An Extreme-Scale Computational Approach to Redistricting Optimization

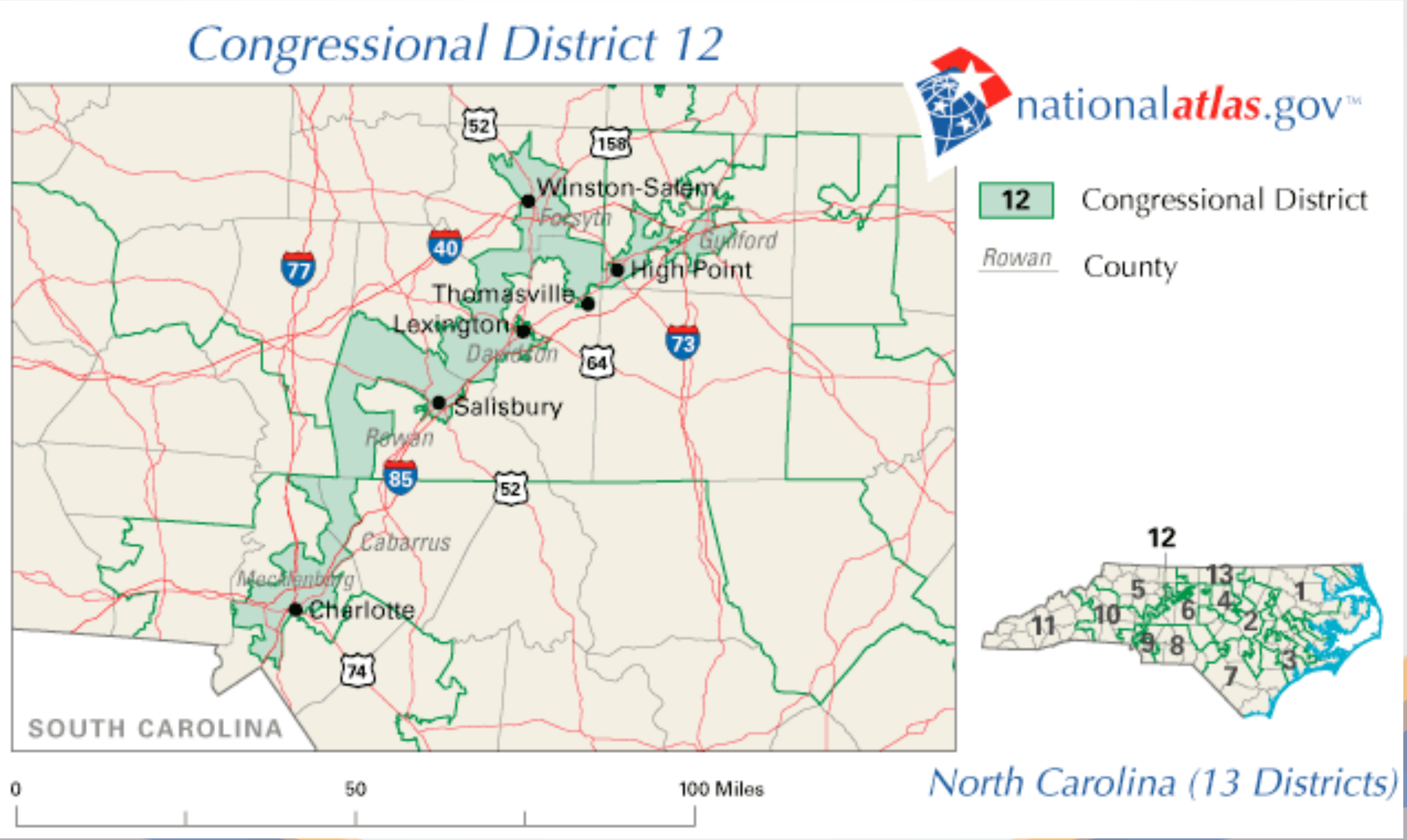
Wendy K. Tam Cho, Yan Liu, and Shaowen Wang (PI)

CyberGIS Center for Advanced Digital and Spatial Studies (CyberGIS Center)
CyberInfrastructure and Geospatial Information Laboratory (CIGI Laboratory)
Department of Geography and Geographic Information Science
Department of Computer Science
Department of Urban and Regional Planning
Graduate School of Library and Information Science
National Center for Supercomputing Applications (NCSA)
University of Illinois at Urbana-Champaign

NCSA Blue Waters Symposium
for Petascale Science and Beyond

May 12-15, 2014

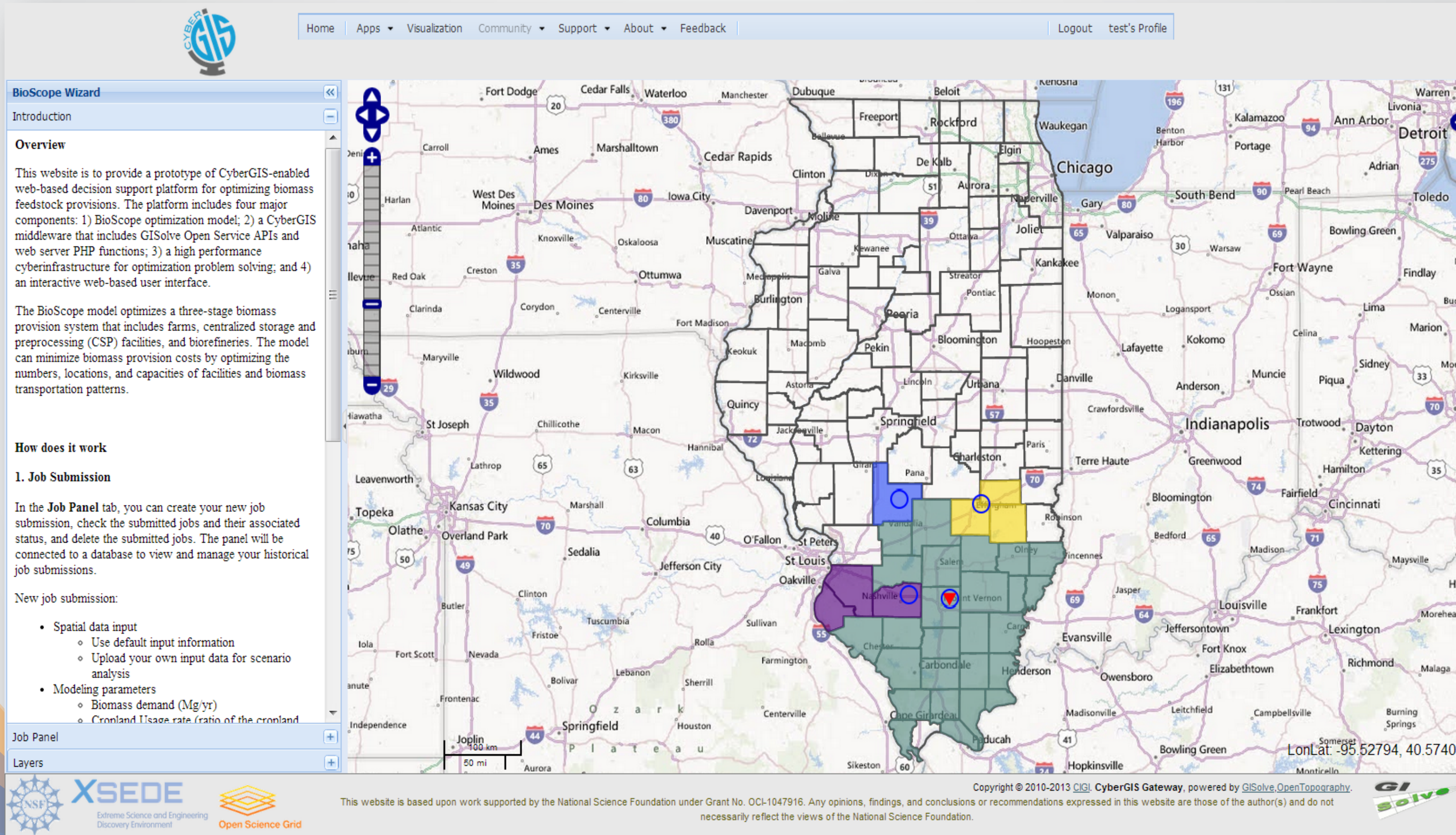
Gerrymandering



Zoning Analytics

- Partitioning a group of indivisible geographic units into a smaller number of districts
  - Objectives and constraints
    - Contiguity, competitiveness, equal population, preservation of communities of interest and local political subdivisions, minority districts
  - Computational complexity
    - Number of possible solutions
      - Stirling number of the second kind: $S(n, k)$
      - Example: $S(55, 6) = 8.7 \times 10^{39}$
    - Computationally intractable
      - *NP*-hard

Complex Spatial Decision Making



Collaborative Work by *Hao Hu, Tao Lin, Yan Liu, Luis F. Rodríguez, and Shaowen Wang*

