Final Exam, Algorithms II 2023-2024

Do not turn the page before the start of the exam. This document is double-sided, has 12 pages, the last ones possibly blank. Do not unstaple.

- The exam consists of three parts. The first part consists of multiple-choice questions, the second part consists of a short open question, and the last part consists of four open-ended questions.
- For the open-ended questions, your explanations should be clear enough and in sufficient detail that a fellow student can understand them. In particular, do not only give pseudocode without explanations. A good guideline is that a description of an algorithm should be such that a fellow student can easily implement the algorithm following the description.
- You are allowed to refer to material covered in the lectures including algorithms and theorems (without reproving them). You are however *not* allowed to simply refer to material covered in exercises/homework.

Good luck!

Problem 1: Multiple Choice Questions (24 points)

For each question, select the correct alternative. Note that each question has **exactly one** correct answer. Wrong answers are not penalized with negative points.

1a. General Knowledge *(8 points).* Select the correct answer.

A. Suppose that a randomized algorithm has success probability of $o(1)$. Then we can boost its success probability to a constant by doing at most $\text{poly}(n)$ repetitions.

B. Solving the standard Linear Program for Vertex Cover on general graphs, requires exponential running time in the worst case.

C. Consider a bipartite graph G and the standard matching polytope for G. There exists an assignment of weights to the edges of G that makes the matching polytope non-integral.

D. Consider an online minimization problem and an online algorithm $\mathcal{A}$. If for every instance I it is true that $\mathcal{A}(I) \leqslant 2 \cdot \text{OPT}(I)$ then $\mathcal{A}$ is 2-competitive.

E. Finding the maximum cut of a graph cannot be reduced to the maximization of a submodular function.

Solution. The correct answer is D.

- Option A is incorrect because a success probability of $1/2^n$ (which is $o(1)$) can not be boosted to a constant only by $\text{poly}(n)$ repetitions.
- Option B is false because any Linear Program can be solved in polynomial time.
- Option C is also incorrect. The weight vector does not influence the integrality of the bipartite matching polytope in any way. It only determines the minimization direction.
- Option D is correct by the definition of the competitive ratio.
- Option E is also false because the function that maps a cut to the number of edges that cross it is a submodular function (see Notes of Lecture 20, Section 1.1).

1b. Matroids *(8 points)*. Consider a ground set E and a family of independent sets $\mathcal{I}$. Which one of the following statements is correct?

A. Suppose that E is the vertex set of some graph G and $\mathcal{I}$ is the collection of all subsets of vertices that share no edges between them. Then, $(E, \mathcal{I})$ is a matroid.

B. Let $w : E \to \mathbb{R}_{>0}$ be a non-negative weight function. We define the weight of a subset of E to be the sum of weights of its elements. If for *any* such function w there exists a maximum cardinality independent set that has maximum weight, then $(E, \mathcal{I})$ is a matroid.

C. Let $\mathcal{I}_{100}$ denote the subcollection of $\mathcal{I}$ that contains all of its sets that have cardinality at most 100 (that is $\mathcal{I}_{100} = \{X \in \mathcal{I} : |X| \leqslant 100\}$). If $(E, \mathcal{I})$ is a matroid then $(E, \mathcal{I}_{100})$ is also a matroid.

D. Suppose that $\mathcal{M} = (E, \mathcal{I})$ is a matroid. Then there exist two bases A and B with $A \neq B$ such that either $A \cup B$ is a base or $A \cap B$ is a base.

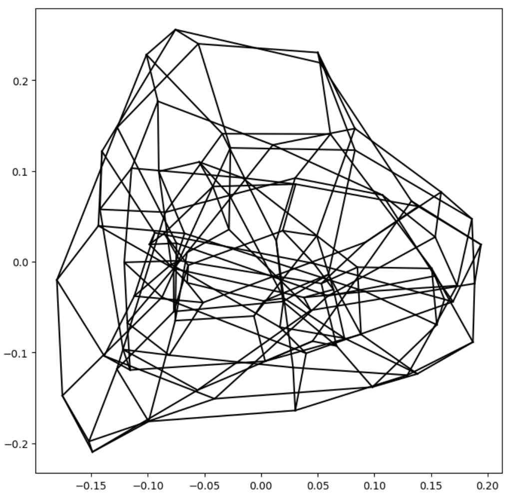
E. Consider a partitioning of E into E_1 and E_2. Let $\mathcal{I}'$ be all the subsets of E that have at least k_1 elements from E_1 and at least k_2 elements from E_2. Then, $(E, \mathcal{I}')$ is a matroid.

