Algorithm for extended persistence diagram

June 15, 2016

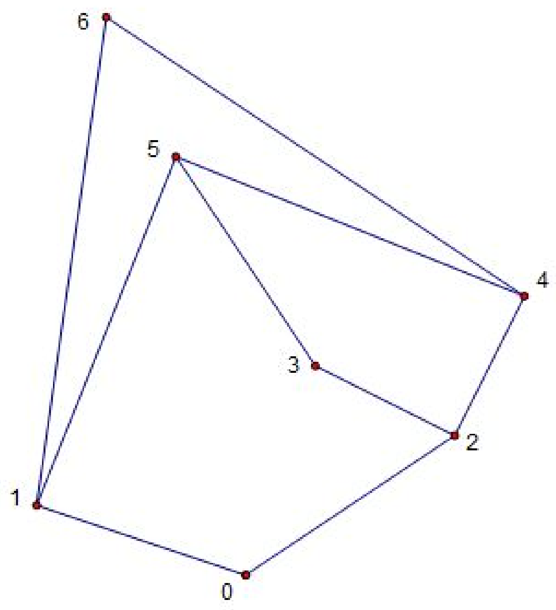

Figure 1: Graph G

Introduction

We study a non-directed graph $G = (V, E)$ with a vertex set V and an edge set E. We employ the numbers $\{0, 1, \ldots\}$ to label the vertices in V and if two vertices i and j are joined by an edge in E, it will be denoted by σ_{ij} with $i < j$. A simple non-directed graph means that there are no self-loops in G and every pair of points has at most one edge connecting them. A path P_{ij}^G from u_i to u_j in a non-directed graph G is an alternative sequence of vertices and edges. It starts with the vertex u, then follows by an edge σ_{ur}. In the middle of the sequence, any two consecutive edges are separated by their adjacent vertex, finally, the path ends with the vertex v. A simple path means that every edge or vertex of G appears at most once in the path. In this note, a path means a simple path.

Suppose that there is a height value function $f : V \mapsto \mathbb{R}$ such that $f(0) < f(1) < \ldots < f(n)$, where $|V| = n + 1$.

Let the number of edges be $|E| = m$. Besides ordering the vertices by a height value function, we queue the edges according to the rules that σ_{ij} lines before σ_{rs} if *“$j < s$” or “$j = s$ and $i < r$”*.

Figure 1 gives an example of a graph G with $|V| = 7$ vertices $\{0, 1, 2, 3, 4, 5, 6\}$ and $|E| = 9$ edges $\{\sigma_{01}, \sigma_{02}, \sigma_{23}, \sigma_{24}, \sigma_{15}, \sigma_{35}, \sigma_{45}, \sigma_{16}, \sigma_{46}\}$. The height value function f is set to be the upwards height of a vertex. The compatible order of edges is

$$\sigma_{01} \to \sigma_{02} \to \sigma_{23} \to \sigma_{24} \to \sigma_{15} \to \sigma_{35} \to \sigma_{45} \to \sigma_{16} \to \sigma_{46}$$

Union-Find algorithm for ordinary part of persistence diagram

The objective is to compute the extended persistence diagram (pdg) of the filtration f with G, in dimension 0 and in dimension 1.

We know in advance that points in pdg can be categorized into three parts: Ordinary, Extended, Relative. Furthermore, the Ordinary and Relative parts of pdg can be computed in linear time complexity with the help of path-compressed union-find algorithms. More precisely, both of them can be computed in $\mathcal{O}(m\alpha(n))$, where $\alpha(n)$ is the Ackermann function.

Take the Ordinary part as an example to see the use of the union-find algorithm. Before moving to the effect caused by vertices and edges, we denote a subgraph $G_k = \{V_k, E_k\}$ of G consisted of vertices $V_k = \{0, \ldots, k\}$ and all edges in E connected these points, namely $E_k = \{\sigma_{ij} \in E | i < j \leq k\}$. The ordinary part of the filtration of f with G is given by

$$\varnothing \subset G_0 \subset G_1 \subset \ldots \subset G_n = G$$

First of all, every vertex in V gives birth to an homology group in dimension 0.

As for an edge $\sigma_{jk} \in E$, it either sentences an homology group in dimension 0 to death in the case that it joints two different connected components $C(j)$ and $C(k)$ in the graph G_{k-1}^{+j} formed by the union of subgraph G_{k-1}, the vertex $\{k\}$, and all edges $\sigma_{ik} \in E_k$ with $i < j$, $G_{k-1}^{+j} = G_{k-1} \cup \{k\} \cup \{\sigma_{ik} \in E_k | i < j\}$; or gives birth to an homology group in dimension 1 by creating a cycle in the connected component in G_{k-1}^{+j} that both vertices j and k belong to.

In the first situation, we find the smallest vertex in each component $C(i)$ and $C(j)$, and call them the oldest ancestor vertex $anc_{c(i)}$ and $anc_{c(j)}$ respectively. Remind that the corresponding homology group in dimension 0 of a connected component is created by its oldest ancestor vertex. If $f(anc_{c(i)}) < f(anc_{c(j)})$, then according the elder rule, the homology group of $anc_{C(j)}$ dies and we have a vertex-edge pair $(anc_{C(j)}, \sigma_{jk})$. This pair represents a point $(f(anc_{C(j)}), f(k))$ in the pdg of dimension 0.

In the second situation, we create a cycle of dimension 1 at point k with edge σ_{jk}. This homology will live till the end of the filtration f, and will eventually die in the relative part of the extended filtration of f with relative homology groups $H(G, G^i)$ where $G^i = \{i, \ldots, n\} \cup \{\sigma_{rs} \in E | s > r \geq i\}$. It corresponds to the abscissa of an extended point in dimension 1. The hard part is to find the matched y-axis coordinate in $\{0, \ldots, k-1\}$.

Therefore, the union-find algorithm serves to keep track on the merge of different connected components along the adhesion of edges which yields information of pdg in dimension 0.

We will focus in the following on how to search efficiently the death moment of a cycle in relative homology groups.

A useful example

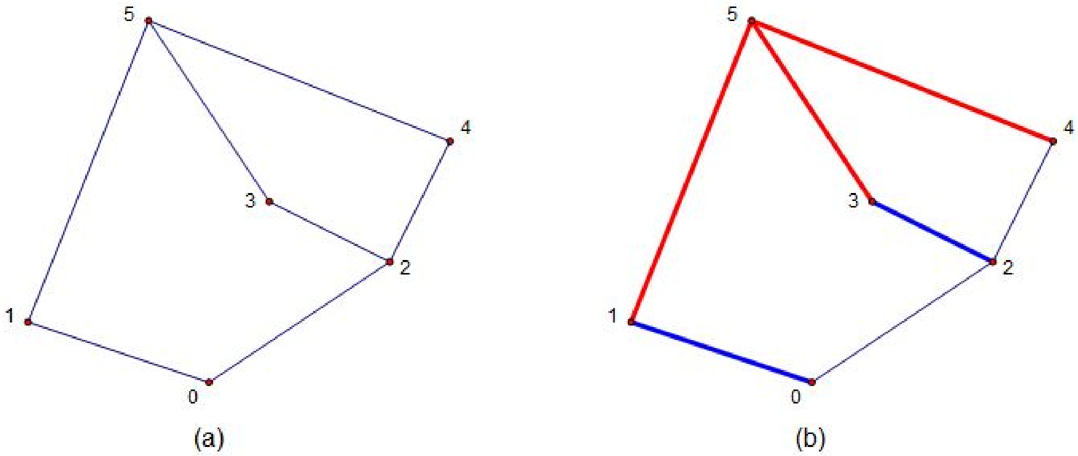

Figure 2: subgraph $G_k = G_5$

Figure 2(a) shows the subgraph $G_k = G_5$ of graph G. We observe that there are three cycles in the subgraph, namely the cycle $[5-3-2-4-5]$, the cycle $[5-1-0-2-3-5]$ and the cycle $[5-1-0-2-4-5]$. two of the three cycles are independent and they correspond to two different points, $(f(5), f(2))$ and $(f(5), f(0))$ in the dimension 1 pdg. Both cycles are born in the ordinary part of the filtration of f and they die in the relative part of the filtration of f.

We denote the lower neighbors of a vertex j in G_k by $\mathcal{N}_k^{j-} = \{i|\sigma_{ij} \in E_k, i < j \leq k\}$, and a upper neighbor of i in G_k by $\mathcal{N}_k^{i+} = \{j|\sigma_{ij} \in E_k, i < j < k\}$. We also denote the subgraph beyond a certain level i of the subgraph G_k by G_k^i with $i = 0, ...k$, more precisely let $G_k^i = \{i, i+1, \ldots, k\} \cup \{\sigma_{rs} \in E | i \leq r < s \leq k\}$ when $i \leq k$, otherwise $G_k^i = \varnothing$.

The first remark is that all the cycles in G_5 are born at point $f(k) = f(5)$, the height value of the largest vertex in the subgraph. Since all the cycles in G_5 must have the vertex 5, the other endpoints of edges σ_{j5} in a cycle will certainly belong to $\mathcal{N}_5^{5-} = \{1, 3, 4\}$. We also observe that these tree edges $(\sigma_{15}, \sigma_{35}, \sigma_{45})$ are equivalent in the sense of forming a cycle with G_4, and two of the three will be named as positive edges by giving birth to two dimension 1 homology classes in G.

