

Impartial Selection with Predictions

Javier Cembrano*

Felix Fischer[†]

Max Klimm[‡]

Abstract

We study the selection of agents based on mutual nominations, a theoretical problem with many applications from committee selection to AI alignment. As agents both select and are selected, they may be incentivized to misrepresent their true opinion about the eligibility of others to influence their own chances of selection. Impartial mechanisms circumvent this issue by guaranteeing that the selection of an agent is independent of the nominations cast by that agent. Previous research has established strong bounds on the performance of impartial mechanisms, measured by their ability to approximate the number of nominations for the most highly nominated agents. We study to what extent the performance of impartial mechanisms can be improved if they are given a prediction of a set of agents receiving a maximum number of nominations. Specifically, we provide bounds on the consistency and robustness of such mechanisms, where consistency measures the performance of the mechanisms when the prediction is accurate and robustness its performance when the prediction is inaccurate. For the general setting where up to k agents are to be selected and agents nominate any number of other agents, we give a mechanism with consistency $1 - O(\frac{1}{k})$ and robustness $1 - \frac{1}{e} - O(\frac{1}{k})$. For the special case of selecting a single agent based on a single nomination per agent, we prove that 1-consistency can be achieved while guaranteeing $\frac{1}{2}$ -robustness. A close comparison with previous results shows that (asymptotically) optimal consistency can be achieved with little to no sacrifice in terms of robustness.

1 Introduction

Majority voting is a simple but very important mechanism for collective decision making. Its use dates back at least to ancient Athens, where it was employed for example to decide on the expulsion of citizens from the city [18]. A much more recent proposal uses majority voting to aggregate the solutions of multiple calls to large language models (LLMs) [14]. Some proposals even go so far as using it in AI alignment, and destroying AI entities if they are perceived as unaligned with human ethics by other AI entities [24]; see also Aaronson [1], Irving et al. [20].

The motivation for using majority decisions in these applications is their superior robustness to outliers compared to decisions made by a single entity. This argument requires, of course, that each entity is incentivized to reveal its true opinion about others rather than following its selfish interests. This is true for voting in general, but even more so in settings like those described above where the set of candidates and the set of voters overlap or are the same. Indeed, it is reasonable to assume that an Athenian citizen in fear of expulsion would have cast their vote for someone they considered likely to receive a large number of nominations, rather than someone they considered worthy of expulsion, in order to minimize their own risk of being expelled. Similarly, it is naïve to assume

*Department of Algorithms and Complexity, Max Planck Institut für Informatik; Department of Industrial Engineering, Universidad de Chile; jcembran@mpi-inf.mpg.de.

[†]School of Mathematical Sciences, Queen Mary University of London; felix.fischer@qmul.ac.uk.

[‡]Institute of Mathematics, Technische Universität Berlin; klimm@math.tu-berlin.de.

that AI entities risking destruction due to misalignment will truthfully report on the misalignment of other entities if this negatively affects their own chances of survival. What is needed are voting mechanisms for which the probability that an entity is selected is independent of the nominations cast by that entity. Such mechanisms are called *impartial* in the literature.

While impartiality is obviously appealing, previous work has established strong impossibility results for mechanisms that satisfy it. Deterministic impartial mechanisms that select a fixed number k of entities must fail natural axioms [12, 19], and the overall number of nominations for the selected entities cannot provide a constant approximation to the maximum number of nominations for any set of k entities [3]. Even randomized impartial mechanisms are relatively limited; for example, for the selection of a single entity they can only approximate the maximum number of nominations to a factor of $\frac{1}{2}$ [3, 17].

To improve the performance of impartial mechanisms, we will assume that the mechanism has access to a *prediction* of the entities most suitable for selection. Depending on the application, the prediction could for example come from another LLM not participating in the voting process or from expert advice. Our work is part of a growing literature on algorithms and mechanisms with advice; a website maintained by Lindermayr and Megow [22] provides an excellent overview of the area. The area is motivated by the fact that LLMs often provide astonishingly accurate answers, but also sometimes fail spectacularly. Mechanisms with advice therefore need to be able to cope with good as well as bad predictions, without a clear way to distinguish between the two. This trade-off is studied formally by considering the *consistency* and *robustness* of a mechanism. The consistency of a mechanism describes its ability to produce good outcomes when the predictions are accurate; the robustness its ability to produce reasonable results even when the predictions are inaccurate. The ability of a mechanism to move gracefully between these extremes is referred to as *smoothness*.

We will specifically consider deterministic and randomized impartial selection mechanisms with predictions. As it is standard in the literature on impartial selection, we formalize nominations among entities as a directed graph, where the set $[n] = \{1, \dots, n\}$ of vertices represent the entities and an edge from i to j indicates that i casts a nomination for j . A deterministic k -selection mechanism with predictions is given such a graph and a prediction $\hat{S} \subseteq [n]$ with $|\hat{S}| = k$, and returns a set of at most k vertices. A randomized k -selection mechanism is a lottery over deterministic mechanisms. Letting Δ_k denote the maximum sum of indegrees of any k vertices in the graph, a mechanism is called α -consistent for some $\alpha \in [0, 1]$ if the (expected) sum of indegrees of the selected vertices is at least $\alpha\Delta_k$ when the prediction is accurate, i.e., when the total indegree of the vertices in $\hat{S}$ is indeed equal to Δ_k . While it is trivial to achieve 1-consistency in an impartial way, by simply returning the predicted set $\hat{S}$, this would lead to arbitrarily bad performance when the predictions are inaccurate. To measure the performance of a mechanism in such cases, a mechanism is called β -robust for some $\beta \in [0, 1]$ if the (expected) sum of indegrees of the selected vertices is at least $\beta\Delta_k$ regardless of the quality of the prediction. We will be interested in the largest possible values of α and β for which impartial α -consistent and β -robust mechanisms can be found.

Our Results. We study impartial mechanisms with predictions in different settings; Figure 1 summarizes our results and compares them with previous work. As we initiate the study of impartial mechanisms with predictions, all previous mechanisms are unable to deal with predictions and consequently have equal robustness and consistency. Comparing our results with the baseline mechanism defined as a lottery between the best known mechanisms from the literature and the trivial mechanism that always selects the predicted set shows significant improvements.

We first study the classic setting of randomized impartial 1-selection mechanisms for approval voting. We propose a family of mechanisms we call ρ -permutation mechanisms that are parametrized

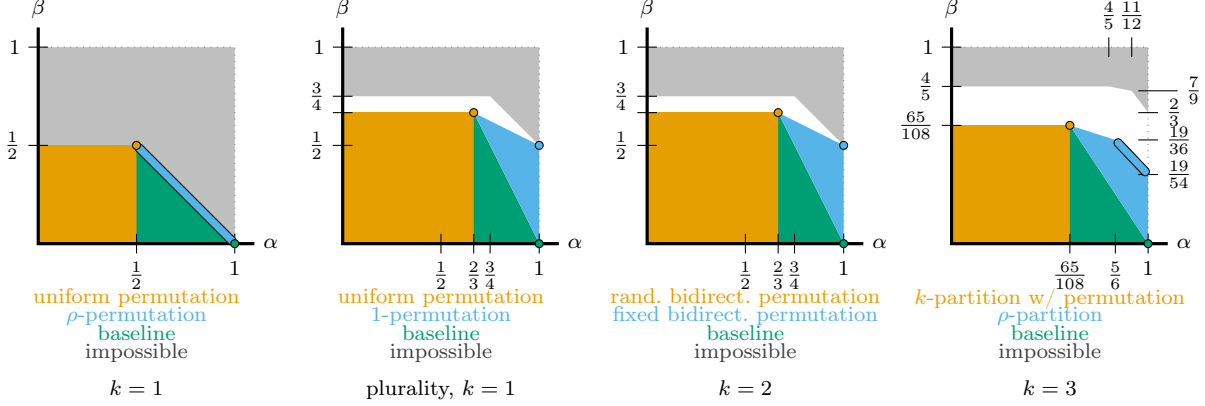


Figure 1: Trade-off between α -consistency and β -robustness of impartial k -selection mechanisms. **Orange dots** are the best mechanisms from previous work; **orange areas** are the whole ranges of possible consistency–robustness combinations implied by them. **Green dots** are the trivial mechanisms always selecting the predicted set; **green areas** are new ranges of consistency–robustness combinations implied by lotteries of them with previous work. **Blue dots** and **blue rounded rectangles** are new mechanisms introduced in this paper; **blue areas** are new ranges of consistency–robustness combinations implied by them or by lotteries of them with previous work. Gray areas are impossible consistency–robustness combinations as shown in Theorem 6.1. Whether the combinations in the white areas are achievable by impartial mechanisms is left for future research.

by a confidence parameter $\rho \in [\frac{1}{2}, 1]$. The mechanisms build upon the so-called (uniform) permutation mechanism [7, 17], which does not use any predictions and is $\frac{1}{2}$ -robust. In a nutshell, this mechanism permutes the vertices uniformly at random and carefully selects a vertex with maximum indegree from vertices that appear previously in the permutation. Our mechanisms favor permutations where the predicted vertex appears towards the end so that most of its incoming edges are likely observed, with a bias that depends on the confidence parameter. We show that for any $\rho \in [\frac{1}{2}, 1]$ the resulting mechanism is ρ -consistent and $(1 - \rho)$ -robust (Proposition 3.1), and that this trade-off between consistency and robustness is best-possible (Theorem 6.1). While this optimal consistency–robustness trade-off can also be achieved by the baseline mechanism that randomizes between the uniform permutation mechanism and the mechanism that selects the predicted vertex, such a mechanism would fail basic fairness notions, as discussed in Section 3.

We then study 1-selection mechanisms for plurality voting, where each vertex has exactly one outgoing edge. In this setting, we establish that the 1-permutation mechanism that puts the predicted vertex at the end of the permutation is 1-consistent and $\frac{1}{2}$ -robust (Theorem 3.3). Prior work had established that the uniform permutation mechanism is $\frac{2}{3}$ -robust, which also implies $\frac{2}{3}$ -consistency [11]. By an appropriate lottery between both mechanisms, we achieve $(\frac{2}{3} + \frac{1}{3}\rho)$ -consistency and $(\frac{2}{3} - \frac{1}{6}\rho)$ -robustness for all $\rho \in [0, 1]$ (Corollary 3.4). We further show in Theorem 6.1 that for any α -consistent and β -robust impartial mechanism, $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.

We next consider 2-selection mechanisms. In this setting, the bidirectional permutation mechanism [7] was shown to achieve the optimal robustness guarantee of $\frac{1}{2}$. We show that, by placing the predicted vertices at both ends of the permutation, we obtain the best-possible consistency guarantee of 1 without any sacrifice of robustness (Theorem 4.1). For randomized mechanisms, an appropriate lottery between this mechanism and the randomized permutation mechanism [7] achieves $(\frac{2}{3} + \frac{1}{3}\rho)$ -consistency and $(\frac{2}{3} - \frac{1}{6}\rho)$ -robustness for all $\rho \in [0, 1]$; see Proposition 4.2. We further show in Theorem 6.1 that, for any α -consistent and β -robust impartial mechanism, $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.

We finally study randomized k -selection mechanisms for an arbitrary number $k \in \mathbb{N}$ and obtain

Theorem 5.2, our most challenging result in terms of technical difficulty. Bjelde et al. [7] proposed the k -partition mechanism with permutation, which partitions the vertices randomly into k sets and selects one vertex from each set in a similar way to the permutation mechanism, but also accounting for edges from outside the set. We propose the ρ -partition mechanism for $\rho \in [\frac{1}{2}, 1]$, that partitions the vertices randomly into k sets but enforces that each set contains exactly one of the predicted vertices. In each set, the predicted vertex is put at position ρ while all other vertices obtain a position drawn uniformly from the unit interval. We then select one vertex from each set, as the k -partition mechanism with permutation. The mechanism achieves higher consistency by avoiding that more than one of the predicted vertices ends up in the same set, but the analysis requires new techniques because the probabilities of two optimal vertices being in the same set are no longer independent.

In the realm of mechanisms with predictions, it is common to also study approximation guarantees as a function of the prediction error, commonly referred to as smoothness. In our context, a natural notion of error of a predicted set of vertices is the difference between the maximum indegree of a set of k vertices and the indegree of the predicted set, normalized by the maximum indegree so it lies in the interval $[0, 1]$. Since all our α -consistent and β -robust mechanisms provide an α -approximation of the indegree of the predicted set, independently of whether this set is or is not optimal, they immediately yield a smoothness guarantee of $\max\{\alpha(1 - \eta), \beta\}$ for an error $\eta \in [0, 1]$.

Related Work. Impartiality, as we study it here, was first considered by de Clippel et al. [15] for the division of a divisible resource among members of a set of agents based on divisions proposed by the agents. In the context of selection, it was first studied by Holzman and Moulin [19] and Alon et al. [3]. Holzman and Moulin studied deterministic mechanisms for the special case of *plurality voting*, where each member casts exactly one nomination for another member of the set. They showed that, even in this restricted setting, impartiality is incompatible with the axioms of negative and positive unanimity, where the former requires that a member receiving no nomination is never selected and the latter that a member nominated by all members except themselves is always selected. Alon et al. studied the more general setting of approval voting, where members may nominate an arbitrary number of other members and a fixed number k of members is to be selected. Call a mechanism an *exact* k -selection mechanism if it always selects exactly k members, and α -optimal for $\alpha \in [0, 1]$ if the (expected) number of nominations that the selected members receive is always at least an α -fraction of the total number of nominations that the k best members receive. In this terminology, Alon et al. showed that no deterministic, impartial, and exact k -selection mechanism can be α -optimal for any fixed $\alpha > 0$. They further provided a randomized impartial $\frac{1}{4}$ -optimal 1-selection mechanism, and a randomized impartial $(1 - o(1))$ -optimal k -selection mechanism for $k \rightarrow \infty$. Fischer and Klimm [17] proposed and analyzed the permutation mechanism and showed that it is $\frac{1}{2}$ -optimal, which is best-possible for 1-selection. They further showed that for plurality votes, the same mechanism is α -optimal for $\alpha = \frac{67}{108} \approx 0.620$. Cembrano et al. [11] gave a tight analysis of the permutation mechanism for plurality votes, showing that it is even $\frac{2}{3}$ -optimal. They further proposed a new mechanism that is $\frac{2105}{3147}$ -optimal, where $\frac{2105}{3147} \approx 0.669$. Bjelde et al. [7] showed that deterministic impartial k -selection mechanisms that are allowed to sometimes select fewer than k members can perform better than exact k -selection mechanisms, and bounded the approximation guarantees of randomized mechanisms that select $k > 1$ members. Caragiannis et al. [9] studied the additive approximation guarantees of impartial selection mechanisms, and Cembrano et al. [12] gave a deterministic mechanism with an improved additive guarantee for plurality votes. Caragiannis et al. [10] considered the additive approximation guarantees of impartial mechanisms that receive *prior information* as additional input. They looked at two different models where members choose their nominations based on a known probability distribution or based on the popularity of a member.

We note that this approach differs from ours, since it does not bound the approximation guarantees if the prior information is inaccurate.

The robustness–consistency framework was first used by Purohit et al. [25] to study the performance of online algorithms with predictions. Predictions have been recently incorporated by Berger et al. [6] into the voting setting of metric distortion, where a candidate is to be selected based on rankings cast by voters with costs given by distances on a common metric space, and the goal is to minimize the ratio between the social cost of the selected candidate and that of the optimal one. More broadly, mechanisms with predictions were first studied by Agrawal et al. [2] for facility location, which has been further considered by Balkanski et al. [5] for randomized mechanisms and different types of predictions, by Fang et al. [16] for a restricted set of candidate locations, and by Istrate and Bonchis [21] for the case where agents’ objective is to maximize rather than minimize their distance to the facilities. Balkanski et al. [4] incorporated predictions into the design of strategyproof mechanisms for makespan minimization in scheduling. Xu and Lu [26] also studied a range of mechanism design problems with and without money, including facility location, scheduling, and auction design.

2 Preliminaries

For $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$ and let $\mathcal{G}_n = \{([n], E) : E \subseteq ([n] \times [n]) \setminus \bigcup_{i \in [n]} \{(i, i)\}\}$ denote the set of simple graphs with vertex set $[n]$ and without self-loops. For $S, T \in 2^{[n]}$, we denote the edges from vertices in S to vertices in T by $N_S^-(T, G) = \{(j, i) \in E : G = ([n], E), j \in S, i \in T\}$, and the number of such edges by $\delta_S^-(T, G)$. We omit S from the previous notation when $S = [n]$, and we write $N^-(i, G)$ instead of $N^-(\{i\}, G)$ and $\delta^-(i, G)$ instead of $\delta^-(\{i\}, G)$. For $k \in [n]$, we write $\Delta_k(G) = \max_{T \subseteq [n] : |T|=k} \delta^-(T, G)$. We omit k when it is equal to 1 and G whenever it is clear from the context. We refer to the graphs $G = ([n], E) \in \mathcal{G}_n$ such that $|\{(i, j) \in E : j \in [n]\}| = 1$ for every $i \in [n]$, in which all vertices have outdegree exactly one, as *plurality graphs*.

