Impartial Peer Review

David Kurokawa
Carnegie Mellon
dkurokaw@cs.cmu.edu

Omer Lev
Hebrew University
omerl@cs.huji.ac.il

Jamie Morgenstern
Carnegie Mellon
jamiemmt@cs.cmu.edu

Ariel D. Procaccia
Carnegie Mellon
arielpro@cs.cmu.edu

Abstract

Motivated by a radically new peer review system that the National Science Foundation recently experimented with, we study peer review systems in which proposals are reviewed by PIs who have submitted proposals themselves. An (m, k)-selection mechanism asks each PI to review m proposals, and uses these reviews to select (at most) k proposals. We are interested in *impartial* mechanisms, which guarantee that the ratings given by a PI to others' proposals do not affect the likelihood of the PI's own proposal being selected. We design an impartial mechanism that selects a k-subset of proposals that is nearly as highly rated as the one selected by the non-impartial (abstract version of) the NSF pilot mechanism, even when the latter mechanism has the "unfair" advantage of eliciting honest reviews.

1 Introduction

The Sensors and Sensing Systems (SSS) program of the National Science Foundation (NSF) recently experimented with a drastically different peer review method. Traditionally, grant proposals submitted to a specific program are evaluated by a panel of reviewers. Potential conflicts of interest play a crucial role in composing the panel; most importantly, principal investigators (PIs) whose proposals are being evaluated by the panel cannot serve on the panel. In stark contrast, the new peer review method — originally designed by Merrifield and Saari [2009] for the review of proposals for telescope time — requires the PIs themselves to review each other's proposals! A "dear colleague letter" [Hazelrigg, 2013] explains the potential merits of the new process:

> *"This pilot is an attempt to find an alternative proposal review process that can preserve the ability of investigators to submit multiple proposals at more than one opportunity per year while encouraging high quality and collaborative research, placing the burden of proposal review onto the reviewer community in proportion to the burden each individual imposes on the system, simplifying the internal NSF review process, ameliorating concerns of conflict-of-interest, maintaining high quality in the review process, and substantially reducing proposal review costs."*

Under the Saari-Merrifield mechanism, each PI must review m proposals submitted by other PIs; in the NSF pilot, $m = 7$. The PI then ranks the m proposals according to their quality. These reviews are aggregated using the Borda count voting rule, so each PI awards $m-i$ points to the proposal she ranks in position i. A proposal's overall rating is the average over the points awarded by the m PIs who reviewed it. Additionally, a PI's own proposal receives a small bonus based on the similarity between the PI's submitted ranking and the aggregate ranking of the proposals she reviewed; this is meant to encourage PIs to make an effort to produce accurate reviews.

The NSF pilot sparked a lively debate amongst mechanism design and social choice researchers in the blogosphere [Procaccia, 2013; Vohra, 2013; Mitzenmacher, 2013]. While most researchers seem to agree that the NSF should be commended for trying out an ambitious peer review method, serious concerns have been raised regarding the pilot mechanism itself. Perhaps most strikingly, while the NSF announcement [Hazelrigg, 2013] states that the "theoretical basis for the proposed review process lies in an area of mathematics referred to as mechanism design", the pilot mechanism provides no theoretical guarantees. In particular, the mechanism is susceptible to strategic manipulation: PIs will often be able to advance their own proposals by giving low scores to competitive proposals (even though they may forfeit some of the small bonus for similarity to others' reviews). Furthermore, while most researchers who sit on NSF panels are well-respected, the pilot mechanism cannot control the quality (or morality) of PIs who submit proposals (and review proposals)— leaving open the very real possibility of game-theoretic mayhem.

In this paper, we alleviate these concerns by proposing a peer review mechanism which is not susceptible to such manipulations. Each PI who submits a proposal or paper will review some other PIs' proposals or papers. Our mechanism is *impartial*: reviewers will not be able to affect the chances of their own proposals being selected. Our research challenge is therefore to *design provably impartial peer review mechanisms that provide formal quality guarantees.*

We believe that solutions to this problem truly matter. The NSF plays a huge role in enabling scientific research in the United States, and its consideration of alternative peer review methods may transform how scientific funding is allocated in the US. The need to build sound foundations for these methods therefore provides a unique opportunity for computational game theory research, and AI research more broadly.

1.1 Our Approach

In our setting there are n PIs, each associated with a proposal. Each PI i has a hypothetical (honest) evaluation of the quality of the proposal j, which is the rating i would give j if she were asked to review that proposal (and could not affect her own chances of selection). The (honest) *score* of a proposal is the average (honest) rating given to it by other PIs. As NSF program directors, if our budget is sufficient to fund k proposals, we would ideally want to select a set of k proposals with maximum honest score.[1] There are two obstacles we must overcome: we cannot possibly ask each PI to review all other proposals, and the reviews may be dishonest.

To address the first problem, we consider only mechanisms which request m reviews per PI (much like the NSF pilot). We define an (m, k)-*selection mechanism* as follows. First, the mechanism asks each PI to review m proposals, in a way that each proposal is reviewed by exactly m PIs; for every such pair (i, j), PI i's evaluation for proposal j is revealed. Based on these elicited reviews, the mechanism selects k vertices. The most natural (m, k)-selection mechanism is an abstract version of the NSF pilot mechanism, which we fondly refer to as the VANILLA mechanism; it chooses m reviews per PI uniformly at random (subject to the constraint that each proposal is reviewed by m PIs), and then selects the k vertices with highest average rating, based only on the sampled reviews.

