Appendix to “Reverse Approximate Nearest Neighbor Queries”

Arif Hidayat, Shiyu Yang, Muhammad Aamir Cheema, David Taniar

1 Influence Zone for RANN Queries

In this section, we improve the extended influence zone algorithm for continuous RANN queries by reducing the number of pruning circles required to construct the influence zone. Suppose that the algorithm has already accessed a set of facilities $F_{accessed}$ and used their pruning circles to prune the search space. Let the pruned area be denoted as $\mathcal{A}$. A facility $f \notin F_{accessed}$ is called a *useless* facility if it does not prune any additional area, i.e., the pruning circle of f is fully contained in $\mathcal{A}$. We propose techniques to identify an entry e of the facility R*-tree (called useless entry) that contains only useless facilities. The algorithm can then create the influence zone by traversing the facility R*-tree (pruning useless entries) and creating the pruning circles for only the facilities that are not useless. First, we present the concept of an *extended rectangle*.

Definition 1 (Extended rectangle)**.** *Let C_p denote the pruning circle of a point p with respect to a query q. Let a, b, c and d denote the four corners of an MBR of a facility R*-tree entry e. Its extended rectangle $ABCD$ is the minimum bounding rectangle of the circles C_a, C_b, C_c, and C_d.*

Consider the MBR $abcd$ in Fig 1. Its extended rectangle $ABCD$ is the minimum bounding rectangle of the pruning circles C_a, C_b, C_c and C_d as shown in Fig. 1.

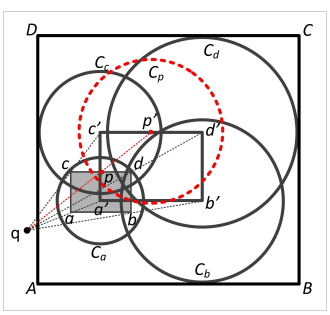

Fig. 1: Extended rectangle

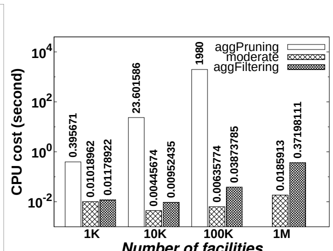

Fig. 2: Comparing strategies

- A. Hidayat is with the Faculty of Information Technology, Monash University, Australia and with Brawijaya University, Indonesia. E-mail: arif.hidayat@monash.edu
- S. Yang is the corresponding author. He is with the School of Computer Science and Engineering, The University of New South Wales, Australia E-mail: yangs@cse.unsw.edu.au
- M. A. Cheema and D. Taniar are with the Faculty of Information Technology, Monash University, Australia E-mail: {aamir.cheema, david.taniar}@monash.edu

Lemma 1. *Let $\mathcal{A}$ denote the area pruned by $F_{accessed}$. Let $abcd$ be the MBR of an unaccessed facility entry e. The entry e is a useless entry if its extended rectangle $ABCD$ is contained by $\mathcal{A}$.*

Proof. Consider MBR $abcd$ and its extended rectangle $ABCD$ in Figure 1. We show that the pruning circle C_p of every point p inside the MBR is contained by $\mathcal{A}$. For a point p, we denote its pruning circle as C_p, the center of the circle as p' and the radius as r_p.

Without loss of generality, let p be a facility point on line cd and C_p be its pruning circle (see the dotted red circle). Since c, d and p are on the same horizontal line, it can be proved that the centers of their pruning circles c', d' and p' are also on a horizontal line. Since $dist(q, p) < dist(q, d)$,according to the definition of pruning circle (Definition 2), $r_p < r_d$. Since $ABCD$ contains C_c and C_d and $r_p < r_d$, r_p can only be outside of $ABCD$ if r_p intersects with the line AD, i.e., r_p can be outside of $ABCD$ only if $r_p > c'p' + r_c$. Next, we prove that this is not possible and $r_p \leq r_c + c'p'$ or $r_c + c'p' - r_p \geq 0$.

