A APPENDIX TO "TOP-k REPRESENTATIVE SPATIAL OBJECTS USING SPATIAL DIVERSIFIED PROPORTIONALITY"

We analyze the approximation bounds of our virtual grid based algorithm and *IAdU* greedy algorithm.

A.1 Bounds of virtual grid based Algorithm

We study the worst-case of the approximation quality of $S^{\mathcal{S}}(\mathcal{S}) = \sum_{pi \in S} S^{\mathcal{S}}(p_i)$ produced by our grid based algorithms. More precisely, we study how the ratio, ap, of the optimal $S^{\mathcal{S}}(\mathcal{S})$ (denoted as $S^{\mathcal{S}}_o(\mathcal{S})$) to the approximated $S^{\mathcal{S}}(\mathcal{S})$ (denoted as $S^{\mathcal{S}}_a(\mathcal{S})$) ranges (i.e., its lower and upper bounds). As we will discuss shortly, our approximation algorithm can either increase or decrease the $s(p_i, p_j)$ score of a pair of places, which consequences to have both an upper and lower bound of ap. We prove that $0.39 \leq ap \leq 2.54$ for $K \rightarrow \infty$ (Theorems 5.2 and 5.3).

Our approximation algorithm can compute either a higher or a lower value compared to the actual score $s(p_i, p_j)$ of a pair of places. Thus, we study the bounds of the worst case of the two cases separately:

- **Error due to $s(p_i, p_j)$ decrement (Case D, Figure 1)**; i.e. after relocation, $s(p_i, p_j)$ scores of pairs are decreased, e.g. the maximum decrease from 1 to 0.
- **Error due to $s(p_i, p_j)$ increment (Case I, Figure 2)**; i.e. after relocation, $s(p_i, p_j)$ scores of pairs are increased, e.g. the maximum increase from 0 to 1.

We prove our bounds by induction. More precisely, we prove two lemmas per case by induction. Our first lemma, (for the decrement case, Lemma A.1), proves that if points are originally co-located at the intersection of four cells, then we get the worst case if these points are evenly relocated in the four adjacent cells (i.e. $K/4$ places per cell). Analogously, for the increment case (lemma A.3), we prove that if all places are relocated to the center of a single cell (as they all reside in it), then we get the worst case if the places were originally evenly located at the four corners of the cell (i.e. $K/4$ places per corner). Our second pair of lemmas proves that these cases also constitute the worst cases of all cases for ap (lemma A.2 and A.4). Then, we calculate the approximation bound for these worst cases (Theorems 5.2 and 5.3).

A.1.1 Estimation of the ap upper bound due to $s(p_i, p_j)$ decrement (Case D).

Our proof is by induction, so we start our proof by studying ap for small values of K. For $K = 2$, we can have the worst case (D_G.1), i.e. $S^{\mathcal{S}}_o(\mathcal{S}) = 2$, $S^{\mathcal{S}}_a(\mathcal{S}) = 0$ and $ap = \infty$; when two places co-located on the intersection of four cells and then relocated to the centres of diametrically opposite cells. This will result to the maximum loss of $s(,)$ of the pair, from 1 to 0. Following the same scenario of co-located places, the addition of a third place will result to the maximum ap, if the three places are relocated in three different cells (case D_G.2). Namely, $S^{\mathcal{S}}_o(\mathcal{S}) = K \cdot (K-1) = 3 \cdot 2 = 6$ and $S^{\mathcal{S}}_a(\mathcal{S}) = 2 \cdot 0 + 4 \cdot \alpha = 1.17$ (where $\alpha = 1 - 1/\sqrt{2} \approx 0.293$); thus $ap = \frac{6}{1.17} = 5.1$. For a fourth place, we get a maximum ap when all places are relocated in four different cells (case D_G.3). Namely, $S^{\mathcal{S}}_o(\mathcal{S}) = 4 \cdot 3 = 12$ and $S^{\mathcal{S}}_a(\mathcal{S}) = 4 \cdot 0 + 8 \cdot \alpha = 2.34$, thus $ap = \frac{12}{2.34} = 5.1$. Note that any other arrangement, e.g. such as D_G.4 or D_G.5 will not give a higher ap. In case D_G.4, where our grid based algorithms co-locates two places, we get $ap = 6/2$. The case of D_G.5, where one place is located on the border with another cell, if our grid algorithm relocates this place to another cell will result to $ap = \frac{2}{1.1} = 1.8$; note that (1) the fact that the third place is not co-located with the two places reduces the $S^{\mathcal{S}}_o(\mathcal{S})$ and (2) the fact that it is relocated in a different cell increases max_D and consequently $S^{\mathcal{S}}_a(\mathcal{S})$. For the following two lemmas we use the same common base case, i.e. D_G.3 which is the worst case for $K = 4$.