We consider selection mechanisms that obtain a prediction for the set of vertices with maximum indegrees. A *k-selection mechanism with predictions* is a family of functions $f : \binom{[n]}{k} \times \mathcal{G}_n \rightarrow [0, 1]^n$ with $\sum_{i \in [n]} f_i(\hat{S}, G) \leq k$ for all $G \in \mathcal{G}_n$, where $f_i(\hat{S}, G)$ denotes the probability assigned by the mechanism to agent i .¹ For a graph $G \in \mathcal{G}_n$ and $i \in [n]$, the number $f_i(\hat{S}, G)$ is the probability that f selects vertex i when $(\hat{S}, G)$ is the input. A mechanism is called *deterministic* if it only assigns probabilities 0 and 1, and is called *impartial* if $f_i(\hat{S}, G) = f_i(\hat{S}, G')$ whenever for two graphs $G = ([n], E)$ and $G' = ([n], E')$ we have $E \setminus \bigcup_{j \in [n]} \{(i, j)\} = E' \setminus \bigcup_{j \in [n]} \{(i, j)\}$.

For $\alpha \in [0, 1]$, we call a *k-selection mechanism with predictions* α -consistent if it achieves an α -approximation when the predictions are accurate, i.e., $\sum_{i \in [n]} f_i(\hat{S}, G) \delta^-(i, G) \geq \alpha \Delta_k(G)$ for all $n \in \mathbb{N}$, $G \in \mathcal{G}_n$, and $\hat{S} \in \binom{[n]}{k}$ with $\delta^-(\hat{S}, G) = \Delta_k(G)$. For $\beta \in [0, 1]$, we call a *k-selection mechanism with predictions* β -robust if it achieves a β -approximation regardless of the predictions’ quality, i.e., $\sum_{i \in [n]} f_i(\hat{S}, G) \delta^-(i, G) \geq \beta \Delta_k(G)$ for all $n \in \mathbb{N}$, $G \in \mathcal{G}_n$, and $\hat{S} \in \binom{[n]}{k}$.

We finally require some notation regarding permutations. For a (vertex) set S , we let $\Pi_S \subset S^{|S|}$ denote the set of permutations of the set S ; we refer to the order induced by a permutation as an order *from left to right* for ease of notation. We write Π_n as a shorthand for $\Pi_{[n]}$. For a permutation $\pi \in \Pi_S$, a set $S' \subseteq S$, and a vertex $i \in S$, we write $\pi_{<i} = \{j \in S : j = \pi_r, i = \pi_t \text{ for some } r < t\}$ for the set of vertices that appear to the left of i , $\pi(S') \in \Pi_{S'}$ for the restriction of π to S' , and $\bar{\pi} \in \Pi_S$ for the reverse of π . Sometimes we fix the position of some vertices in the permutation. For a set of distinct

¹It is not hard to see that such a distribution over vertices can be translated into a probability over sets of size at most k via the Birkhoff-von Neumann Theorem; see Bjelde et al. [7, Lemma 2.1] for the details.

vertices $\{i_j : j \in [m]\}$ and distinct positions $\{r_j : j \in [m]\}$, we write $\Pi_S(i_1 \rightarrow r_1, \dots, i_m \rightarrow r_m)$ for the set of permutations $\pi \in \Pi_S$ such that $i_j = \pi_{r_j}$ for every $j \in [m]$.

3 Selecting a Single Vertex

In this section, we study 1-selection mechanisms with predictions. For ease of notation, we denote the predicted set by $\hat{S} = \{\hat{i}\}$ and write $\Delta(G)$ instead of $\Delta_1(G)$ for the maximum indegree.

It is well known that deterministic mechanisms cannot achieve any constant approximation in the classic setting without predictions, which for our setting has the direct implication that no deterministic mechanism with predictions can be β -robust for a constant $\beta > 0$. Thus, the trivial answer to the best-possible trade-off between consistency and robustness is given by the mechanism that selects the predicted vertex $\hat{i}$ and achieves 1-consistency and 0-robustness.

The problem becomes more interesting with randomization, as the best-known mechanism for the setting without predictions achieves a $\frac{1}{2}$ -approximation. We refer to the mechanism achieving this approximation, introduced by Fischer and Klimm [17], as the *uniform permutation mechanism*. This mechanism sorts the vertices uniformly at random and considers them one by one according to this order while maintaining a candidate vertex, initially the first vertex. A vertex is taken as the new candidate if its observed indegree is larger than that of the current candidate, where *observed indegree* refers to the indegree when only considering incoming edges from previous vertices and omitting a potential edge from the current candidate. The vertex that is the candidate in the end is selected.

For later use, we define a more general version of this mechanism, where in addition to the graph $G = ([n], E)$, the mechanism receives a subset of vertices $S \subseteq [n]$ and a vector $x \in [0, 1]^S$. Vertices in S are those taken into account for the permutation, while all other vertices in $[n] \setminus S$ are not eligible for selection and the incoming edges from these vertices are always considered. The vector $x \in [0, 1]^S$ defines the permutation $\pi \in \Pi_S$: i comes before j if its associated value x_i is smaller than x_j . Formally, for every $i, j \in S$ we have $i \in \pi_{<j}$ if and only if either $x_i < x_j$ or both $x_i = x_j$ and $i < j$ hold (we break ties in favor of vertices with smaller indices). We denote the permutation $\pi \in \Pi_S$ constructed in this way from $x \in [0, 1]^S$ by $\pi(x)$.

The permutation mechanism for a fixed set S and vector $x \in [0, 1]^S$ is formally described in Algorithm 1; we refer to its output for a graph G , a set S , and a vector x by $\text{Pm}(G, S, x)$. The uniform permutation mechanism, providing the best-possible guarantee among randomized 1-selection mechanisms without prediction, corresponds to the mechanism that receives a graph G and returns $\text{Pm}(G, [n], x)$, where $x_i \in [0, 1]$ is taken uniformly at random for each $i \in [n]$.

Instead of the uniform permutation mechanism, we consider in the setting with predictions the ρ -permutation mechanism, given in Algorithm 2. This mechanism receives a graph $G = ([n], E)$ and a predicted vertex $\hat{i} \in [n]$, and returns $\text{Pm}(G, [n], x)$, where now $\hat{i}$ has an associated value $x_{\hat{i}} = \rho$ and all values x_i for $i \in [n] \setminus \{\hat{i}\}$ are sampled uniformly at random. The value ρ then has the natural interpretation of a confidence parameter: Taking $\rho = 1$ ensures seeing all incoming edges of the predicted vertex, while smaller values of ρ increase the probability of seeing potential outgoing edges of $\hat{i}$. This mechanism attains any convex combination of α -consistency and β -robustness between the points $(\alpha, \beta) \in \{(1, 0), (\frac{1}{2}, \frac{1}{2})\}$. In Section 6, we will see that this trade-off is actually best-possible.

Proposition 3.1. *For any confidence parameter $\rho \in [\frac{1}{2}, 1]$ the ρ -permutation mechanism is impartial, ρ -consistent and $(1 - \rho)$ -robust.*

We need some notation. For a fixed graph $G = ([n], E) \in \mathcal{G}_n$, set $S \subseteq [n]$, and vector $x \in [0, 1]^S$, we let $i^{\text{Pm}}(G, S, x)$ denote the outcome of $\text{Pm}(G, S, x)$. Whenever x is fixed, we write π for the

Algorithm 1 Permutation mechanism $\text{Pm}(G, S, x)$

Input: graph $G = ([n], E)$, set $S \subseteq [n]$, $x \in [0, 1]^S$.

Output: vertex $i^{\text{Pm}} \in [n]$.

```
 $\pi \leftarrow \pi(x) \in \Pi_S$ 
initialize  $i^{\text{Pm}} \leftarrow \pi_1$  and  $d \leftarrow \delta_{[n] \setminus S}^-(\pi_1)$ 
for  $r \in \{2, \dots, |S|\}$  do
     $i \leftarrow \pi_r$ 
    if  $\delta_{([n] \setminus S) \cup (\pi_{< i} \setminus \{i^{\text{Pm}}\})}^-(i) \geq d$  then
        update  $i^{\text{Pm}} \leftarrow i$  and  $d \leftarrow \delta_{([n] \setminus S) \cup \pi_{< i}}^-(i)$ 
return  $i^{\text{Pm}}$ 
```

Algorithm 2 ρ -permutation mechanism $\text{Pm}^\rho(\hat{i}, G)$

Input: graph $G = ([n], E)$, predicted vertex $\hat{i} \in [n]$.

Output: vertex $i^{\text{Pm}} \in [n]$.

```
 $x_{\hat{i}} \leftarrow \rho$ 
sample  $x_i \in [0, 1]$  uniformly at random  $\forall i \in [n] \setminus \{\hat{i}\}$ 
return  $\text{Pm}(G, [n], x)$ 
```

induced permutation instead of $\pi(x)$. As a key property for the analysis of the (uniform) permutation mechanism, Bousquet et al. [8] showed that, for any fixed permutation, it selects a vertex with maximum indegree from the left. Bjelde et al. [7] extended this result to the case where we restrict to a set of vertices and consider all incoming edges from other vertices. We phrase the latter result with our notation as the following lemma, which we apply in Appendix A.1 to prove Proposition 3.1.

Lemma 3.2 (Bjelde et al. [7]). *For every $G = ([n], E) \in \mathcal{G}_n$, $S \subseteq [n]$, and $x \in [0, 1]^S$, it holds that $i^{\text{Pm}}(G, S, x) \in \arg \max \{\delta_{([n] \setminus S) \cup \pi_{< i}}^-(i, G) : i \in [n]\}$.*

It is worth noting that the consistency and robustness guarantees of Proposition 3.1 are also achieved by a baseline mechanism that returns the predicted vertex with probability ρ and runs the uniform permutation mechanism with probability $1 - \rho$. However, the baseline mechanism fails a basic *unanimity* notion introduced by Holzman and Moulin [19]: If a vertex v is such that all other vertices have a single outgoing edge to v , then v should be selected. Whenever v is not the predicted vertex, the baseline mechanism fails to select v with constant probability, while the ρ -permutation mechanism returns v as long as it is not first or second in the permutation, i.e., with probability $1 - O(\frac{1}{n})$.

Plurality Voting. A usual restriction in voting is that each member nominates one other member, which in our graph representation implies having vertices with outdegree one. This paradigm of *plurality voting*, extensively considered in the impartial selection literature [11, 19, 23], has been shown to enable better approximation guarantees for randomized mechanisms.² In particular, Cembrano et al. [11] proved that the uniform permutation mechanism provides an improved approximation ratio of $\frac{2}{3}$ in this case.

In our setting, we show that the ρ -permutation mechanism with $\rho = 1$, where the predicted vertex is deterministically placed at the end of the permutation and all other vertices are sorted uniformly at random, achieves 1-consistency and $\frac{1}{2}$ -robustness. The following theorem provides a more fine-grained bound on the robustness of this mechanism as a function of the maximum indegree Δ of the input graph; the bound of $\frac{1}{2}$ follows by taking the worst case over Δ .

Theorem 3.3. *The 1-permutation mechanism is impartial, 1-consistent, and $\beta(\Delta)$ -robust on plurality*

²The impossibility of providing a constant approximation of the maximum indegree with deterministic mechanisms remains true in this restricted setting [19].

graphs with maximum indegree $\Delta \geq 2$, where

$$\beta(\Delta) = \begin{cases} \frac{3\Delta-2}{4\Delta} & \text{if } \Delta \text{ is even,} \\ \frac{3\Delta^2-2\Delta-1}{4\Delta^2} & \text{if } \Delta \text{ is odd.} \end{cases}$$

Moreover, this function β is increasing, implying that this mechanism is impartial, 1-consistent, and $\frac{1}{2}$ -robust on plurality graphs.

We prove this theorem in Appendix A.2, using a strengthened version of a lemma of Cembrano et al. [11] establishing a negative correlation between the indegree from the left of the maximum-indegree vertex and that of all other vertices. We show as Lemma A.1 that this result remains true for the non-uniform distribution over permutations induced by the vector x defined in the 1-permutation mechanism, and that it holds not only for the maximum-indegree vertex but for any fixed vertex. The proof adapts that of Cembrano et al. [11], defining an injective function between sets of permutations to couple the probabilities that certain indegrees are observed in the permutation taken in the mechanism. We then use the lemma to prove Theorem 3.3. The most challenging case, which ultimately leads to a worse robustness guarantee than in the setting without predictions, is when the maximum-indegree vertex has an incoming edge from the predicted vertex, as this edge is never considered by the mechanism when observing the indegrees from the left. However, since all outdegrees are 1, we can still obtain a lower bound on the probability of selecting this maximum-indegree vertex or another vertex with high indegree.

We now state the implications of Theorem 3.3, in terms of the trade-off between consistency and robustness we can achieve by combining the 1-permutation mechanism with the uniform permutation mechanism. The proof of this result is deferred to Appendix A.3. In Section 6, we will see that this trade-off is not far from tight.

Corollary 3.4. *For every $\rho \in [0, 1]$, there exists a randomized 1-selection mechanism with predictions that is impartial, α -consistent, and β -robust on plurality graphs with maximum indegree Δ , where*

$$\alpha(\Delta) = \begin{cases} \frac{3\Delta+2}{4(\Delta+1)} + \frac{\Delta+2}{4(\Delta+1)}\rho & \text{if } \Delta \text{ is even,} \\ \alpha(\Delta-1) & \text{if } \Delta \text{ is odd,} \end{cases} \quad \beta(\Delta) = \begin{cases} \frac{3\Delta+2}{4(\Delta+1)} - \frac{\Delta+2}{4\Delta(\Delta+1)}\rho & \text{if } \Delta \text{ is even,} \\ \frac{3\Delta-1}{4\Delta} - \frac{\Delta+1}{4\Delta^2}\rho & \text{if } \Delta \text{ is odd.} \end{cases}$$

In particular, for every $\rho \in [0, 1]$, there exists a randomized 1-selection mechanism with predictions that is impartial, $(\frac{2}{3} + \frac{1}{3}\rho)$ -consistent, and $(\frac{2}{3} - \frac{1}{6}\rho)$ -robust on plurality graphs.

4 Selecting Two Vertices

In this brief section, we state our results for the selection of two vertices. In terms of mechanisms without predictions, the best-known deterministic and randomized impartial mechanisms achieve $\frac{1}{2}$ - and $\frac{2}{3}$ -optimality, respectively. While the bound for deterministic mechanisms is best-possible, only an upper bound of $\frac{3}{4}$ is known for randomized mechanisms [7]. For compactness, throughout this section we denote the predicted set by $\hat{S} = \{\hat{i}_1, \hat{i}_2\}$.

The deterministic mechanism achieving $\frac{1}{2}$ -optimality is based on the permutation mechanism. It runs, for an arbitrarily fixed permutation π , the permutation mechanism for both π and its reverse $\bar{\pi}$, and returns the selected vertices for each direction (potentially the same vertex). A natural approach to incorporate the prediction is to run this mechanism with the predicted vertices at both extremes of the fixed permutation. The resulting mechanism, which we call *fixed bidirectional permutation*, maintains the best-possible robustness of $\frac{1}{2}$ while achieving 1-consistency. The formal description of the mechanism as Algorithm 4 and the proof of this result are deferred to Appendix B.1.

Theorem 4.1. *The fixed bidirectional permutation mechanism is impartial, 1-consistent, and $\frac{1}{2}$ -robust.*

In terms of randomized mechanisms, convex combinations of the best-known mechanism without prediction, achieving $\frac{2}{3}$ -robustness [7], and the fixed bidirectional permutation mechanism, achieving 1-consistency and $\frac{1}{2}$ -robustness, allows us to attain combinations of α -consistency and β -robustness between $(\alpha, \beta) = (\frac{2}{3}, \frac{2}{3})$ and $(\alpha, \beta) = (1, \frac{1}{2})$. We state this simple fact in the following proposition; we will see in Section 6 that this combination of consistency and robustness is not far from tight.

Proposition 4.2. *For every $\rho \in [0, 1]$, there exists a randomized 2-selection mechanism with predictions that is impartial, $(\frac{2}{3} + \frac{1}{3}\rho)$ -consistent, and $(\frac{2}{3} - \frac{1}{6}\rho)$ -robust.*

5 Selecting $k \geq 3$ Vertices

In this section, we study the impartial selection of $k \geq 3$ vertices when the mechanism is equipped with a prediction on the optimal set.