Exact Algorithms *vs.* Heuristics

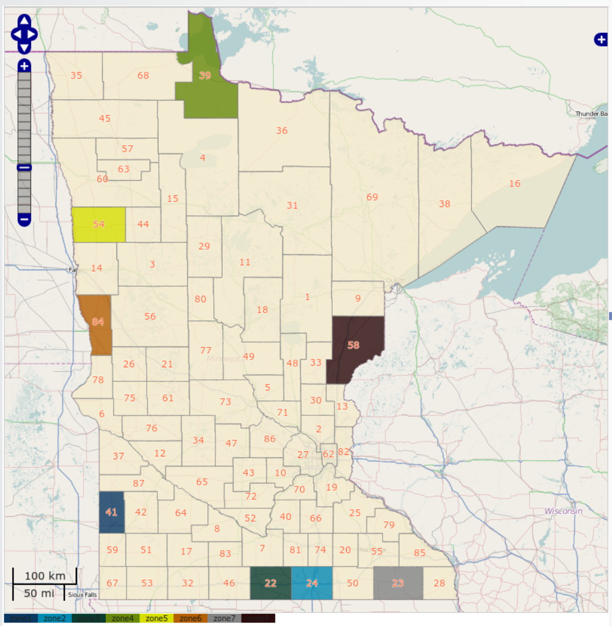
- Exact algorithms
  - Guarantee to find an optimal solution
  - Methods
    - Branch & bound
    - Branch & cut
    - Etc.
  - Computationally intractable

- Heuristics
  - Algorithms that produce optimal or near-optimal solutions within a reasonable amount of time
  - Population-based heuristics
    - **Genetic algorithm**
    - Swarm
    - Ant colony
    - Etc.

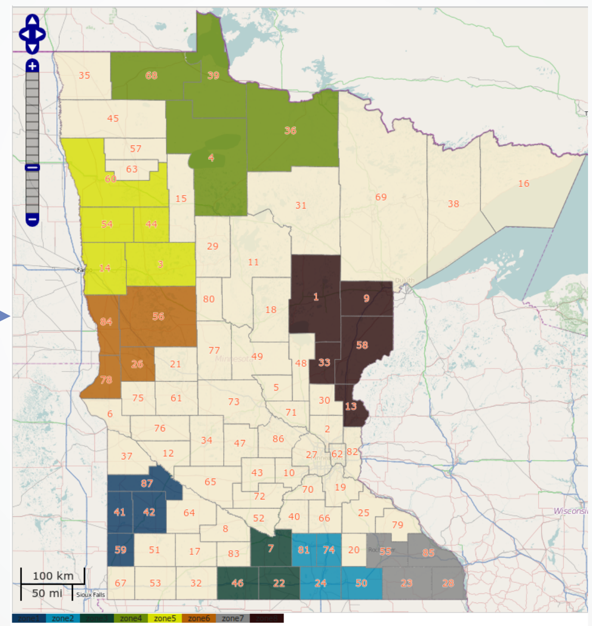
Genetic Algorithm (GA)

- Principles
  - Evolutionary process
    - “survival of the fittest”
    - Iterative algorithm
  - Solution population: a diverse set of initial solutions
  - GA operators
    - Selection, crossover, mutation, replacement
  - Stopping criteria
    - Solution quality
    - Time or the number of iterations
- Spatial GA operators
  - Solution generation
  - Crossover
  - Mutation

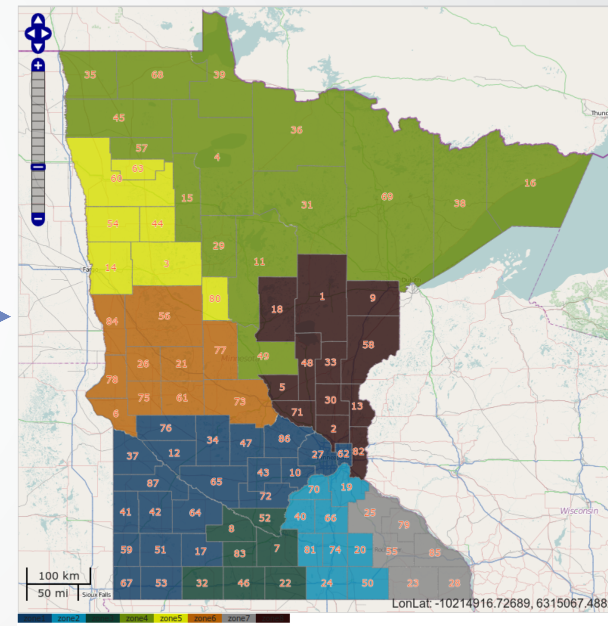
Spatial GA Operators – Feasible Solution Generation



Seeding



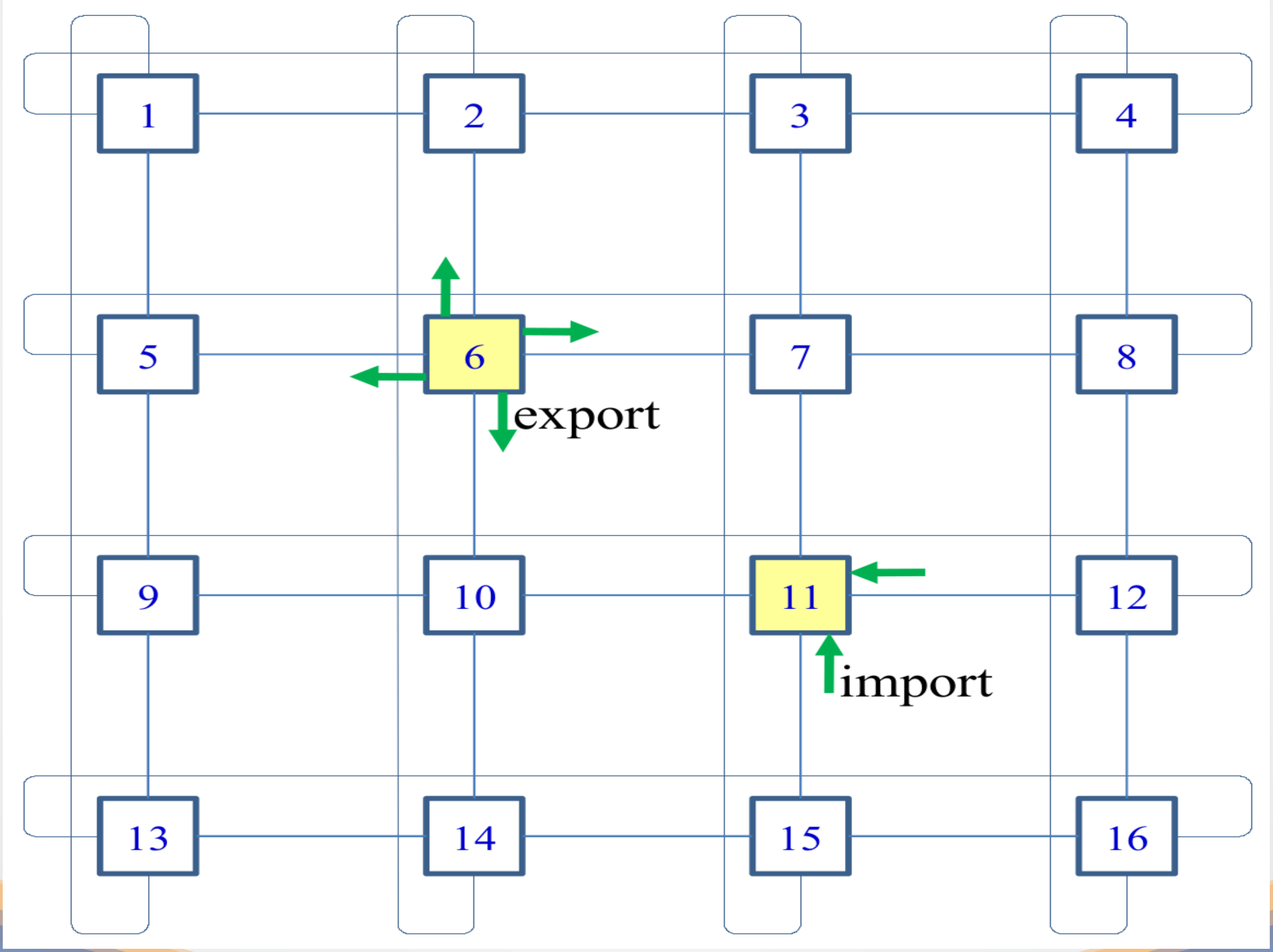
Expansion



Completion

- Liu, Y.Y. and Wang, S. 2014. "A Scalable Parallel Genetic Algorithm for the Generalized Assignment Problem." *Parallel Computing*, http://dx.doi.org/10.1016/j.parco.2014.04.008

