

# RELATIONAL DATA CLUSTERING ALGORITHMS WITH BIOMEDICAL APPLICATIONS

Mohammed Khalilia  
Dr. Mihail Popescu, Dissertation Supervisor

## ABSTRACT

Relational data is not limited to social networks, but it has a wide presence in medical and biomedical domains. Relationships among patients can be measured based on their common diagnosis, medications, vital signs, etc. Such relational data is one of the key components to many machine learning algorithms such as clustering and prediction. Besides structured data, the electronic health records contain vast amounts of complex unstructured data, but mining such data comes with many challenges as I will demonstrate using nursing visit notes.

I will focus on two algorithms related to relational data clustering and analysis: (i) adapting the Kohonen Self-Organizing Maps (SOM) to relational data and incorporating a fuzzy membership function, which results in a new algorithm called Relational Fuzzy SOM (RFSOM); (ii) proposing a new technique to measure the topology preservation in RFSOM; and (iii) extending the well-known Relational Fuzzy  $c$ -Means (RFCM) to handle non-Euclidean relational datasets.

Furthermore, I will present applications based on various medical and biomedical datasets: (i) clustering patients described by activity of daily living trajectories extracted from the Minimum Data Set; (ii) predicting patient diagnosis based on patient current diseases using the National Inpatient Sample dataset; and (iii) mining nursing unstructured data, gathered from Senior Care services in Columbia, Missouri. The notes are parsed using natural language processing techniques aided by domain-specific ontology to extract care coordination activities and measure the care coordination dose.