Scaling Agent-based Simulation of Contagion Diffusion over Dynamic Networks on Petascale Machines

Keith Bisset

Jae-Seung Yeom, Ashwin Aji

kbisset@vbi.vt.edu

ndssl.vbi.vt.edu

Network Dynamics and Simulation Science Lab

Virginia Bioinformatics Institute

Virginia Tech

Contagion Diffusion

- The problem we are trying to solve
  - Contagion propagation across large interaction networks
  - ~300 million nodes, ~1.5/70 billion edges
- Examples
  - Infectious Disease
  - Norms and Fads (e.g., Smoking, Obesity)
  - Digital viruses (e.g., computer viruses, cell phone worms)
  - Human immune system modeling

EpiSimdemics

- Episimdemics is an individual-based modeling environment
  - Each individual is represented based on synthetic population of US
  - Each interaction between two co-located individuals is represented
- Uses a people-location bipartite graph as the underlying network.
- Planned: Add people-people graph for direct interactions
- Features
  - Time dependent and location dependent interactions
  - A scripting language to specify complex interventions
  - PTTS representation of disease and behavior

Example Person-Person Graph

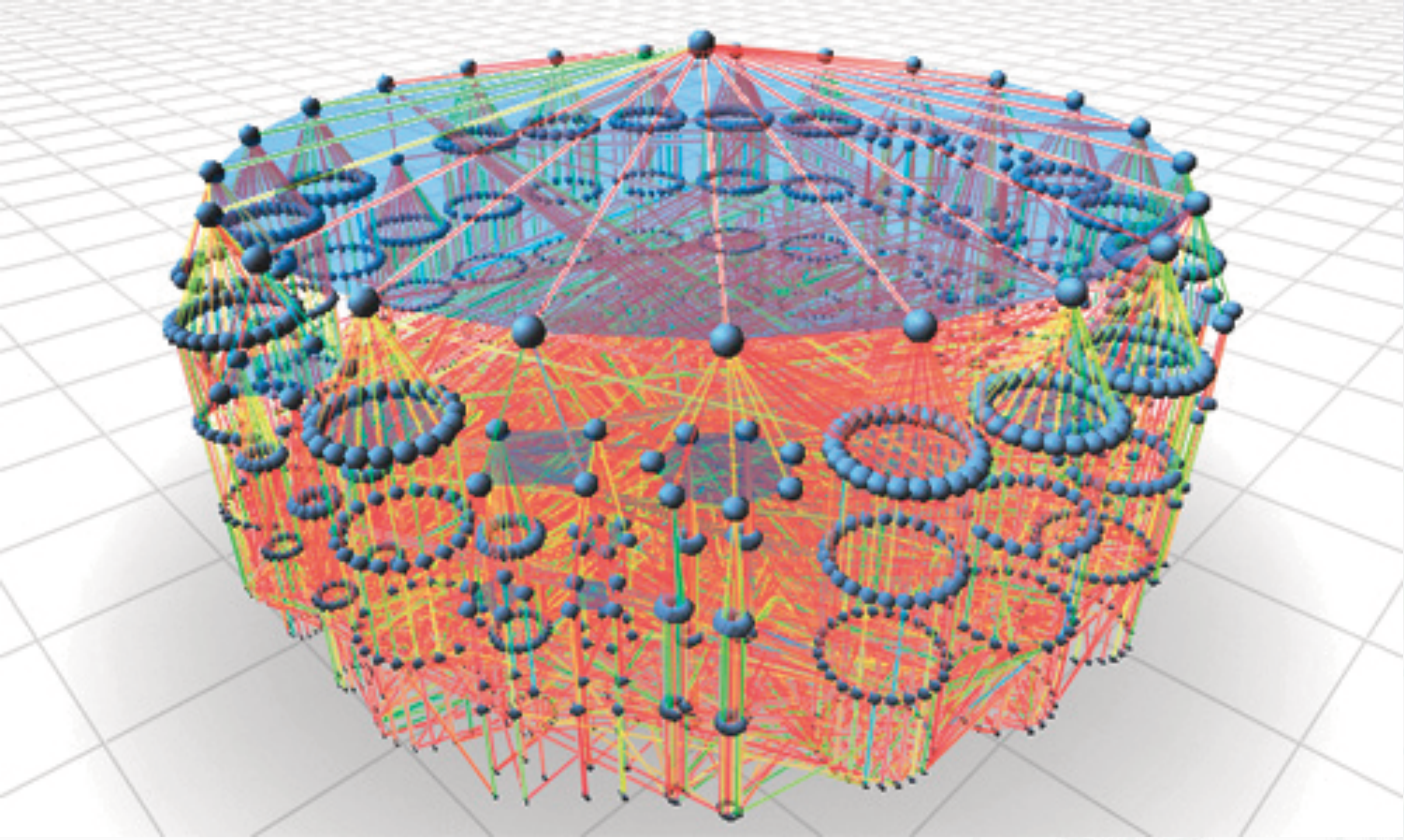
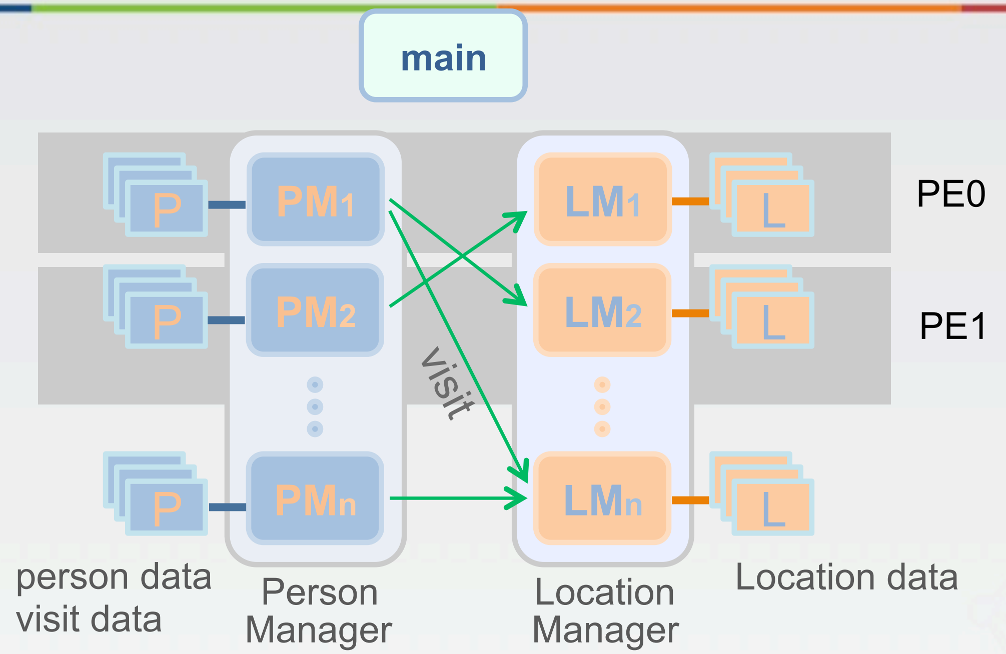