Solution. The correct answer is C.

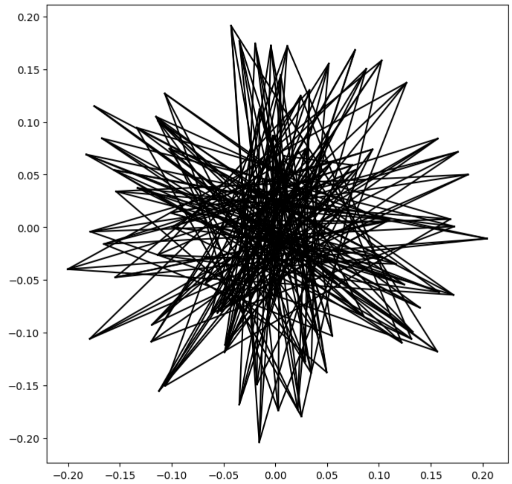
- Option A is wrong because the augmentation property does not hold (e.g. on a star graph the neighbors of a star node is an independent set and the central node is an independent set but you can not augment it with any one of its neighbors). Also, if it was true then P would be equal to NP because we could solve the maximum independent set problem in polynomial time (as a matroid maximization problem).
- Option B is false. Consider $E = \{a, b\}$ and $\mathcal{I} = \{\{a\}, \{a, b\}\}$. The set $\{a, b\}$ will always have bigger weight than $\{a\}$ under any non-negative weight function, however $(E, \mathcal{I})$ is not a matroid since $\mathcal{I}$ is not downwards closed.
- Option C is correct, $(E, \mathcal{I}_{100})$ is a truncated matroid (see Notes of Lecture 1, Section 3.1.4).
- Option D is incorrect, both $A \cup B$ and $A \cap B$ will have different sizes than $|A|$ and therefore can not be bases.
- Option E is false because $\mathcal{I}'$ is not downwards closed.

1c. Spectral Graph Theory *(8 points).*

Let G be a connected d-regular graph with n vertices and let M be its normalized adjacency matrix. Recall that the normalized adjacency matrix is equal to $\frac{1}{d}A$, where A is the adjacency matrix of the graph. Let $1 = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ be the eigenvalues of M and $e_1, \ldots e_n \in \mathbb{R}^n$ be the corresponding unit-norm eigenvectors.
Each of the two images below were generated using the following procedure: Select two eigenvectors of M, let them be a and b. Place the i-th vertex of the graph at the position $(a(i), b(i))$ where $a(i)$ is the value of the i-th dimension of eigenvector a (correspondigly for b). If vertices i and j are connected with an edge, draw a line between $(a(i), b(i))$ and $(a(j), b(j))$.



(a)



(b)

Select the correct answer.

A. Figure (a) was generated using the eigenvectors e_{n-1} and e_n, while (b) was generated from e_2 and e_3.

B. Figure (a) was generated using the eigenvectors e_{n-1} and e_n, while (b) was generated from e_1 and e_2.

C. Figure (a) was generated using the eigenvectors e_2 and e_3, while (b) was generated from e_{n-1} and e_n.

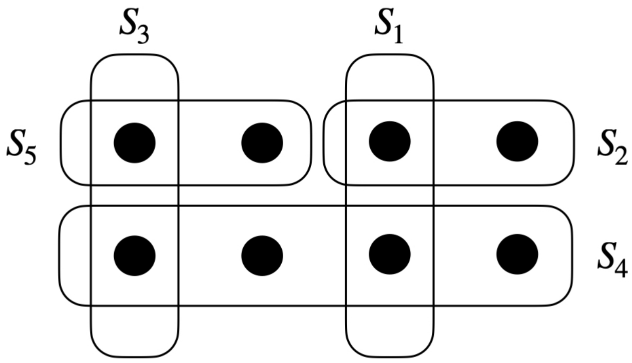
D. Figure (a) was generated using the eigenvectors e_1 and e_2, while (b) was generated from e_{n-1} and e_n.

E. The selection of eigenvectors does not influence the embedding in a systematic way.

Solution. The correct answer is C. When we draw the vertices using eigenvectors that correspond to large eigenvalues of the normalized adjecency matrix, neighboring vertices tend to get mapped to nearby positions whereas the oposite happens with small eigenvalues (see the Notes of Lecture 23).

