

## Data Science Seminar Series

### “Geometric and Topological Graph Analysis for Machine Learning Applications”



#### **Tina Eliassi-Rad, Ph.D.**

**Professor**

**Khoury College of Computer Sciences  
Northeastern University**

**Date:** Wednesday, April 21st, 2021

**Time:** 4:00 PM – 5:00 PM EDT

**Location:** Zoom Virtual Room

**Web Link:** <https://njit-institute-for-data-science.eventbrite.com>

This talk has two parts: (1) geometric analysis for graph embedding and (2) topological analysis for graph distances. First, graph embedding seeks to build an accurate low-dimensional representation of a graph. This low-dimensional representation is then used for various downstream tasks such as link prediction. One popular approach is Laplacian Eigenmaps, which constructs a graph embedding based on the spectral properties of the Laplacian matrix of a graph. The intuition behind it, and many other embedding techniques, is that the embedding of a graph must respect node similarity: similar nodes must have embeddings that are close to one another. We dispose of this distance-minimization assumption. In its place, we use the Laplacian matrix to find an embedding with geometric properties (instead of spectral ones) by leveraging the simplex geometry of the graph. We introduce Geometric Laplacian Eigenmap Embedding (or GLEE for short) and demonstrate that it outperforms various other techniques (including Laplacian Eigenmaps) in the tasks of graph reconstruction and link prediction. This work is joint with Leo Torres and Kevin Chan, and was published in the Journal of Complex Networks in March 2020. Second, measuring graph distance is a fundamental task in graph mining. For graph distance, determining the structural dissimilarity between networks is an ill-defined problem, as there is no canonical way to compare two networks. Indeed, many of the existing approaches for network comparison differ in their heuristics, efficiency, interpretability, and theoretical soundness. Thus, having a notion of distance that is built on theoretically robust first principles and that is interpretable with respect to features ubiquitous in complex networks would allow for a meaningful comparison between different networks. We rely on the theory of the length spectrum function from algebraic topology, and its relationship to the non-backtracking cycles of a graph, in order to introduce the Non-Backtracking Spectral Distance (NBD) for measuring the distance between undirected, unweighted graphs. NBD is interpretable in terms of features of complex networks such as presence of hubs and triangles. We showcase the ability of NBD to discriminate between networks in both real and synthetic data sets. This work is joint with Leo Torres and Pablo Suarez-Serrato, and was published in the Journal of Applied Network Science in June 2019.

Tina Eliassi-Rad is a Professor of Computer Science at Northeastern University in Boston, MA. She is a core faculty member at Northeastern's Network Science Institute. Prior to joining Northeastern, Tina was an Associate Professor of Computer Science at Rutgers University; and before that she was a Member of Technical Staff and Principal Investigator at Lawrence Livermore National Laboratory. Tina earned her Ph.D. in Computer Sciences (with a minor in Mathematical Statistics) at the University of Wisconsin-Madison. Her research is at the intersection data mining, machine learning, and network science. She has over 100 peer-reviewed publications; and has given over 200 invited talks and 14 tutorials. Tina's work has been applied to personalized search on the World-Wide Web, statistical indices of large-scale scientific simulation data, fraud detection, mobile ad targeting, cyber situational awareness, and ethics in machine learning.