Namen: Tim Suchan, Nico Schum, Linus Palm

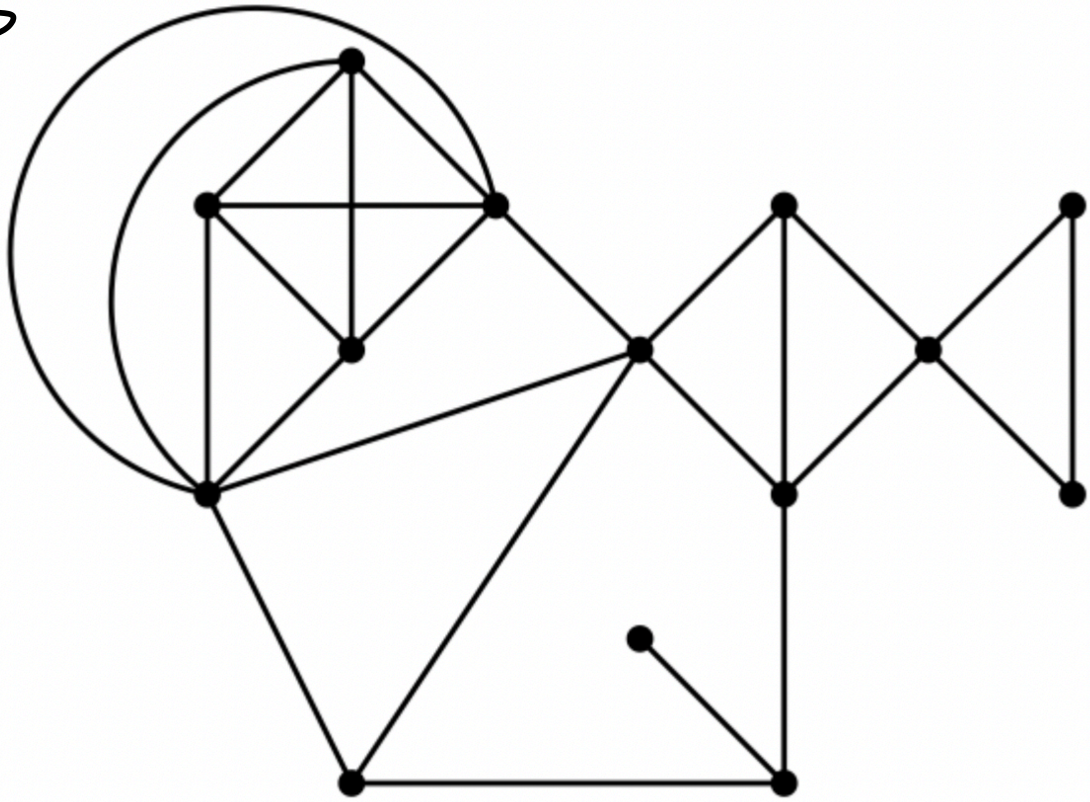
17/20

Ex. 1 4/4

Samstag, 21. Dezember 2024 16:18

1-strongly connected components

G

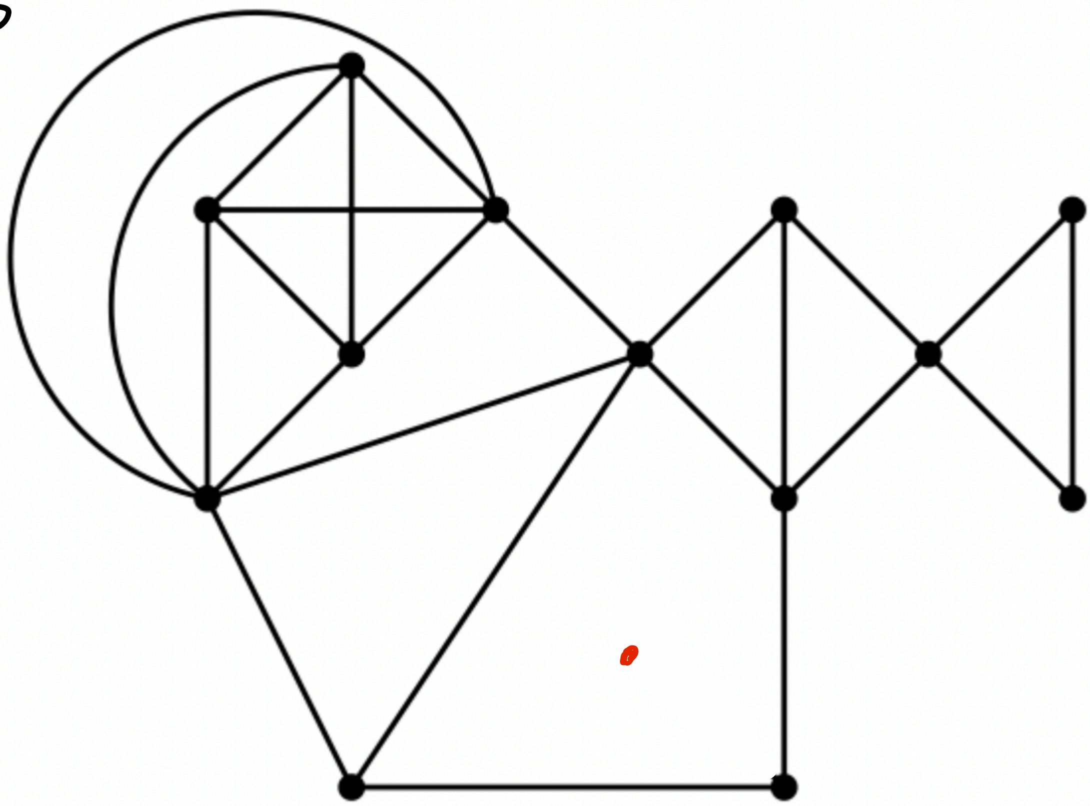


Graph itself, since it is 1-connected.

✓

2 - strongly connected components

G'

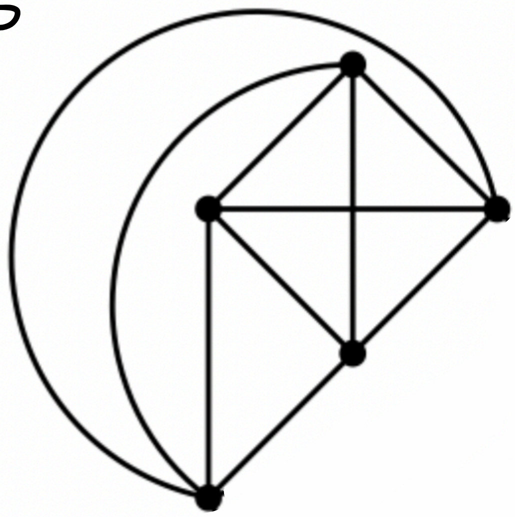


edge towards

Remove vertex with just one edge. Result is maximum 2-connected induced subgraph. ✓

3 - strongly connected components

G''



✓

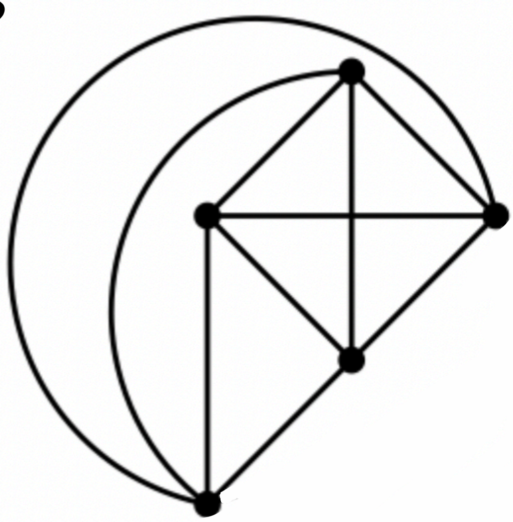
Every other vertex of the graph is its own strongly-connected component, s.t. they form a partition of V.

Remove all vertices v_i with degree$(v_i) = 2$. Min-Cut(G'') is now ≥ 3.

✓

4- strongly connected components

G''



s. a.

No changes, because G'' is also maximal 4-connected induced subgraph.

No 5-strongly connected component possible, since there exists no connected component in G where for all vertices v_i holds degree(v_i) ≥ 5.

Ex 2

4/4

a)

Let $\bar{S} = S_1 \cup S_2$. Now for any set of centers C:

$$cost(\bar{S}, C) = \sum_{s \in \bar{S}} (w_1(s) + w_2(s)) \times d(s, C) \quad ✓$$

$$cost(\bar{S}, C) = \sum_{s \in S_1} w_1(s) \times d(s, C) + \sum_{s \in S_2} w_2(s) \times d(s, C) \quad ✓$$

And now since S_1 and S_2 are (k, ϵ)-coresets for P_1, P_2

$$(1 - \epsilon)(\sum_{p \in P_1} w(p) \times d(p, C) + \sum_{p \in P_2} w(p) \times d(p, C)) \leq$$

$$\sum_{s \in S_1} w_1(s) \times d(s, C) + \sum_{s \in S_2} w_2(s) \times d(s, C)$$

$$\leq \cancel{\geq} (1 + \epsilon)(\sum_{p \in P_1} w(p) \times d(p, C) + \sum_{p \in P_2} w(p) \times d(p, C)) \quad ✓$$