Image courtesy SDSC

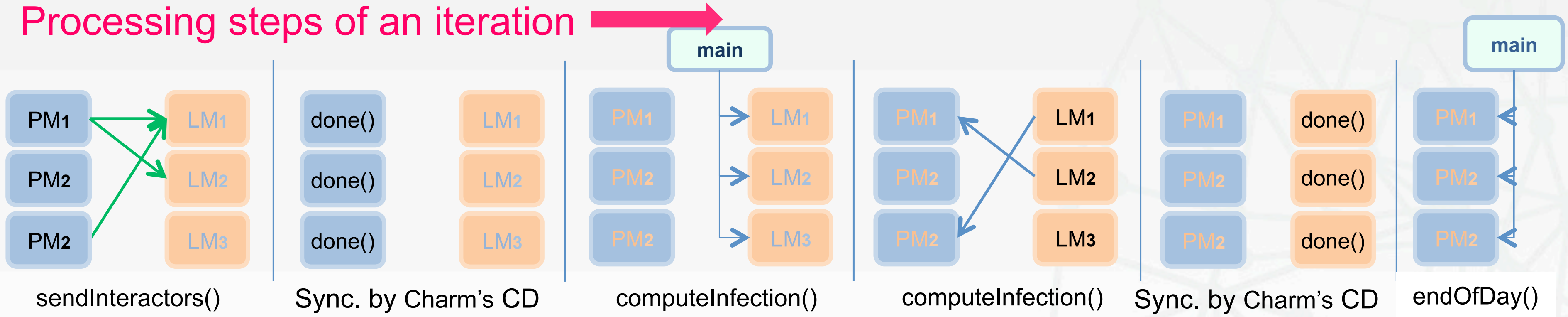
EpiSimdemics Algorithm

- For each timestep (e.g., a day)
  - In parallel, each person
    - Determines where they will go
    - Send a message to each location they will visit
  - In parallel, each location
    - Converts each message into an event pair
    - Calculates probability of infection between each co-located pair of infectious, susceptible people
    - Sends message to each newly infected person
  - In parallel, each infected person updates state
  - Global simulation state is updated

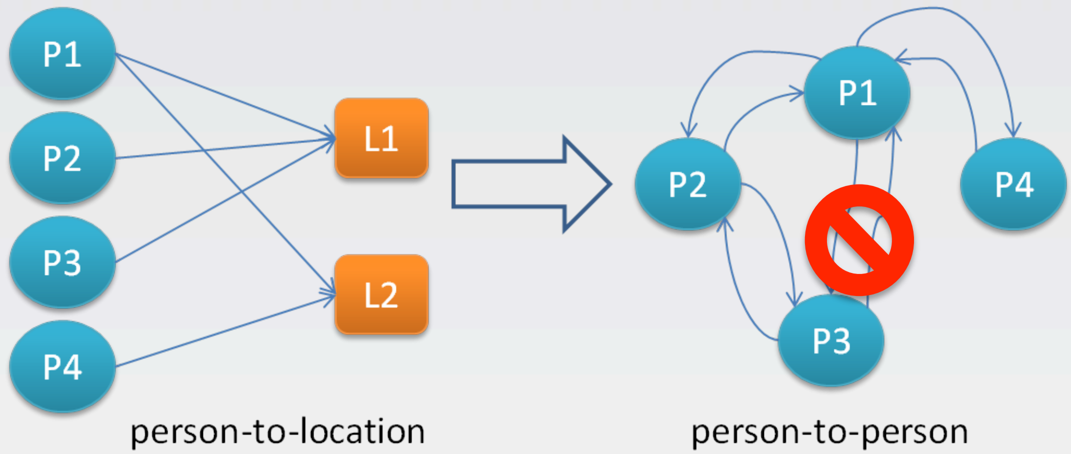
Charm Implementation



Processing steps of an iteration



Data organization



- P-L graph explicit, defines communication
- P-P graph implicit, defines computation, 50x more edges
- Both graphs evolve over time
- US Population
  - 270 million people, 70 million locations
  - 1.5 billion edges P-L graph
  - ~75 billion edges P-P graph (potential interactions/step)

