Geometric and Topological Graph Analysis for Machine Learning Applications

By Tina Eliassi-Rad

This course 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. We introduce Geometric Laplacian Eigenmap Embedding and demonstrate that it outperforms various other techniques in the tasks of graph reconstruction and link prediction. Second, measuring graph distance is a fundamental task in graph mining. I will present 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. Lastly, I will discuss recent work on non-backtracking eigenvalues under node removal. In network epidemiology, the reciprocal of the largest eigenvalue of the nonbacktracking matrix is a good approximation for the epidemic threshold of certain network dynamics. We develop techniques that identify which nodes have the largest impact on this leading eigenvalue. From this analysis we derive two new centrality measures: X-degree and X-nonbacktracking centrality. We perform extensive experimentation with targeted immunization strategies derived from these two centrality measures.