The second remark is that a point in the pdg may be associated to different cycles, even without counting the multiplicity of this point in the diagram. For instance, $(f(5), f(0))$ has multiplicity 1 but could be matched to the cycle $[5-1-0-2-3-5]$ or to the cycle $[5-1-0-2-4-5]$. This ambiguity reveals the insight fact that an exact representation of cycles by point pairs or by edge pairs will become meaningless or even intractable due to the non-deterministic way to choose a base of cycles to generate all cycles in the graph. This is similar to the non deterministic way to choose a base of vectors in a real vector space.

The third remark is that G_4 contains no cycles, in another word, it is a tree, so that every pair of points (i, j) in G_4 has a unique non-directed path form i to j, vice-versa. This point is important, though it can be implied from the first remark, because every cycle can be broken into two paths, a path $[i-5-j]$ with $i, j \in \mathcal{N}_5^{5-}$ and a unique path P_{ij} from i to j in G_4. Henceforth, the smallest vertex in P_{ij}, denote it by w, splits the path P_{ij} into two sub-paths whose vertices are larger than w and smaller than k. These two connected components belongs to G_{k-1}^{w+1}. Inversely, if a point r with two of its edges in $\{\sigma_{rs} | s \in \mathcal{N}_k^{r+}\}$ (upward edges of r) that join two different connected components in G_{k-1}^{r+1} which contains respectively the vertex i and j, then r must on the path P_{ij} because there is only one path from i to j. Furthermore, because r is smaller than vertices in components of G_{k-1}^{r+1}, it is thus smaller than any of the vertices on the path P_{ij}, so we have $r = w$.

This property provides us a method to determine the lowest vertex in a cycle by using the union-find algorithm to keep track on the merge of connected components in V_4 by adding one by one the edges in the order $\sigma_{24} \to \sigma_{23} \to \sigma_{02} \to \sigma_{01}$. It also provides a way to use an vertex-edge pair to characterize a cycle after we actually find that cycle. We choose a positive edge σ_{i5} between the two edges $(\sigma_{i5}, \sigma_{j5})$ that join at 5 such that $i < j$, the vertex-edge pair for the cycle becomes (w, σ_{i5}).

The forth remark is that "an essential dimension 1 homology class (i.e. a cycle) of G_k^i gets born at the same time that a dimension 1 relative homology class of (G_k, G_k^i) dies". This indicates that every death moment of a cycle in (G_5, G_5^i) has a one-to-one match to the birth time of a cycle in G_5^i (we use $-f$ as the filtration function to count the birth time of cycles in G_5^i).

By excision theorem, the relative homology groups in $\mathbf{H}_1(G_5, G_5^i)$ are isomorphism to the relative homology groups in $\mathbf{H}_1(G, G^i)$. Therefore, the birth time $f(w)$ of a cycle in G_5^i can be associated to the death moment of a cycle in $\mathbf{H}_1(G, G^i)$ (indeed, these two cycles are identical except for the direction we consider the cycle.) and gives raise to a point $(f(5), f(w))$ in pdg.

An important thing needed to be keep in mind with the union-find algorithm in the third remark is that it avoids the representation problem of a cycle before actually finding that cycle. The things that really count in a union-find algorithm is the related position of connected components that contains lower neighbor vertices $\mathcal{N}_5^{5-}$. Therefore, the order of merging elements in $\mathcal{N}_5^{5-}$ into connected components will finally determine the death sequence as well as the positive edges of dimension 1 homology classes in G_5.

In our example,see Figure 2(b), $\mathcal{N}_5^{5-} = \{1, 3, 4\}$. Following the union-find algorithm, we descend from 4 to 0. At first, vertices $\{4, 3\}$ creates respectively a connected component. When we reach at vertex 2, we merge $\{3, 4\}$ together after adding the edge σ_{24} and σ_{23}. This gives death to a cycle which can be represented by $(2, \sigma_{35})$. After that, we keep on descending to 1 and the 0. We then merge the connected component $\{1\}$ and $\{3, 4, 2\}$ together. It kills another cycle which is represented by $(0, \sigma_{15})$. We observe that the fusion order of vertices in $\mathcal{N}_5^{5-}$ is $((4, 3), 1)$ and the positive edges are $\{\sigma_{35}, \sigma_{15}\}$. However, we don not know this order in advance until we finish running the union-find algorithm.

Moving one step forward

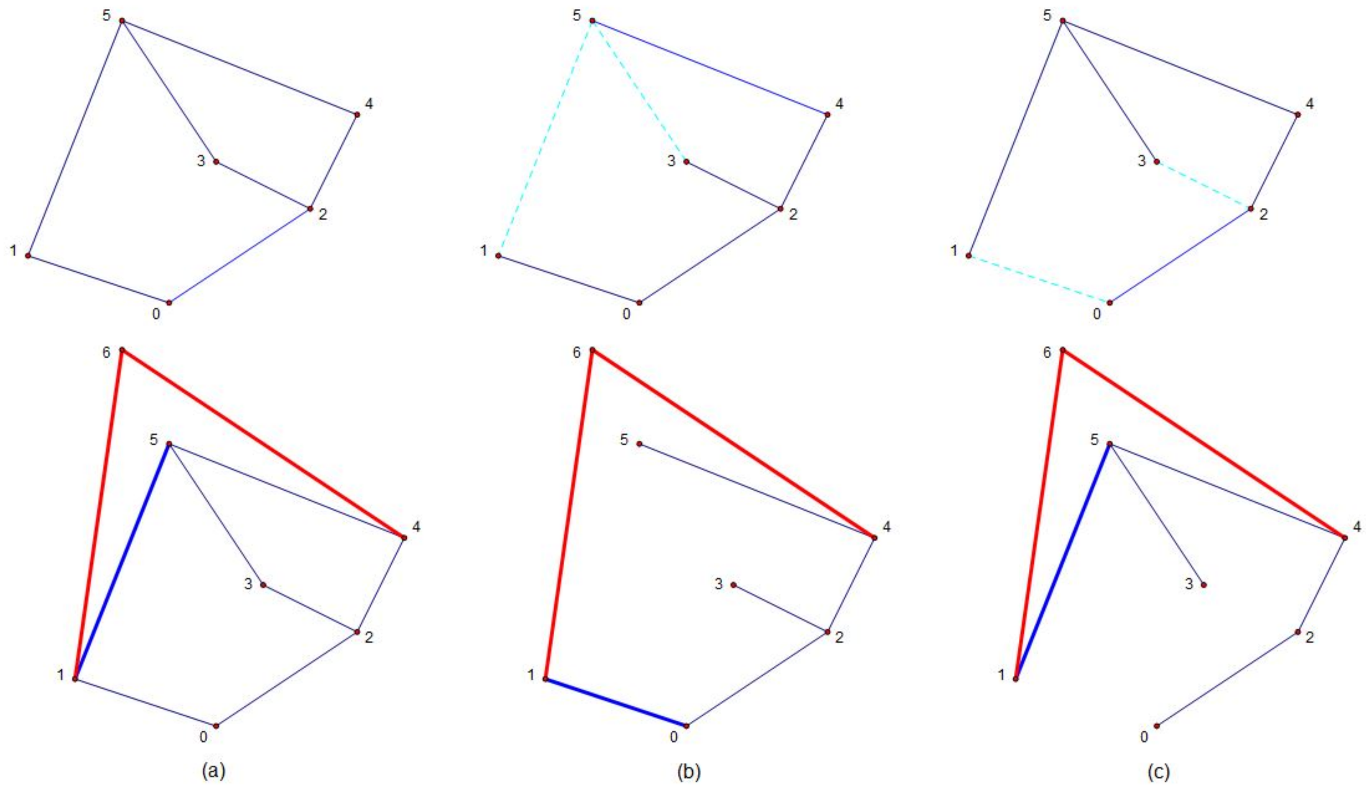

Figure 3: (a) graph G_5 and G_6; (b) spanning trees in G_5 with negative edges for dimension 0 homology classes T_5^- and its union with $\{\sigma_{16}, \sigma_{46}\}$; (c) tree-type subgraph T_5 of G_5 and the subgraph $F_6 = T_5 \cup \{6\} \cup \{\sigma_{16}, \sigma_{46}\}$ of G_6.

After determining the death times of cycles born at $f(5)$, we consider now the graph $G_6 = G$. We can see in the Figure 3(a) that there is one cycle born at $f(6)$ and this cycle dies at time $f(1)$ in the relative part of filtration f, $\mathbf{H}_1(G, G^1)$.

Following the same notation in the previous section, we denote the lower neighbors of 6 by $\mathcal{N}_6^{6-} = \{1, 4\}$. If we look at directly the graph G_6 in Figure 3(a), we find that G_5 is not a tree so that there are in total 3 cycles in G_6. These cycles are not all born at $f(6)$. Two of them are born at $f(5)$ and have already been studied in the previous step. Nevertheless, we can still apply a modified union-find algorithm on G_5 to determine the death time of the dimension 1 homology class born at $f(6)$.