Parallel GA

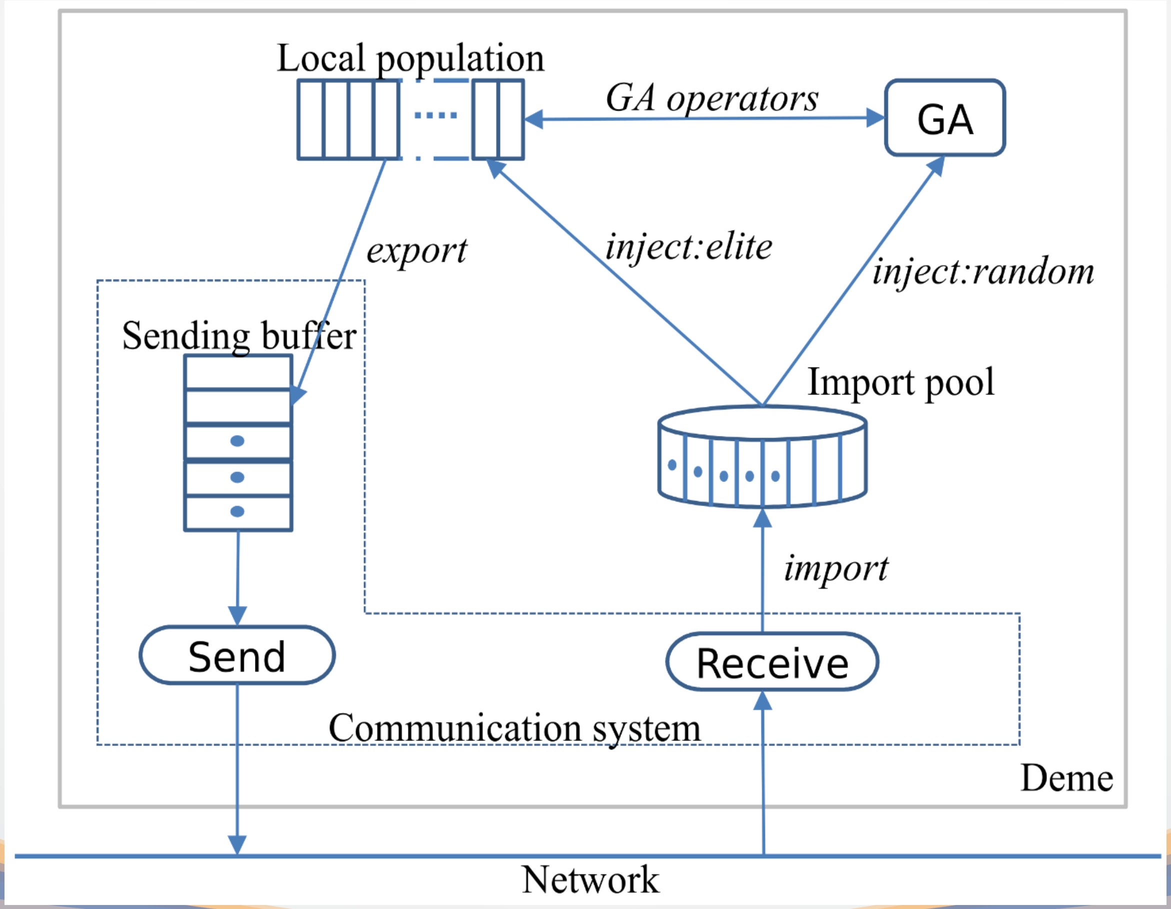


Challenges for PGA

- Scalability to a large number of processor cores
  - Solution migration
    - Migration interval, rate
    - Traditional implementation: global barrier
  - Communication cost increases significantly when using a large number of cores
- Synchronous *vs*. Asynchronous
  - Goals of asynchronous migration
    - Mimic the natural behavior of GA
    - Increase the overlapping of computing and communication
  - Breaking the global barrier
    - Asynchronous migration
    - Buffer handling is inevitable

Asynchronous Migration

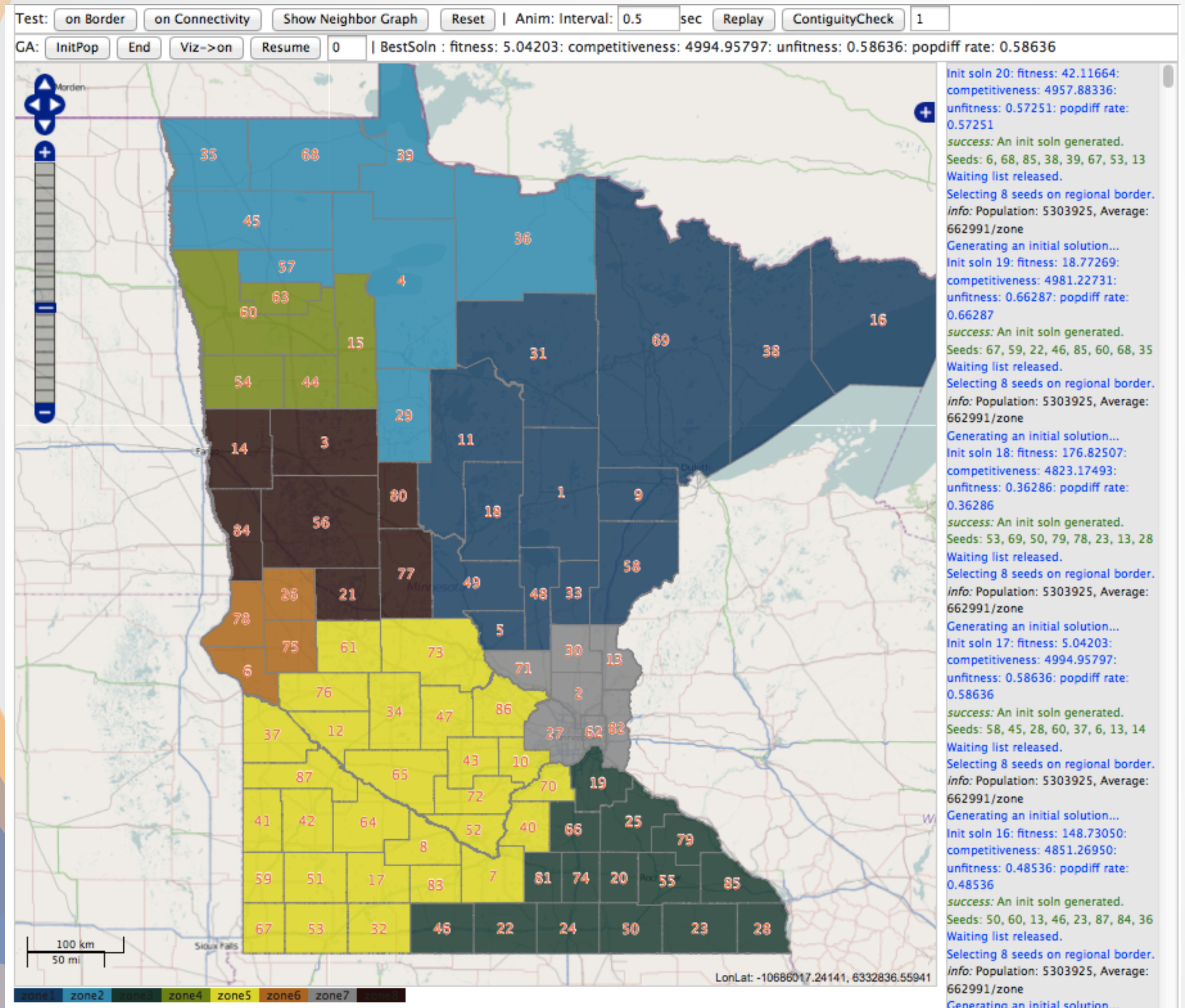
- Operators
  - export, import, inject

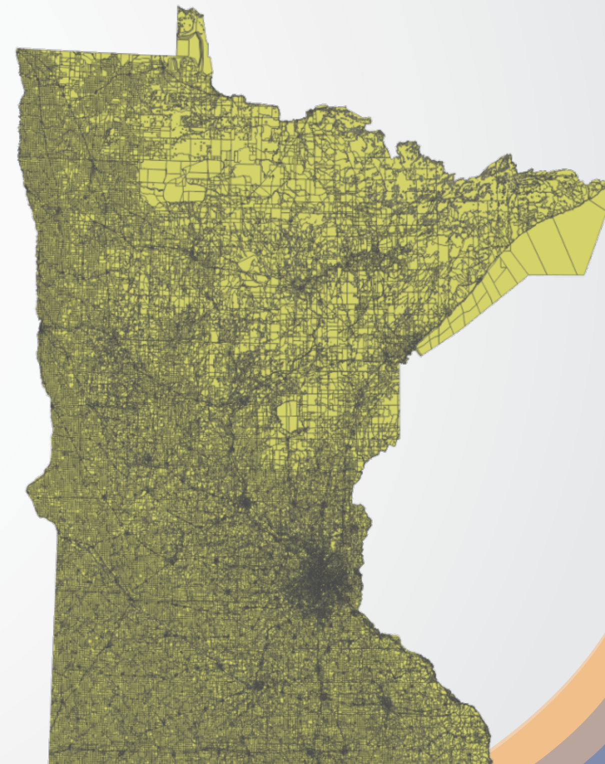


PGA Parameters

Parameters	Settings
Population size per deme	100
Initial population generation	Random with feasibility improvement or constraint-based improvement
Selection	Binary tournament
Crossover	1-point. Probability: 0.8
Mutation	1-item mutation. Probability: 0.2
Replacement	Replacing the unfittest or worst
Elitism	Yes
Stopping rules	No solution improvement, bounded solution quality reached, or fixed number of iterations
Connectivity d	4
Migration rate r	2
Export interval M_{expt}	100
Import interval M_{impt}	50
Probability of holding	1/20 (the probability to export when no better solution found during a previous export interval)
Sending buffer size $K_{sendbuf}$	20 solutions. Actual memory requirement is $(20 \times n \times 4 + buffer\_overhead)$ bytes
Import pool size K_{impt}	80 solutions. Actual memory requirement is $(80 \times n \times 4)$

