Topological ranks reveal functional knowledge encoded in biological networks: a comparative analysis

M. Bonomo, D. Greco, R. Giancarlo, S. E. Rombo

Briefings in Bioinformatics, 2022
Biological Graphs

Widely used to model many biological problems
Molecular Interactions
Genotype Phenotype Associations
To what extent function may be retrieved from the graph encoded topology
Views and Node/Edge Ranking

Monuments in Paris
Views and Node/Edge Ranking

Level 1
The most famous monuments in Paris
Views and Node/Edge Ranking

Level 2
The most famous monuments in Paris and other important monuments
Views and Node/Edge Ranking

... Level k
All monuments in Paris
Topological Measures

Assign a real weight to nodes or edges
(Topological Overlap Measure, Edge Clustering Value, Dispersion, Edge Betweenness, Clustering Coefficient, Eigenvector Centrality, Subgraph Centrality, κ-Path Centrality...)

Edge Rank

An ordered list of subsets of nodes/edges such that they are a partition of all nodes/edges

Provide different-levels graph views
Topological Views

Incremental Views

Display the subgraphs induced by incrementally considering sets in the edge rank, according to the priority, i.e., ‘relevance’, of the edges given by the rank.

Decremental Views

Remove edges according to the edge rank, such that the priority of the rank indicates irrelevance.
Static vs dynamic edge rank

Static edge rank

In one pass, the value of the input topological measure is computed for each of the edges of the input graph.

Dynamic edge rank

At each step, edges with the highest score are appended to the partial solution and deleted from the graph. This allows to intercept edges with an important role in their topological context, yet hidden by other edges scoring much higher values.
Datasets and Gold Standards

Homo sapiens

- Human Gene Network
  Gene mutations associated with the occurrence/progress of diseases
- Human Disease Network

Caenorhabditis elegans

- Worm Gene Network
  A gene is connected to a defect if its inhibition (via breakdown experiments) is involved in the development of the defective phenotype

Saccaromices cerevisiae

- PPI Networks
  Some obtained by filtering networks from the literature to delete unreliable interactions, one built upon high-throughput yeast two-hybrid screening

# of commons:

- GS1 - diseases implied by single-nucleotide polymorphisms (SNPs)
- GS2 - GO terms between two genes
- GS3 - SNPs between diseases
- GS4 - defects of phenotype between genes
- GS5/GS6/GS7 - biological processes between proteins
Comparison of different ranks

K-haus

Distance function on partial ranks which counts the number of inversions in two ranks, excluding ties. It is normalized so that it has value in $[0, 1]$, where zero indicates identity.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>.1 Total Random</strong></td>
<td>Null model: A random permutation of the edges of the network is generated</td>
</tr>
<tr>
<td><strong>.2 Equal Classes</strong></td>
<td>Null model: each class of edges is assigned the same number of edges it has, but this time chosen randomly, without replacement, from the set of edges of the network</td>
</tr>
</tbody>
</table>

Statistical significance of a rank comparison
Performance and statistical significance
Best performing measures

- Clustering Coefficient
  - H. sapiens: ECC3
  - C. elegans: ECC3
  - S. cerevisiae: ECC3

- Neighborhoods
  - H. sapiens: GTOM2, TOM
  - C. elegans: EB
  - S. cerevisiae: EB

- Modularity
  - H. sapiens: EB
  - C. elegans: EB
  - S. cerevisiae: KB1, KB2, KB3

- Dispersion
Thank You