Similar to the previous union-find algorithm, we queue the edges in G_5 such that σ_{ij} lines before σ_{rs} if *"$i > r$" or "$i = r$ and $j > s$"*. In another word

$$\sigma_{45} \to \sigma_{35} \to \sigma_{24} \to \sigma_{23} \to \sigma_{15} \to \sigma_{02} \to \sigma_{01}.$$

We keep track on the merge of two connected components in which contains respectively at least one vertex in $\mathcal{N}_6^{6-}$ and label it with a vertex-edge pair. More precisely, if we merge two connected components $C(i)$ and $C(j)$ with edge σ_{rs} such that i (resp j) is the largest vertex in $\mathcal{N}_6^{6-} \cap C(i)$ (resp $\mathcal{N}_6^{6-} \cap C(j)$) and $i < j$, we give this fusion a label (r, σ_{i6}). Along the union-find algorithm for connected components finding, we do not label any fusion of connected components that at most one of them contains vertices in $\mathcal{N}_6^{6-}$, neither do we label the creation of any cycles.

In our example, the moment of adding the edge σ_{23} creates a cycle in the connected component $C(4) = \{5, 4, 3, 2\} \cup \{\sigma_{45}, \sigma_{35}, \sigma_{24}\}$. We do not record this redundant modification in an already connected component as if the edge σ_{23} doesn't exist. After that, the merge of connected component $C(1)$ and $C(4)$ raises a label of vertex-edge pair $(1, \sigma_{16})$. We know that a cycle born at $f(6)$ has been killed since we find a path from 1 to 4 through in the new merged connected component from $C(1)$ and $C(4)$. Since we know in advance there is only one cycle born at $f(6)$, and we have already found its death time at $f(1)$, there is no need to study the fusion of connected components caused by the remaining edges $\{\sigma_{02}, \sigma_{01}\}$.

Since G_5 is not a tree, so we do not have a single path from any two vertices $i, j \in \mathcal{N}_6^{6-}$. Therefore, we need to check if the height value of the vertex $f(w) = f(1)$ in the vertex-edge pair $(1, \sigma_{16})$ corresponds to the death time of a cycle born at $f(6)$. This can be verified by the order that we add edges progressively to merge connected components contains vertices of the lower neighbors of 6 in order to build paths between these vertices.

Review the case of G_5, we observe that G_4 is a tree and all the cycles in G_5 contain the vertex 5. In this setting, things turn out to be much more easier to understand because all the cycles we found, all the death times we determined, and all the points in the dimension 1 pdg of this subgraph will be undoubtedly matched to the a single birth time $f(5)$.

While we are in the case of G_6, we don not has a tree structure with G_5. If we try to erase some edges in G_5 to form a tree or a forest, we may return to our favorable situation. However, a random remove of any number of edges will lose a unpredictable amount of information and distort too much the graph structure. In consequence, we mismatch the birth and death time of cycles appeared in the origin unmodified graph G_6.

For example, in Figure 3 (b), we try to use a spanning tree formed with negative edges of dimension 0 homology classes in G_5 with all the edges of type $\{\sigma_{j6}\}$ (equivalently, removing all positive edges of cycles in G_5) to capture the death times of cycles born at $f(6)$. What we find is a large cycle $[6 - 1 - 0 - 2 - 4]$ whose death time is $f(0)$, as a combination of two cycles $[6 - 1 - 5 - 4]$ and $[5 - 1 - 0 - 2 - 4]$ in the origin subgraph G_6. Indeed, the only dimension 1 homology class, the cycle, born at $f(6)$ in the graph G_6 has already been dead when we arrive at the homology groups $\mathbf{H_1}(G_6, G_6^1)$ in the extended filtration of f. The erase of edge σ_{15} turns out to lose the information that we search for that cycle. On the other hand, in Figure 3 (c), we try to use another tree-type subgraph of G_5, T_5, formed by removing negative edges $\{\sigma_{23}, \sigma_{01}\}$ of cycles in G_5, in another words, by removing one edge of type σ_{rs} in the connected component that contains edge σ_{i5} for each vertex-edge label (r, σ_{i5}) of cycle in G_5. Let $F_6 = T_5 \cup \{6\} \cup \{\sigma_{j6}\}$. We observe that in there is still a cycle $[6-1-5-4]$ in F_6 and it dies at time $f(1)$.

It seems like a coincidence, but we will prove in the following that the subgraph $F_k = T_{k-1} \cup \{k\} \cup \{\sigma_{jk}\}$ of G_k that captures all the death times of independent cycles born at $f(k)$ in the original graph G.

Definition of T_k and F_k

Let $|V| = n+1$ and we have an compatible order of edges such that σ_{ij} lines before σ_{rs} if "$j < s$" *or* "$j = s$ *and* $i < r$". The lower neighbors of a vertex k is the set of endpoint of edges σ_{jk} with $j < k$, denoted by $\mathcal{N}_k^- = \{j | \sigma_{jk} \in E, j < k\}$. We denote the subgraph of G_k beyond certain level $i < k$ the graph $G_k^i = \{i, \ldots, k\} \cup \{\sigma_{rs} \in E | i \leq r < s \leq k\}$.

For $0 \leq k \leq n-1$, T_k is a subgraph of G_k with vertices set $V_k = \{0, \ldots, k\}$. We call that T_k is traceable if it satisfied the following three properties:

- If there is a path from u to v in G_k, then there must be a path from u to v in T_k. In another word, T_k preserves the connected components of G_k.
- T_k contains no cycles, so is can be viewed a collection of tree-type connected components of vertices in V_k.
- For any two vertices $u_i, u_j \in \mathcal{N}_{k+1}^-$, if there is a path $P_{ij}^{T_k}$ from u_i to u_j in T_k, then the minimum vertex w_{ij} on this unique path $P_{ij}^{T_k}$ in T_k is the maximum vertex among all the minimum vertices on paths $\{P_{ij}^{q,G_k}\}$ from u_i to u_j in G_k. Namely

$$\begin{aligned} w_{ij} &= \arg\min_{v} \{f(v) | v \in P_{ij}^{T_k}\} \\ &= \arg\max_{w_{ij}^q} \{f(w_{ij}^q) | f(w_{ij}^q) = \min\{f(v)\}, \forall v \in P_{ij}^{q,G_k}\} \end{aligned}$$

For $k = 1, \ldots, n$, F_k is a subgraph of G_k defined as a union of T_{k-1}, the vertex k, and all the edges in G_k with an endpoint k, namely

$$F_k = T_{k-1} \cup \{k\} \cup \{\sigma_{jk} \in E | j < k\} \subset G_k.$$

Figure 3 (c) gives an example of T_5 and F_6 in the graph G.

Our objective is to prove that if a cycle is found in the graph F_k, its largest vertex k and its smallest vertex w must correspond to the a point $(f(k), f(w))$ in the dimension 1 pdg of G. Before seeing the demonstration of this property, we present in the first place the recursive construction of graphs T_k and F_k and the algorithm based on these subgraphs that finds the points of type extended in the dimension 1 pdg of G.

Algortihm for T_k and F_k and points of type extended in G

Let $T_0 = F_0 = \{0\}$. For $k = 1$, if $\sigma_{01} \in E$ then $T_1 = F_1 = \{0, 1\} \cup \{\sigma_{01}\}$, otherwise, $T_1 = F_1 = \{0, 1\}$.

Suppose now that we have already constructed a traceable graph T_{k-1} and F_k. We denote the subgraph of T_{k-1} beyond the level r by $T^r_{k-1} = \{r, \ldots, k-1\} \cup \{\sigma_{j(k-1)} \in E | r \leq j \leq k-1\}$.

First of all, we will find a basis $\mathcal{CB}^{F_k}_k$ of cycles in F_k. In other words, we find a collection of cycles in F_k such that any other cycle in F_k can be written as a formal sum modulo 2 of cycles in the collection, and none of the cycle in the collection can be expressed as a formal sum of other cycles in this collection. A formal sum of modulo 2 of cycles is a cycle formed by removing all edges that appear in even number of times in the sum, as well as the adjacent vertices between any two of these removed edges.

The method is to use the union-find algorithm to trace down the fusion of certain connected components. Starting from the vertex set V_{k-1}, we descend from the level $r = k-1$ to $r = 0$ and add progressively edges to rebuild the graph T_{k-1}. The order of adhesion of edges is that σ_{ij} is added before σ_{rs} if "$j > s$" *or* "$j = s$ *and* $i > r$".

While adding σ_{rs}, we check whether the addition of such an edge merges two connected components in the subgraph $T^{rs}_{k-1} = T^{r+1}_{k-1} \cup \{r\} \cup \{\sigma_{rt} | s < t\}$. If so, we merge these two components together.

Furthermore, in the situation that both connected components contain at least one element of $\mathcal{N}^-_k$, we find a cycle in the graph F_k. Denote the largest vertex of the intersection between $\mathcal{N}^-_k$ and these two connected components by u_i and u_j. We actually find the unique path $P^{T_{k-1}}_{ij}$ in T_{k-1} from u_i to u_j passing r and a cycle $cyc^k_{ij} = [k - u_i - \ldots - r - \ldots - u_j - k]$ in F_k.

