Provable Sensor Sets for Epidemic Detection over Networks with Minimum Delay

Jack Heavey[1], Jiaming Cui[2], Chen Chen[1], B. Aditya Prakash[2], Anil Vullikanti[1]
University of Virginia[1] and Georgia Institute of Technology[2]

Introduction

- The efficient detection of outbreaks within networks has become a common topic of study. A method of surveilling these networks is choosing a "sensor set," a subset of size k of the original network, to monitor consistently.
- Choosing the optimal sensor set is a common area of research, with the most common method being using a greedy approach proposed by Leskovec et al.[1], who use submodularity to maximize the probability of detecting an infection within the network.
- We study the related problem of finding a sensor set that minimizes the delay of detection of an infection within a network, which we call the Minimum Delay Sensor Set (MinDelSS) problem.
- We show that the MinDelSS problem cannot be approximated within an $O\left(n^{1-1/\gamma}\right)$-factor approximation, and solutions given by the greedy algorithm can be a factor of $\Omega(nT(S^*))$, where S^* is the optimal solution
- We introduce a bicriteria algorithm that gives a worst case $O(\log n)$-factor for the average delay while violating the budget by a factor of $O(log^2 n)$.

Our Approach

We define a linear program that finds the optimal sensor set to solve the MinDelSS Problem:

$$\min \sum_{d=0}^{n+1} \frac{1}{N} \sum_{i=1}^{N} y_{id} \cdot d \qquad (1)$$

$$\text{for all } i, d\text{:} \sum_{u \in V_{id}} x_u \geq y_{id} \qquad (2)$$

$$\sum_u x_u \leq k \qquad (3)$$

$$\text{for all } i\text{:} \sum_d y_{id} = 1 \qquad (4)$$

$$x_u, y_{id} \in \{0,1\} \qquad (5)$$

The binary constraints in (5) make this linear program infeasible for large networks. Instead, we define a new algorithm RoundSensor.

RoundSensor

1) Solve the linear programming problem while relaxing (5) to x_u, $y_{id} \in [0,1]$ for all u, i, d.
2) For each node u, add u to sensor set S_r with probability $x_u' = \min\{1, x_u \log(n+1) \log(Nn)\}$
3) Return S_r