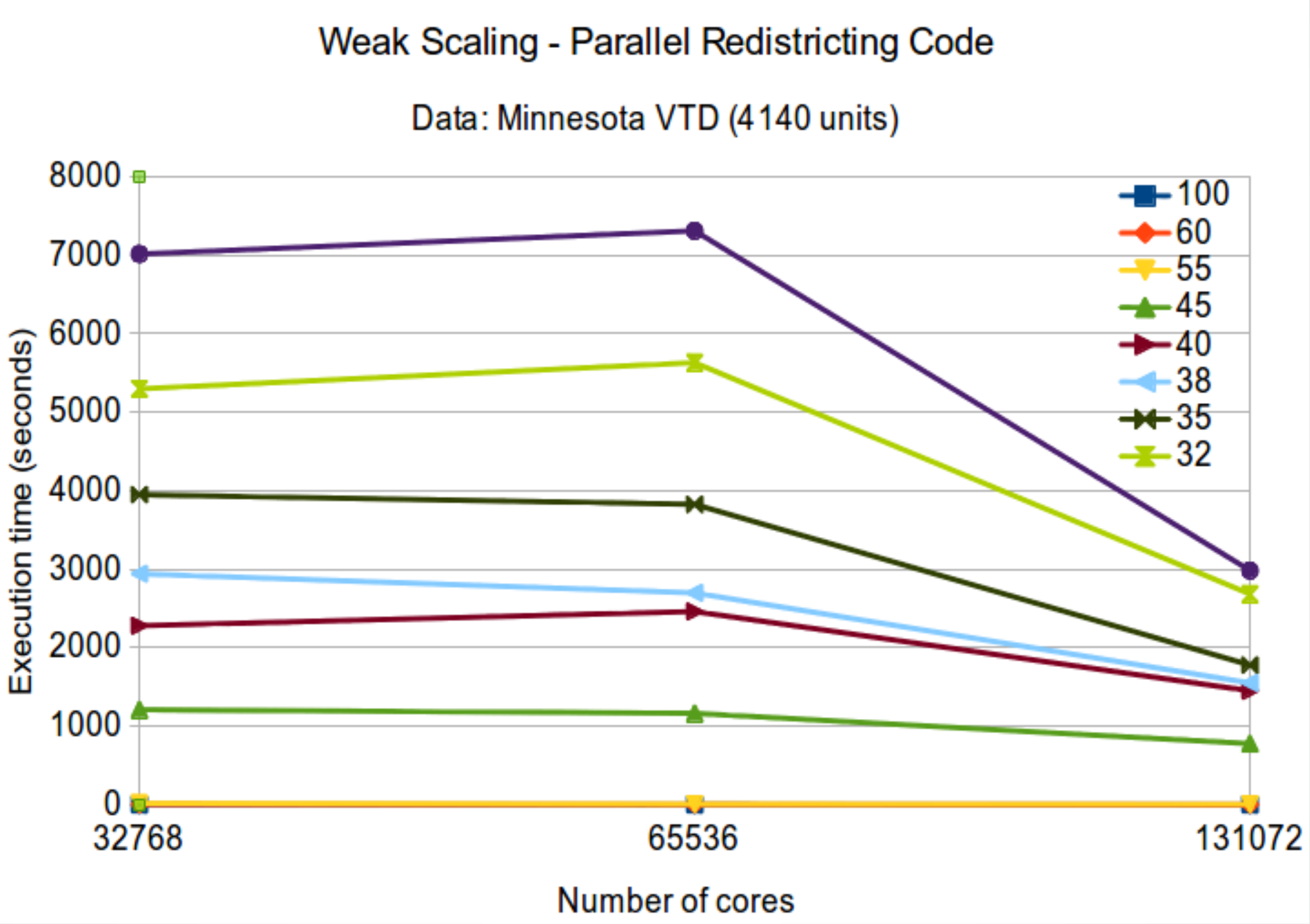
Case Study


Test: on Border | on Connectivity | Show Neighbor Graph | Reset | Anim: Interval: 0.5 sec | Replay | ContiguityCheck | 1
GA: InitPop | End | Viz->on | Resume | 0 | BestSoln : fitness: 5.04203: competitiveness: 4994.95797: unfitness: 0.58636: popdiff rate: 0.58636
Init soln 20: fitness: 42.11664: competitiveness: 4957.88336: unfitness: 0.57251: popdiff rate: 0.57251
success: An init soln generated.
Seeds: 6, 68, 85, 38, 39, 67, 53, 13
Waiting list released.
Selecting 8 seeds on regional border.
info: Population: 5303925, Average: 662991/zone
Generating an initial solution...
Init soln 19: fitness: 18.77269: competitiveness: 4981.22731: unfitness: 0.66287: popdiff rate: 0.66287
success: An init soln generated.
Seeds: 67, 59, 22, 46, 85, 60, 68, 35
Waiting list released.
Selecting 8 seeds on regional border.
info: Population: 5303925, Average: 662991/zone
Generating an initial solution...
Init soln 18: fitness: 176.82507: competitiveness: 4823.17493: unfitness: 0.36286: popdiff rate: 0.36286
success: An init soln generated.
Seeds: 53, 69, 50, 79, 78, 23, 13, 28
Waiting list released.
Selecting 8 seeds on regional border.
info: Population: 5303925, Average: 662991/zone
Generating an initial solution...
Init soln 17: fitness: 5.04203: competitiveness: 4994.95797: unfitness: 0.58636: popdiff rate: 0.58636
success: An init soln generated.
Seeds: 58, 45, 28, 60, 37, 6, 13, 14
Waiting list released.
Selecting 8 seeds on regional border.
info: Population: 5303925, Average: 662991/zone
Generating an initial solution...
Init soln 16: fitness: 148.73050: competitiveness: 4851.26950: unfitness: 0.48536: popdiff rate: 0.48536
success: An init soln generated.
Seeds: 50, 60, 13, 46, 23, 87, 84, 36
Waiting list released.
Selecting 8 seeds on regional border.
info: Population: 5303925, Average: 662991/zone
100 km
50 mi
LonLat: -10686017.24141, 6332836.55941
zone1 zone2 zone3 zone4 zone5 zone6 zone7 zone8



