# Comparing Link Filtering Backbone Techniques in Real-World Networks

Ali Yassin[1], Hocine Cherifi[2], Hamida Seba[3] and Olivier Togni[1]

[1]LIB EA 7534 - Univ. Bourgogne - Franche-Comté, Dijon, France
[2]ICB UMR 6303 CNRS - Univ. Bourgogne - Franche-Comté, Dijon, France
[3]LIRIS UMR 5205 - Univ Lyon, UCBL, CNRS, INSA Lyon, F-69622 Villeurbanne, France

Networks are valuable representations of complex systems. They can be analyzed for various purposes, such as identifying communities, influential nodes, and network formation. However, large networks can be computationally challenging. Multiple techniques have been developed to reduce the network size while keeping its main properties. One can distinguish two approaches to deal with this issue: 1) structural and 2) statistical methods. Structural techniques reduce the network while preserving a set of essential properties. In contrast, statistical techniques tend to filter nodes or links that blur the original network. They rely on a statistical hypothesis testing model or estimate to filter noisy edges or nodes.

In this study[1], we carry out a comprehensive comparison of seven statistical filtering techniques on a collection of 39 weighted real-world networks of various sizes (Number of nodes ranging from 18 from to 13,000) (number of links ranging from 78 to 5,574,233) and origins (character, web, biological, economic, infrastructural, and offline/online social). First, we investigate the similarities between the filtering techniques. Indeed, each link has an associated probability value (P-value), allowing us to compare the methods through correlation analysis. In a second set of experiments, we investigate the relationship between the basic local properties of the nodes and the underlying statistical model through the P-values. Then we turn to the global backbone properties. More precisely, we compare the weight distribution of the extracted backbones to that of the original network for a given significance level ($\alpha = 0.05$). Finally, we study the backbone's criticality. We iteratively remove edges in ascending order of their P-value from the original network and measure the size of its largest connected component (LCC). Fig 1 illustrates the results.

The first panel presents the mean Pearson correlation between pairs of filtering techniques across all networks. The couples (LANS, Disparity filter) and (Noise Corrected, ECM ) are well correlated (0.8). Conversely, the Polya Urn filter does not exhibit a noticeable correlation with any other filtering method. The second panel showcases the typical pattern of the cumulative weight distribution in the Fr-HS network. The distribution in the Polya Urn backbone closely matches the original network. However, it deviates noticeably from the Disparity, LANS, and GloSS filters. The other techniques are in between these two extremes. The third panel displays the typical pattern of network fragmentation in the Fr-HS network when one removes the top $X$ most significant edges using each filtering method. The ECM and Noise Corrected filters cause the network to break apart more rapidly than the other techniques. On the other hand, the Polya Urn and GloSS filters demonstrate a slower fragmentation rate. The remaining filtering techniques' fragmentation patterns fall between these two extremes. Removing the core edges from the network would hasten network fragmentation. Thus, an ideal backbone would hold these critical edges. Overall, the Polya Urn filter departs from its alternatives. Indeed, it is the only method that preserves the weight distribution. The Noise Corrected and ECM filters' backbones are more critical since they hold the network's binding edges. This work allows a deeper understanding of each filtering technique's similarities and unique properties.

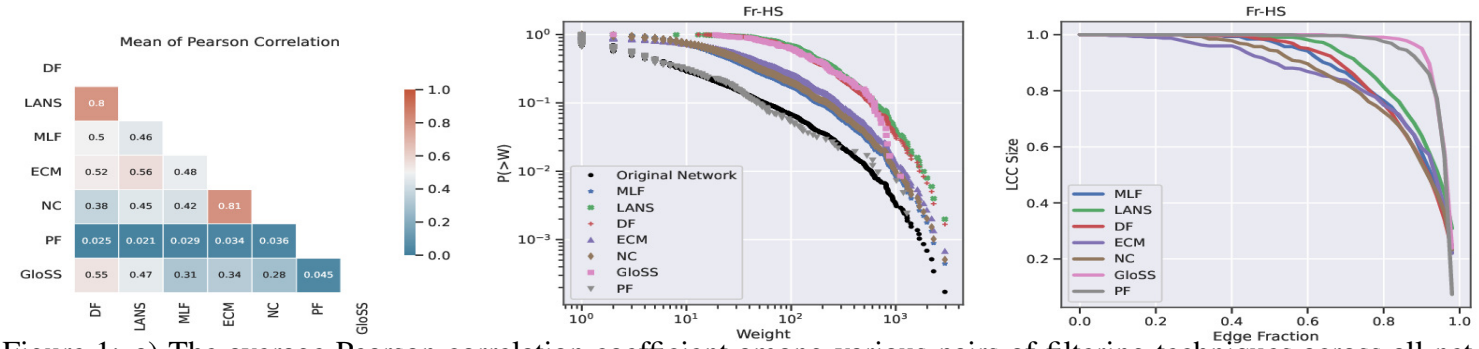


Figure 1: a) The average Pearson correlation coefficient among various pairs of filtering techniques across all networks. b) A typical pattern of the cumulative weight distribution in the Fr-HS network. c) A typical outcome of network fragmentation resulting from removing the top $X$ most significant edges in the Fr-HS network. MLF is Marginal Likelihood Filter. DF is: Disparity Filter, LANS: Local Adaptive Network Sparsification, PF is: Polya Urn Filter, NC: Noise Corrected Filter, and GloSS is: Global Statistical Significance Filter.

[1]Agence Nationale de Recherche funds this work under grant ANR-20-CE23-0002.