Returning to the second problem — dishonest reviewing — we will consider only mechanisms where reviewers cannot affect their chances of being selected by misreporting their reviews. A selection mechanism is *impartial* if the probability of proposal i being selected is independent of the ratings given by PI i. The motivation for our work stems from the observation that the VANILLA mechanism is not impartial: we seek mechanisms that are.

How should we evaluate the impartial mechanisms we design? Without any assumptions, competing with an omniscient mechanism that maximizes underlying scores is clearly impossible.[2] We therefore use the VANILLA mechanism as our performance benchmark. Competing with VANILLA is nontrivial, because we give it the "unfair" advantage of assuming that reviews are honest, even though it is not impartial. Specifically, we say that an impartial mechanism α-*approximates* VANILLA if, in the worst case over reviews, the ratio between the expected score (based on the largely unseen set of all possible reviews) of the set of proposals selected by the impartial mechanism, and the expected score of the set of proposals selected by VANILLA, is at least α.

The choice of VANILLA as a benchmark has two main advantages. First, since the VANILLA Mechanism is an abstraction of the NSF pilot mechanism, our choice of benchmark allows us to quantify how much the NSF must sacrifice to achieve impartiality — and our results show that this sacrifice is negligible. Moreover, innovations that are closest to the current accepted practice are the most likely to be adopted.

Second, modulo its lack of impartiality, VANILLA is intuitively the "right" mechanism: it selects those nodes with the highest sampled scores. Furthermore, in an average-case model where each proposal has an intrinsic quality, and reviews are drawn from a well-behaved distribution whose expectation is the true quality of a proposal, VANILLA will pinpoint the best proposals given a sufficiently large m. Even when we assume reviews are worst-case, we can obtain an excellent approximation of VANILLA via an impartial mechanism, and that guarantee immediately extends to the average case model.

1.2 Our Results

In §3 we present an impartial (m, k)-selection mechanism, CREDIBLE SUBSET, which (usually) selects k proposals at random from a slightly larger pool (of size $k + m$) of eligible proposals. We prove that CREDIBLE SUBSET gives an approximation ratio of $\frac{k}{k+m}$ to VANILLA. We think of m, the number of reviews per PI, as being a small constant, and we would like to think of k, the number of proposals to be selected, as significantly larger. In particular, when $m = o(k)$, the approximation ratio goes to 1 as k goes to infinity (in an ideal world, where growth in funding outpaces growth in the quantity of work for reviewers, see §5).

In §4, we show that CREDIBLE SUBSET is the *optimal* impartial mechanism, in the sense that its approximation ratio of $\frac{k}{k+m}$ is asymptotically tight (when $k = m^2$ is a constant and the number of PIs n grows).

1.3 Related Work

Our paper is closely related to the work of Alon et al. [2011]. In parallel with Holzman and Moulin [2013], Alon et al. introduced the notion of impartial selection mechanisms (using the term "strategyproofness" for impartiality). Their model can be interpreted as a special case of our model, where $m = n - 1$ (i.e., each PI reviews all other proposals) and all the ratings are in $\{0, 1\}$. The main result of Alon et al. is the design of an impartial mechanism that approximates *the score of the optimal subset of k vertices* to a factor that goes to 1 as k grows. When $m = n - 1$ and all ratings are in $\{0, 1\}$, this is equivalent to approximating Vanilla: Vanilla can see all ratings and will select the optimal subset. But when $m \ll n - 1$ we cannot reason about scores directly, as Alon et al. do. In fact, in this regime, which is typical for a peer review setting, our results are incomparable to theirs: our mechanisms use far less information, but the performance of these mechanisms is (necessarily) measured against a weaker benchmark.

[1] We distill the strategic aspects of the NSF reviewing setting and abstract away some other practical aspects, such as the fact that PIs may submit multiple proposals to the same program. However, our model and results easily extend.

[2] Indeed, even VANILLA with truthful reviews will be unable to do so!

Other papers on impartial mechanisms include the ones by de Clippel et al. [2008], Holzman and Moulin [2013], Fischer and Klimm [2014], Berga and Gjorgjiev [2014], Tamura and Ohseto [2014], and Mackenzie [2014].

Merrifield and Saari [2009] are not the first researchers to suggest improvements to the peer review process, although most other papers focus on conference reviewing [Nierstrasz, 2000; Haenni, 2008; Douceur, 2009; Roos *et al.*, 2011]. For example, in a AAAI'11 paper, Roos et al. [2011] propose a method for calibrating the ratings of potentially biased reviewers via a maximum likelihood estimation (MLE) approach.

2 The Model

Let $N = \{1, 2, \ldots, n\}$ be the set of proposals and also the set of strategizing reviewers. Each reviewer i has an estimate of the quality of every other proposal $j \neq i$ — the score i would give j if i honestly reviewed j. We represent this setting as a weighted, complete, directed graph $G = (N, E, w_G)$ where $E = \{(i, j) \mid i, j \in N, i \neq j\}$, and $w_G(i, j) \in \mathbb{R}^+$ is the quality of j according to i's evaluation. We call G the *underlying graph*.