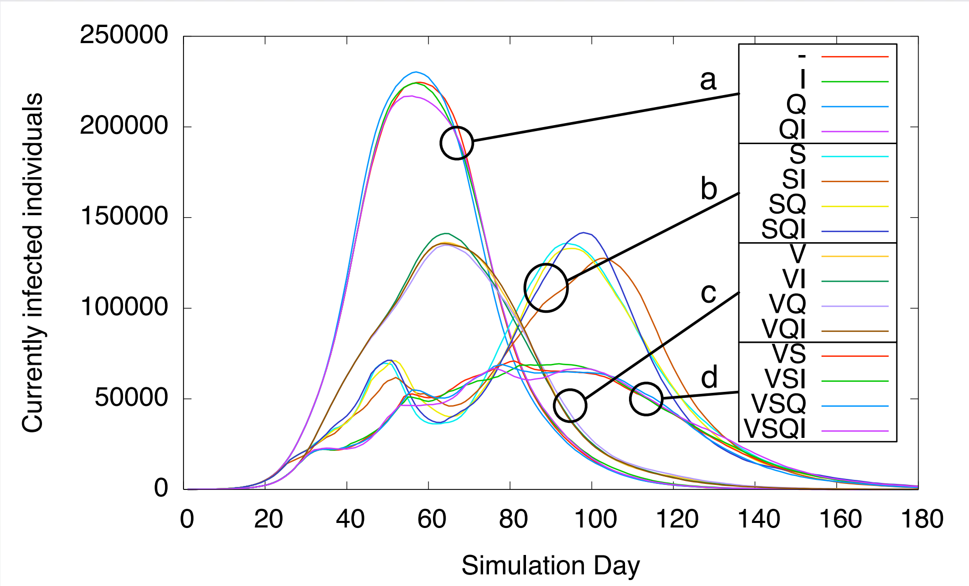
Complex, Layered Interventions

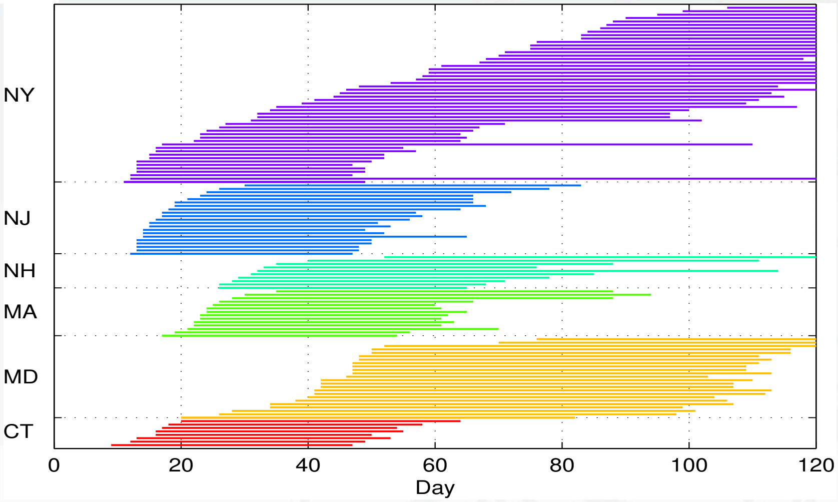
Intervention	Population	Compliance	When
Vaccination	Adult Children Crit Workers	25% 60% 100%	Day 0
School	Closure Reopen	60%	1.0% Children diagnosed (by county)
Quarantine	Crit Workers	100%	1.0% adults diagnosed
Self Isolate	All	20%	2.5% adults diagnosed

```
# stay home when symptomatic.
intervention symptomatic
  set num_symptomatic++
  apply diagnose with prob=0.60
  schedule stayhome 3

trigger disease.symptom >= 2
        apply symptomatic
```

```
# vaccinate 25% of adults
intervention vaccinate_adult
        treat vaccine
        set num_vac_adult++

trigger person.age > 18
  apply vaccinate_adult with prob=0.25
```

