A Negative Control Study

On Network Randomization Methods

2012 NSF Bio-Grid (University of Connecticut)
REU Research Fellow
University of Connecticut
Max Espinoza (Fairfield University 2013)
Saad Quadar
Dr. Chun-His Huang
Mathematical Abstraction

- Graphs

\[ G = (V, E) \]

- \( V \) is a set of vertices/nodes
- \( E \) is a set of edges
Graphs

UNDIRECTED GRAPH

DIRECTED GRAPH
They are ubiquitous

In scientific disciplines

In everyday life

In industry
Internet Graph
Yeast PPI Network
Transcription Factor Networks
Social Networks
Negative Control

Scientific Control

• Positive Control
  • Phenomenon is expected
  • Effect when there should be an effect

• **Negative** Control
  • No phenomenon is expected
  • No effect when there *should* be no effect
Network Motifs

- Are hypothesized to be simple building blocks of complex networks

- Are defined as statistically over represented subgraphs in a given network

- Cross disciplinary motifs exploration
Negative Control and Motif Exploration

- Motif Detection Software
  - Internally counts all subgraphs frequencies
  - Determines statistical significance
    - Uses negative control in the form of a set of randomized graphs
Randomized Graphs

A) real network

B) randomized networks

motif:
Graph Randomization Algorithms

“Switching” Method

“Matching” Method
Switching Method

- 2 edges randomly chosen to be swapped
- Monte Carlo Markov Chain
- No Theoretical Bound on Switching
Matching Method

- Vertex assigned set of “stubs”
- In and out stubs are picked randomly to create network edges
- Generally generates biased sample of possible networks
Mfinder 1.2

- Weizmann Institute of Science (2005)
- Network Centric
- Implements both randomization methods needed
Input networks

E. Coli Transcription Network

Yeast Transcription Network
Methodology

- Compare motifs detected using both randomization algorithms
- Increase data reliability with multiple trials
## Data Collection

<table>
<thead>
<tr>
<th>Node</th>
<th>E. Coli TF</th>
<th>Yeast TF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Switching</td>
<td>Matching</td>
</tr>
<tr>
<td>3 Node</td>
<td>6 trials</td>
<td>6 trials</td>
</tr>
<tr>
<td>4 Node</td>
<td>6 trials</td>
<td>6 trials</td>
</tr>
<tr>
<td>5 Node</td>
<td>6 trials</td>
<td>6 trials</td>
</tr>
</tbody>
</table>
Motifs Found

Motif Size vs Motifs Detected

Avg Num of Motifs Found

- 3 Node
- 4 Node
- 5 Node

E. Coli  Yeast

- E. Coli
- Yeast
## Runtime Data (mfinder 1.2)

<table>
<thead>
<tr>
<th>Node</th>
<th>E. Coli TF (Runtime in Min)</th>
<th>Yeast TF (Runtime in Min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Switching</td>
<td>Matching</td>
</tr>
<tr>
<td>3 Node</td>
<td>2.41</td>
<td>0.75</td>
</tr>
<tr>
<td>4 Node</td>
<td>19.3</td>
<td>17.18</td>
</tr>
<tr>
<td>5 Node</td>
<td>1132.50</td>
<td>1009.10</td>
</tr>
</tbody>
</table>
3 Node Data

- E. Coli
  - Had a feed forward loop in all occurrences
  - Motif ID: 38

- Yeast
  - Had no motifs
Motif IDs

- Feed Forward Loop ID: 38

38 → 100110 → 000100110 → 000|100|110
   → 011|001|000

```
0 1 1
0 0 1
0 0 0
```
## 4 Node Data

### E. COLI (TF) NETWORK

<table>
<thead>
<tr>
<th>Matching Algorithm (Motif IDs)</th>
<th>Switching Algorithm (Motif IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>204</td>
<td>204</td>
</tr>
<tr>
<td>206</td>
<td>206</td>
</tr>
<tr>
<td>2186</td>
<td>2186</td>
</tr>
</tbody>
</table>

### YEAST (TF) NETWORK

<table>
<thead>
<tr>
<th>Matching Algorithm (Motif IDs)</th>
<th>Switching Algorithm (Motif IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>92</td>
<td>92</td>
</tr>
<tr>
<td>204</td>
<td>204</td>
</tr>
<tr>
<td>206</td>
<td>206</td>
</tr>
<tr>
<td>2186</td>
<td>2186</td>
</tr>
</tbody>
</table>
E. COLI (TF) NETWORK

- Motif (ID: 188) appeared 5/6 trials using the switching method
  - Appeared only once with the matching method

YEAST (TF) NETWORK

- Motif (ID: 8730) appeared 2/6 trials using the switching method
  - Appeared no times with the matching method
Difficulties

- Exponential Time Increase
- Limited amount of resources
Importance of Reasearch

- Lay the groundwork for further investigation of randomized graphs as negative control

- Let's use ask valuable questions in as to why some motifs are detected while others are not
  - Could the sampling bias in the matching algorithm effect this
Future Research

- 6 + Node Motifs
- Additional Graph Randomization Algorithms
- Alternate Parameters
Citations

2. Structure of yeast protein interactions (nodes - proteins, edges - reactions) (Barabasi et. al., 2003)