Because all edges added to the graph up till this moment are beyond the level r, we know for certain that r is the minimum vertex of the path $P^{T_{k-1}}_{ij}$, as well as for the cycle cyc^k_{ij}. We assign to the cycle cyc^k_{ij} an edge-pair $(\sigma_{rs}, \sigma_{u_i k})$ if $u_i < u_j$, otherwise $(\sigma_{rs}, \sigma_{u_j k})$. The level r corresponds to a death time of a dimension 1 homology class in G born at $f(k)$.

We stop once we reach the level $r = 0$ or we have already merged all the vertices in $\mathcal{N}^-_k$ into a single connected component. The collection of basis cycle is $\mathcal{CB}^{F_k}_k = \{cyc^k_{ij}\}$. There should be $d_k - c_{k-1}$ cycles in $\mathcal{CB}^{F_k}_k$, where $d_k = |\mathcal{N}^-_k$ and c_{k-1} is the number of connected components of the subgraph G_{k-1}. The edge-pair of a cycle in $\mathcal{CB}^{F_k}_k$ gives us a point $(f(k), f(w))$ in the dimension 1 pdg of G.

To construct T_k, we delete the edge σ_{rs} of each edge-pair $(\sigma_{rs}, \sigma_{u_i k})$ in the graph F_k. The new graph is also traceable, and it is T_k.

$$
\begin{aligned}
T_k &= F_k \backslash \{\sigma_{rs} \in T_{k-1} | \sigma_{rs} \text{ merges two cc of } T^{rs}_{k-1} \text{ contains vertices in } \mathcal{N}^-_k\} \\
&= F_k \backslash \{\sigma_{rs} \in T_{k-1} | f(r) \text{ is the death time of a cycle in } F_k\} \\
&= G_k \backslash \{\sigma_{rs} \in G_{k-1} | f(r) \text{ is the death time of a cycle in } G_k\}
\end{aligned}
$$

Consequently, $F_{k+1} = T_k \cup \{k+1\} \cup \{\sigma_{j(k+1)} \in E | j < k+1\}$.

The traceableness of T_k assures that we find the death time of all dimension 1 homology classes born at $f(k)$ in the graph G. As k goes to n, we find all points of type extended in G.

Correctness of the algorithm

We consider a simple non-directed graph G and the simple paths in G as well as in its subgraphs G_k. A cycle in a simple non-directed graph contains at least 3 vertices.

Every cycle in G is a dimension 1 homology class in the homology group $\mathbf{H}_1(G)$. To draw the points with abscissa $f(k)$ of type extended in the pdg, we choose independent dimension 1 homology classes born at time $f(k)$.

Definition 1

A modulo 2 formal sum of a set of cycles in G are a subgraph in G that contains all the edges as its two endpoints that appear with an odd number of times in these cycles.

Definition 2

n cycles contains the vertex k in G_k are said to be independent if the modulo 2 formal sum of any subset of these cycles cannot be expressed as a modulo 2 formal sum of cycles in G_{k-1}.

Proposition 1

A graph with n vertex and n edges must contain a cycle.

Proof:

If these n edges form a path in the graph, then there are at least $n+1$ different vertices on the path. However, the graph contains only n vertices, contradiction.

Corollary 1.1

For any given n pairs of numbers (i, j) for $i, j \in \{1, \dots, n\}$, there must be $k < n$ pairs which for a circle of pairs. In another words, there are k pairs of numbers form a circle like $(a, b), (b, c), \dots, (j, k), (k, a)$.

Proof:

We build a graph G with n vertices denoted as $u_1, \dots u_n$. An edge σ_{ij} is in the graph if (i, j) is a pair of numbers. Then there are n edges in the graph G. By Proposition 0, there is a cycle in G. the pairs of number correspond to edges on this cycle form a circle of pairs like $(a, b), (b, c), \dots, (k, a)$.

Propostion 2

A cycle contains the vertex k must contains exactly two vertices u_i and u_j in the lower neighbors of k $\mathcal{N}_k^-$.

Proof:

Because a cycle in a non-directed simple graph contains at least 3 vertices, there exist two different vertices v_i and v_j on the cycle other than k. Following the cycle, we find a single simple path from v_i to v_j passing k. Thus, there are two edges of type $\{\sigma_{ik}, \sigma_{jk}\}$ with the vertex $\{u_i, u_j\}$ which form a sub-path $[u_i - \sigma_{ik} - k - \sigma_{jk} - u_j]$ in the cycle. We then conclude that $u_i, u_j \in \mathcal{N}_k^-$. The uniqueness derives from the definition of simple path in a non-directed graph.

Corollary 2.1

In the graph G_k of G, let C_{k-1} be a connected component of G_{k-1}. Suppose that there are r vertices in the intersection of $\mathcal{N}_k^- \cap C_{k-1}$, denoted by $\{v_1, \ldots, v_r\}$. Then there exist, and at most, $r-1$ cycles in $C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\}$ contained the vertex k which are independent in G_k.

Proof:

By proposition 2, a cycle in $C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\}$ contained the vertex $\{k\}$ must contains exactly two vertices $v_i, v_j \in \{v_1, \ldots, v_r\}$.

Let $\{cyc_k^q\}_{q=1,\ldots,r}$ be r cycles in $C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\}$. Then, each cycle $cyc_k^q, \forall q = 1, \ldots, r$ has a pair of vertices (v_{i_q}, v_{j_q}).

By Corollary 1.1, since there are only r vertices in $\mathcal{N}_k^- \cap C_{k-1}$, there thus exists a circle of vertex-pair of length $s \leq r$ in $\{(v_{i_q}, v_{j_q})\}_{q=1,\ldots,r}$. Denote this circle by $(u_1, u_2), (u_2, u_3), \ldots, (u_s, u_1)$.

Then the modulo 2 formal sum of these s cycles is a subgraph H in G_{k-1}. Furthermore, let the path from u_l to u_{l+1} on the related cycle in G_{k-1} be $P_{l,l+1}^{G_{k-1}}$ $(u_{s+1} = u_1)$. Because all edges $\{\sigma_{u_lk}\}_{l=1,\ldots,s}$ appear exactly two times in the modulo 2 formal sum of these cycles, the subgraph H equals to the modulo 2 formal sum of s paths $\{P_{l,l+1}^{G_{k-1}}\}_{l=1,\ldots,s}$. They form a circle of paths head and tail docking. Thus, the subgraph $H = \sum_{l=1}^{s} P_{l,l+1}^{G_{k-1}}(mod2)$ will be a collection of cycles in G_{k-1}.

Therefore, by definition, these r cycles containing the vertex k are not independent in G_k, nor do they be independent in a subgraph $C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\} \subset G_k$. In another word, there are at most $r-1$ cycles contained k in $C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\}$ that are independent in G_k.

On the other hand, let $P_{1q}^{G_{k-1}}$ be a path in C_{k-1} from v_1 to v_q for $q = 2, \ldots, r$. We build $r-1$ cycles

$$cyc_k^{1q} = P_{1q}^{G_{k-1}} \cup \{k\} \cup \{\sigma_{1k}, \sigma_{v_qk}\} \subset (C_{k-1} \cup \{k\} \cup \{\sigma_{v_1k}, \ldots, \sigma_{v_rk}\}), \quad \forall q = 2, \ldots, r.$$

We observe that the cycle cyc_k^{1q} contains an edge σ_{v_qk} which does not belong to any other $r-2$ cycles $cyc_k^{G_{k-1}}$. Hence, the modulo 2 formal sum contains cycle $cycl_k^{1q}$ will never be a subgraph in G_{k-1}. By definition, these $r-1$ cycles contained the vertex k are independent in G_k.

Corollary 2.2

In the graph G_k, denote the number of vertices in $\mathcal{N}_k^-$ by d_k, the number of connected component in G_{k-1} by c_{k-1}. Then there are exactly $d_k - c_{k-1}$ independent cycles in G_k contained the vertex k which are independent in G_k. In another words, there are $d_k - c_{k-1}$ dimension 1 homology classes born at time $f(k)$ in G_k.

Proof:

Let d_k^i be the number of vertices in $\mathcal{N}_k^-$ and in a connected component C_{k-1}^i of the graph G_{k-1}.

By applying the corollary 2.1 to each connected component of $C_{k-1}^i \subset G_{k-1}$ for $i = 1, \ldots, c_{k-1}$, we have $d_k^i - 1$ cycles created with vertices of C_{k-1}^i that are independent in G_k. Besides, a cycle created with vertices in C_{k-1}^i contains an edge σ_{vk} that does not belong to any other cycles created with another connected component $C_{k-1}^j \subset G_{k-1}$ with $j \neq i$.

Therefore, the total $\sum_{i=1}^{c_{k-1}} d_k^i - 1 = d_k - c_{k-1}$ cycles in G_k contained the vertex k are independent in G_k.

Since any cycle in G_k contained the vertex k must have two lower neighbor of k, $(v_r, v_s) \in (\mathcal{N}_k^-)^2$ belong to the same connected component C_{k-1}^i, it be then be expressed as modulo 2 formal sum of the $d_k^i - 1$ cycles with C_{k-1} independent in G_k and some cycles in G_{k-1}.