In terms of deterministic mechanisms, the setting without predictions is far from well understood. Indeed, a large gap remains between the best-known lower and upper bounds of $\frac{1}{k}$ and $\frac{k-1}{k}$ on the approximation guarantee that impartial mechanisms can achieve [7]. Recently, Cembrano et al. [13] improved the lower bound for cases where k is larger than (approximately) $2\sqrt{n}$, but the lower bound of $\frac{1}{k}$ remains the best-known bound for an arbitrary number of agents n . This guarantee comes from the bidirectional permutation mechanism explained in the previous section, whose $\frac{1}{2}$ -approximation of the optimal set of two agents translates into a $\frac{1}{k}$ -approximation of the optimal committee of k agents. Similarly to the previous section, we can modify this mechanism to maintain its robustness guarantee and achieve 1-consistency. Specifically, we select $k - 2$ vertices from the predicted set and one or two more vertices through our fixed bidirectional permutation mechanism, with the remaining two predicted vertices at the extremes of the permutation. We state the properties of this simple mechanism in the following proposition, proven in Appendix C.1.

Proposition 5.1. *There exists a deterministic k -selection mechanism with predictions that is impartial, 1-consistent, and $\frac{1}{k}$ -robust.*

Regarding randomized mechanisms, the best-known mechanism for k -selection was developed by Bjelde et al. [7] and provides an approximation guarantee of $\frac{k}{k+1}(1 - (\frac{k-1}{k})^{k+1})$, which starts at $\frac{7}{12} \approx 0.5833$ for $k = 2$, $\frac{65}{108} \approx 0.6019$ for $k = 3$, and approaches $1 - \frac{1}{e} \approx 0.6321$ as k grows. The mechanism assigns each vertex to one out of k sets uniformly at random. It then selects one vertex from each set via the permutation mechanism restricted to that set with an internal permutation taken uniformly at random. While its impartiality is easy to see, the approximation guarantee requires a careful analysis of the expected observed indegree of optimal vertices in each set. In the following, we develop a randomized mechanism with predictions inspired by this mechanism that achieves almost optimal robustness while losing very little in terms of consistency, especially as k grows.

As in the mechanism by Bjelde et al., vertices are assigned to one of k sets, and one vertex is selected from each set by running the permutation mechanism restricted to the set. However, both the assignment to sets and the permutation are not taken independently and uniformly for each vertex anymore. Instead, we assign one predicted vertex to each set; all other vertices are still assigned to a set chosen independently and uniformly at random. Within each set, the permutation is sampled as in the ρ -permutation mechanism from Section 3: For each set A_j with a predicted vertex $\hat{i}_j$, we take a vector $x \in [0, 1]^{A_j}$ such that $x_{\hat{i}_j} = \rho$ and $x_i \in [0, 1]$ is taken uniformly at random

Algorithm 3 ρ -partition mechanism $\text{Pt}^\rho(\hat{S}, G)$

Input: graph $G = ([n], E)$,
 predicted set $\hat{S} = \{\hat{i}_1, \dots, \hat{i}_k\} \in \binom{[n]}{k}$.
Output: set $S^{\text{Pt}} \in \binom{[n]}{k}$.
 sample $j_i \in [k]$ uniformly $\forall i \in [n] \setminus \hat{S}$
 assign i to $A_{j_i} \forall i \in [n] \setminus \hat{S}$
for $j \in [k]$ **do**
 $A_j \leftarrow A_j \cup \{\hat{i}_j\}, x_{i_j} \leftarrow \rho$
 sample $x_i \in [0, 1]$ uniformly $\forall i \in A_j \setminus \{\hat{i}_j\}$
 $i_j^{\text{Pt}} \leftarrow \text{Pm}(G, A_j, x)$
 $S^{\text{Pt}} \leftarrow S^{\text{Pt}} \cup \{i_j^{\text{Pt}}\}$
return S^{Pt}

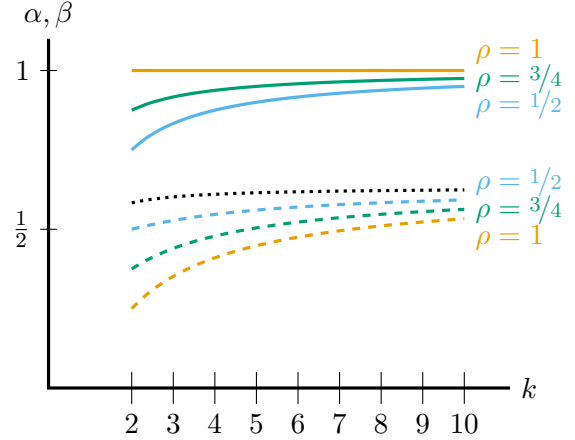


Figure 2: The ρ -partition mechanism (left) and a plot of its α -consistency (solid) and β -robustness (dashed) for the values $\rho = 1$, $\rho = \frac{3}{4}$, and $\rho = \frac{1}{2}$ as a function of k (right). The dotted black line is the consistency and robustness of the k -partition mechanism of Bjelde et al. [7].

for each $i \in A_j \setminus \{\hat{i}_j\}$. Intuitively, these changes allow the mechanism to see most incoming edges of the predicted vertices while only mildly affecting the distributions to keep a strong robustness guarantee.

The mechanism, which we refer to as the ρ -partition mechanism, is formally presented in Algorithm 3. For $\rho \in [0, 1]$, we denote its output by $\text{Pt}^\rho(\hat{S}, G)$ for each graph $G = ([n], E)$ and predicted set $\hat{S}$. By tuning the confidence parameter ρ between $\frac{1}{2}$ and 1, we achieve a consistency between $1 - \frac{1}{2k}$ and 1 while only losing $O(\frac{1}{k})$ in robustness compared to the best-known mechanism without prediction.

Theorem 5.2. *For any confidence parameter $\rho \in [\frac{1}{2}, 1]$, the ρ -partition mechanism is impartial, α -consistent, and β -robust, where $\alpha = 1 - \frac{1-\rho}{k}$ and $\beta = (1 - \frac{2\rho}{k+1})(1 - (\frac{k-1}{k})^k)$.*

For example, when taking $\rho = \frac{1}{2}$ to prioritize robustness, our mechanism achieves a robustness guarantee of $\frac{1}{2}$ for $k = 2$, $\frac{19}{36} \approx 0.5278$ for $k = 3$, $\frac{35}{64} \approx 0.5469$ for $k = 4$, and approaching $1 - \frac{1}{e} \approx 0.6321$ for $k \rightarrow \infty$. The consistency guarantee for this value of ρ and any $k \geq 2$ is $1 - \frac{1}{2k}$, which is $\frac{3}{4} = 0.75$ for $k = 2$, $\frac{5}{6} \approx 0.8333$ for $k = 3$, $\frac{7}{8} = 0.875$ for $k = 4$, and approaches 1 for $k \rightarrow \infty$. When taking $\rho = 1$ to maximize consistency, the mechanism is 1-consistent for any k and achieves a robustness guarantee of $\frac{1}{4} = 0.25$ for $k = 2$, $\frac{19}{54} \approx 0.3519$ for $k = 3$, $\frac{105}{256} \approx 0.4102$ for $k = 4$, and again approaching $1 - \frac{1}{e} \approx 0.6321$ for $k \rightarrow \infty$. Figure 2 illustrates the performance of the ρ -partition mechanism for $\rho \in \{\frac{1}{2}, \frac{3}{4}, 1\}$ and $k \in \{2, \dots, 10\}$, and compares it with the k -partition mechanism of Bjelde et al. [7].

The proof of Theorem 5.2 is deferred to Appendix C.2; here we briefly describe the main ideas behind the robustness guarantee, which constitutes the most difficult part of the proof. For the analysis we consider an optimal set S^* and $j \in [k]$ such that A_j contains an optimal vertex, i.e., $S^* \cap A_j \neq \emptyset$, and sample a vertex i^* from $S^* \cap A_j$ uniformly at random. We then bound the expected indegree of i^* that the mechanism observes by bounding the probability that each in-neighbor i of i^* lies in a set other than A_j or in the set A_j but before i^* according to the internal permutation. What complicates the analysis is that, unlike in the mechanism without predictions, the events $i^* \in A_j$ and $i \in A_j$ are not independent. However, it is not difficult to see that when $i \notin S^* \cup \hat{S}$, the probability of i being in A_j is the same as in the independent case. We show further that when $i \in S^*$ or $i^* \in \hat{S}$, the probability of i being in A_j cannot increase much, and the only difference is

given by the position of the predicted vertex in the internal permutation. The most intricate part of the proof is the case where $i^* \in S^* \setminus \hat{S}$ and $i \in \hat{S} \setminus S^*$, because the events of i^* being sampled in the set A_j and i being in this set can be strongly correlated. Indeed, the probability of the former event conditional on $i \in A_j$ can be as large as 1 if, for example, all predicted vertices except i belong to S^* , as in this case $i \in A_j$ implies that i^* is the unique vertex in $S^* \cap A_j$. We tackle this difficulty by directly computing a lower bound on the (unconditional) probability of i^* being sampled as the optimal vertex in A_j .

6 Upper Bounds

To put our consistency and robustness results into perspective, we will now give upper bounds on the values α and β for which an impartial selection mechanism with predictions can simultaneously guarantee α -consistency and β -robustness. We do so for k -selection with $k \in \{1, 2, 3\}$, and for 1-selection from plurality graphs. The upper bounds are shown in Figure 1 alongside the lower bounds obtained in earlier sections.

Theorem 6.1. *The following statements hold:*

- (i) *If a randomized 1-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{1}{2}$ and $\alpha + \beta \leq 1$.*
- (ii) *If a randomized 1-selection mechanism with predictions is impartial, α -consistent, and β -robust on plurality graphs, then $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.*
- (iii) *If a randomized 2-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.*
- (iv) *If a randomized 3-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{4}{5}$, $4\alpha + 3\beta \leq 6$, and $4\alpha + 21\beta \leq 20$.*

We prove these results in Appendix D.1. To this end we consider appropriate families of graphs and for each vertex in these graphs introduce a variable for the probability with which some impartial, α -consistent, and β -robust k -selection mechanism selects that vertex. We generalize a lemma of Holzman and Moulin [19] to show that one can restrict attention to symmetric mechanisms, and use impartiality, consistency, robustness, and the fact that the probabilities for each graph must sum up to k to obtain a set of linear inequalities involving the probability variables, α , and β . We then show that any values of α and β *not* satisfying the statements violate the linear inequalities.

7 Discussion

We have initiated the study of impartial selection mechanisms with predictions. Unlike majority voting, these mechanisms are not prone to strategic manipulation. While we have made substantial progress regarding the approximation guarantees achievable by such mechanisms, in most settings a moderate gap remains between the upper and lower bounds. We leave closing these gaps for future work. In addition, it would be interesting to test the mechanisms we have proposed in practical applications, for example in the aggregation of outputs of different LLMs.

Acknowledgements

Research was supported by the Deutsche Forschungsgemeinschaft under project number 431465007, by the Engineering and Physical Sciences Research Council under grant EP/T015187/1, and by a Structural Democracy Fellowship through the Brooks School of Public Policy at Cornell University.

A Proofs Deferred from Section 3

A.1 Proof of Proposition 3.1

Impartiality follows directly from the impartiality of the permutation mechanism with any (fixed or randomly chosen) permutation, established by Fischer and Klimm [17]. This is because the permutation π constructed from x in our setting only depends on the identity of the predicted vertex, which does not depend on the outgoing edges, and is otherwise sampled randomly and independently of any vertex's identity or outgoing edges.

For the approximation guarantees, we fix $G = ([n], E)$ and $\hat{i} \in [n]$. For the consistency guarantee, we further assume that $\delta^-(\hat{i}) = \Delta$ and observe that

$$\mathbb{E}[i^{\text{Pm}}(G, [n], x)] \geq \mathbb{E}[\delta_{\pi_{<\hat{i}}}^-(\hat{i})] = \sum_{i \in N^-(\hat{i})} \mathbb{P}[i \in \pi_{<\hat{i}}] = \rho \delta^-(\hat{i}) = \rho \Delta,$$

where the first inequality follows from Lemma 3.2 and the second equality from the fact that $x_{\hat{i}} = \rho$ and $x_i \in [0, 1]$ is taken uniformly at random for every $i \in [n] \setminus \{\hat{i}\}$. We conclude that, in this case, $\frac{1}{\Delta} \mathbb{E}[i^{\text{Pm}}(G, [n], x)] \geq \rho$, so the mechanism is ρ -consistent.

Similarly, for the robustness guarantee we denote by $i^* \in [n]$ any vertex with $\delta^-(i^*) = \Delta$ and observe that

$$\begin{aligned} \mathbb{E}[i^{\text{Pm}}(G, [n], x)] &\geq \mathbb{E}[\delta_{\pi_{<i^*}}^-(i^*)] = \sum_{i \in N^-(i^*)} \mathbb{P}[i \in \pi_{<i^*}] \\ &= \mathbb{1}_{i \in N^-(i^*)} \mathbb{P}[i \in \pi_{<i^*}] + \sum_{i \in N^-(i^*) \setminus \{i\}} \mathbb{P}[i \in \pi_{<i^*}] \\ &\geq (1 - \rho) \mathbb{1}_{i \in N^-(i^*)} + \frac{1}{2} \delta_{N^-(i^*) \setminus \{i\}}^-(i^*) \geq (1 - \rho) \Delta, \end{aligned}$$

where the first inequality follows from Lemma 3.2, the first inequality from the fact that $x_{\hat{i}} = \rho$ and $x_i \in [0, 1]$ is taken uniformly at random for every $i \in [n] \setminus \{\hat{i}\}$, and the last inequality due to $\rho \geq \frac{1}{2}$. We conclude that, regardless of $\delta^-(\hat{i})$, we have $\frac{1}{\Delta} \mathbb{E}[i^{\text{Pm}}(G, [n], x)] \geq 1 - \rho$, so the mechanism is $(1 - \rho)$ -robust. $\square$

A.2 Proof of Theorem 3.3

Throughout this appendix, we let $\sigma_\pi(i, G) = \delta_{\pi_{<i}}(i, G)$ denote the indegree of vertex i from its left, either for a fixed permutation π or for the random variable corresponding to the permutation $\pi(x)$. Note that this permutation is taken uniformly at random from within all permutations that have $\hat{i}$ in the last position, i.e., from the set $\Pi_n(\hat{i} \rightarrow n)$. For a graph $G = ([n], E)$ and a vertex $i \in [n]$, we let

$$A_r(i, G) = [\sigma_\pi(i, G) = r] \quad \text{for } r \in [\delta^-(i, G)]$$

denote the event that i has indegree r from the left, and

$$B_r(i, G) = \bigcup_{j \in [n] \setminus \{i\}} [\sigma_\pi(j, G) \geq r] \quad \text{for } r \in [\Delta(G)]$$

denote the event that a vertex other than i has indegree r or higher from the left. We omit the graph G from the previous notation when clear from context.

Before proceeding with the proof of Theorem 3.3, state and prove the following lemma.

Lemma A.1. For every graph $G = ([n], E) \in \mathcal{G}_n$, vertices $\hat{i} \in [n]$, $i \in [n] \setminus \hat{i}$, and values $r, s \in [\delta^-(i, G) - \mathbf{1}_{\hat{i} \in N^-(i)}]$ with $r > s$,

$$\mathbb{P}[B_r(i, G) \mid A_s(i, G)] \geq \mathbb{P}[B_r(i, G) \mid A_r(i, G)],$$

where the probabilities are taken over $x \in [0, 1]^n$, with $x_{\hat{i}} = 1$ and x_j taken uniformly at random for every $j \in [n] \setminus \{\hat{i}\}$.

Proof. Let G , $\hat{i}$, i , r , and s be as in the statement. From the definition of the events, we have that

$$\begin{aligned} \mathbb{P}[B_r(i) \mid A_s(i)] &= \frac{\mathbb{P}[\sigma_\pi(i) = s \text{ and } \exists j \in [n] \setminus \{i\} : \sigma_\pi(j) \geq r]}{\mathbb{P}[\sigma_\pi(i) = s]} \\ \mathbb{P}[B_r(i) \mid A_r(i)] &= \frac{\mathbb{P}[\sigma_\pi(i) = r \text{ and } \exists j \in [n] \setminus \{i\} : \sigma_\pi(j) \geq r]}{\mathbb{P}[\sigma_\pi(i) = r]}. \end{aligned}$$

Note that $\mathbb{P}[\sigma_\pi(i) = s] = \mathbb{P}[\sigma_\pi(i) = r]$. Indeed, these probabilities are both equal to $\frac{1}{\delta^-(i)+1}$ if $(\hat{i}, i) \notin E$ and to $\frac{1}{\delta^-(i)}$ if $(\hat{i}, i) \in E$. Thus, it suffices to show the inequality for the numerators. Letting

$$\begin{aligned} \Pi_n^{rs} &= \{\pi \in \Pi_n(\hat{i} \rightarrow n) : \sigma_\pi(i) = s \text{ and } \exists j \in [n] \setminus \{i\} : \sigma_\pi(j) \geq r\} \\ \Pi_n^r &= \{\pi \in \Pi_n(\hat{i} \rightarrow n) : \sigma_\pi(i) = r \text{ and } \exists j \in [n] \setminus \{i\} : \sigma_\pi(j) \geq r\}, \end{aligned}$$

we only need to prove that $|\Pi_n^{rs}| \geq |\Pi_n^r|$, since the permutation is chosen uniformly at random from $\Pi_n(\hat{i} \rightarrow n)$. We prove this inequality by constructing an injective function $f: \Pi_n^r \rightarrow \Pi_n^{rs}$.