Problem 2: Short Open Question (11 points)

Consider a coverage problem on a ground set E of 8 elements and a collection $\mathcal{U}$ of 5 sets as depicted below.



We define the coverage function $f : \{0,1\}^5 \to \mathbb{N}$ as

$$f(x_1, x_2, x_3, x_4, x_5) = \left| \bigcup_{\substack{j=1,\ldots,5 \\ \text{s.t. } x_j=1}} S_j \right| .$$

In other words, f accepts a binary vector x with 5 terms, which denotes a subset of the family $\mathcal{U}$. The i-th set is included in the collection described by x if and only if $x_i = 1$. For example, $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 1, 0, 1)$ corresponds to the collection $\{S_3, S_5\}$. The coverage function then maps each such vector to the cardinality of the union of its sets. In the previous example, $f(0, 0, 1, 0, 1) = 3$, because the union of S_3 and S_5 contains 3 elements.

Let $\hat{f} : [0,1]^5 \to \mathbb{R}$ denote the Lovász extension of f. Calculate $\hat{f}(0.2, 0.9, 0, 0.8, 0.1)$.

Solution. Let us first rearrange the coordinates in non-increasing order: let z be the vector $z = (0.9, 0.8, 0.2, 0.1, 0)$ and let $S'_1 = S_2$, $S'_2 = S_4$, $S'_3 = S_1$, $S'_4 = S_5$, $S'_5 = S_3$. Also let $f' : \{0,1\} \to \mathbb{N}$ be the coverage function corresponding to the set system $S'_1, \ldots, S'_5$, and let $z_6 = 0$ for notation's sake. Note that $f(0,0,0,0,0) = f'(0,0,0,0,0) = 0$. Then we have

$$\begin{aligned}
\hat{f}(0.2, 0.9, 0, 0.8, 0.1) &= \hat{f}'(z) \\
&= \sum_{i=1}^{5} (z_i - z_{i+1}) f'(\{1, \ldots, i\}) \\
&= 0.1 \cdot |S_2| + 0.6 \cdot |S_2 \cup S_4| + 0.1 \cdot |S_2 \cup S_4 \cup S_1| + 0.1 \cdot |S_2 \cup S_4 \cup S_1 \cup S_5| \\
&= 0.1 \cdot 2 + 0.6 \cdot 6 + 0.1 \cdot 6 + 0.1 \cdot 8 \\
&= 5.2 \,.
\end{aligned}$$

Problem 3: Estimating the mean (20 points)

Consider n values $x_1, \ldots, x_n \in [-1, 1]$ between -1 and 1. In this problem, our goal is to approximate their mean $\mu = \sum_{i=1}^{n} x_i/n$ by sampling.

A. *(3 points)* One estimator for μ is the following: sample one of the n points, let it be x_i, uniformly at random. Output $\hat{\mu} = x_i$. Show that this estimator is unbiased, that is $\mathbb{E}[\hat{\mu}] = \mu$.

B. *(5 points)* Show that for the variance of $\hat{\mu}$ the following is true: $\mathrm{Var}(\hat{\mu}) \leqslant 1$.

C. *(12 points)* Given $\varepsilon > 0$, use the unbiased estimator $\hat{\mu}$ to construct an estimator $\hat{\mu}^+$ such that the following holds with probability at least 0.9,

$$|\mu - \hat{\mu}^+| \leqslant \varepsilon\,.$$

Your estimator $\hat{\mu}^+$ should use an asymptotically optimal number of samples.

Solution.

A. By the definition of the expectation,

$$\mathrm{E}[\hat{\mu}] = \sum_{i=1}^{n} Pr[\hat{\mu} = x_i] \cdot x_i = \sum_{i=1}^{n} \frac{x_i}{n} = \mu.$$

B. For the variance we have the following inequality,

$$\mathrm{Var}[\hat{\mu}] = \mathrm{E}[\hat{\mu}^2] - \mathrm{E}[\hat{\mu}]^2 \leqslant \mathrm{E}[\hat{\mu}^2].$$

Since $\hat{\mu} \in [-1, 1]$, $\hat{\mu}^2 \leqslant 1$. The bound on the variance follows.

