Mathematical Statistics
Stockholm University
Master Thesis **2023:9**
http://www.math.su.se

Pruning of complex networks in psychiatric symptomatology

Vera Andersson*

June 2023

Abstract

Psychiatric symptoms can be conceptualized as nodes in a directed network with edges representing the perceived causality between symptoms, and the visualizations can be useful for identification of interventions and diagnosis of symptoms and mental disorders. However, these networks are often large with complex structures, and we therefore aim to develop simplification methods that remove superfluous edges while maintaining significant structures. We propose three centrality based pruning methods that iteratively remove edges with the lowest edge betweenness, and a generalization of the "brute force" approach in Zhou, Malher Toivonen [23] for directed networks that allows for disconnected components, and this method prunes the edge retaining the connectivity the most in each step. All methods incorporate the node and edge weights, and the second and third centrality based methods compute the PageRank and updated PageRank centrality of all nodes, respectively. We evaluate the methods by comparing the simplifications of pairs of similar, empirical networks from a self-reported survey completed twice by Swedish teenagers having screened positive for depression, and by using synthetic networks with added noise to test reliability. We also let two sets of psychologists who know and do not know the patients visually evaluate the pruned results to select the most useful visualizations. The findings suggest that all methods have trouble with the empirical network pairs, but the basic edge betweenness approach and the non-updating PageRank approach maintain more similarity in general. However, all methods significantly increase the similarity between the synthetic and pruned noisy networks, especially when adding many edges with smaller weights. The psychologists overall prefer the method using connectivity, but those who know the patients prefer the simple edge betweenness approach and the PageRank approach. Therefore, selecting the best method is difficult and might need to be based on the tradeoff between practical usefulness, accuracy, and time complexity.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: vera.c.andersson@gmail.com. Supervisor: Tom Britton.