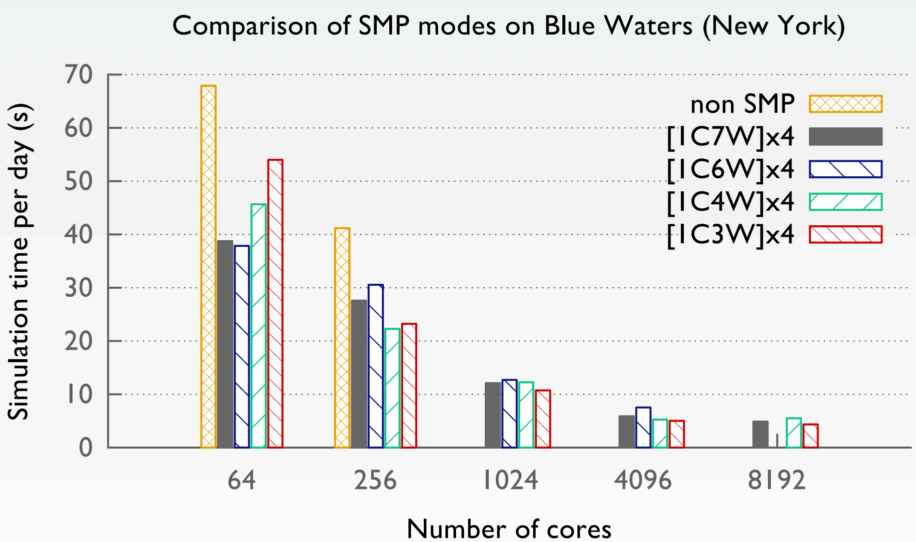
Effects of Interventions



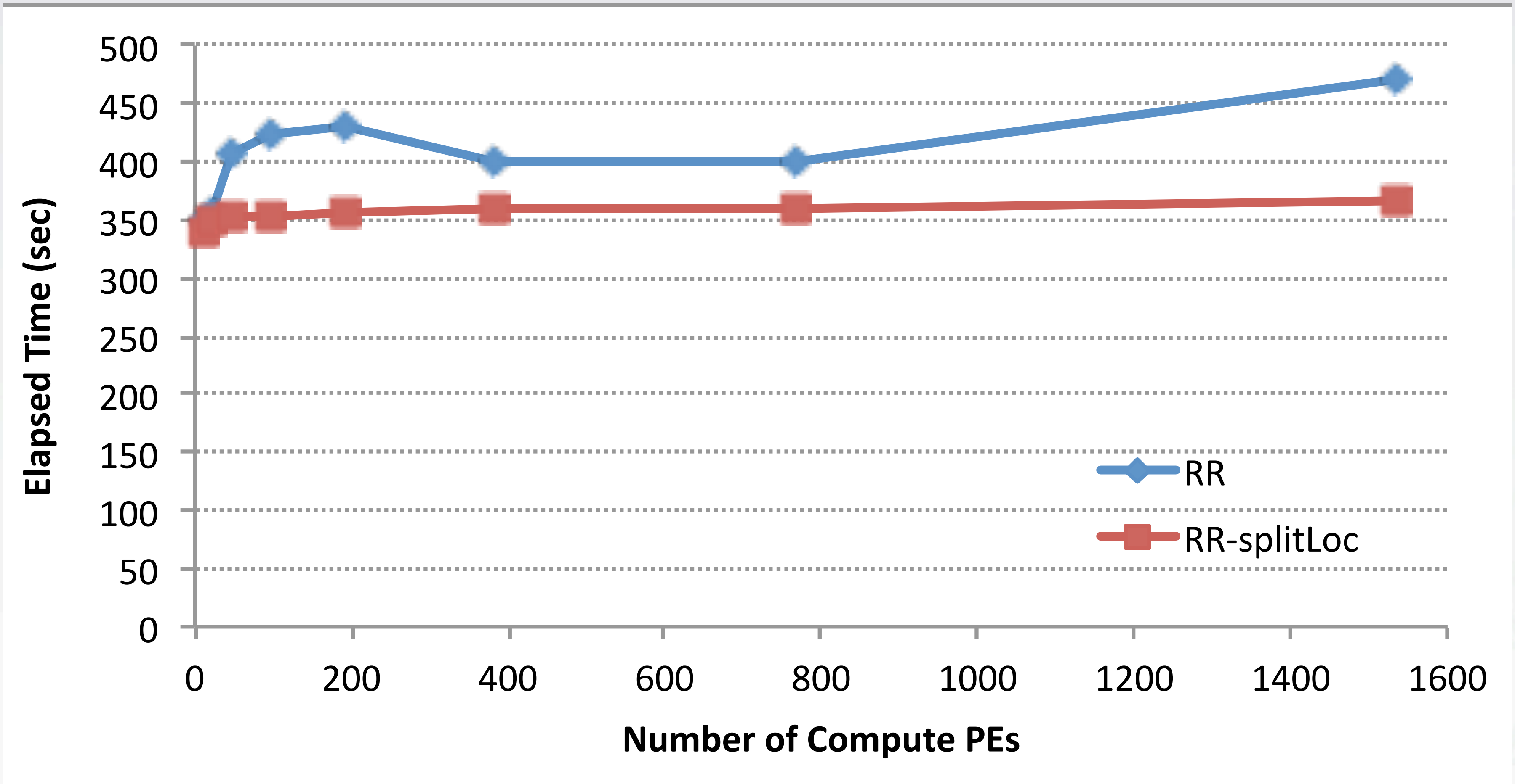


BlueWaters Setup

- Charm++ SMP mode
- Gemini network layer
- 4 processes/node
- 3 compute 1 comm threads per process
- Application based message coalescence