Let m be the number of proposals that each PI can review, which must equal to the number of reviews each proposal receives (we assume each PI submits one proposal). In our model, m is the number of outgoing edges from each vertex and the number of incoming edges to each vertex. Slightly abusing terminology, we say that a directed graph is m*-regular* if it satisfies these properties.

A peer review process is governed by an (m, k)*-selection mechanism*, which works in two stages:

1. The mechanism selects (possibly randomly) a directed m-regular graph $G^m = (N, E(G^m))$, called the *sampled graph*. We assume this graph is drawn prior to the next step: that the sampling is done all at once independent of the edge weights.
2. Given the underlying graph G, the weight $w_G(i, j)$ is revealed for each edge $(i, j) \in E(G^m)$. The mechanism then maps these elicited ratings to a subset of selected vertices of size at most k.

Step 1 corresponds to the mechanism assigning m proposals to each PI. Based on the reviews $w_G(i, j)$ for $(i, j) \in E(G^m)$, in Step 2, the mechanism selects a subset of at most k proposals that will receive funding.

Let us reinterpret the NSF pilot mechanism [Hazelrigg, 2013] in this framework, abstracting away details such as the use of Borda count and the bonus component for accurate reviews. To this end, let $\mathcal{G}^m$ denote the uniform distribution over m-regular graphs. Given a weighted m-regular graph G^m, let

$$\text{top}_k(G^m) \in \operatorname*{arg\,max}_{Y \subseteq N:\ |Y|=k} \sum_{i \in Y} \sum_{j:(j,i) \in E(G^m)} w_G(j, i),$$

breaking ties lexicographically (i.e., the k nodes with the largest sum of incoming edge weights in the graph). Now, the *Vanilla* mechanism, denoted $\mathcal{M}^v$, is defined as follows:

VANILLA (G, m, k)

1. Draw $G^m \sim \mathcal{G}^m$.
2. Return $\text{top}_k(G^m)$.

Intuitively, the mechanism assigns proposals to PIs for review based on the graph G^m, and then returns the k highest-rated reviews based on the sampled reviews (for convenience we look at the sum of ratings, which is equivalent to the average).

For a mechanism $\mathcal{M}$ and an underlying graph G, let $\mathcal{M}(G)$ be a random variable, which takes the value $X \subseteq N$ with the same probability that $\mathcal{M}$ outputs X when the underlying graph is G. Then we can use $\mathbb{P}[i \in \mathcal{M}(G)]$ to denote the probability that $\mathcal{M}$ selects $i \in N$ when the underlying graph is G. We say that $\mathcal{M}$ is *impartial* if for any $i \in N$ and any two underlying graphs G and G' that differ only in the weights on the outgoing edges of i, $\mathbb{P}[i \in \mathcal{M}(G)] = \mathbb{P}[i \in \mathcal{M}(G')]$.

Unfortunately, VANILLA is clearly not impartial. To see this, let $k = 1$, $m = 1$, and define the weights of G and G' as follows:

$$w_G(i, j) = \begin{cases} n+1 & i = 1 \\ 1 & j = 1, i \neq 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$w_{G'}(i, j) = \begin{cases} 0 & i = 1 \\ 1 & j = 1, i \neq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Then $\mathbb{P}[1 \in \mathcal{M}^v(G)] = 0$, whereas $\mathbb{P}[1 \in \mathcal{M}^v(G')] = 1$ (using lexicographic tie-breaking, 1 would be selected even if only 0-weight edges are sampled).

The purpose of this paper is to design (m, k)-selection mechanisms that are simultaneously impartial (unlike VANILLA), yet similarly practical in terms of the number of reviews per proposal and similar in the quality of the output. We measure the quality of a mechanism by the expected score of the vertices it selects. Formally, let $\text{sc}(i, G) = \sum_{(j,i) \in E} w_G(j, i)$ be the *score of vertex i in G*, and let $\text{sc}(X, G) = \sum_{i \in X} \text{sc}(i, G)$ be the score of a set of vertices $X \subseteq N$ in G. We can now define

$$\text{sc}(\mathcal{M}, G) = \mathbb{E}_{X \sim \mathcal{M}(G)}[\text{sc}(X, G)].$$

This is our optimization objective.

Note that for some underlying graphs G, VANILLA itself may do poorly in terms of $\text{sc}(\mathcal{M}, G)$. As an extreme example, let $k = 1$, $m = 1$, and define the weights of the underlying graph G as follows:

$$w_G(i, j) = \begin{cases} 1000 & i = 1 \wedge j = 2 \\ 1/n & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

It is very likely that the edge $(1, 2)$ will not be sampled by VANILLA, and therefore the mechanism will likely select vertex 1, the only one with non-zero score. However, $\text{sc}(1, G) = \frac{n-1}{n} < 1$, whereas $\text{sc}(2, G) = 1000$. This is not a shortcoming of VANILLA specifically — it is clear that such examples can be constructed for any (m, k)-selection mechanism when m is much smaller than n.

Nevertheless, we can use VANILLA as a benchmark. We wish to design impartial mechanisms whose quality guarantee is quite close to that of VANILLA pointwise (assuming all reviews given to VANILLA were truthful). We say that an (m,k)-selection mechanism $\mathcal{M}$ α-approximates VANILLA, for $\alpha = \alpha(m,n,k) \leq 1$, if for *every* underlying graph G,

$$\frac{\text{sc}(\mathcal{M},G)}{\text{sc}(\mathcal{M}^v,G)} \geq \alpha.$$

3 The Credible Subset Mechanism

In this section we present and analyze an (m,k)-selection mechanism, the *Credible Subset* mechanism. The mechanism relies on two ideas:

1. Every vertex that has the potential to be among the top k by changing its outgoing edges must have a chance to be selected. Such vertices are called *credible*. There are not too many of them, and they include the actual top k.
2. A credible vertex can potentially affect the number of credible vertices (by giving a low score to another credible vertex), and therefore the probability of selecting a credible vertex must be independent of the number of credible vertices.