Thus, there are at most $d_k - c_{k-1}$ cycles in G_k contained the vertex k which are independent in G_k.

Proposition 3 [a sufficient condition]

Let $\mathcal{N}_0^+$ be the neighbors of vertex 0, $\mathcal{N}_0^+ = \{j > 0 | \sigma_{0j} \in E\}$. Suppose That $(u_i, u_j) \in (\mathcal{N}_0)^2$, and r is the minimum vertex among all maximum vertices on different paths from u_i to u_j in G without passing 0. We show that there is no path from u_i to u_j in G_t for $0 < t < r$ which does not pass 0. Furthermore, there is a dimension 1 homology class contained u_i, u_j and the vertex 0 in G_r gets born at time $f(r)$.

Proof:

Suppose that there is a path from u_i to u_j without passing 0 in G_t with $t < r$, then the maximum vertex on the path must be smaller than r, which results in a contradiction to the definition of r. Hence, there is no cycle in G_t that contains u_i, u_j and the vertex 0.

On the other hand, by combining the path from u_i to u_j passing r and the path $[u_i - \sigma_{0i} - 0 - \sigma_{0j} - u_j]$, we create a cycle in G_r that contains both u_i, u_j and the vertex 0.

Therefore, the conclusion holds.

Corollary 3.1

Let $\mathcal{N}_k^-$ be the lower neighbors of the vertex k, $\{j < k | \sigma_{jk} \in E\}$. Suppose $(u_i, u_j) \in (\mathcal{N}_k^-)^2$, and w is the maximum vertex among all minimum vertices on different paths from u_i to u_j in G_k. Then there is no path from u_i to u_j in G_k^t for $k > t > w$. Furthermore, there is a dimension 1 homology class in G_k^w contains u_i, u_j and k in G_k^w that gets born at time $f(w)$, or equivalently, a dimension 1 homology class born at time $f(k)$ in G_k dies at time $f(w)$ in (G_k, G_k^w).

Proof:

We replace i by $k-i$ for all $i = 0, \ldots, k$, the graph G_k by G, the subgraph G_k^t by G_t, G_k^w by G_r, and w by r. By applying the proposition 3, the conclusion holds.

Corollary 3.2

Suppose $(f(k), f(w))$ is a point of type extended in the dimension 1 pdg with the highest y-coordinate among all points with $f(k)$ as abscissa. Then there is no path between any pair of vertices $(v_r, v_s) \in (\mathcal{N}_k^-)^2$ in the graph G_{k-1}^{w+1}.

Proof:

If there is a path connects a pair of vertices $(v_r, v_s) \in (\mathcal{N}_k^-)^2$ in the graph G_{k-1}^{w+1}. Let w_{rs} be the maximum vertex among all minimum vertices on path from v_r to v_s in G_{k-1}. We then have, $w_{rs} \geq w+1$. On the other hand, by the Corollary 3.1, $f(w_{rs})$ is the death time of a dimension 1 homology class born at $f(k)$ in G_k. This means that $(f(k), f(w_{rs}))$ is a point of type extended in the dimension 1 pdg with a y-coordinate larger than $f(w)$, contradiction.

Proposition 4

If $(f(k), f(w))$ is a point of type extended in the dimension 1 pdg of G, then there must be a cycle in G_k that contains both vertices k and w. Furthermore, k and w are respectively the largest and the smallest vertex on this cycle.

Proof:

From the excision theorem, the dimension 1 homology group $\mathbf{H}_1(G, G^i)$ is isomorph to the dimension 1 homology group $\mathbf{H}_1(G_k, G_k^i)$ for $i < k$. Furthermore, an essential dimension 1 homology class of G_k^i gets born at the same time that a dimension 1 relative homology class of (G_k, G_k^i) dies.

We know that every dimension 1 homology class of $G_k = G_k^0$ are essential, because the graph contains only vertices and edges. If $(f(k), f(w))$ is a point of type extended in the dimension 1 pdg of G, then by its definition, there is a dimension 1 homology class of G born at $f(k)$ and this homology class has its image in $\mathbf{H}_1(G, G^w)$ with $w < k$ for the inclusion map $g : \mathbf{H}_1(G_k) \mapsto \mathbf{H}_1(G, G^w)$.

Since the death moment $f(w)$ of this dimension 1 homology class in $\mathbf{H}_1(G, G^w) \simeq \mathbf{H}_1(G_k, G_k^w)$ corresponds to the birth time $f(w)$ of a dimension 1 homology class in $\mathbf{H}_1(G_k^w)$, we have a creation of at least one new cycle that contains k and w in $G_k^w \subseteq G_k$.

Since k and w are the largest and the smallest vertices in G_k^w, they are surely the largest and the smallest vertices of the cycle.

Corollary 4.1

By proposition 2, the cycle created above contains exactly two lower neighbors of k, $(u_i, u_j) \in (\mathcal{N}_k^-)^2$. If the point $(f(k), f(w))$ is with multiplicity $\mu_{kw} = 1$, then w is the maximum vertex among all minimum vertices on paths from u_i to u_j in G_{k-1}.

Proof:

Denote the maximum vertex among all minimum vertices on paths from u_i to u_j in G_{k-1} by w_{ij}. The cycle provides a path from u_i to u_j in $G_{k-1}^w \subset G_{k-1}$, we have $w_{ij} \geq w$. By Proposition 2, a cycle is created at time $f(w_{ij})$ with the vertex u_i, u_j and k. As there is only on cycle created, this cycle is the same dimension 1 homology class corresponded to the point $(f(k), f(w))$ with multiplicity $\mu_{kw} = 1$. Therefore, $w_{ij} = w$.

Proposition 5 [a necessary condition]

Suppose that $(f(k), f(w))$ with $w < k$ is a point of type extended in the dimension 1 pdg. Let $u_i \in \mathcal{N}_k^-$ be a vertex in the lower neighbors of k such that there exists at least one path from k to w passing u_i in the subgraph G_k^w. We collect all the u_i in a set $\mathcal{U}_k = \{u_1, u_2, \ldots, u_l\} \subset G_{k-1}$. Therefore, there is at least one pair of vertices (u_i, u_j) in $\mathcal{U}_k$ such that $u_i \neq u_j$, the maximum vertex among all minimum vertices on different paths $\{P_{ij}^{q,G_{k-1}}\}_q$ from u_i to u_j in G_{k-1} will be w, namely

$$w = w_{ij} := \max\{w_{ij}^{q,G_{k-1}} | f(w_{ij}^{q,G_{k-1}}) = \min\{f(v)\}, \forall v \in P_{ij}^{q,G_{k-1}}\}.$$

Proof:

For any $u_i \in \mathcal{U}_k$, there exists a path $P_{iw}^{G_{k-1}^w}$ from u_i to w in G_{k-1}^w. Thus, all vertices on $P_{iw}^{G_{k-1}^w}$ are larger than w because the path belongs to G_{k-1}^w. Besides, for any two vertices $u_i, u_j \in \mathcal{U}_k$, u_i and u_j are connected in G_{k-1}^w as they all connected to w in G_{k-1}^w by paths $P_{iw}^{G_{k-1}^w}$ and $P_{jw}^{G_{k-1}^w}$. Thus, there exists a path $P_{ij}^{G_{k-1}^w}$ from u_i to u_j in G_{k-1}^w. Certainly, the minimum vertex on $P_{ij}^{G_{k-1}^w}$ is larger than w. Since $P_{ij}^{G_{k-1}^w} \in \{P_{ij}^{q,G_{k-1}}\}_q$, we have

$$w_{ij} := \max\{w_{ij}^{q,G_{k-1}} | f(w_{ij}^{q,G_{k-1}}) = \min\{f(v)\}, \forall v \in P_{ij}^{q,G_{k-1}}\} \geq w \quad \forall (u_i, u_j) \in (\mathcal{U}_k)^2.$$

Because $(f(k), f(w))$ is a point of type extended in pdg, by the Proposition 2, there must be a cycle *cyc* contains k and w in G_k such that k and w are the largest and the smallest vertices on this cycle. Let u_r, u_s be the two vertices in $\mathcal{N}_k^-$ that belong to this cycle.

We firstly point out that u_r and u_s are in the same connected component of G_{k-1} that contains w, because u_r and u_s are connected by a path passing w in the cycle *cyc* of G_k. Thus, we denote the path from u_r to w in the cycle *cyc* by $P_{rw}^{G_{k-1}}$, respectively $P_{sw}^{G_{k-1}}$ for the path from u_s to w in the cycle. We know that $P_{rw}^{G_{k-1}}$ and $P_{sw}^{G_{k-1}}$ do not share any common edges in G_{k-1}.

We prove the converse negative proposition. Suppose that there is no pair of vertices $(u_i, u_j) \in (\mathcal{U}_k)^2$ such that $u_i \neq u_j$ and w is the maximum vertex among all minimum vertices on different paths $\{P^{q,G_{k-1}}\}_q$ from u_i to u_j in G_{k-1}. In another words, for every pair of $(u_i, u_j) \in (\mathcal{U}_k)^2$, $w_{ij} > w$. We try to show that no dimension 1 homology class in G_k born at time $f(k)$ will die at time $f(w)$, or equivalently, no dimension 1 homology class is created at time $f(w)$ in $\mathbf{H}_1(G_k^w)$.