C. For $t \in \mathrm{N}$, let $\hat{\mu}_1, \ldots, \hat{\mu}_t$ be t independent realizations of the estimator $\hat{\mu}$. Define,

$$\hat{\mu}^+ = \frac{1}{t} \sum_{i=1}^{t} \hat{\mu}_i.$$

The estimator μ^+ remains unbiased by linearity of expectation. Moreover, the variance decreases by a factor of $\frac{1}{t}$,

$$\mathrm{Var}[\hat{\mu}^+] = \frac{1}{t^2} \sum_{i=1}^{t} \mathrm{Var}[\hat{\mu}_i] \leqslant \frac{1}{t}.$$

For the last inequality, we used the bound from Part **B**. By Chebyshev's inequality,

$$Pr\left[|\hat{\mu}^+ - \mu| \geqslant \epsilon\right] \leqslant \frac{\mathrm{Var}[\hat{\mu}^+]}{\epsilon^2} \leqslant \frac{1}{\epsilon^2 t}.$$

Thus, choosing $t = \frac{10}{\epsilon^2}$ gives the required bound on the probability.

Problem 4: A Special Vertex Cover (15 points)

After taking the Algorithms II course, your friend Bob has become obsessed with the vertex cover problem. In his latest attempt to tackle this question, he has achieved the following:

For a graph $G = (V, E)$, Bob has constructed a distribution $\mathcal{D}$ over subsets of vertices $T \subseteq V$ of size k ($|T| = k$), such that if you sample a vertex set from $\mathcal{D}$, then for every edge $e = (u, v) \in E$ it holds that

$$\Pr[u \in T \text{ or } v \in T] \geqslant 0.99\,.$$

That is, if we select a random set T from $\mathcal{D}$ then we are *almost* but not quite a vertex cover since every edge is likely but not necessarily covered.

Your friend needs your help to analyze the size of a real vertex cover solution. Specifically, you have to prove that, *assuming the graph G is bipartite*, his construction implies that the minimum vertex cover of G has size at most $\frac{k}{0.99}$.

1st Solution. Let p_u be the marginal probability that u is picked when sampling a vertex set T from the distribution $\mathcal{D}$, that is

$$p_u = \Pr_{T \sim \mathcal{D}}[u \in T].$$

By the union bound, for every edge (u, v) the following is true

$$p_u + p_v \geqslant \Pr_{T \sim \mathcal{D}}[u \in T \text{ or } v \in T] \geqslant 0.99,$$

or equivalently

$$\frac{p_u}{0.99} + \frac{p_v}{0.99} \geqslant 1. \tag{1}$$

Recall the LP relaxation of the Vertex Cover problem:

$$\begin{aligned} \text{Minimize} \quad & \sum_{u \in V} x_u \\ \text{Subject to} \quad & x_u + x_v \geqslant 1 \quad \forall (u, v) \in E \end{aligned}$$

Notice that the assignment $x_u = p_u/0.99$ is feasible because of (1). Let $VC(G)$ denote the size of the minimum vertex cover of G. We know that the Vertex Cover LP is integral on bipartite graphs and that the optimal solution is smaller or equal to any feasible solution, therefore

$$VC(G) \leqslant \sum_{u \in V} \frac{p_u}{0.99}.$$

Finally, we can upper bound the size of the feasible solution that we constructed as

$$\sum_{u \in V} \frac{p_u}{0.99} = \frac{1}{0.99} \cdot \sum_{u \in V} 1 \cdot p_u = \frac{1}{0.99} \mathbb{E}_{T \sim D}[|T|] = \frac{k}{0.99}.$$

By the former two equations we get that

$$VC(G) \leqslant \frac{k}{0.99}.$$

2nd Solution. Let $\mathcal{M}$ be a maximum matching of G and $V(\mathcal{M})$ be the set of vertices of $\mathcal{M}$. We will calculate the expected number of vertices of $V(\mathcal{M})$ that belong to a sampled vertex $T \sim \mathcal{D}$.

$$\mathbb{E}_{T\sim\mathcal{D}}\left[|T \cap V(\mathcal{M})|\right] \geqslant \sum_{(u,v)\in\mathcal{M}} 1 \cdot \Pr_{T\sim\mathcal{D}}[u \in T \text{ or } v \in T] \geqslant 0.99 \cdot |\mathcal{M}|.$$