For $\pi \in \Pi_n^r$, we construct $g(\pi)$ by exchanging i with the $(s+1)$ th vertex among its in-neighbors, i.e., by exchanging $i = \pi_{r+1}(\{i\} \cup N^-(i))$ with $i' = \pi_{s+1}(\{i\} \cup N^-(i))$. This function is clearly injective and, moreover, $\sigma_{g(\pi)}(i) = s$. To conclude that $g(\pi) \in \Pi_n^{rs}$, it only remains to show that $\sigma_{g(\pi)}(j) \geq r$ for some $j \in [n] \setminus \{i\}$. To see this, observe that $\sigma_\pi(j') \geq r$ for some $j' \in [n] \setminus \{i\}$, because $\pi \in \Pi_n^r$. We claim that $\sigma_{g(\pi)}(j') \geq r$ holds as well; i.e., that the indegree from the left of this vertex j' does not decrease after applying g . This is the case because $g(\pi)$ only differs from π in the position of the vertices i and i' . Since i' moved to the right, its indegree from the left cannot decrease. The indegree from the left of i' outneighbor i decreases by one, but we know that $j' \neq i$. The indegree from the left of i 's outneighbor, finally, may or may not increase by one, but cannot decrease. The indegree from the left of all other vertices remains constant. Thus, we have indeed $\sigma_{g(\pi)}(j') \geq r$, hence $g(\pi) \in \Pi_n^{rs}$ and we conclude. $\square$

We now proceed with the proof of Theorem 3.3.

Proof of Theorem 3.3. Impartiality of the 1-permutation mechanism follows directly from the impartiality of the permutation mechanism with any fixed permutation, established by Fischer and Klimm [17], since the 1-permutation mechanism samples the permutation independently from all outgoing edges.

For the approximation guarantees, we fix $G = ([n], E) \in \mathcal{G}_n$ and $\hat{i} \in [n]$, and denote $\Delta = \Delta(G)$. We let x be the random vector taken when running the mechanism and $\pi = \pi(x) \in \Pi_n(\hat{i} \rightarrow n)$ the associated random permutation. We write i^{Pm} instead of $i^{\text{Pm}}(G, S, x)$ for the (random) vertex selected by the mechanism.

To see that the mechanism is 1-consistent, we assume that $\delta^-(\hat{i}) = \Delta$ and observe that $\sigma_\pi(\hat{i}) = \delta^-(\hat{i})$ because $\pi_n = \hat{i}$. Thus, we obtain

$$\delta^-(i^{\text{Pm}}) \geq \max\{\sigma_\pi(i) : i \in [n]\} \geq \delta^-(\hat{i}) = \Delta,$$

where the first inequality follows from Lemma 3.2. We conclude that the mechanism is 1-consistent.

For the robustness guarantee, we assume that $\delta^-(\hat{i}) < \Delta$ since otherwise the bound follows trivially. We let $i^* \in \arg \max\{\delta^-(i) : i \in [n]\}$ be a maximum-indegree vertex and note that $i^* \neq \hat{i}$. Importantly, if there is more than one maximum indegree vertex, we fix i^* such that $(\hat{i}, i^*) \notin E$, whose existence is guaranteed in this case as $\hat{i}$ has outdegree one. For $r \in [\Delta]$, we write A_r and B_r instead of $A_r(i^*)$ and $B_r(i^*)$ for compactness.

We aim to bound the expectation of $\delta^-(i^{\text{Pm}})$ in terms of conditional expectations of disjoint events, generalizing the proof by Cembrano et al. [11] to the case where the permutation is no longer taken uniformly at random but with a fixed vertex at the end. We denote $X = \delta^-(i^{\text{Pm}})$ for compactness. Note that we can assume that $\Delta \geq 2$, since otherwise $\delta^-(i) = 1$ for every $i \in [n]$ and $\mathbb{E}[X]$ is trivially equal to Δ .

We observe that the following pairs of events are disjoint:

- (i) $[A_r \cap \neg B_r]$ and $[A_{r'} \cap \neg B_{r'}]$ are disjoint for $r \neq r'$, because $A_r \cap A_{r'} = \emptyset$;
- (ii) $[A_s \cap B_r \cap \neg B_{r+1}]$ and $[A_{s'} \cap B_r \cap \neg B_{r+1}]$ are disjoint for $s \neq s'$ and any r, r' , because $A_s \cap A_{s'} = \emptyset$;
- (iii) $[A_s \cap B_r \cap \neg B_{r+1}]$ and $[A_s \cap B_{r'} \cap \neg B_{r'+1}]$ are disjoint for $r \neq r'$ and any s , because the former implies $\max\{\sigma_\pi(i) : i \in [n] \setminus \{i^*\}\} = r$ and the latter implies $\max\{\sigma_\pi(i) : i \in [n] \setminus \{i^*\}\} = r'$;
- (iv) $[A_r \cap \neg B_r]$ and $[A_s \cap B_{r'} \cap \neg B_{r'+1}]$ are disjoint for $s \leq r'$, because the former implies $\sigma_\pi(i^*) > \max\{\sigma_\pi(i) : i \in [n] \setminus \{i^*\}\}$ and the latter implies the opposite inequality.

In what follows, we distinguish two cases, depending on whether $(\hat{i}, i^*) \in E$ or not.

We first consider the case where $(\hat{i}, i^*) \in E$ and thus i^* has indegree at most $\Delta - 1$ from the left. Importantly, because of the way i^* was fixed, it is the unique maximum indegree vertex in this case, so that $\sigma_\pi(i) \leq \delta^-(i) \leq \Delta - 1$ for every $i \in [n]$. Since $\pi \in \Pi_n(\hat{i} \rightarrow n)$ is taken uniformly at random besides the fixed vertex i^* , we have $\mathbb{P}[A_r] = \frac{1}{\Delta}$ for every $r \in \{0, 1, \dots, \Delta - 1\}$. Thus, for every $r \in \{0, 1, \dots, \Delta - 1\}$ and $s \in \{0, 1, \dots, r\}$

$$\mathbb{P}[A_r \cap \neg B_r] = \mathbb{P}[\neg B_r | A_r] \mathbb{P}[A_r] = \frac{1}{\Delta} (1 - \mathbb{P}[B_r | A_r]), \quad (1)$$

$$\mathbb{P}[A_s \cap B_r \cap \neg B_{r+1}] = \mathbb{P}[B_r \cap \neg B_{r+1} | A_s] \mathbb{P}[A_s] = \frac{1}{\Delta} (\mathbb{P}[B_r | A_s] - \mathbb{P}[B_{r+1} | A_s]), \quad (2)$$

where we used that $B_{r+1} \subseteq B_r$ in the second chain of equalities. By Lemma 3.2, we further know that $i^{\text{Pm}} = i^*$ and thus $\delta^-(i^{\text{Pm}}) = \Delta$ whenever $\sigma_\pi(i^*) > \sigma_\pi(i)$ holds for every $i \neq i^*$, and that $\delta^-(i^{\text{Pm}}) \geq r$ whenever $\sigma_\pi(i) \geq r$ holds for some r . Thus,

$$\mathbb{E}[X | A_r \cap \neg B_r] = \Delta, \quad \text{and} \quad \mathbb{E}[X | A_s \cap B_r \cap \neg B_{r+1}] \geq r \quad (3)$$

for every $r \in \{0, 1, \dots, \Delta - 1\}$ and $s \in \{0, 1, \dots, r\}$.

We now combine the previous observations to obtain the following chain of inequalities:

$$\begin{aligned} \mathbb{E}[X] &\geq \sum_{r=1}^{\Delta-1} \mathbb{E}[X | A_r \cap \neg B_r] \mathbb{P}[A_r \cap \neg B_r] + \sum_{r=0}^{\Delta-1} \sum_{s=0}^r \mathbb{E}[X | A_s \cap B_r \cap \neg B_{r+1}] \mathbb{P}[A_s \cap B_r \cap \neg B_{r+1}] \\ &\geq \frac{1}{\Delta} \left(\Delta \sum_{r=1}^{\Delta-1} (1 - \mathbb{P}[B_r | A_r]) + \sum_{r=0}^{\Delta-1} r \sum_{s=0}^r (\mathbb{P}[B_r | A_s] - \mathbb{P}[B_{r+1} | A_s]) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\Delta} \left(\Delta(\Delta - 1) - \Delta \sum_{r=1}^{\Delta-1} \mathbb{P}[B_r | A_r] + \sum_{r=1}^{\Delta-1} r \sum_{s=0}^r \mathbb{P}[B_r | A_s] - \sum_{r=2}^{\Delta} (r-1) \sum_{s=0}^{r-1} \mathbb{P}[B_r | A_s] \right) \\
&= \frac{1}{\Delta} \left(\Delta(\Delta - 1) - \Delta \sum_{r=1}^{\Delta-1} \mathbb{P}[B_r | A_r] + \sum_{r=1}^{\Delta-1} r \sum_{s=0}^r \mathbb{P}[B_r | A_s] - \sum_{r=1}^{\Delta-1} (r-1) \sum_{s=0}^{r-1} \mathbb{P}[B_r | A_s] \right) \\
&= \frac{1}{\Delta} \left(\Delta(\Delta - 1) - \Delta \sum_{r=1}^{\Delta-1} \mathbb{P}[B_r | A_r] + \sum_{r=1}^{\Delta-1} \sum_{s=0}^{r-1} \mathbb{P}[B_r | A_s] + \sum_{r=1}^{\Delta-1} r \mathbb{P}[B_r | A_r] \right).
\end{aligned}$$

The first inequality holds because the sum is over disjoint events, as argued in items (i) to (iv); the second inequality because of inequalities (1), (2), and (3); the equalities from observing that B_Δ never holds and from rearranging terms. We can now apply Lemma A.1 to bound $\mathbb{P}[B_r | A_s]$ from below by $\mathbb{P}[B_r | A_r]$ for each $r \in [\Delta - 1]$ and $s \in \{0, 1, \dots, r\}$, and obtain

$$\begin{aligned}
\mathbb{E}[X] &\geq \frac{1}{\Delta} \left(\Delta(\Delta - 1) - \Delta \sum_{r=1}^{\Delta-1} \mathbb{P}[B_r | A_r] + \sum_{r=1}^{\Delta-1} \sum_{s=0}^{r-1} \mathbb{P}[B_r | A_r] + \sum_{r=1}^{\Delta-1} r \mathbb{P}[B_r | A_r] \right) \\
&= \frac{1}{\Delta} \left(\Delta(\Delta - 1) - \sum_{r=1}^{\Delta-1} (\Delta - 2r) \mathbb{P}[B_r | A_r] \right).
\end{aligned}$$

Since $\Delta \geq 2$ and $\mathbb{P}[B_r | A_r] \leq 1$, we can bound the sum from above as follows:

$$\sum_{r=1}^{\Delta-1} (\Delta - 2r) \mathbb{P}[B_r | A_r] \leq \sum_{r=1}^{\lfloor \Delta/2 \rfloor} (\Delta - 2r) \mathbb{P}[B_r | A_r] \leq \sum_{r=1}^{\lfloor \Delta/2 \rfloor} (\Delta - 2r) = \left\lfloor \frac{\Delta}{2} \right\rfloor \left(\Delta - \left\lfloor \frac{\Delta}{2} \right\rfloor - 1 \right). \quad (4)$$

Thus, if Δ is even,

$$\frac{\mathbb{E}[X]}{\Delta} \geq \frac{1}{\Delta^2} \left(\Delta(\Delta - 1) - \frac{\Delta}{2} \left(\frac{\Delta}{2} - 1 \right) \right) = \frac{3\Delta - 2}{4\Delta},$$

and if Δ is odd,

$$\frac{\mathbb{E}[X]}{\Delta} \geq \frac{1}{\Delta^2} \left(\Delta(\Delta - 1) - \frac{\Delta - 1}{2} \left(\frac{\Delta + 1}{2} - 1 \right) \right) = \frac{3\Delta^2 - 2\Delta - 1}{4\Delta^2}.$$

We now consider the case where $(i, i^*) \notin E$. Since $\pi \in \Pi_n(i \rightarrow n)$ is taken uniformly at random besides the fixed vertex i^* , we now have $\mathbb{P}[A_r] = \frac{1}{\Delta+1}$ for every $r \in \{0, 1, \dots, \Delta\}$. Thus, for every $r \in \{0, 1, \dots, \Delta\}$ and $s \in \{0, 1, \dots, r\}$

$$\mathbb{P}[A_r \cap \neg B_r] = \mathbb{P}[\neg B_r | A_r] \mathbb{P}[A_r] = \frac{1}{\Delta+1} (1 - \mathbb{P}[B_r | A_r]), \quad (5)$$

$$\mathbb{P}[A_s \cap B_r \cap \neg B_{r+1}] = \mathbb{P}[B_r \cap \neg B_{r+1} | A_s] \mathbb{P}[A_s] = \frac{1}{\Delta+1} (\mathbb{P}[B_r | A_s] - \mathbb{P}[B_{r+1} | A_s]), \quad (6)$$

where we used that $B_{r+1} \subseteq B_r$ in the second chain of equalities. By Lemma 3.2, we further know that $i^{\text{Pm}} = i^*$ and thus $\delta^-(i^{\text{Pm}}) = \Delta$ whenever $\sigma_\pi(i^*) > \sigma_\pi(i)$ holds for every $i \neq i^*$, and that $\delta^-(i^{\text{Pm}}) \geq r$ whenever $\sigma_\pi(i) \geq r$ holds for some r . Thus,

$$\mathbb{E}[X | A_r \cap \neg B_r] = \Delta, \quad \text{and} \quad \mathbb{E}[X | A_s \cap B_r \cap \neg B_{r+1}] \geq r \quad (7)$$

for every $r \in \{0, 1, \dots, \Delta\}$ and $s \in \{0, 1, \dots, r\}$.

We now combine the previous observations to obtain the following chain of inequalities:

$$\begin{aligned}
\mathbb{E}[X] &\geq \sum_{r=1}^{\Delta} \mathbb{E}[X \mid A_r \cap \neg B_r] \mathbb{P}[A_r \cap \neg B_r] \\
&\quad + \sum_{r=0}^{\Delta} \sum_{s=0}^r \mathbb{E}[X \mid A_s \cap B_r \cap \neg B_{r+1}] \mathbb{P}[A_s \cap B_r \cap \neg B_{r+1}] \\
&\geq \frac{1}{\Delta+1} \left(\Delta \sum_{r=1}^{\Delta} (1 - \mathbb{P}[B_r \mid A_r]) + \sum_{r=0}^{\Delta} r \sum_{s=0}^r (\mathbb{P}[B_r \mid A_s] - \mathbb{P}[B_{r+1} \mid A_s]) \right) \\
&= \frac{1}{\Delta+1} \left(\Delta^2 - \Delta \sum_{r=1}^{\Delta} \mathbb{P}[B_r \mid A_r] + \sum_{r=1}^{\Delta} r \sum_{s=0}^r \mathbb{P}[B_r \mid A_s] - \sum_{r=2}^{\Delta+1} (r-1) \sum_{s=0}^{r-1} \mathbb{P}[B_r \mid A_s] \right) \\
&= \frac{1}{\Delta+1} \left(\Delta^2 - \Delta \sum_{r=1}^{\Delta} \mathbb{P}[B_r \mid A_r] + \sum_{r=1}^{\Delta} r \sum_{s=0}^r \mathbb{P}[B_r \mid A_s] - \sum_{r=1}^{\Delta} (r-1) \sum_{s=0}^{r-1} \mathbb{P}[B_r \mid A_s] \right) \\
&= \frac{1}{\Delta+1} \left(\Delta^2 - \Delta \sum_{r=1}^{\Delta} \mathbb{P}[B_r \mid A_r] + \sum_{r=1}^{\Delta} \sum_{s=0}^{r-1} \mathbb{P}[B_r \mid A_s] + \sum_{r=1}^{\Delta} r \mathbb{P}[B_r \mid A_r] \right).
\end{aligned}$$

The first inequality holds because the sum is over disjoint events, as argued in items (i) to (iv); the second inequality because of inequalities (5), (6), and (7); the equalities from observing that $B_{\Delta+1}$ never holds and from rearranging terms. We can apply Lemma A.1 to bound $\mathbb{P}[B_r \mid A_s]$ from below by $\mathbb{P}[B_r \mid A_r]$ for each $r \in [\Delta]$ and $s \in \{0, 1, \dots, r\}$, and obtain

$$\begin{aligned}
\mathbb{E}[X] &\geq \frac{1}{\Delta+1} \left(\Delta^2 - \Delta \sum_{r=1}^{\Delta} \mathbb{P}[B_r \mid A_r] + \sum_{r=1}^{\Delta} \sum_{s=0}^{r-1} \mathbb{P}[B_r \mid A_r] + \sum_{r=1}^{\Delta} r \mathbb{P}[B_r \mid A_r] \right) \\
&= \frac{1}{\Delta+1} \left(\Delta^2 - \sum_{r=1}^{\Delta} (\Delta - 2r) \mathbb{P}[B_r \mid A_r] \right).
\end{aligned}$$

Since $\Delta \geq 2$ and $\mathbb{P}[B_r \mid A_r] \leq 1$, the bound on the sum in the final expression established in inequality (4) remains valid. Thus, if Δ is even,

$$\frac{\mathbb{E}[X]}{\Delta} \geq \frac{1}{\Delta(\Delta+1)} \left(\Delta^2 - \frac{\Delta}{2} \left(\frac{\Delta}{2} - 1 \right) \right) = \frac{3\Delta+2}{4(\Delta+1)},$$

and if Δ is odd,

$$\frac{\mathbb{E}[X]}{\Delta} \geq \frac{1}{\Delta(\Delta+1)} \left(\Delta^2 - \frac{\Delta-1}{2} \left(\frac{\Delta+1}{2} - 1 \right) \right) = \frac{3\Delta-1}{4\Delta}.$$

We conclude that for any given Δ , the lower bounds on $\frac{\mathbb{E}[X]}{\Delta}$ in this case are larger than in the case with $(\hat{i}, i^*) \in E$. Those obtained in the previous case are thus valid bounds on the robustness of the mechanism.

