Project 1.8

# Discovering Key Entities in Large Complex Systems via Scalable Submodular Optimization

## Concurrent Design and Control of Complex Systems

Arun Sathanur, Shaobu Wang, Renke Huang and Mahantesh Halappanavar

CCSI Control of Complex Systems Initiative

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

U.S. DEPARTMENT OF ENERGY

PNNL-SA-XXXXX

## Objective

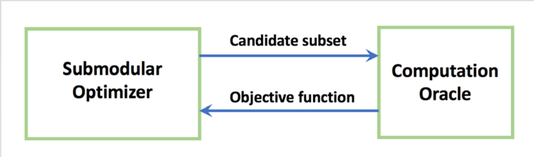
Development of *scalable* combinatorial optimization approaches by leveraging *sub-modular objective functions* to address observability and controllability in large complex systems - spatial and networked.

Example submodular optimization scenarios

- Sensor placement in spatial and networked systems to maximize mutual information between observed and unobserved locations
- Computation of a summary set of entities from a large data-set
- Localizing the spread of power outage via controlled islanding
- Opinion shaping and incentivizing on social networks for user behavior adoption (such as EV/Solar Panel ownership)

## Methodology

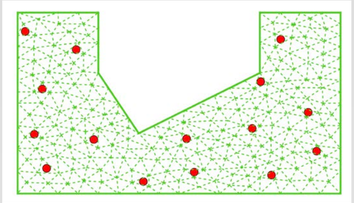
Leveraging submodular maximization as a scalable combinatorial optimization engine




Two pieces to scale:

- Scale the greedy algorithm. Ordinary Greedy Algorithm: O(*kn*) oracle calls. Active area of research in ML.
- Scale the computation oracle via advances in linear algebraic techniques/HPC. Oracle Complexity varies between applications.

## Example Use Case: Spatial Sensor Placement

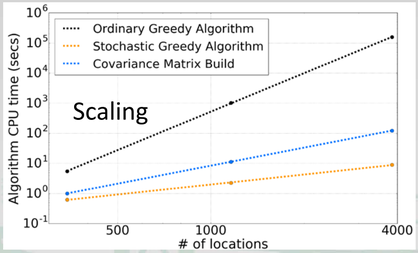


Maximizing the coverage by choosing 'k' placement locations from 'N' possible locations in a 2D GP-based model for the field.

The underlying oracle is the marginal gain in mutual information between un-observed and observed locations.

The oracle complexity is $O(N^3)$

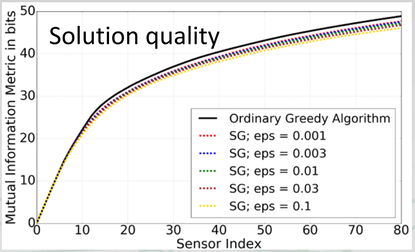
## Informative Locations in Spatial Systems




$O(N^{\alpha})$

| | |
|---|---|
| Ordinary Greedy: | $\alpha = 4.25$ |
| Stochastic Greedy: | $\alpha = 1.11$ |
| Covariance Build: | $\alpha = 1.99$ |

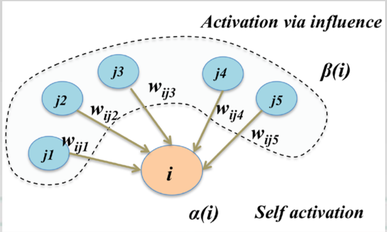
## Key Entities in Networked Systems




Parameter choices in Stochastic-Greedy allows polling a constant but sufficiently large number of random candidates during each iteration

Renders the algorithm linear in the number of locations but with a large constant

Currently exploring two strategies for multi-stage optimization to circumvent the $O(N^2)$ dependency of the pre-processing step




Our contribution extends the Influence Maximization algorithm to include intrinsic nodal activation

This extension allows for realistic modeling of information spread in complex networked systems

A. V. Sathanur, M. Halappanavar, Y. Shi and Y. Sagduyu, "Exploring the Role of Intrinsic Nodal Activation on the Spread of Influence in Complex Networks," Accepted, *Springer Lecture Notes on Social Network Analysis*

## Ongoing/Future Work

1. Multi-stage methods to further scale submodular optimization
   a. Active Set Pre-Selection
   b. Distributed strategy based on dataset partitioning
2. Systematic comparison with traditional MINLP solvers
3. Scalable methods for subset selection in control applications for power grids
4. Scalable algorithms for Influence Maximization
5. Incorporation of uncertainty and dynamics for real-world control applications (i.e., robust and online methods)