Filtrations

Saturday, January 16, 2021 11:06 AM

$\mathbb{T}$: topological space

$f: \mathbb{T} \to \mathbb{R}$ a real-valued function

$\mathbb{T}_a$: $f^{-1}(-\infty, a]$ sublevel set for value $a \in \mathbb{R}$

$\mathbb{T}_a \subseteq \mathbb{T}_b$ for $a \leq b$.

$a_1 < a_2 < \cdots < a_n$ distinct values, $a_0 = -\infty$, $\mathbb{T}_{a_0} = \phi$.

$$\mathcal{F}_f: \mathbb{T}_{a_0} \hookrightarrow \mathbb{T}_{a_1} \hookrightarrow \cdots \hookrightarrow \mathbb{T}_{a_n} \quad \cdots \quad (1)$$

$\mathcal{F}_f$: a space filtration.

* The inclusion in spaces induces homomorphisms in the singular homology groups

* $i: \mathbb{T}_{a_i} \hookrightarrow \mathbb{T}_{a_j} \Rightarrow h_p^{i,j}: H_p(\mathbb{T}_{a_i}) \to H_p(\mathbb{T}_{a_j})$

$$0 = H_p(\mathbb{T}_{a_0}) \to H_p(\mathbb{T}_{a_1}) \to \cdots \to H_p(\mathbb{T}_{a_n}) \text{ from (1)}$$

* Classes in $H_p(\mathbb{T}_{a_i})$ are mapped to classes in $H_p(\mathbb{T}_{a_j})$. Some classes become trivial (They die). Some classes survive.

$\operatorname{Im} h_p^{i,j}$ contains this information.

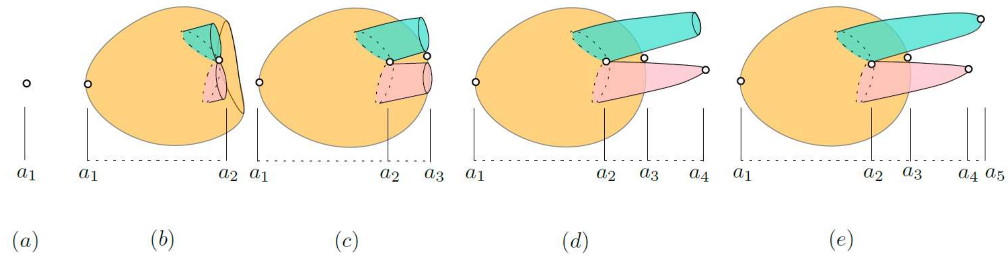


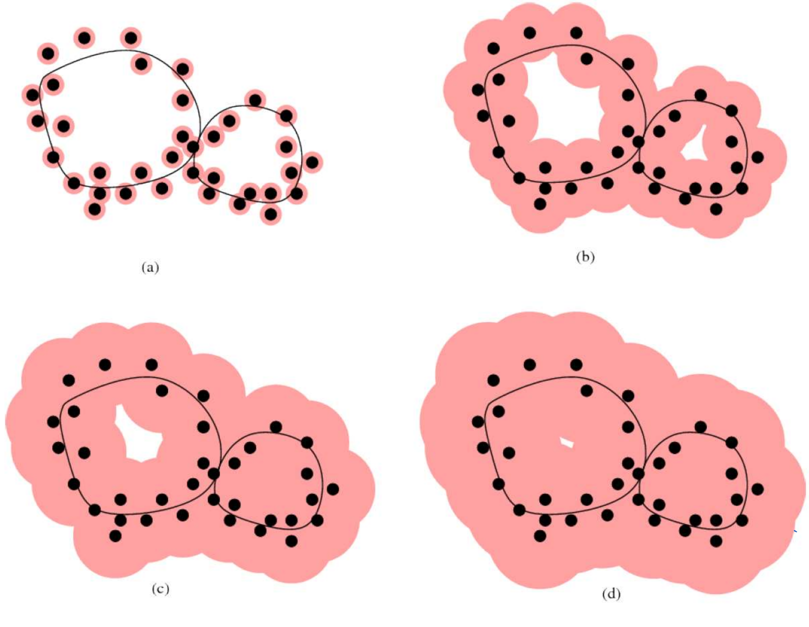
Figure 3.1: Persistence of a function on a topological space that has five critical values: (a) $\mathbb{T}_{a_1}$: only a new class in H_0 is created, (b) $\mathbb{T}_{a_2}$: two new independent classes in H_1 are created, (c) $\mathbb{T}_{a_3}$: one of the two classes in H_1 dies, (d) $\mathbb{T}_{a_4}$: the single remaining class in H_1 dies, (e) $\mathbb{T}_{a_5}$: a new class in H_2 is created.