From triangle $\triangle qc'p'$, we have

$$c'p'^2 = qc'^2 + qp'^2 - 2.qc'.qp' \cos\theta \quad (1)$$

where $\theta = \angle c'qp'$. Since $qc' = \frac{x^2.qc}{x^2-1}$ and $qp' = \frac{x^2.qp}{x^2-1}$ (Definition 2), we have

$$\begin{aligned} c'p'^2 &= (\frac{x^2 \cdot qc}{x^2-1})^2 + (\frac{x^2 \cdot qp}{x^2-1})^2 - 2 \cdot (\frac{x^2 \cdot qc}{x^2-1}) \cdot \\ &\quad (\frac{x^2.qp}{x^2-1}) \cdot cos(\theta) \\ &= \frac{x^4}{(x^2-1)^2}(qc^2 + qp^2 - 2 \cdot qc \cdot qp \cdot cos(\theta)) \end{aligned}$$

Since $qc^2 + qp^2 - 2 \cdot qc \cdot qp \cdot cos(\theta) = cp$, we have

$$c'p' = \sqrt{(\frac{x^4 \cdot cp^2}{(x^2-1)^2})} = \frac{x^2 \cdot cp}{x^2-1} \quad (2)$$

Since $r_c = \frac{x \cdot qc}{x^2-1}$, $r_p = \frac{x \cdot qp}{x^2-1}$ (Definition 2) and $c'p' = \frac{x^2.cp}{(x^2-1)}$ (Eq. (2)),

$$\begin{aligned} r_c + c'p' - r_p &= \frac{x \cdot qc}{x^2-1} + \frac{x^2 \cdot cp}{(x^2-1)} - \frac{x \cdot qp}{x^2-1} \\ &= \frac{x}{x^2-1}(qc + x.cp - qp) \end{aligned}$$

From triangle inequality, $qc + cp > qp$. Since $x > 1$, $\frac{x}{x^2-1}(qc + x \cdot cp - qp) > 0$ which completes the proof. $\square$

2 Additional Experimental Results

2.1 Effectiveness of pruning/filtering strategies

As discussed in Section 3.3 of the manuscript, different strategies could be used to shortlist the facility entries for the filtering phase. In this section, we briefly describe three strategies and compare their effectiveness.

Aggressive pruning approach. The aggressive pruning approach shortlists every facility entry that prunes at least one point in the search space not pruned by the previously considered facility entries. This approach ensures that the pruned space is maximized, i.e., any user u that lies outside the pruned space is guaranteed to be an answer.

Moderate pruning approach. The moderate pruning approach shortlists only the facility entries that lie outside the existing pruned space. This is inspired by the observation that a facility entry that lies inside the pruned area is less likely to prune additional area and, therefore, is less effective.

Aggressive filtering approach. We designed another strategy called *aggressive filtering* which uses the moderate pruning approach but filtering is done aggressively. Specifically, in our main approach, if a user entry U is not filtered by the shortlisted facilities, all its children are considered to be candidates which need to be verified. In the *aggressive* filtering approach, if a user entry U is not filtered by the shortlisted facilities, we traverse facility R*-tree to see if this entry can be pruned by considering other facilities in the facility R*-tree. This is done using the same idea used in the improved range query algorithm discussed in Section 5.1 of the conference version of this paper. The idea is to traverse facility R*-tree by visiting only those nodes that may filter the user entry. Since this aggressive filtering is done using the facility R*-tree instead of the shortlisted facilities, its filtering power is the same as the filtering done by the facilities shortlisted by aggressive pruning approach. In other word, every user u not filtered by aggressive filtering is an answer.

In Figures 2, we compare aggressive pruning approach (shown as aggPruning). aggressive filtering approach (shown as aggFiltering) and moderate pruning approach. Specifically, we study the effect of the number of facilities where other settings are the same as default settings in our experiments, e.g., # users is $100,000$, # of queries is 100 and average cost is displayed. Figure 2 shows that moderate pruning approach outperforms both aggressive pruning and aggressive filtering approaches – the cost of aggPruning is not shown for 1 Million facilities because it failed to return results even after a couple of days.

