Exercise 1

Definition 1. *A matroid is a pair $(E, \mathcal{I})$, where E is a finite set ("ground set"), I is a collection of subsets of E ("independent sets") such that*

(A1) $\emptyset \in \mathcal{I}$

(A2) If $L' \subseteq L$ and $L \in \mathcal{I}$, then $L' \in \mathcal{I}$

(A3) If $I, J \in \mathcal{I}$ and $|I| < |J|$, then exists $j \in J - I$ such that $I \cup j \in \mathcal{I}$ (exchange axiom)

Proposition 1. *Let E be a finite set of vectors in a vector space V. Let I be the collection of linearly independent subsets of E. Then (E, I) is a matroid.*

Proof. The proof of the first and the second axioms is trivial (A1, A2) since the empty set is linearly independent and taking a linear combination of less vectors that were linear independent is still a linear independent combination. In the following we will prove by contradiction that also the third axiom (A3) holds:
Let $I = \{i_1, \dots, i_a\}$, $J = \{j_1, \dots, j_b\}$, $|I| < |J|$. Suppose $I \cup j$ is not independent for any $j \in J - I$. This means that there exist a linear dependent relation

$$c_1 i_1 + \dots + c_a i_a + cj = 0$$

where c_i, c are scalars. Moreover it holds that $c \neq 0$ because if it would have been zero, then we would have had a linear relation in the i's. Therefore we have that

$$j = (-c_1 i_1 - \dots - c_a i_a)/c \Rightarrow j \in \text{span}(i_1, \dots, i_a) \Rightarrow J - I \subseteq \text{span}(I)$$

But also $I \subseteq \text{span}(I)$. Since both $J - I$ and I are in the $\text{span}(I)$, then also their union must be: $J \subset \text{span}(I)$. Therefore

$$\underbrace{\text{span}(J)}_{|J| \text{ dimensional}} \subseteq \underbrace{\text{span}(I)}_{|I| \text{ dimensional}} \Rightarrow |J| \leq |I|$$

which is a contradiction to $|I| < |J|$. □

Proposition 2. *Let $G = (V, E)$ be a graph. Let $\mathcal{I}$ the collection of independent sets of edges (forests of G). Then $(E, \mathcal{I})$ is a matroid.*

Lemma 1. *Let $G = (V, E)$. If $I \subset E$ is independent, then G[I] has $|V| - |I|$ connected components.*

Proof of Lemma 1. Consider $G = (V, \emptyset)$. G has exactly $|V|$ connected components. Every time I add an edge $e \in I$ i lose exactly one connected component since not losing a connected component by adding e would mean to create a cycle which cannot be done since I is an independent set of edges. □

Proof of Proposition 2. The proof of A1 is trivial since if I have no edgesI have no cycles (empty set is independent). The second axiom says: "if I have two sets of edges J and I with J bigger or equal I and J forms no cycles, the also I forms no cycles". This is also trivially true since removing edges from a graph with no cycles cannot create cycles. We will now prove by contradiction that also A3 holds. Suppose $I, J \in \mathcal{I}$ (independent sets) and $|I| < |J|$. Due to Lemma 1, $G[I]$ has $|V| - |I|$ connected components and $G[J]$ has $|V| - |J|$ respectively. Suppose that every edge $j \in J - I$ forms a cycle in $G[I \cup \{j\}]$. Since every edge from $J - I$ added to $G[I]$ closes some cycle, it doesn't decrease the number of connected components and so it holds that $G' = (V, I \cup J)$ has $|V| - |I|$ connected components. By removing from G' the edges in I, the number of connected components can only increase, thus it holds that $G'' = (V, J)$ has at least $|V| - |I|$ connected components. Given the observation before it must so hold that $|J| \leq |I|$ which is a contradiction to $|I| < |J|$. □

Given a matroid $M = (E, I)$ and $w : E \mapsto \mathbb{R}_{>0}$ we are interested into finding a basis of minimum weight (sum of weight of edges). Similar to Kruskal's MST greedy algorithm, we have

Algorithm 1 Matroid Greedy Algorithm for Minimum Weight Basis

```
function MinWeightBasis(M = (E, I), w)
    B ← {}
    S ← Sort(E, w)                    ▷ Sort elements in ascending order of weight
    for e ∈ S do
        if B ∪ {e} ∈ I then
            B ← B ∪ {e}
        end if
    end for
    return B
end function
```

Proof of correctess of Algorithm 1. Assume that the greedy algorithm gave the basis $B = \{i_1, \dots, i_n\}$ with $w(i_1) \leq w(i_2) \leq \dots \leq w(i_n)$ (order is enforced by the fact that the algorithm select at each step the cheaper "i"). Assume that $C = \{j_1, \dots, j_n\}$ with $w(j_1) \leq w(j_2) \leq \dots \leq w(j_n)$ is cheaper than B, hence

$$w(C) < w(B) \Leftrightarrow w(j_1) + w(j_2) + \dots + w(j_n) < w(i_1) + w(i_2) + \dots + w(i_n)$$