- Point cloud: For a finite sample $P \subseteq (M,d)$ of a metric space M.

- $f : M \to R$, $x \mapsto d(x,p)$ where $p \in \underset{p \in P}{\operatorname{argmin}}\, d(x,p)$

  x has value = its distance to the closest point in P

- Sublevel sets M_a are union of metric balls of radius a centering points in P.



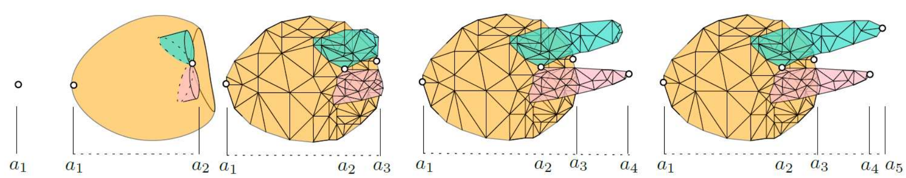
Simplicial Filtrations

- $\mathbb{T}$ is replaced with a simplicial complex K
- Singular homology is replaced with simplicial homology.
- $\mathbb{T}$ can be triangulated; for PCD, we can consider Čech or VR-complex.

Def (Simplicial filtration): A nested sequence of subcomplexes

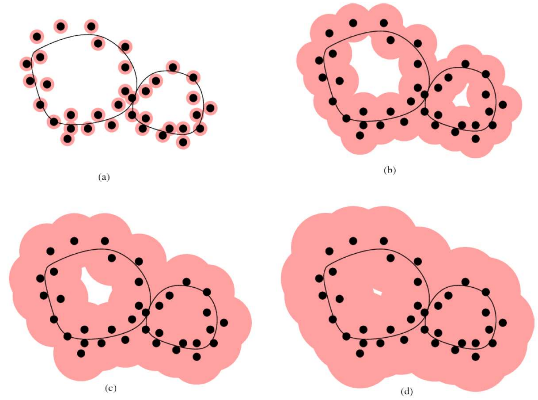
$$\mathcal{F}: K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$$

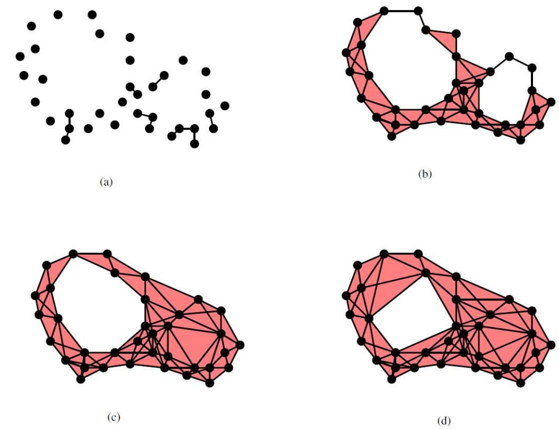
It is <u>simplex-wise</u> if $K_i \setminus K_{i-1} = \sigma_i$, a single simplex for every $i \in [1, n]$



(a) (b) (c) (d) (e)

Filtration on triangulated space





Union of balls $\Rightarrow$ Čech complex

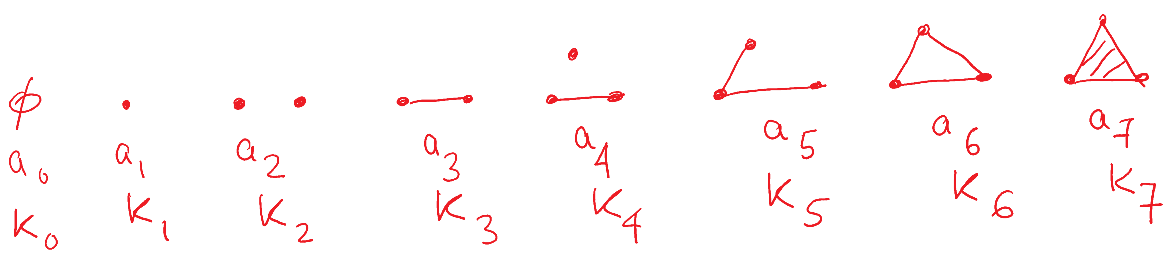
Filtration functions

- Simplex-wise monote function: $f: K \to \mathbb{R}$ is SMF if
  $\forall \sigma' \subseteq \sigma, \quad f(\sigma') \leq f(\sigma)$

$\Downarrow$

Sublevel sets $f^{-1}(-\infty, a_i] = K_i$ are subcomplexes.