For the last part of the statement, we show that β is an increasing function in $\Delta \geq 2$. To show this property, we distinguish between even and odd values of Δ . If $\Delta \geq 2$ is even, we have that

$$\beta(\Delta+1) - \beta(\Delta) = \frac{3(\Delta+1)^2 - 2(\Delta+1) - 1}{4(\Delta+1)^2} - \frac{3\Delta-2}{4\Delta} = \frac{\Delta+2}{4\Delta(\Delta+1)^2} > 0.$$

Similarly, if $\Delta \geq 3$ is odd, we have that

$$\beta(\Delta + 1) - \beta(\Delta) = \frac{3(\Delta + 1) - 2}{4(\Delta + 1)} - \frac{3\Delta^2 - 2\Delta - 1}{4\Delta^2} = \frac{3\Delta + 1}{4\Delta^2(\Delta + 1)}.$$

We conclude that β is increasing in $\Delta \geq 2$, as claimed. Since $\Delta \geq 2$ and $\beta(2) = \frac{1}{2}$, we conclude that the mechanism is $\frac{1}{2}$ -robust on plurality graphs. $\square$

A.3 Proof of Corollary 3.4

We claim the result for the mechanism that runs the 1-permutation mechanisms with probability ρ and the random permutation mechanism with probability $1 - \rho$. The former is 1-consistent and $\beta_1(\Delta)$ -robust on plurality graphs with maximum indegree Δ , where

$$\beta_1(\Delta) = \begin{cases} \frac{3\Delta-2}{4\Delta} & \text{if } \Delta \text{ is even,} \\ \frac{3\Delta^2-2\Delta-1}{4\Delta^2} & \text{if } \Delta \text{ is odd,} \end{cases}$$

as established in Theorem 3.3. The random permutation mechanism was shown by Cembrano et al. [11] to be $\beta_2(\Delta)$ -robust on plurality graphs with maximum indegree Δ , where

$$\beta_2(\Delta) = \begin{cases} \frac{3\Delta+2}{4(\Delta+1)} & \text{if } \Delta \text{ is even,} \\ \frac{3\Delta-1}{4\Delta} & \text{if } \Delta \text{ is odd.} \end{cases}$$

Thus, the mixture between these mechanisms with parameter ρ is $\alpha(\Delta)$ -consistent and $\beta(\Delta)$ -robust on graphs with maximum indegree Δ , where

$$\begin{aligned} \alpha(\Delta) &= \begin{cases} \rho + \frac{3\Delta+2}{4(\Delta+1)}(1-\rho) = \frac{3\Delta+2}{4(\Delta+1)} + \frac{\Delta+2}{4(\Delta+1)}\rho & \text{if } \Delta \text{ is even,} \\ \rho + \frac{3\Delta-1}{4\Delta}(1-\rho) = \frac{3\Delta-1+(\Delta+1)\rho}{4\Delta} = \alpha(\Delta-1) & \text{if } \Delta \text{ is odd,} \end{cases} \\ \beta(\Delta) &= \begin{cases} \frac{3\Delta-2}{4\Delta}\rho + \frac{3\Delta+2}{4(\Delta+1)}(1-\rho) = \frac{3\Delta+2}{4(\Delta+1)} - \frac{\Delta+2}{4\Delta(\Delta+1)}\rho & \text{if } \Delta \text{ is even,} \\ \frac{3\Delta^2-2\Delta-1}{4\Delta^2}\rho + \frac{3\Delta-1}{4\Delta}(1-\rho) = \frac{3\Delta-1}{4\Delta} - \frac{\Delta+1}{4\Delta^2}\rho & \text{if } \Delta \text{ is odd.} \end{cases} \end{aligned}$$

This concludes the first claim in the statement.

For the second claim, we observe that the functions α and β are non-decreasing in $\Delta \geq 2$ for any fixed $\rho \in [0, 1]$. Indeed, β_1 is non-decreasing in $\Delta \geq 2$, as proven in Theorem 3.3, and β_2 is non-decreasing in $\Delta \geq 2$, as proven by Cembrano et al. [11]. Since α is a fixed convex combination of two non-decreasing functions (the constant function with value 1 and β_2), its monotonicity follows. Similarly, that β is non-decreasing follows from it being a fixed convex combination of two non-decreasing functions (β_1 and β_2). Thus, the guarantees for $\Delta = 2$ are valid for any plurality graph: The $(\frac{2}{3} + \frac{1}{3}\rho)$ -consistency and $(\frac{2}{3} - \frac{1}{6}\rho)$ -robustness follow immediately by computing the previous expressions for $\Delta = 2$. $\square$

B Proofs Deferred from Section 4

B.1 Proof of Theorem 4.1

Impartiality follows directly from the impartiality of the bidirectional permutation mechanism by Bjelde et al. [7], since the outgoing edges play no role in fixing the permutation.

For the approximation guarantees, we fix an arbitrary graph $G = ([n], E)$ and predicted set $\hat{S} = \{\hat{i}_1, \hat{i}_2\}$. We also fix any vector $x \in [0, 1]^n$ with $x_{\hat{i}_1} = 0$, $x_{\hat{i}_2} = 1$, and $x_i \in (0, 1)$ for every

Algorithm 4 Fixed bidirectional permutation mechanism, $\text{Pm}_{\text{bi}}(\hat{S}, G)$

Input: graph $G = ([n], E)$, predicted set $\hat{S} = \{\hat{i}_1, \hat{i}_2\} \subseteq [n]$.

Output: set $S \subseteq [n]$ with $|S| \leq 2$.

Fix $x_{\hat{i}_1} \leftarrow 0$ and $x_{\hat{i}_2} \leftarrow 1$

fix $x_i \in (0, 1)$ arbitrarily for each $i \in [n] \setminus \hat{S}$

$\bar{x}_i \leftarrow 1 - x_i$ for every $i \in [n]$

return $\text{Pm}(G, [n], x) \cup \text{Pm}(G, [n], \bar{x})$

$\in [n] \setminus \{\hat{i}_1, \hat{i}_2\}$, and denote the permutation induced by x by $\pi \in \Pi_n$. In particular, we have $\pi_1 = \hat{i}_1$ and $\pi_n = \hat{i}_2$. Note that π corresponds to the permutation used by the mechanism when running $\text{Pm}(G, [n], x)$ and its reverse $\bar{\pi}$ to the permutation used by the mechanism when running $\text{Pm}(G, [n], \bar{x})$. We denote the vertex output by the former by i_1^{Pm} and that output by the latter by i_2^{Pm} , so that $\text{Pm}_{\text{bi}}(\hat{S}, G) = \{i_1^{\text{Pm}}, i_2^{\text{Pm}}\}$.

To show consistency, we suppose that $\delta^-(\hat{S}) = \Delta_2$ and observe that, in this case

$$\delta^-(\{i_1^{\text{Pm}}, i_2^{\text{Pm}}\}) \geq \delta_{\pi_{< i_2}}^-(i_2) + \delta_{\bar{\pi}_{< i_1}}^-(i_1) = \delta^-(i_2) + \delta^-(i_1) = \Delta_2,$$

where the first inequality follows from Lemma 3.2, the second one from $\pi_1 = \hat{i}_1$ and $\pi_n = \hat{i}_2$, and the third one from the assumption that $\delta^-(\hat{S}) = \Delta_2$. Thus, we conclude that $\frac{1}{\Delta_2} \delta^-(\{i_1^{\text{Pm}}, i_2^{\text{Pm}}\}) \geq 1$; i.e., the mechanism is 1-consistent.

For the robustness guarantee, we let $i^* \in \arg \max\{\delta^-(i) : i \in [n]\}$ be a maximum-indegree vertex. Note that, in particular, this implies that $\delta^-(i^*) \geq \frac{\Delta_2}{2}$. We can bound the indegree selected by the mechanism as follows:

$$\delta^-(\{i_1^{\text{Pm}}, i_2^{\text{Pm}}\}) \geq \delta_{\pi_{< i^*}}^-(i^*) + \delta_{\bar{\pi}_{< i^*}}^-(i^*) = \delta^-(i^*) \geq \frac{\Delta_2}{2},$$

where the first inequality follows from Lemma 3.2, the second one from the fact that $\pi_{< i^*} \cup \bar{\pi}_{< i^*} = [n] \setminus \{i^*\}$, and the third one from the assumption that i^* is a maximum-indegree vertex. We conclude that $\frac{1}{\Delta_2} \delta^-(\{i_1^{\text{Pm}}, i_2^{\text{Pm}}\}) \geq \frac{1}{2}$; i.e., the mechanism is $\frac{1}{2}$ -robust. $\square$

C Proofs Deferred from Section 5

C.1 Proof of Proposition 5.1

We claim the result for the mechanism that, for an input graph $G = ([n], E)$ and predicted set $\hat{S} \in \binom{[n]}{k}$, returns $\{\hat{i}_1, \dots, \hat{i}_{k-2}\} \cup \text{Pm}_{\text{bi}}(\{\hat{i}_{k-1}, \hat{i}_k\}, G)$.

Impartiality follows directly from the impartiality of the bidirectional permutation mechanism by Bjelde et al. [7], since the other $k - 2$ selected vertices are fixed predicted vertices, independent of the input graph.

For the approximation guarantees, we fix a graph $G = ([n], E)$ and a predicted set $\hat{S} \in \binom{[n]}{k}$. To show consistency, we suppose that $\delta^-(\hat{S}) = \Delta_k$ and observe that the indegree of the set selected by the mechanism is

$$\delta^-(\{\hat{i}_1, \dots, \hat{i}_{k-2}\}) + \delta^-(\text{Pm}_{\text{bi}}(\{\hat{i}_{k-1}, \hat{i}_k\}, G)) \geq \delta^-(\{\hat{i}_1, \dots, \hat{i}_k\}) = \delta^-(\hat{S}) = \Delta_k,$$

where the inequality follows from the 1-consistency of the bidirectional permutation mechanism, established in Theorem 4.1. We conclude that the mechanism is 1-consistent. To show robustness,

we observe that the indegree of the set selected by the mechanism is

$$\begin{aligned}
\delta^-(\{\hat{i}_1, \dots, \hat{i}_{k-2}\}) + \delta^-(\text{Pmbi}(\{\hat{i}_{k-1}, \hat{i}_k\}, G)) &\geq \delta^-(\text{Pmbi}(\{\hat{i}_{k-1}, \hat{i}_k\}, G)) \\
&\geq \frac{1}{2} \max\{\delta^-(S) : S \subseteq [n], |S| = 2\} \\
&\geq \frac{1}{2} \cdot \frac{2}{k} \max\{\delta^-(S) : S \subseteq [n], |S| = k\} = \frac{1}{k} \Delta_k,
\end{aligned}$$

where the second inequality follows from the $\frac{1}{2}$ -robustness of the bidirectional permutation mechanism, established in Theorem 4.1. We conclude that the mechanism is $\frac{1}{2}$ -robust. $\square$

C.2 Proof of Theorem 5.2

Impartiality follows directly from the impartiality of the k -partition mechanism, established by Bjelde et al. [7], as both the placement of vertices in the sets and the internal permutations are independent of the outgoing edges of the vertices.

To show both guarantees, we fix an arbitrary graph $G = ([n], E)$, a value $k \in \{2, \dots, n-1\}$, and a predicted set $\hat{S} \in \binom{[n]}{k}$. We let $S^{\text{Pt}} = \{i_1^{\text{Pt}}, \dots, i_k^{\text{Pt}}\}$ denote the set output by the mechanism for this input, where i_j^{Pt} is the vertex selected from each set A_j . For each $j \in [k]$, we denote by x^j and π^j the (random) vector constructed by the mechanism on this set and its associated permutation in Π_{A_j} , respectively.

For the consistency guarantee, we assume that $\delta^-(\hat{S}) = \Delta_k$. From Lemma 3.2, we know that

$$\begin{aligned}
\mathbb{E}[\delta^-(S^{\text{Pt}})] &= \sum_{j=1}^k \mathbb{E}[\delta^-(i_j^{\text{Pt}})] \geq \sum_{j=1}^k \mathbb{E}\left[\delta^-_{([n] \setminus A_j) \cup \pi_{< i_j}^j}(\hat{i}_j)\right] \\
&= \sum_{j=1}^k \sum_{i \in N^-(\hat{i}_j)} \mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i_j}^j]. \tag{8}
\end{aligned}$$

We now observe that, for each $j \in [k]$ and $i \in \hat{S} \setminus \{\hat{i}_j\}$, we have $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i_j}^j] = 1$, because each predicted vertex is assigned to a different set. Furthermore, for each $j \in [k]$ and $i \in [n] \setminus \hat{S}$,

$$\begin{aligned}
\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i_j}^j] &= \mathbb{P}[i \in [n] \setminus A_j] + \mathbb{P}[i \in \pi_{< i_j}^j] = \frac{k-1}{k} + \frac{1}{k} \mathbb{P}[x_i < x_{\hat{i}_j}] \\
&= \frac{k-1}{k} + \frac{\rho}{k} = 1 - \frac{1-\rho}{k},
\end{aligned}$$

since all vertices in $[n] \setminus \hat{S}$ are assigned independently and uniformly at random to a set among $A_1, \dots, A_k$, $x_{\hat{i}_j} = \rho$, and x_i is taken uniformly from the interval $[0, 1]$. We conclude from the previous inequalities that $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i_j}^j] \geq 1 - \frac{1-\rho}{k}$ $j \in [k]$ and every $i \in N^-(\hat{i}_j)$. Replacing in inequality (8), we conclude that

$$\frac{\mathbb{E}[\delta^-(S^{\text{Pt}})]}{\Delta_k} \geq \frac{1}{\Delta_k} \sum_{j=1}^k \left(1 - \frac{1-\rho}{k}\right) \delta^-(\hat{i}_j) = 1 - \frac{1-\rho}{k},$$

i.e., the mechanism is $(1 - \frac{1-\rho}{k})$ -consistent.

For the robustness guarantee, we fix an optimal set of k agents $S^* \in \binom{[n]}{k}$ such that $\delta^-(S^*) = \Delta_k$. We let $p = |S^* \cap \hat{S}|$ denote the number of optimal vertices among the predicted ones and assume

that $p \in \{0, \dots, k-1\}$, since the guarantee follows trivially from the consistency guarantee when $p = k$. We let $j \in [k]$ be an arbitrary index such that $S^* \cap A_j \neq \emptyset$. For $i \in S^* \cap A_j$, we let D_i denote the event that i is chosen among vertices $S^* \cap A_j$ when choosing a vertex in this set uniformly at random, and we write i_j^* for the (random) vertex in $A_j \cap S^*$ taken uniformly at random. We further denote by i_j^{Pt} the vertex selected by the mechanism in this set. From Lemma 3.2, we know that

$$\mathbb{E}[\delta^-(i_j^{\text{Pt}})] \geq \mathbb{E}\left[\delta^-_{([n] \setminus A_j) \cup \pi_{< i_j^*}^j}(i_j^*)\right] = \sum_{i \in N^-(i_j^*)} \mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i_j^*}^j]. \quad (9)$$

In what follows, we proceed to bound the probabilities on the right-hand side of this inequality for each in-neighbor i of i_j^* . To do so, we distinguish whether $i_j^* \in \hat{S}$ or not, and whether its in-neighbor i belongs to the optimal set S^* , to the predicted set $\hat{S}$, or to neither of them. This is necessary because, as we will see, the probability that i belongs to the set $([n] \setminus A_j) \cup \pi_{< i_j^*}^j$ depends on these facts. Since j will remain fixed, we write i^* , x , and π instead of i_j^* , x_j , and π_j for compactness.