The Credible Subset mechanism, denoted $\mathcal{M}^{cs}$, formally works as follows.

CREDIBLE SUBSET(G,m,k)

1. Draw $G^m \sim \mathcal{G}^m$.
2. $P \leftarrow \{i \notin \text{top}_k(G^m) \mid \text{if } i \text{ reported } \forall j : w(i,j) = 0,$ $i \text{ would be in top}_k(G^m)\}$
3. $S \leftarrow \text{top}_k(G^m) \cup P$.
4. With probability $\frac{|S|}{k+m}$ return a random k-subset of S, and with probability $1 - \frac{|S|}{k+m}$ return $\emptyset$.

Let us verify that CREDIBLE SUBSET is well-defined, in the sense that $\frac{|S|}{k+m} \leq 1$. Recall that for the purpose of computing $\text{top}_k(G^m)$, ties are broken lexicographically. This implies that, for a given $i \notin \text{top}_k(G^m)$, the only way for i to enter P would be to reduce weights on outgoing edges to some of the top k vertices. It can reduce its outgoing weights to at most m vertices; thus, any vertex that makes it into the top k after reducing weights must have been in the top $k+m$ to begin with, where $k+m$ is defined with respect to the tie-breaking order. We conclude that there cannot be more than m vertices that can enter $\text{top}_k(G^m)$ by reducing their outgoing weights. That is, $|P| \leq m$, and hence

$$|S| = |\text{top}_k(G^m)| + |P| \leq k + m.$$

Theorem 1. CREDIBLE SUBSET *is an impartial* (m,k)*-selection mechanism which* $\frac{k}{k+m}$*-approximates* VANILLA.

Proof. We first establish impartiality. The mechanism is clearly impartial with respect to vertices $i \in N \setminus S$: for any G and G' that differ only in the weights of outgoing edges from i,

$$\mathbb{P}[i \in \mathcal{M}^{cs}(G) \mid i \notin S] = 0 = \mathbb{P}[i \in \mathcal{M}^{cs}(G') \mid i \notin S].$$

The mechanism is also impartial for $i \in S$. Indeed, some k-subset of S is selected with probability $\frac{|S|}{k+m}$. Given that some k-subset of S is selected, the probability that $i \in S$ is selected is $\frac{k}{|S|}$. Thus,

$$\mathbb{P}[i \in \mathcal{M}^{cs}(G)] = \frac{\mid S \mid}{k+m} \cdot \frac{k}{\mid S \mid} = \frac{k}{k+m}. \qquad (1)$$

In other words, for two graphs G and G' as above,

$$\mathbb{P}[i \in \mathcal{M}^{cs}(G) \mid i \in S] = \frac{k}{k+m} = \mathbb{P}[i \in \mathcal{M}^{cs}(G') \mid i \in S],$$

and we conclude that for all $i \in N$,

$$\mathbb{P}[i \in \mathcal{M}^{cs}(G)] = \mathbb{P}[i \in \mathcal{M}^{cs}(G')].$$

Next we establish the approximation guarantees of CREDIBLE SUBSET. Notice that CREDIBLE SUBSET samples from $\mathcal{G}^m$, just as VANILLA does. In addition, for a fixed sampled graph $G^m \sim \mathcal{G}^m$, VANILLA outputs $\text{top}_k(G^m)$. Thus, for every underlying graph G, the approximation ratio given by CREDIBLE SUBSET is

$$\begin{aligned}
&\frac{\text{sc}(\mathcal{M}^{cs},G)}{\text{sc}(\mathcal{M}^v,G)} \\
&= \frac{\sum_{G^m} \mathbb{P}[G^m] \cdot \sum_{i \in N} \mathbb{P}[i \in \mathcal{M}^{cs}(G) |\, G^m] \cdot \text{sc}(i,G)}{\sum_{G^m} \mathbb{P}[G^m] \cdot \sum_{i \in N} \mathbb{P}[i \in \mathcal{M}^{v}(G) |\, G^m] \cdot \text{sc}(i,G)} \\
&\geq \frac{\sum_{G^m} \mathbb{P}[G^m] \cdot \sum_{i \in N} \mathbb{I}[i \in \text{top}_k(G^m)] \cdot \frac{k}{k+m} \cdot \text{sc}(i,G)}{\sum_{G^m} \mathbb{P}[G^m] \cdot \sum_{i \in N} \mathbb{I}[i \in \text{top}_k(G^m)] \cdot \text{sc}(i,G)} \\
&= \frac{k}{k+m},
\end{aligned}$$

where the second transition follows from Equation (1), and $\mathbb{I}[E]$ is an indicator variable that takes that value 1 if the event E is true and 0 if E is false. □

We remark that the mechanism may return subsets of size smaller than k — empty subsets, in fact! Choosing empty subsets is not necessary: the same approximation guarantee can be achieved by defining a finer distribution over subsets preserving that each vertex in S is selected with probability $\frac{k}{k+m}$ (this is the insight that drives the proof of Theorem 1). We focus on the simpler formulation of the mechanism for ease of exposition, and further discuss this point in §5.