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \cdots \subseteq K_n = K$$



Vertex function , PI-function

Saturday, January 16, 2021 1:48 PM

- Vertex function: $f : V(K) \to \mathbb{R}$
- Build a filtration out of f

<u>Lower/upper star</u> filtration

- Order vertices $f(v_1) \le f(v_2) \le \cdots f(v_n)$
- $\underset{\text{lower star}}{Lst(v)} = \{\sigma \in st(v) \mid$ all vertices of σ except v appear before v in the order$\}$
- $\underset{\text{upper star}}{Ust(v)} = \{\sigma \in st(v) \mid$ " .... after .... " $\}$
- $\overline{Lst(v)}, \overline{Ust(v)}$: closures
- $K_{f(v_i)} = \{\sigma \mid$ all vertices of σ are contained in $\{v_1, v_2, \dots v_i\}\}$
- v_0 a dummy vertex with $f(v_0) = -\infty$

$$\phi = K_{f(v_0)} \subseteq K_{f(v_1)} \subseteq K_{f(v_2)} \subseteq \cdots \subseteq K_{f(v_n)}$$

$$K_{f(v_i)} \setminus K_{f(v_{i-1})} = Lst(v_i), \quad i \in [1, n]$$

This is why it is called <u>lower star</u> filtration.

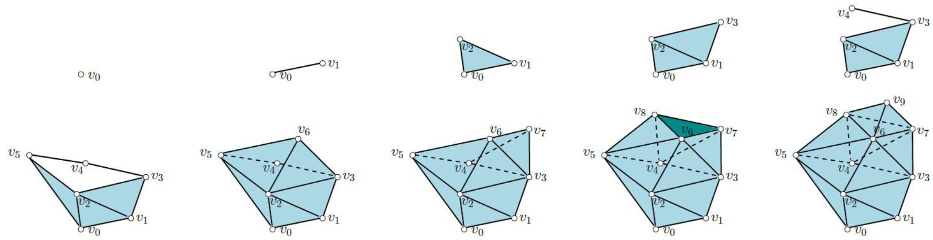


Figure 3.5: The sequence shows a simplex-wise *lower-star* filtration of K induced by a vertex function which is a 'height function' that records the vertical height of a vertex increasing from bottom to top here.

- Similarly one can have an Upper star filtration
- v_{n+1}, a dummy vertex, $f(v_{n+1}) = \infty$
- $K_{f(v_i)} = \{\sigma \mid \text{all vertices of } \sigma \text{ are in } \{v_i, v_{i+1}, \cdots v_n\}\}$

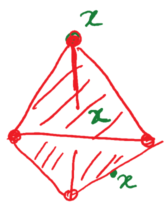
$$\phi = K_{f(v_{n+1})} \subseteq K_{f(v_n)} \subseteq K_{f(v_{n-1})} \subseteq \cdots \subseteq K_{f(v_1)}$$

$$K_{f(v_i)} \setminus K_{f(v_{i+1})} = \underbrace{Ust(v_i)}_{\text{upper star}}$$

* Both lower and upper star filtrations can be made simplex-wise by ordering the simplices in stars to obey the face relation

Def (PL-functions): $\bar{f} : |K| \to \mathbb{R}$ is a linear extension of a vertex function $f: V(K) \to \mathbb{R}$

$\forall x \in |K|$: $\bar{f}(x) = \sum_{i=1}^{k+1} \alpha_i f(v_i)$ where $\sigma_i = \{v_1, \ldots, v_{k+1}\}$ is the unique lowest dimensional simplex containing x and α_i's are the barycentric coordinates of x in σ_i.



* A simplex-wise lower star filtration for f is also a simplex-wise monotone function filtration for $\bar{f}: K \to \mathbb{R}$ where $\bar{f}(\sigma) = \max_{v \in \sigma} f(v)$
* for upper star filtration, it is $\bar{f}(\sigma) = \max_{v \in \sigma} (-f(v))$

* Persistence of f can be studied from the persistence of $\bar{f}$. (book).