Firstly, every dimension 1 homology class in G_k born at time $f(k)$ will eventually dies at some time $f(t)$ with $t < k$. It corresponds to a point of type extended in dimension 1 pdg $(f(k), f(t))$. Thus, with the proposition 4, it creates a cycle cyc_t in G_k containing t and k.

Secondly, after the proposition 2, any dimension 1 homology class in G_k born at time $f(k)$ must contain exactly two different vertices in $\mathcal{N}_k^-$, denoted by $(v_r, v_s) \in (\mathcal{N}_k^-)^2$.

Apart from the path $[v_r - \sigma_{rk} - k - \sigma_{sk} - v_s]$ in the cycle, there is another path from v_r to v_s in G_{k-1}. Thus, these two vertices belong to the same connected component C of the subgraph G_{k-1}. Moreover, the smallest vertex on this second path from v_r to v_s is t.

What we are going to show is that t will never be w. And thus, this dimension 1 homology class will not die at $f(w)$.

An obvious situation is that if the connected component C doesn't contain the vertex w, then any cycles with v_r and v_s will not contains w, as they all belong to the same connected component.

Suppose now $w \in C$, then we must have $\mathcal{U}_k \subset C \cap \mathcal{N}_k^-$. Let $\mathcal{V}_k = (C \cap \mathcal{N}_k^-) \backslash \mathcal{U}_k$. If $\mathcal{V}_k \neq \varnothing$, we distinguish three cases by whether or not v_r and v_s belong to $\mathcal{U}_k$. Otherwise, we move directly to the third case.

- If $(v_r, v_s) \in (\mathcal{V})_k^2$, then every path from v_r to w in G_{k-1} contains at least one vertex strictly smaller than w, so as to v_s. We also have $v_r \neq w$ and $v_s \neq w$.

  We consider a path $P_{rs}^{G_{k-1}}$ in G_{k-1} from v_r to v_s.

  If $w \in P_{rs}^{G_{k-1}}$, then the sub-path of $P_{rs}^{G_{k-1}}$ from v_r to w contains a vertex strictly smaller than w, which results in the fact that the minimum vertex on the path $P_{rs}^{G_{k-1}}$ is not w.

  If $w \neq P_{rs}^{G_{k-1}}$, then the minimum vertex on $P_{rs}^{G_{k-1}}$ must not be w.

- If $v_r \in \mathcal{V}_k$ and $v_s \in \mathcal{U}_k$, then every path from v_r to w in G_{k-1} contains at least one vertex strictly smaller than w.

  We use the same argument to show that the minimum element of a path from v_r to v_s in G_{k-1} is not w.

  Things go exactly the same in case $v_s \in \mathcal{V}_k$ and $v_r \in \mathcal{U}_k$.

- If $(v_r, v_s) \in (\mathcal{U}_k)^2$, by the definition of $\mathcal{U}_k$, we know that there exists at least a path from v_r to v_s in G_{k-1}^w.

  According to the hypothesis, among all paths from v_r to v_s in G_{k-1}^w, there is a path $P_{rs}^{G_{k-1}^w}$ in G_{k-1}^w whose minimum vertex is $w_{rs} > w$.

  By attaching to $P_{rs}^{G_{k-1}^w}$ the edges $\{\sigma_{rk}, \sigma_{sk}\}$, we form a new cycle $cyc_k^{w_{rs}}$ in G_k whose minimum vertex is $w_{rs} > w$.

  This cycle can be also viewed as a dimension 1 homology class in $G_k^{w_{rs}}$ created at time $f(w_{rs})$, because there is no path from u_r to u_s in G_{k-1}^i for $k > i > w_{rs}$ according to the maximization of w_{rs}. In consequence, no cycle with v_r and v_s in G_k^i is created before we reach w_{rs} along the direction from $i = k-1$ to $i = 0$.

  Hence, the death time $f(w_{ij})$ of this dimension 1 homology class $cyc_k^{w_{rs}}$ born at $f(k)$ in G_k is before $f(w)$.

In conclusion, in the first and the second cases, the minimum vertex of any path from u_r to u_s in G_{k-1} is not w, so none of the dimension 1 homology class in G_k born at $f(k)$ and with vertices $v_r \in \mathcal{V}_k$ or $v_s \in \mathcal{V}_k$ will die at time $f(w)$. In the third case, we show directly that all the dimension 1 homology classes born at time $f(k)$ and with vertices $(v_r, v_s) \in (\mathcal{U}_k)^2$ will die before $f(w)$. Therefore, the choice of some independent homology classes born at $f(k)$ to draw points of type extended in the dimension 1 pdg of G_k will never give raise to a point with y-coordinate $f(w)$. The converse negative statement of the proposition 5 holds, so as for the proposition 5.

Corollary 5.1

When $f(w)$ is the largest y-coordinate among all points of type extended with $f(k)$ abscissa in the dimension 1 pdg of G. In this case, $\forall(u_i, u_j) \in (\mathcal{U}_k)^2$, w is the maximum vertex among all minimum vertices on paths from u_i to u_j in G_{k-1}.

Proof: By Corollary 3.2, there is no path for any pair of vertices $(v_r, v_s) \in (\mathcal{N}_k^-)^2$ in the graph G_{k-1}^{w+1}. For a pair of vertices $(u_i, u_j) \in (\mathcal{U}_k)^2 \subset (\mathcal{N}_k^-)^2$, let w_{ij} be the maximum vertex among all minimum vertices on paths from u_i to u_j in G_k. We have $w_{ij} < w + 1$. On the other hand, u_i and u_j are connected by w in G_{k-1}^w, we then have $w_{ij} > w$. Therefore, $w_{ij} = w$.

Definition

A highest path from u_i to u_j in G_k, denoted by P_{ij}^{*,G_k} is a path in G_k from u_i to u_j whose minimum vertex w_{ij} is the largest among all minimum vertices on paths from u_i to u_j in G_k. The minimum vertices of all highest paths from u_i to u_j in G_k are equal by its definition.

By proposition 5, if $(f(k), f(w))$ is a point of type extended in the dimension 1 pdg, then there exists a pair (u_i, u_j) and a highest path , $P_{ij}^{*,G_{k-1}}$, from u_i to u_j in G_{k-1} whose minimum vertex is w.

Corollary 5.2

Let P_{iw} and P_{jw} be the sub-paths from u_i to w and from u_j to w in the highest path $P_{ij}^{*,G_{k-1}}$. Then P_{iw} and P_{jw} share no common edge.

Proof:

Firstly, P_{iw} and P_{jw} belong to the graph $G_{k-1}^w \subset G_k^w$.

Suppose conversely that P_{iw} and P_{jw} share some common edges. Let v be the maximum endpoint among all endpoints of shared edges. Then the sub-path P_{iv} from u_i to v in P_{iw} and the sub-path P_{jv} from u_j to v in P_{jw} share no edges. Therefore, we are able to build a new path from u_i to u_j passing v. This new path has a minimum vertex larger than w, because it belongs to G_k^w and does not contain the vertex w. This results in a contradiction to the maximization of w.

Corollary 5.3

For any vertex in $u \in \mathcal{U}_k$, a from u to w in G_{k-1}^w, $P_{uw}^{G_{k-1}^w}$ can only share edges with one of the two paths P_{iw} and P_{jw}. Besides, all paths $P_{uw}^{G_{k-1}^w}$ share their edges with the same path between P_{iw} or P_{jw}.

Proof:

Suppose that there is a path $P^1_{uw} \subset G^w_{k-1}$ shares edges with P_{iw}, and another path $P^2_{uw} \subset G^w_{k-1}$ shares edges with P_{jw}. P^1_{uw} and P^2_{vw} can be identical. For example, they both equal to the path $P^{G^w_{k-1}}_{uw}$.

Let x be the maximum endpoint among endpoints of shared edges in $P_{iw} \cap P^1_{uw}$; similarly, y is the maximum endpoint among endpoints of shared edges in $P_{jw} \cap P^2_{uw}$. Denote

- the sub-path from u_i to x in P_{iw} by P_{ix};
- the sub-path from u_j to x in P_{jw} by P_{jy};
- the sub-path from u to x in P^1_{uw} by P_{ux};
- the sub-path from u to y in P^2_{uw} by P_{uy}.

We have $x \neq w$ because $P_{iw} \cap P^1_{uw} \neq \varnothing$. Similarly, $y \neq w$. Thus $P_{ix}, P_{jy} \in G^{w+1}_{k-1}$.

Since $P_{iw} \cap P_{jw} = \{w\}$, we have $x \neq y$. Moreover $P_{ux}, P_{uy} \in G^w_{k-1}$, and $w \notin P_{ux}, w \notin P_{uy}$, we then know that x and y are connected in G^w_{k-1} by a path without w. We denote this path by $P_{xy} \in G^{w+1}_{k-1}$.

Therefore, u_i is connected to x by P_{ix}, x is connected to y by P_{xy}, y is connected to u_j by P_{jy}, and all these three path is in G^{w+1}_{k-1}, we have u_i is connected to u_j in G^{w+1}_{k-1}. In another word, there is a path from u_i to u_j in G^{w+1}_{k-1} whose minimum vertex is larger than w. This results in a contradiction to the maximization of w.