Computational Experiments

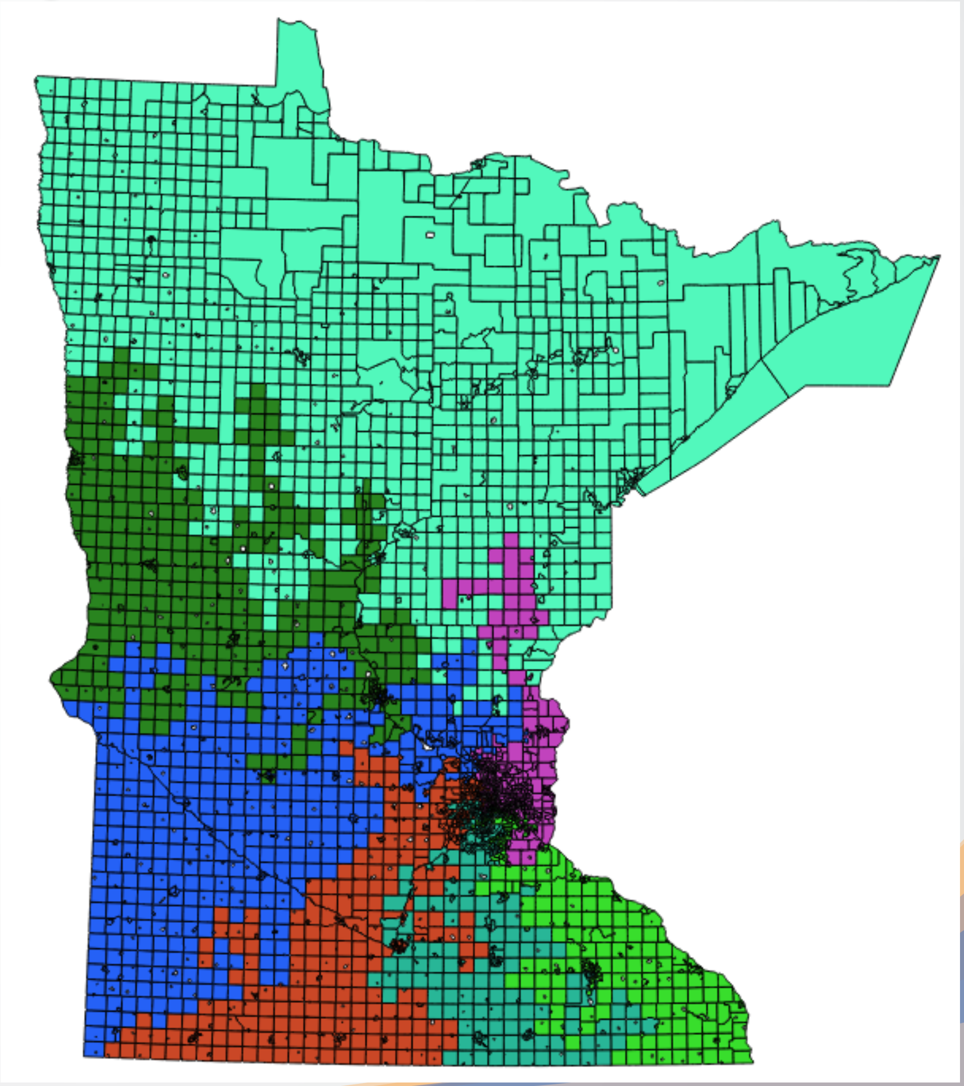
- Problem size
  - 4140 voting districts
  - Number of districts: 8
- Data
  - Input: shape file; rook and queen neighborhood files
  - Output: shape file; each core outputs one
- Number of cores
  - 32768, 65536, 131072



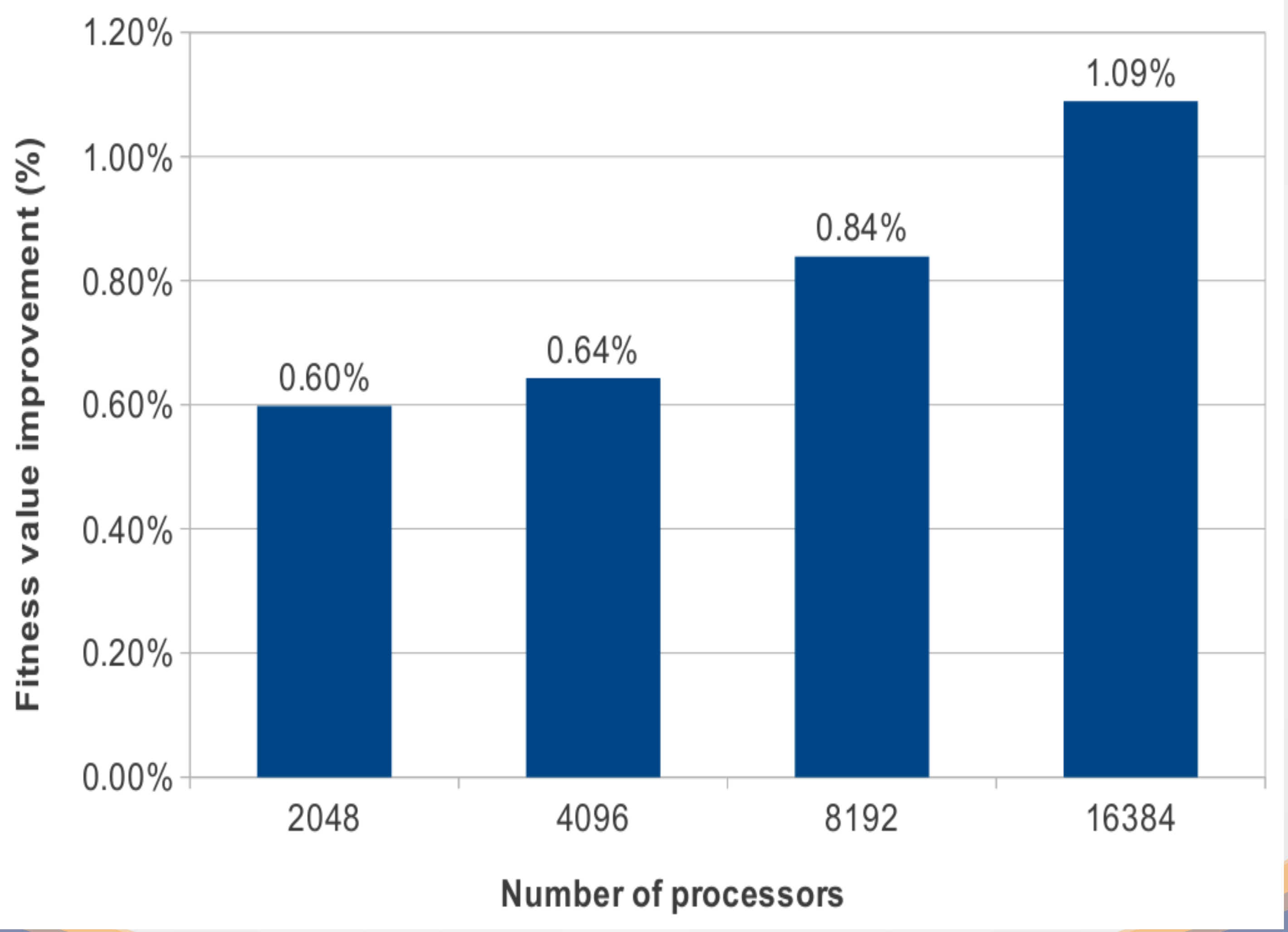
- Measurement: time taken to achieve different solution bounds (the smaller, the better)
- Using 131072 cores led to significant improvement

An Example Solution

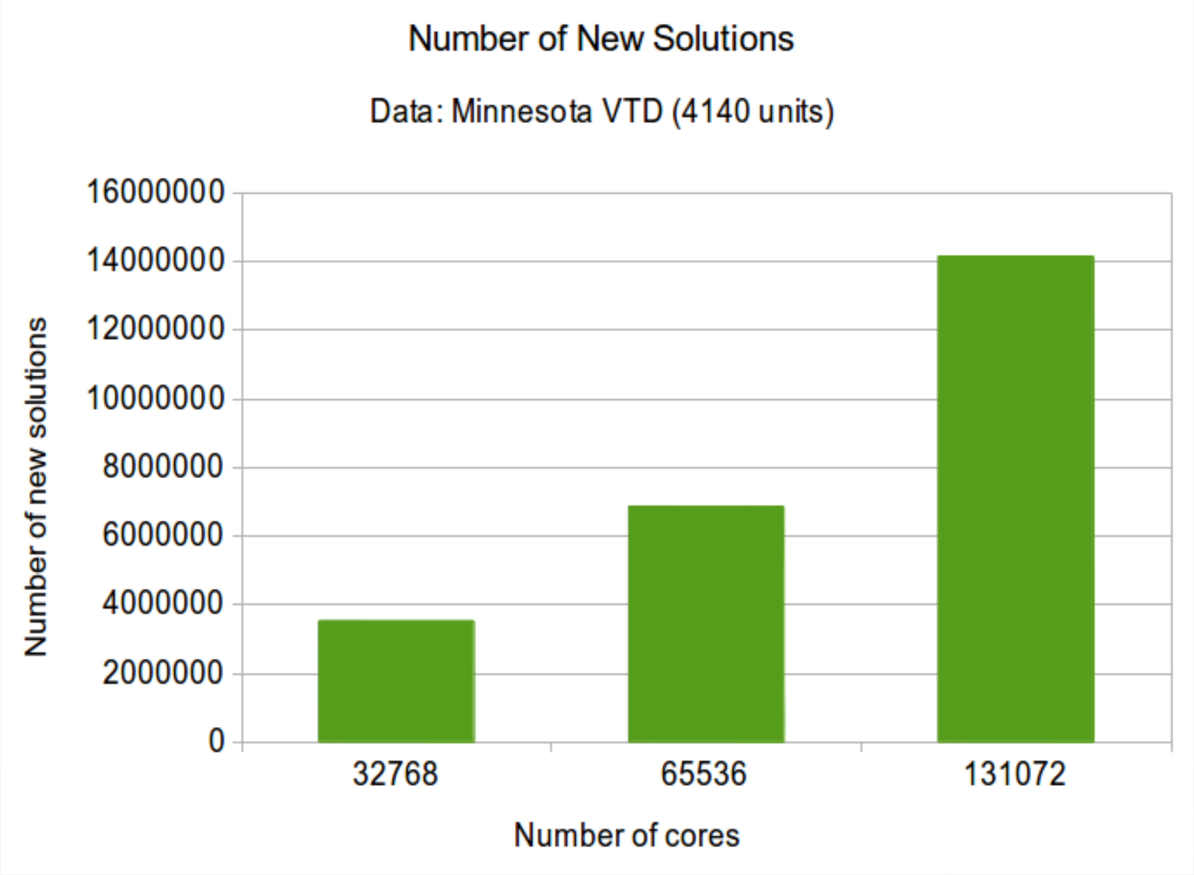
- Found by the run using 131072 cores
- Competitiveness as objective
  - R / (R + D)
  - 0.499998
- Constraint: population deviation



Solution Quality Improvement



Number of New Solutions



What is GIS?