Persistence

Saturday, January 16, 2021 2:47 PM

Filtration to homology module

- In both space and simplicial filtrations

$$H_pF: 0 = H_p(X_0) \to H_p(X_1) \to \cdots H_p(X_i) \xrightarrow{h_p^{i,j}} \cdots \to H_p(X_j) \to \cdots \to H_p(X_n)$$

$$X_i = \mathbb{T}_{a_i} \text{ or } X_i = K_i, \quad X = \mathbb{T} \text{ or } K$$

- F: filtration, H_pF: homology module

* p-th persistent homology group $H_p^{i,j} := \operatorname{Im} h_p^{i,j}$

* p-th persistent Betti number $\beta_p^{i,j} := \dim H_p^{i,j}$

* Nontrivial classes of $H_p^{i,j}$ exist from X_i to X_j

* They are not quotiented out by boundaries in X_j

<u>Fact</u> $H_p^{i,j} = Z_p(X_i) / (Z_p(X_i) \cap B_p(X_j))$

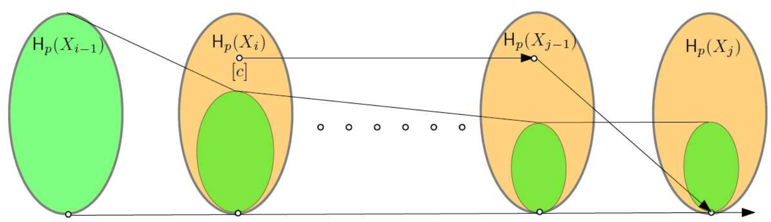
Birth & Death of classes

- A class appearing first time is called <u>born</u>
- A class that ceases to exist is called to <u>die</u>

Def (Birth): $[c]$ is born at X_i if

$[c] \in H_p(X_i)$ but $[c] \notin H_p^{i-1,i}$

Def (Death): $[c]$ dies entering X_j if $[c] \in H_p(X_{j-1})$ is not zero but $h_p^{j-1,j}([c]) = 0$

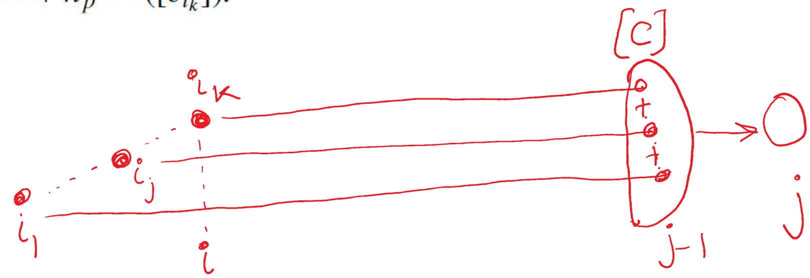


* When a class $[c]$ is born, many classes that are sum of $[c]$ and existing classes are also born

* When a class $[c]$ ceases to exist, many other classes may cease to exist along with it.

* So, we need some <u>canonical</u> pairing of birth and death.

Fact . *Let $[c]$ be a p-th homology class that dies entering X_j. Then, it is born at X_i if and only if there exists $i_1 \le i_2 \le \cdots \le i_k = i$ for some $k \ge 1$ so that $[c_{i_\ell}]$ is born at X_{i_ℓ} for every $\ell \in \{1, \ldots, k\}$, the classes $h_p^{i_1, j-1}([c_{i_1}]), \cdots, h_p^{i_k, j-1}([c_{i_k}])$ remain linearly independent, and $[c] = h_p^{i_1, j-1}([c_{i_1}]) + \cdots + h_p^{i_k, j-1}([c_{i_k}])$.*



Persistence Diagram (PD)

Sunday, January 17, 2021 10:21 AM

Persistence Diagram (PD)

Consider Extended plane $\mathbb{R} \times (\mathbb{R} \cup \{\infty\})$.

- A class born at a_i and dying at a_j is represented by point (a_i, a_j) in this plane
- To geneate pairs (a_i, a_j), we use a persistence pairing function (Assume $X = X_n = X_{n+1} \cdots = X_\infty$)

<u>Def</u>:

$$\mu_p^{ij} = \begin{cases} \left(\beta_p^{i,j-1} - \beta_p^{i,j}\right) - \left(\beta_p^{i-1,j-1} - \beta_p^{i-1,j}\right) & \text{when } i < j < \infty \\ \beta_p^{i,j} - \beta_p^{i-1,j} & \text{when } i < \infty,\ j = \infty \end{cases}$$

- <u>When $j < \infty$</u> (finite death), $\left(\beta_p^{i,j-1} - \beta_p^{i,j}\right)$ counts the # of independent classes that are born at or before X_i but die entering X_j.
  - The second term $\left(\beta_p^{i-1,j-1} - \beta_p^{i-1,j}\right)$ counts the # of independent classes that are born at or before X_{i-1} and die entering X_j.
  - So, the two terms with subtraction counts the # of independent classes born at X_i and dying entering X_j

* **When $j=\infty$** (classes never die)
  - The term $(\beta_p^{i,j} - \beta_p^{i-1,j})$ exactly counts the # of independent classes that are born at X_i and exist all the way to X_∞.

