Integrating multiple heterogeneous networks for novel IncRNA-disease association inference

Presenter: Jingpu Zhang
Supervised by: Lei Deng
Accumulating evidences have demonstrated the important associations between IncRNAs and a broad range of complex human diseases.

Some computational methods have been developed to predict novel IncRNA-disease associations.

Previous methods don’t allow to consider networks with more than two types of biological entities.

We predict novel IncRNA-disease associations by integrating multiple heterogeneous biological entities.
Pipeline

A

\[ \text{IncRNADisease} \rightarrow \text{MeSH} \rightarrow \text{OMIM} \rightarrow \text{HPRD} \]

\[ \text{LncRNA similarity network} \rightarrow \text{Disease similarity network} \rightarrow \text{Protein interaction network} \]

\[ \text{link} \rightarrow \text{NPinter} \rightarrow \text{OMIM} \]

Construct global network

B

infer potential IncRNA-disease associations

C

\[ \text{IncRNA} \rightarrow \text{disease} \rightarrow \text{protein} \]
Our problem is to determine how related the query set and the target set are based on known relations between elements.
Two kinds of networks

a) networks representing interactions or similarities between entities of the same type.
b) networks representing interactions or similarities between entities of different type.

Type b networks are used to interconnect type a networks.
Value propagation inside networks

$F$ is both smooth over the network and also respects the prior knowledge

\[
F(v) = a \left[ \sum_{u \in N(v)} F(u) w'(v, u) \right] + (1 - a) Y(v)
\]

\[
F^i = a W' F^{i-1} + (1 - a) Y
\]

$w'$ is a normalized matrix whose values are given by the adjacency matrix of the network

$Y$ represents a prior knowledge function

$N(v)$ denotes the direct neighborhood of $v$
Value propagation between networks

\[ \Psi(v) = \frac{\sum_{x \in \text{neig}(v)} \Psi(x)}{|	ext{neig}(v)|} \]

\text{neig}(v) \text{ is the set of nodes from the current network which are connected with node } v \text{ in the following network.}
- The nodes in the networks which are adjacent to the target network present values that determine their degree of relationship to the query set.
- Also, the nodes in the target network are assigned a value that determines the degree of relationship with the target set.
- We can indirectly measure the relationship between the query set and the target set by measuring the similarity between the values of the nodes in the target network and those that are directly connected to them in adjacent networks.
For each path $p_i$ with length $l$, a vector $y_i$ is computed as:

$$y_i = S_{(l-1)l} \hat{y}_{(l-1)i}$$

where $S_{(l-1)l}$ is the normalized adjacency matrix of $N_{(l-1)(l)}$ and $\hat{y}_{(l-1)i}$ is the vector obtained after propagating values inside the network $N_{(l-1)}$.

$$y = concat(y_i) \quad \forall i \in [1, |P|]$$
the target network after the propagation process are represented by \( \hat{y}_t \)

\[
\hat{y}_t = \text{concat}(\hat{y}_t, L \hat{y}_t)
\]

the correlation value which derives a measure for the relationship between the query and target sets is computed as:

\[
s = \text{corr}(y, t)
\]
For a given disease regarded as a query set in the query network, we compute the correlation value between the disease and each IncRNA, i.e. a target set, utilizing the above method in turn. The all IncRNAs in the target network are sorted by the correlation values in decreasing order after the values are obtained.
Results

The benefit of protein interaction network

LOOCV

5-fold CV
The benefit of protein interaction network

<table>
<thead>
<tr>
<th>Disease-cutoff</th>
<th>IncRNA-cutoff</th>
<th>LRDNetFlow-3N</th>
<th>LRDNetFlow-2N</th>
</tr>
</thead>
<tbody>
<tr>
<td>5nn</td>
<td>0.5</td>
<td>0.906</td>
<td>0.894</td>
</tr>
<tr>
<td>5nn</td>
<td>0.6</td>
<td>0.906</td>
<td>0.894</td>
</tr>
<tr>
<td>5nn</td>
<td>0.1(5nn)</td>
<td>0.849</td>
<td>0.830</td>
</tr>
<tr>
<td>5nn</td>
<td>0.2(5nn)</td>
<td>0.841</td>
<td>0.821</td>
</tr>
<tr>
<td>5nn</td>
<td>0.3(5nn)</td>
<td>0.817</td>
<td>0.790</td>
</tr>
<tr>
<td>0.3</td>
<td>0.5</td>
<td>0.898</td>
<td>0.889</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
<td>0.896</td>
<td>0.883</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1(5nn)</td>
<td>0.835</td>
<td>0.818</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2(5nn)</td>
<td>0.829</td>
<td>0.814</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3(5nn)</td>
<td>0.806</td>
<td>0.786</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
<td>0.902</td>
<td>0.891</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>0.900</td>
<td>0.886</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1(5nn)</td>
<td>0.844</td>
<td>0.825</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2(5nn)</td>
<td>0.836</td>
<td>0.816</td>
</tr>
<tr>
<td>0.4</td>
<td>0.3(5nn)</td>
<td>0.811</td>
<td>0.787</td>
</tr>
</tbody>
</table>
Results

Comparison with state-of-the-art methods

LOOCV

5-fold CV
Results

AUC value

- LncRDNetFlow-3N
- LncRDNetFlow-2N
- KATZLDA
- RWRHLD

Different pre-processing and cutoffs of lncRNA and disease score

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

AUC value

0.4 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95
Thank you