The results can be explained by Figure 3 and Figure 4. Specifically, Figure 3 shows the number of facility entries shortlisted in the pruning phase by each approach. Since aggressive pruning approach aims to maximize the pruned space, it ends up shortlisting a much higher number of facilities, e.g., for $100,000$ facilities, the average number of shortlisted facilities by aggressive pruning is more than $58,000$ as compared to only around 77 facilities shortlisted by the moderate pruning approach. On the other hand, Figure 4 shows that the number of candidate users (i.e., the users that cannot be filtered by the shortlisted facilities) are comparable for both approaches. Specifically, for $100,000$ facilities, the average number of candidates for aggressive pruning approach is 3 as compared to 3.42 for moderate pruning approach. Even for 1000 facilities, the number of candidates for moderate pruning approach is around 2.5 times of the number of candidates generated by aggressive pruning which uses around 20 times more facilities for pruning. Therefore, the pruning and filtering cost of aggressive approach is significantly higher because it uses a much higher number of entries for pruning and filtering. However, the number of candidates generated by moderate approach is comparable.

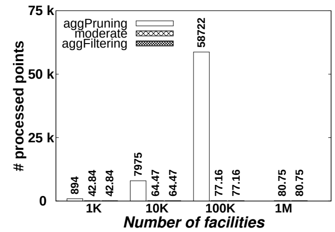

Fig. 3: # shortlisted entries

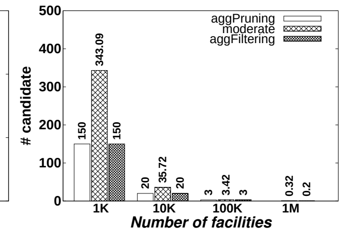

Fig. 4: # of candidates

Note that the aggressive filtering (shown as aggFiltering) is significantly more effective than the aggressive pruning approach but it is still worse than moderate approach. However, we remark that, in the data sets that may exhibit worst-case scenarios where most of users are failed to be pruned by the facilities shortlisted by moderate pruning approach, the aggressive filtering approach may be the right choice. Therefore, the applications where avoiding the worst-case scenarios is important, the aggressive filtering approach may be preferable.

2.2 Comparing # of user updates

Whenever a user leaves its safe zone, it sends its updated location to the server. In this section, we compare the total number of user updates for all 100 timestamps for our approach and InfZone. Note that the total number of updates reflect the size of safe zones.

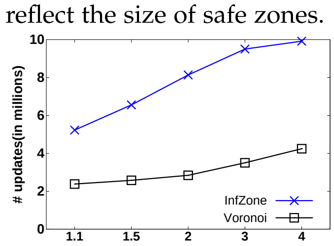

Fig. 5: Effect of the x factor

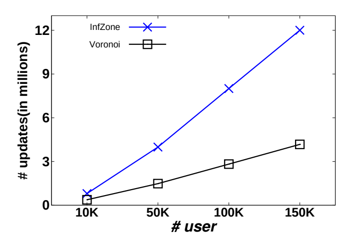

Fig. 6: Effect of # of users

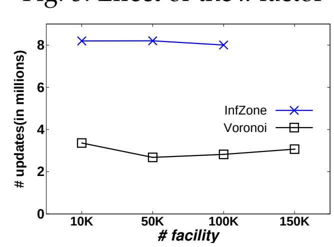

Fig. 7: Effect of # of facilities

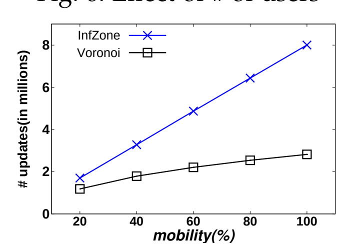

Fig. 8: Effect of mobility

Figures 5 - 8 compare the total number of updates for both approaches for varying values of x, the total number of users, the total number of facilities and the mobility. As expected, the number of updates for our approach is significantly smaller than the number of updates by InfZone which shows the effectiveness of our approach in creating bigger safe zones. Figure 7 shows that the number of updates do not necessarily increase as the number of facilities increases. This is because when there are too few facilities, the Voronoi cells are bigger and hence the number of circles overlapping the Voronoi cells increases (resulting in smaller safe zones). On the other hand, when the number of facilities is too big, the Voronoi cells are too small resulting in smaller safe zones.