We first consider the simplest case when $i \in N^-(i^*)$ is such that $i \notin S^* \cup \hat{S}$. Since vertices in $[n] \setminus \hat{S}$ are assigned independently and uniformly at random to one of the k sets, and since the event D_{i^*} does not affect the distribution of i due to $i \notin S^*$, we have

$$\begin{aligned} \mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{< i^*} \mid D_{i^*}] &= \mathbb{P}[i \in ([n] \setminus A_j) \mid D_{i^*}] + \mathbb{P}[i \in A_j, x_i < x_{i^*} \mid D_{i^*}] \\ &\geq \frac{k-1}{k} + \frac{1}{2k}. \end{aligned} \quad (10)$$

Indeed, the equality follows directly from the definition of the sets, while the inequality follows from two facts. First, we use that the sets to which i^* and i belong distribute independently and uniformly at random because $i \notin S^* \cup \hat{S}$. Second, we use that i^* is taken uniformly at random among optimal agents in A_j and i 's position in the permutation is taken uniformly at random from $[0, 1]$. Thus, conditional on $i \in A_j$, we have $x_i < x_{i^*}$ with probability $\frac{1}{2}$ if $i^* \notin \hat{S}$ and with probability ρ if $i^* \in \hat{S}$; the inequality then follows since $\rho \geq \frac{1}{2}$. In what follows, we only need to consider cases with $i \in S^* \cup \hat{S}$.

We next consider the case where $i^* \in \hat{S}$ and start by bounding $\mathbb{P}[i \in A_j \mid D_{i^*}]$. If $i \in \hat{S} \setminus \{i^*\}$, we have that $\mathbb{P}[i \in A_j \mid D_{i^*}] = 0$ since predicted vertices are assigned to different sets. If $i \in S^* \setminus \hat{S}$, on the other hand, we have

$$\begin{aligned} &\mathbb{P}[i \in A_j \mid D_{i^*}] \\ &= \frac{\mathbb{P}[D_{i^*} \mid i \in A_j]}{\mathbb{P}[D_{i^*}]} \mathbb{P}[i \in A_j] \\ &= \frac{\sum_{\ell=0}^{k-p-1} \mathbb{P}[D_{i^*} \mid i \in A_j, (S^* \setminus (\hat{S} \cup \{i\})) \cap A_j = \ell] \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i\})) \cap A_j = \ell]}{\sum_{\ell=0}^{k-p} \mathbb{P}[D_{i^*} \mid (S^* \setminus \hat{S}) \cap A_j = \ell] \mathbb{P}[(S^* \setminus \hat{S}) \cap A_j = \ell]} \mathbb{P}[i \in A_j] \\ &= \frac{\mathbb{E}\left[\frac{1}{X+1} \mid X \geq 1\right]}{\mathbb{E}\left[\frac{1}{X+1}\right]} \mathbb{P}[i \in A_j] \leq \mathbb{P}[i \in A_j] = \frac{1}{k}, \end{aligned}$$

where $X \sim B(k-p, \frac{1}{k})$ represents a binomial random variable. Indeed, the first equality follows from Bayes rule; the second and third equalities follow from the facts that $i^* \in \hat{S}$ and $i \in S^* \setminus \hat{S}$, the definition of the event D_{i^*} , and the fact that the vertices in $[n] \setminus \hat{S}$ are assigned to a set independently and uniformly at random. The last equality follows from this last fact as well. We conclude that, no

matter whether i is in $\hat{S}$ or in $S^* \setminus \hat{S}$, we have $\mathbb{P}[i \in A_j \mid D_{i^*}] \leq \frac{1}{k}$. Therefore,

$$\begin{aligned}
\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] &= \mathbb{P}[i \in ([n] \setminus A_j) \mid D_{i^*}] + \mathbb{P}[i \in A_j, x_i < x_{i^*} \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \mathbb{P}[x_i < x_{i^*} \mid i \in A_j] \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \rho \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - (1 - \rho) \mathbb{P}[i \in A_j \mid D_{i^*}] \geq 1 - \frac{1 - \rho}{k}.
\end{aligned} \tag{11}$$

Indeed, the second equality holds because the events $x_i < x_{i^*}$ and D_{i^*} are independent since i^* is chosen uniformly at random, the third one because $x_{i^*} = \rho$ and, conditional on $i \in A_j$, x_i is sampled uniformly in $[0, 1]$, and the last inequality because of the previous bound on $\mathbb{P}[i \in A_j \mid D_{i^*}]$.

In what follows, we consider cases with $i^* \in S^* \setminus \hat{S}$ and $i \in S^* \cup \hat{S}$, and we make use of an explicit expression for $\mathbb{P}[D_{i^*}]$. From the way the partition is computed and since i^* is chosen uniformly at random among the vertices in $S^* \cap A_j$, we have

$$\begin{aligned}
\mathbb{P}[D_{i^*}] &= \mathbb{P}[D_{i^*} \mid \hat{i}_j \in S^*] \mathbb{P}[\hat{i}_j \in S^*] + \mathbb{P}[D_{i^*} \mid \hat{i}_j \notin S^*] \mathbb{P}[\hat{i}_j \notin S^*] \\
&= \frac{p}{k} \sum_{\ell=0}^{k-p-1} \mathbb{P}[D_{i^*} \mid \hat{i}_j \in S^*, (S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell] \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell] \\
&\quad + \frac{k-p}{k} \sum_{\ell=0}^{k-p-1} \mathbb{P}[D_{i^*} \mid \hat{i}_j \notin S^*, (S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell] \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell] \\
&= \frac{p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\
&\quad + \frac{k-p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell}.
\end{aligned} \tag{12}$$

Note that, in particular, we have used for the second equality that the events $(S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell$ and $\hat{i}_j \in S^*$ are independent, because vertices in $S^* \setminus \hat{S}$ are assigned to sets independently and uniformly at random. We will make use of this expression in the cases that follow.

For the cases where $i \in S^* \cap \hat{S}$ or $i \in S^* \setminus \hat{S}$, we state the corresponding bounds on the probability $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}]$ in the following claims and defer their respective proofs to Sections C.2.1 and C.2.2.

Claim C.1. *If $i^* \in S^* \setminus \hat{S}$ and $i \in N^-(i^*) \cap S^* \cap \hat{S}$, then $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] \geq 1 - \frac{\rho}{k}$.*

Claim C.2. *If $i^* \in S^* \setminus \hat{S}$ and $i \in (N^-(i^*) \cap S^*) \setminus \hat{S}$, then $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] \geq 1 - \frac{1}{2k}$.*

We finally consider the case with $i^* \in S^* \setminus \hat{S}$ and $i \in \hat{S} \setminus S^*$. Since $\mathbb{P}[D_{i^*} \mid i \in A_j]$ cannot be bounded from above by $\mathbb{P}[D_{i^*}]$, we proceed by directly computing a lower bound on $\mathbb{P}[D_{i^*}]$. We start by computing both sums on the right-hand side of equality (12) to reach a simpler expression for this probability. The following claim states the result of this computation, which can be found in Section C.2.3.

Claim C.3. $\mathbb{P}[D_{i^*}] = \frac{k(k-2p+1) - (k^2 + k - 3pk + p^2) \left(\frac{k-1}{k}\right)^{k-p}}{(k-p+1)(k-p)}.$

The following claim allows us to compute a lower bound for the expression we have computed for $\mathbb{P}[D_{i^*}]$.

Claim C.4. For any $k \in \mathbb{N}$, the function $g_k: \{0, \dots, k-1\} \rightarrow \mathbb{R}$ defined as

$$g_k(p) = \frac{k(k-2p+1) - (k^2 + k - 3pk + p^2)\left(\frac{k-1}{k}\right)^{k-p}}{(k-p+1)(k-p)}$$

is non-increasing in p .

The proof of this claim is deferred to Section C.2.4. It proceeds by showing that the expression $g_k(p) - g_k(p+1)$ is non-negative for $p \in \{0, \dots, k-2\}$.

Equipped with the previous claims, we can now bound $\mathbb{P}[D_{i^*}]$ by the value of g_k at $p = k-1$. Indeed, we conclude from Claim C.3 and Claim C.4 that

$$\begin{aligned} \mathbb{P}[D_{i^*}] \geq g(k-1) &= \frac{k(k-2(k-1)+1) - (k^2 + k - 3k(k-1) + (k-1)^2)\left(\frac{k-1}{k}\right)}{2} \\ &= \frac{k^2(-k+3) - (k-1)(-k^2+2k+1)}{2k} = \frac{k+1}{2k}, \end{aligned}$$

and proceed as in the other cases. Specifically, we observe that

$$\mathbb{P}[i \in A_j \mid D_{i^*}] = \frac{\mathbb{P}[D_{i^*} \mid i \in A_j]}{\mathbb{P}[D_{i^*}]} \mathbb{P}[i \in A_j] \leq \frac{1}{\frac{k+1}{2k}} \mathbb{P}[i \in A_j] = \frac{2k}{k+1} \cdot \frac{1}{k} = \frac{2}{k+1},$$

where the first equality follows from Bayes' rule, the inequality from the previous bound on $\mathbb{P}[D_{i^*}]$, and the second equality from the fact that i is assigned to a set uniformly at random. We obtain

$$\begin{aligned} \mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] &= \mathbb{P}[i \in ([n] \setminus A_j) \mid D_{i^*}] + \mathbb{P}[i \in A_j, x_i < x_{i^*} \mid D_{i^*}] \\ &= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \mathbb{P}[x_i < x_{i^*} \mid i \in A_j] \mathbb{P}[i \in A_j \mid D_{i^*}] \\ &= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + (1 - \rho) \mathbb{P}[i \in A_j \mid D_{i^*}] \\ &= 1 - \rho \mathbb{P}[i \in A_j \mid D_{i^*}] \geq 1 - \frac{2\rho}{k+1}. \end{aligned} \tag{13}$$

Indeed, the second equality holds because the events $x_i < x_{i^*}$ and D_{i^*} are independent, the third one because $x_i = \rho$ and x_{i^*} is sampled independently and uniformly in $[0, 1]$, and the last inequality because of the previous bound on $\mathbb{P}[i \in A_j \mid D_{i^*}]$.

We can now combine all lower bounds on $\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}]$ we have computed to conclude the robustness guarantee in the statement. Specifically, by combining inequalities (10), (11), and (13), as well as Claim C.1 and Claim C.2, we obtain

$$\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] \geq \min \left\{ 1 - \frac{1}{2k}, 1 - \frac{1-\rho}{k}, 1 - \frac{\rho}{k}, 1 - \frac{2\rho}{k+1} \right\} = 1 - \frac{2\rho}{k+1},$$

where we used that $\rho \geq \frac{1}{2}$ and $k \geq 2$. Replacing in inequality (9), we obtain

$$\mathbb{E}[\delta^-(i_j^{\text{Pt}})] \geq \frac{1}{k} \sum_{i^* \in S^*} \sum_{i \in N^-(i^*)} \mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*}^j \mid D_{i^*}] \geq \left(1 - \frac{2\rho}{k+1}\right) \frac{\Delta_k}{k},$$

where the first inequality holds since i^* distributes uniformly among all vertices in S^* . Finally, since

the previous analysis is valid for all $j \in [k]$ with $S^* \cap A_j \neq \emptyset$, we conclude that

$$\begin{aligned}
\frac{\mathbb{E}[\delta^-(S^{\text{Pt}})]}{\Delta_k} &\geq \frac{1}{\Delta_k} \sum_{j=1}^k \mathbb{E}[\delta^-(i_j^{\text{Pt}})] \mathbb{P}[S^* \cap A_j \neq \emptyset] \\
&= \left(1 - \frac{2\rho}{k+1}\right) \sum_{\ell=1}^k \mathbb{P}[|S^* \cap A_j| = \ell] \\
&= \left(1 - \frac{2\rho}{k+1}\right) \sum_{\ell=1}^k \binom{k}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-\ell} \\
&= \left(1 - \frac{2\rho}{k+1}\right) \left(1 - \left(\frac{k-1}{k}\right)^k\right). \quad \square
\end{aligned}$$

C.2.1 Proof of Claim C.1

We consider $i^* \in S^* \setminus \hat{S}$ and $i \in N^-(i^*)$ with $i \in S^* \cap \hat{S}$. We observe that

$$\begin{aligned}
&\mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= \frac{\mathbb{P}[D_{i^*} \mid i \in A_j]}{\mathbb{P}[D_{i^*}]} \mathbb{P}[i \in A_j] \\
&= \frac{\sum_{\ell=0}^{k-p-1} \mathbb{P}[D_{i^*} \mid i \in A_j, (S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell] \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i^*\})) \cap A_j = \ell]}{\mathbb{P}[D_{i^*}]} \mathbb{P}[i \in A_j] \\
&= \frac{\sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \cdot \mathbb{P}[i \in A_j]}{\frac{p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} + \frac{k-p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell}} \\
&\leq \mathbb{P}[i \in A_j] = \frac{1}{k},
\end{aligned}$$

Similar to previous cases addressed in the proof of Theorem 5.2, the first equality follows from Bayes' rule. The second and third equalities now follow from the facts that $i^* \notin \hat{S}$ and $i \in S^*$, the fact that the vertices in $[n] \setminus \hat{S}$ are assigned to a set independently and uniformly at random, and equality (12). The last equality follows from this last fact as well. The inequality is now straightforward, as every term in the second sum in the denominator is larger than every term in the other sum since $\frac{1}{\ell+1} \geq \frac{1}{\ell+2}$, and the expression in the denominator is a convex combination of both sums. We conclude that $\mathbb{P}[i \in A_j \mid D_{i^*}] \leq \frac{1}{k}$. Therefore,

$$\begin{aligned}
\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] &= \mathbb{P}[i \in ([n] \setminus A_j) \mid D_{i^*}] + \mathbb{P}[i \in A_j, x_i < x_{i^*} \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \mathbb{P}[x_i < x_{i^*} \mid i \in A_j] \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + (1 - \rho) \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - \rho \mathbb{P}[i \in A_j \mid D_{i^*}] \geq 1 - \frac{\rho}{k}.
\end{aligned}$$

As in previous cases, the second equality holds because the events $x_i < x_{i^*}$ and D_{i^*} are independent since i^* is chosen uniformly at random, the third one because $x_i = \rho$ and x_{i^*} is sampled uniformly in $[0, 1]$, and the last inequality because of the previous bound on $\mathbb{P}[i \in A_j \mid D_{i^*}]$. $\square$

C.2.2 Proof of Claim C.2

We consider $i^* \in S^* \setminus \hat{S}$ and $i \in N^-(i^*)$ such that $i \in S^* \setminus \hat{S}$. We let $X \sim B(k-p-1, \frac{1}{k})$ represent a binomial random variable and first observe that, similarly to inequality (12), we have

$$\begin{aligned}
& \mathbb{P}[D_{i^*} \mid i \in A_j] \\
&= \mathbb{P}[D_{i^*} \mid i \in A_j, \hat{i}_j \in S^*] \mathbb{P}[\hat{i}_j \in S^*] + \mathbb{P}[D_{i^*} \mid i \in A_j, \hat{i}_j \notin S^*] \mathbb{P}[\hat{i}_j \notin S^*] \\
&= \frac{p}{k} \sum_{\ell=0}^{k-p-2} \mathbb{P}[D_{i^*} \mid i \in A_j, \hat{i}_j \in S^*, (S^* \setminus (\hat{S} \cup \{i, i^*\})) \cap A_j = \ell] \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i, i^*\})) \cap A_j = \ell] \\
&\quad + \frac{k-p}{k} \sum_{\ell=0}^{k-p-2} \mathbb{P}[D_{i^*} \mid i \in A_j, \hat{i}_j \notin S^*, (S^* \setminus (\hat{S} \cup \{i, i^*\})) \cap A_j = \ell] \cdot \\
&\quad \mathbb{P}[(S^* \setminus (\hat{S} \cup \{i, i^*\})) \cap A_j = \ell] \\
&= \frac{p}{k} \mathbb{E}\left[\frac{1}{X+2} \mid X \geq 1\right] + \frac{k-p}{k} \mathbb{E}\left[\frac{1}{X+1} \mid X \geq 1\right].
\end{aligned}$$