Lemma A.1. *If all K places are co-located at the intersection of four cells, then the worst case will be when the places are relocated evenly to the adjacent four cells (i.e. $K/4$ per cell).*

Proof. We prove this by induction. Our base case is case $\mathrm{D}_G.3$ (for $K = 4$), where the four co-located points are then relocated to the four adjacent cells. Let's assume we have a case where we have K places at an intersection and the worst case is when $K/4$ points are relocated in each cell center. Without loss of generality, we want to prove that this also holds when we add four new places on $\mathcal{S}$ generating the inductive set $\mathcal{S}^+$. We have the following cases of $\mathcal{S}^+$: (1) we relocate one point in each cell, (2) we relocate two points in two cells, (3) we relocate two points in a cell and the other two points in two different cells, (4) we relocate three points in one cell and another point in another cell. Below, we will estimate the worst case for the four cases by calculating $S_a^{\mathcal{S}}(\mathcal{S}^+)$. We also show that the first case achieves the worst $S_a^{\mathcal{S}}(\mathcal{S}^+)$ and therefore the worst *ap*.

Given the four cells (Fig 1), $c_{1,1}, ..., c_{2,2}$, we have $s(c_{1,1}, c_{1,1}) = s(c_{1,2}, c_{1,2}) = s(c_{2,1}, c_{2,1}) = s(c_{2,2}, c_{2,2}) = 1$, $s(c_{1,1}, c_{1,2}) = s(c_{1,1}, c_{2,1}) = s(c_{2,2}, c_{2,1}) = s(c_{2,2}, c_{1,2}) = \alpha$ and $s(c_{1,1}, c_{2,2}) = s(c_{1,2}, c_{2,1}) = 0$.

For the first case, we now have $\frac{K}{4} + 1$ points in each cell. Thus we have $S_a^{\mathcal{S}}(c_{1,1}) = S_a^{\mathcal{S}}(c_{1,2}) = S_a^{\mathcal{S}}(c_{2,1}) = S_a^{\mathcal{S}}(c_{2,2})$. For instance, for cell $c_{1,1}$ we have $s(c_{1,1}, c_{1,1})$ for $\frac{K}{4} \cdot (\frac{K}{4} + 1)$ times, $s(c_{1,1}, c_{1,2})$ and $s(c_{1,1}, c_{2,1})$ for $(\frac{K}{4} + 1)^2$ times respectively. Thus, we have $S_a^{\mathcal{S}}(c_{1,1}) = S_a^{\mathcal{S}}(c_{1,2}) = S_a^{\mathcal{S}}(c_{2,1}) = S_a^{\mathcal{S}}(c_{2,2}) = \frac{K}{4} \cdot (\frac{K}{4} + 1) + 2 \cdot \alpha(\frac{K}{4} + 1)^2$ which collectively gives us $S_a^{\mathcal{S}}(\mathcal{S}^+) = 4 \cdot (\frac{K}{4} \cdot (\frac{K}{4} + 1) + 2 \cdot \alpha \cdot (\frac{K}{4} + 1)^2)$, i.e.:

$$S_a^{\mathcal{S}}(\mathcal{S}^+) = 0.396 \cdot K^2 + 2.168 \cdot K + 2.34 \tag{1}$$

The above formula also verifies the $S_a^{\mathcal{S}}(\mathcal{S})$ of our base case with four places. Note that the above formula is for $\mathcal{S}^+$ with $K + 4$, thus if we want to calculate the score for only 4 places we need $K = 0$ which will give us 2.34.

