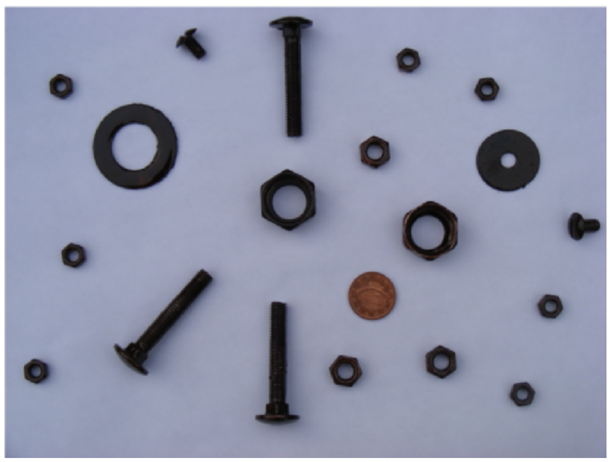
The 2-hour written exam for *CS4103 Geometric Foundations in Data Analysis II* will consist of four questions: all four from Graham. Students will be required to attempt all questions.

Students registered for *MA500 Geometric Foundations of Data Analysis* will take both the CS4102 paper and the CS4103 paper.

The following are examples of the kinds of things that Graham could ask.

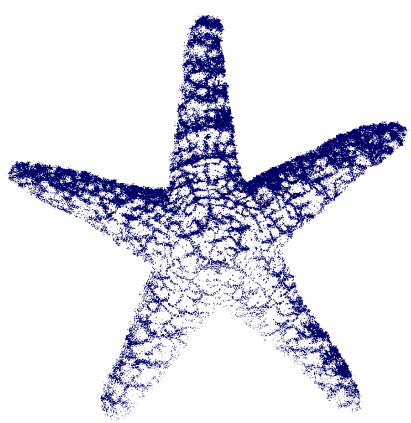
1 Persistence and digital image analysis

1. Explain how cluster analysis and barcodes could be used to estimate the number of objects in a digital photograph such as the following.



   Explain how cluster analysis could also be used to estimate the number of objects with holes.

2. Explain how cluster analysis and barcodes could be used to estimate the number of 'limbs' of an object such as a starfish from a digital image of the object.



2 Simplicial complexes, clique complexes, and Mapper clustering

1. Define what is meant by a *geometric k-simplex*, a *simplicial complex K*, and the *geometric realization* $|K|$ of a finite simplicial complex K. Illustrate your answers with a **simple** example.

2. Which of the following K are simplicial complexes? For those that are, sketch the simplicial complex, and determine the Euler characteristic

$$\chi(K) = \alpha_0 - \alpha_1 + \alpha_2 - \alpha_3 + \cdots$$

where α_k denotes the number of k-simplexes.

   (a) $V = \{1,2,3,4,5,6,7,8\}, K = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}, \{1,2\},\{1,6\},\{2,3\}, \{2,4\},\{2,5\},\{3,4\},\{3,5\},\{4,5\},\{4,6\},\{5,6\}, \{2,3,4\},\{2,3,5\},\{2,4,5\},\{3,4,5\}\}$.

   (b) $V = \{1,2,3,4,5,6,7,8\}, K = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\}, \{1,2\},\{1,6\},\{2,3\}, \{2,4\},\{1,5\},\{3,4\},\{3,5\},\{4,5\},\{4,6\},\{5,6\}, \{2,3,4\},\{2,3,5\},\{2,4,5\},\{3,4,5\}\}$.

3. Explain how, for each $\epsilon \geq 0$, one can associate a clique complex K_ϵ to a symmetric $n \times n$ matrix of distances between n items. For $\epsilon = 2.5$ and for the following 6×6 matrix of distances

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 3 & 3 \\ 1 & 0 & 1 & 3 & 3 & 3 \\ 2 & 1 & 0 & 1 & 3 & 3 \\ 3 & 3 & 1 & 0 & 1 & 2 \\ 3 & 3 & 3 & 1 & 0 & 1 \\ 3 & 3 & 3 & 2 & 1 & 0 \end{pmatrix}$$

determine the simplicial complex K_ϵ; then sketch the geometric realization $|K_\epsilon|$ and calculate the Euler characteristic $\chi(K_\epsilon)$.