Indeed, this follows from the way the partition is computed, the fact that i^* is chosen uniformly at random among the vertices in $S^* \cap A_j$, the independence between the events $i \in A_j$ and $\hat{i}_j \in S^*$ due to $i \notin \hat{S}$, and the independence between the events $(S^* \setminus (\hat{S} \cup \{i, i^*\})) \cap A_j = \ell$ and $\hat{i}_j \in S^*$ are independent because vertices in $S^* \setminus \hat{S}$ are assigned to sets independently and uniformly at random. This implies

$$\begin{aligned}
\mathbb{P}[i \in A_j \mid D_{i^*}] &= \frac{\mathbb{P}[D_{i^*} \mid i \in A_j]}{\mathbb{P}[D_{i^*}]} \mathbb{P}[i \in A_j] \\
&= \frac{\frac{p}{k} \mathbb{E}\left[\frac{1}{X+2} \mid X \geq 1\right] + \frac{k-p}{k} \mathbb{E}\left[\frac{1}{X+1} \mid X \geq 1\right]}{\frac{p}{k} \mathbb{E}\left[\frac{1}{X+2}\right] + \frac{k-p}{k} \mathbb{E}\left[\frac{1}{X+1}\right]} \mathbb{P}[i \in A_j] \leq \mathbb{P}[i \in A_j] = \frac{1}{k},
\end{aligned}$$

Indeed, the first equality follows from Bayes' rule and the last equality from the fact that vertices in $[n] \setminus \hat{S}$ are assigned to sets uniformly at random. The second equality now follows from the previous bound on $\mathbb{P}[D_{i^*} \mid i \in A_j]$ and from equality (12). The inequality holds because $\mathbb{E}\left[\frac{1}{X+2} \mid X \geq 1\right] \leq E\left[\frac{1}{X+2}\right]$ and $\mathbb{E}\left[\frac{1}{X+1} \mid X \geq 1\right] \leq E\left[\frac{1}{X+1}\right]$. We conclude that $\mathbb{P}[i \in A_j \mid D_{i^*}] \leq \frac{1}{k}$, and thus,

$$\begin{aligned}
\mathbb{P}[i \in ([n] \setminus A_j) \cup \pi_{<i^*} \mid D_{i^*}] &= \mathbb{P}[i \in ([n] \setminus A_j) \mid D_{i^*}] + \mathbb{P}[i \in A_j, x_i < x_{i^*} \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \mathbb{P}[x_i < x_{i^*} \mid i \in A_j] \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - \mathbb{P}[i \in A_j \mid D_{i^*}] + \frac{1}{2} \mathbb{P}[i \in A_j \mid D_{i^*}] \\
&= 1 - \frac{1}{2} \mathbb{P}[i \in A_j \mid D_{i^*}] \geq 1 - \frac{1}{2k}.
\end{aligned}$$

The second equality holds because the events $x_i < x_{i^*}$ and D_{i^*} are independent since i^* is chosen uniformly at random, the third one because both x_i and x_{i^*} are sampled independently and uniformly in $[0, 1]$, and the last inequality because of the previous bound on $\mathbb{P}[i \in A_j \mid D_{i^*}]$. $\square$

C.2.3 Proof of Claim C.3

From equality (12), we know that

$$\begin{aligned}\mathbb{P}[D_{i*}] &= \frac{p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &\quad + \frac{k-p}{k} \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell}.\end{aligned}$$

In what follows, we compute both terms on the right-hand side of this equality to conclude the equality in the statement.

For the second term, we observe that

$$\begin{aligned}& \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &= \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \cdot \frac{(k-p-1)!}{\ell!(k-p-1-\ell)!} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &= \frac{k}{k-p} \sum_{\ell=0}^{k-p-1} \frac{(k-p)!}{(\ell+1)!(k-p-1-\ell)!} \left(\frac{1}{k}\right)^{\ell+1} \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &= \frac{k}{k-p} \sum_{\ell=0}^{k-p-1} \binom{k-p}{\ell+1} \left(\frac{1}{k}\right)^{\ell+1} \left(\frac{k-1}{k}\right)^{k-p-(\ell+1)} \\ &= \frac{k}{k-p} \sum_{\ell=1}^{k-p} \binom{k-p}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-\ell} \\ &= \frac{k}{k-p} \left(1 - \left(\frac{k-1}{k}\right)^{k-p}\right).\end{aligned}\tag{14}$$

For the first term, the computation is similar but more demanding. We start replacing $\frac{1}{\ell+2}$ by $\frac{1}{\ell+1} - \frac{1}{(\ell+2)(\ell+1)}$ to obtain

$$\begin{aligned}& \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &= \sum_{\ell=0}^{k-p-1} \frac{1}{\ell+1} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\ &\quad - \sum_{\ell=0}^{k-p-1} \frac{1}{(\ell+2)(\ell+1)} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell}\end{aligned}\tag{15}$$

$$= \frac{k}{k-p} \left(1 - \left(\frac{k-1}{k}\right)^{k-p}\right) - \sum_{\ell=0}^{k-p-1} \frac{1}{(\ell+2)(\ell+1)} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell},\tag{16}$$

where the second equality follows from equality (14). For the other sum on the right-hand side, we obtain

$$\sum_{\ell=0}^{k-p-1} \frac{1}{(\ell+2)(\ell+1)} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell}$$

$$\begin{aligned}
&= \sum_{\ell=0}^{k-p-1} \frac{1}{(\ell+2)(\ell+1)} \cdot \frac{(k-p-1)!}{\ell!(k-p-1-\ell)!} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\
&= \frac{k^2}{(k-p+1)(k-p)} \sum_{\ell=0}^{k-p-1} \frac{(k-p+1)!}{(\ell+2)!(k-p-1-\ell)!} \left(\frac{1}{k}\right)^{\ell+2} \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\
&= \frac{k^2}{(k-p+1)(k-p)} \sum_{\ell=0}^{k-p-1} \binom{k-p+1}{\ell+2} \left(\frac{1}{k}\right)^{\ell+2} \left(\frac{k-1}{k}\right)^{k-p+1-(\ell+2)} \\
&= \frac{k^2}{(k-p+1)(k-p)} \sum_{\ell=2}^{k-p+1} \binom{k-p+1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p+1-\ell} \\
&= \frac{k^2}{(k-p+1)(k-p)} \left(1 - \left(\frac{k-1}{k}\right)^{k-p+1} - \frac{k-p+1}{k} \left(\frac{k-1}{k}\right)^{k-p}\right) \\
&= \frac{k}{(k-p+1)(k-p)} \left(k - (2k-p) \left(\frac{k-1}{k}\right)^{k-p}\right)
\end{aligned}$$

Replacing in equality (16), we obtain

$$\begin{aligned}
&\sum_{\ell=0}^{k-p-1} \frac{1}{\ell+2} \binom{k-p-1}{\ell} \left(\frac{1}{k}\right)^\ell \left(\frac{k-1}{k}\right)^{k-p-1-\ell} \\
&= \frac{k}{(k-p+1)(k-p)} \left((k-p+1) - (k-p+1) \left(\frac{k-1}{k}\right)^{k-p} - k + (2k-p) \left(\frac{k-1}{k}\right)^{k-p} \right) \\
&= \frac{k}{(k-p+1)(k-p)} \left((k-1) \left(\frac{k-1}{k}\right)^{k-p} - (p-1) \right).
\end{aligned}$$

Combining this equality with equalities (12) and (14), we conclude that

$$\begin{aligned}
\mathbb{P}[D_{i^*}] &= \frac{p}{(k-p+1)(k-p)} \left((k-1) \left(\frac{k-1}{k}\right)^{k-p} - (p-1) \right) + 1 - \left(\frac{k-1}{k}\right)^{k-p} \\
&= \frac{(k-p+1)(k-p) - p(p-1) - ((k-p+1)(k-p) - p(k-1)) \left(\frac{k-1}{k}\right)^{k-p}}{(k-p+1)(k-p)} \\
&= \frac{k(k-2p+1) - (k^2 + k - 3pk + p^2) \left(\frac{k-1}{k}\right)^{k-p}}{(k-p+1)(k-p)}. \quad \square
\end{aligned}$$

C.2.4 Proof of Claim C.4

We compute the difference between the value of $g_k(p)$ at two consecutive values of p . For each $p \in \{0, \dots, k-2\}$, we have

$$\begin{aligned}
&g_k(p) - g_k(p+1) \\
&= \frac{k(k-2p+1) - (k^2 + k - 3pk + p^2) \left(\frac{k-1}{k}\right)^{k-p}}{(k-p+1)(k-p)} \\
&\quad - \frac{k(k-2p-1) - (k^2 - 2k - 3pk + p^2 + 2p+1) \left(\frac{k-1}{k}\right)^{k-p-1}}{(k-p)(k-p-1)} \\
&= \frac{k(k-p-1)(k-2p+1) - k(k-p+1)(k-2p-1)}{(k-p+1)(k-p)(k-p-1)}
\end{aligned}$$

$$\begin{aligned}
& - \frac{(k-1)(k^2+k-3pk+p^2)(k-p-1) - k(k^2-2k-3pk+p^2+2p+1)(k-p+1)}{k(k-p+1)(k-p)(k-p-1)} \left(\frac{k-1}{k}\right)^{k-p-1} \\
& = \frac{2pk}{(k-p+1)(k-p)(k-p-1)} - \frac{(k(5k-4p-3)+p^2+p)p}{k(k-p+1)(k-p)(k-p-1)} \left(\frac{k-1}{k}\right)^{k-p-1} \\
& = \frac{p}{(k-p+1)(k-p)(k-p-1)(k-1)} \left(2k(k-1) - \left(\frac{k-1}{k}\right)^{k-p} (k(5k-4p-3)+p^2+p)\right).
\end{aligned}$$

To conclude, we now prove that the last expression is non-negative for every $p \in \{0, \dots, k-2\}$. To see this, let

$$h(p, k) = 2k(k-1) - \left(\frac{k-1}{k}\right)^{k-p} (k(5k-4p-3)+p^2+p).$$

Then

$$h(k, k) = 2k(k-1) - \left(\frac{k-1}{k}\right)^{k-k} (k(5k-4k-3)+k^2+k) = 2k(k-1) - (2k^2-2k) = 0$$

and

$$\begin{aligned}
h(p-1, k) - h(p, k) & = 2k(k-1) - \left(\frac{k-1}{k}\right)^{k-p+1} (k(5k-4(p-1)-3) + (p-1)^2 + (p-1)) \\
& \quad - 2k(k-1) + \left(\frac{k-1}{k}\right)^{k-p} (k(5k-4p-3) + p + p^2) \\
& = \left(\frac{k-1}{k}\right)^{k-p} (5k^2 - 4kp - 3k + p^2 + p) \\
& \quad - \left(\frac{k-1}{k}\right)^{k-p} \frac{k-1}{k} (5k^2 + k - 4kp + p^2 - p) \\
& = \left(\frac{k-1}{k}\right)^{k-p} \frac{1}{k} (5k^3 - 3k^2 - 4k^2p + kp^2 + kp) \\
& \quad - \left(\frac{k-1}{k}\right)^{k-p} \frac{1}{k} (5k^3 - 4k^2 - k - 4k^2p + 3kp + kp^2 - p^2 + p) \\
& = \left(\frac{k-1}{k}\right)^{k-p} \frac{1}{k} (k^2 + k - 2kp + p^2 - p) \\
& = \left(\frac{k-1}{k}\right)^{k-p} \frac{(k-p)(k-p+1)}{k} \geq 0.
\end{aligned}$$

We conclude that $h(p, k) \geq 0$ for every $p \in \{0, \dots, k\}$ and thus $g_k(p) - g_k(p+1) \geq 0$ for every $p \in \{0, \dots, k-2\}$. $\square$

D Proofs Deferred from Section 6

D.1 Proof of Theorem 6.1

We first state and prove a lemma that allows us to restrict to mechanisms that assign the same probabilities to symmetric vertices in the input graph, as long as they all belong or all do not belong

to the predicted set. This is a natural extension to our setting of a lemma first formulated by Holzman and Moulin [19] and used extensively to prove upper bounds on the performance guarantees of impartial mechanisms [7, 11, 17].

For a set of vertices $[n]$ and a predicted set $\hat{S} \in \binom{[n]}{k}$, we say that a permutation $\pi = (\pi_1, \dots, \pi_n)$ of $[n]$ is $\hat{S}$ -invariant if, for every $j \in [n]$, $j \in \hat{S} \Leftrightarrow \pi_j \in \hat{S}$; we denote the set of such permutations by $\Pi_n^{\hat{S}}$. A k -selection mechanism with predictions f is *symmetric* if it is invariant with respect to renaming of the predicted and non-predicted vertices: For every $G = ([n], E) \in \mathcal{G}_n$, every $\hat{S} \in \binom{[n]}{k}$, every $i \in [n]$, and every $\hat{S}$ -invariant permutation $\pi = (\pi_1, \dots, \pi_n)$ of $[n]$, it holds that $(f(\hat{S}, G_\pi))_{\pi_i} = (f(\hat{S}, G))_i$, where $G_\pi = ([n], E_\pi)$ with $E_\pi = \{(j, j') : (j, j') \in E\}$. For a given k -selection mechanism with predictions f and a given predicted set $\hat{S}$ of size k , we denote by f_s the mechanism obtained by applying an $\hat{S}$ -invariant permutation π taken uniformly at random to the vertices of the input graph, invoking f , and permuting the result by the inverse of π . Thus, for all $n \in \mathbb{N}$, $G \in \mathcal{G}_n$, and $i \in [n]$,

$$(f_s(\hat{S}, G))_i = \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} (f(\hat{S}, G_\pi))_{\pi_i}.$$

Lemma D.1. *Let f be a k -selection mechanism with predictions that is impartial, α -consistent, and β -robust on $\mathcal{G}_n$. Then, f_s is symmetric, impartial, α -consistent, and β -robust on $\mathcal{G}_n$.*

Proof. Let f be as in the statement. To prove that f_s is impartial, let $G = ([n], E)$, $G' = ([n], E') \in \mathcal{G}_n$, $\hat{S} \in \binom{[n]}{k}$, and $i \in [n]$ such that $E \setminus (\{i\} \times [n]) = E' \setminus (\{i\} \times [n])$. Since f is impartial,

$$(f_s(\hat{S}, G))_i = \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} (f(\hat{S}, G_\pi))_{\pi_i} = \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} (f(\hat{S}, G'_\pi))_{\pi_i} = (f_s(\hat{S}, G'))_i,$$

and thus f_s is impartial.

To prove that f_s is α -consistent, let $G = ([n], E) \in \mathcal{G}_n$ and $\hat{S} \in \binom{[n]}{k}$ be such that $\delta^-(\hat{S}, G) = \Delta_k(G)$. Then,

$$\begin{aligned} \sum_{i \in [n]} (f_s(\hat{S}, G))_i \delta^-(i, G) &= \sum_{i \in [n]} \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} (f(\hat{S}, G_\pi))_{\pi_i} \delta^-(i, G) \\ &= \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} \sum_{i \in [n]} (f(\hat{S}, G_\pi))_{\pi_i} \delta^-(i, G) \\ &\geq \alpha \Delta_k(G), \end{aligned}$$

where the last inequality holds because f is α -consistent.

Similarly, to prove that f_s is β -robust, we fix $G = ([n], E) \in \mathcal{G}_n$ and $\hat{S} \in \binom{[n]}{k}$. Then,

$$\begin{aligned} \sum_{i \in [n]} (f_s(\hat{S}, G))_i \delta^-(i, G) &= \sum_{i \in [n]} \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} (f(\hat{S}, G_\pi))_{\pi_i} \delta^-(i, G) \\ &= \frac{k!(n-k)!}{n!} \sum_{\pi \in \Pi_n^{\hat{S}}} \sum_{i \in [n]} (f(\hat{S}, G_\pi))_{\pi_i} \delta^-(i, G) \\ &\geq \beta \Delta_k(G), \end{aligned}$$

where the last inequality holds because f is β -robust. $\square$



Figure 3: Impartial 1-selection from n -vertex graphs. The predicted vertex is shown in white. Only 2 vertices are shown; the remaining $n - 2$ vertices do not have any incident edges.

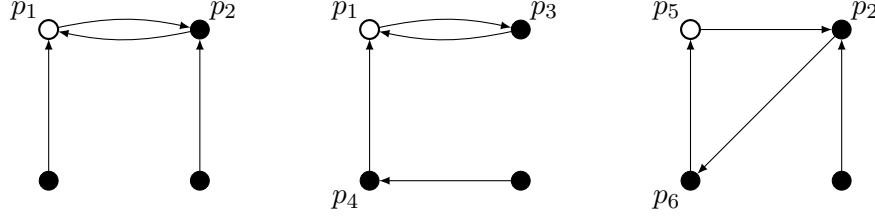


Figure 4: Impartial 1-selection from 4-vertex plurality graphs. The predicted vertex is shown in white.

