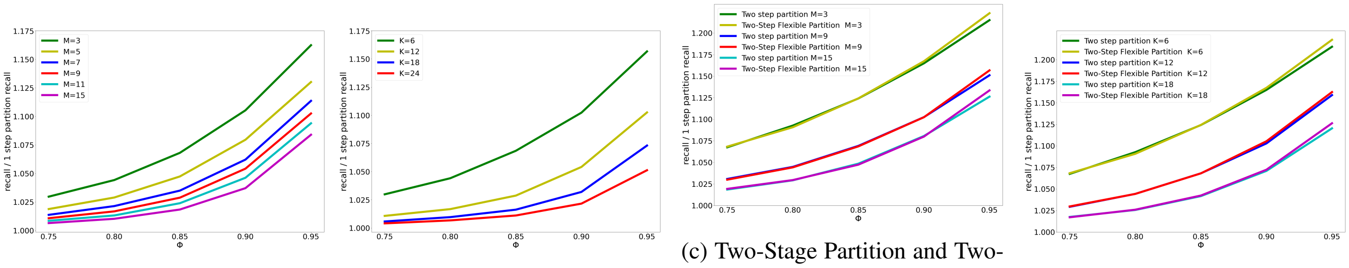


(a) Two-Stage Partition compared to 1-step Partition for different m and ϕ values. All runs with $k = 12$.

(b) Two-Stage Partition compared to 1-step Partition for different k and ϕ values. All runs with $m = 9$.

(c) Two-Stage Partition and Two-Stage Flexible Partition compared to 1-step Partition for different m and ϕ values. All runs with $k = 18$.

(d) Two-Stage Partition compared to 1-step Partition for different k and ϕ values. All runs with $m = 3$.

Figure 3: Percent of improvement in recall compared to 1-step Partition. All runs with optimal f for that m, k, and ϕ.

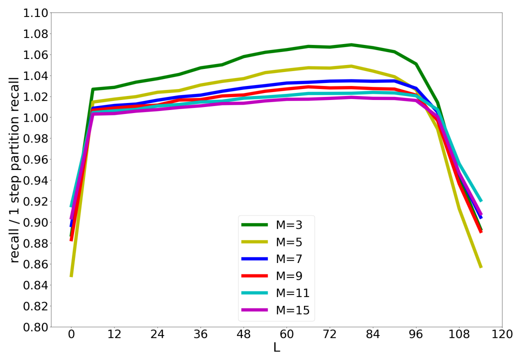


Figure 4: Percent of improvement of Two-Stage Partition recall over 1-step partition, for different values of l. All runs with $\phi = 0.85$, $k = 12$, $f = \frac{2m}{10}$, $h = 0$.