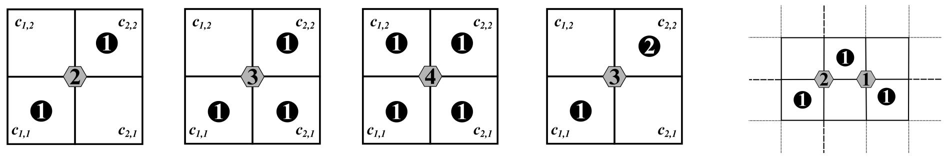
For the second case, we consider the worst case where the two pairs of points are relocated in diametrically opposite cells (e.g. $c_{1,1}$ and $c_{2,2}$; then we have $S_a^{\mathcal{S}}(c_{1,1}) = S_a^{\mathcal{S}}(c_{2,2})$). For instance, for $c_{1,1}$ we get $S_a^{\mathcal{S}}(c_{1,1}) = s(c_{1,1}, c_{1,1}) \cdot (\frac{K}{4} + 2)(\frac{K}{4} + 1) + (s(c_{1,1}, c_{1,2}) + s(c_{1,1}, c_{2,1})) \cdot \frac{K}{4}(\frac{K}{4} + 2)$. We also have $S_a^{\mathcal{S}}(c_{1,2}) = S_a^{\mathcal{S}}(c_{2,1})$, where $S_a^{\mathcal{S}}(c_{1,2}) = s(c_{1,2}, c_{1,2}) \cdot \frac{K}{4} \cdot (\frac{K}{4} - 1) + (s(c_{1,2}, c_{1,1}) + s(c_{1,2}, c_{2,2})) \cdot \frac{K}{4} \cdot (\frac{K}{4} + 2)$. Then, $S_a^{\mathcal{S}}(\mathcal{S}^+) = 2((\frac{K}{4} + 2)(\frac{K}{4} + 1) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 2)) + 2((\frac{K}{4} \cdot (\frac{K}{4} - 1) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 2))$ which finally gives us $pSS_a(\mathcal{S}^+) = 0.396 \cdot K^2 + 2.168 \cdot K + 4$. We see that this is larger than the score for case 1.

For the third case, we consider the worst case when the single points are relocated in diametrically opposite cells (e.g. $c_{1,1}$ and $c_{2,2}$) and the pair of places in $c_{1,2}$. Then for $c_{1,1}$ and $c_{2,2}$ we get $S_a^{\mathcal{S}}(c_{1,1}) = S_a^{\mathcal{S}}(c_{2,2}) = \frac{K}{4} \cdot (\frac{K}{4} + 1) + 2 \cdot \alpha \cdot (\frac{K}{4} + 1)(\frac{K}{4} + 2)$, for $c_{1,2}$ $S_a^{\mathcal{S}}(c_{1,2}) = (\frac{K}{4} + 2)(\frac{K}{4} + 1) + 2 \cdot \alpha \cdot (\frac{K}{4} + 1)(\frac{K}{4} + 1)$, and for $c_{2,1}$ $S_a^{\mathcal{S}}(c_{2,1}) = \frac{K}{4} \cdot (\frac{K}{4} - 1) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 1)$. Then, $pSS_a(\mathcal{S}^+) = 0.396 \cdot K^2 + 2.30 \cdot K + 4.9$. We see that this is larger than the score for case 1 and 2.

For the fourth case, we consider the worst case when the three places are relocated in one cell and the fourth place is relocated in another cell diametrically opposite (e.g. $c_{1,1}$ and $c_{2,2}$). Then for $c_{1,1}$ we get $S_a^{\mathcal{S}}(c_{1,1}) = (\frac{K}{4} + 3)(\frac{K}{4} + 2) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 3)$, for $c_{2,2}$ $S_a^{\mathcal{S}}(c_{2,2}) = \frac{K}{4}(\frac{K}{4} + 1) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 1)$, and for $c_{1,2}, c_{2,1}$ $S_a^{\mathcal{S}}(c_{1,2}) = S_a^{\mathcal{S}}(c_{2,1}) = \frac{K}{4} \cdot (\frac{K}{4} - 1) + 2 \cdot \alpha \cdot \frac{K}{4}(\frac{K}{4} + 2))$. Then, $pSS_a(\mathcal{S}^+) = 0.396 \cdot K^2 + 2.16 \cdot K + 6$. We see that this is larger than the score for all previous cases.