Since $\mathbb{E}[|T \cap V(\mathcal{M})|] \leqslant \mathbb{E}[|T|]$, we get that

$$\mathbb{E}_{T\sim\mathcal{D}}[|T|] = k \geqslant 0.99 \cdot |\mathcal{M}|.$$

Finally, from König's theorem we know that $|\mathcal{M}| = VC(G)$. Therefore,

$$VC(G) \leqslant \frac{k}{0.99}.$$

Problem 5: Alice and Bob (15 points)

In this problem, your goal is to design an efficient communication protocol where Alice sends a short message to Bob, allowing him to output the correct answer with good probability. The setup is as follows. Alice is given a vector $a \in \{1, 2, \ldots, n\}^n$. Bob is given a vector $b \in \{1, 2, \ldots, n\}^n$ and an index i. We emphasize that Alice has no information about Bob's input and Bob has no information regarding Alice's input.

The protocol is as follows. As a function of her input, Alice randomly selects a message m, consisting of $O(\log n)$ bits, and sends it to Bob. Then, based on his input and the message m, Bob outputs "Identical" or "Far". Your task is to explain and analyze a strategy of Alice and Bob (also called a communication protocol) so that the output of Bob satisfies the following guarantees:

- If $a_i = b_i$ then Bob outputs "Identical" with probability at least 2/3.
- If $|a_i - b_i| > \frac{\|a-b\|_2}{10}$ then Bob outputs "Far" with probability at least 2/3.

Here, the probability is over the randomness of Alice, i.e., over the randomly selected message m. We further remark that the length of the message m should always be at most $O(\log n)$ bits and Bob may output anything if $a_i \neq b_i$ and $|a_i - b_i| \leqslant \frac{\|a-b\|_2}{10}$.

Solution. Let $t = 10^6$ be a parameter. Alice samples t random and independent 4-wise hash functions $\sigma_r : [n] \to \{\pm 1\}$ for $r \in [t]$ that assign signs of each of the n coordinates. Then, for each $r \in [t]$ she computes a sketch S_r^A of her own vector, defined as $S_r^A := \sum_{j=1}^n a_j \sigma_r(j)$. Then, she sends $(\sigma_r)_{r\in[t]}$ and $(S_r^A)_{r\in[t]}$ to Bob. This message consists of $O(\log n)$ bits, since each σ_r and S_r^A require $O(\log n)$ bits to represent and t is a constant.

Upon receiving Alice's message, for each $r \in [t]$ Bob computes a sketch of his own vector $S_r^B := \sum_{j=1}^n b_j \sigma_r(j)$, and a "shifted" sketch $Q_r := \sigma_r(i) \cdot (S_r^A - S_r^B)$, and averages it over $r \in [t]$ to get $Q := \frac{1}{t}\sum_{r=1}^t Q_r$. Bob also computes the difference of his and Alice's sketches $W_r = S_r^A - S_r^B$, and averages the square of these over $r \in [t]$ to get $W := \frac{1}{t}\sum_{r=1}^t W_r^2$. Then, Bob outputs "Identical" if $Q^2 \leqslant W/100$, and he outputs "Far" otherwise.

Finally we discuss correctness. We claim that Q is an unbiased estimator of $a_i - b_i$: note that by 4-wise independence of the σ_r's (and in particular their 2-wise independence), for any $r \in [t]$ one has

$$\mathbb{E}[Q] = \mathbb{E}[Q_r] = \sum_{j=1}^n \mathbb{E}\left[\sigma_r(i)\cdot\sigma_r(j)\right]\cdot(a_j - b_j) = a_i - b_i\,.$$

We cannot use a Chernoff bound to show that Q_r concentrates around its mean because the signs $(\sigma_r(j))_{j\in[n]}$ are not independent. Instead we apply Chebyshev's inequality to the average Q of the Q_r's. To do so, first note that

$$\mathrm{Var}\left[Q_r\right] \leqslant \mathbb{E}\left[Q_r^2\right] = \sum_{j,h\in[n]} \mathbb{E}\left[\sigma_r^2(i)\cdot\sigma_r(j)\cdot\sigma_r(h)\right]\cdot(a_j - b_j)(a_h - b_h) = \|a-b\|_2^2\,.$$