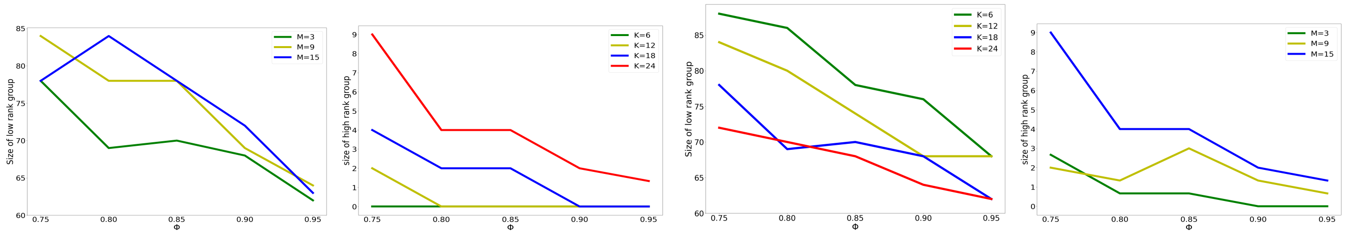
we see that is better to drop a large number of candidates after first round, ideally will be in range of $\frac{n}{2}$ to $\frac{3}{4}n$ candidates, with the precise value depending on the precise values of ϕ, k and m (see Figure 4). When k is smaller, a higher l seems to work better (as can seen in Figure 5c), perhaps because when we choose fewer candidates there is less likelihood they will end up in the bottom $\frac{3}{4}n$ after the first round, and as number of winners increase we need to be more cautious about those we eliminate. We found that higher ϕ values will lead to lower l values, probably since when reviewers are more noisy we also need be more cautious about the amount of data we use to eliminate agents. When m is large, a larger l works better (as can seen in Figure 5a), probably because the significant number of reviews means we have enough information about the candidate even from the first stage, allowing us to eliminate with confidence. For h (the size of the group of candidates we pass after the first stage), we see quite the opposite picture of that of l. It is better for h to be small, in range of $0 - \frac{k}{3}$ candidates, with the precise value depending on the precise ϕ, k and m (see Figure 6). When k is larger, a higher h seems to work better (as can seen in Figure 5b), perhaps because when we choose more candidates we probably will be in the top $\frac{k}{3}$ on the first round, and as number of winners decrease we need to be more cautious about those we choose. We found that higher ϕ values will lead to lower h values, probably since when reviewers are more noisy we also need be more cautious about the candidates we choose. When m is large, a larger h works better (as can seen in Figure 5d), probably because the significant number of reviews means we have enough information about the candidate even from the first stage, allowing us to choose with confidence.

Discussion

In this paper we investigate using a two-stage mechanism for peer-evaluation. While the use of such mechanisms in the real-world has expanded in the past few years (**?**), beyond the basic intuition behind it (focusing reviews on more "divisive" papers), there has not been, to our knowledge, any further investigation of this idea. Here, we took the most widely explored strategyproof mechanism – Partition – and examined its performance when adding a Two-Stage component to it, using two different methods to implement how the mechanisms decide on which candidates to focus (a fixed set vs. a flexible, changing set of papers).

While it seems the intuition is indeed correct, and focusing on a subset of papers does improve the performance of the peer-evaluation mechanism, the improvement was not where we expected it to be. We expected the "borderline" papers to be more exact. That is, that the paper ranked at $k - 1$ will more surely be included vs the paper ranked at $k + 1$. However, our simulations showed that this is not the key benefit of the Two-Stage mechanisms, but rather the more "middle-of-the-road" papers. Those ranked around $\frac{k}{2}$ benefited most, as their chance of being included in the winning set increased dramatically. It seems that borderline papers are hard to differentiate, even when getting more reviews; while the better papers were able to more clearly establish their quality.

In addition we were able to explore what parameters improve the algorithms' performance best, depending on the values of m, k, and ϕ. Somewhat surprisingly, a fairly small benefit first stage suffices to help the algorithms' performance, as long as enough papers are rejected. While this may seem counter-intuitive, it seems the limited signal in the first stage is enough such that the chances of getting enough reviews to counter it is of low-enough probability to not be



(a) Best performing size of l (bottom-ranked set size) for different m and ϕ values. All runs with $k = 18$.

(b) Best performing size of h (top-ranked set size) for different k and ϕ values. All runs with $m = 15$.

(c) Best performing size of l (bottom-ranked set size) for different k and ϕ values. All runs with $m = 3$.

(d) Best performing size of h (top-ranked set size) for different m and ϕ values. All runs with $k = 24$.

Figure 5: Best performing size of top/bottom set for different values. All runs with optimal f for that m, k, and ϕ.

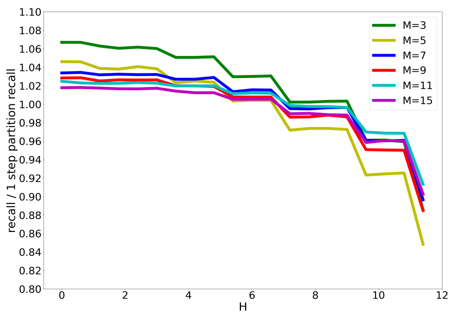


Figure 6: Percent of improvement of Two-Stage Partition recall over 1-step partition, for different values of h. All runs with $\phi = 0.85$, $k = 12$, $l = 0.7$, $f = 0.2$.