In summary, we see that the worst case is the first case (when the points are evenly relocated to the four cells). This completes the proof. □

Lemma A.2. *Given a set $\mathcal{S}$ with K places, the worst case of ap can happen only when all places are co-located at the intersection of four cells and then relocated evenly in the adjacent cells' centers.*



(a) D_G.1 ($ap = \infty$) (b) D_G.2 ($ap = 5.1$) (c) D_G.3 ($ap = 5.1$) (d) D_G.4 ($ap = 3$) (e) D_G.5 ($ap = 1.8$)

Fig. 1. Cases D: Error Decrement (hexagons and circles indicate the original and new location of places respectively; numbers indicate the amount of co-located places)

Proof. We proves this lemma by induction. Let D$_G$.3 (Fig 1(c)) for $K = 4$ be our base case again. Let assume the induction case where we have K points co-located at an intersection and give us indeed the worst case ap. Thus, we have $S_o^{\mathcal{S}}(\mathcal{S}) = K \cdot (K-1)$ (since all places are collocated) and $S_a^{\mathcal{S}}(\mathcal{S}) = \sum_{p_i \in \mathcal{S}} S^{\mathcal{S}}(p_i) = \sum_{p_i,p_j \in \mathcal{S}, i \neq j} s(p_i, p_j) = \sum_{p_i,p_j \in \mathcal{S}, i \neq j} \left(1 - \frac{\|p_i,p_j\|}{\max_D}\right) = K \cdot (K-1) - \sum_{p_i,p_j \in \mathcal{S}, i \neq j} \frac{\|p_i,p_j\|}{\max_D}$, where $\max_D = \sqrt{2}$ (i.e. the largest distance of the centers of four adjacent cells). According to Lemma A.1, the worst approximation for this case is when the points are evenly relocated in the neighboring cells so we have $K/4$ points per cell. Thus, we have $\sum_{p_i,p_j \in \mathcal{S}, i \neq j} \|p_i, p_j\| = 4 \cdot (2(\frac{K}{4})^2 + 2 \cdot \sqrt{2}(\frac{K}{4})^2) = 1.20K^2$.

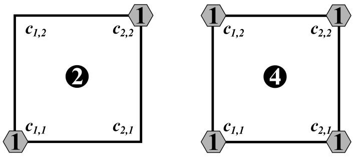
Now, let's study the two cases when we add a new place to $\mathcal{S}$ producing inductive set $\mathcal{S}^+$ (i.e. we have $K+1$ places). Case A where we add the new place on the same intersection and Case B where we add a place in such a way that is relocated in a different square. In case A, we have $ApS_o = S_o^{\mathcal{S}}(\mathcal{S}^+) = K \cdot (K+1)$. Assuming even distributions of places (According to Lemma A.1), then we have the additional score $2\frac{K}{4} + 4\alpha \cdot \frac{K}{4} = 0.79K$ to $ApS_a = S_a^{\mathcal{S}}(\mathcal{S})$. Thus, we have $ApS_a = S_a^{\mathcal{S}}(\mathcal{S}^+) = K \cdot (K-1) - \frac{1.20K^2}{\sqrt{2}} + 0.79K$.

In case B, we have $BpS_o = S_o^{\mathcal{S}}(\mathcal{S}^+) = K \cdot (K-1) + 2 \cdot \frac{K \cdot d_i}{\max_D}$, where $\max_D = \sqrt{5}$ which is the maximum distance on two adjacent squares and $d_i = 1$ is the distance of the centers of two adjacent cells. Note that as $\max_D$ increases ap decreases, thus this is the next smallest $\max_D$ after $\sqrt{2}$ (the $\max_D$ of a single cell). Apparently, $ApS_o > BpS_o$ since ApS_o corresponds to the maximum value for $K+1$ points. Assuming even distributions of places (According to Lemma A.1), then we have the additional score $2\frac{K}{4}(1 - \frac{2}{\sqrt{5}}) + 2\frac{K}{4}(1 - \frac{\sqrt{2}}{\sqrt{5}}) + 2\frac{K}{4}(1 - \frac{1}{\sqrt{5}}) = 0.51K$ (i.e. $max_D = \sqrt{5}$) to $S_a^{\mathcal{S}}(\mathcal{S})$. Thus, we have $BpS_a = S_a^{\mathcal{S}}(\mathcal{S}^+) = K \cdot (K-1) - \frac{1.20K^2}{\sqrt{5}} + 0.51K$. We can easily see that $ApS_a < BpS_a$. In summary, since $ApS_o > BpS_o$ and $ApS_a < BpS_a$, the inequality $\frac{ApS_o}{ApS_a} > \frac{BpS_o}{BpS_a}$ holds. This completes the proof of the lemma. □

Theorem 5.2 The ap of the algorithm is upper bounded by 2.54 for $K \longrightarrow \infty$.