Hence, we have

$$\mathrm{Var}\left[Q\right] \leqslant \frac{1}{t}\|a-b\|_2^2\,,$$

since the σ_r's are independent of each other. Similarly, we have that W is an unbiased estimator of $\|a-b\|_2^2$: for any $r \in [t]$ we have

$$\mathbb{E}[W] = \mathbb{E}[W_r^2] = \sum_{j,h\in[n]} \mathbb{E}\left[\sigma_r(j)\cdot\sigma_r(h)\right]\cdot(a_j - b_j)(a_h - b_h) = \|a-b\|_2^2\,,$$

and

$$\mathrm{Var}\,[W] = \frac{1}{t}\mathrm{Var}[W_r^2] = \frac{1}{t}\left(\mathbb{E}[W_r^4] - \mathbb{E}[W_r^2]^2\right) \leqslant \frac{2}{t}\|a-b\|_2^4\,,$$

where the last inequality follows by the same case analysis that we did for the AMS algorithm. By Chebyshev's inequality, we get that with probability at least $98/100 \geqslant 2/3$ one has

$$|Q - (a_i - b_i)| \leqslant \frac{\|a-b\|_2}{100} \quad \text{and} \quad |W - \|a-b\|_2^2| \leqslant \frac{\|a-b\|_2^2}{100}\,.$$

We then have two cases to distinguish: if $a_i = b_i$ then $Q^2 \leqslant \|a-b\|_2^2/10000 \leqslant W/100$, so Bob outputs "Identical"; if $|a_i - b_i| > \|a-b\|_2^2/10$ then $Q^2 \geqslant \|a-b\|_2^2/20 > W/100$, so Bob outputs "Far".

Alternative solution. We briefly outline an alternative solution, which is more aligned with what many students attempted. Alice computes 3 independent AMS sketches S_1, S_2, S_3 of her vector a, and sends them to Bob together with the random bits r_1, r_2, r_3 used to construct these AMS transforms. All in all, this takes $O(\epsilon^{-2}\log 1/\delta \log n)$ bits, so choosing $\epsilon, \delta \in (0,1)$ to be small enough constants satisfies the communitcation complexity requirement.

Then Bob uses S_1 and r_1 to compute an AMS sketch of $a - b$ (this is possible since the AMS sketch is a linear sketch), thus obtaining a $(1 \pm \epsilon)$-approximation W to $\|a-b\|_2$ with probability $1 - \delta$ (note that with here W is an estimator of the ℓ_2 norm of $a - b$ without squaring). Then, Bob uses S_2 and r_2 to compute an AMS sketch of the vector $a - b + W \cdot e_i$ (again, this is possible since the AMS sketch is a linear sketch), where e_i is the i-th coordinate vector, thus obtaining a $(1 \pm \epsilon)$-approximation Q_{+1} to $\|a - b + W/10 \cdot e_i\|_2^2$ with probability $1 - \delta$. Analogously, from S_3 and r_3 Bob computes a $(1 \pm \epsilon)$-approximation Q_{-1} to $\|a - b - W/10 \cdot e_i\|_2^2$ with probability $1 - \delta$. If $a_i = b_i$, then $Q_{+1} = Q_{-1} = (1 + 1/100)\|a-b\|_2^2$. If $\xi \cdot (a_i - b_i) > \|a-b\|_2/10$ for $\xi \in \{\pm 1\}$, then $Q_\xi = \|a-b\|_2^2 - 2(a_i - b_i)\xi \cdot W/10 + W^2/100 \leqslant (1 - (1 - 2\epsilon)/100)\|a-b\|_2^2$. He can then distinguish the two cases.

Problem 6: Recycling (15 points)

The administration of Lausanne decided to install $k \geqslant 1$ recycling centers around the city. We represent Lausanne as an undirected and connected graph $G := (V, E)$, in which every vertex $v \in V$ represents an area of the city and every edge $e \in E$ connects a pair of distinct areas. You can assume that there are no self-loops or parallel edges in G. Furthermore, every area $v \in V$ has a potential $P(v) \in \mathbb{R}_{\geqslant 0}$ of reducing pollution by recycling.

The administration has to choose a subset $C \subseteq V$ of $|C| = k$ areas in which recycling centers will be installed. They ordered a study which led to the following insight: Assume that the closest recycling center to an area $v \in V$ is $C(v) \in C$, and let $\operatorname{dist}(v, C(v))$ be the distance between v and $C(v)$ in the graph G. In particular, if $v = C(v)$, i.e., v itself contains a recycling center, then $\operatorname{dist}(v, C(v)) = 0$. The study proved that the pollution reduction from area v is given by $P(v)/2^{\operatorname{dist}(v,C(v))}$. Hence, the total amount of pollution reduction is:

$$\operatorname{success}(C) := \sum_{v \in V} \frac{P(v)}{2^{\operatorname{dist}(v,C(v))}}.$$

The administration wants to find a solution as close as possible to an optimal solution $C^* \subseteq V$ (consisting of k areas) which maximizes the above objective. However, the computational resources of the administration are limited, and they cannot brute-force over all possible $\binom{n}{k}$ solutions.

