**Constructing Graph Structures from Datasets: A Graph Theory and Graph Signal Processing Approach**

**Introduction and Motivation**

Many real-world datasets inherently contain hidden structures that can be naturally represented as graphs. However, these structures are often not directly observable and must be inferred. Constructing an appropriate graph from a dataset unlocks the ability to apply powerful tools from **Graph Theory** and **Graph Signal Processing (GSP)** for tasks such as clustering, dimensionality reduction, signal denoising, and semi-supervised learning.  
This research aims to systematically explore methods to construct directed or undirected graphs from datasets, leveraging graph-theoretic principles and signal smoothness priors from GSP. By doing so, we bridge the gap between raw data and graph-based modeling, enhancing our ability to extract meaningful insights from complex data.

**2. Related Work**

Several research directions inform this work:

* **Graph Learning from Data**:  
  *Dong et al. (2016)* proposed methods to learn the graph Laplacian from smooth signal assumptions. Similarly, *Kalofolias (2016)* introduced sparsity-promoting graph learning models.
* **Graph Signal Processing**:  
  *Ortega et al. (2018)* provided a comprehensive overview of GSP, emphasizing how graph structures can be used to process and analyze signals that reside on irregular domains.
* **Classical Graph Construction**:  
  k-NN graphs, ε-neighborhood graphs, and fully connected graphs with weighted edges have long been used to approximate data manifolds (e.g., in manifold learning like Isomap and Laplacian Eigenmaps).
* **Topology Inference**:  
  More recently, topology inference models (e.g., *Egilmez et al., 2017*) have emerged, focusing on inferring graph structure via optimization formulations that fit observed data properties.

**3. Research Objectives**

* **Objective 1:** Develop systematic methods to map a given dataset into an optimal directed or undirected graph structure.
* **Objective 2:** Investigate the application of Graph Signal Processing tools (e.g., graph Fourier transforms, filtering) on the constructed graphs.
* **Objective 3:** Evaluate how different graph construction methods influence downstream tasks (e.g., clustering, prediction, denoising).

**4. Proposed Methods**

* **Graph Construction Techniques:**
  + Distance-based methods (e.g., k-NN, ε-graphs).
  + Data-driven graph learning using optimization (e.g., Laplacian learning).
  + Probabilistic and information-theoretic approaches (e.g., mutual information graphs).
* **Graph Theory Tools:**
  + Spectral properties (eigenvalues of Laplacians/adjacency matrices).
  + Centrality and connectivity measures.
  + Graph sparsification and robustness analysis.
* **Graph Signal Processing Tools:**
  + Signal smoothness measures across constructed graphs.
  + Graph Fourier analysis of dataset signals.
  + Designing graph filters for denoising and feature extraction.

**5. Potential Contributions**

* A unified framework for dataset-to-graph conversion based on theoretical foundations and empirical properties.
* Novel metrics to assess the "quality" of constructed graphs for signal processing tasks.
* New applications in fields such as sensor networks, biomedical data, and social network analysis.

**6. Expected Challenges**

* Balancing graph sparsity and connectivity.
* Sensitivity to noise and outliers in the dataset.
* Scalability to large datasets.