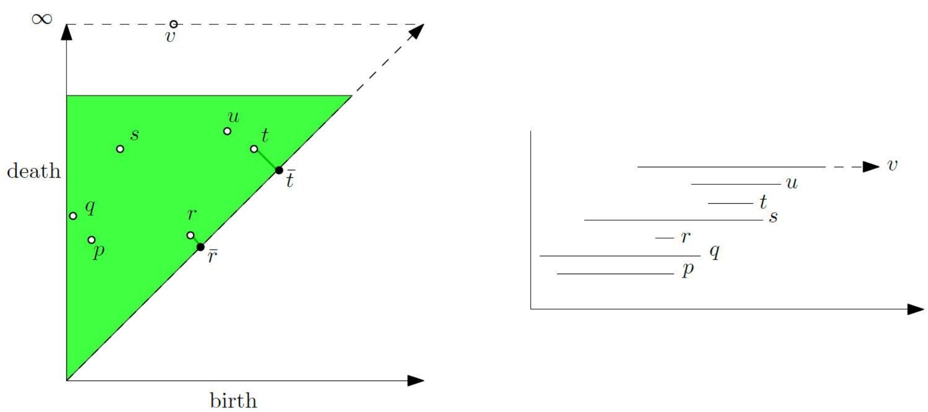
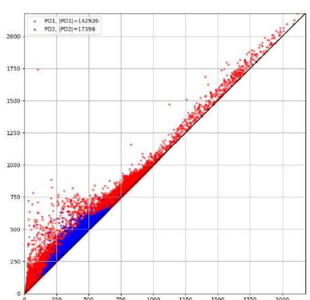
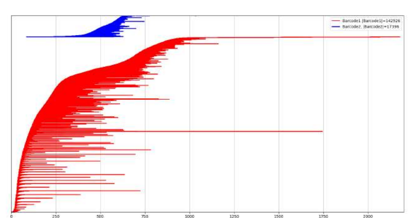


Figure 3.7: Persistence diagram and the corresponding barcode.





Def (Persistence diagram (PD)): $\mathrm{Dgm}_p(\mathcal{F})$ is a set of points (a_i, a_j) with multiplicity $\mu_p^{i,j}$, $i<j$.
We add diagonal $\Delta: \{(x,x)\}$ with infinite multiplicity.

Def (Bar code): Alternate representation:

(a_i, a_j): line segment $[a_i, a_j]$.
(a_i, ∞): ray emanating from a_i.

Def (class persistence): $\text{pers}([c]) = a_j - a_i$,
$\text{pers}([c]) = \infty$ if $[c]$ never dies

Fact: 1. if $[c]$ has persistence s, then point representing it is within $s/\sqrt{2}$ distance from Δ

2. Since $i < j$, all points lie on or above Δ

3. If m_i is multiplicity of (a_i, ∞) in Dgm_p then $\sum_i m_i = \dim H_p(X)$.

Theorem 16. *For every pair of indices* $0 \leq k \leq \ell \leq n$ *and every* p, *the* p*-th persistent Betti number satisfies* $\beta_p^{k,\ell} = \sum_{i \leq k} \sum_{j > \ell} \mu_p^{i,j}$.

Stability of Persistence Diagram

- How stable $Dgm_p(\mathcal{F}_f)$ is wrt perturbations of f?
- Let $Dgm_p(\mathcal{F}_f)$ and $Dgm_p(\mathcal{F}_g)$ be two PDs with two filtration functions.
- Goal is to define a distance $d(Dgm_p(\mathcal{F}_f), Dgm_p(\mathcal{F}_g))$ and determine its relation to the difference

$$\|f-g\|_\infty = \max_x |f(x)-g(x)|$$

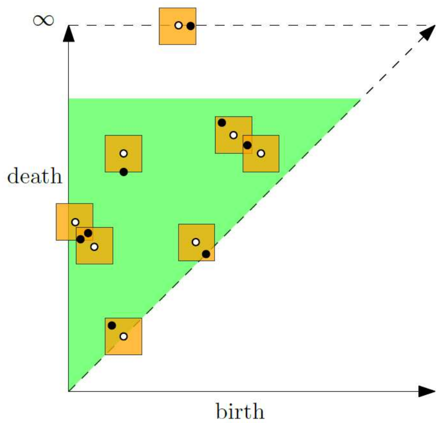
Def. (Bottleneck distance):

