Effect of pruning on co-authorship network topology

P.A.D. Pasion, G.A. Tapang

National Institute of Physics, University of the Philippines, Diliman, Quezon City, Philippines

We applied two pruning schemes, the marginal likelihood filter (MLF) and the thresholding by bill author count, on the bill co-authorship networks of the Philippine Congress and determined the effect of each scheme on their topologies. MLF, proposed by Dianati, evaluates weighted edges by calculating their p-values based from their proposed null model. For the MLF, the higher the weighted degrees of the nodes are, the larger the edge weights should be for these edges to be significant. The thresholding by bill author count scheme deletes in-degree nodes of the bipartite origin of the network (bill-author network) based from an in-degree threshold value (author count of bills) as the maximum value to be included in creating the unipartite network. To characterize the network, the measures calculated were the node count N, density D, average weighted degree kw, weighted and unweighted average clustering coefficient C and Cw, connected component count, modularity Q and degree assortativiy R.

We found that both pruning schemes effectively increase Q and decrease C, D and kw as the number of retained edges decreases. In fact, their graphs for D are coinciding. They also preserve the connectivity of the network until 10 percent of the edges are retained. These suggest that at region with connected network, the networks are sparser and the qualities of community partition are better than the original graphs.

Comparing the two pruning schemes, subtle differences are found. In MLF, N is constant until around 6 percent of the edges are retained. Also, the networks become positively assortative as the edges retained decreases. Furthermore, Q increases smoothly and rapidly. In kw, it decreases linearly but the Cw is preserved.

In the thresholding by bill author count, a pattern of constant N at long irregular intervals and subtle decrease occurs repetitively until a sudden drop of N at 12 percent of retained edges. Unlike in MLF, the networks' dissassortativeness is preserved in this scheme. In Q, it increases but small second peaks are found. In kw, there is a distinct "elbow" found at around 90 percent of edges retained. The elbow in kw is seen in the Cw. The Cw in this scheme suddenly decreases and then stays constant at values very close to 0.

[l] N. Dianati, Phys. Rev. E **93 (1)**, 012304 (2016).

[2] P. Gerlee, L. Lizana, K. Sneppen, Bioinformatics **25**, 3282 (2009).

[3] S. Mengiste, A. Aertsen, A. Kumar, Scientific Reports **5**, 18145 (2015).