Corollary 5.4

Suppose that $(f(k), f(w))$ is a point of type extended with multiplicity μ_{kw} in the dimension 1 pdg, then there are at least μ_{kw} different pairs of $(u_i, u_j) \in (\mathcal{U}_k)^2$ such that w is the maximum vertex among all minimum vertices on different paths from u_i to u_j in G_{k-1}.

Proof:

We made a stronger statement that for a point $(f(k), f(w))$ with multiplicity at least μ_{kw}, there are at least μ_{kw} different pairs of $(u_i, u_j) \in (\mathcal{U}_k)^2$ satisfied the conclusion.

We prove it by induction. If $\mu_{kw} \geq 1$, by the proposition 5, the conclusion holds.

Suppose that the corollary holds for $\mu_{kw} \geq \mu \geq 1$, we prove the conclusion for $\mu_{kw} = \mu + 1$. Still by the proposition 5, we have a pair $(u_i, u_j) \in (\mathcal{U}_k)^2$ and a highest path from u_i to u_j in G_{k-1}, $P^{*,G_{k-1}}_{ij}$, whose minimum vertex is w.

We remove the edge σ_{ik} that connects the vertices u_i and k in the graph G_k, and denote the new graph by $\tilde{G}_k$. In the new graph $\tilde{G}_k$, the number of vertices in the lower neighbors of k is decreased by 1, as u_i no longer belongs to $\mathcal{N}^-_k$. Therefore, the total number of independent dimension 1 homology classes created at time $f(k)$ decreases 1, because the removal of an edge σ_{ik} no in G_{k-1} has no effect on the connected components of G_{k-1}.

For the reason that any cycle contains k in the subgraph $\tilde{G}_k \subset G_k$ is also a cycle in G_k, and the points in the dimension 1 pdg of $\tilde{G}_k$ is drawn with a choice of independent dimension 1 homology classes in $\tilde{G}_k$. Hence, these homology classes are also independent in the graph G_k. In consequence, the points of type extended in the dimension 1 pdg of the subgraph $\tilde{G}_k \subset G_k$ is a subset of the points of type extended in the dimension 1 pdg of the graph G_k.

Since the total number of points of type extended counted with multiplicity and with $f(k)$ abscissa in the dimension 1 pdg of the new graph $\tilde{G}_k$ is drop by 1, the multiplicity of the point $(f(k), f(w))$ is at most decreased by 1, namely $\mu_{kw}^{\tilde{G}_k} \geq \mu_{kw}^{G_k} - 1 = \mu_{kw} - 1 = \mu$.

By induction, there are at least μ pairs of vertices $(u_r, u_s) \in \mathcal{U}_k^{\tilde{G}_k}$ such that w is the minimum vertex of the highest path from u_r to u_s in G_{k-1}. The pair (u_i, u_j) is different with any of the above μ pairs (u_r, u_s) because $u_i \notin \mathcal{U}_k^{\tilde{G}_k} = \mathcal{U}_k \backslash \{u_i\}$. Thus, the conclusion holds.

Corollary 5.5

Denote the μ_{kw} pairs of vertices in $\mathcal{U}_k$ found in the corollary 5.4 by (u_{i_q}, u_{j_q}) for $q = 1, \ldots, \mu_{kw}$. Then, every pair $(u_r, u_s) \in \{(u_{i_q}, u_{j_q})\}_{q=1,\ldots,\mu_{kw}}$ has a vertex, either u_r or u_s, that is different to at least one vertice in the remaining $(\mu_{kw} - 1)$ vertex-pairs.

Furthermore, for a vertex-pair (u_{i_q}, u_{j_q}), Let $P_{i_q j_q}^{*,G_{k-1}}$ be a highest path from u_{i_q} and u_{j_q} in G_{k-1} whose minimum vertex is w. Let $cyc_{i_q j_q}^{G_k^w}$ be a cycle contained both u_{i_q}, u_{j_q} in G_k^w whose minimum and maximum vertices are k and w, for example

$$cyc_{i_q j_q}^{G_k^w} = P_{i_q j_q}^{*,G_{k-1}} \cup \{k\} \cup \{\sigma_{i_q k}, \sigma_{j_q k}\}.$$

In this way, these μ_{kw} cycles $\{cyc_{i_q j_q}^{G_k^w}\}_{q=1,\ldots\mu_{kw}}$ in G_k^w contained the vertex k are independent in G_k.

Proof:

From the demonstration of Corollary 5.4, we remove an edge $\sigma_{i_1 k}$ from the graph G_k after we find a vertex-pari $(u_{i_1}, u_{j_1}) \in (\mathcal{U}_k^{G_k})^2$, and all the other $\mu_{kw} - 1$ pairs $(u_{i_q}, u_{j_q}), q = 2, \ldots, \mu_{kw}$ belong to $(\mathcal{U}_k^{G_k} \backslash \{u_{i_1}\})^2$. Therefore, in the pair (u_{i_1}, u_{j_1}), the vertex u_{i-1} is different to all other $2.(\mu_{kw} - 1)$ vertices in the remaining $\mu_{kw} - 1$ vertex-pairs $\{(u_{i_q}, u_{j_q})\}_{q=2,\ldots,\mu_{kw}}$. Then the first conclusion holds.

Furthermore, we have a stronger statement that the μ_{kw} vertex-pairs in $(\mathcal{U}_k^{G_k})^2$ does not form a circle of vertex-pairs of any length $l \leq \mu_{kw}$.

We prove this statement by induction on μ_{kw}. $\mu_{kw} = 1$ is a trivial case. Suppose the statement holds for $\mu_{kw} = \mu$. When the multiplicity μ_{kw} of $(f(k), f(w))$ equals $\mu_{kw} = \mu + 1$, we find a pair of vertex $(u_{i_1}, u_{j_1}) \in (\mathcal{U}_k^{G_k})^2$ and the other $\mu_{kw} - 1 = \mu$ vertex-pairs $(u_{i_q}, u_{j_q}) \in (\mathcal{U}_k^{G_k} \backslash \{u_i\})^2, q = 2, \ldots, \mu + 1$ as before.

By induction, there are no circle of vertex-pairs of length $l \leq \mu$ form with $\{(u_{i_q}, u_{j_q})\}_{q=2,\ldots,\mu+1}$.

Since u_{i_1} is different to all the 2μ vertices in $\{(u_{i_q}, u_{j_q})\}_{q=2,\ldots,\mu+1}$, no circle of length $l \leq \mu+1$ will be form with (u_{i_1}, u_{j_1}) and $l-1$ vertex-pairs in $\{(u_{i_q}, u_{j_q})\}_{q=2,\ldots,\mu+1}$.

Otherwise, u_{i_1} must belong to a vertex-pair $(u_{i_t}, u_{j_t}) \in \{(u_{i_q}, u_{j_q})\}_{q=2,\ldots,\mu+1}$, contradiction.

The above statement is equivalent to said that the μ_{kw} cycles $\{cyc_{i_q j_q}^{G_k^w}\}_{q=1,\ldots\mu_{kw}}$ in G_k^w contained the vertex k are independent in G_k.

Proposition 6

Suppose the points of type extended associated to the $d_k - c_{k-1}$ dimension 1 homology classes born at time $f(k)$ in the pdg are $\{(f(k), f(w_l))\}_{l=1,\ldots L}$ with $w_1 \geq \ldots \geq w_L$. The multiplicity of a point $(f(k), f(w_l))$ is μ_{kw_l}. Then $\sum_{l=1}^{L} \mu_{kw_l} = d_k - c_{k-1}$.

In the graph G_k, there exists $d_k - c_{k-1}$ cycles $\{cyc_{i_q j_q}^{G_k^{w_l}}\}_{q=1,\ldots\mu_{kw_l}, l=1,\ldots,L}$ contained the vertex k that are independent in G_k.

In another words, we interpret the points of type extended with abscissa $f(k)$ in the dimension 1 pdg of G with the help of independent cycles and highest paths between vertex-pairs in $(\mathcal{N}_k^-)^2$.

Proof:

We prove the proposition by induction on L. When $L = 1$, the proposition holds after the Corollary 5.5. Suppose the proposition holds for $L - 1 \leq 1$.

We study firsty the independent cycles born at time $f(k)$ in G_k and dies at time $f(w_1)$ in $(G_k, G_k^{w_1})$ where $f(w_1)$ is the highest y-coordinate of points with $f(k)$ abscissa in the pdg.

Let $d_k^{w_1} = |\mathcal{U}_{kw_1}^{G_k}| = |\mathcal{N}_k^- \cap C_{k-1}^{w_1}(w_1)|$ where $C_{k-1}^{w_1}(w_1)$ is the connected component in $G_{k-1}^{w_1}$ that contains the vertex w_1. Let $\mathcal{U}_{kw_1}^{G_k} = \{u_1, u_2, \ldots, u_{d_k^{w_1}}\}$.

After the Corollary 5.5, there are exists μ_{kw_1} cycles in $G_k^{w_1}$ which are independent in G_k. With the Corollary 1.1, there are at least $\mu_{kw_1} + 1$ vertices in the set $\mathcal{U}_{kw_1}^{G_k}$. Thus $d_k^{w_1} - 1 \geq \mu_{kw_1}$.