4 Impossibility Results

In §3 we proved that CREDIBLE SUBSET approximates VANILLA to a factor of $\frac{k}{k+m}$. When $m = o(k)$, this is $1 - o(1)$. But when both k and m are constants, this ratio is bounded away from 1 even when $n \to \infty$. It is natural to wonder, though, if an impartial (m,k)-selection mechanism can approximate VANILLA to a factor of $1 - o(1)$ when k and m are constants and n grows. After all, in this regime the performance of VANILLA will be very poor in the worst case (as G^m gives an extremely incomplete picture of G), so VANILLA becomes easier to approximate. We answer this question in the negative: we show below that the $\frac{k}{k+m}$ ratio is essentially the best possible for impartial (m,k)-selection mechanisms.

Let us start with an informal discussion of a simple upper bound of $\frac{k}{k+1}$ that only assumes that $k \le m$ (that is, it gives a constant upper bound for $k = O(1)$ even if m grows). Let G be an underlying graph such that

$$w_G(i,j) = \begin{cases} \epsilon & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

VANILLA will certainly select vertex 1. Consider an impartial (m,k)-selection mechanism $\mathcal{M}$, and let $\mathbb{P}[1 \in \mathcal{M}(G)] = p$. Since 1 is the only vertex with nonzero score, the approximation ratio of $\mathcal{M}$ on G is p.

Next, consider the underlying graph G' with weights:

$$w_{G'}(i,j) = \begin{cases} \epsilon & j = 1 \\ 1 & i = 1 \\ 0 & \text{otherwise} \end{cases}$$

For $\epsilon \ll \frac{1}{n-1}$, VANILLA will certainly select k vertices with score 1, so $\text{sc}(\mathcal{M}^v, G') = k$. By impartiality, $\mathbb{P}[1 \in \mathcal{M}(G')] = p$, hence

$$\text{sc}(\mathcal{M}, G') \le (1-p)k + p(k-1+(n-1)\epsilon).$$

Since ϵ is arbitrarily small, the approximation ratio is upper-bounded in the limit by

$$\alpha = \min\left\{p, (1-p) + \frac{p(k-1)}{k}\right\}.$$

Maximizing α over all $p \in [0,1]$ gives $p = \frac{k}{k+1}$ as an upper bound on the approximation ratio.

Let us now turn to our more intricate upper bound.

Theorem 2. *Let $c \in (0, \frac{1}{4})$, $k = m^2$, and $m \le n^c$. Then any impartial mechanism at best $\left(\frac{k}{k+m} + \epsilon(n)\right)$-approximates* VANILLA, *for $\epsilon(n) = o(1)$.*

We require the following probabilistic lemma, whose easy proof is omitted due to lack of space.

Lemma 1. *Let $c \in (0, 1/4)$. Suppose n^c distinct elements are drawn from a universe of size n uniformly at random and independently. Suppose this experiment is repeated n^c times, and let the selected set in round t be denoted N_t. Then, with high probability, $N_t \cap N_{t'} = \emptyset$, for all $t \ne t'$.*

Proof of Theorem 2. Let $\mathcal{M}$ be an impartial mechanism. Consider a set $X \subset N$ of size m. We will build up a matching μ between X and $N \setminus X$, such that the probability $\mathcal{M}$ samples the edge $(\mu(i), i)$ is small (roughly m/n) for all i. This will imply that $\mathcal{M}$ will have to select i with similar probability on two graphs which differ only in the weight of the edge $(\mu(i), i)$.

We will now select vertices and relabel them, adding them to X as we progress. Select an arbitrary vertex and label it 1. Let $\mu(1) = \text{argmin}_j \mathbb{P}[\mathcal{M} \text{ samples } (j,1)]$ (the vertex with the smallest probability of $(j,1)$ being sampled by $\mathcal{M}$). Let $q_1 = \mathbb{P}[\mathcal{M} \text{ samples } (\mu(1),1)]$; note that $q_1 \le \frac{m}{n-1}$ by a simple averaging argument. Then, for each $i \in [2, \dots, m]$, select another arbitrary vertex and label it i such that $i \notin \{1, \dots, i-1\} \cup \{\mu(1), \dots, \mu(i-1)\}$, and let

$$\mu(i) = \text{argmin}_{j \notin \{1,\dots,i\} \cup \{\mu(1),\dots,\mu(i-1)\}} \mathbb{P}[\mathcal{M} \text{ samples } (j,i)],$$

be the vertex such that $(\mu(i), i)$ has the smallest probability of being sampled by $\mathcal{M}$ which is not already part of the matching, and

$$q_i = \mathbb{P}[\mathcal{M} \text{ samples } (\mu(i), i)]$$

be that probability. Note that $q_i \le \frac{m}{n-2(i-1)-1}$, else the expected number of edges incident to i would be larger than m.

