Spectral and Local Clustering for Hypergraphs

Abstract

This thesis advances the theory and algorithms for analyzing large-scale hypergraphs and temporal networks, focusing on spectral methods, diffusion processes, dynamic Personalized PageRank (PPR), and community detection.

First, we develop a spectral framework for non-uniform hypergraphs by extending normalized Laplacian theory and establishing Cheeger's inequality with provable guarantees. We show that the second smallest eigenvalue of the normalized Laplacian of a non-covering hypergraph is at most one, leading to improved and tight Cheeger inequalities. The framework also yields higher-order Cheeger inequalities, stronger guarantees for diffusion-based local clustering and spectral partitioning, and a new family of optimal hypergraph expanders that attain the Alon–Boppana bound.

Next, we present the first Lovász–Simonovits analysis of averaging-based diffusion on hypergraphs. We introduce Averaging-based Personalized PageRank for Hypergraphs (APPRH), whose linearity enables efficient computation using the Forward Push algorithm. Our local clustering algorithm, A-HyperCut, achieves stronger theoretical guarantees and is an order of magnitude faster than existing methods on million-scale hypergraphs while producing competitive cluster quality.

We then address the previously unexplored problem of maintaining Personalized PageRank on dynamic hypergraphs. Extending dynamic local push techniques to averaging-based hypergraph diffusion, we develop the first framework for efficiently updating PPR under hyperedge insertions and deletions without graph expansion, establishing its correctness and update complexity.

Finally, we investigate community stability in temporal networks through the concept of **core-invariant nodes**—vertices that remain in the *K*-core throughout a time window. We propose **Kwiq**, an efficient algorithm based on a novel orientation graph data structure, reducing computation substantially and achieving more than a fivefold speedup over baseline methods on large real-world temporal networks.

Together, these contributions establish new theoretical foundations and scalable algorithms for analyzing the structure, dynamics, and evolution of large-scale higher-order networks.