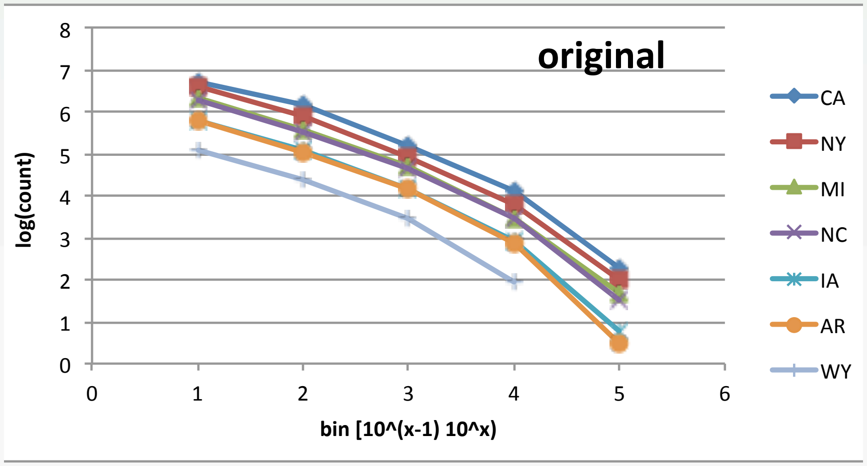


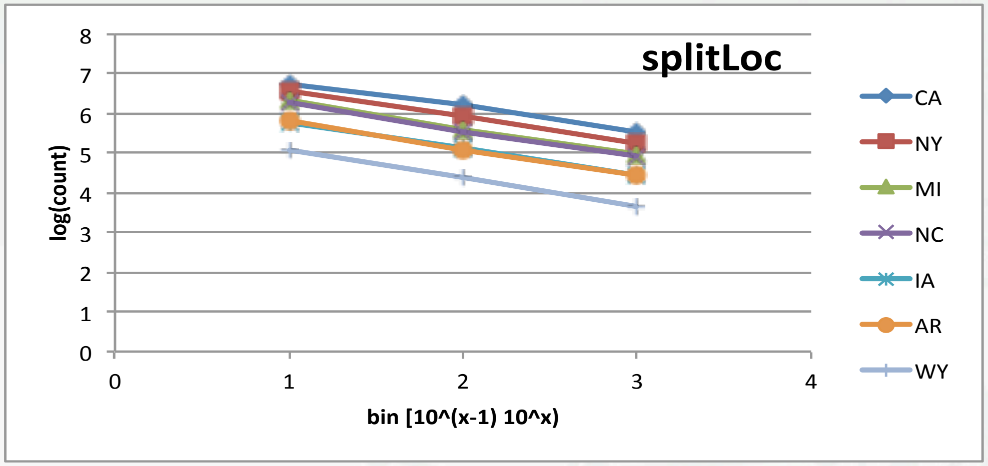
Weak Scaling



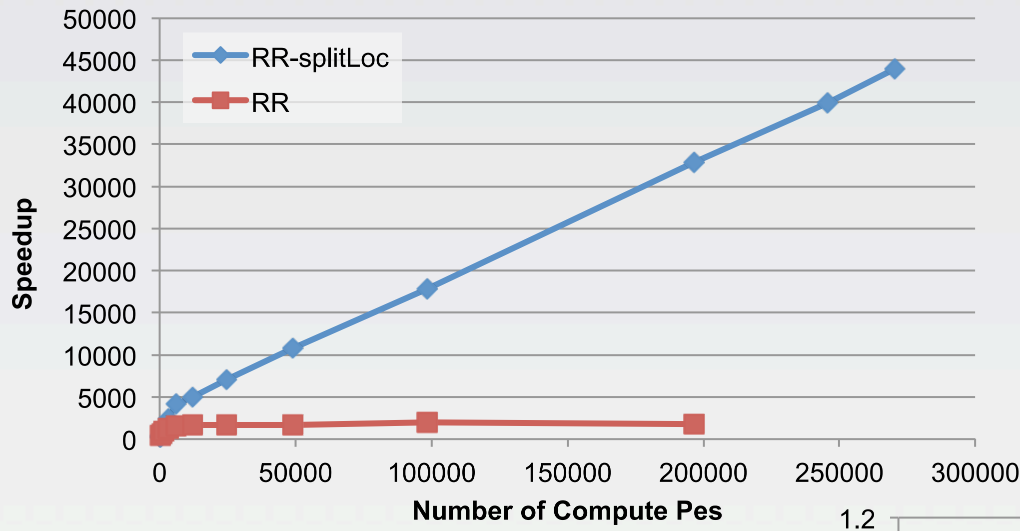
Location Granularity

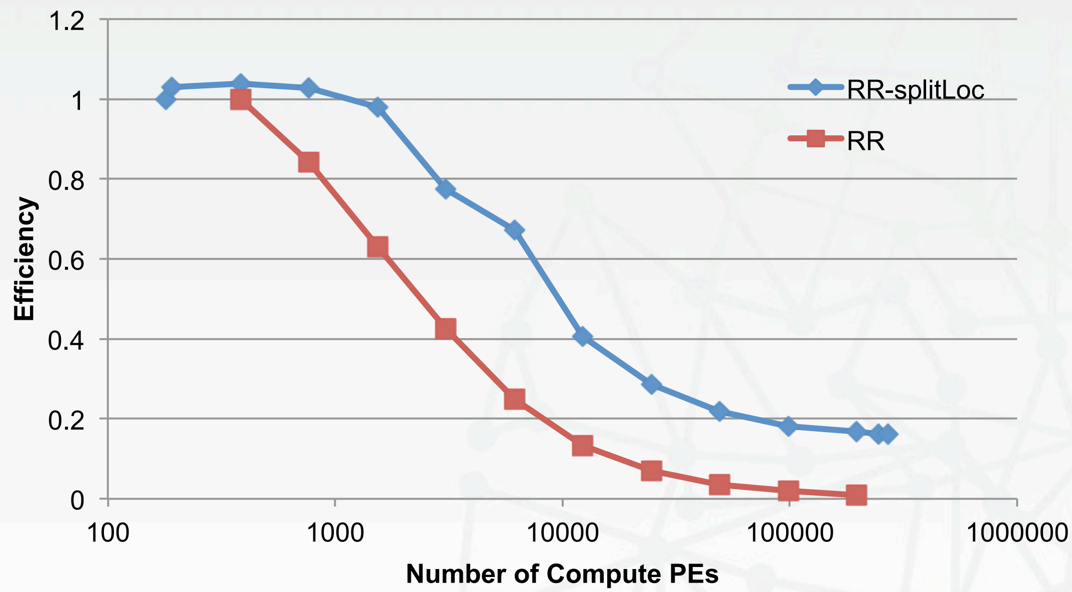
- Location load depends on number of visits
- Location size follows power law
- Not apparent until running at scale





Scaling for US Population



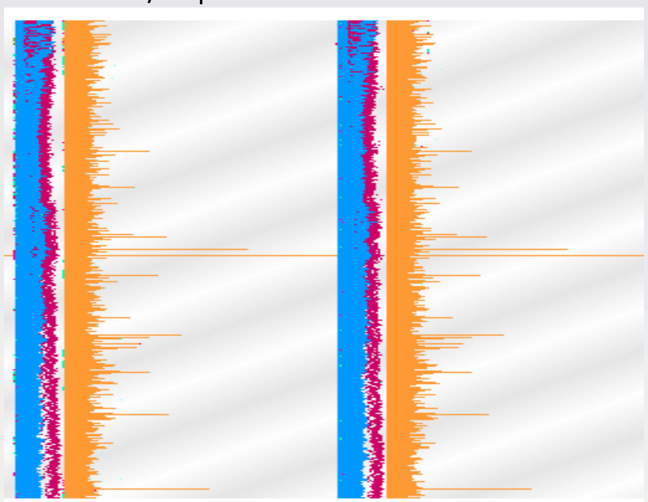


Static Partitioning

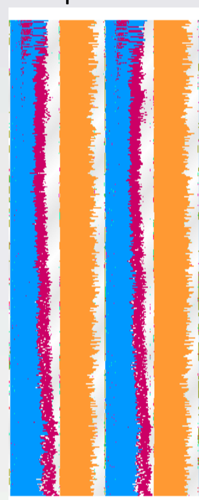
- Round Robin
  - Random distribution
  - Low overhead
  - Works well for small geographic areas (metro area)
- Graph Partitioner
  - Metis based partitioning
  - Multi-constraint (two phases separated by sync)
  - Higher Overhead
  - Helps as geographic area increases (state, national)