On the other hand, the Corollary 5.1 says that every pair (u_i, u_j) has a highest path in $G_k^{w_1}$ from u_i to u_j whose minimum vertex is w_1. With the proposition 3, the cycle created with the vertex u_i and u_j in $G_k^{w_1}$ gets born at time $f(w_1)$. Because we have in total $d_k^{w_1}$ vertices in $\mathcal{U}_{kw_1}^{G_k}$, after the Corollary 2.1, we can construct $d_k^{w_1} - 1$ cycles in $G_k^{w_1}$ which are independent in G_k, namely

$$cyc_{u_1 u_q}^{G_k^{w_1}} = P_{u_1 u_q}^{*,G_k} \cup \{k\} \cup \{\sigma_{u_1 k}, \sigma_{u_q k}\}, \text{ for } q = 2, \ldots, d_k^{w_1}.$$

All these cycles are born at the same time $f(w_1)$. Thus, $\mu_{kw_1} \leq d_k^{w_1} - 1$.

Hence $d_k^{w_1} = \mu_{kw_1} + 1$. Denote the $d_k^{w_1} - 1$ cycles $\{cyc_{u_1 u_q}^{G_k^{w_1}}\}_{q=2,\ldots,d_k^{w_1}} = \{cyc_{i_q j_q}^{G_k^{w_1}}\}_{q=1,\ldots\mu_{kw_1}}$, namely $u_{i_q} = u_1, u_{j_q} = u_{q+1}, q = 1, \ldots, \mu_{kw_1}$.

Now, in order to pass to the induction hypothesis, we remove $\mu_{kw_1} = d_k^{w_1} - 1$ edges $\sigma_{u_q k}$ for $q = 2, \ldots d_k^{w_1}$ from the graph G_k. Denote $\tilde{G}_k = G_k \backslash \{\sigma_{u_q k}\}$. Thus, we left only one vertex of $\mathcal{U}_{kw_1}^{G_k}$ in the new graph $\tilde{G}_K$.

In the new graph, no dimension 1 homology classes gets born at time $f(k)$ in G_k and dies at time $f(w_1)$ in $(G_k, G_k^{w_1})$. Otherwise, it must contains two vertices $(\tilde{u}_i, \tilde{u}_j) \in (\mathcal{U}_{kw_1}^{\tilde{G}_k})^2$, a set of vertices in the lower neighbors of k connected to w_1 in $G_k^{w_1}$. However, by definition of $\tilde{G}_k$, there is only one vertex in this set.

Because $\tilde{G}_k \subset G_k$, we have that all the points of type extended in the pdg of $\tilde{G}_k$ must be the points in the pdg of G_k. $\tilde{G}_k$ contains no points like $(f(k), f(w_1))$, so there are at most $(d_k - c_{k-1}) - \mu_{kw_1} = \sum_{l=2}^{L} \mu_{kw_l}$ points in the pdg of $\tilde{G}_k$.

On the other hand, the connected component of G_{k-1} does not change, only the size of $\mathcal{N}_k^-$ diminues from d_k to $d_k - (d_k^{w_1} - 1) = d_k - \mu_{kw_1}$. By Corollary 2.2, there are contains $(d_k - \mu_{kw_1}) - c_{k-1}$ cycles in $\tilde{G}_k$ which are independent in $\tilde{G}_k$. These cycles will eventually die at some time $f(w_l)$ for $2 \leq l \leq L$ because they are also cycles in G_k. Consequentely, there are at least $(d_k - \mu_{kw_1}) - c_{k-1}$ points in the pdg of $\tilde{G}_k$.

Moreover, because the subgraph G_{k-1} does not change after the removal of μ_{kw_1} edges, so as the connected componented in any $G_{w_l+1}^{k-1}$ for any $2 \leq l \leq L$. If a connected component $C_{k-1}^{w+1}(u_i)$ in G_{w+1}^{k-1} contains a vertex $u_i \in \mathcal{U}_{kw_1}^{G_k}$, it must contains all the elements in $\mathcal{U}_{kw_1}^{G_k}$ since they are all connected to w_1, so these $d_k^{w_1}$ will belong to one single connected component in G_{w+1}^{k-1} for any $w < w_1$. So $C_{k-1}^{w+1}(u_i)$ contains u_1, otherwise, C_{k-1}^{w+1} contains no elements in $\mathcal{U}_{kw_1}^{G_k}$. In another word, the collection of connected components in $G_{k-1}^{w_1+1}$ for $2 \leq l \leq L$ which contains at least one vertex in $\mathcal{N}_k^-$ does not change. This means that the number of dimension 1 homology classes born at time $f(w_l)$ in $G_k^{w_l}$ and dies at time $f(k)$ which are independent in $G_k^{w_l}$ does not change. Therefore, the multiplicity of points $(f(k), f(w_l))$ in the new graph $\tilde{G}_k$ is still μ_{kw_l} for $2 \leq l \leq L$.

Hence, the pdg of $\tilde{G}_k$ consists of $L-1$ points $(f(k), f(w_l))$ with multiplicy μ_{kw_l} for $l = 2, \ldots, L$.

By induction, we have $d_k - \mu_{kw_1} - c_{k-1}$ cycles in $\tilde{G}_k$ that are independent in $\tilde{G}_k \subset G_k$, denote them by $\{cyc_{i_q j_q}^{G_k^{w_l}}\}_{q=1,\ldots\mu_{kw_l}}$ for $l = 2, \ldots, L$. The $d_k - \mu_{kw_1} - c_{k-1}$ vertex-pair in each cycle that connects k is denoted by $\mathcal{A} = \{(u_{i_q}^l, u_{j_q}^l)\}_{q=1,\ldots,\mu_{kw_l}, l=2,\ldots,L}$. By construction $\mathcal{A}$ contains only u_1, an element in $\mathcal{U}_{kw_1}^{G_k}$. Because the μ_{kw_1} vertex-pair for cycles in $\{cyc_{i_q j_q}^{G_k^{w_1}}\}_{q=1,\ldots\mu_{kw_1}}$ are $\{(u_1, u_q)\}_{q=2,\ldots,\mu_{kw_1}}$, any subset of these μ_{kw_1} vertex-pairs contains at least one vertex $u_t \in \{u_2, \ldots, u_{\mu_{kw_1}}\}$. Thus, there is no circle of vertex-pairs in $\mathcal{A} \cup \{(u_1, u_q)\}_{q=2,\ldots,\mu_{kw_1}}$.

Therefore, the $d_k - c_{k-1}$ cycles

$$cyc_{i_q j_q}^{G_k^{w_l}} = P_{u_{i_q}^l u_{j_q}^l}^{*,G_k} \cup \{k\} \cup \{\sigma_{u_{i_q}^l k}, \sigma_{u_{j_q}^l k}\} \quad \forall q = 1, \ldots \mu_{kw_l}, l = 1, \ldots, L$$

in G_k contained the vertex k are independent in G_k.

Among these cycles, there are exactly μ_{kw_l} cycles whose minimum vertex on the cycle is w_l.

Proposition

For $0 < k \leq n$, every cycle in F_k is a cycle in G, and every dimension homology class of F_k gets born at time $f(k)$.

Proof: This is obvious because F_k is a subgraph of G and T_{k-1} contains no cycles by definition.

Proposition 6

Suppose that $(f(k), f(w))$ is a point of type extended in the dimension 1 pdg. There is at least one cycle in F_k such that w is the minimum vertex of this cycle.

Proof:

Proposition 7

In the graph G, for any $k > 0$, there are exactly $d_k - c_{k-1}$ number of points of type extended whose abscissa is $f(k)$ in the dimension 1 pdg, where $d_k = |\mathcal{N}_k^-|$ and c_{k-1} is the connected components of subgraph G_{k-1}. In another words, there are $d_k - c_{k-1}$ different dimension 1 essential homology classes created at time $f(k)$.

Furthermore, there exists a set of $d_k - c_{k-1}$ cycles which form a basis $\mathcal{CB}_k$ such that all the other cycles created at time $f(k)$ in the graph G can be represented as a formal sum of these cycles, and non of the cycle in $\mathcal{CB}_k$ can be expressed as a formal sum of other elements in $\mathcal{CB}_k$.

Proof:

Proposition 8

Suppose $k > 0$. In the graph F_k, there exists at most a set of $d_k - c_{k-1}$ cycles $\mathcal{CB}_k^F$ that all the cycles in F_k can be represented as a formal sum of these cycles, and any of these cycles cannot be expressed as a formal sum of others in $\mathcal{CB}_k^F$.

Proof:

Corollary 8.1

We can construct a base $\mathcal{CB}_k^F$ of size $d_k - c_{k-1}$ with the help of the points $\{f(k), f(w_l)\}_{l=1,\dots,d_k-c_{k-1}}$ of type extended in the dimension 1 pdg.

Proof:

Proposition 9

The points of type extended in the dimension 1 pdg of the subgraph F_k represent the persistence of all the dimension 1 homology classes in G which are born at time $f(k)$.

Proof:

Proposition 10

T_k constructed above is traceable.