And with P_1, P_2 *disjoint*: ✓

$$(1 - \epsilon) cost(P, C) = (1 - \epsilon) \sum_{p \in P} w(p) \times d(p, C) \leq$$

$$\sum_{s \in S_1} w_1(s) \times d(s, C) + \sum_{s \in S_2} w_2(s) \times d(s, C) = \sum_{s \in \bar{S}} (w_1(s) + w_2(s)) \times d(s, C) = cost(\bar{S}, C)$$

$$\leq \cancel{\geq} (1 + \epsilon) \sum_{p \in P} w(p) \times d(p, C) = (1 + \epsilon) cost(P, C) \quad ✓$$

2/4

b)

In a 1-dimensional metric space let $P_1 = \{1, 2, 4\}$, $P_2 = \{5, 6, 8, 9\}$, $S_1 = \{1, 2\}$, $S_2 = \{8, 9\}$, $k = 4$, $b = 2$. Now P_2 is at least a (4,5)-coreset since the points 5 and 6 can be reached by triangulation through 8. f Following the same argument P_1 is at least a (4, 2)-coreset. Therefore both P_1 and P_2 are (4, 5)-coresets. f (s.b.) Now let $c = \{1, 5, 8, 100\}$ be a set of four cluster-centers. And $P = P_1 \cup P_2 = \{1, 2, 4, 5, 6, 8, 9\}$, $S = S_1 \cup S_2 = \{1, 2, 8, 9\}$ then obviously the clusters for S will be $C_s = \{\{1, 2\}\{8, 9\}\}$ with centers $c_s = \{1, 8\}$ and overall cost 2. Now since $b = 2$ at least one of the points in P has to be put into a cluster with the center 100 ✓ therefore $cost(P, C) \geq min_{p \in P} ||p - 100||_2 = 91$ and with $\frac{91}{2} > 5$ S can not be a (4,5)-coreset for P.

You found the problem, but forgot to weight points in S. Your coresets also aren't correct:

Ass. centers $C = \{1, 2, 8, 9\}$.

$\Rightarrow cost(S_1, C) = 0 = cost(S_2, C)$, but

$cost(P_1, C) = 2 > (1 + \varepsilon) \cdot 0 \quad \forall \varepsilon$ (especially $2 > 6 \cdot 0$)

and $cost(P_2, C) = 5 > (1 + \varepsilon) \cdot 0 \quad \forall \varepsilon$

Ex. 3

a)

4/4

The idea is to partition the space into grid cells of side length proportional to $\epsilon \cdot L$, where L is the length of the longest side of the bounding box enclosing the point set. Then we select one representative point from each non-empty grid cell. These representative points form the coreset. We first compute the minimum and maximum x- and y-coordinates of the points in P. The bounding box enclosing P is $[x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$. Let:

$$L = \max(x_{\max} - x_{\min}, y_{\max} - y_{\min})$$

denote the length of the bounding box's longest side. We then divide the bounding box into a grid of cells with side length:

$$\ell = \frac{\epsilon}{\sqrt{2}} \cdot L.$$

Each grid cell ensures that the distance between any two points within the same cell is at most $\epsilon \cdot L$. Assign each point $p \in P$ to a grid cell based on its coordinates. For each non-empty grid cell, select one point p from P in that cell as a representative point. Let S denote the set of all selected points, the we return S as the ϵ-coreset.

b)

2/2

Computing the bounding box and assigning points to grid cells takes $O(n)$ time, since we have to iterate over all points to find the maximum and minimum. Collecting representative points takes $O(1/\epsilon^2)$ time. Therefore, the total time complexity is $O(n)$. (✓)

$O(1/\epsilon^2)$

c)

1/2

For any center c, the true cost of the center for the full set is:

$$\text{cost}(P, c) = \max_{p \in P} \|p - c\|.$$

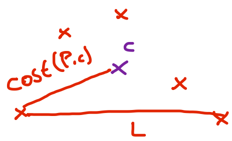
Using the grid, the distance from any point $p \in P$ to its representative $s \in S$ satisfies:

$$\|p - s\| \leq \sqrt{2} \cdot \ell.$$

Thus, the approximation error is bounded by:

$$\sqrt{2} \cdot \ell = \epsilon \cdot L \leq \epsilon \cdot \text{cost}(P, c),$$

f



ensuring that:

$$(1 - \epsilon) \cdot \text{cost}(P, c) \leq \text{cost}(S, c) \leq (1 + \epsilon) \cdot \text{cost}(P, c).$$

Since the grid uses cells of side length ℓ, the total number of cells is proportional to the area of the bounding box divided by the cell area:

$$\text{number of cells} = O\left(\frac{L^2}{\ell^2}\right) = O\left(\frac{1}{\epsilon^2}\right).$$

Each non-empty grid cell contributes at most one representative point to S, so the size of the coreset is $O(1/\epsilon^2)$.