Now, we construct an underlying graph G that is defined using the following weights:

$$w_G(i,j) = \begin{cases} 1 & i \in X, j \notin X \\ \epsilon \ll \frac{1}{m} & i \notin X, j \in X \\ 0 & \text{otherwise} \end{cases}$$

For each $i \in X$, let the graph G'_i on n vertices be as follows:

$$w_{G'_i}(j,j') = \begin{cases} M \gg 1 & j = \mu(i), j' = i \\ 1 & j \in X, j \ne i, j' \notin X \\ \epsilon \ll \frac{1}{m} & j \notin X, j' \in X, (j,j') \ne (\mu(i), i) \\ 0 & \text{otherwise} \end{cases}$$

Notice that G'_i differs from G in two ways: it has one high-weight edge to i, and the outgoing edges from i have weight 0 rather than weight 1.

We begin by showing that

$$\text{sc}(\mathcal{M}^v, G) \ge |X| k (1 - o(1)). \tag{2}$$

To prove (2), denote the set of vertices adjacent to a set Y in the sampled graph G^m by $\mathcal{N}_{G^m}(Y)$. Notice that the vertices $j \in \mathcal{N}_{G^m}(X)$ have strictly higher sampled ratings than all other vertices in G^m. Moreover, $|\mathcal{N}_{G^m}(X)| \le k$, so VANILLA will select all $j \in \mathcal{N}_{G^m}(X)$. Thus,

$$\begin{aligned} \text{sc}(\mathcal{M}^v, G) &= \sum_j \mathbb{P}[j \in \text{top}_k(G^m)] \text{sc}(j, G) \\ &\ge \sum_{j \notin X} \mathbb{P}[j \in \text{top}_k(G^m)] \text{sc}(j, G) \\ &\ge \sum_{j \notin X} \mathbb{P}[j \in \mathcal{N}_{G^m}(X)] \text{sc}(j, G) \\ &\ge |X| \sum_{j \notin X} \mathbb{P}[j \in \mathcal{N}_{G^m}(X)] = |X| \cdot \mathbb{E}\left[|\mathcal{N}_{G^m}(X)|\right] \\ &\ge |X|(k(1 - o(1))), \end{aligned}$$

where the final transition follows from Lemma 1 and the assumption that $c \in (0, \frac{1}{4})$ and $m \le n^c$.

Next, we claim that

$$\text{sc}(\mathcal{M}^v, G'_i) \ge M. \tag{3}$$

Let G^m denote the sampled graph. Then, notice that there is a trivial upper bound on the size of $|\mathcal{N}_{G^m}(X \setminus \{i\})|$:

$$|\mathcal{N}_{G^m}(X \setminus \{i\})| \le m(|X| - 1) = k - m. \tag{4}$$

Therefore,

$$\begin{aligned} \text{sc}(\mathcal{M}^v, G'_i) &= \sum_j \mathbb{P}[j \in \text{top}_k(G^m)] \text{sc}(j, G'_i) \\ &\ge M \cdot \mathbb{P}[i \in \text{top}_k(G^m)] \ge M \cdot \mathbb{P}[X \subset \text{top}_k(G^m)] \\ &= M \cdot \mathbb{P}[|\mathcal{N}_{G^m}(X \setminus \{i\})| \le k - m] = M. \end{aligned}$$

The fourth transition follows from the observation that the only vertices with nonzero sampled ratings are in $X \cup \mathcal{N}_{G^m}(X \setminus \{i\})$ (which implies VANILLA will select all of them, if there are not more than k), and the final equality comes from from (4).

Now, we revisit the impartial mechanism $\mathcal{M}$. We show the probability i is selected by $\mathcal{M}$ in G cannot be too different from the probability i is selected by $\mathcal{M}$ in G'_i. Let $p_i = \mathbb{P}[i \in \mathcal{M}(G)]$. Consider the "intermediate" graph G''_i such that

$$w_{G''_i}(j,j') = \begin{cases} M \gg 1 & j = \mu(i), j' = i \\ 1 & j \in X, j' \notin X \\ \epsilon \ll \frac{1}{m} & j \notin X, j' \in X, (j,j') \neq (\mu(i), i) \\ 0 & \text{otherwise} \end{cases}$$

That is, G''_i is the graph G with the added heavy-weight edge to i, or the graph G'_i with the outgoing edges from i set to 1.

Let G^m be the graph sampled by $\mathcal{M}$. If $(\mu(i), i) \notin E(G^m)$, $\mathcal{M}$ cannot distinguish between G and G''_i, and thus must select i with the same probability in those cases. Then, by impartiality, $\mathcal{M}$ must select i with equal (unconditional) probability in G'_i, G''_i, since they differ only in the outgoing edges from i.

In more detail, let us denote $p_i = \mathbb{P}[i \in \mathcal{M}(G)]$. We have

$$\begin{aligned} p_i =& \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \in E(G^m)]\mathbb{P}[(\mu(i), i) \in E(G^m)] + \\ & \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \notin E(G^m)]\mathbb{P}[(\mu(i), i) \notin E(G^m)] \\ =& \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \in E(G^m)]\mathbb{P}[(\mu(i), i) \in E(G^m)] + \\ & \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \notin E(G^m)](1 - \mathbb{P}[(\mu(i), i) \in E(G^m)]). \end{aligned}$$

Then, we explicitly write p_i in terms of q_i:

$$\begin{aligned} p_i =& \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \in E(G^m)]q_i \\ & + \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \notin E(G^m)]\,(1 - q_i)\,. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{P}[i \in \mathcal{M}(G''_i) \mid (\mu(i), i) \notin E(G^m)] \\ &= \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \notin E(G^m)] \\ &= \frac{p_i - q_i \mathbb{P}[i \in \mathcal{M}(G) \mid (\mu(i), i) \in E(G^m)]}{(1 - q_i)} \leq \frac{p_i}{(1 - q_i)}. \end{aligned}$$