Worst-Case Bounds

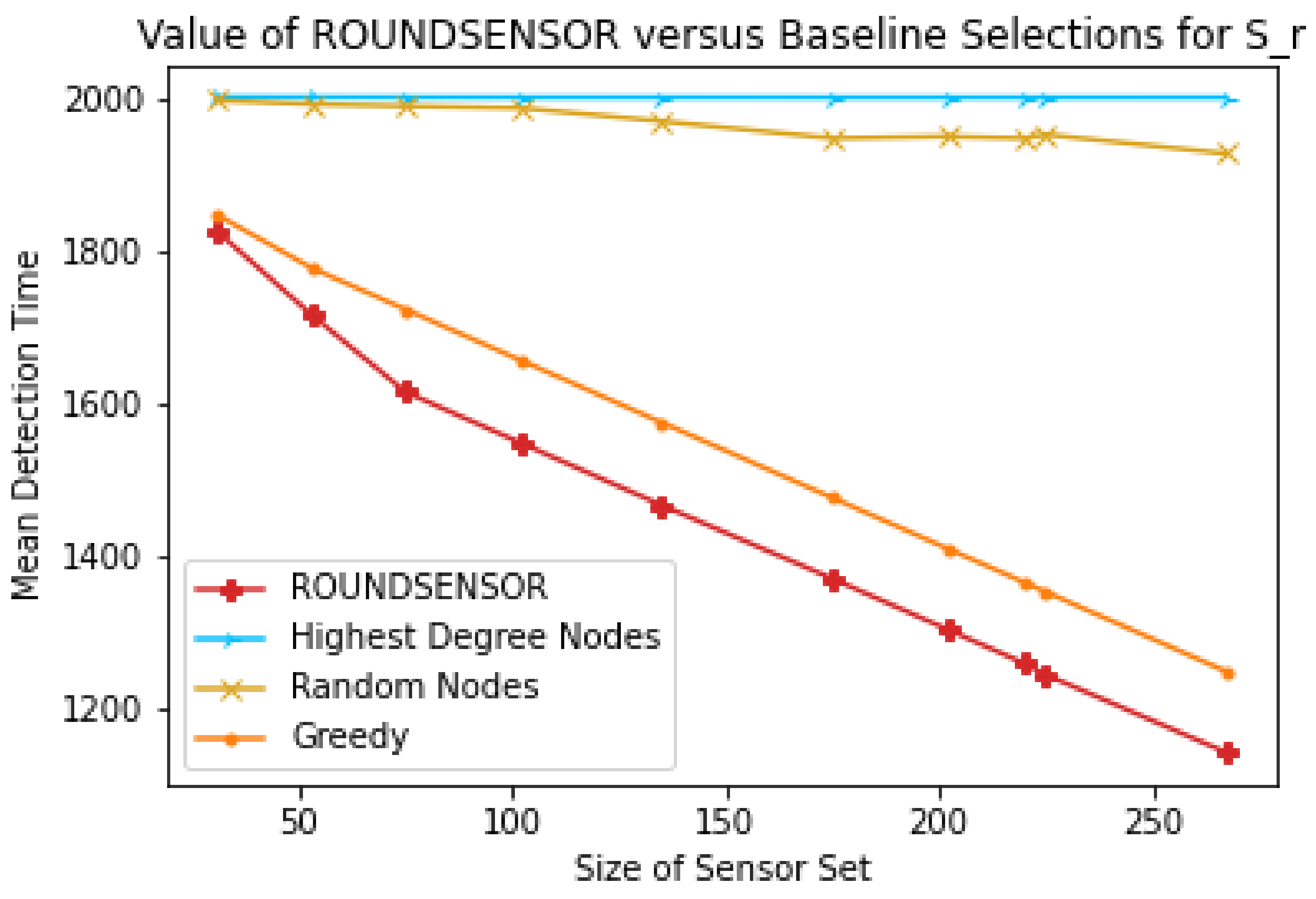
We prove rigorous worst case bounds for the sensor set produced by RoundSensor depending on whether cascades are provided to us or not. If N cascades are provided, with probability of at least $1 - \frac{2}{n}$:

1) We have $|S_r| \leq k \cdot 2\log(Nn)\log(n+1)$
2) We have $T_{avg}(S_r) \leq 2\log(n+1)\, T_{avg}(S^*)$ where S^* is the optimal solution.

If cascades are not provided, N cascades can be sampled using an SIR methodology. Let $N \geq \frac{3}{\varepsilon^2} n(n+1)\log(n)$ for $\varepsilon \in (0,1)$. We show that with at least probability $1 - \frac{3}{n}$:

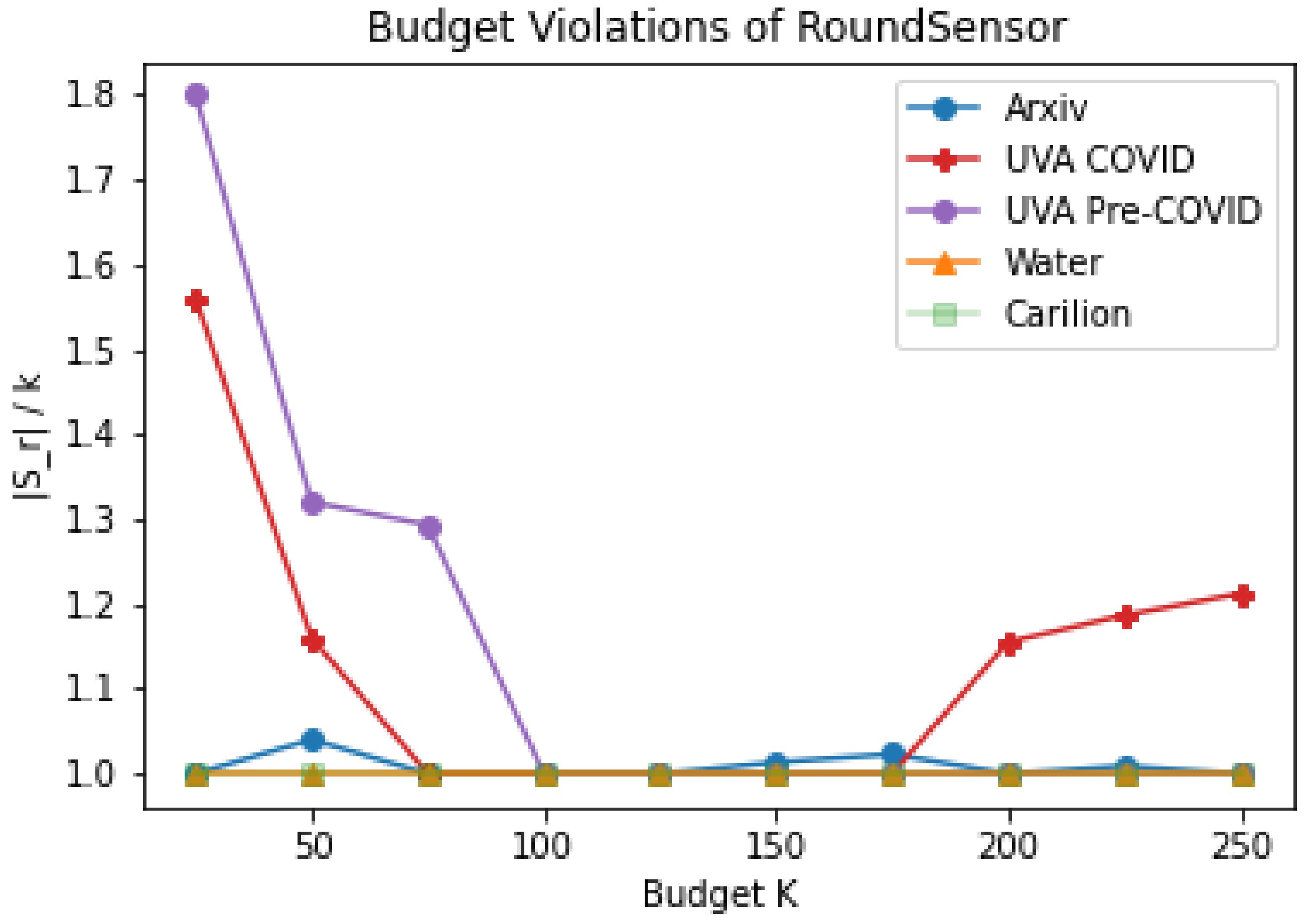
1) We have $|S_r| \leq k \cdot 7\log\left(\frac{1}{\varepsilon}\right) log^2(n)$
2) We have $E[T(S_r)] \leq 2(1+\varepsilon)\log(n+1)\, E[T(S^*)]$ where S^* is the optimal solution.

Experiments with RoundSensor



RoundSensor outperforms the baselines of selecting random nodes or selecting the highest degree nodes by a large margin, while also outperforming the greedy algorithm proposed by Leskovec et al.[1], even after accounting for possible budget violations. The mean detection time was up to 9% lower than our greedy baselines on a network of 9949 nodes and 399495 edges.

Experiments cont.



Across the five datasets used for testing, our randomized rounding approach never violated our initial budget k by a factor of more than 2. This is much better than the worst-case bounds that we prove within the paper.

Discussion and Conclusion

We show that our algorithm produces an output closer to the optimal sensor set for daily testing of individuals within a network, with experiments showing results that are even better than the worst-case bounds proven.

While our algorithm has the downside of not providing certainty in the testing budget, we believe that this tradeoff for a more optimal solution is worth it.

Future work would involve generalizing the linear program used in RoundSensor to handle looser periodic testing guidelines rather than daily testing.

Acknowledgements

This paper is based on work partially supported by NSF (Expeditions CCF-1918770 and CCF-1918656, CAREER IIS-2028586, RAPID IIS-2027862, Medium IIS-1955883, Medium IIS-2106961, CCF-2115126, IIS-1931628, IIS-1955797), CDC MInD U01CK000589, NIH R01GM109718, DTRA subcontract/ARA S-D00189-15-TO-01-UVA, ORNL, faculty award from Facebook, and funds/computing resources from Georgia Tech.

---

[1] Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; and Glance, N. 2007. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 420–429. ACM.

jch7jm@virginia.edu, jiamingcui1997@gatech.edu, zrh6du@virginia.edu, badityap@gatech.edu, vsakumar@virgina.edu