4. Consider the collection $\mathcal{U} = \{\{1,2,3\},\{2,3,4\},\{3,4,5\},\{4.6\},\{1,6\}\}$ of sets. Sketch the nerve $N\mathcal{U}$ and calculate its Euler characteristic $\chi(N\mathcal{U})$.

5. Define what is means for two maps to be *homotopic* and for two spaces to be *homotopy equivalent.*

6. Let $Y \subset \mathbb{E}^n$ be an arbitrary convex subset of Euclidean space and let X be an arbirary topological space. Prove that any two continuous maps $f, g \colon X \to Y$ are homotopy equivalent.

7. Prove that any convex subspace $Y \subset \mathbb{E}^n$ is homotopy equivalent to the space consisting of a single point.

8. Prove that homotopy equivalence of maps $f \simeq g$ is an equivalence relation on the set of continuous maps $X \to Y$ from a given space X to a given space Y.

9. Give a careful proof that $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ is homotopy equivalent to $\mathbb{C} \setminus \{0\}$.

10. State Leray's theorem about the nerve $N\mathcal{U}$ of an open cover $\mathcal{U}$ of a space X. Illustrate the theorem by using an appropriate open cover of the annulus $X = \{z \in \mathbb{C} : 1 \leq |z| \leq 2\}$. Calculate the Euler characteristic $N\mathcal{U}$ in your illustration.

11. Give an account of the Mapper clustering algorithm. In your account, illustrate the algorithm on a set of points chosen from an annulus.

12. Give an example of a finite data set $S \subset \mathbb{R}^3$ for which 'topological information' will be lost under any linear projection $\mathbb{R}^3 \to \mathbb{R}^2$, but for which this information might be preserved under a map $S \to K$ to a 2-dimensional simplicial complex K produced from the Mapper clustering procedure.

3 Homology and Persistent Homology

1. Consider the mod-2 chain complex C_*K of the simplicial complex with $V = \{1, 2, 3, 4\}$, $K = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3\}\}$.

   (a) Determine the matrices for the linear homomorphisms $d_2 \colon C_1K \to C_0K$ and $d_1 \colon C_2K \to C_1K$.

   (b) Use these matrices to compute $H_0(C_*K)$ and $H_1(C_*K)$.

2. Consider the simplicial complex with $V = \{1, 2, 3, 4, 5, 6\}$, and with K consisting of the six 2-simplices $\{1, 2, 3\}, \{2, 3, 4\}, \{3, 4, 5\}, \{4, 5, 6\}, \{1, 5, 6\}, \{1, 2, 6\}$ together with all non-empty subsets of these 2-simplices. Sketch the geometric realization K. Then exhibit a simplicial subcomplex L of K for which the inclusion $|L| \hookrightarrow |K|$ is a homotopy equivalence. (Don't prove the homotopy equivalence.) Hence determine the mod-2 homology $H_n(K)$ for all $n \geq 0$.

3. Explain what is meant by:

   (a) a *simplicial complex* K,

   (b) a *filtered* simplicial complex K,

   (c) the degree n *persistent Betti numbers* $\beta_n^{s,t}$ of a filtered simplicial complex K.

4. The degree 0 and degree 1 persistent Betti numbers of a certain filtered simplicial complex K are given by the matrices

$$\beta_0^{s,t} = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix} \quad \beta_1^{s,t} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

   How many 'persistent connected components' and how many 'persistent 1-dimensional holes' does K have?

4 Essay type questions

1. Describe the Mapper clustering algorithm. Explain how it was used in this research paper of Nicolau, Levine and Carlsson to interpret data about a certain cancer.

2. Explain how persistent homology was used in this research paper by Carlsson, Ishkanov, de Silva and Zomorodian to study the local behavious of spaces of natural images.