Static Partitioning - Results

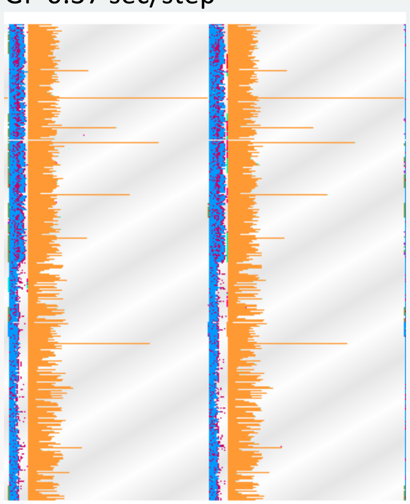
RR 0.42 sec/step



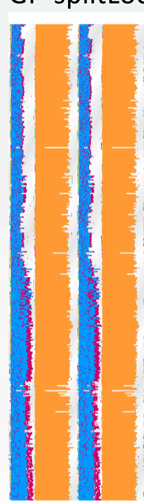
RR-splitLoc 0.12 sec/step



GP 0.37 sec/step



GP-splitLoc 0.09 sec/step



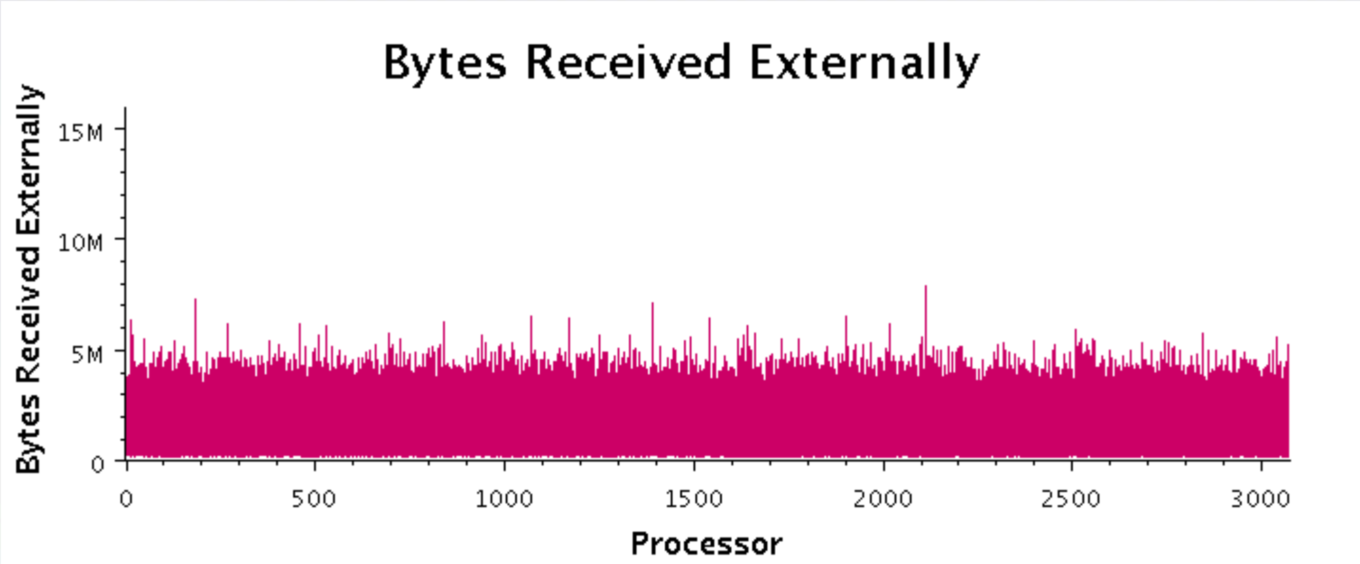
SendInteractor(). Person computation to generate visit messages

AddVisitMessage(). Location side message receive handling.

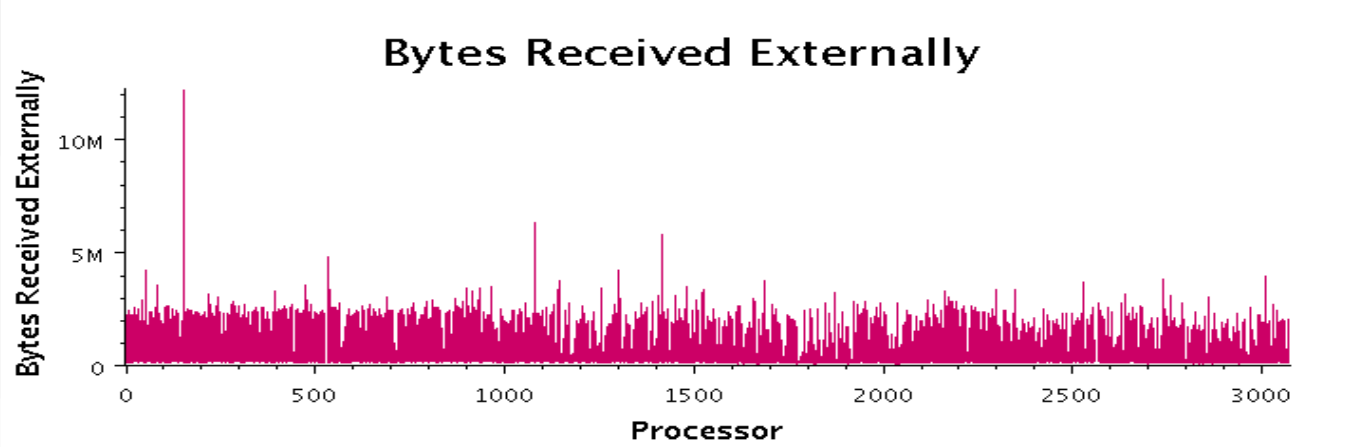
ComputeInfections(). Location computation of interaction among visitors