Proof. According to our two lemmas, we have the worst case when places are originally co-located at an intersection of four cells, and then evenly relocated to the four adjacent cells. We can proceed and calculate the ap for this case as follows. Let for $K+4$ places (since in lemma A.1 we calculated $S_a^{\mathcal{S}}(\mathcal{S}^+)$ for $K+4$), we have $S_o^{\mathcal{S}}(\mathcal{S}^+) = (K+4)(K+3) = K^2 + 6K + 12$ and $S_a^{\mathcal{S}}(\mathcal{S}^+) = 0.396 \cdot K^2 + 2.168 \cdot K + 2.33$ (Eq. 1) which gives us $ap = \frac{K^2+6K+12}{0.396 \cdot K^2 + 2.168 \cdot K + 2.33}$; for $K \longrightarrow \infty$ ap converges to $1/0.396 = 2.54$. This completes the proof. □

A.1.2 *Estimation of the ap lower bound due to $s(p_i, p_j)$ increment (Case I).*



(a) $I_G.1$ ($ap = 0$) (b) $I_G.2$ ($ap = 1.95$)

Fig. 2. Cases I: Error Increment (hexagons and circles indicate the original and new location of places respectively; numbers indicate the amount of co-located places)

Again our proof is by induction, so we start our proof by studying ap for small values of K. In this case, we assume that originally places were located in such a way that minimises $S_o^{\mathcal{S}}(\mathcal{S})$. Then, our algorithm relocates all places on the same location which will give us the maximum $S_a^{\mathcal{S}}(\mathcal{S}) = K \cdot (K-1)$ score. For $K = 2$, we can have the worst case (I$_G$.1), i.e. $S_o^{\mathcal{S}}(\mathcal{S}) = 0$, $S_a^{\mathcal{S}}(\mathcal{S}) = 2$ and $ap = 0$; when the two places are located in two opposite corners of a cell (i.e., $s(,) = 0$). After relocation both will be relocated on the centre of the cell, thus $S_a^{\mathcal{S}}(\mathcal{S}) = 2$. Following the same scenario, let assume that we have four places located in the four corners of the cell (case I$_G$.2). We can easily see that this arrangement will result to the minimum possible summation of $s(,)$ among these four places, so $S_o^{\mathcal{S}}(\mathcal{S}) = 2.34$ and $S_a^{\mathcal{S}}(\mathcal{S}) = 12$, thus $ap = 1.95$.

From the discussion so far, we can easily see that the increment case is the reverse of the decrement case. More precisely, the $S_o^{\mathcal{S}}(\mathcal{S})$ of the increment case corresponds to $S_a^{\mathcal{S}}(\mathcal{S})$ of the decrement case. Where in the increment case, points are originally located in the corners of a cell (denoted as $c_{1,1}, ..., c_{2,2}$) whereas in the decrement case in the four adjacent cells (denoted also as $c_{1,1}, ..., c_{2,2}$). The $S_a^{\mathcal{S}}(\mathcal{S})$ of the increment case corresponds to $S_o^{\mathcal{S}}(\mathcal{S})$ of the decrement case. Where in both cases the points are co-located either at the center of the cell (increment case) or at the intersection of four cells (decrement case). Considering these reductions and proofs of the decrement case the following two lemmas and theorem also hold.

Lemma A.3. *If all K places are relocated to the center of a single cell (as they all reside within a single cell), then the worst case will be when the places were originally located evenly to the four corners of the cell (i.e. $K/4$ per corner).*

Lemma A.4. *Given a set $\mathcal{S}$ with K places, the worst case of ap can happen when all places are located evenly in the four corners of a single cell and then they are all relocated at the center of the cell.*

Theorem 5.3 The ap of the algorithm is lower bounded by 0.39 for $K \longrightarrow \infty$.

