.
Visual encodings

• Time: heat maps OR spaghetti plots (considered as ensembles)
• Spatial data: radar charts
Multiple radar charts
Selected radar charts with time
1D projection and half-radar charts
Spaghetti plots for time