

# DSCC: DISEASE SUBTYPING USING COMMUNITY DETECTION FROM CONSENSUS NETWORK

Hung Nguyen\*, Monikrishna Roy\*, Ha Nguyen\*, Tin Nguyen\*

\*Department of Computer Science and Engineering, University of Nevada, Reno

## Abstract:

**Motivation:** Well-defined subtypes for a disease and condition can help develop precision treatment and personalized medicine, especially in cancer and complex diseases. With the abundance of genomic-scale data of different levels generated in recent years, disease subtyping has shifted from analyzing data of a single omic to integrating multi-omics data. Integrative analysis of multi-omics data can differentiate subtypes from a holistic perspective that reveals connections that otherwise cannot be detected using observations from a single data type. Current state-of-the-art methods for disease subtyping using multi-omics data are, however, can only efficiently handle a certain kind of omics (i.e., mRNA, miRNA, and methylation), and are unable to cluster data with different numbers of patients.

**Method:** We developed a robust disease subtyping method, Disease Subtyping using Community detection from Consensus network (DSCC), that (i) is capable of handling missing data where some patients have data for only a subset of omics (e.g., a patient has mRNA but not methylation), (ii) can integrate categorical and continuous data, and (iii) takes advantages of pathway information to improve disease subtyping results.

**Results:** We validated the method on 33 TCGA cancer datasets containing 10 data types with a total of 11,085 samples and clinical data for each patient. DSCC achieved the best results among six state-of-the-art multi-omics clustering methods in discovering novel subtypes with significantly different survival profiles.

**Future directions:** We will continue to test the method using data of different diseases. We will combine this method with other developed techniques in the context of biological networks, single-cell, genomics and epigenomics, and drug repurposing. The different omics are highly correlated, so future work is needed to deal with the correlation among various omics to get more accurate results.