# Experiments

In the following experiments, the number $n$ of data points or graph vertices is no less than $1,000$ in at least two datasets/graphs.

1. Objective: feature-to-graph conversion.

   Provided a feature data (vectors/points) $X$ in a $d$-dimensional metric space, construct a near-neighbor graph $G(V, E)$ such that the $X$-to-$V$ mapping is one to one and that $E$ preserves spatial proximity to some extent.

   (a) Gather at least 2 feature datasets (such as color images or other types of feature points);

   (b) Construct a **rnn** graph (with the radius parameter $r$) or a **knn** graph ($k$ is the number of nearest neighbors;

   (c) Convert the pairwise distances to pairwise similarity scores with a chosen weight kernel function;

   Weight kernel examples: the Hinton-Roweis kernel function in SNE or t-SNE (2008) (the $P$-function, not the $Q$-function); the Shi-Malik kernel (1997)

   (d) Identify the types of the degree distribution and describe the variation in degree distribution as $r$ or $k$ changes;

   (e) `Optional.` Obtain and observe distributions of additional measures and their variation.

   (f) `Optional.` If the feature data are categorical, find or propose an approach to converting the categorical data into numerical.

2. Objective: vertex-to-vector encoding and embedding.

   Provided with a graph $G(V, E)$, weighted or unweighted, map the vertices in $V$ to a vector set $X$ in a metric space so that the pairwise adjacency is preserved in the pairwise distances, to some extent.

   (a) Use at least two graphs (directly available or converted from feature data), at least one is weighted

   (b) Spectral embedding space via a normalized Laplacian, with the dimension $d > 2$;

   (c) For a weighted graph, show the difference in pairwise distances (in matrix or in distribution);

   (d) Use the Fiedler vector of each graph to identify two dominant communities (two labels/colors) and the (high-centrality) cut edges between the two communities, which is a bipartite subgraph of $G$; Show the communities (in blue and red colors) in a 2D/3D scatter plot, and show the cut edges;

   (e) [Optional.] Apply the Fiedler cut to the subgraphs induced by the subcommunities; add the new labels/colors to the previous 2D/3D space;

   (f) [Optional.] Stochastic vertex embedding in a 2D/3D space.

3. Objective: empirical analysis of a digraph $G$ representing a real-world network, which may be a knn graph.

   (a) Preprocessing: describe the entities and relationship represented by $G$; find the number of (weakly) connected components; determine whether or not the largest connected component (LCC) is strongly connected;

   (b) Computationally get the Perron distribution $x_p$ of the LCC in two cases: case b-1: the LCC is strongly connected; case b-2: the LCC is not strongly connected (for instance, it has sink or source nodes). Use the BP-approach to get a variational, and conditional Perron distribution $x_p(\alpha_i, b_j)$, $\alpha_i \in (0.85, 1)$, $i = 1:4$, $b_j$ is a probing vector $b_j > 0$, $e^{\mathrm{T}} b_{=} 1$, $j = 1:5$. Use a figure or plot to describe the structure and structural variation with $\alpha$ and $b$.

   (c) [Optional to undergrads:] Apply the Laplacian embedding to a digraph.

4. [Optional to undergrads:] Differential description of a graph sequence $G_i(V_i, E_i)$, $i = 1:q$, $q > 1$. There are overlaps in the vertex sets: $V_i \cap V_j \neq \emptyset$. For example, a characteristic property of a BA graph is the growth; a WS graph can be obtained by a rewiring process, the vertex set does not change.

   Let $G(V, E)$ be the graph to serve as a global reference, $V = \cup V_i$, $E = \cup E_i$. Let $X$ be the vertex encoding (embedding, mapping) of $V$ in a 2D/3D spatial space. Show the difference and connection between $G_i$ and $G_{i+1}$ on the global spatial reference map.