An Empirical Investigation of Statistical Backbone Filtering Techniques
Ali Yassin, Hocine Cherifi, Hamida Seba, Olivier Togni

To cite this version:
Ali Yassin, Hocine Cherifi, Hamida Seba, Olivier Togni. An Empirical Investigation of Statistical Backbone Filtering Techniques. Conference on Complex Systems, Oct 2022, Palma de Mallorca, Spain. hal-04321035

HAL Id: hal-04321035
https://hal.science/hal-04321035
Submitted on 4 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
An Empirical Investigation of Statistical Backbone Filtering Techniques

Ali Yassin
University of Burgundy, FRANCE
ali_yassin@etu.u-bourgogne.fr

Hocine Cherifi
University of Burgundy, FRANCE
hocine.cherifi@u-bourgogne.fr

Hamida Seba
University of Lyon, FRANCE
hamida.seba@univ-lyon1.fr

Olivier Togni
University of Burgundy, FRANCE
olivier.togni@u-bourgogne.fr

This work is funded by Agence Nationale de Recherche under grant ANR-20-CE23-0002

What is a Network Backbone?
Filtering Noise?
Preserving Topological Properties?

No Universal Definition!

Types of Backbone Filtering Techniques
1. Structural Backbone Filtering Techniques.
2. Statistical Backbone Filtering Techniques.

Statistical Backbone Filtering Techniques

- **Noise Corrected Filter**
  - edge weights are drawn from a binomial distribution. However, it estimates the probability of observing a weight connecting two nodes using a Bayesian framework.

- **Marginal Likelihood Filter**
  - each unit edge randomly chooses two nodes, which results in a binomial distribution.

- **ECM Filter**
  - Employs a null model based on the canonical maximum-entropy ensemble of weighted networks having the same degree and strength distribution as the original network.
  - Employs the empirical cumulative density function to judge statistical significance.

- **LANS Filter**
  - a node distributes its strength (summation of weights) uniformly on the edges.

- **Disparity Filter**
  - a node distributes its strength at random through a Polya process, which results in a Beta-binomial distribution.
  - edge weights are randomly drawn from the empirical weight distribution.

- **Polya Urn Filter**
  - a node distributes its strength (summation of weights) uniformly on the edges of the network.

- **GloSS Filter**
  - edge weights are drawn from the empirical weight distribution.

Evaluation Using Four Indicators

<table>
<thead>
<tr>
<th>Filter</th>
<th>Technique Behavior</th>
<th>Giant Component</th>
<th>Isolated Nodes</th>
<th>Weights Retained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Likelihood</td>
<td>Conservative</td>
<td>Always</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Noise Corrected</td>
<td>Conservative</td>
<td>Always</td>
<td>Low</td>
<td>Small &amp; High</td>
</tr>
<tr>
<td>ECM</td>
<td>Flexible</td>
<td>$\alpha &gt; 10^{-2}$</td>
<td>$\alpha &lt; 10^{-2}$</td>
<td>Small &amp; High</td>
</tr>
<tr>
<td>LANS</td>
<td>Aggressive</td>
<td>Always</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Disparity</td>
<td>Aggressive</td>
<td>$\alpha &gt; 10^{-2}$</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Polya Urn</td>
<td>Aggressive</td>
<td>$\alpha &gt; 10^{-2}$</td>
<td>High</td>
<td>Small &amp; High</td>
</tr>
<tr>
<td>GloSS</td>
<td>Aggressive</td>
<td>$\alpha &gt; 10^{-2}$</td>
<td>High</td>
<td>Small &amp; High</td>
</tr>
</tbody>
</table>

Conclusion

Worldwide Air Transportation Network
Noise Corrected Filter
Marginal Likelihood Filter
ECM Filter
LANS Filter
Disparity Filter
Polya Urn Filter
GloSS Filter

What about the distributions?