A.2 Bounds of the Greedy *IAdU* Algorithm

We know from previous work that *IAdU* algorithm can achieve approximation ratio of 4 when $H(u, v)$ (Equation 8) satisfies triangle inequality [2], [1]. For this purpose, we investigate when $H(u, v)$ satisfies the triangle inequality. Then, by using this key observation, we can trivially prove the approximation loss.

Theorem 6.1 $H(u, v)$ satisfies the Triangle Inequality when $r(v) \geq w \cdot \frac{k-1}{K-k}$

Proof. By expanding $H(u, v)$ we get:
$\frac{K-k}{k-1} \cdot (r(u) + r(v)) + \frac{1}{k-1} \cdot (S^{\mathcal{S}}(u) + S^{\mathcal{S}}(v)) - 2 \cdot w \cdot s(u, v) +$
$\frac{K-k}{k-1} \cdot (r(v) + r(w)) + \frac{1}{k-1} \cdot (S^{\mathcal{S}}(v) + S^{\mathcal{S}}(w)) - 2 \cdot w \cdot s(v, w)) \geq$
$\frac{K-k}{k-1} \cdot (r(u) + r(w)) + \frac{1}{k-1} \cdot (S^{\mathcal{S}}(u) + S^{\mathcal{S}}(w)) - 2 \cdot w \cdot s(u, w)) \implies \frac{K-k}{k-1} \cdot r(v) + \frac{1}{k-1} \cdot S^{\mathcal{S}}(v)$
$-w \cdot s(u, v) - w \cdot s(v, w) \geq -w \cdot s(u, w) \implies$

$\frac{K-k}{k-1} \cdot r(v) + \frac{1}{k-1} \cdot S^{\mathcal{S}}(v) - w \cdot (s(u,v) + w \cdot s(v,w) - w \cdot s(u,w)) \geq 0 \implies$
$\frac{K-k}{k-1} \cdot r(v) + \frac{1}{k-1} \cdot S^{\mathcal{S}}(v) - w \cdot (1 - \frac{||p_u,p_v||}{max_D} - \frac{||p_v,p_w||}{max_D} + \frac{||p_u,p_w||}{max_D}) \geq 0.$

Since $||p_u, p_v||$ satisfies triangle inequality (euclidean distance) then $\frac{||p_u,p_v||}{max_D}$ do as well. Since $\frac{||p_u,p_v||}{max_D}$ ranges in $[1, 0]$ and satisfies triangle inequality, then the minimum value for $\frac{||p_u,p_v||}{max_D} + \frac{||p_v,p_w||}{max_D} - \frac{||p_u,p_w||}{max_D}$ is 0. Then we have: $\frac{K-k}{k-1} \cdot r(v) + \frac{1}{k-1} \cdot S^{\mathcal{S}}(v) - w \geq 0 \implies (K-k) \cdot r(v) \geq w \cdot (k-1) \implies r(v) \geq w \cdot \frac{k-1}{K-k}$. □

For further simplification, we have dropped $S^{\mathcal{S}}(v)$ (which is the summation of $K-k$ places (including $s(u,v)$ and $s(v,w)$) and thus should be a significant value.

If we see more carefully this inequality, it holds in most pragmatic cases and our default settings. For $K = 1000, k = 20, w = 0.25K/k = 12.5$, then we get: $r(v) \geq 12.5 \cdot \frac{19}{980} = 0.24$. This is a pragmatic case as results with smaller $r(v)$ are not really relevant and they never make it in the $\mathcal{S}$. For instance, in OS cases we include objects with at most 3 hops from the root and thus $r(p_i) \geq 0.28$. Hence, in most pragmatic cases this holds.

REFERENCES

[1] Refael Hassin, Shlomi Rubinstein, and Arie Tamir. 1997. Approximation algorithms for maximum dispersion. *Operations Research Letters* 21, 3 (1997), 133–137.

[2] S. S. Ravi, Daniel J. Rosenkrantz, and Giri Kumar Tayi. 1991. *Facility dispersion problems: Heuristics and special cases.* 431–450.