Help them out by providing a polynomial-time approximation algorithm (in both n and k) which computes a solution $C \subseteq V$ with the following guarantee:

$$\operatorname{success}(C) \geqslant \left(1 - \frac{1}{e}\right) \operatorname{success}(C^*).$$

In other words, your algorithm should compute a $(1-1/e)$ approximation of the optimal solution. For full credit, you are required to explain your algorithm and to justify that its approximation ratio is $(1 - 1/e)$.

Solution. We define the function success : $2^V \to \mathbb{Z}_{\geqslant 0}$ as in the problem statement for any subset $C \subseteq V$ of areas. If $|C| = 0$ and thus $C = \emptyset$, we define $\operatorname{dist}(v, C(v)) = \infty$ for any v and therefore $\operatorname{success}(C) = 0$. We will prove that "success" is submodular and monotone, such that the problem reduces to cardinality constrained monotone maximization of a submodular function, with parameter k. We know from the lecture that Greedy is a $(1-1/e)$-approximation algorithm for this problem. Hence, it remains to prove the submodularity and monotonicity of "success".

We start with the latter. Let $C_1 \subseteq C_2 \subseteq V$. Then, for any v, we trivially have $\operatorname{dist}(v, C_1(v)) \geqslant \operatorname{dist}(v, C_2(v))$, such that:

$$\frac{P(v)}{2^{\operatorname{dist}(v,C_1(v))}} \leqslant \frac{P(v)}{2^{\operatorname{dist}(v,C_2(v))}}.$$

By summing over the areas $v \in V$, this implies immediately $\operatorname{success}(C_1) \leqslant \operatorname{success}(C_2)$.

We continue by proving submodularity. Let again $C_1 \subseteq C_2 \subseteq V$ and consider an area $w \in V \setminus C_2$. Let $C_3 := C_1 \cup \{w\}$ and $C_4 := C_2 \cup \{w\}$. We need to prove that:

$$\operatorname{success}(C_3) - \operatorname{success}(C_1) \geqslant \operatorname{success}(C_4) - \operatorname{success}(C_2). \quad (2)$$

By fixing a particular area $v \in V$, it suffices to prove:

$$\frac{P(v)}{2^{\text{dist}(v,C_3(v))}} - \frac{P(v)}{2^{\text{dist}(v,C_1(v))}} \geqslant \frac{P(v)}{2^{\text{dist}(v,C_4(v))}} - \frac{P(v)}{2^{\text{dist}(v,C_2(v))}}, \tag{3}$$

as then (2) follows by summing the above over areas $v \in V$. If $P(v) = 0$, then (3) follows trivially, so assume w.l.o.g. that $P(v) \neq 0$. Also, for simplifying the notation, let $c_i := C_i(v)$ and $d_i := \text{dist}(v, c_i)$ for $i \in \{1, 2, 3, 4\}$. Then, (2) is equivalent to:

$$2^{-d_3} - 2^{-d_1} \geqslant 2^{-d_4} - 2^{-d_2}. \tag{4}$$

To prove (4), we consider two cases. For case 1, assume that $\text{dist}(v, w) \geqslant d_2$. Then, $d_4 = \min\{d_2, \text{dist}(v, w)\} = d_2$, and therefore the RHS of (4) is 0. On the other hand, we have trivially $d_3 = \min\{d_1, \text{dist}(v, w)\} \leqslant d_1$, which implies that the LHS of (4) is $\geqslant 0$.

For case 2, assume the complementary event that $\text{dist}(v, w) < d_2$, which implies that $d_4 = \min\{d_2, \text{dist}(v, w)\} = \text{dist}(v, w)$ and also $d_3 = \min\{d_1, \text{dist}(v, w)\} = \text{dist}(v, w)$. As $d_3 = d_4$, (4) is equivalent to:

$$-2^{-d_1} \geqslant -2^{-d_2},$$

i.e. $d_1 \geqslant d_2$, which is obvious because $C_1 \subseteq C_2$.