We now proceed to prove each of the items in Theorem 6.1 as lemmas; the theorem then follows directly.

Lemma D.2. *If a randomized 1-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{1}{2}$ and $\alpha + \beta \leq 1$.*

Proof. Consider the graphs in Figure 3. It is easily verified that any symmetric impartial mechanism must assign probabilities as shown, and the symmetry assumption is without loss of generality due to Lemma D.1.

By the first graph, $\alpha \leq p_1$ and $\beta \leq p_1$. By the second graph, $\beta \leq p_2$. By the third graph, $p_1 + p_2 \leq 1$. Thus $2\beta \leq p_1 + p_2 \leq 1$ and $\alpha + \beta \leq p_1 + p_2 \leq 1$, as claimed. $\square$

Lemma D.3. *If a randomized 1-selection mechanism with predictions is impartial, α -consistent, and β -robust on plurality graphs, then $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.*

Proof. Consider the plurality graphs in Figure 4. It is easily verified that any symmetric impartial mechanism must assign probabilities as shown, and the symmetry assumption is without loss of generality due to Lemma D.1.

By the first graph,

$$p_1 + p_2 \leq 1.$$

By the second graph,

$$\begin{aligned} p_1 + p_3 + p_4 &\leq 1 \\ 2\alpha &\leq 2p_1 + p_3 + p_4, \quad \text{and} \\ 2\beta &\leq 2p_1 + p_3 + p_4. \end{aligned}$$

By the third graph,

$$\begin{aligned} p_2 + p_5 + p_6 &\leq 1 \quad \text{and} \\ 2\beta &\leq 2p_2 + p_5 + p_6. \end{aligned}$$

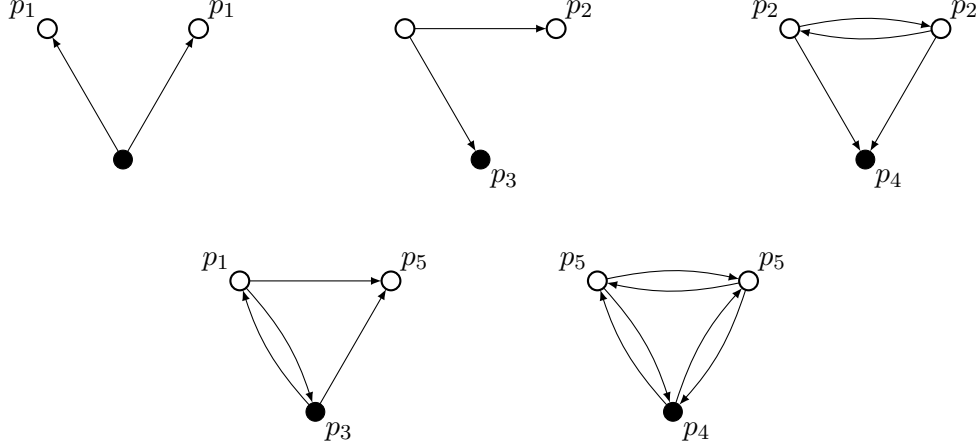


Figure 5: Impartial 2-selection from n -vertex graphs. Predicted vertices are shown in white. Only 3 vertices are shown; the remaining $n - 3$ vertices do not have any incident edges.

Then

$$4\beta \leq 2p_1 + 2p_2 + p_3 + p_4 + p_5 + p_6 \leq 3, \quad \text{and thus} \quad \beta \leq \frac{3}{4}.$$

Similarly

$$2\alpha + 2\beta \leq 2p_1 + 2p_2 + p_3 + p_4 + p_5 + p_6 \leq 3, \quad \text{and thus} \quad \alpha + \beta \leq \frac{3}{2}. \quad \square$$

Lemma D.4. *If a randomized 2-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{3}{4}$ and $\alpha + \beta \leq \frac{3}{2}$.*

Proof. Consider the graphs in Figure 5. It is easily verified that any symmetric impartial mechanism must assign probabilities as shown, and the symmetry assumption is without loss of generality due to Lemma D.1.

By the first graph,

$$2\alpha \leq 2p_1 \quad \text{and} \\ 2\beta \leq 2p_1.$$

By the second graph,

$$2\beta \leq p_2 + p_3.$$

By the third graph,

$$2p_2 + p_4 \leq 2.$$

By the fourth graph,

$$p_1 + p_3 + p_5 \leq 2.$$

Finally, by the fifth graph,

$$4\alpha \leq 2p_4 + 4p_5 \quad \text{and}$$

$$4\beta \leq 2p_4 + 4p_5.$$

Then

$$\begin{aligned} 8\beta &= 2\beta + 4\beta + 2\beta \\ &\leq 2p_1 + 2(p_2 + p_3) + (p_4 + 2p_5) \\ &= (2p_2 + p_4) + 2(p_1 + p_3 + p_5) \\ &\leq 2 + 4 = 6, \end{aligned}$$

and thus

$$\beta \leq \frac{3}{4}.$$

Similarly,

$$\begin{aligned} 4\alpha + 4\beta &= 2\alpha + 4\beta + 2\alpha \\ &\leq 2p_1 + 2(p_2 + p_3) + (p_4 + 2p_5) \\ &= (2p_2 + p_4) + 2(p_1 + p_3 + p_5) \\ &\leq 2 + 4 = 6, \end{aligned}$$

and thus

$$\alpha + \beta \leq \frac{3}{2}.$$

□

Lemma D.5. *If a randomized 3-selection mechanism with predictions is impartial, α -consistent, and β -robust, then $\beta \leq \frac{4}{5}$, $4\alpha + 3\beta \leq 6$, and $4\alpha + 21\beta \leq 20$.*

Proof. Consider the graphs in Figure 6. It is easily verified that any symmetric impartial mechanism must assign probabilities as shown, and the symmetry assumption is without loss of generality due to Lemma D.1.

By the first graph,

$$3\alpha \leq 3p_1.$$

By the second graph,

$$\begin{aligned} 3\alpha &\leq 3p_2 + p_3 \quad \text{and} \\ 3\beta &\leq 3p_2 + p_3. \end{aligned}$$

By the third graph,

$$3\beta \leq 2p_4 + 2p_5.$$

By the fourth graph

$$\begin{aligned} p_1 + p_5 + 2p_6 + p_7 &\leq 3 \quad \text{and} \\ 5\alpha &\leq p_1 + p_5 + 4p_6 + p_7. \end{aligned}$$

By the fifth graph,

$$2p_3 + 3p_8 \leq 3,$$

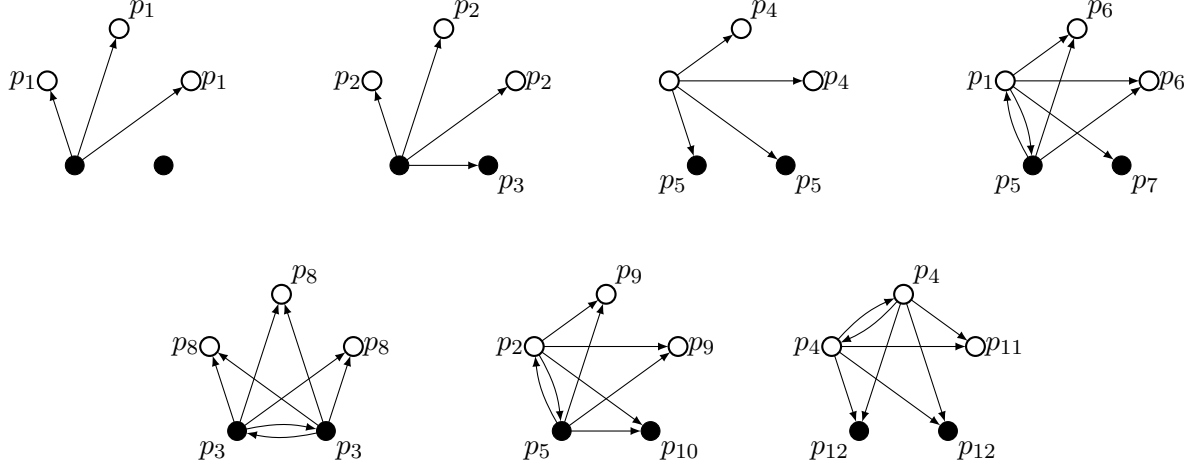


Figure 6: Impartial 3-selection from n -vertex graphs. Predicted vertices are shown in white. Only 5 vertices are shown; the remaining $n - 5$ vertices do not have any incident edges.

$$6\alpha \leq 2p_3 + 6p_8, \quad \text{and} \quad 6\beta \leq 2p_3 + 6p_8.$$

By the sixth graph,

$$p_2 + p_5 + 2p_9 + p_{10} \leq 3 \quad \text{and} \\ 6\beta \leq p_2 + p_5 + 4p_9 + 2p_{10}.$$

Finally, by the seventh graph,

$$2p_4 + p_{11} + 2p_{12} \leq 3 \quad \text{and} \\ 6\beta \leq 2p_4 + 2p_{11} + 4p_{12}.$$

Then

$$\begin{aligned} 75\beta &= 2(3\beta) + 3(3\beta) + 6\beta + 6(6\beta) + 3(6\beta) \\ &= 2(3p_2 + p_3) + 3(2p_4 + 2p_5) + (2p_3 + 6p_8) \\ &\quad + 6(p_2 + p_5 + 4p_9 + 2p_{10}) + 3(2p_4 + 2p_{11} + 4p_{12}) \\ &= 2(2p_3 + 3p_8) + 12(p_2 + p_5 + 2p_9 + p_{10}) + 6(2p_4 + p_{11} + 2p_{12}) \\ &\leq 2 \cdot 3 + 12 \cdot 3 + 6 \cdot 3 = 60, \end{aligned}$$

and thus

$$\beta \leq \frac{4}{5}.$$

Also

$$\begin{aligned} 36\alpha + 27\beta &= 2(3\alpha) + 6(5\alpha) + 3(3\beta) + 3(6\beta) \\ &\leq 2(3p_1) + 6(p_1 + p_5 + 4p_6 + p_7) + 3(2p_4 + 2p_5) + 3(2p_4 + 2p_{11} + 4p_{12}) \\ &\leq 12(p_1 + p_5 + 2p_6 + p_7) + 6(2p_4 + 2p_{11} + 4p_{12}) \\ &\leq 12 \cdot 3 + 6 \cdot 3 = 54, \end{aligned}$$

and thus

$$4\alpha + 3\beta \leq 6.$$

Finally

$$\begin{aligned} 12\alpha + 63\beta &= 2(3\alpha) + 6\alpha + 3(3\beta) + 6(6\beta) + 3(6\beta) \\ &\leq 2(3p_2 + p_3) + (2p_3 + 6p_8) + 3(2p_4 + 2p_5) \\ &\quad + 6(p_2 + p_5 + 4p_9 + 2p_{10}) + 3(2p_4 + 2p_{11} + 4p_{12}) \\ &= 2(2p_3 + 3p_8) + 12(p_2 + p_5 + 2p_9 + p_{10}) + 6(2p_4 + p_{11} + 2p_{12}) \\ &\leq 2 \cdot 3 + 12 \cdot 3 + 6 \cdot 3 = 60, \end{aligned}$$

and thus

$$4\alpha + 21\beta \leq 20. \quad \square$$

References

- [1] S. Aaronson. My AI safety lecture for UT effective altruism, 2022. available under <https://scottaaronson.blog/?m=202211>; last accessed May 16, 2025.
- [2] P. Agrawal, E. Balkanski, V. Gkatzelis, T. Ou, and X. Tan. Learning-augmented mechanism design: Leveraging predictions for facility location. *Mathematics of Operations Research*, 49(4): 2626–2651, 2024.
- [3] N. Alon, F. Fischer, A. Procaccia, and M. Tennenholtz. Sum of us: Strategyproof selection from the selectors. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 101–110, 2011.
- [4] E. Balkanski, V. Gkatzelis, and X. Tan. Strategyproof scheduling with predictions. In *Proceedings of the 14th Conference on Innovations in Theoretical Computer Science*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.
- [5] E. Balkanski, V. Gkatzelis, and G. Shahkarami. Randomized strategic facility location with predictions. In A. Globersons, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. M. Tomczak, and C. Zhang, editors, *Proceedings of the 38th Annual Conference on Neural Information Processing Systems*, 2024.
- [6] B. Berger, M. Feldman, V. Gkatzelis, and X. Tan. Learning-augmented metric distortion via (p, q) -veto core. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 984–984, 2024.
- [7] A. Bjelde, F. Fischer, and M. Klimm. Impartial selection and the power of up to two choices. *ACM Transactions on Economics and Computation*, 5(4):1–20, 2017.
- [8] N. Bousquet, S. Norin, and A. Vetta. A near-optimal mechanism for impartial selection. In *Proceedings of the 10th International Conference on Web and Internet Economics*, pages 133–146. Springer, 2014.
- [9] I. Caragiannis, G. Christodoulou, and N. Protopapas. Impartial selection with additive approximation guarantees. In *Proceedings of the 12th International Symposium on Algorithmic Game Theory*, pages 269–283. Springer, 2019.

- [10] I. Caragiannis, G. Christodoulou, and N. Protopapas. Impartial selection with prior information. In Y. Ding, J. Tang, J. F. Sequeda, L. Aroyo, C. Castillo, and G. Houben, editors, *Proceedings of the ACM Web Conference 2023*, pages 3614–3624, 2023.
- [11] J. Cembrano, F. Fischer, and M. Klimm. Improved bounds for single-nomination impartial selection. In *Proceedings of the 24th ACM Conference on Economics and Computation*, page 449, 2023.
- [12] J. Cembrano, F. Fischer, D. Hannon, and M. Klimm. Impartial selection with additive guarantees via iterated deletion. *Games and Economic Behavior*, 144:203–224, 2024.
- [13] J. Cembrano, S. M. Griesbach, and M. J. Stahlberg. Deterministic impartial selection with weights. *ACM Transactions on Economics and Computation*, 12(3):10:1–10:22, 2024.
- [14] L. Chen, J. Q. Davis, B. Hanin, P. Bailis, I. Stoica, M. A. Zaharia, and J. Y. Zou. Are more LLM calls all you need? towards the scaling properties of compound AI systems. In A. Globersons, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. M. Tomczak, and C. Zhang, editors, *Proceedings of the 38th Annual Conference on Neural Information Processing Systems*, 2024.
- [15] G. de Clippel, H. Moulin, and N. Tideman. Impartial division of a dollar. *Journal of Economic Theory*, 139(1):176–191, 2008.
- [16] J. Fang, Q. Fang, W. Liu, and Q. Nong. Mechanism design with predictions for facility location games with candidate locations. In X. Chen and B. Li, editors, *Proceedings of the 18th Annual Conference on Theory and Applications of Models of Computation*, volume 14637 of *Lecture Notes in Computer Science*, pages 38–49, 2024.
- [17] F. Fischer and M. Klimm. Optimal impartial selection. *SIAM Journal on Computing*, 44(5):1263–1285, 2015.
- [18] J. G. Heinberg. History of the majority principle. *The American Political Science Review*, 20(1):52–68, 1926.
- [19] R. Holzman and H. Moulin. Impartial nominations for a prize. *Econometrica*, 81(1):173–196, 2013.
- [20] G. Irving, P. Christiano, and D. Amodei. AI safety via debate. arXiv Preprint; available under <https://arxiv.org/abs/1805.00899>, 2018.
- [21] G. Istrate and C. Bonchis. Mechanism design with predictions for obnoxious facility location. arXiv preprint, available under <https://arxiv.org/abs/2212.09521>, 2022.
- [22] A. Lindermayr and N. Megow. Algorithms with predictions, 2025. Available under <https://algorithms-with-predictions.github.io>; last accessed February 10, 2025.
- [23] A. Mackenzie. Symmetry and impartial lotteries. *Games and Economic Behavior*, 94:15–28, 2015.
- [24] Moebius314. Multiple AIs in boxes, evaluating each other’s alignment, 2022. Available under <https://www.lesswrong.com/posts/biskschef2zSNgKkz/multiple-ais-in-boxes-evaluating-each-other-s-alignment>; last accessed May 15, 2025.

- [25] M. Purohit, Z. Svitkina, and R. Kumar. Improving online algorithms via ML predictions. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Proceedings of the 31th Annual Conference on Neural Information Processing Systems*, pages 9684–9693, 2018.
- [26] C. Xu and P. Lu. Mechanism design with predictions. In L. D. Raedt, editor, *Proceedings of the 31st International Joint Conference on Artificial Intelligence*, pages 571–577, 2022.