Identification of optimal parameter ranges in building and assessing correlation networks built from gene expression.

Qianran Li, Kate Cooper
College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182

Objective
- Through analysis of network parameters, identify if there are differences among tissues in mice.

Method
- Get data from Gene Expression Omnibus (GEO).
- Build the network on their correlation coefficient.
- Identify optimal parameter ranges.

Data Sources
- Free online sources from GEO in platforms GPL1261
  - 45,101 Genes
  - 5 series (tissues)
  - 43 Networks

Process
- Get data from GEO
- Build the network
  - Using gene as node, edge is defined as correlation coefficient between each pair genes.
  - Select edges by conditions of correlation coefficient range from 0.7 to 1.0 and p-value less or equal than 0.0005.
- Analyze the networks individually
  - Get number of nodes
  - Get number of edges
  - Build degree distribution
  - Determine if it is scale-free network by Kolmogorov–Smirnov test
    - Generate a “random” scale-free network from Barabási-Albert model as a standard network
    - Using two sample Kolmogorov–Smirnov test to determine the distance between standard network and sample network and the p-value of the test.
  - Calculate the assortativity degree
  - Calculate the clustering coefficient.
- Identify parameter ranges of each series

Results
- Output1: list of 43 networks.
<table>
<thead>
<tr>
<th>Series</th>
<th>Networks</th>
<th>Description</th>
<th>Samples per Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE1999</td>
<td>2</td>
<td>Neuroprotective effects of erythropoietin</td>
<td>5</td>
</tr>
<tr>
<td>GSE26299</td>
<td>11</td>
<td>Gene expression profiling in DBA/2J glaucoma</td>
<td>10</td>
</tr>
<tr>
<td>GSE27563</td>
<td>4</td>
<td>murine PBMCs from mice with advanced mammary tumors and their tumor-free counterparts</td>
<td>14</td>
</tr>
<tr>
<td>GSE6514</td>
<td>18</td>
<td>mouse brain during spontaneous sleep and prolonged wakefulness</td>
<td>5</td>
</tr>
<tr>
<td>GSE12413</td>
<td>8</td>
<td>Prediction of left ventricle systolic dysfunction in mice</td>
<td>10</td>
</tr>
</tbody>
</table>
- Output2: network file

<table>
<thead>
<tr>
<th>Gene A</th>
<th>Correlation Coefficient</th>
<th>Gene B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1425694 at 0.999750664743794 1425616 at</td>
<td>2 1425694 at 0.99771724236767 1426251 at</td>
<td></td>
</tr>
<tr>
<td>3 1425694 at 0.994355837874856 1427931 at</td>
<td>4 1425694 at 0.994355837874856 1427931 at</td>
<td></td>
</tr>
<tr>
<td>5 1425694 at 0.993422952717746 1427734 at</td>
<td>6 1425694 at 0.993422952717746 1427734 at</td>
<td></td>
</tr>
<tr>
<td>7 1425694 at 0.99440881856098 1428589 at</td>
<td>8 1425694 at 0.99440881856098 1428589 at</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions
- Degree Distribution of networks
  - The two graphs below show the degree distributions of 8 networks as example, each network has a plot graph of their degree distribution.
- Assortativity
- Clustering Coefficient

References:

Acknowledgements: This work was supported by the National Institute of General Medical Sciences (NIGMS) of the U.S. National Institutes of Health (NIH) under award number R01GM119791 to D.N. Additional support was provided by the National Heart, Lung, and Blood Institute of the NIH under award number R01HL128820 to D.N.