- Systems
- Science
- Services
- Society
- Synthesis
- Geo and spatial are special

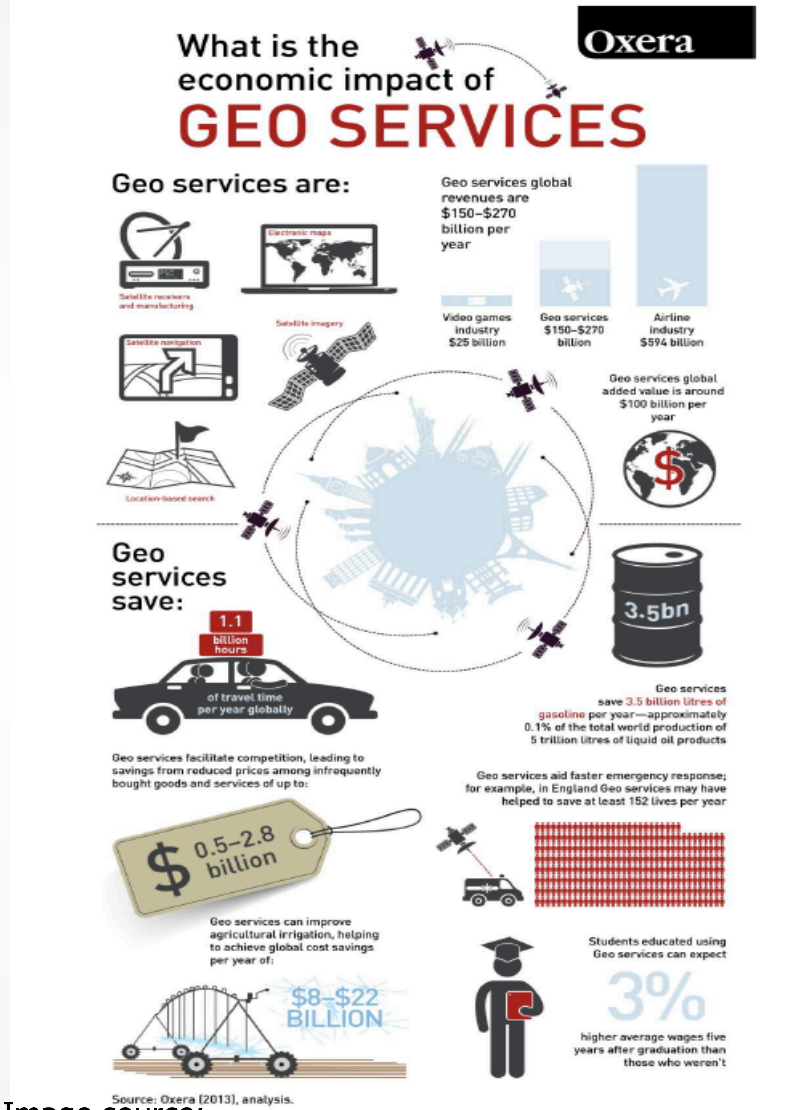
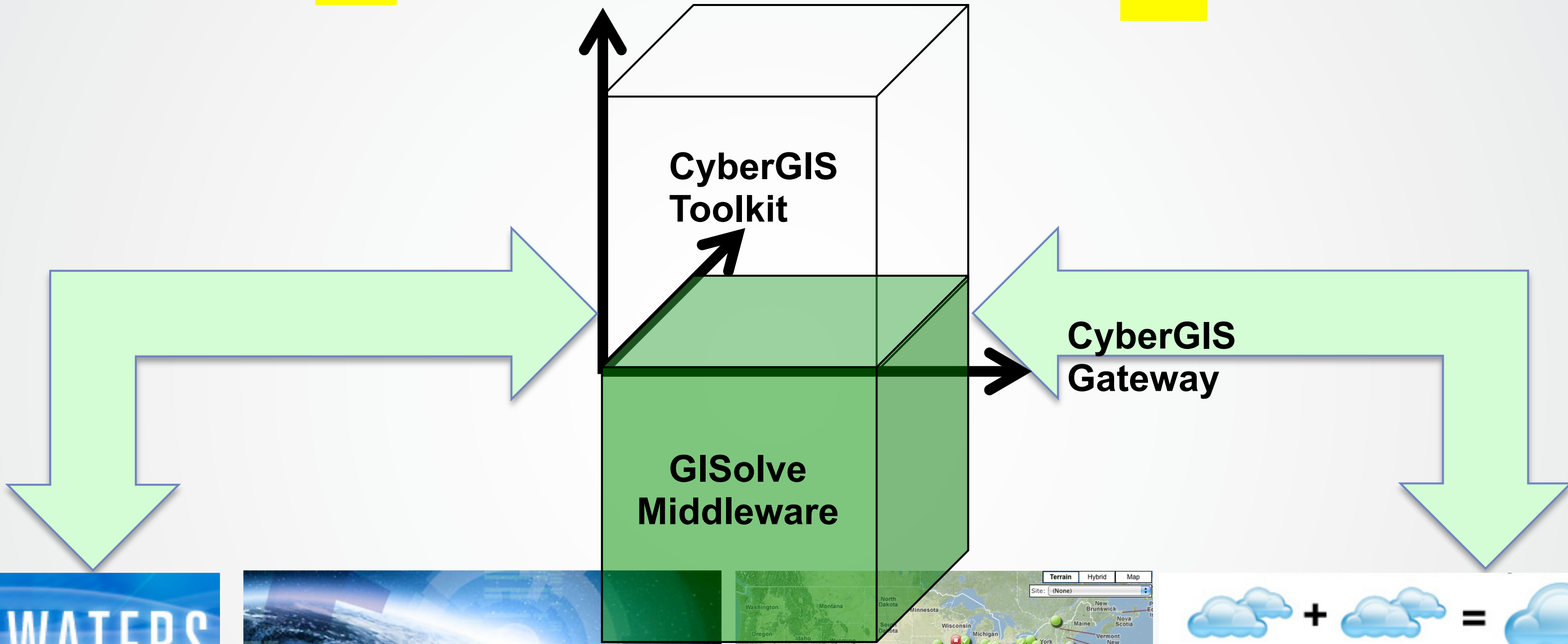


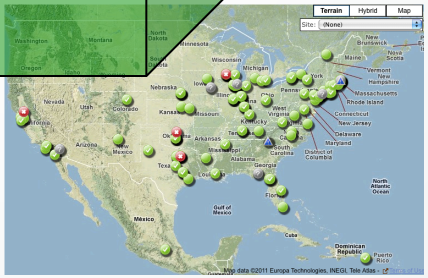
Image source:
http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services_1.pdf