Message Volume

Round Robin



Graph Partitioner



VirginiaTech
Virginia Bioinformatics Institute

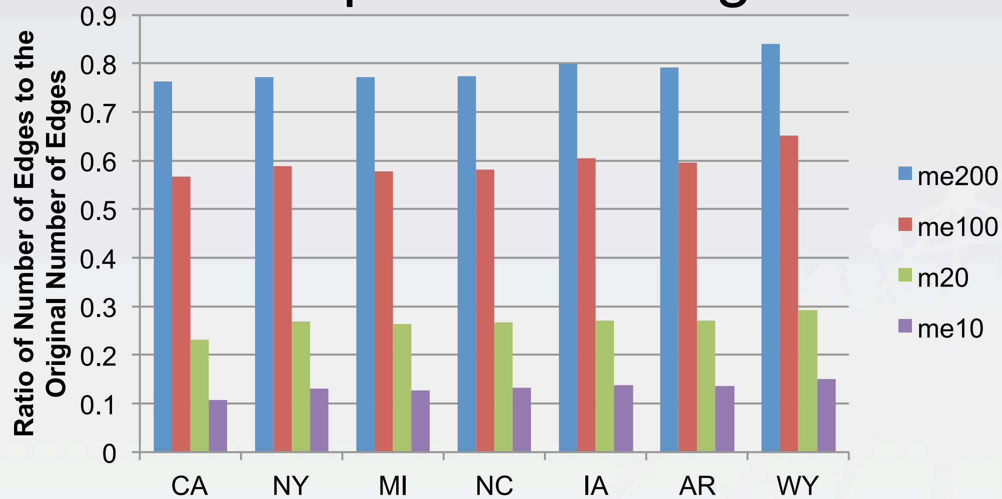
256 nodes, 10 million people

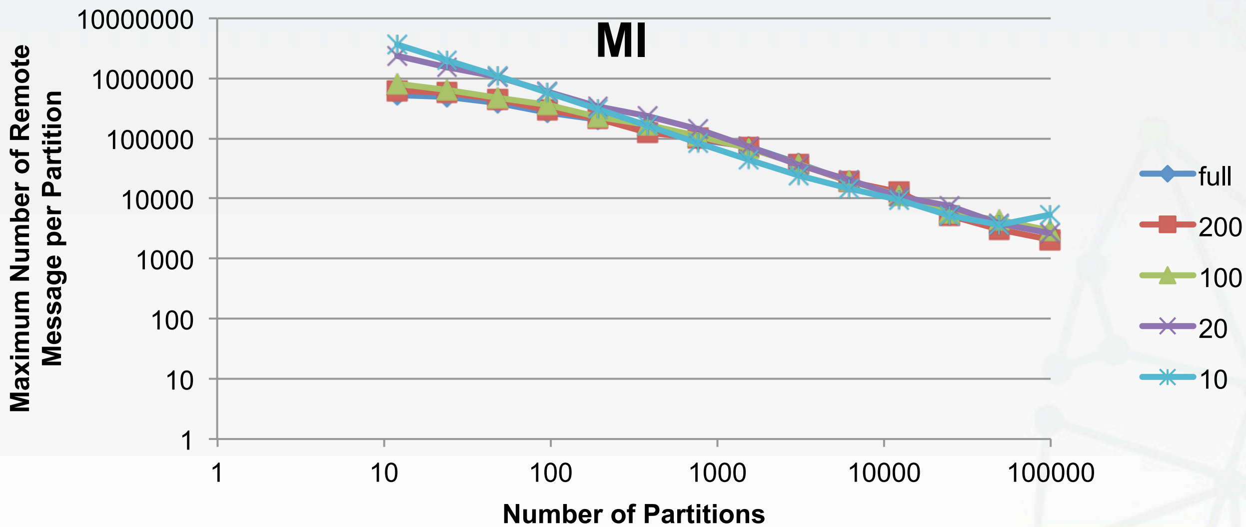
Graph Sparsification

Goal: Improve runtime of Graph Partitioning

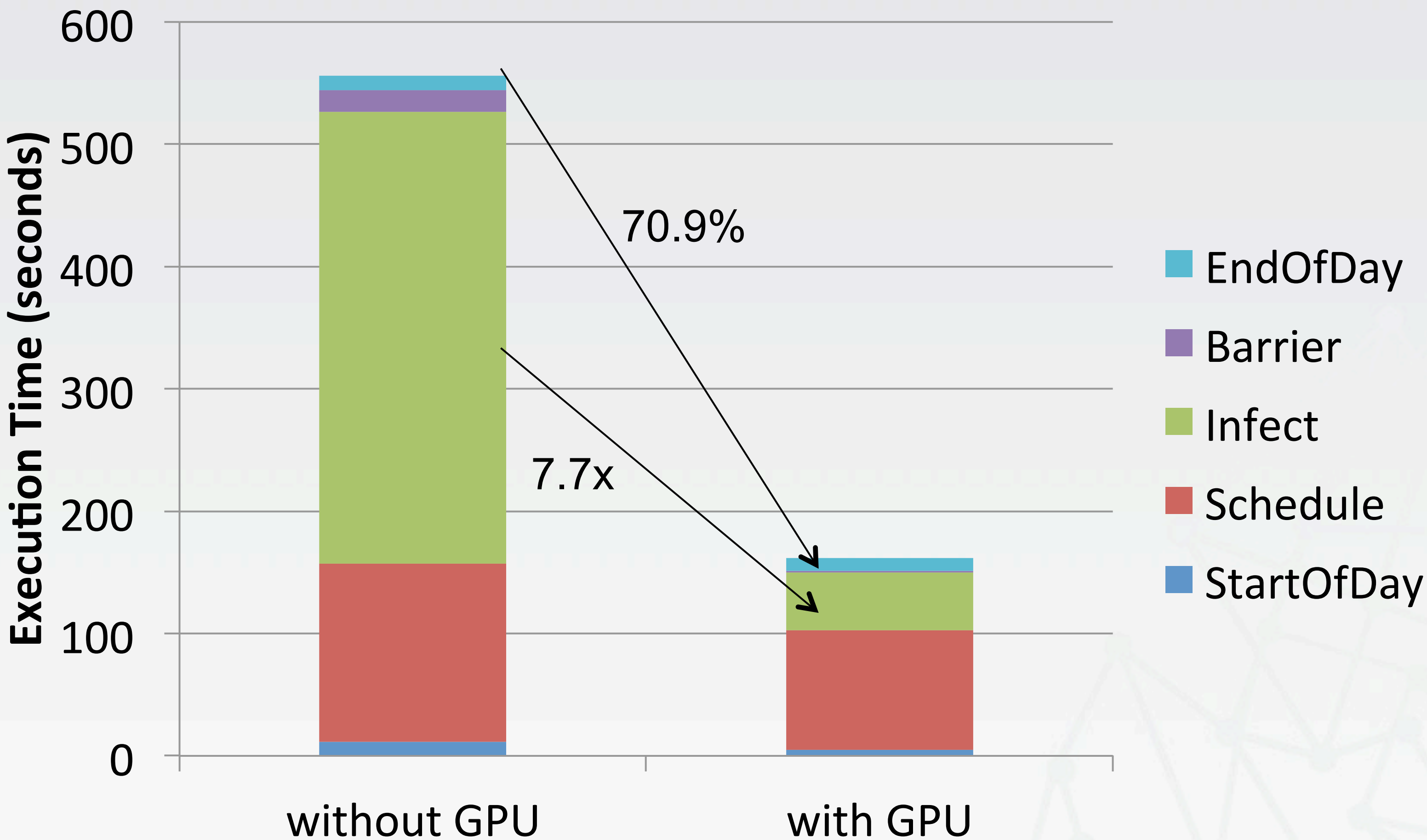
Procedure

- Randomly remove edges from high degree nodes
- Partition sparse graph
- Use full graph for execution



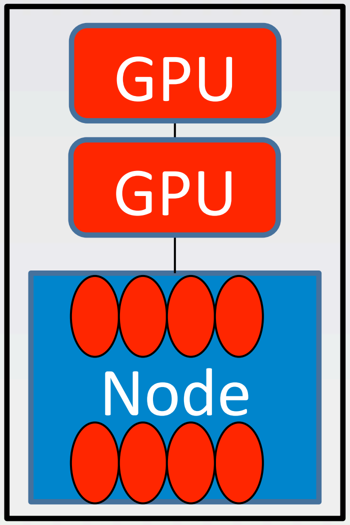


Impact of GPU Acceleration on Execution Profile

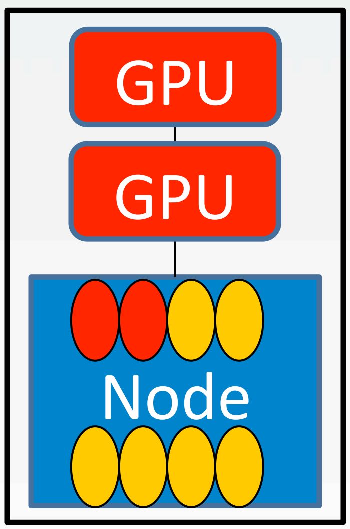


Assume 1CPU cores per GPU devices, in practice, CPU > GPU

GPU-CharmSimdemics Scenarios



- Scenario 1 – All chares from all CPU processes offload simultaneously to GPU
  - GPUs (Kepler) maintain tasks queue from different processes
  - Inefficient: CPUs will be idle waiting for GPU execution to complete



- Scenario 2 – Chares from only some select CPU processes offload to GPU
  - 1:1 ratio can be maintained between "GPU" processes and GPUs
  - But, "GPU" chares will finish sooner than "CPU" chares, i.e. load imbalance