We can use this inequality to derive an upper bound on the probability that $i \in \mathcal{M}(G''_i)$:

$$\begin{aligned} \mathbb{P}[i \in \mathcal{M}(G''_i)] &= (1 - q_i)\mathbb{P}[i \in \mathcal{M}(G''_i) | (\mu(i), i) \notin E(G^m)] \\ &\quad + q_i \mathbb{P}[i \in \mathcal{M}(G''_i) | (\mu(i), i) \in E(G^m)] \\ &\leq (1 - q_i)\frac{p_i}{1 - q_i} + q_i = p_i + q_i. \end{aligned}$$

Then, by impartiality, $\mathbb{P}[i \in \mathcal{M}(G'_i)] = \mathbb{P}[i \in \mathcal{M}(G''_i)] \leq p_i + q_i$. It follows that

$$\begin{aligned} & \frac{\text{sc}(\mathcal{M}, G'_i)}{\text{sc}(\mathcal{M}^v, G'_i)} \\ &\leq \frac{(p_i + q_i)(M + (k-1)(|X| - 1)) + (1 - p_i - q_i)k(|X| - 1)}{M} \\ &= p_i + q_i + \frac{((p_i + q_i)(k-1) + (1 - p_i - q_i)k)(|X| - 1)}{M} \\ &\leq p_i + q_i + \frac{((p_i + q_i)k + (1 - p_i - q_i)k)(|X| - 1)}{M} \\ &= p_i + q_i + \frac{k(|X| - 1)}{M} \end{aligned} \tag{5}$$

where the first inequality comes from a simple calculation of scores, Equation (3), and the bound $p_i + q_i \geq \mathbb{P}[i \in \mathcal{M}(G'_i)]$.

On the other hand, let $p = \frac{\sum_{i \in X} p_i}{m}$. Then

$$\begin{aligned} \frac{\text{sc}(\mathcal{M}, G)}{\text{sc}(\mathcal{M}^v, G)} &\leq \frac{(k - \sum_{i \in X} p_i)|X| + \epsilon(n - |X|)\sum_{i \in X} p_i}{(1 - o(1))|X|k} \\ &= \frac{(k - \sum_{i \in X} p_i)m + \epsilon(n - m)\sum_{i \in X} p_i}{(1 - o(1))mk} \\ &= \frac{(k - pm)m + \epsilon\,(n - m)\,pm}{(1 - o(1))mk} \\ &= \frac{\left(1 - \frac{pm}{k}\right) + \epsilon\,(n - m)\,\frac{p}{k}}{(1 - o(1))} \leq \frac{\left(1 - \frac{pm}{k}\right) + \epsilon n \frac{p}{k}}{(1 - o(1))}. \end{aligned} \tag{6}$$

Now, some $p_i \leq p$, by a simple averaging argument; consider that i. In the construction of μ above, we showed the upper bound $q_i \leq \frac{m}{n - 2(i-1) - 1}$ on the probability that $(\mu(i), i)$ is sampled by $\mathcal{M}$. Notice that the approximation ratio for $\mathcal{M}$ is at most

$$\begin{aligned} \alpha &\leq \min\left\{ p_i + q_i + \frac{k(|X| - 1)}{M}, \frac{\left(1 - \frac{pm}{k}\right) + \epsilon n \frac{p}{k}}{(1 - o(1))} \right\} \\ &\leq \min\left\{ p + q_i + \frac{k(|X| - 1)}{M}, \frac{\left(1 - \frac{pm}{k}\right) + \epsilon n \frac{p}{k}}{(1 - o(1))} \right\}, \end{aligned}$$

by (5) and (6). Since ϵ is arbitrarily small, M is arbitrarily large, and $q_i = o(1)$, $\alpha \leq \min\left\{p, \left(1 - \frac{pm}{k}\right)\right\} + o(1)$. We derive an upper bound on the minimum by equalizing the two expressions and solving for p, which yields $p = \frac{k}{k+m}$. It follows that $\alpha \leq \frac{k}{k+m} + o(1)$. □

We remark that Alon et al. [2011] prove an upper bound of $\frac{k^2+k-1}{k^2+k}$ for their setting, which is the special case of ours in the regime $m = n - 1$. They do this by creating a graph where all edges have weight 0 except for a cycle of length $k + 1$ of edges of weight 1. One of the vertices in this cycle — call it i — is selected with probability at most $k/(k + 1)$. The upper bound is obtained by reducing the weight on i's outgoing edge to 0. In this new graph, i is still selected with probability at most $\frac{k}{k+1}$ by impartiality, so the mechanism's score is at most $\frac{k}{k+1}k + \frac{1}{k+1}(k-1)$, whereas the optimal solution (which is equivalent to VANILLA in this regime) achieves score k. It is interesting to note that this argument does not extend to the case of $m \ll n$, because VANILLA is unlikely to see the cycle of valuable edges.

5 Discussion

From a practical point of view, with NSF reviewing in mind, Theorem 1, and CREDIBLE SUBSET itself, are quite compelling. To implement the insights behind Theorem 1, one should slightly expand the set of eligible winners to include all "credible" proposals (associated with PIs who can manipulate their way into the top k), and randomly choose k among them. This seems justifiable, because it is difficult to distinguish between proposals at the very top.