- $\Pi = \{\pi\}$ set of all bijections $\pi: Dgm_p(\mathcal{F}_f) \to Dgm_p(\mathcal{F}_g)$
- Bijections are possible because of Δ (diagonal)

$$d_b(Dgm_p(\mathcal{F}_f), Dgm_p(\mathcal{F}_g)) = \inf_{\pi \in \Pi} \sup_{x \in Dgm_p(\mathcal{F}_f)} \|x - \underbrace{\pi(x)}_{= y}\|_\infty$$

Δ

$\max(|x_1 - x_2|, |y_1 - y_2|)$



Fact d_b is a metric on space of PDs.

- $d_b(X, Y) = 0$ iff $X = Y$
- $d_b(X, Y) = d_b(Y, X)$
- $d_b(X, Y) \leq d_b(X, Z) + d_b(Z, X)$

* No classes are born and die at the same time. Otherwise d_b becomes a pseudometric because the first axiom fails.

Theorem (Stability): If $f, g : X \to \mathbb{R}$ are two functions giving two filtrations $\mathcal{F}_f$ and $\mathcal{F}_g$. Then,

$$d_b(Dgm_p(\mathcal{F}_f), Dgm_p(\mathcal{F}_g)) \leq \|f - g\|_\infty .$$

Computing d_b

A, B non-diagonal points of $Dgm(\mathcal{F}_f)$ and

- A, B non-diagonal points of $Dgm(F_f)$ and $Dgm(F_g)$
- $a \in A$, $\bar{a}$ = nearest pt. on Δ, similarly, define $\bar{b}$ for $b \in B$.
- $\bar{A} = \{\bar{a}\}$, $\bar{B} = \{\bar{b}\}$, $\tilde{A} = A \cup \bar{B}$, $\tilde{B} = B \cup \bar{A}$
- Want to match $\tilde{A}$, $\tilde{B}$ bijectively; $|\tilde{A}| = |\tilde{B}| = n$
- $$d_b = \min_{\pi \in \Pi} \max_{a \in \tilde{A}, \pi(a) \in \tilde{B}} \|a - \pi(a)\|_\infty$$
- d_b must equal $\|a - b\|_\infty$ for some $a \in \tilde{A}$, $b \in \tilde{B}$.
- We consider complete matchings between $\tilde{A}$ & $\tilde{B}$ and do a binary search to find the optimal matching.
- $O(n^2)$ possible distances $\delta_0, \delta_1, \delta_2 \ldots \delta_{n'}$ between pairs of points in $\tilde{A}$, $\tilde{B}$.

Algorithm 1 BOTTLENECK($\mathrm{Dgm}_p(\mathcal{F}_f)$, $\mathrm{Dgm}_p(\mathcal{F}_g)$)

Input:
Two persistent diagrams $\mathrm{Dgm}_p(\mathcal{F}_f)$, $\mathrm{Dgm}_p(\mathcal{F}_g)$
Output:
Bottleneck distance $\mathsf{d}_b(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g))$

Compute sorted distances $\delta_0 \le \delta_1 \le \cdots \le \delta_{n'}$ from $\mathrm{Dgm}_p(\mathcal{F}_f)$ and $\mathrm{Dgm}_p(\mathcal{F}_g)$
$\ell := 0$; $u = n'$
while $\ell < u$ **do**
    $i := \lfloor \frac{(u+\ell)}{2} \rfloor$; $\delta := \delta_i$
    Compute graph $G = (\tilde{A} \cup \tilde{B}, E)$ where $\forall e \in E$, $weight(e) \le \delta$
    if $\exists$ complete matching in G **then**
        $u := i$
    else
        $\ell := i$
    end if
end while
Output δ

- A complete matching in G is a set of n edges so that every vertex in $\tilde{A}$ & $\tilde{B}$ is incident on exactly one edge.
- Complete matching can be computed in $O(n^{2.5})$ time (Hopcroft-Tarjan)
- Total time $O(n^{2.5} \log n)$.