CyberGIS Analytics @ Scale



www.ncsa.illinois.edu/BlueWaters/

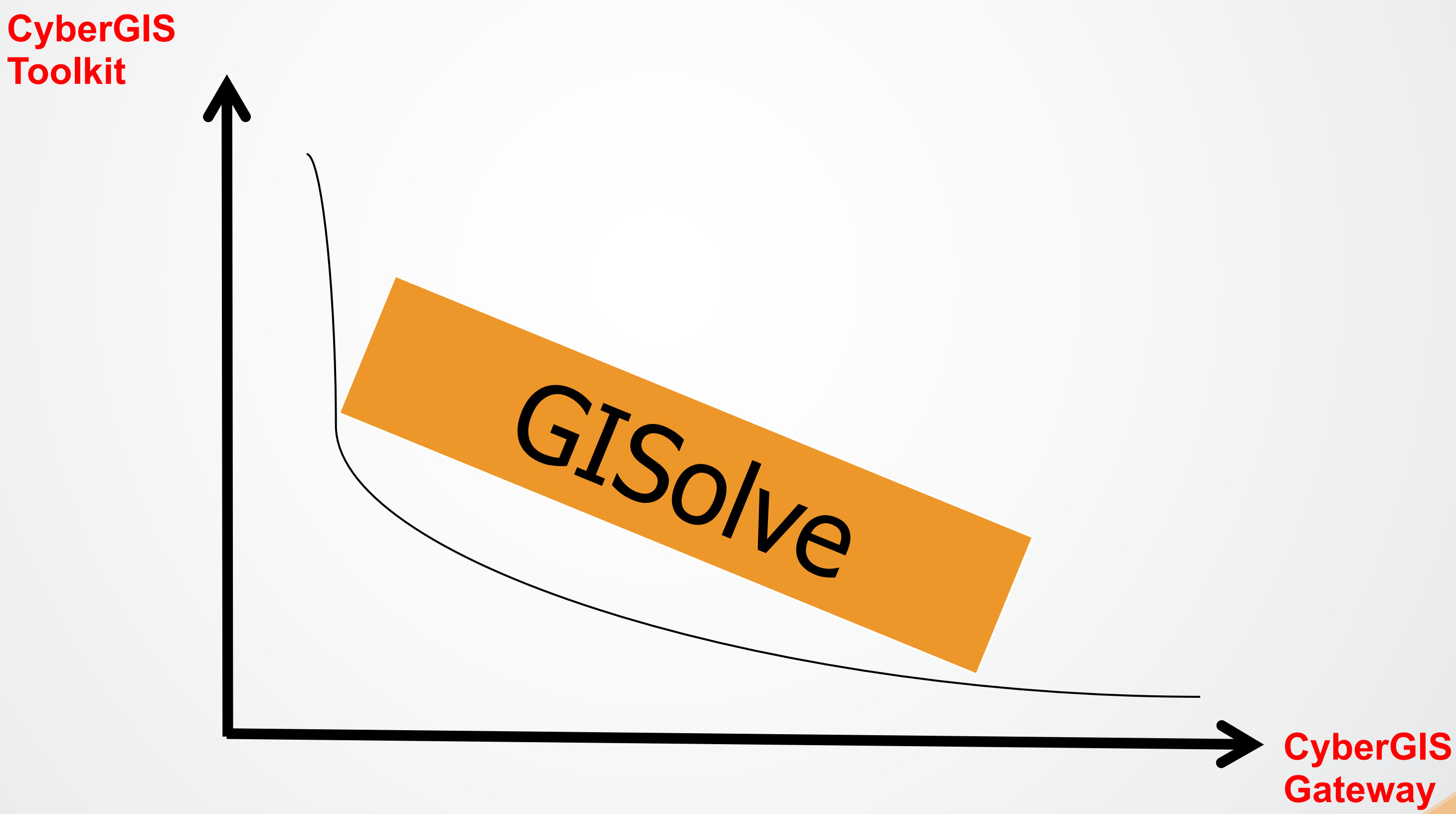
www.xsede.org



www.opensciencegrid.org

http://lakjeewa.blogspot.com/2011/09/what-is-cloud-computing.html

CyberGIS for What and Whom?



NSF SI2-SSI: CyberGIS Project
$4.43 million, Year: 2010-2015

Principal Investigator
- Shaowen Wang

Co-Principal Investigators
- Luc Anselin
- Budhendra Bhaduri
- Timothy Nyerges
- Nancy Wilkins-Diehr

Senior Personnel
- Michael Goodchild
- Sergio Rey
- Marc Snir
- E. Lynn Usery

Chair of the Science Advisory Committee
- Michael Goodchild

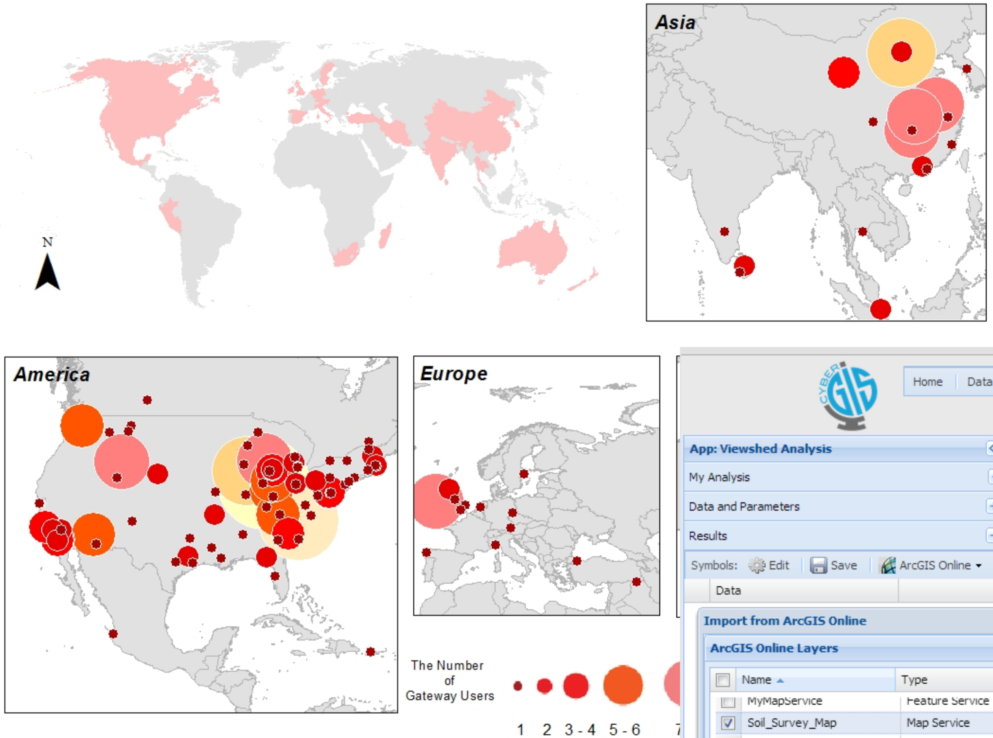
Project Manager
- Anand Padmanabhan

Project Staff
- ASU: Wenwen Li and Rob Pahle
- ORNL: Ranga Raju Vatsavai
- SDSC: Choonhan Youn
- UIUC: Yan Liu and Anand Padmanabhan
- USGS: Michael Finn and David Mattli
- Graduate and undergraduate students

Industrial Partner: Esri
- Steve Kopp and Dawn Wright

CyberGIS Communities

- **Science and Technology Communities**
  - Advanced cyberinfrastructure
  - Climate change impact assessment
  - Emergency management
  - Geographic information science
  - Geography and spatial sciences
  - Geosciences
  - Social sciences
  - Etc.
- **User Communities**
  - Biologists
  - Geographers
  - Geoscientists
  - Social scientists
  - General public
  - Broad GIS users
  - Etc.


N
Asia
America
Europe
The Number of Gateway Users
1
2
3 - 4
5 - 6


Home
Data Access
Apps
Visualization
Community
Support
About
Feedback
Logout
App: Viewshed Analysis
My Analysis
Data and Parameters
Results
Symbols:
Edit
Save
ArcGIS Online
Data
Import from ArcGIS Online
ArcGIS Online Layers
Name
Type
Description
MyMapService
Feature Service
Soil_Survey_Map
Map Service
This map shows the Soil Survey Geo...
USA_Topo_Maps
World_Imagery
for-grp-mtg
operations
operations
sids2
Add to Gateway
Soil_Survey_Map
Type: Map Service
Description: This map shows the Soil Survey Geographic (SSURGO) by the United States Department of Agriculture's Natural Resources Conservation Service. It also shows data that was developed by the National Cooperative Soil Survey and supersedes the State Soil Geographic (STATSGO) dataset published in 1994. SSURGO digitizing duplicates the original soil survey maps. This level of mapping is designed for use by landowners, townships, and county natural resource planning and management. The user should be knowledgeable of soils data and their characteristics. The smallest scale map shows the Global Soil Regions map by the United States Department of Agriculture's Natural Resources Conservation Service. For more information on this map, visit us online.
Source: Source: USDA Natural Resources Conservation Service
URL:
http://services.arcgisonline.com/ArcGIS/rest/services/Speci
Beaufort Sea
Chukchi Sea
Hudson Bay
Gulf of Mexico
Caribbean Sea
LonLat: -57.62155, 65.91701
Source: USDA Natural Resources Conservation Service
1000 km
1000 mi
XSEDE
Open Science Grid
Copyright © 2010-2012 CIGI. CyberGIS Gateway, powered by GISolve, OpenTopography, PySAL.
This website is based upon work supported by the National Science Foundation under Grant No. OCI-1047916. Any opinions, findings, and conclusions or recommendations expressed in this website are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