I can find $w(j_r) < w(i_r)$. Consider the set $B' = \{i_1, \dots, i_{r-1}\}$ is independent since $B' \subseteq B$. For the same reasoning also $\{j_1, \dots, j_r\}$ is independent. Due to the exchange axiom we can find a j_s such that $\{i_1, \dots, i_{r-1}, j_s\}$ is independent. Now $w(j_s) \leq w(j_r) < w(i_r)$. The greedy algorithm would have chosen j_s instead of i_r. We therefore have a contradiction. $\square$

Remark 1. *The possible outcomes of the greedy algorithm are all the bases of minimum weight (by running many times the algorithm we could recover them all!).*

Remark 2. *Matroids are exactly the simplicial complexes where the greedy algorithm always works. A simplicial complex can also be defined as matroid with only A1, A2 hold.*

Exercise 2

The solution to this exercise is taken from the CTDA book, proposition 5.14 and [1][Proposition 3.2].

Definition 2. *A simplicial complex K is a weak $(p+1)$-*pseudomanifold *if each p-simplex is a face of no more than two $(p+1)$-simplices in K.*

Definition 3. *An* undirected flow network (G, s_1, s_2) *consists of an undirected graph G with vertex set $V(G)$ and edge set $E(G)$, a capacity function $c : E(G) \to [0, +\infty]$, and two non-empty disjoint subsets s_1 and s_2 of $V(G)$. Vertices in s_1 are referred to as* sources *and vertices in s_2 are referred to as* sinks.

Definition 4. *A* cut (S, T) *of (G, s_1, s_2) consists of two disjoint subsets S and T of $V(G)$ such that $S \cup T = V(G)$, $s_1 \subseteq S$, and $s_2 \subseteq T$. The set of edges that connect a vertex in S and a vertex in T are referred as the* edges across the cut (S, T) *and is denoted as $\xi(S, T)$. The* capacity of a cut (S, T) *is defined as $c(S, T) = \sum_{e \in \xi(S,T)} c(e)$. A* minimal cut of (G, s_1, s_2) *is a cut with the minimal capacity.*

Definition 5. *Let K be a simplicial complex, for $q \geq 1$, two q-simplices σ and σ' of K are q-*connected *in K if there is a sequence of q-simplices of K, $(\sigma_0, \dots, \sigma_l)$, such that $\sigma_0 = \sigma$, $\sigma_l = \sigma'$, and for all $0 \leq i < l, \sigma_i$ and σ_{i+1} share a $(q-1)$-face. The property of q-connectedness defines an equivalence relation on q-simplices of K. Each set in the partition induced by the equivalence relation constitutes a q-*connected component *of K.*

Suppose that the input weak $(p+1)$-pseudomanifold is K which is associated with a simplex-wise filtration $\mathcal{F} : \varnothing = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_n$ and the task is to compute an optimal persistent cycle of a finite interval $[b, d) \in \mathrm{Dgm}_p(\mathcal{F})$. Let $\sigma_b^{\mathcal{F}}$ and $\sigma_d^{\mathcal{F}}$ be the creator and destructor pair for the interval $[b, d)$. We first construct an undirected dual graph G for K where vertices of G are dual to $(p+1)$-simplices of K and edges of G are dual to p-simplices of K. One dummy vertex v_∞ termed as infinite vertex which does not correspond to any $(p+1)$-simplices is added to G for graph edges dual to those boundary p-simplices, i.e., the p-simplices that are faces of at most one $(p+1)$-simplex.

We then build an undirected flow network on top of G where the source is the vertex dual to $\sigma_d^{\mathcal{F}}$ and the sink is the infinite vertex along with the set of vertices dual to those $(p+1)$-simplices which are added to $\mathcal{F}$ after $\sigma_d^{\mathcal{F}}$. If a p-simplex is $\sigma_b^{\mathcal{F}}$ or added to $\mathcal{F}$ before $\sigma_b^{\mathcal{F}}$, we let the capacity of its dual graph edge be its weight; otherwise, we let the capacity of its dual graph edge be $+\infty$.

In the dual graph, an edge is created for each p-simplex. If a p-simplex is a face of two $(p+1)$ simplicies, we simply let its dual graph edge connect the two vertices dual to its two $(p+1)$-simplicies; otherwise, its dual graph edge has to connect to the infinite vertex on one end. A problem about this construction is that some weak $(p+1)$-pseudomanifolds may have p-simplices being face of no $(p+1)$-simplices and these p-simplices may create self loops around the infinite vertex. To avoid self loops, we simply ignore these p-simplices. The reason why we can ignore these p-simplices is that they cannot be on the boundary of a $(p+1)$-chain and hence cannot be on a persistent cycle of minimal weight. Algorithmically, we ignore these p-simplices by constructing the dual graph only from what we call the $(p+1)$-connected component of K containing $\sigma_d^{\mathcal{F}}$.

If (S,T) is a cut for a flow network built on G, we denote with $\theta^{-1}(E(S,T))$ the set of p-simplices dual to the edges across the cut. Since simplicial chains with $\mathbb{Z}_2$ coefficients can be interpreted as sets, $\theta^{-1}(E(S,T))$ is also a p-chain.