  - Use LB methods of Charm++ to rebalance chares

Future Work

- Dynamic Load Balancing with semantic information
  - Prediction model based on past runs
  - Information from simulation state variables
  - Use dynamic interventions – more variable load
- Try Charm++ Meta Load Balancer
- Further improvements to initial partitioning
  - Minimize message imbalance as well as edge-cut
- Message reduction
- Sequential replicates to amortize data load time
- Scale to global population - 10 billion people

Acknowledgements

UIUC Parallel Programming Lab

NSF HSD Grant SES-0729441, NSF PetaApps Grant OCI-0904844, NSF NETS Grant CNS-0831633, NSF CAREER Grant CNS-0845700, NSF NetSE Grant CNS-1011769, NSF SDCI Grant OCI-1032677, DTRA R&D Grant HDTRA1-09-1-0017, DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, DOE Grant DE-SC0003957, PSU/DOE 4345-VT-DOE-4261, US Naval Surface Warfare Center Grant N00178-09-D-3017 DEL ORDER 13, NIH MIDAS Grant 2U01GM070694-09, NIH MIDAS Grant 3U01FM070694-09S1, LLNL Fellowship SubB596713 DOI Contract D12PC00337

VIRGINIA BIOINFORMATICS INSTITUTE AT VIRGINIA TECH

VirginiaTech Invent the Future