CyberGIS Gateway and ArcGIS Online Integration

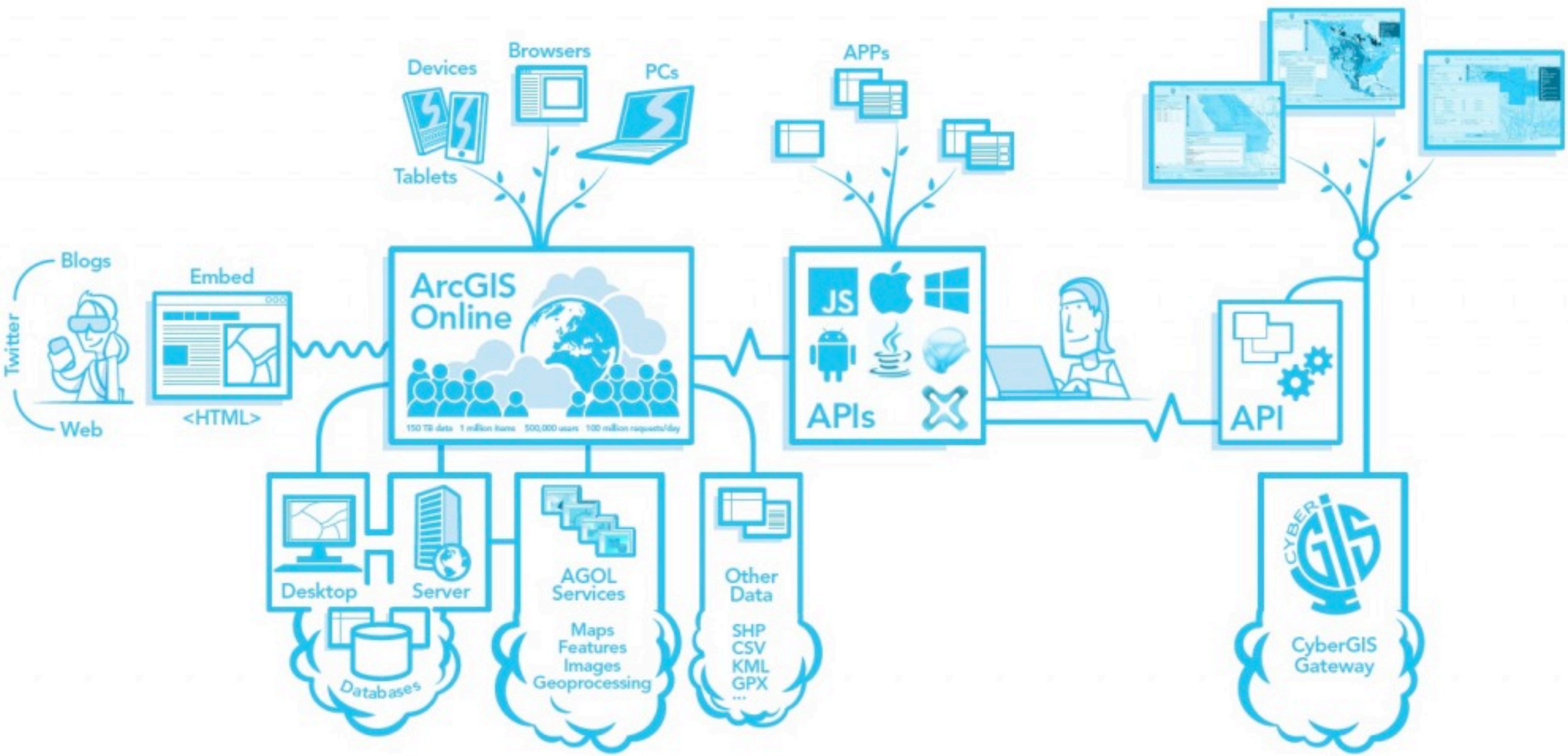
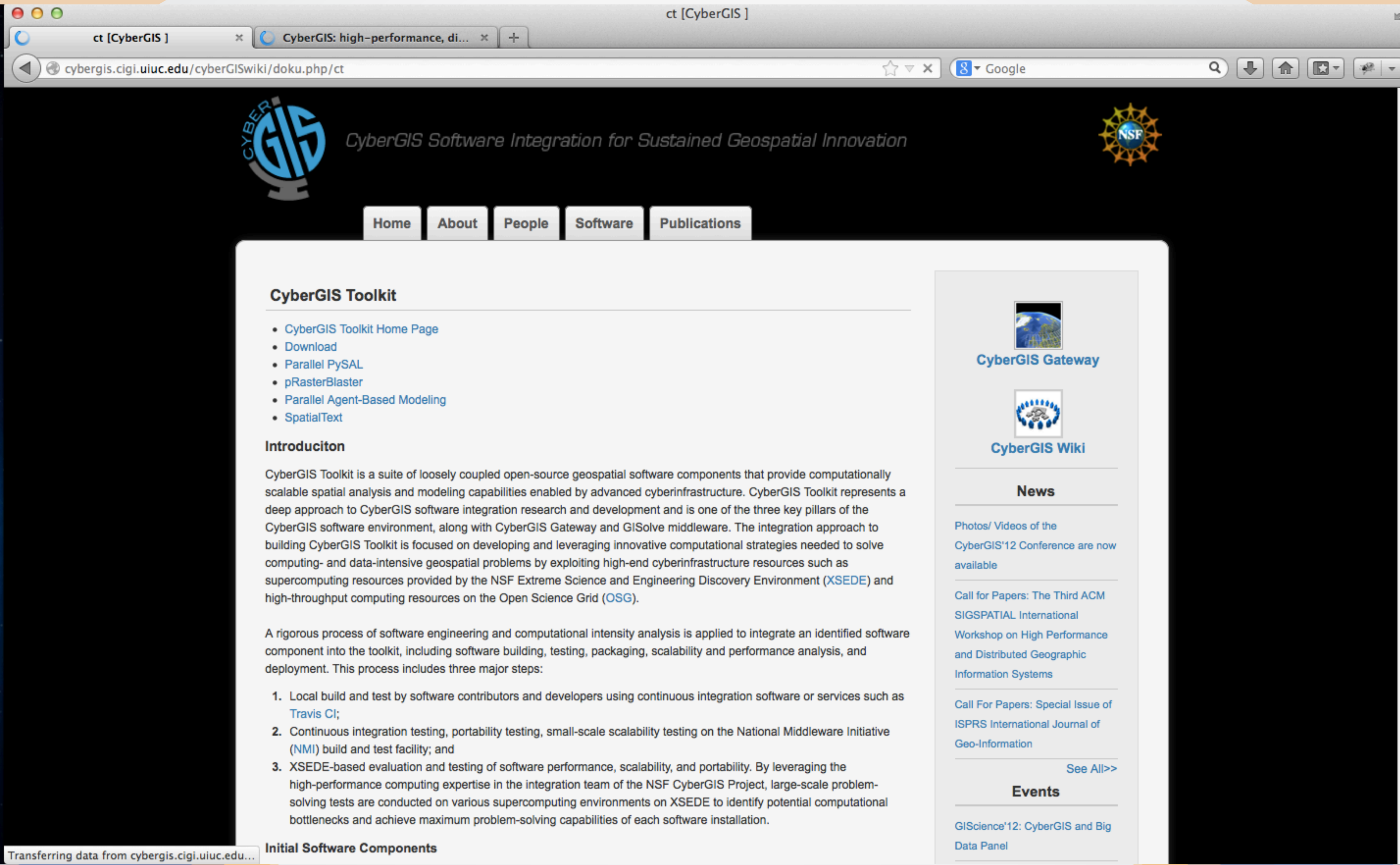
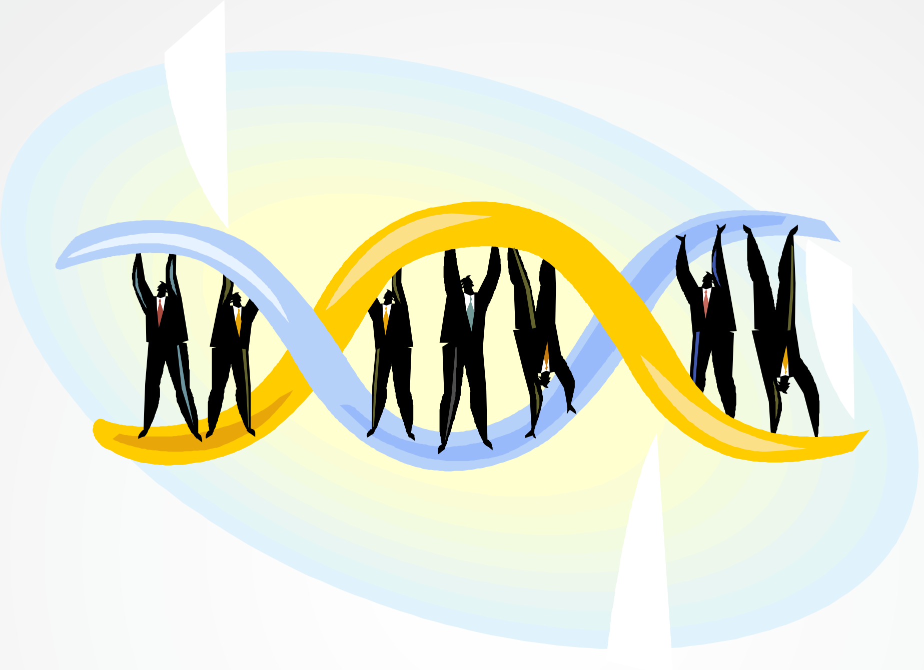


Image source: http://blogs.esri.com/esri/arcgis/2013/10/01/what-is-cybergis/

CyberInfrastructure and Geospatial Information Laboratory





- Curriculum and pedagogy
- Partnerships
- Open

Acknowledgments

- **National Science Foundation**
  - Blue Waters
  - BCS-0846655
  - OCI-1047916
  - XSEDE SES070004

Thanks!

- Comments / Questions?

- Email: shaowen@illinois.edu