

Expression of Interest for Special Issue on Network Medicine in the era of Big Data in Science and Healthcare

Topic: Applications of Node-Based Resilience Graph Theoretic Framework to Clustering Autism Spectrum Disorders Phenotypes

Authors: John Matta^{*}, Thy Nguyen[†], Gunes Ercal^{*}, Tayo Obafemi-Ajayi[†]

^{*}Dept. of Computer Science, Southern Illinois University Edwardsville
Edwardsville, IL 62026, USA
(jmatta, gercal)@siue.edu

[†]Dept. of Computer Science/+Engineering Program, Missouri State University
Springfield, MO 65897, USA
Nguyen318@live.missouristate.edu, tayoobafemiajayi@missouristate.edu

Abstract:

With the growing ubiquity of data in network form, clustering in the context of a network, represented as a graph, has become increasingly important. Clustering is a very useful data exploratory machine learning tool that allows us to make better sense of heterogeneous data by grouping data with similar attributes based on some criteria. This paper investigates the application of a novel graph theoretic clustering method, Node-Based Resilience clustering algorithm (NBR-Clust), to address the heterogeneity of Autism Spectrum Disorder (ASD) and identify meaningful subgroups. The hypothesis is that analysis of these subgroups would reveal relevant biomarkers that would provide a better understanding of ASD phenotypic heterogeneity useful for further ASD studies. We address the appropriate graph constructions suited for representing the ASD phenotype data. The sample population is drawn from two large and rigorous datasets: Simon Simplex Collection (SSC) and the Autism Genetic Resource Exchange (AGRE). We provide a comparison of the results obtained from these two diverse datasets. Analysis of the results performed using internal cluster validation measures and clinical analysis outcome demonstrate the potential usefulness of resilience measure clustering for biomedical datasets.

(1) How this topic contributes to the special issue:

This topic is relevant to the special issue as it investigates application of complex network theory (node-based resilience measures) to addressing a medical problem: Autism Spectrum Disorders. The datasets to be utilized are very large > 2500 patients from the SSC database and >1500 from the AGRE database.

(2) Key concepts:

- addresses an open and increasingly relevant problem of sound and robust graph-theoretic clustering of noisy, complex data with applications to medical datasets.
- Systematic investigation of appropriate graph representations of biomedical data in relation to statistical properties of the type of data.

(3) Methods (if the paper will be based on empirical research)

- Node-Based Resilience Clustering Framework
- Varied statistical tests to confirm significance of clinical and behavioral results

(4) Expected conclusions and results.

We will apply graph theoretical clustering models to analyze a biomedical problem. The clustering models will categorize ASD patients into more meaningful homogenous groups. We will also apply the overlapping clustering algorithm in particular to determine notable overlap between the groups, and analyze consequences of such overlap. The results of this paper will be a significant contribution to understanding the biological mechanisms and clinical outcomes of children with ASD. A better understanding of ASD heterogeneity, based on scientifically rigorous approaches centered on systematic evaluation of the clinical and research utility of phenotypes will generate useful information for the study of the etiology and prognosis of the disorder. The goal is that the resulting clusters would provide a better understanding of ASD phenotypic heterogeneity and delineate subgroups useful for further ASD studies.