An Extreme-Scale Computational Approach to Redistricting Optimization

Wendy K. Tam Cho*
Shaowen Wang†
Yan Liu‡
University of Illinois at Urbana-Champaign

1 Computational Redistricting Reform

Redistricting occurs every 10 years following the decennial census. Ideally, the resulting districts are created to provide fair representation for each citizen. In the current system, the redistricting process is easily manipulated so that future election results are essentially known even before votes are cast. Drawing district lines to favor a specific outcome is generally called gerrymandering. In a gerrymandered district, politicians essentially hand-pick their voters. It is not, as we have come to assume in a democracy—the voters choosing their representative. Gerrymanders are not looked upon favorably, and the practice of gerrymandering is obviously controversial, but regulating the practice is quite challenging. There are so many ways to draw gerrymanders within the legal constraints that the phenomenon continues scarcely impeded by regulation.

We approach redistricting from a non-legal perspective. Rather than proposing regulations intended to constrain map-making, we develop a tool to illuminate and open up the redistricting process. Although the process will always involve partisan and interested parties, a widely accessible computational tool that provides access to relevant data and enables users to explore the universe of possible redistricting plans satisfying certain criteria would engage a much broader array of interested citizens and make the process eminently more fair and transparent. In the realm of public policy, such a tool has the potential to fundamentally transform the redistricting process by providing greater flexibility and enhanced capabilities for designing redistricting plans than has ever existed. Powerful computational tools will provide a quantum leap in understanding that will be instrumental for identifying previously unseen insights into the redistricting process and its impact on democratic rule. These goals would be otherwise inconceivable.

Our project takes advantage of the massive computational power provided by the Blue Waters supercomputer for computationally intensive redistricting analysis by extending and enhancing a scalable parallel genetic algorithm (PGA) library that we have developed (Liu and Wang, 2013). Access to Blue Waters enables us to examine this problem at considerably finer spatial scales than ever before. Exploiting high-performance computing is crucial for redistricting optimization since the magnitude of the problem rises exponentially with the number of geographic units. The redistricting problem amounts to arranging a finite number of indivisible geographic units into a smaller number of aggregated areas or districts. Since every unit must belong to exactly one district, a map is a partition of the set of all units into a fixed number of non-empty district. Consider that the total number of possible maps that can be drawn by placing n units into k districts is a Stirling number of the second kind, $S(n,k)$, defined combinatorially as the number of partitions of an n-element set into k non-empty subsets. Even with a modest number of units, the scale of

*Department of Political Science; Department of Statistics; National Center for Supercomputing Applications
†Department of Geography and Geographic Information Science; Department of Computer Science; Department of Urban and Regional Planning; National Center for Supercomputing Applications
‡Department of Geography and Geographic Information Science; National Center for Supercomputing Applications

unconstrained map-making is awesome. There are 8.7×10^{39} possibilities for dividing $n = 55$ units into $k = 6$ districts. The redistricting problem in Minnesota involves partitioning 8 districts from 259,777 census blocks—an obviously astronomically larger problem.

The redistricting problem is an application of the set-partitioning problem that is known to be *NP*-complete. We define the problem as a multi-objective discrete optimization problem with a configuration of objectives and constraints that are reflective of the legal environment in which redistricting proceeds. Because an exact optimal solution is computationally intractable, we focus on developing a heuristic by combining the idiosyncrasies of the redistricting process with a genetic algorithm (GA) to produce near-optimal redistricting maps. The legal constraints of the problem reduce the solution space, and our scalable algorithms for exploiting the massive Blue Waters computing power help us implement our heuristic.

2 Algorithm and Scalability

Our PGA parallelizes the GA computation by running a large number of PGA processes simultaneously. Each process conducts independent GA computation with a migration strategy that exchanges solutions between any two directly connected PGA processes at regular intervals. On each PGA process, a set of solutions forms a local population. The mutation operator randomly selects and swaps two adjacent units that belong to adjacent districts.

The PGA approach employs an asynchronous migration strategy based on an island GA model, which effectively removes the costly global synchronization on migration operations and allows for improved overlapping of GA computation and migration communication. The scalability of this library has been tested in strong and weak scaling experiments on up to 16,384 processor cores on a cutting edge high-performance computer. Here, we extend and enhance this PGA library to scale the computation capabilities to the 724,480 integer cores available on the NSF Blue Waters supercomputer and achieve sustained petaflop performance.

The redistricting problem provides both an important substantive arena and ideal large optimization instances for Blue Waters. Working with this large size problem will shed new insights into search patterns among a large number of PGA processes and enable further research and refinement of search heuristics and computational strategies. The cyberGIS environment and computational capabilities developed thus have a mutualistic relationship with the redistricting problem, each nicely benefitting and advancing the other.

Technologically, the world has undergone an unmistakable and immense transformation. The power of information and computing has already demonstrated its extensive and often surprising reach in many realms of life. We must take advantage of these technological advances to facilitate societal tasks. We are at the threshold of making unprecedented use of statistical and mathematical modeling and computing technology in the redistricting process. Instead of merely tinkering with endless possibilities, we will be able to develop computationally intensive models to synthesize and organize massive amounts of computation/data to help us evaluate redistricting schemes and tailor them to our notions of "fairness" and democratic rule.

References

Liu, Y.Y. and S. Wang. 2013. "A Scalable Parallel Genetic Algorithm for the Generalized Assignment Problem." *Journal of Parallel Computing.* Accepted.