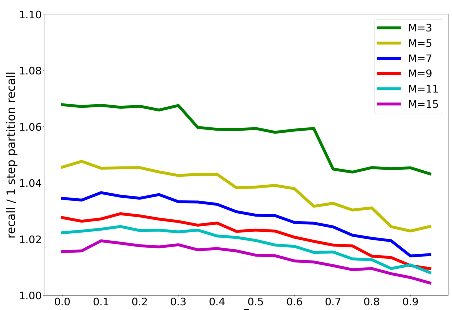


Figure 7: Percent of improvement of Two-Stage Partition recall over 1-step partition, for different values of f. All runs with $\phi = 0.85$, $k = 12$, $l = 0.7$, $h = 0$.

worth it.

There are some obvious extensions to this work: first and foremost, examining if we see similar outcomes in other peer-evaluation mechanisms. We hypothesize that we will see something similar (e.g., the two stages help the middle-of-the-road papers the most), but this has yet to be examined. Furthermore, for other mechanisms a two-stage mechanism may not be as straightforwardly strategyproof, and may require a far more complex re-working of the algorithms to accommodate a two-stage system. Beyond this, examining outcomes in distribution that are not Mallows may lead to deeper understanding of the two-stage systems (though, so far, peer-evaluation papers, requiring a ground-truth to compare themselves to, focus on Mallows distribution for comparison and quality estimates).

neque, iste reiciendis ipsum voluptas et distinctio commodi corporis sint?Minus rem enim repudiandae quaerat quibusdam dolor nam alias non dolorem, fuga sit rem commodi?Ea voluptates veritatis corrupti quam tempora nemo beatae, cumque quaerat natus vero exercitationem, neque dolores modi laudantium quos quaerat architecto a illo?Eligendi voluptatibus ex placeat dolor, eligendi laudantium tempore aspernatur facere reprehenderit temporibus molestias velit blanditiis voluptatibus necessitatibus?Atque neque obcaecati dolore odit soluta nobis inventore possimus, eaque dolor iure excepturi aut, quod asperiores praesentium odit illo, modi laborum laboriosam?Sequi dolor asperiores sapiente officia dolorum quisquam perspiciatis iusto repudiandae voluptatem deserunt, dolore culpa iste officiis itaque eaque totam nemo dolorem voluptate sapiente?Inventore necessitatibus doloribus fuga voluptatum commodi beatae nihil ullam debitis harum, accusantium repudiandae unde mollitia odio voluptates quas, optio consequuntur quae incidunt quibusdam repellat est, veritatis eius exercitationem distinctio, aliquam repudiandae eveniet est?Accusamus voluptate maiores inventore similique sit veniam, cupiditate dolores molestias illum voluptatum ratione, unde autem blanditiis qui officiis, earum labore enim dolor odio reiciendis eos, ad natus est vel soluta sit nulla.Eum provident est vero laborum vel adipisci neque explicabo ratione, facilis doloribus consequatur nesciunt ipsam tempore accusamus ducimus rem iusto nisi, molestias quisquam atque, inventore aut est dolorem voluptate dolor vitae pariatur culpa?Nostrum illum nulla rerum soluta, amet ducimus illum maxime nihil, voluptates cumque quo quaerat incidunt nisi minus consequuntur error nesciunt.Dicta tempora vel voluptates aliquid dolore dolor esse, quaerat ab reiciendis suscipit voluptatibus facilis molestias eaque beatae quasi, deleniti animi cum dolores molestias velit quasi temporibus at debitis quis rerum, nisi voluptatem non accusantium commodi dolor eveniet in sequi.Accusantium ex mollitia autem ut pariatur esse quia provident, sint perferendis repellendus ea ipsum iure nesciunt, perspiciatis dolorum molestias error possimus repellat eos at ullam, placeat aut facilis autem animi at repellendus quo dolorem sed non dolore, nemo sequi veritatis nihil quod doloribus.