Our formulation of CREDIBLE SUBSET selects empty subsets with small probability to achieve impartiality. As noted

above, we can replace this with a distribution over nonempty subsets. Moreover, in practice, this aspect of the mechanism can perhaps be ignored: PIs would be able to ever-so-slightly increase the probability of their own proposals being accepted by decreasing the number of credible vertices, but the incentives for manipulation under this almost impartial version of CREDIBLE SUBSET would be weak compared to VANILLA.

One of the ways in which the mechanism of Merrifield and Saari [2009] differs from our setting is that reviewers are restricted to ranking the proposals. Since Borda count is used to aggregate the rankings, this is equivalent to limiting the reviewers to handing out the ratings $m-1, m-2, \ldots, 0$ (exactly one of each) — even though their true ratings may be different. Our ideas readily extend to this setting.

Finally, while we have focused on NSF reviewing in the introduction (and, indeed, this is the real-world setting that motivated us), our results can certainly be applied to conference reviewing. For example, in large conferences such as AAAI and IJCAI, the PC includes hundreds of people — a large fraction of the researchers who actually submit papers to the conference. These conferences are a great fit with our model and results, because: (i) VANILLA is, essentially, the mechanism that is typically used (modulo choosing the m-regular graph in a way that matches reviewers with suitable papers), and (ii) k (the number of papers selected for presentation and publication) is much larger than m (the number of reviews per PC member) — in IJCAI'13 (the previous IJCAI), the values were $k = 413$ and $m < 10$, making the CREDIBLE SUBSET Mechanism (or a variation thereof) eminently practical.

Acknowledgments

Kurokawa and Procaccia were partially supported by the NSF under grants CCF-1215883 and IIS-1350598, and by a Sloan Research Fellowship. Morgenstern was partially supported by NSF grants CCF-1116892 and CCF-1101215. Lev was partially supported by Microsoft Research through its PhD Scholarship Program, Israel Science Foundation grant #1227/12, the Israel Ministry of Science and Technology — Knowledge Center in Machine Learning and Artificial Intelligence grant #3-9243. This work has also been partly supported by COST Action IC1205 on Computational Social Choice.

References

[Alon *et al.*, 2011] N. Alon, F. Fischer, A. D. 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 (TARK)*, pages 101–110, 2011.

[Berga and Gjorgjiev, 2014] D. Berga and R. Gjorgjiev. Impartial social rankings. Manuscript, 2014.

[de Clippel *et al.*, 2008] G. de Clippel, H. Moulin, and N. Tideman. Impartial division of a dollar. *Journal of Economic Theory*, 139:176–191, 2008.

[Douceur, 2009] J. Douceur. Paper rating vs. paper ranking. *Operating Systems Review*, 43:117–121, 2009.

[Fischer and Klimm, 2014] F. Fischer and M. Klimm. Optimal impartial selection. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*, pages 803–820, 2014.

[Haenni, 2008] R. Haenni. Aggregating referee scores: an algebraic approach. In *Proceedings of the 2nd International Workshop on Computational Social Choice (COMSOC)*, pages 277–288, 2008.

[Hazelrigg, 2013] G. A. Hazelrigg. Dear colleague letter: Information to principal investigators (PIs) planning to submit proposals to the Sensors and Sensing Systems (SSS) program October 1, 2013, deadline. http://www.nsf.gov/pubs/2013/nsf13096/nsf13096.jsp?WT.mc_id=USNSF_25#reference1, 2013. Retrieved on June 17, 2014.

[Holzman and Moulin, 2013] R. Holzman and H. Moulin. Impartial nominations for a prize. *Econometrica*, 81(1):173–196, 2013.

[Mackenzie, 2014] A. Mackenzie. Impartiality and symmetry. Manuscript, 2014.

[Merrifield and Saari, 2009] M. Merrifield and D. Saari. Telescope time without tears: a distributed approach to peer review. *Astronomy and Geophysics*, 50(4):2–6, 2009.

[Mitzenmacher, 2013] M. Mitzenmacher. NSF reviewing trial run. http://mybiasedcoin.blogspot.com/2013/06/nsf-reviewing-trial\\-run.html, 2013. Retrieved on June 17, 2014.

[Nierstrasz, 2000] O. Nierstrasz. Identify the champion. In N. Harrison, B. Foote, and H. Rohnert, editors, *Pattern Languages of Program Design*, volume 4, pages 539–556. Addison-Wesley, 2000.

[Procaccia, 2013] Ariel D. Procaccia. NSF (actually) reviewing via social choice. http://agtb.wordpress.com/2013/06/10/nsf-actually-reviewing-via-social-choice/, 2013. Retrieved on June 17, 2014.

[Roos *et al.*, 2011] M. Roos, J. Rothe, and B. Scheuermann. How to calibrate the scores of biased reviewers by quadratic programming. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence (AAAI)*, pages 255–260, 2011.

[Tamura and Ohseto, 2014] S. Tamura and S. Ohseto. Impartial nomination correspondences. *Social Choice and Welfare*, 43:47–54, 2014.

[Vohra, 2013] R. V. Vohra. A mechanism design approach to peer review. http://theoryclass.wordpress.com/2013/06/06/a-mechanism-design-approach-to-peer-review/, 2013. Retrieved on June 17, 2014.