Proposition 3. *For any cut (S,T) of (G,s_1,s_2) with finite capacity, the p-chain $c=\theta^{-1}(E(S,T))$ is a persistent p-cycle of $[b,d)$ and $w(c)=C(S,T)$.*

Proof. We proceed by steps.
1. $c=\partial(A)$. Let $A=\theta^{-1}(S)$, we first want to prove $c=\partial(A)$, so that c is a cycle.
1.1. $c\subseteq\partial(A)$. Let σ^p be any p-simplex of c, then $\theta(\sigma^p)$ connects a vertex $u\in S$ and a vertex $v\in T$. If $v=v_\infty$, then σ^p cannot be a face of another $(p+1)$-simplex in K other than $\theta^{-1}(u)$. So, σ^p is a face of exactly one $(p+1)$-simplex of A. If $v\neq v_\infty$, then σ^p is also a face of exactly one $(p+1)$-simplex of A (the other face is not in A). Therefore, $\sigma^p\in\partial(A)$.
1.2. $\partial(A)\subseteq c$. Let σ^p be any p-simplex of $\partial(A)$, then σ^p is a face of exactly one $(p+1)$-simplex σ_0^{p+1} of A. If σ^p is a face of another $(p+1)$-simplex σ_1^{p+1} in K, then $\sigma_1^{p+1}\notin A$, otherwise σ^p would not be a boundary in A. In particular, $\theta(\sigma_1^{p+1})\notin S$ and so $\theta(\sigma_1^{p+1})\in T$, since (S,T) is a partition of $V(G)$, and $\theta(\sigma^p)$ connects the vertex $\theta(\sigma_0^{p+1})\in S$ and the vertex $\theta(\sigma_1^{p+1})\in T$ in the graph G, thus $\sigma^p\in c=\theta^{-1}(E(S,T))$.
If instead σ^p is a face of exactly one $(p+1)$-simplex in $K,\theta(\sigma^p)$ must connect $\theta(\sigma_0^{p+1})\in S$ and $v_\infty\in T$ in G, since, by construction, the infinite vertex belongs to the sink $s_2\subseteq T$. So we have $\theta(\sigma^p)\in E(S,T)$, i.e., $\sigma^p\in\theta^{-1}(E(S,T))$.
2. $s_2\neq\emptyset$ and c cannot be empty. The fact that $s_2\neq\emptyset$ follows from the fact that s_2 contains the dummy veertex and so it cannot be empty. To prove that c is not empty, we notice that, since $s_2\subseteq T$ and $s_2\neq\emptyset$, we have that $T\neq\emptyset$ and so, since the dual graph is connected, necessarily there will exist edges crossing the cut (S,T) and so the pre-image under the bijection θ of these edges is not empty.
3. c gets created at b and becomes a boundary at d. We first show that c is created by $\sigma_b^{\mathcal{F}}$. Suppose that c is created by a p-simplex $\sigma^p\neq\sigma_b^{\mathcal{F}}$. Since $C(S,T)$ is finite, we have σ^p must be created before b (index $(\sigma^p)<b$). We can let c' be a persistent cycle of $[b,d)$. Since c' is killed at d, $c'=\partial A'$ where A' contains the simplex $\sigma_d^{\mathcal{F}}$. Then we have $c+c'=\partial(A+A')$. We now look at the $(p+1)$-chains that make c and c' a boundary, which are A and A'. Since c is the boundary of A, and A corresponds to the simplices on the source side of the cut, and since $\sigma_d^{\mathcal{F}}$ is a source, also A must contain $\sigma_d^{\mathcal{F}}$. Thus, if you add A and A' together, you get a chain that definitely does not contain $\sigma_d^{\mathcal{F}}$ (since we work over $\mathbb{Z}_2$!), and only contains simplices occuring before d. So $c+c'$ is a p-cycle created by $\sigma_b^{\mathcal{F}}$ (since c' is created at b) which becomes a boundary before $\sigma_d^{\mathcal{F}}$ is added. This means that $\sigma_b^{\mathcal{F}}$ is already paired when $\sigma_d^{\mathcal{F}}$ is added, contradicting the fact that $\sigma_b^{\mathcal{F}}$ is paired with $\sigma_d^{\mathcal{F}}$.

Similarly, we can prove that c is not a boundary until $\sigma_d^{\mathcal{F}}$ is added, so c is a persistent cycle of $[b,d)$.
4. $w(c)=C(S,T)$. We have

$$C(S,T)=\sum_{e\in\theta(c)}C(e)=\sum_{\theta^{-1}(e)\in c}w(\theta^{-1}(e))=w(c)$$

where the first equality holds since the capacity of (S, T) is finite, the second equality follows by the definition of capacity for our undirected flow network and the last equality follows by the definition of weight of a chain. □

References

[1] Tamal K. Dey, Tao Hou, and Sayan Mandal. Computing minimal persistent cycles: Polynomial and hard cases. *SODA '20: Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, page 2587–2606, 2020.