Visualization
Prof. Bernhard Schmitzer, Uni Göttingen, summer term 2024

Problem sheet 8

- *Submission by **2024-06-17** 18:00 via StudIP **as a single PDF/ZIP**. Please combine all results into one PDF or archive. If you work in another format (markdown, jupyter notebooks), add a PDF converted version to your submission.*
- *Use Python 3 for the programming tasks as shown in the lecture. If you cannot install Python on your system, the GWDG jupyter server at https://jupyter-cloud.gwdg.de/ might help. Your submission should contain the final images as well as the code that was used to generate them.*
- *Work in groups of up to three. Clearly indicate names and enrollment numbers of all group members at the beginning of the submission.*

Exercise 8.1: intrinsic dimension and spectral embedding with the Laplace operator.
In this problem we study in more detail the apparent intrinsic dimension of point data with the spectral embedding via the Laplace operator.

1. First we generate an example dataset. Let $n_1 = 25$, $n_2 = 20$. Let X be a set of $n_1 \cdot n_2$ points in $\mathbb{R}^2$, lying on a regular two-dimensional Cartesian grid with n_1 and n_2 points along the two axes, with distance 1 between points along each axis. So X should be an array of shape $(n_1 \cdot n_2) \times 2$. Generate X and c´ompute the matrix D of squared Euclidean distances between the points in X. D should be of shape $(n_1 \cdot n_2) \times (n_1 \cdot n_2)$.
2. In addition, create an array $c \in \mathbb{R}^{(n_1 \cdot n_2) \times 3}$, such that c`[i,:]` $\in \mathbb{R}^3$ is an RGB color code that encodes the position of point X`[i,:]` $\in \mathbb{R}^2$ in the grid. The exact choice of the encoding is up to you.
3. Based on the squared distances and the length scale parameter $\varepsilon = 1$, compute now the matrix A, defined as follows: $A_{i,j} = \exp(-D_{i,j}/\varepsilon^2)$ for $i \neq j$, $A_{i,i} = 0$. Then from A generate the graph Laplacian L and its eigendecomposition, as in the lecture.
4. Let `eigvec` be the list of eigenvectors of L, sorted by increasing eigenvalue. Show a spectral embedding of X as a scatter plot of `eigvec[`k_1`]` versus `eigvec[`k_2`]`, $k_1 = 1$, $k_2 = 2$, and use the color codes c to encode the original grid structure. You should find that this indeed recovers approximately the original two-dimensional grid structure.
5. Now set $n_1 = 100$, $n_2 = 10$ and re-run the above code. $k_1 = 1$, $k_2 = 2$ now does no longer recover the two-dimensional grid structure. Find the best choices of k_1 and k_2 that do.

Exercise 8.2: Visualizing a relational database as decorated graph.
Consider a simple relational database that represents and online newspaper. *Journalists* can author *articles* (for simplicity, each article will be written by precisely one author) and articles can be *assigned* to (multiple) *categories*. *Readers* can create accounts and write *comments* on articles, they can express *reactions* to other readers' comments (such as 'agree' or 'disagree'),

and they can *follow* certain authors (to be automatically informed, when they publish a new article). Assume that each of the concepts above that were highlighted in *italics* is represented by a separate table, and that relations between the concepts is encoded by simple key/foreign key references.

1. For each of the above tables, except for the *follow* and *assignment* tables, list at least two examples of columns that these tables should have (beyond keys and foreign keys).

2. Draw a graph that represents the above database, in particular the table columns and reference relations. Possibly you can choose a separate visual representation for the relations encoded by the auxiliary tables *follow* and *assigned.* You can do this in any software you want, or with a simple hand drawing (scanned or on a tablet).